ELSEVIER

Remote Sensing of Environment 82 (2002) 258–270

Remote Sensing of Environment

www.elsevier.com/locate/rse

Impacts of vegetation dynamics on the identification of land-cover change in a biologically complex community in North Carolina, USA

Ross S. Lunetta[a,*], Jayantha Ediriwickrema[b], David M. Johnson[b], John G. Lyon[c], Alexa McKerrow[b]

[a]U.S. Environmental Protection Agency, National Exposure Research Laboratory, Research Triangle Park, NC 27711, USA
[b]Lockheed Martin Services, Inc., Research Triangle Park, NC 27709, USA
[c]U.S. Environmental Protection Agency, National Exposure Research Laboratory, Las Vegas, NV 89193-3478, USA

Received 12 October 2001; received in revised form 3 April 2002; accepted 6 April 2002

Abstract

A land-cover (LC) change detection experiment was performed in the biologically complex landscape of the Neuse River Basin (NRB), North Carolina using Landsat 5 and 7 imagery collected in May of 1993 and 2000. Methods included pixel-wise Normalized Difference Vegetation Index (NDVI) and Multiband Image Difference (MID) techniques. The NDVI method utilized non-normalized (raw) imagery data, while the MID method required normalized imagery. Image normalization techniques included both automatic scattergram-controlled regression (ASCR) and localized relative radiometric normalization (LRRN) techniques. Change/no-change thresholds for each method were optimized using calibration curves developed from reference data and a series of method-specific binary change masks. Cover class-specific thresholds were derived for each of the four methods using a previously developed NRB-LC classification (1998–1999) to support data stratification. An independent set of accuracy assessment points was selected using a disproportionate stratified sampling strategy to support the development of error matrices. Area-weighted conditional probability accuracy statistics were calculated based on the areal extent of change and no change for each cover class. All methods tested exhibited acceptable accuracies, ranging between 80% and 91%. However, change omission errors for woody cover types were unacceptably high, with values ranging between 60% and 79%. Overall commission errors in the change category were also high (42–51%) and strongly affected by the agriculture class. There were no significant differences in the Kappa coefficient between the NDVI, MID ASCR, and LRRN normalization methods. The MID non-normalized method was inferior to both the NDVI and MID ASCR methods. Stratification by major LC type had no effect on overall accuracies, regardless of method.

1. Introduction

The characterization of land-cover (LC) type, extent, and distribution over time is required to monitor ecosystem condition and for the study of numerous ecosystem processes including habitat suitability, wetland functions, the allocation of nonpoint nutrient sources, and the study of erosion and sedimentation potentials. Of particular importance is the application of LC data for the generation of landscape-based assessment metrics to evaluate relative ecosystem condition over a wide range of analytical scales (watershed–national) to assess the impacts attributable to human land-use activities (Jones et al., 1997). Past, present, and emerging remote sensor technologies provide a cost-effective high quality source of data to characterize changing LC conditions and to monitor changes at multiple scales.

A major advance in the application of remote sensing technologies for change detection studies has been the development of consistent imagery databases and the implementation of LC mapping efforts at global, continental, and national scales. Global LC mapping efforts have been accomplished using National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) biweekly Normalized Difference Vegetation Index (NDVI) composites (Loveland, Merchant, Ohlen, & Brown, 1991) developed and compiled by the U.S. Geological Survey's (USGS) EROS Data Center (Belward, Estes, & Klene, 1999; Eidenshink, 1992). The North American Landscape Characterization (NALC) Landsat

* Corresponding author. Tel.: +1-919-541-4256; fax: +1-919-541-1138.
E-mail address: lunetta.ross@epa.gov (R.S. Lunetta).

0034-4257/02/$ - see front matter
PII: S0034-4257(02)00042-1

Pathfinder project assembled moderate-resolution imagery over three epoch periods (1970s to 1990s) for the intended purpose of supporting LC change analysis (Lunetta, Lyon, Guindon, & Elvidge, 1998). Recently, the Multi-Resolution Land Characteristic (MRLC) project, has provided the first consistent high-resolution National Land Cover Dataset (NLCD) for the conterminous United States (Van Driel, 2001; Vogelmann, Sohl, & Howard, 1998).

To date, large-area LC characterization efforts have primarily focused on the development of baseline data. With numerous baseline inventories now complete, many researchers are now focusing on the development of regional scale change detection methods to update existing LC data sets and also determine the type, extent, and distribution of LC changes. Two impediments that must be overcome in developing practical methods for regional scale change detection monitoring using remote sensor data include: (1) the reduction of vegetation phenology-mediated omission (Type I) and commission (Type II) errors and (2) robust methods to assess the accuracy of LC change results.

As part of the natural processes associated with vegetation dynamics, plants undergo intra-annual variations or cycles (phenology). Physical drivers including temperature, solar radiation, and water availability mediate these cycles in large part. For example, in higher latitudes, plant cycles are principally driven by temperature and photoperiod. While, in the arid lower latitudes, water availability, and photoperiod are the main drivers. During different stages of growth, plant structures, and associated pigment assemblages can vary significantly. Thus, the same vegetation type can appear significantly different at various growth cycle stages.

Changes in landscape composition and function result from both acute LC conversions and/or chronic landscape changes. LC conversions are typically mediated by human land-use activities (e.g., conversion from forest to agriculture), while more subtle chronic landscape changes can result from either natural processes (e.g., insect infestations, successional processes, and climatic changes) or human land-use activities (e.g., forest thinning). Numerous efforts have recently been undertaken in an attempt to document LC conversions using moderate-resolution remote sensor data for both the United States (Cohen & Fiorella, 1998; Elvidge, Miura, Jansen, Groeneveld, & Ray, 1998) and Mexico (Lunetta et al., in press; Lyon, Yuan, Lunetta, & Elvidge, 1998). However, current spectral-based change detection techniques using either pre- or post-classification approaches have tended to be performance limited in biologically complex ecosystems, owing in part to vegetation phenology induced errors. Phenology errors are largely associated with the temporal displacement (vegetation phenology versus data collection) inherent with sequential satellite data collections and commonly result in an unacceptable level of omission and commission errors.

2. Background

Remote sensing change detection techniques can be broadly classified into as either post-classification change methods or pre-classification spectral change detection. The post-classification approach involves the analysis of differences between two independent categorization products using GIS-based analytical tools (Lunetta, 1998). Originally, this approach was considered to be the most reliable and was used to evaluate emerging methodologies (Weismiller, Kristof, Scholz, Anuta, & Momin, 1977). Factors that limit the application of post-classification change detection techniques can include (a) cost, (b) consistency, and (c) error propagation. Thus, this approach can produce large numbers of erroneous change indications since an error in either data gives a false indication of change (Singh, 1989). Categorization errors associated with postclassification techniques can be characterized using an error propagation model (Lunetta et al., 1991).

Numerous change detection techniques and methodologies have been developed and evaluated to achieve optimal performance over the greatest possible range of ecosystem conditions. Among the first of the semi-automated digital data processing approaches was image-based composite analysis. The technique performs a single analysis of a multidate data set using standard pattern recognition and spectral classification techniques to identify LC change areas (Weismiller et al., 1977). Although this technique requires only a single classification, data should be collected under similar conditions (i.e., vegetation phenology, solar zenith angles, and atmospheric conditions), and the interpretation of LC versus change classes can be very complex.

A variation to the image composite analysis technique that minimizes manual interpretation requirements involves the use of principal components analysis (PCA). PCA is a powerful data transformation technique useful for information extraction with multispectral and/or multitemporal data (Lillesand & Keifer, 1979). Experiments have demonstrated when PCA is applied in a multitemporal data analysis, that the PC3 and later principal components tend to contain the changed LC information (Byrne, Crapper, & Mayo, 1980; Richards, 1984).

Recent research on the development of various image differencing techniques has focused on two general categories: (1) simple image subtraction and thresholding and (2) data transformation, subtraction, and thresholding. The most commonly applied data transformations include: (a) band ratioing, (b) NDVI, and (c) the tasseled-cap transformation. Simple image subtraction and thresholding generally requires that input imagery first be normalized (or otherwise corrected) for solar illumination, atmospheric scattering and absorption, and detector performance. However, some transformations (e.g., NDVI) incorporate data normalization directly into their calculation. Others, such as the tasseled-cap transformation, typically require data normalization prior to transformation calculation. Also, the tasseled-cap transformation has been determined to be influenced by sensor calibration, and thus is sensor dependent (Crist, 1985).

A robust analysis technique that can be applied to interpret data transformation results is change vector analysis (CVA). More powerful than a simple threshold technique, the change vector can be used to compare differences in the time trajectory of a biophysical indicator such as the NDVI or tasseled-cap transformation. The change vector represents the multidimensional vector difference between successive time trajectories, with length indicating the magnitude of change and direction indicating the nature of change (Lambin & Strahler, 1994a). When the time trajectory of an indicator departs from that expected, an LC change can be detected (Lambin & Strahler, 1994b). This provides a particularly powerful technique to analyze many successive dates of data. The Multiband Image Difference (MID) technique used in this study was a derivation of the above described CVA, because neither a coordinate transformation or change direction vector analysis were performed.

Past research evaluating the comparative performance of competing methods for LC change detection over large geographic areas has been limited. Cohen and Fiorella (1998) found that composite analysis out performed both image differencing and CVA in a two-date experiment for detecting conifer forest change using Landsat Thematic Mapper (TM) imagery in the Oregon Cascades. Lyon et al. (1998) reported that the NDVI was the best performing vegetation index (VI) for detecting general LC changes in the biologically complex vegetation communities of Chiapas, Mexico. Analysis of LC data for the country of Mexico using three dates of Landsat Multi-Spectral Scanner (MSS) imagery over a two-decade (1972–1992) study period demonstrated that ecosystem variability combined with classification errors precluded scene- or pixel-wise change detection (Lunetta et al., in press). Elvidge et al. (1998) clearly demonstrated the value of using high-temporal resolution MSS imagery to monitor changes in wetland vegetation.

2.1. Study site

The Neuse River Basin (NRB) is one of three river basins located entirely within the state boundaries of the state of North Carolina (Fig. 1). The NRB boundaries were delineated to correspond with the USGS, six-digit hydrologic unit code (HUC), code number 030202. The NRB is 14,582 km^2 in area and contains 16,900 km (1:24,000 scale) of stream length. The upper (northwestern) third of the basin is located in the Piedmont physiographic region and the remainder in the coastal plain region. The Piedmont portion of the basin is characterized by highly erodible clay soils, rolling topography with broad ridges and sharply indented stream valleys, and low gradient streams composed of a series of sluggish pools separated by riffles and occasional small rapids. In contrast, the coastal plain is characterized by flat terrain, “blackwater streams”, low-lying wetlands, and productive estuarine areas (NCEMC, 1993). Elevations within the NRB range from 276 m in the Piedmont to sea level at the confluence of the Neuse River and Pamlico Sound, which is defined by a series of barrier islands known as the Outer Banks.

2.2. Study objectives

The objectives of this study were to investigate the impacts of vegetation dynamics (phenology) associated with the detection of LC conversions using both NDVI difference and MID-based approaches in the biologically diverse

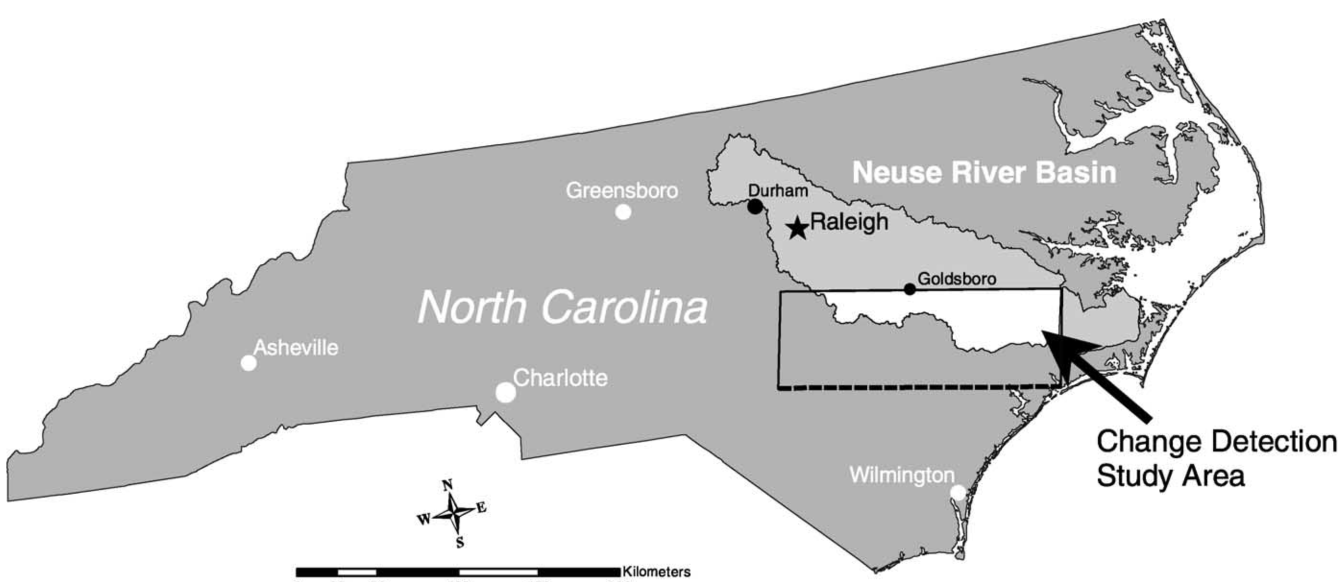

Fig. 1. Neuse River change detection study area within Landsat (WRS-2) path 15/row 36.

vegetation communities in the NRB of North Carolina. Specific study objectives included the following: (1) determine the utility of Landsat TM and Enhanced Thematic Mapper Plus (ETM^+) to characterize LC conversions, (2) compare the performance of NDVI difference and MID change detection techniques, (3) evaluate the impact(s) of vegetation dynamics on the identification of LC conversions, and (4) explore possible remedies to overcome obstacles and optimize performance.

3. Methods

This study was designed to compare change detection methods, stratification approaches, and image normalization techniques. To evaluate methodology performance, both the NDVI pixelwise differencing and MID techniques were applied to the same imagery data. Threshold levels were standardized by incorporating the use of calibration curves for selecting optimum values (Fung & LeDrew, 1988; Morisette & Khorram, 2000). Commonly referred to as accuracy curves, here we refer to them as "calibration curves". Numerous threshold levels were used to create several change masks for both methods. The threshold represented by the highest Kappa was selected as the optimum and was applied to develop change/no-change images. Stratifications were performed using the previously developed (1998–1999) NRB-LC classification (Lunetta et al., in press) to evaluate the benefits of selecting cover class-specific thresholds. Finally, a comparison involving the application of the MID change detection method to non-normalized and normalized data was performed. Both automatic scattergram-controlled regression (ASCR) and localized relative radiometric normalization (LRRN) techniques were employed in the normalization.

An accuracy assessment based on a stratified random sampling was performed for the eight final change images. Accuracy statistics were "conditioned" based on the areal extent of each strata and the number of samples each contained, relative to the totals for the study area (Stehman & Czaplewski, 1998). Standardization based on these conditional probabilities made the direct comparison of the error matrices between methods possible.

3.1. Data preprocessing

A subset of the NRB represented by a single Landsat scene (path 15/row 36, WRS-2) defined the extent of the study area (Fig. 1). Two dates of imagery, representing a 7-year time differential, were used to support the analysis. Imagery included a Landsat 5 TM image from May 16, 1993 and a Landsat 7 ETM^+ image from May 11, 2000. Phenological variations associated with the study area were expected to have been minimized due to the near "anniversary" imagery collection dates. Both scenes were maintained at the nominal pixel resolution of 30 m and were of high quality with little to no haze or cloud cover. Both images were georeferenced independently using an image-to-image registration based on a 1998 SPOT 4 (XS) imagery mosaic previously developed for the NRB.

3.2. Calibration reference data set development

Reference data were developed for the change detection threshold analysis using a random stratified sampling frame. The stratification was based on an estimated change/no-change mask created using a NDVI difference image and the LC classes from a 1998 to 1999 NRB-LC classification (Lunetta et al., in press). The change mask was created by setting the NDVI differencing threshold at 2.00 standard deviations. The major LC types in the study area included urban, agriculture, woody, and wetland and they represented over 98% of the scene (Table 1). The final stratified image was created by intersecting the change mask with the LC image. The resulting strata were then used in a stratified random sampling design for selection of 340 reference points. The number of sample points within each stratum was proportional to its areal extent. For each point, the change/no-change status was labeled based on analyst's interpretation of the 1.0-m resolution 1993 panchromatic Digital Orthophoto Quarter Quadrangles (DOQQs) and 1998 color infrared DOQQs. Reference points were then compared directly to the corresponding 30-m change/no-change pixels. Edge of patch pixels were avoided to minimize the impacts registration errors and to avoid ambiguities associated with LC transition areas.

3.3. NDVI change analysis

NDVI products were derived from the 1993 TM and 2000 ETM^+ images based on Richardson and Everitt (1992). Difference images were created by subtracting the 1993 from 2000 NDVI images. Stratification by cover class was included by intersecting this difference image with the NRB-LC (1998–1999) classification. The resulting difference image could contain data values ranging from 2.00 to − 2.00. Negative values suggested that pixels had decreased in green vegetation biomass over the index period, while positive values suggested a green biomass increase. Near-zero differences indicated little or no change in biomass. The difference values ranged from + 1.072 to − 1.292 and

Table 1
Statistics by LC class for the NDVI difference image (2000–1993)

Class	Pixels	Area (ha)	Areal %	Mean	S.D.
Urban	528,561	47,570.5	10.9	− 0.18	0.15
Agriculture	1,681,959	151,376.3	34.6	0.19	0.19
Woody	1,706,452	153,580.7	35.1	− 0.09	0.14
Water	63,741	5736.7	01.3	− 0.19	0.20
Wetland	866,821	78,013.9	17.8	− 0.09	0.13
Barren	19,952	1795.7	00.4	− 0.26	0.22
Total/average	4,867,486	438,073.7	100.00	− 0.14	0.17

were near-normally distributed around the mean (− 0.141) with a standard deviation of 0.166.

Assuming pixels with near-zero difference represented areas of no change, we attempted to separate the change/no-change areas by selecting a threshold value that maximized sample accuracy (Fung & LeDrew, 1988; Morisette & Khorram, 2000). Thresholds were selected based on increments of the standard deviation from the mean. To objectively test the standard deviation threshold levels, the change/no-change calibration reference data set described above was used to evaluate the relative performance of increments ranging from 0.50 to 3.00 standard deviations. Calibration curves were then developed for both the non-stratified (Fig. 2a) and the stratified (Fig. 2b–e) NDVI images. Because the Kappa statistic (KHAT) represents a compromise between commission and omission errors, it was selected as the statistic for use in identifying the optimum threshold. Thresholds exhibiting the highest Kappa values were applied to derive the binary change image.

3.4. MID change analysis

The MID analysis incorporated a tasseled-cap transformation to detect shifts in three dimensions—brightness, wetness, and greenness—between the two imagery dates (Crist & Kauth, 1986; Dwyer, Sayler, & Zylstra, 1996). In addition to quantifying the directional change magnitude, a noise image was applied as described by Dwyer et al. (1996), to modulate the magnitude of change in relation to the noise levels. In addition, two types of radiometric normalization were applied to compensate for differences in sun angle, atmospheric conditions, and detector calibration.

Both the ASCR (Elvidge, Yuan, Weerakoon, & Lunetta, 1995) and an LRRN (Ediriwickrema, Lunetta, Dulaney, McKerrow, & Johnson, 2002) were compared. The normalized/non-normalized images were used to develop three separate change vector images and subsequent binary change masks. ASCR was applied to the 1993 image using the 2000 image as the reference base (Elvidge et al., 1995). The ASCR procedure located stable land and water data clusters using near- and shortwave–infrared scattergrams and established an initial regression line relating the digital numbers (DNs) for both imagery dates. The process identified "no-change" pixels as those that fall within a set perpendicular distance from the regression line. For this study, pixels falling within a perpendicular distance of 10 DNs were considered invariant. Using only the invariant pixels, slope, and intercept regression coefficients were calculated and applied to the subject image creating a spectrally comparable image to that of the reference.

The LRRN is an automated method developed to account for problematic theoretical concepts and issues discussed by Heo and FitzHugh (2000) and Yang and Lo (2000) that occur in standard relative radiometric normalization. Initially, the method identifies the pixels with the greatest probability for being temporally invariant (no-change) in two temporal images. The total number of

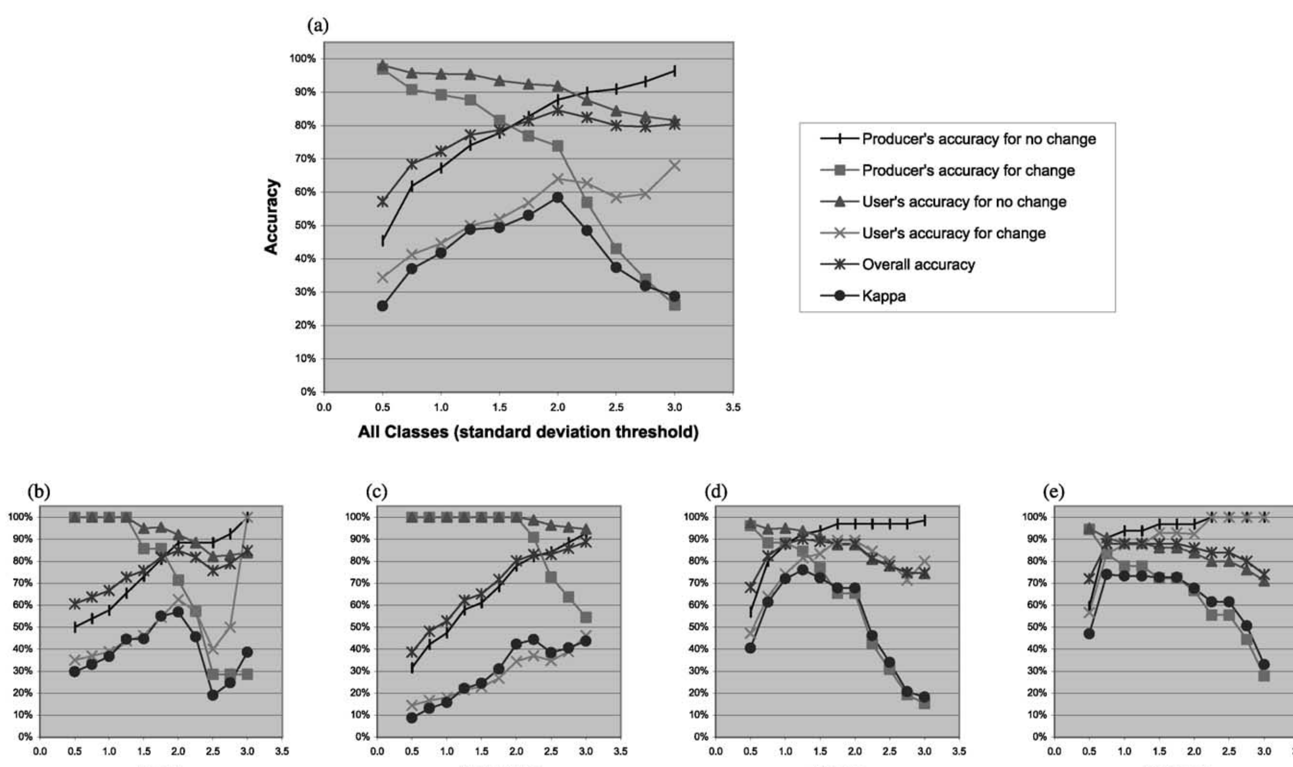

Fig. 2. Threshold assessment curves for standard deviations of NDVI differencing results.

invariant pixels is then ordered into spectral groups (bins). Then, the LRRN method normalized each source pixel locally (bin) on a pixel-by-pixel basis using spectrally similar temporarily invariant neighboring pixels. In this analysis, the 1993 image was normalized relative to the 2000 image using eight bin levels and a minimum of 50 pixels/bin.

Imagery used to support the MID based analysis included the May 11, 2000 ETM$^+$ image and three versions of the May 16, 1993 TM image; non-normalized, ASCR, and LRRN normalized. For each image, a tasseled-cap transformation was applied creating brightness, greenness, and wetness bands. Standard tasseled-cap transformation coefficients were applied, thus negating the requirement for scene-specific locations representing the extremes in brightness, greenness, and wetness (Jackson, 1983). The brightness, greenness, and wetness bands from both dates of imagery were then used to calculate the total magnitude of change in each of three dimensions using the following equation.

$$\text{Magnitude of change (DN)} = [(\text{brightness 1993} - \text{brightness 2000})^2 + (\text{greenness 1993} - \text{greenness 2000})^2 + (\text{wetness 1993} - \text{wetness 2000})^2]^{1/2} \quad (1)$$

A noise image was calculated based on a systematic sampling of every tenth pixel in each companion image. For each pixel sampled, one of its eight neighbors was randomly selected and the digital DN mean and variance was calculated for all spectral bands (Dwyer et al., 1996). Difference magnitudes were divided by the estimated noise to create the modulated magnitude image. However, the noise correction variable tended to increase the magnitude of change for parts of the image where variability was extremely low, such as in open water areas. A modification to the original method outlined by Dwyer et al. (1996) was made to insure that the denominator would be ≥ 1.0 DN and therefore, not magnify differences in low noise areas. The denominator of the noise image was rescaled using the following equation.

$$\text{Modulated magnitude} = \text{magnitude(DN)}/(\text{total noise(DN)} + 1\text{(DN)}) \quad (2)$$

Lastly, a set of change images was created based on the application of selected change thresholds. MID magnitudes that exceeded established thresholds were considered change pixels. For example, a pixel threshold of 10 DNs would indicate that the magnitude of change calculated for a pixel would have to exceed the total noise(DN) + 1 by a factor of 10. Optimum thresholds were determined using calibration curves based on the same 340 reference data

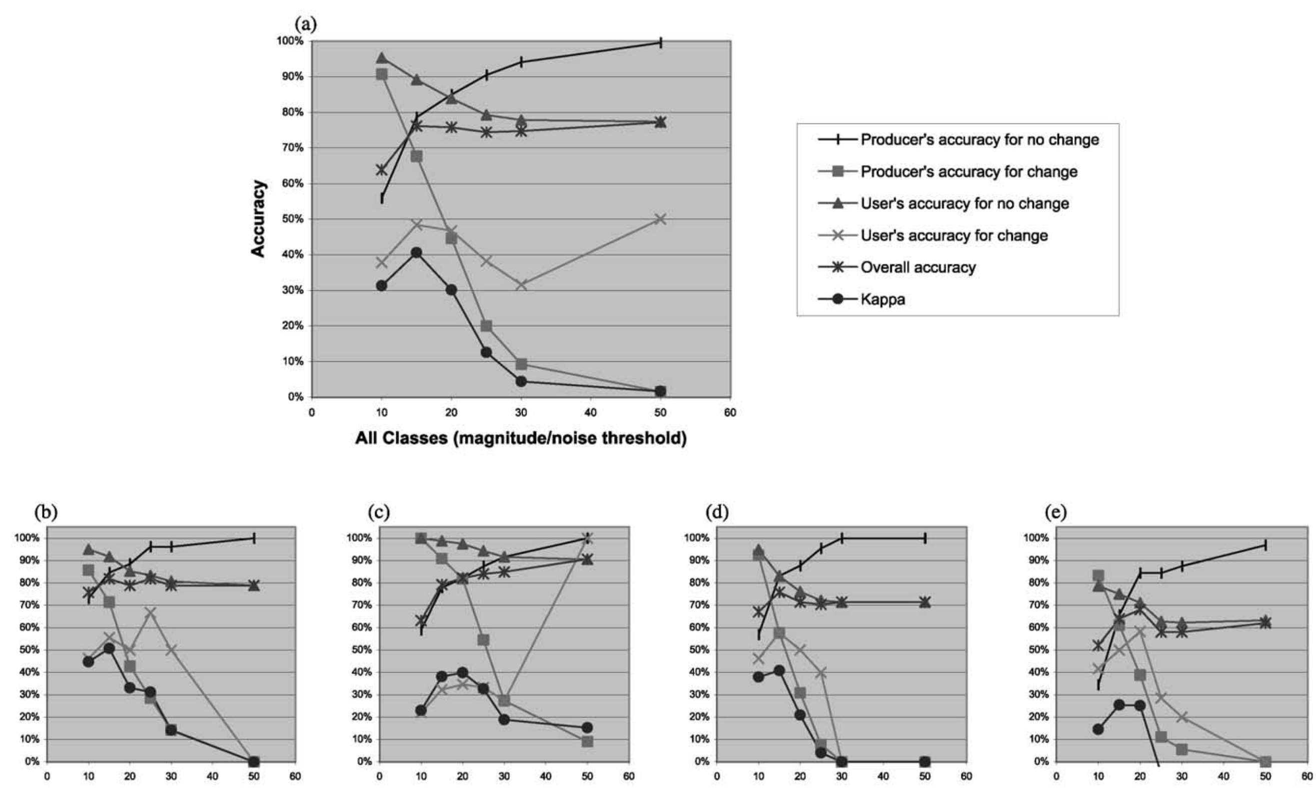

Fig. 3. Threshold assessment for MID with non-normalized data.

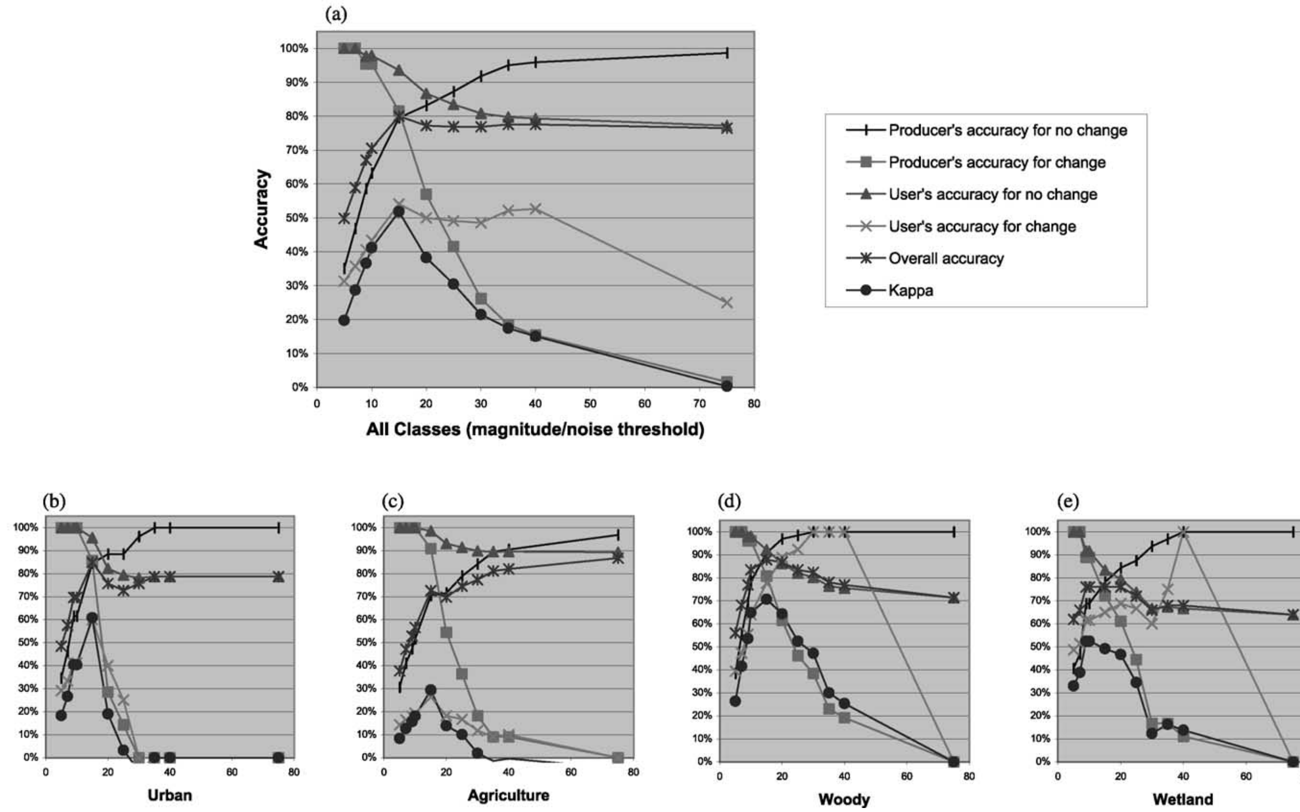

Fig. 4. Threshold assessment curves for MID with ASCR.

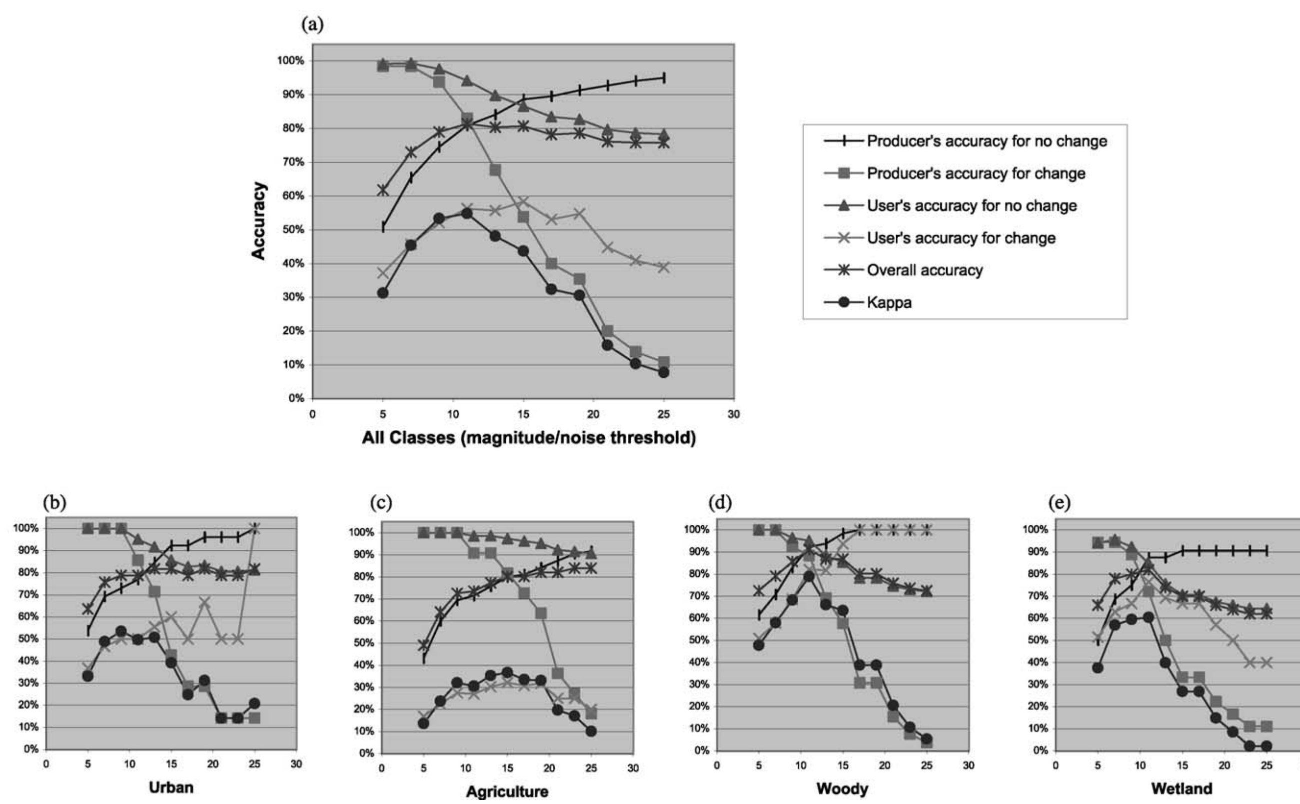

Fig. 5. Threshold assessment curves for MID with LRRN.

points described above and used in the NDVI thresholding. Calibration curves for the non-stratified (Figs. 3a, 4a, and 5a) and the stratified cases (Figs. 3b–e, 4b–e, 5b–e) were used to select the threshold for the final change mask based on the maximum Kappa. Table 2 summarizes the thresholds for each case and the estimated percentage of LC change.

3.5. Accuracy assessment of change detection images

An accuracy assessment was performed to evaluate the performance of the various change detection methods. Assessment points were collected independently from those used to construct the calibration curves that were applied to establish thresholds. A disproportionate stratified sampling design was employed (Khorram et al., 1999; Lunetta & Elvidge, 1998). Stratifications were performed using the 2.00 standard deviation NDVI change mask and the 1998–1999 NRB-LC classification. Determining the minimum sampling size was done according to Khorram et al. (1999) using the following formula.

$$n = [t^2 pq]/\delta^2 \tag{3}$$

where n is the sample size, t is the standard error, p is the probability of change, $q = 1 - p$, and δ is the error tolerance.

The standard error was estimated using the reference data set. An error tolerance of 0.05 was used and the probability of change was set to 25.0%. Because the variance in a population is highest when the probabilities for change and no-change approach 50% (Khorram et al., 1999), the use of a liberal change estimate insured an adequate sampling size for assessing the change detection. A total of 384 points, selected in proportion to the areal extent of each stratum, were interpreted to obtain sufficient reference data to adequately represent rare agriculture and barren change categories. Conditional probability measures including overall accuracy, omission, and commission were calculated for each of the change detection methods based on Stehman and Czaplewski's (1998) equations for a stratified sampling design. Pairwise comparisons to test for differences between methods were performed with contingency matrices using a Kappa test Z statistic (Congalton & Green, 1999). Kappa variances were calculated using the estimators proposed by Stehman (1996) for the stratified sampling design. Matrices were compared for both the overall change masks and for each of the LC classes. Because the same assessment points were used in generating each of the error matrices, the assumption of independence was not valid, thus the tests must be interpreted cautiously (Stehman, 1997).

Table 2
Best MID thresholds with predicted change estimates derived from reference data set utilizing non-normalized, ASCR normalized, and LRRN normalized data

Class	Peak Kappa	Magnitude/ (noise + 1) threshold	Strata change %	Image change %
Non-normalized				
All	0.41	15	N/A	20.5
Non-normalized stratified				
Urban	0.51	15	12.5	1.4
Agriculture	0.40	20	8.8	3.0
Woody	0.41	15	26.8	9.4
Water	N/A	15	59.5	0.8
Wetland	0.26	15	34.4	6.1
Barren	N/A	15	35.2	0.1
Non-normalized stratified total				20.8
ASCR				
All	0.52	15	N/A	18.2
ASCR stratified				
Urban	0.61	15	14.3	1.6
Agriculture	0.30	15	21.9	7.6
Woody	0.71	15	19.1	6.7
Water	N/A	15	37.5	0.5
Wetland	0.53	10	38.4	6.8
Barren	N/A	15	33.3	0.1
ASCR stratified total				23.3
LRRN				
All	0.55	11	N/A	16.0
LRRN stratified				
Urban	0.54	9	17.8	1.9
Agriculture	0.37	15	10.8	3.7
Woody	0.79	11	19.0	6.7
Water	N/A	11	53.9	0.7
Wetland	0.61	11	21.6	3.9
Barren	N/A	11	34.5	0.1
LRRN stratified total				17.0

4. Results

The overall and per class type accuracies of change/no-change for each of the eight methods tested are listed in Table 3. All methods did reasonably well when considering overall accuracy. The stratified MID ASCR method had the highest accuracy (91.7%, $n = 384$), and the stratified MID nonnormalized technique had the lowest (84.5%, $n = 384$). Because overall accuracies were calculated based on total areal extent, the value indicates the probability (e.g., 91.7% for the stratified MID ASCR) that a randomly selected pixel within the change image was correctly classified (Stehman & Czaplewski, 1998). Accuracies among individual LC types were generally 85.0% or higher, with the exception of the woody class that ranged between 62.5% and 83.4%. Because the woody class represented 35.1% of the area in the NRB (Table 1), depressed accuracies had a large impact on the overall change/no-change accuracies for the entire image. Among the other classes, the non-stratified NDVI was 100% accurate for both wetland ($n = 41$) and water ($n = 40$) classes. Change in urban LC was best measured using the MID LRRN stratified approach (92.9%, $n = 76$) while barren accuracy was high (97.0%, $n = 27$) for both

Table 3
Overall accuracy (%) based on the probability that any point is correctly classified

Change detection method	All	Urban	Agriculture	Woody	Water	Wetland	Barren
NDVI	90.5	87.2	90.1	79.1	100.0	100.0	91.1
NDVI stratified	91.3	87.2	89.2	83.4	100.0	99.4	93.0
MID	85.8	84.5	90.0	62.5	95.6	84.2	92.3
MID stratified	84.5	84.5	83.5	62.5	95.6	84.2	92.3
MID ASCR	91.2	90.6	89.5	82.4	98.7	92.9	93.8
MID ASCR stratified	91.7	90.6	89.5	82.4	98.7	99.0	93.8
MID LRRN	89.5	89.0	89.6	76.6	96.2	88.2	97.0
MID LRRN stratified	90.3	92.9	87.1	76.6	96.2	88.2	97.0
Assessment samples	384	76	154	46	40	41	27

MID LRRN methods. Agricultural change was best estimated with the non-stratified NDVI differencing method (90.1%, $n=154$). The agriculture and water classes had the lowest variability between methods with differences ranging just a few percentage points. The woody strata had the highest variability ranging more than 20.0% ($n=46$).

Commission and omission errors give greater insight into method performance than the percent accuracy values because the rate of change was relatively low compared to the overall areal extent over the 7-year change period (approximately 1.0% per year). The commission rates were high for all methods (Table 4). The NDVI stratified method had the lowest change commission error at 42.5%, while the non-normalized MID method reached 53.5%. These commission rates were primarily the result of high rates of change commission in the agriculture class. All methods resulted in greater than 82.0% change commission in agriculture. Since agriculture represents about 34.6% of the study area (Table 1), these errors had a large effect on overall commission rates. Among other class types, both NDVI methods performed well with overall change commission values below 15.4% in all cases. The MID LRRN methods demonstrated a marginal improvement for the urban class with omission error rates of 13.3% and 12.5%.

Commission rates for the no-change pixel predictions were much better than that for the change class (Table 4). The MID ASCR methods performed best with 5.8% and 5.3% no-change commission error for the strata combined. The MID stratified was the poorest performer with an error of 13.0%. Low no-change commission rates are found throughout the individual LC strata, with the notable exception of the woody class (16.7–38.1%). No-change commission for the remaining classes (urban, agriculture, water, wetland, and barren) were all less than 16.0% for each of the methods.

Omission probabilities were dramatically different between the change and the no-change strata (Table 5). Overall change omissions ranged from 59.8% (MID ASCR stratified) to 78.5% (MID stratified). Change omission error was always high within the urban, agriculture, and woody classes, regardless of method, with errors ranging from 49.1% to 91.3%. The NDVI and MID ASCR stratification methods did well at detecting omission change in the wetland class (0.0% error) and MID methods did well with

Table 4
Probability of change and no-change commission (%) for the eight change detection methods

Method	All	Urban	Agriculture	Woody	Water	Wetland	Barren
Change commission							
NDVI	44.4	13.8	83.5	4.3	0.0	0.0	15.4
NDVI stratified	42.5	13.8	82.4	14.3	0.0	11.5	8.3
MID	53.5	20.0	84.2	24.0	43.5	25.0	36.8
MID stratified	50.3	20.0	86.5	24.0	43.5	25.0	36.8
MID ASCR	50.2	15.6	85.5	17.2	13.3	15.4	29.4
MID ASCR stratified	50.4	15.6	85.5	17.2	13.3	20.7	29.4
MID LRRN	50.9	13.3	85.2	20.7	38.1	12.5	14.3
MID LRRN stratified	47.9	12.5	84.9	20.7	38.1	12.5	14.3
Assessment samples	123	31	17	27	13	23	12
No-change commission							
NDVI	7.0	12.8	1.8	21.7	0.0	0.0	7.1
NDVI stratified	6.3	12.8	2.9	16.7	0.0	0.0	6.7
MID	11.4	15.2	1.9	38.1	0.0	15.4	0.0
MID stratified	13.0	15.2	8.8	38.1	0.0	15.4	0.0
MID ASCR	5.8	9.1	2.3	17.6	0.0	6.7	0.0
MID ASCR stratified	5.3	9.1	2.3	17.6	0.0	0.0	0.0
MID LRRN	7.6	10.9	2.2	23.5	0.0	11.8	0.0
MID LRRN stratified	7.0	6.8	4.9	23.5	0.0	11.8	0.0
Assessment samples	261	45	137	19	27	18	15

Table 5
Probability of change and no-change omission (%) for the eight change detection methods

Method	All	Urban	Agriculture	Woody	Water	Wetland	Barren
Change omission							
NDVI	63.8	73.0	49.1	82.7	0.0	0.0	24.1
NDVI stratified	60.5	73.0	59.9	80.4	0.0	0.0	21.5
MID	77.5	77.7	52.0	91.3	0.0	80.4	0.0
MID stratified	78.5	77.7	85.5	91.3	0.0	80.4	0.0
MID ASCR	62.1	66.4	58.7	81.8	0.0	61.1	0.0
MID ASCR stratified	59.8	66.4	58.7	81.8	0.0	0.0	0.0
MID LRRN	68.5	69.7	57.1	86.2	0.0	72.8	0.0
MID LRRN stratified	65.4	58.8	74.8	86.2	0.0	72.8	0.0
Assessment samples	123	31	17	27	13	23	12
No-change omission							
NDVI	3.3	0.9	8.6	0.3	0.0	0.0	4.2
NDVI stratified	3.1	0.9	8.5	0.8	0.0	0.6	2.3
MID	4.1	1.3	8.6	1.8	4.6	1.5	8.9
MID stratified	4.0	1.3	9.5	1.8	4.6	1.5	8.9
MID ASCR	3.7	0.9	8.8	1.0	1.5	0.8	7.2
MID ASCR stratified	3.7	0.9	8.8	1.0	1.5	1.0	7.2
MID LRRN	3.8	0.8	8.8	1.3	4.1	0.7	3.7
MID LRRN stratified	3.5	0.7	9.0	1.3	4.1	0.7	3.7
Assessment samples	261	45	137	19	27	18	15

change omission in the barren class (0.0%). Overall omission values for the no-change category were very low (less then 10.0%) for all of the methods with little variance between methods among the individual strata. Because the overall scene was dominated by the no-change category, it means these low omission errors were a dominant factor in leading to the high overall accuracies. Because there was little variation between methods in these no-change omission accuracies, most of the overall classification accuracy differences were dependent on the no-change commission results (Table 4).

The stratified sampling estimator of the Kappa statistic can be used to rank the general performance of each change detection method tested (Table 6). Overall, the NDVI stratified methods did the best job of balancing omission and commission errors (Kappa = 0.42). MID non-normalized stratified methods performed the worst (0.23). When considering only individual LC classes, urban change was best monitored using the MID LRRN stratified methods (0.53). NDVI performed best for agriculture, but Kappa was low (0.21) given both its high rates of change omission and change commission errors. The woody class was nearly equally poor with its high levels of change omission and relatively high rates of no-change commission being reflected in Kappa values that range from 0.08 for the non-normalized MID method to a high of only 0.27 for the NDVI stratified method. The NDVI methods ranked highest for the water and wetland classes, with Kappa results ranging between 0.94 and 1.00. The MID ASCR stratified and LRRN methods performed equally best for the barren class (0.91). Of all the classes, wetland showed the most variance between methods ranging from a low of 0.26 to a high of 1.00.

The estimated variance of Kappa was used to determine differences (P=.05) between methods (Stehman, 1996). Table 7 shows the matrix for which methods are statistically different when accounting for all strata. Overall, the NDVI methods performed significantly better than the non-normalized MID methods. There were also significant differences between the ASCR MID stratified versus the MID non-normalized stratified, and between both the stratified and non-stratified non-normalized MID versus the MID ASCR stratified. The MID ASCR methods performed better than the non-normalized methods (Table 6). The MID non-

Table 6
Overall Kappa based on the probability that any point is correctly classified

Change detection method	All	Urban	Agriculture	Woody	Water	Wetland	Barren
1. NDVI	0.39	0.36	0.21	0.23	1.00	1.00	0.74
2. NDVI stratified	0.42	0.36	0.20	0.27	1.00	0.94	0.80
3. MID	0.24	0.29	0.20	0.08	0.70	0.26	0.73
4. MID stratified	0.23	0.29	0.05	0.08	0.70	0.26	0.73
5. MID ASCR	0.38	0.44	0.17	0.24	0.92	0.50	0.79
6. MID ASCR stratified	0.40	0.44	0.17	0.24	0.92	0.88	0.79
7. MID LRRN	0.33	0.40	0.18	0.17	0.75	0.37	0.91
8. MID LRRN stratified	0.37	0.53	0.12	0.17	0.75	0.37	0.91
Assessment samples	384	76	154	46	40	41	27

Table 7
Significant differences found between methods (×) including all strata combined

Change detection method	1	2	3	4	5	6	7	8
1. NDVI			×	×				
2. NDVI stratified			×	×				
3. MID								
4. MID stratified								
5. MID ASCR				×				
6. MID ASCR stratified			×	×				
7. MID LRRN								
8. MID LRRN stratified								

Methods listed in row headings are significantly better than counterparts listed in columns.

normalized methods performed significantly poorer than all other methods. When considering the image-wide change detection, there were no differences between the stratified and non-stratified cases.

Comparing the results for the individual strata by method showed no significant differences within the urban, agriculture, woody, or barren classes. However, there were significant differences within the water and wetland classes. NDVI non-stratified performed significantly better than all MID methods in classifying wetland change. The stratified MID ASCR performed significantly better than both of the MID LRRN and non-normalized MID methods. For the water class, there were no differences between stratified and nonstratified cases for any of the methods. Both NDVI methods did better than the MID non-normalized and LRRN methods for water classes. MID ASCR also performed better than the MID non-normalized method.

5. Discussion

Agriculture and woody class types represented the largest components of the change detection error budget in this study. Errors of change commission for agriculture were much greater than for the other classes (Table 4). The net result was an overestimate of change from agricultural to other LC type. Because LC variations occur at frequent temporal intervals, it is clear that change detection methods using simple image subtraction (i.e., NDVI) are not appropriate techniques for these agricultural settings. MID methods, relying only on magnitude of change, were equally poor at estimating true change for the agricultural class. Considering the direction of the change vector may help more accurately identify true change pixels within the agricultural class.

The woody class had relatively low commission errors but high omission errors (Tables 4 and 5). Thus, the conversion of the woody class to other LC class was underestimated by all methods tested. Additionally, within the woody class there was a great deal of spectral variability between the deciduous and evergreen cover types as well as between stands of varying ages. This confounding spectral variability and the existence of large patches of relatively homogeneous woody vegetation, resulted in the change detection efforts being sensitive to small shifts in selected thresholds.

The results of the cover type stratification based analysis were mixed. Stratification allowed us to identify and quantify the problem with respect to the agricultural and woody classes. It also improved change estimates for certain cover classes within this study. However, stratification did not significantly improve the overall or per class change detection results. This was contrary to the expected, but may be related to several factors. First, the NRB is dominated by two cover types (woody and agriculture) that represented approximately 70% of the total study area. Secondly, misclassification errors inherent to the LC map may have outweighed the gains of determining unique thresholds by cover type. Additionally, even a highly accurate LC map can yield temporal displacement errors. For example, the NRB-LC classification used in this study had a woody class accuracy of 85%, but the 1-year difference between the LC map and the 2000 image lead to several clear cuts being lumped in with the woody cover class.

The results of the normalization study supported the continued use of normalization as a part of the MID methodology. The non-normalized MID method performed poorly by most measures, and in some comparisons, it was significantly worse than the others. The woody class showed the greatest improvement with normalization, improving in its accuracy by over 10.0%.

NRB-LC change was estimated to be 0.5–1.0% or approximately 0.75% LC change per year. Thus, approximately a 5.3% change was thought to have occurred over the 7-year study period. All but one of the methods predicted a significantly greater amount of change than was expected (Tables 2 and 8). The MID methods all estimated 16.0% or greater LC change. The NDVI method estimated a plausible 6.7% and 11.5% change rates corresponding to the non-stratified and stratified methods, respectively. Based on the high rate of no-change commission errors, these rates appear to be overestimates. Eliminat-

Table 8
Best standard deviation thresholds with predicted change estimates for NDVI difference

Class	Peak Kappa	S.D. threshold	Strata change %	Image change %
NDVI				
All	0.58	2.00	N/A	6.7
NDVI stratified				
Urban	0.57	2.00	3.4	0.4
Agriculture	0.44	2.25	7.0	2.4
Woody	0.76	1.25	12.1	4.2
Water	N/A	2.00	10.0	0.1
Wetland	0.74	0.75	24.0	4.3
Barren	N/A	2.00	16.9	0.1
Stratified total				11.5

ing the agricultural LC class from the NDVI stratified computation resulted in a 9.1% LC change. The three MID methods: non-normalized, ASCR, and LRRN predicted 17.8%, 15.7%, and 13.3%, respectively.

6. Conclusions

A goal of this study was to evaluate the application of the NDVI difference and MID change detection methods using Landsat TM imagery in the biologically complex NRB North Carolina. Previous results in biologically complex ecosystems determined that the NDVI was likely the best performing of the VIs for change detection discrimination (Lyon et al., 1998). The results of this study indicated that NDVI difference performed the best, followed by MID ASCR. Although, MID ASCR stratified ranked highest in overall accuracy, the NDVI stratified performed significantly better based on the Kappa test. The non-normalized MID method performed the worst. For individual cover classes, the non-stratified NDVI performed best for accuracy with respect to woody, wetland, and water classes. MID LRRN stratified performed best with respect to urban and barren. The poor results obtained for woody vegetation could probably have been improved by increasing imagery temporal resolution.

The application of a two-date change detection approach was inadequate based on reported errors of commission and omission. These errors were largely attributed to annual changes in vegetation phenology (Type I error) and the rapid revegetation (Type II error) of areas undergoing LC conversion between imagery collection dates. The application of change vector directional information could facilitate a better identification of the true change areas. However, greater temporal imagery resolution would be required to support the analysis. Additional research is needed to determine minimal temporal imagery resolution (annual) needs to support CVA based change detection in biologically complex ecosystems. Additionally, a better understanding of intra-annual variations in vegetation phenology and the resulting contributions to change detection error budgets would facilitate the development of temporal data collection requirements.

Acknowledgements

The authors would like to acknowledge Dr. Steve Stehman for his statistical analysis assistance. The U.S. Environmental Protection Agency funded and conducted the research described in this paper. It has been subject to the Agency's programmatic review and has been approved for publication. Mention of any trade names or commercial products does not constitute endorsement or recommendation for use.

References

Belward, A. S., Estes, J. E., & Kline, K. D. (1999). The IGBP-DIS global 1 km land-cover data set discover: a project overview. *Photogrammetric Engineering and Remote Sensing*, *65*(9), 1013–1020.

Byrne, G. F., Crapper, P. F., & Mayo, K. K. (1980). Monitoring land-cover change by principal component analysis of multitemporal Landsat data. *Remote Sensing of Environment*, *10*, 175–184.

Cohen, W. B., & Fiorella, M. (1998). Comparison of methods for detecting conifer forest change with Thematic Mapper imagery. In R. S. Lunetta, & C. D. Elvidge (Eds.), *Remote sensing change detection: environmental monitoring methods and applications* (pp. 89–102). Chelsea, MI: Ann Arbor Press (318 pp. + illustrations, co-published in Europe by Taylor & Francis, UK).

Congalton, R. G., & Green, K. (1999). *Assessing the accuracy of remotely sensed data: principles and practice*. Boca Raton, FL: Lewis (137 pp.).

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

Crist, E. P., & Kauth, R. J. (1986). The tassel cap de-mystified. *Photogrammetric Engineering and Remote Sensing*, *52*(1), 81–86.

Dwyer, J. L., Sayler, K. L., & Zylstra, G. J. (1996). Landsat pathfinder data sets for landscape change analysis. *Proceeding of the International Geoscience and Remote Sensing Symposium, Lincoln, NE, May 27–31*, *1*, 547–550.

Ediriwickrema, J., Lunetta, R. L., Dulaney, D., McKerrow, A., & Johnson, D. M. (2002). *Localized relative radiometric normalization for change detection* (Internal Report). Research Triangle Park, NC: U.S. Environmental Protection Agency.

Eidenshink, J. C. (1992). The 1990 conterminous U.S. AVHRR data set. *Photogrammetric Engineering and Remote Sensing*, *58*(6), 809–813.

Elvidge, C. D., Miura, T., Jansen, W. T., Groeneveld, D. P., & Ray, J. (1998). Monitoring trends in wetland vegetation using a Landsat MSS time series. In R. S. Lunetta, & C. D. Elvidge (Eds.), *Remote sensing change detection: environmental monitoring methods and applications* (pp. 191–210). Chelsea, MI: Ann Arbor Press (318 pp. + illustrations, co-published in Europe by Taylor & Francis, UK).

Elvidge, C. D., Yuan, D., Weerakoon, R. D., & Lunetta, R. S. (1995). Relative radiometric normalization of Landsat Multispectral Scanner data using automatic scattergram-controlled regression. *Photogrammetric Engineering and Remote Sensing*, *61*, 1255–1260.

Fung, T., & LeDrew, E. (1988). The determination of optimal threshold levels for change detection using various accuracy indices. *Photogrammetric Engineering and Remote Sensing*, *54*, 1449–1454.

Heo, J., & FitzHugh, T. W. (2000). A standardized radiometric normalization method for change detection using remotely sensed imagery. *Photogrammetric Engineering and Remote Sensing*, *66*(2), 173–181.

Jackson, R. D. (1983). Spectral indices in *n*-space. *Remote Sensing of Environment*, *13*, 409–421.

Jones, B. A., Ritters, K. H., Wickham, J. D., Tankersley Jr., R. D., O'Neill, R. V., Chaloud, D. J., Smith, E. R., & Neale, A. C. (1997). *An ecological assessment of the United States mid-Atlantic region: a landscape atlas* (U.S. Environmental Protection Agency Report No. EPA/600/R-97/130, 104 pp.). Washington, DC: U.S. Printing Office.

Khorram, S. K., Biging, G. S., Chrisman, N. R., Colby, D. R., Congalton, R. G., Dobson, J. E., Ferguson, R. F., Goodchild, M. F., Jensen, J. R., & Mace, T. H. (1999). Accuracy assessment of remote sensing-derived change detection. *American Society for Photogrammetry and Remote Sensing, Monograph Series* (64 pp.).

Lambin, E. F., & Strahler, A. H. (1994a). Change–vector analysis in multispectral space: a tool to detect and categorize land-cover change processes using high temporal-resolution satellite data. *Remote Sensing of Environment*, *48*, 231–244.

Lambin, E. F., & Strahler, A. H. (1994b). Indicators of land-cover change for change–vector analysis in multitemporal space at coarse spatial scales. *International Journal of Remote Sensing*, *15*(10), 2099–2119.

Lillesand, T. M., & Keifer, R. W. (1979). *Remote sensing and image interpretation* (2nd ed.). New York: Wiley (721 pp.).

Loveland, T. R., Merchant, J. W., Ohlen, D. O., & Brown, J. F. (1991). Development of a land-cover characteristics data base for the conterminous U.S. *Photogrammetric Engineering and Remote Sensing*, *57*(11), 1453–1463.

Lunetta, R. S. (1998). Project formulation and analysis approaches. In R. S. Lunetta, & C. D. Elvidge (Eds.), *Remote sensing change detection: environmental monitoring methods and applications* (pp. 1–19). Chelsea, MI: Ann Arbor Press (318 pp. + illustrations, co-published in Europe by Taylor & Francis, UK).

Lunetta, R. S., Alvarez, R., Edmonds, C. M., Lyon, J. G., Elvidge, C. D., Bonifaz, R., García, C., Gómez, G., Castro, R., Bernal, A., & Cabrera, A. L. (2002). An assessment of NALC/Mexico land cover mapping results: implications for assessing landscape change. *International Journal of Remote Sensing* (in press).

Lunetta, R. S., Congalton, R. G., Fenstermaker, L. K., Jensen, J. R., McGwire, K. C., & Tinney, L. R. (1991). Remote sensing and geographic information system data integration: error sources and research issues. *Photogrammetric Engineering and Remote Sensing*, *57*(6), 677–687.

Lunetta, R. S., Ediriwickrema, J., Iiames, J., Johnson, D., Lyon, J. G., McKerrow, A., & Pilant, D. (2002). A quantitative assessment of a combined spectral and GIS rule-based land cover classification in the Neuse River Basin of North Carolina. *Photogrammetric Engineering and Remote Sensing* (in press).

Lunetta, R. S., & Elvidge, C. D. (Eds.) (1998). *Remote sensing change detection: environmental monitoring methods and applications*. Chelsea, MI: Ann Arbor Press (318 pp. + illustrations, co-published in Europe by Taylor & Francis, UK).

Lunetta, R. S., Lyon, J. G., Guindon, B., & Elvidge, C. D. (1998). North American Landscape Characterization: dataset development and data fusion issues. *Photogrammetric Engineering and Remote Sensing*, *64*(8), 821–829.

Lyon, J. G., Yuan, D., Lunetta, R. S., & Elvidge, C. D. (1998). A change detection experiment using vegetation indices. *Photogrammetric Engineering and Remote Sensing*, *64*(2), 143–150.

Morisette, J. T., & Khorram, S. (2000). Accuracy assessment curves for satellite-based change detection. *Photogrammetric Engineering and Remote Sensing*, *66*, 875–880.

North Carolina Environmental Management Commission (NCEMC) (1993). Neuse River Basin wide water quality management plan. Raleigh, NC: North Carolina Division of Water Quality Management (134 pp.).

Richards, J. A. (1984). Thematic mapping from multitemporal image data using the principal components transformation. *Remote Sensing of Environment*, *16*, 25–46.

Richardson, A., & Everitt, J. (1992). Using spectral vegetation indices to estimate rangeland productivity. *Geocarto International*, *1*, 63–67.

Singh, A. (1989). Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, *10*(6), 989–1003.

Stehman, S. V. (1996). Estimating the kappa coefficient and its variance under stratified random sampling. *Photogrammetric Engineering and Remote Sensing*, *62*(4), 401–407.

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment*, *62*, 77–89.

Stehman, S. V., & Czaplewski, R. L. (1998). Design and analysis for thematic map accuracy assessment: fundamental principles. *Remote Sensing of Environment*, *64*, 331–344.

Van Driel, N. (1990). Completion of the, 1990's national land cover data set. *Photogrammetric Engineering and Remote Sensing*, *67*(6), 650–662.

Vogelmann, J. E., Sohl, T., & Howard, S. M. (1998). Regional characterization of land cover using multiple sources of data. *Photogrammetric Engineering and Remote Sensing*, *64*(1), 45–57.

Weismiller, R. A., Kristof, S. J., Scholz, D. K., Anuta, P. E., & Momin, S. A. (1977). Change detection in coastal zone environments. *Photogrammetric Engineering and Remote Sensing*, *43*(12), 1533–1539.

Yang, X., & Lo, C. P. (2000). Relative radiometric normalization performance for change detection from multi-date satellite images. *Photogrammetric Engineering and Remote Sensing*, *66*(8), 967–980.