Available online at www.sciencedirect.com
ScienceDirect
Remote Sensing of Environment 112 (2008) 735–749
Remote Sensing of Environment
www.elsevier.com/locate/rse

Estimating proportional change in forest cover as a continuous variable from multi-year MODIS data

Daniel J. Hayes [a,*], Warren B. Cohen [b], Steven A. Sader [c], Daniel E. Irwin [d]

[a] Department of Forest Science, Oregon State University, 321 Richardson Hall, Corvallis, OR 97331 USA
[b] USDA Forest Service, PNW Research Station Forestry Sciences Laboratory, 3200 SW Jefferson Way, Corvallis, OR 97331 USA
[c] School of Forest Resources, University of Maine, 260 Nutting Hall, Orono, ME 04469 USA
[d] Global Hydrology and Climate Center, NASA - Marshall Space Flight Center, 320 Sparkman Dr., Huntsville, AL 35805 USA

Received 11 January 2007; received in revised form 4 June 2007; accepted 7 June 2007

Abstract

This article describes a series of fundamental analyses designed to test and compare the utility of various MODIS data and products for detecting land cover change over a large area of the tropics. The approach for estimating proportional forest cover change as a continuous variable was based on a reduced major axis regression model. The model relates multispectral and multi-temporal MODIS data, transformed to optimize the spectral detection of vegetation changes, to reference change data sets derived from a Landsat data record for several study sites across the Central American region. Three MODIS data sets with diverse attributes were evaluated on model consistency, prediction accuracy and practical utility in estimating change in forest cover over multiple time intervals and spatial extents.

A spectral index based on short-wave infrared information (normalized difference moisture index), calculated from half-kilometer Calibrated Radiances data sets, generally showed the best relationships with the reference data and the lowest model prediction errors at individual study areas and time intervals. However, spectral indices based on atmospherically corrected surface reflectance data, as with the Vegetation Indices and Nadir Bidirectional Reflectance Distribution Function - Adjusted Reflectance (NBAR) data sets, produced consistent model parameters and accurate forest cover change estimates when modeling over multiple time intervals. Models based on anniversary date acquisitions of the one-kilometer resolution NBAR product proved to be the most consistent and practical to implement. Linear regression models based on spectral indices that correlate with change in the brightness, greenness and wetness spectral domains of these data estimated proportional change in forest cover with less than 10% prediction error over the full spatial and temporal extent of this study.

Keywords: MODIS; Regression models; Tropical forests; Land cover and land use change

1. Introduction

Complete coverage, accurate and repeated measures of tropical forest and land cover variables at the regional scale are being increasingly relied upon for natural resource and conservation planning, habitat and biodiversity assessments, and carbon accounting for international agreements on climate change mitigation strategies. Scientists and practitioners have an over 30-year history of using Landsat data to monitor vegetation at landscape and regional scales. The Landsat data set has several advantages that account for it being so widely-used in ecological applications, including a long-running historical time-series, a spatial resolution appropriate to land cover and land use change (LCLUC) investigations, and a spectral coverage appropriate to studies of vegetation properties (Cohen & Goward, 2004). However, the low repeat frequency (with recurrent cloud coverage) and small area coverage per scene preclude its use for large-area monitoring in the tropics. The current sensor malfunctions with the Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and the unavailability of Landsat 5 Thematic Mapper (TM) data for some regions, lends further credence to considering other sources of satellite data for regional monitoring.

* Corresponding author. Current address: Department of Biology and Wildlife University of Alaska Fairbanks 215 Irving I Building Fairbanks, AK 99775 USA.

E-mail addresses: Daniel.Hayes@Oregonstate.edu, ffdjh1@uaf.edu (D.J. Hayes).

0034-4257/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2007.06.003

1.1. Monitoring tropical forest ecosystems with satellite data

Local and landscape-scale monitoring of tropical deforestation has been accomplished using complete coverage of regularly updated Landsat data (e.g. Sader et al., 2001a; Skole & Tucker, 1993). However, in tropical regions, persistent cloud cover during the rainy season and smoke and haze from biomass burning during the dry season can preclude the annual or biennial acquisition of a clear scene of Landsat data for the same area (Asner, 2001). This problem, along with monetary cost, data storage, and processing time limitations, is exacerbated when multiple scenes are required to cover larger areas on regional and global scales. While a sampling of high resolution data can serve to validate mapping efforts and provide insight into processes such as deforestation over a region (FAO, 1993; Richards et al., 2000), these efforts have critical limitations, including the misrepresentation of rates and extent of change because deforestation is often spatially clustered (Fearnside, 1986; Tucker & Townshend, 2000). Techniques need to be developed for extracting LCLUC data from coarse resolution satellite data sets, which generally have the advantages of larger area coverage, greater frequency of acquisitions, and lower cost.

Estimates of forest cover and land cover change have been carried out over full spatial coverage of tropical regions using coarse resolution data (1 km or greater) from the Advanced Very High Resolution Radiometer (AVHRR) (DeFries et al., 2000; Malingreau & Tucker, 1988; Mayaux et al., 1998). Although previous studies have demonstrated the utility of multi-temporal AVHRR data for identifying deforestation 'hot spots' or tropical forest clearing at broad spatial scales along forest/agriculture boundaries (Lambin & Ehrlich, 1997), finer-scale changes associated with anthropogenic LCLUC are not directly detectable at this resolution (Ehrlich et al., 1997). Many of these land cover changes due to human activities occur at spatial scales near 250 m or less (Cohen et al., 2003a; Hayes et al., 2002; Townshend & Justice, 1988). With its improved radiometric properties, higher spatial and spectral resolutions and twice-daily coverage, the freely-available suite of products generated from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) hold the promise of becoming the most important data source for future land cover characterization and operational monitoring at regional and global scales (Justice et al., 1998; Townshend & Justice, 2002).

1.2. Land cover change detection with MODIS data

Despite the potential for monitoring LCLUC at moderate resolution (250 and 500 m) with MODIS data, the Vegetative Cover Conversion product is designed to serve only as a global 'alarm' system for prioritizing closer inspection of land cover change (Zhan et al., 2000). Prototype studies of change algorithms using 250 m MODIS radiance data suggest satisfactory results for extreme events such as large-scale flooding and burning, yet the detection of deforestation was less reliable where forest clearing is patchy (Zhan et al., 2002). Fewer studies have attempted to characterize forest regeneration with coarse resolution data, and those based on AVHRR data have shown similar limitations where the spectral signal of forests regenerating on small, fragmented clearings was mixed with other vegetation and land cover types within the same 1.1 km pixel (Lucas et al., 2000a,b). Research is needed to develop methodologies and models for extracting accurate information on both forest clearing and vegetation regeneration from coarse spatial resolution data.

The majority of studies reporting on the use of satellite imagery to map changes in land cover have employed traditional classification algorithms that apply a single, discrete change/no change category to each pixel in a multi-temporal data set (see Coppin et al., 2004). Other techniques, such as linear mixture models that estimate the sub-pixel composition land cover features (e.g. Adams et al., 1995; DeFries et al., 2000) and regression techniques for modeling biophysical properties as continuous variables (e.g. Cohen et al., 2003b), have been designed to more fully exploit the information about the variability of the features of interest inherent in the spectral signal of remotely sensed data. These techniques become particularly important when the processes of interest operate at scales below the resolution of the sensor, as will often be the case when studying land cover changes with more coarse resolution imagery such as that from the MODIS data set. Because of the advantages involved with mapping sub-pixel cover, Hansen et al. (2002) speculated that the MODIS-based Vegetation Continuous Fields products could be used to measure changes in proportional forest cover over time, yet few studies have evaluated either this concept or those data in a change detection context.

Hayes and Cohen (2007) reported on the development of a modeling framework to estimate proportional change in tropical forest cover as a continuous variable from multi-temporal spectral data sets. With the ability to detect proportional forest cover change relatively invariant to the spatial grain size of the analysis, their results demonstrated the utility of the half-kilometer MODIS data, which contains short-wave infrared reflectance — information that when included in these models improved the ability to estimate forest cover change. The results showed that spectral response in the Calibrated Radiances Swath data set followed more closely with the expected patterns of forest cover change, as compared to the reference data, than did the spectral response in the gridded Surface Reflectance product. This result may be attributed to the geolocation offset problem encountered in gridded MODIS products described by Tan et al. (2006), which they found to cause "pixel shift" in the MODIS data when compared against higher resolution reference data. These authors observed the problem to be most prevalent with the finer resolution MODIS products (quarter- and half-kilometer) and suggested that, to improve the correspondence between the location of reference data observations and the grid cells, data should be aggregated to coarser resolutions (i.e. 1 km or greater).

The research described above has shown the potential of models based on these data for estimating forest cover change at the landscape scale for a give time interval. The development of these models in pilot studies for specific regions and time periods also highlighted some of the challenges and limitations in using these data sets, including those related to spatial resolution, geometric precision, atmospheric and seasonal normalization and spectral variable selection (Hayes & Cohen, 2007). Further research is needed to test the utility of these models for application in a larger spatial and temporal context.

1.3. Objectives

Attributes of the product stream generated by the MODIS sensor suggest that it is a key data set for monitoring large areas at high temporal frequency and low cost. This study was designed to resolve some of the issues involved in expanding a stand-alone, MODIS-based forest monitoring system, based on these models, over larger spatial extents and over time. The primary goal of this study was to develop an accurate, consistent, and efficient methodology for creating detailed, full coverage estimates of forest cover change across a large tropical study site. Specifically, the objectives of this research were:

1) Model evaluation: compare key data sets within the MODIS data stream for their utility in predicting tropical forest cover change over space and time through evaluation of single, individual spectral index models at varying spatial grain size and processing level; and
2) Model selection: identify a regression model, based on multiple spectral indices, which optimizes the detection of forest cover change and allows for consistent, practical implementation over large areas and multiple time periods.

2. Methods

The Mesoamerican region covers approximately 768,990 km^2, stretching from the southern Yucatán Peninsula of Mexico through Panamá. The Isthmus of Central America is comprised of more than 20 distinct ecological life zones (Holdridge et al., 1971) ranging from coral reefs, coastal wetlands, and Atlantic lowland wet forests to pine savannas, Pacific dry forests, and montane cloud forests. Much of the region experiences strong seasonality, with a distinct dry season from December through April. The strength of this seasonality generally follows a latitudinal gradient, with more extreme dry seasons and seasonally semi-deciduous forest phenology found at the northern end of this region. Dry season strength, along with associated forest

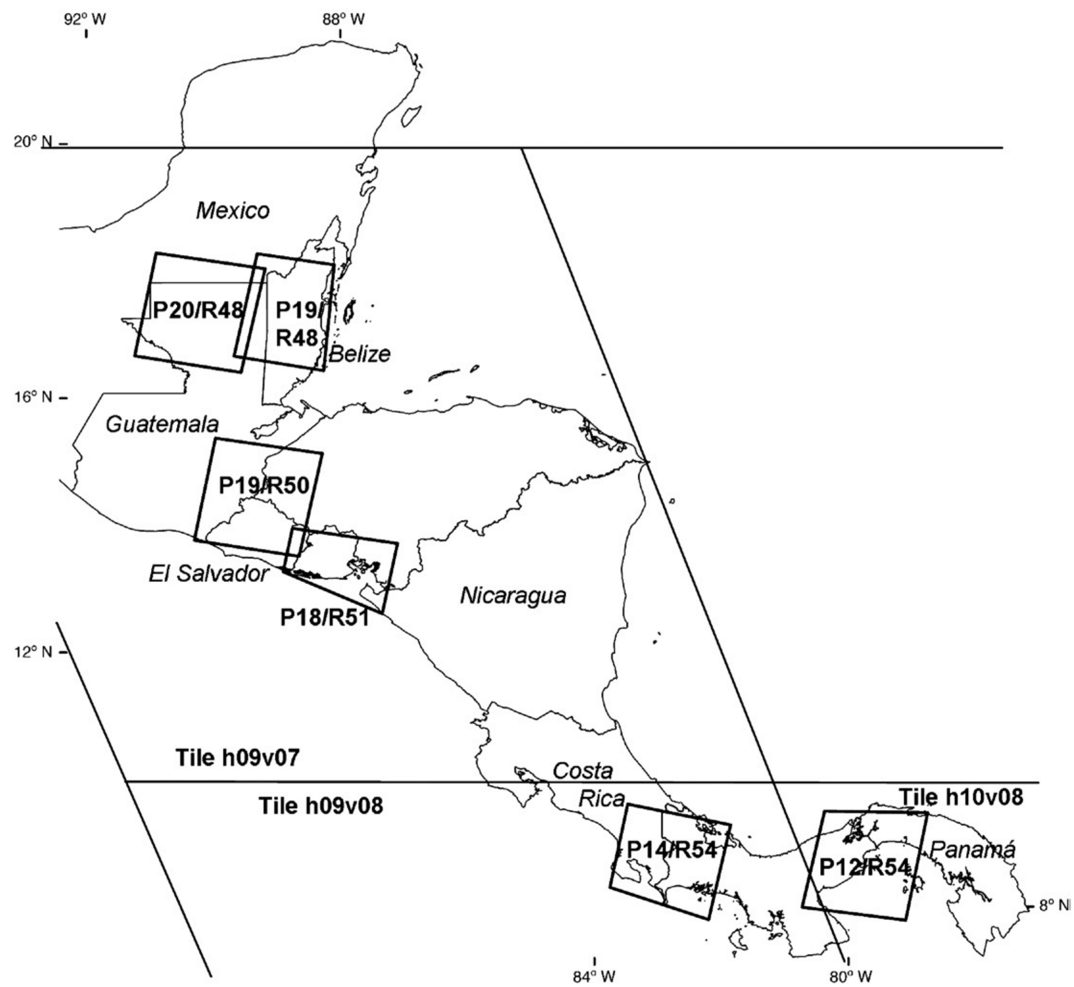

Fig. 1. Map illustrating the location of the MODIS grid tiles and the reference forest cover change study sites (based on Landsat WRS scenes) in relation the boundaries of the nations of the Central American region.

fires, is accentuated during strong El Niño affected climate conditions. The Mesoamerican land bridge connecting the North and South American continents contains only one-half of a percent of the world's total land surface, yet its geographic position and variety of ecosystems combine to harbor about 7% of the planet's biological diversity (Miller et al., 2001). However, the rates of deforestation in Central America during the 1980's were among the highest in the world (FAO, 1993). Rapid population growth, human migration, and slash-and-burn agriculture have had detrimental effects on what forest remains as high rates of deforestation have continued throughout the region in the 1990's (Sader et al., 2001b).

The calibration and testing of regional scale models of forest cover change with coarse resolution MODIS data is based on a series of high resolution land cover change databases developed from multi-temporal Landsat data for 6 selected scenes across Central America (Fig. 1). The analyses described in this study span a three-year time interval (2000 to 2003) in which MODIS data could be acquired concurrently with Landsat 5 and pre-scan line corrector (SLC)-off Landsat 7 data. The land cover change is based on the 6 Landsat World-wide Reference System (WRS)-based study sites and 6 time intervals (2000 to 2001, 2000 to 2002, 2000 to 2003, 2001 to 2002, 2001 to 2003, and 2002 to 2003), for a combined total of 11 reference scene — time interval pairs (Table 1). The loss of forest cover observed over the study areas during the time period of interest most commonly included forest harvest, slash-and-burn agricultural and pasture establishment, forest burning, insect defoliation and conversion of mangrove forest to aquaculture. These land cover changes are manifested in spectral changes in the multi-temporal satellite data record, which are also influenced by the interannual variation in land cover condition resulting from seasonal and climatic effects (Hayes & Cohen, 2007).

2.1. Reference forest cover change classifications

The reference forest cover change data sets were developed by multivariate clustering and interpretation of vegetation index composites created from multi-date Landsat imagery, aggregated to coarse resolution continuous variables matching the grid cell sizes of the MODIS data and products. Two-date Landsat images were paired and classified for each WRS study site and time interval pair used in this study (Table 1). At each site location, the Landsat GeoCover image acquisition was used as the "base-line" for both geometric correction and radiometric normalization of the corresponding "subject" Landsat image. The subject images were geo-rectified to the base-line via the automated selection of image tie-points, as described in Kennedy and Cohen (2003), and first-order polynomial resampling. The geometrically corrected Landsat images retained the original 30 m pixel resolution projected in

Table 1
Acquisition dates (given as Julian days) for the images used in this study

Study site		Year 2000		Year 2001		Year 2002		Year 2003	
Column	Row	Sensor	Date	Sensor	Date	Sensor	Date	Sensor	Date
P12	R54	TM	2000081			ETM+	2002148[d]		
–[a]	–	MOD02	2000084			MOD02	2002148		
v08	h10	MOD13	2000081[b]			MOD13	2002145		
v08	h10	MOD43	2000081[b]			MOD43	2002145		
P14	R54	ETM+	2000045[d]	ETM+	2001031				
–	–	MOD02	2000082	MOD02	2001031				
v08	h09	MOD13	2000081	MOD13	2001017				
v08	h09	MOD43	2000081	MOD43	2001017				
P18	R51			ETM+	2000361[c]	ETM+	2002126[d]		
–	–			MOD02	2000361	MOD02	2002126		
v07	h09			MOD13	2000353	MOD13	2002113		
v07	h09			MOD43	2000353	MOD43	2002113		
P19	R48	TM	2000088			ETM+	2002261[d]		
–	–	MOD02	2000087			MOD02	2002261		
v07	h09	MOD13	2000081			MOD13	2002273		
		MOD43	2000081			MOD43	2002273		
P19	R50			TM	2001090	ETM+	2002213[d,e]	ETM+	2003104
–	–			MOD02	2001087			MOD02	2003104
v07	h09			MOD13	2001081			MOD13	2003097
v07	h09			MOD43	2001081			MOD43	2004097
P20	R48	ETM+	2000087[d]	TM	2001081	ETM+	2002076	ETM+	2003127
–	–	MOD02	2000087	MOD02	2001089	MOD02	2002071	MOD02	2003127
v07	h09	MOD13	2000081	MOD13	2001081	MOD13	2002081	MOD13	2003065
		MOD43	2000081	MOD43	2001081	MOD43	2002081	MOD43	2003065

[a]Swath data coverage is variable and not acquired in predefined tiles.
[b]MOD13 and MOD43 were acquired for the same 16-day composite period in each analysis.
[c]The first date of reference data for P18 R51 was acquired in the calendar year 2000, but analyzed as 2001 land cover.
[d]Acquired from NASA's GeoCover Data Set (Tucker et al., 2004).
[e]GeoCover data used as the geometric/radiometric base-line, but not in change detection.

UTM (zone 16 or 17). Overall root mean square error (RMSE) was less than half a pixel (<15 m) for each resampled image. Digital numbers were converted to reflectance values in the reference image using the COST model (Chavez, 1996). The subsequent Landsat scenes were radiometrically normalized, band-by-band, to the reference scene using linear regression techniques (Hall et al., 1991). The algorithm was based on the selection of end-point targets from the wet and dry non-vegetated extremes in each band at each date, as outlined in Hayes and Sader (2001).

The atmospherically-corrected Landsat reflectance data were transformed to the tasseled cap indices at each date using coefficients derived by Crist (1985) for ground-based spectral data, as discussed by Cohen et al. (2003b). For each image pair (each study site and time interval), tasseled cap brightness, greenness and wetness for each date were combined into a two-date, 6-layer data set that was transformed to spectrally similar clusters via the ISODATA routine (ERDAS, 2005). Final reference LCLUC classifications were arrived at through a combination of cloud, smoke and haze removal, "cluster-busting" (Jensen, 1996), visual interpretation, comparison with ancillary data, and hand-editing. Hayes and Sader (2001) reported an overall accuracy of 85% (Kappa = 0.83) across all change categories using a similar methodology, based on clustering of multi-date normalized difference vegetation index (NDVI) composite imagery, for mapping forest clearing and regeneration in a tropical study area.

Each classification resulted in four LCLUC categories: 1) unchanged forest, 2) unchanged non-forest, 3) loss of forest cover, and 4) regeneration of forest cover. For this study, both mature forest types and young forests (≥1 year-old) are considered forest cover, and non-forest includes agriculture and pasture lands, natural savannas and grasslands, and low shrubs and wetlands. Forest regeneration within each two-date time interval was identified on fallow plots, abandoned agricultural areas, and post-harvest and burned sites. The Landsat-derived reference LCLUC was aggregated to a continuous variable representing the net proportional change in forest cover (ΔFC) to match the different Sinusoidal Projection Swath and Grid cell resolutions of the MODIS data and products, as described in Hayes and Cohen (2007).

2.2. MODIS spectral data and products

Amongst the suite of products generated by the MODIS data stream, this study considered the daily Calibrated Radiances (MOD02) data and the 16-day composite Vegetation Indices product (MOD13) at half- and one-kilometer spatial resolution (HKM and 1 km, respectively), as well as the Nadir Bidirectional Reflectance Distribution Function (BRDF)-Adjusted Reflectance (NBAR) product (MOD43), which is available at 1 km resolution and 16-day composites. Collection 4 Terra MODIS data and products were used in these analyses and acquired from the Earth Observing System Data Gateway of the USGS Land Processes Distributed Active Archive Center. These three different MODIS data sets vary in spectral properties, temporal compositing and spatial resolution, allowing for analysis of these effects on the ability to model change in forest cover. The daily or composite dates of the MODIS data sets were timed to coincide with the acquisition dates of the Landsat imagery used to generate the reference data (Table 1).

The MOD02 data consist of at-satellite calibrated radiance values, prior to any atmospheric correction algorithm, and are available as the original Swath observations with 250 m, 500 m and 1 km cell sizes. Radiance values are collected in the red and near-infrared (NIR) wavelengths (channels 1 and 2) at 250 m, as well as blue, green, and three short-wave infrared (SWIR) bands (channels 3 through 7) at 500 m resolution. In this study, two MOD02 data sets were evaluated: 1) MOD02 HKM in which channels 1 and 2 are resampled to 500 m to match the other 5 bands, and 2) MOD02 1 km in which all 7 bands have been resampled to 1 km cells. The MODIS Reprojection Tool for Swath data (MRT Swath) required latitude and longitude data for geolocation of MOD02, which is available as part of the MODIS data stream (MOD03) at 1 km resolution. MOD02 data was output from the MRT Swath in the Sinusoidal projection to maintain consistency with the other MODIS products used here, but retained the original 500 m and 1 km cell sizes.

The MOD13 product includes two vegetation indices calculated from MODIS surface reflectance data: the NDVI and the Enhanced Vegetation Index (EVI), which is a modification of the NDVI to adjust for soil reflectance and other background surface properties (Huete et al., 2002). The MOD13 algorithm applies a filter that computes the indices from the highest quality surface observation (based on cloud cover, aerosols and viewing geometry) over each 16 day period. This study included these products at half- and one-kilometer spatial resolution (MOD13 HKM and MOD13 1 km, respectively). The MOD43 NBAR product is the result of the BRDF algorithm, which is computed for the MODIS spectral reflectance in each of the bands 1 through 7 at the mean solar zenith angle of each 16 day period, designed to remove the view angle effects of the original Swath observations (Schaaf et al., 2002). Unlike the MOD02 Swath observations, the MOD13 and MOD43 products are resampled to the MODIS Global Sinusoidal Grid projection system. The grid has actual cell sizes of 463.31 m and 926.63 m at the half- and one-kilometer resolutions, respectively.

2.3. Model development

Models were constructed using a multiple linear regression approach to predict proportional change in forest cover (i.e. ΔFC) between two dates as a continuous variable from each multi-temporal MODIS spectral data set. Spectral indices (considered here as independent variables) were derived from two dates of coarse resolution MODIS data and used to model ΔFC (the dependent variable), which is based on reference data from the higher resolution Landsat LCLUC classifications for the same time interval. The model development methodology, employed here across the Central American region, follows that detailed by Hayes and Cohen (2007) for a study site in northern Guatemala.

2.3.1. The dependent variable

The final reference ΔFC data sets were created by overlaying a grid matching the location and size of the MODIS pixels on each 30 m LCLUC classification, which had been re-projected to Sinusoidal to match the MODIS data geographically. A count of the relative number of forest loss, regeneration, and no change pixels was used to calculate the reference ΔFC, which has a possible range of −100% (total loss of cover on completely forested land) to +100% (complete regeneration of forest cover on previously non-forest land) for each grid cell. Reference ΔFC data sets were calculated according to the grid cells matching the MODIS spectral data set or product being investigated, so that separate ΔFC data were individually created using the Swath cells (for the MOD02) as well as the Sinusoidal Grid cells (for the MOD13 and MOD43) at both half- and one-kilometer resolution.

2.3.2. The independent variables

Several different spectral indices derived from the three MODIS data sets were evaluated in the model exercises for their utility and accuracy in estimating ΔFC. The NDVI was used because of its common application in vegetation studies and its ease in calculation from the spectral data available in all three data sets at both half- and one-kilometer resolution. For the MOD02 and MOD43 data, the NDVI was calculated as:

$$\mathrm{NDVI} = \frac{(\mathrm{Band2} - \mathrm{Band1})}{(\mathrm{Band2} + \mathrm{Band1})} \tag{1}$$

The NDVI is already calculated and available as part of the MOD13 data set. Similarly, it was also possible to calculate the normalized difference moisture index (NDMI) for all the data sets. It was derived by the equation:

$$\mathrm{NDMI} = \frac{(\mathrm{Band2} - \mathrm{Band6})}{(\mathrm{Band2} + \mathrm{Band6})} \tag{2}$$

for the MOD02 and MOD43 data. The bands labeled “NIR” (near-infrared) and “MIR” (mid-infrared) were used to calculate the NDMI from the MOD13 data set, in place of Band 2 and Band 6, respectively. Studies by Wilson and Sader (2002) and Hayes and Cohen (2007) have demonstrated improved results in detecting forest cover changes with the NDMI index over the NDVI.

Several other spectral indices were employed in the modeling effort: the EVI from the MOD13 product and the Tasseled Cap brightness (TCB), greenness (TCG), and wetness (TCW) indices from the MOD43 data set. The EVI has been shown to be more responsive to vegetation canopy structure and is less subject to saturation in high biomass environments than the original NDVI (Huete et al., 2002). The Tasseled Cap indices were calculated for the MOD43 data set according to the transformation coefficients developed by Lobser and Cohen (in press). The TCW index has been shown to be sensitive to the moisture content and structural characteristics of vegetation (Cohen & Spies, 1992), as well as a good index for detecting changes in forest cover (e.g. Healey et al., 2005). Currently, Tasseled Cap transformation coefficients have not been developed for the other MODIS data sets. For comparison, surrogate variables of the EVI and Tasseled Cap indices were calculated from the MOD43 spectral bands in the model selection process. The EVI is essentially a standard measure of greenness that attempts to remove background brightness by subtracting reflectance in the blue band from the NDVI. This was simulated with the MOD43 bands by including a change in band 3 index (ΔBand3) along with ΔNDVI in step-wise regression analyses. Similarly, MOD43-derived indices ΔBand3, ΔNDVI, and ΔNDMI represent the changes in the “brightness”, “greenness” and “wetness” spectral domains of these data, respectively, and thus can be compared with the change in Tasseled Cap indices for the ability to estimate ΔFC.

2.3.3. Model structure

The relationship between reference ΔFC and the change in spectral response between two dates was investigated for each data set, spectral index, and time interval using a Reduced Major Axis (RMA) approach, a class of orthogonal regression models (Curran & Hay, 1986; Van Huffel, 1997). The use of RMA in the modeling of biophysical variables with remotely sensed data is discussed by Cohen et al. (2003b) and demonstrated in a change detection context by Hayes and Cohen (2007) for a study site in Central America. The regression models evaluated here used the approach of this latter study, in which models are constructed based on stratified random samples of spatially independent grid cells for each study site and time interval. In this study, a simple thresholding of the spectral change index values from the set of the sample cells is used so that the two model sets (forest regeneration and forest clearing) could be separated easily and consistently for expanded ΔFC modeling in the absence of detailed initial reference data. For instance, any particular sample was included in the forest regeneration model set if its spectral change index value was greater than the mean of that index for all samples, or modeled as forest loss if it had a value less than the sample set mean.

The number of samples used in constructing the models varied by spatial resolution, with the higher resolution data (HKM) having more potential sample cells, as well as by the data set. The MOD13-based models had lower number of samples used because only the “best” data were used in model construction, as per the quality assurance (QA) files available with this product. The same criteria were used in sample selection for the models using MOD43 data, although the QA criteria for this product differ from that of the MOD13, primarily because of the differences in the compositing algorithms. As a result, the MOD43 data tended to have fewer poor quality pixels masked from analysis. The MOD02 data are single-day acquisitions and not composited, and therefore do not contain QA information that can be readily used to remove data from the potential data set. This was done manually, through unsupervised clustering and hand-editing for cloud removal (see Hayes & Cohen, 2007). The MOD02 models generally had the highest sample size as a result.

2.4. Model evaluation

The model evaluation process (Table 2) included four steps of analysis: 1) individual study areas and time intervals were analyzed *by scene–by interval*; 2) multiple study areas were combined and analyzed *across scenes–by interval*; 3) multiple time intervals were combined and analyzed *by scene–across intervals*; and 4) multiple scenes and multiple time intervals were combined and analyzed *across scenes–across intervals*. The step 1 analysis involved all 11 Landsat scene and time interval pairs available in the reference data used in this study, with results summarized using each pair as an analysis unit. While the step 1 analysis involved developing a unique model for each individual scene/interval pair, the analyses in steps 2, 3 and 4 were based on models developed from pooled sample points generated from scene and interval pairs organized by spatial and temporal coverage. Step 2 used scenes (P12 R54, P19 R48, P20 R48) that were available for the same time interval (2000 to 2002), while step 3 evaluated models from a single scene (P20 R48) across 3 time intervals (2000 to 2001, 2001 to 2002 and 2002 to 2003). Step 4 used the full set of 11 scene and time interval pairs as with the step 1 analysis but compared models built from the pooled sample points rather than summarizing the site-by-site results.

The first step in model evaluation was to analyze the model predictions and test statistics amongst the different data sets and spectral indices on an individual basis, for each study site and time interval. There were 11 individual reference data sets, by scene–by interval, which served as sample units for summarizing model test statistic results among the various data sets and spectral indices considered in this study. In addition to assessing their accuracy in predicting ΔFC at individual study sites, the different MODIS data sets and spectral indices were also evaluated in terms of their potential for expanded modeling over larger spatial extents. Samples from reference scenes that covered the same time interval were grouped into a new, multiple scene sample set for which the data sets and spectral indices could be evaluated across scenes–by interval. For this study, the time interval of 2000 to 2002 had three reference scenes available: 12/54, 19/48 and 20/48.

The MODIS data sets and spectral indices were further evaluated for their consistency and accuracy in model predictions over multiple time intervals. All of the reference sample data for the three time intervals available for the 20/48 study area (2000 to 2001, 2001 to 2002 and 2002 to 2003) were combined for the by scene–across intervals analysis. Ultimately, the combined effect of expanding the models over both time and space was investigated by developing and testing models based on the MODIS data sets and indices using a pooled reference data set from all of the study sites and time intervals and analyzed across scenes–across intervals.

Table 2
The Landset-based reference data used in this study, organized by the 4 steps of analysis that make up the "Model Evaluation" of the MODIS data sets and spectral indices, described in Section 2.4

Analysis step	Analysis units	Landset scene(s)	Time interval(s)	Results presented in…
Step 1: by scene–by interval		*11 scene/interval pairs from 2000 to 2003*		Table 3
	Landsat scene and time interval pairs ($n=11$)	P12 R54	2000–02	
		P14 R54	2000–01	
		P18 R51	2001–02	
		P19 R48	2000–02	
		P19 R50	2001–03	
			2000–01, 2000–02, 2000–03	
		P20 R48	2001–02, 2001–03, 2002–03	
Step 2: across scenes–by interval		*3 scene/interval pairs from 2000 to 2002*		Table 4
	Pooled MODIS-derived sample points (n varies by data set)	P12 R54	2000–02	
		P19 R48	2000–02	
		P20 R48	2000–02	
Step 3: By Scene– Across Intervals		*3 Scene / Interval Pairs from 2000 to 2003*		Table 5
	Pooled MODIS-derived Sample Points (n varies by data set)	P20 R48	2000–01	
		P20 R48	2001–02	
		P20 R48	2002–03	
Step 4: across scenes–across intervals[a]		*3 scene/interval pairs from 2000 to 2003*		Fig. 2
	Pooled MODIS-derived Sample Points (n varies by data set)	P12 R54	2000–02	
		P14 R54	2000–01	
		P18 R51	2001–02	
		P19 R48	2000–02	
		P19 R50	2001–03	
			2000–01, 2000–02, 2000–03	
		P20 R48	2001–02, 2001–03, 2002–03	

[a] The reference sample points produced in Step 4 were also used in the "Model Evaluation" process, described in Section 2.5 and the results of which are presented in Table 7 and Fig. 3 ($n=1208$ based on the pooled MOD43 sample points).

At each step, models (based on each data set and spectral index) were compared on the basis of the variability in their parameters, the accuracy in separating forest cover loss from regeneration by thresholding the spectral index (threshold accuracy %), the correlation (R) with ΔFC for each the forest loss and forest regeneration model sets, and the amount of variation explained (the coefficient of determination, R^2) by the spectral change index. Model results were compared based on % error, the error in the model predictions as a percent of the range of observed ΔFC (see Hayes & Cohen, 2007). Percent error was based on the RMSE of the prediction versus observed data set, calculated using cross-validation of the sample set (see Cohen et al., 2003b), divided by the range of values in observed ΔFC for that sample set.

2.5. Model selection

After evaluating the models based on individual spectral indices for their consistency and accuracy in modeling ΔFC over time and space, the next step was to develop a regression model that incorporated multiple spectral indices. The data set used in this analysis was comprised of anniversary dates of MOD43 acquisitions, corresponding to the 16-day composite period ending on the 81st day of each year in this analysis. These models optimized the relationship to ΔFC and the prediction accuracy while simultaneously demonstrating a consistent and practical methodology for future ΔFC modeling in the absence of detailed reference data. Model selection and development involved a combination of intentional variable selection, based on the relationships of the different indices and ΔFC evaluated here, and step-wise regression to select additional significant variables for modeling ΔFC. A variable set including the spectral change for the 7 MOD43 bands, the NDVI, NDMI, TCB, TCG, and TCW indices was used in the step-wise regression analysis based on Bayesian Information Criteria. The different RMA multiple linear regression models were evaluated according to the standard errors of the parameter estimates, model R^2, and the percent error of the predictions, as compared to the pooled reference sample data sets covering all of the study areas and time intervals (across scenes–across intervals).

3. Results and discussion

In addition to 36 channels of spectral information at varying spatial resolution, there are numerous derived products designed for specific applications that are available from the MODIS data stream. A subset of this information relevant to forest and land cover characterization was considered in this study as potential primary data sources for a tropical forest cover monitoring scheme that could be applied over large areas and multiple time periods. These data were evaluated based on regression model statistics testing their ability to detect proportional change in forest cover, including both clearing and regeneration. These statistics were based on comparing variable correlation and model predictions against forest cover change reference data developed by well-established classification methods using higher resolution Landsat data. These classifications do have error, however, and so are used as here as “reference” rather than being considered “truth”. As such, any discussion of improvements in the test statistics from one data set to another, especially marginal ones (e.g. 1 to 3% prediction accuracy), must be considered within this context. Consequently, the model selection process used here gives at least equal weight to other important attributes of these data sets, including the practical aspects of their implementation, as to the test statistic comparisons alone.

3.1. Model evaluation

3.1.1. By scene–by interval analysis

The results of the model analysis based on the individual study sites and time intervals helped to evaluate the performance of the different data sets and spectral indices in terms of estimating ΔFC on a scene-by-scene basis. The different data sets, spatial resolutions and individual spectral indices were evaluated by reporting the averages of model test statistics across the 11 sample units (Table 3). Of the MODIS data and products evaluated here, models based on the ΔNDMI index gave good results for the MOD02 and MOD43 data, and the ΔEVI for the MOD13 data. On average, the better model test statistics among the different tests on a scene-by-scene basis

Table 3
Results of the by scene–by interval analysis showing the mean RMA regression model test statistics for each data set, summarized for all scenes and time intervals analyzed in this study

Data set	Variable	n	Threshold accuracy %	Regeneration set R	Clearing set R	Combined model R^2	Model error
MOD02 HKM	ΔNDVI	11	91.8%	0.67	0.69	0.57	10.0%
	ΔNDMI	11	94.6%	0.75	0.73	0.62	8.8%
MOD02 1 km	ΔNDVI	11	91.0%	0.65	0.61	0.47	10.6%
	ΔNDMI	11	91.0%	0.73	0.65	0.55	9.9%
MOD13 HKM	ΔNDVI	11	89.8%	0.64	0.52	0.47	13.7%
	ΔEVI	11	90.3%	0.66	0.60	0.50	12.4%
	ΔNDMI	11	81.4%	0.56	0.53	0.45	13.9%
MOD13 1 km	ΔNDVI	11	88.7%	0.57	0.59	0.47	12.5%
	ΔEVI	11	90.1%	0.67	0.49	0.48	11.5%
	ΔNDMI	11	79.8%	0.50	0.43	0.46	12.9%
MOD43 1 km	ΔNDVI	11	92.2%	0.65	0.54	0.49	10.8%
	ΔNDMI	11	92.6%	0.68	0.63	0.53	10.2%
	ΔTCW	11	85.2%	0.63	0.57	0.50	11.4%

resulted from using the ΔNDMI index calculated with the MOD02 HKM data set.

Within the MOD02 data sets, the ΔNDMI outperformed the models based on ΔNDVI, with an approximate increase in error of 1% when moving from half-kilometer (MOD02 HKM) to one-kilometer resolution (MOD02 1 km). On the other hand, models based on the MOD13 ΔEVI index actually showed an approximately 1% improvement in mean error with the more coarse resolution product. Compared to the mean test statistics of the MOD02 models, the MOD13 ΔNDVI had 3.7% greater combined model error at half-kilometer resolution and 2.5% at one-kilometer and the MOD13 ΔNDMI error increased by 5.1% and 3.0%, respectively. The MOD43 ΔNDMI model had only slightly greater error (0.3%) on average than that of the same index calculated from Calibrated Radiances data at equivalent spatial resolution (MOD02 1 km ΔNDMI). Relatively good correlations with ΔFC were also found with the MOD43-derived ΔNDVI and the ΔTCW indices, but the ΔTCW had generally lower threshold accuracies and higher combined model errors.

The utility of the ΔNDMI index for estimating ΔFC was consistent with the findings of Hayes and Cohen (2007), who demonstrated the importance of including information from SWIR wavelengths in these models. In spite of this, the ΔNDMI index calculated from the MOD13 data sets performed comparatively poorly in modeling ΔFC. The MOD13 NDMI differs from the same index used in the MOD02 and MOD43 data sets in several ways, most noteworthy being the use of band 7 reflectance, as opposed to band 6 in the MOD02 and MOD43, and the difference in compositing algorithms. The MOD13 EVI, on the other hand, applies correction factors to the NDVI so it is more responsive to properties of the vegetation itself by removing background soil brightness and moisture variation (Huete et al., 2002). Models predicting ΔFC from ΔEVI performed best among those derived from the MOD13 data set tested here.

MODIS spectral indices generally showed good correlation with ΔFC from the regeneration model set in the by scene–by interval analysis. Indeed, the summary statistics revealed higher correlations between spectral indices and ΔFC from the regeneration model set than that from the clearing model set. These findings are consistent with the model results of Hayes and Cohen (2007), which showed that forest regeneration (the $\Delta FC \geq 0\%$ model set) could be predicted with good accuracy when developing models for an individual Landsat study site at a particular time interval. The relative correlations for the regeneration against the clearing model sets varied with the time interval, dependent on the spectral response to the interannual variability in land cover condition.

3.1.2. Across scenes–by interval analysis

Comparing data sets and spectral indices when modeling over multiple scenes for the same time interval showed similar results to the by scene–by interval analysis. For example, Table 4 gives the resulting test statistics for the models based on the sample set that combined the data from the three scenes that covered the 2000 to 2002 time interval (12/54, 19/48 and 20/48). The best model performance was found with the MOD02 ΔNDMI at half-kilometer resolution, which explained a much greater amount of the variation in ΔFC and had lower prediction error than the same index calculated from the MOD13 data at equivalent spatial resolution. The MOD13 products are based on MOD09 data, which has been shown to result in a similar deterioration of model performance when compared to MOD02 data (Hayes & Cohen, 2007). The lower correlation and higher error in the MOD13 models, then, may be in part based on the geolocation offset problem that exists for gridded MODIS products (Tan et al., 2006).

The analyses of Hayes and Cohen (2007) demonstrated that MOD02 data, which are available as the original Swath observations, more closely matched the spatial patterns of higher resolution reference data than did higher-level products that had been resampled to the Sinusoidal Grid projection system (as with MOD09, MOD13 and MOD43, for example). Other differences between the MOD02 and MOD13 data sets include the temporal resolution of the data and the level of atmospheric correction processing. The MOD02 data are a daily product with no quality screening and consist of at-satellite calibrated radiances values, prior to the application of any atmospheric correction algorithm. The MOD13 indices, on

Table 4
Results of the across scenes–by interval analysis showing the RMA regression model test statistics for each data set and variable resulting from combining all of the sample data for the 3 study areas (12/54, 19/48 and 20/48) from the 2000 to 2002 time interval

Data Set	Variable	*n*	Threshold accuracy %	Regeneration set *R*	Clearing set *R*	Combined model R^2	Model error
MOD02 HKM	ΔNDVI	429	89.4%	0.52	0.64	0.53	10.2%
	ΔNDMI	429	92.3%	0.69	0.83	0.77	9.2%
MOD02 1 km	ΔNDVI	342	83.1%	0.41	0.46	0.46	13.2%
	ΔNDMI	342	83.8%	0.62	0.76	0.68	11.5%
MOD13 HKM	ΔNDVI	247	77.4%	0.30	0.42	0.39	16.2%
	ΔEVI	247	89.9%	0.34	0.52	0.46	12.6%
	ΔNDMI	247	85.0%	0.34	0.44	0.42	14.5%
MOD13 1 km	ΔNDVI	131	82.4%	0.38	0.45	0.48	13.2%
	ΔEVI	131	86.8%	0.48	0.62	0.40	12.0%
	ΔNDMI	131	83.1%	0.47	0.43	0.33	13.4%
MOD43 1 km	ΔNDVI	266	87.1%	0.45	0.62	0.51	12.2%
	ΔNDMI	266	88.1%	0.67	0.79	0.71	11.5%
	ΔTCW	266	84.5%	0.48	0.57	0.49	12.5%

the other hand, are based on the surface reflectance algorithm (from the MOD09 product) and are a result of a 16-day compositing of the "best quality" samples that maximize the EVI signal. Tan et al. (2006) found the lack of consistency of gridded MODIS products with higher resolution data (due to the geolocation offset problem) to be exacerbated by multi-temporal compositing.

The results of the study by Tan et al. (2006) further imply that the geolocation offset problem diminishes as the size of the grid cell increases, and the authors suggested that better results, in terms of matching MODIS products to higher resolution data, may be obtained at 1 km resolution. In this study, the similar results with respect to model performance between the MOD43 ΔNDMI and the MOD02 1 km ΔNDMI support this contention. If the geolocation offset problem was significant at 1 km resolution, we would expect to see an improvement in model performance with the MOD02 index, based on Swath data, over the gridded MOD43 product. It is important to note that there are other unique properties of the MOD43 product, particularly the NBAR algorithm and 16-day mean solar zenith compositing, which distinguish it from the other data sets evaluated. These differences are likely to account for the improved model performance of the MOD43 when compared with the same index from the MOD13 data, which are also a gridded but do not compare as well to the MOD02 data. With an improved index such as the ΔEVI, however, the MOD13 model at 1 km resolution compared better to MOD02 and MOD43 models at the same spatial resolution. Visual analysis of the MOD13 ΔNDMI imagery suggests that, in terms of its utility in predicting ΔFC, this index suffers from artifacts of the compositing algorithm, which results in spatial discontinuity with regard to pixel-by-pixel spectral response.

In the previous (by scene–by interval) analysis, we saw that spectral indices showed good correlation with forest regeneration for a given study site and time interval pair. With spatial variability introduced in the across scenes–by interval analysis, these correlations decrease. The relationship between changes in spectral indices and forest clearing appeared to be more consistent when the models were expanded over space. In fact, the ability to model forest clearing improved with the addition of the larger set of reference sample points created by combining multiple scenes over a larger spatial extent.

3.1.3. By scene–across intervals analysis

Models based on a combined sample that grouped all of the data from one scene (20/48) over multiple time intervals (2000 to 2001, 2001 to 2002, and 2002 to 2003) give different results (Table 5), in terms of model performance amongst the different data sets and spectral indices, than did models based on sample data from a single time interval. While the best indices for each data set remained consistent (i.e. ΔNDMI for MOD02 and MOD43 and ΔEVI for MOD13), the MOD13 and MOD43 models showed results that were similar to, or better than, those of the MOD02 data when tested against multi-temporal reference data. Spectral indices calculated from reflectance data (MOD13 and MOD43) resulted in consistent model parameters and good overall correlation and prediction accuracy when modeled over multiple time intervals.

While the MOD02 data provided the best model results when applied for a given time interval, indices based on radiance data were not as consistent from year-to-year, which caused the MOD02 models to degrade over multiple time intervals. As such, the removal of atmospheric variation and view angle differences between dates, as with the MOD13 and MOD43 data, was important in building a general model that was consistent over time. However, the poor performance of the half-kilometer MOD13 models, relative to the MOD02 HKM, suggested that the geolocation offset problem still outweighs any advantage from using the reflectance data over radiances at this spatial resolution. At one-kilometer resolution, on the other hand, the MOD43 models showed improvement over those based on MOD02 data, highlighting the importance of the NBAR algorithm and mean solar zenith angle compositing in modeling ΔFC over multiple time intervals.

The addition of temporal variability in this analysis resulted in a decrease in correlation between spectral indices and ΔFC (both clearing and regeneration) as compared with the analyses based on one time interval. Still, the relationship between change in the spectral indices and forest clearing appeared to be more consistent over time than that for forest regeneration. In

Table 5
Results of the by scene–across interval analysis showing the RMA regression model test statistics for each data set and variable resulting from combining all of the sample data from the 3 time intervals (2000–01–02–03) for the 24/48 study site

Data Set	Variable	*n*	Threshold accuracy %	Regeneration set *R*	Clearing set *R*	Combined model R^2	Model error
MOD02 HKM	ΔNDVI	462	89.4%	0.35	0.62	0.44	11.7%
	ΔNDMI	462	92.3%	0.45	0.68	0.62	9.6%
MOD02 1 km	ΔNDVI	396	83.1%	0.32	0.57	0.43	12.9%
	ΔNDMI	396	83.8%	0.43	0.62	0.61	10.3%
MOD13 HKM	ΔNDVI	261	77.4%	0.31	0.44	0.37	16.8%
	ΔEVI	261	89.9%	0.62	0.49	0.43	15.7%
	ΔNDMI	261	85.0%	0.39	0.43	0.38	15.8%
MOD13 1 km	ΔNDVI	142	82.4%	0.48	0.50	0.57	10.9%
	ΔEVI	142	86.8%	0.65	0.58	0.62	10.3%
	ΔNDMI	142	83.1%	0.44	0.42	0.39	11.5%
MOD43 1 km	ΔNDVI	303	87.1%	0.55	0.65	0.65	8.9%
	ΔNDMI	303	88.1%	0.59	0.78	0.68	8.8%
	ΔTCW	303	84.5%	0.64	0.63	0.64	10.7%

general, models developed for individual scene and time interval pairs do better in predicting forest regeneration than do models expanded over time and/or space. The models based on the MOD43-derived indices showed the smallest decrease in correlation for both model sets when the temporal variability was introduced in this analysis.

3.1.4. Across scenes–across intervals analysis

The scatter of points in each plot illustrating the relationship of reference ΔFC to the different data sets and spectral change indices (Fig. 2) indicated that the distinct relationships separating forest loss and forest regeneration, apparent in the scene-by-scene analyses described above and reported by Hayes and Cohen (2007), was less conspicuous when the sample data from multiple time intervals and study sites was combined. As such, the different data sets and individual indices (MOD02 HKM ΔNDMI, MOD02 1 km ΔNDMI, MOD13 1 km ΔEVI and MOD43 ΔNDMI) were compared on the basis of a single RMA regression line for all sample data across the range of reference ΔFC values. The amount of variation explained and the prediction error in these models fell within the range of the test statistics calculated for the previous *across scenes–by time interval* (Table 4) and *by scene–across time intervals* analyses (Table 5), suggesting that a single linear model without thresholding did not significantly affect model performance. These combined sample set relationships were better illustrated as single, best-fit RMA lines on scatter plots of the spectral variables against the reference data than with the tabulated summary statistics of the relationships in the previous by scene or by time period analyses.

Comparisons among the model relationships with ΔFC for the best individual spectral indices from the different MODIS data sets (Fig. 2) were evidence for the relative effects of the different spatial, spectral and temporal properties of these data

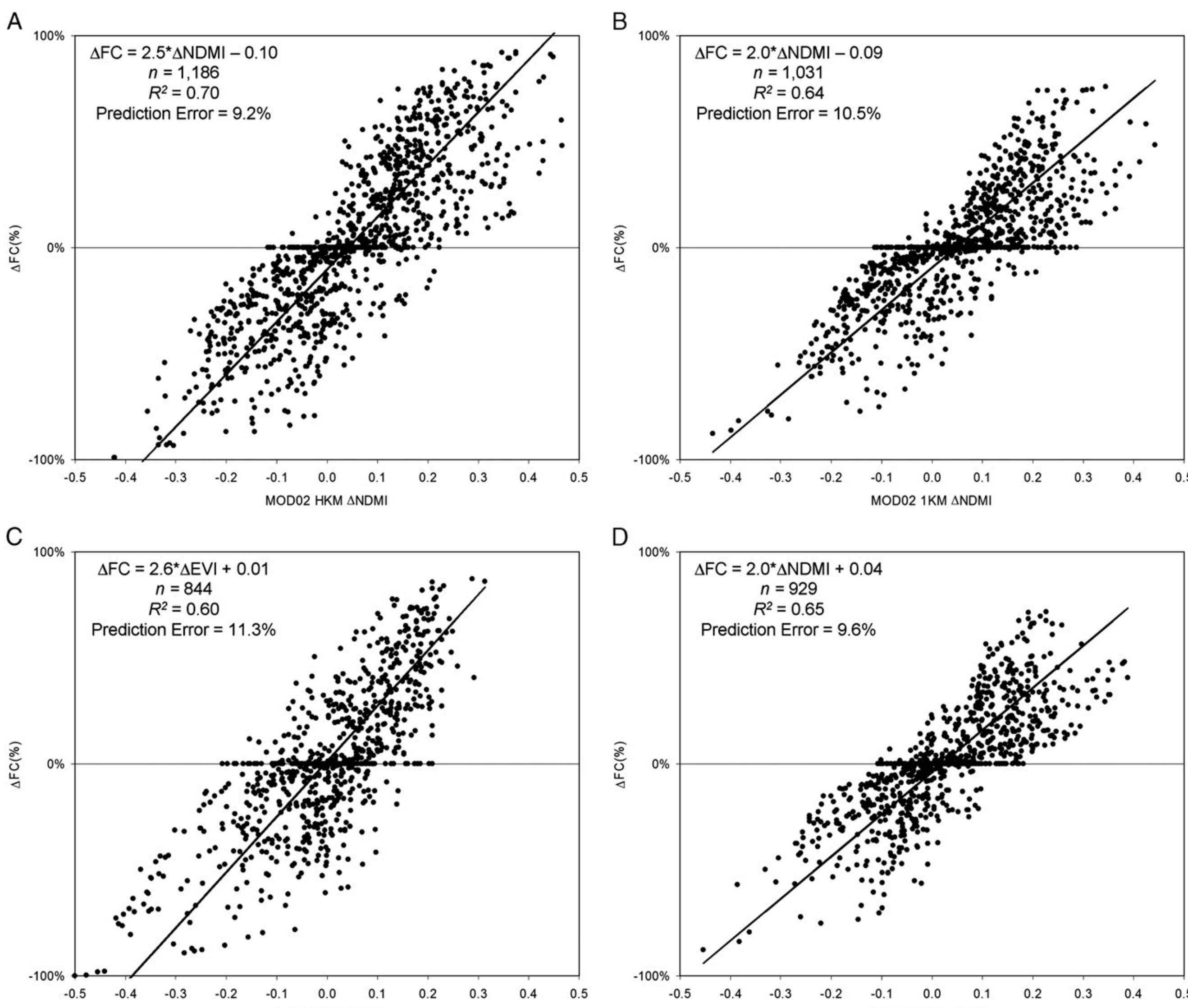

Fig. 2. The relationship between reference ΔFC for the pooled sample set and MODIS spectral change indices: A) MOD02 HKM ΔNDMI, B) MOD02 1 km ΔNDMI, C) MOD13 1 km ΔEVI and D) MOD43 ΔNDMI.

Table 6
Important attributes of each MODIS data set affecting its utility as the primary data source in a regional forest cover change monitoring system

Attribute	MOD02	MOD13	MOD43
Spatial resolution	Quarter-, half-, and one-kilometer	Quarter-, half-, and one-kilometer	One-kilometer
Projection and geometry	Swath data	Sinusoidal grid	Sinusoidal grid
Spatial coverage	Variable	Grid tiles	Grid tiles
Temporal coverage	Daily	16-day quality screened composite	16-day mean solar zenith angle composite
Radiometric properties	At-satellite radiance	Surface reflectance	Nadir-adjusted bidirectional reflectance
User quality assurance	None	Water and cloud masks, aerosol quality	Water and some cloud masking
Bands and indices	Visible, NIR, SWIR	NDVI, EVI, and blue, red, NIR, SWIR	Visible, near IR, SWIR, tasseled cap

on the amount of variation explained and the prediction error in modeling ΔFC. A reduction in spatial resolution from half- to one-kilometer for the same data set and spectral index (MOD02 ΔNDMI) resulted in 6% less of the variation ΔFC explained and a 1.3% increase in prediction error. At one-kilometer resolution, the ΔNDMI based on the NBAR product explained a similar amount of the variation in ΔFC as did that based on the radiance data, with an improvement in prediction error of nearly 1%. The best index from the MOD13 data set, on the other hand, explained less variation in ΔFC with a higher prediction error than both the MOD02 and MOD43 models. By using the one-kilometer MOD43 ΔNDMI to model ΔFC for the combined sample set, only a slight increase in prediction error (0.4%) occurred in comparison with the best model based on the half-kilometer MOD02 ΔNDMI.

3.2. Model selection

The MOD43 data set was chosen for use in estimating annual ΔFC for the Central American region by weighing the advantages and disadvantages of the different MODIS data sets evaluated in this study (Table 6). Overall, the MOD43 product proved to be the most consistent data set, was practical to implement, and showed similar results in terms of predicting ΔFC to higher resolution models based on the radiance Swath data (MOD02). The main advantage of the MOD02 data is its locational precision when compared to higher resolution reference data, which contributed to its better model performance in these analyses. However, these data are not corrected for atmospheric conditions and are not accompanied by any user-level quality assurance information or cloud or water masks. The higher order products (e.g. MOD13 and MOD43) offer more extensive and useful quality assurance information designed for the user. Another difficulty with the repeated use of the MOD02 data is the Swath format for which it is distributed. The spatial coverage of each Swath is variable on a day-to-day basis and an extra data set (MOD03 geolocation fields) is required for geometric correction of these data. Both the MOD13 and MOD43 data sets, on the other hand, are available in spatially consistent Grid tile format and are readily imported into image processing software as already geometrically corrected images.

Anniversary date acquisitions of MOD43 data sets were used to develop a general model for application across the Central American region that estimates annual change in forest cover as a continuous variable. Here, the emphasis was on choosing the best combination of spectral variables and indices in multiple linear regression models for predicting ΔFC, developed through a series of intentional selection and step-wise regression analyses (Table 7). Among the models based on single MOD43 indices (Models 1, 2 and 3), the ΔNDMI model had low standard error of the parameter estimates, explained the greatest amount of variation in the reference ΔFC, and had the lowest error in predictions. Compared to the models based on ΔNDVI and ΔNDMI, modeling ΔFC on the ΔTCW by itself resulted in a relatively low coefficient of determination and high prediction error. The ΔBand3 index was a significant variable in the forward step-wise regression analysis when the ΔNDVI and/or ΔNDMI variables were first added to the model. Adding ΔBand3 to the ΔNDVI (Model 4) contributed to a slight improvement in the amount of variation explained in the model, as well as in the error in the ΔFC predictions. A multiple linear regression model including ΔBand3, ΔNDVI and ΔNDMI (Model 5), all significant variables in the model, had the highest R^2 value (0.68) and lowest error (9.0%) of the six MOD43 anniversary date models evaluated in this exercise. The average error in the model predictions was distributed relatively evenly over the range of observed ΔFC values, as illustrated in a plot of

Table 7
RMA regression test statistics for the combined sample set (n = 1208) based on the MOD43 anniversary date spectral changes

Model	Variables	Value	SE	Model R^2	Prediction % error
Model 1				0.6015	10.65%
	Intercept	−0.0707	0.0063		
	ΔNDVI	2.3530	0.0552		
Model 2				0.6162	9.88%
	Intercept	−0.0696	0.0062		
	ΔNDMI	2.2914	0.0521		
Model 3				0.3930	13.85%
	Intercept	−0.1397	0.0092		
	ΔTCW	3.9664	0.1419		
Model 4				0.6182	9.48%
	Intercept	−0.0815	0.0067		
	ΔNDVI	2.5240	0.0664		
	ΔBand3	0.0007	0.0002		
Model 5				0.6792	9.00%
	Intercept	−0.0875	0.0064		
	ΔBand3	0.0009	0.0001		
	ΔNDVI	1.1854	0.1347		
	ΔNDMI	1.4487	0.1287		
Model 6				0.6641	9.30%
	Intercept	−0.0980	0.0073		
	ΔTCB	−1.3047	0.0854		
	ΔTCG	2.6724	0.0896		
	ΔTCW	4.0473	0.1077		

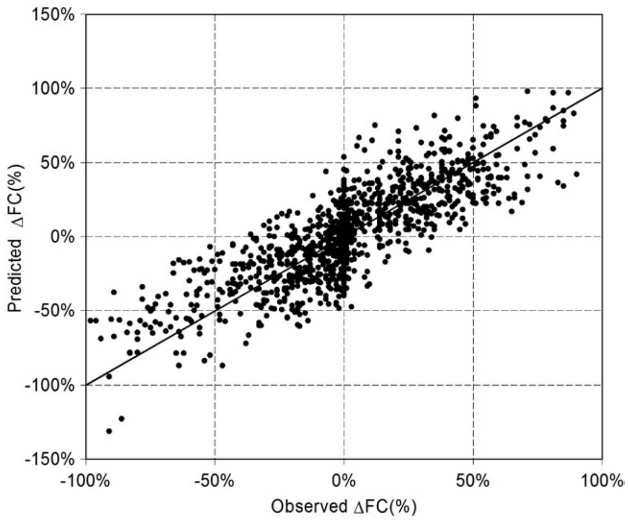

Fig. 3. Plot illustrating the relationship between the observed reference ΔFC and Model 5 predicted values, based on the pooled sample set (n=1208).

model results against reference values in relation to the 1:1 line (Fig. 3). The three tasseled cap change indices (Model 6) were all significant variables when included together in the regression model, which explained 66.4% of the variation in ΔFC and had a prediction error of 9.3%.

This analysis demonstrated the importance of the "brightness", "greenness" and "wetness" components of the spectral domain in detecting and measuring change. From the MOD43 anniversary dates, the variable combination that produced good predictions against the reference data used here included the change in Band 3, the NDVI and the NDMI. MODIS Band 3 measures reflectance in the visible blue wavelength region, which is also used in the ΔEVI index to compensate for background soil brightness and moisture. The NDVI is a common measure of vegetation greenness, and the NDMI incorporates SWIR spectral information and correlates to vegetation structure and moisture content (Jin & Sader, 2005; Wilson & Sader, 2002). A similar model, based on the TCB, TCG and TCW indices gave similar results in terms of model performance, as well as in net ΔFC estimates and error statistics when applied at the regional scale and over multiple time periods.

4. Summary and conclusions

An RMA-transformed linear regression model based on multiple indices calculated from annual anniversary dates of MOD43 data, applied region-wide in Central America for this study, demonstrated a practical methodology for estimating proportional change in forest cover as a continuous variable with good accuracy. A model incorporating spectral changes in the NDMI and NDVI indices and the visible blue channel explained 68% of the variability in ΔFC with a prediction error of 9%. Taken together, the results of the various analyses summarized above demonstrated a number of the different issues involved in using coarse resolution MODIS data and products to detect land cover change and estimate proportional changes in forest cover at landscape to regional scales, and over multiple time intervals. The analyses were designed to allow evaluation and comparison of model performance over varying time intervals and spatial extents amongst models based on different MODIS data sets and spectral indices. With respect to the research objectives outlined above, some of the key findings of this study include:

- When estimating change in forest cover at landscape to regional scales for a particular time interval, the models based on the half-kilometer MOD02 ΔNDMI index had the best relationship with the reference data and the lowest prediction errors. The main advantage of the MOD02 data is the geometric locational precision of the Swath observations at 500 m native resolution with respect to the higher resolution reference data.
- Spectral indices based on atmospherically corrected surface reflectance data, as with the MOD13 and MOD43 data sets, produced more consistent model parameters and accurate forest cover change estimates when modeled over multiple time intervals.
- The MOD43 product proved to be the most consistent data set, was practical to implement, and allowed for the computation of a number of important spectral vegetation indices. Models based on these indices showed similar results in terms of predicting forest cover change to higher resolution models based on radiance swath data.
- The most accurate estimates of forest cover change resulted from models that included measures of the year-to-year change in the brightness (e.g. visible bands or TCB), greenness (NDVI or TCG) and wetness (NDMI or TCW) spectral domains of the MOD43 data.

Selecting the best data upon which to build a forest cover change model (objective 1) was realized through comparisons among three MODIS data sets, each of which had potential advantages and disadvantages in terms of model performance. The MOD02 data set has been shown to have a strong relationship with the spatial pattern and magnitude of forest cover change (Hayes & Cohen, 2007). However, these data are not corrected for day-to-day atmospheric conditions, thereby putting into question the stability of these data over time in the face of temporal variability. Other MODIS data sets such as the Vegetation Indices product and the NBAR data set are corrected based on potentially more consistent surface reflectance algorithms. Each data set, grain size and spectral index was evaluated by individual study sites (Landsat scenes) and time period interval, across scenes by interval, by scene across intervals, and across all scenes and intervals. Evaluation among these different data sets based on the strength and consistency of their relationship with forest cover change for a sample of high resolution reference data provided the criteria for choosing the best spectral variables in the model (objective 2). Models that were practical to implement and produced accurate and consistent results were suggested for application over the full regional data set to assess the quality and usefulness of the forest cover change estimates, and to identify any limitations and areas for improvement in model development.

The subject matter of this study is important and timely as scientists and users are investigating the issues involved in moving from the use of MODIS data as a global land characterization and productivity sensor to one geared toward more regional and local investigations. The results presented here shed light on important MODIS data properties such as the effects of compositing on the geometric quality of the data, of atmospheric and BRDF corrections on spectral vegetation change indices, and on the extraction of detailed, continuous forest cover change variables from coarse resolution data. This study provides a foundation for the further investigation of the use of these data at local and global scales. Site-specific studies of the spectral and spatial properties of different LCLUC types, including seasonal, atmospheric and BRDF effects, will inform future model development for the monitoring of different ecosystems around the globe. Such models will provide key data sources for resource management and conservation decisions as well as input to models estimating the drivers and consequences of global environmental change.

Acknowledgements

This research was funded by a NASA Earth System Science Fellowship (NGT5-30526). Image and field data support were provided by the Laboratory for Applications of Remote Sensing in Ecology at Oregon State University, the NASA/USAID SERVIR Data Center in Panamá and the ProPetén Foundation in Guatemala. We greatly appreciate the indispensable assistance in research development from W. David Ritts of Oregon State University and Thomas K. Maiersperger at the USGS EROS Data Center. Anonymous reviewers provided valuable comments which ultimately greatly improved the organization, clarity and quality of the manuscript.

References

Adams, J. B., Sabol, D., Kapos, V., Filho, R. A., Roberts, D. A., Smith, M. O., et al. (1995). Classification of multispectral images based on fractions of endmembers: application to land-cover change in the Brazilian Amazon. *Remote Sensing of Environment*, *52*, 137−154.

Asner, G. P. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, *22*, 3855−3862.

Chavez Jr., P. S. (1996). Image-based atmospheric corrections — Revisited and improved. *Photogrammetric Engineering and Remote Sensing*, *62*, 1025−1036.

Cohen, W. B., & Goward, S. N. (2004). Landsat's role in ecological applications of remote sensing. *Bioscience*, *54*(6), 535−545.

Cohen, W. B., Maiersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, *84*, 561−571.

Cohen, W. B., & Spies, T. A. (1992). Estimating structural attributes of Douglas-fir/western hemlock forest stands from LANDSAT and SPOT imagery. *Remote Sensing of Environment*, *41*, 1−17.

Cohen, W. B., Spies, T. A., Alig, R. J., Oetter, D. R., Maiersperger, T. K., & Fiorella, M. (2003). Characterizing 23 years (1972–95) of stand replacement disturbance in Western Oregon forests with Landsat imagery. *Ecosystems*, *5*, 122−137.

Coppin, P., Jonckheere, I., Nackaerts, K., Muysand, B., & Lambin, E. (2004). Digital change detection methods in ecosystem monitoring: A review. *International Journal of Remote Sensing*, *25*, 1565−1596.

Crist, E. P. (1985). A TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sensing of Environment*, *17*, 301−306.

Curran, P. J., & Hay, A. (1986). The importance of measurement error for certain procedures in remote sensing at optical wavelengths. *Photogrammetric Engineering and Remote Sensing*, *52*, 229−241.

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

Ehrlich, D., Lambin, E. F., & Malingreau, J. P. (1997). Biomass burning and broad-scale land cover changes in Western Africa. *Remote Sensing of Environment*, *61*, 201−209.

[ERDAS] (2005). *ERDAS Imagine version 8.7.* Leica Geosystems Inc.

[FAO] Food and Agricultural Organization. (1993). Forest Resources Assessment, 1990, Tropical Countries. FAO Forestry Paper 112. Rome: Food and Agricultural Organization of the United Nations.

Fearnside, P. M. (1986). Spatial concentration of deforestation in the Brazilian Amazon. *Ambio*, *15*, 74−81.

Hall, F. G., Strebel, D. E., Nickeson, J. E., & Goetz, S. J. (1991). Radiometric rectification: toward a common radiometric response among multidate, multisensor images. *Remote Sensing of Environment*, *35*, 11−27.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Sohlberg, R., Dimiceli, C., & Carroll, M. (2002). Towards an operational MODIS continuous field of percent tree cover algorithm: Examples using AVHRR and MODIS data. *Remote Sensing of Environment*, *83*, 303−319.

Hayes, D. J., & Cohen, W. B. (2007). Spatial, spectral and temporal patterns of tropical forest cover change as observed with multiple scales of optical satellite data. *Remote Sensing of Environment*, *106*, 1−17.

Hayes, D. J., & Sader, S. A. (2001). Change detection techniques for monitoring forest clearing and regrowth in a tropical moist forest. *Photogrammetric Engineering and Remote Sensing*, *67*(9), 1067−1075.

Hayes, D. J., Sader, S. A., & Schwartz, N. B. (2002). Developing a forest conversion history database to explore the temporal and spatial characteristics of land cover change in the Maya Biosphere Reserve, Guatemala. *Landscape Ecology*, *17*, 299−314.

Healey, S. P., Cohen, W. B., Zhiqiang, Y., & Krankina, O. (2005). Comparison of Tasseled Cap-based Landsat data structures for forest disturbance detection. *Remote Sensing of Environment*, *97*, 301−310.

Holdridge, L. R., Grenke, W. C., Hatheway, W. H., Liang, T., & Tosi, J. A. (1971). *Forest environments in tropical life zones, a pilot study.* Pergamon Press: New York.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., & Ferreira, L. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*, 195−213.

Jensen, J. R. (1996). *Introductory digital image processing.* Englewood Cliffs: Prentice-Hall.

Jin, S., & Sader, S. A. (2005). Comparison of time-series tasseled cap wetness and the normalized difference moisture index in detecting forest disturbances. *Remote Sensing of Environment*, *94*, 364−372.

Justice, C. O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. P., Hall, D. K., et al. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, *36*, 1228−1249.

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.

Lambin, E. F., & Ehrlich, D. (1997). Land-cover changes in Sub-Saharan Africa (1982–1991): Application of a change index based on remotely sensed surface temperature and vegetation indices at a continental scale. *Remote Sensing of Environment*, *61*, 181−200.

Lobser, S. E. & Cohen, W. B. (in press). The MODIS Tasseled Cap: Land cover characteristics expressed through transformed MODIS data. *International Journal of Remote Sensing*.

Lucas, R. M., Honzak, M., Curran, P. J., Foody, G. M., Milne, R., Brown, T., et al. (2000). Mapping the regional extent of tropical forest regeneration stages in the Brazilian Legal Amazon using NOAA AVHRR data. *International Journal of Remote Sensing*, *21*, 2855−2881.

Lucas, R. M., Honzak, M., Curran, P. J., Foody, G. M., & Nguele, D. T. (2000). Characterizing tropical forest regeneration in Cameroon using NOAA AVHRR data. *International Journal of Remote Sensing*, *21*, 2831−2854.

Malingreau, J. P., & Tucker, C. J. (1988). Large-scale deforestation in the southeastern Amazon basin of Brazil. *Ambio*, *17*, 49−55.

Mayaux, P., Achard, F., & Mallingreau, J. P. (1998). Global tropical forest area measurements derived from coarse resolution maps at a global level: A comparison with other approaches. *Environmental Conservation*, *25*, 37−52.

Miller, K., Chang, E., & Johnson, N. (2001). *Defining common ground for the Mesoamerican Biological Corridor*. Washington, DC: World Resources Institute.

Richards, T., Gallego, J., & Achard, F. (2000). Sampling for forest cover change assessment at the pan-tropical scale. *International Journal of Remote Sensing*, *21*, 1473−1490.

Sader, S. A., Hayes, D. J., Hepinstall, J. A., Coan, M., & Soza, C. (2001). Forest change monitoring of a remote biosphere reserve. *International Journal of Remote Sensing*, *22*, 1937−1950.

Sader, S. A., Hayes, D. J., Irwin, D. E., & Saatchi, S. S. (2001). Preliminary forest cover estimates for Central America (1990's) with reference to the proposed Mesoamerican Biological Corridor. *2001 Annual Meeting, St. Louis, Missouri, April 23–26. ASPRS, Bethesda, Maryland USA*: American Society for Photogrammetry and Remote Sensing (ASPRS).

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., et al. (2002). First operational BRDF, albedo and nadir reflectance products from MODIS. *Remote Sensing of Environment*, *83*, 135−148.

Skole, D., & Tucker, C. (1993). Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988. *Science*, *260*, 1905−1910.

Tan, B., Woodcock, C. E., Hu, J., Zhang, P., Ozdogan, M., Huang, D., et al. (2006). The impact of gridding artifacts on the local spatial properties of MODIS data: Implications for validation, compositing, and band-to-band registration across resolutions. *Remote Sensing of Environment*, *105*, 98−114.

Townshend, J. R. G., & Justice, C. O. (1988). Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations. *International Journal of Remote Sensing*, *9*, 187−236.

Townshend, J. R. G., & Justice, C. O. (2002). Towards operational monitoring of terrestrial systems by moderate-resolution remote sensing. *Remote Sensing of Environment*, *83*, 351−359.

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.

Tucker, C. J., & Townshend, J. R. G. (2000). Strategies for monitoring tropical deforestation using satellite data. *International Journal of Remote Sensing*, *21*, 1461−1471.

Van Huffel, S. (Ed.). (1997). *Recent advances in total least squares techniques and errors-in-variables modelling.* Philadelphia: Society for Industrial and Applied Mathematics.

Wilson, E. H., & Sader, S. A. (2002). Detection of forest harvest type using multiple dates of Landsat TM imagery. *Remote Sensing of Environment*, *80*, 385−396.

Zhan, X., DeFries, R. S., Townshend, J. R. G., Dimiceli, C., Hansen, M., Huang, C., et al. (2000). The 250 m global land cover change product from the moderate resolution imaging spectroradiometer of NASA's earth observing system. *International Journal of Remote Sensing*, *21*, 1433−1460.

Zhan, X., Sohlberg, R., Townshend, J. R. G., Dimiceli, C., Carroll, M. L., Eastman, J. C., et al. (2002). Detection of land cover changes using MODIS 250 m data. *Remote Sensing of Environment*, *83*, 336−350.