2011

Utilizing temporally invariant calibration sites to classify multiple dates and types of satellite imagery

Joe Fortier
Clark University

John Rogan
Clark University, jrogan@clarku.edu

Curtis E. Woodcock
Boston University

Daniel Miller Runfola
Clark University

Utilizing Temporally Invariant Calibration Sites to Classify Multiple Dates and Types of Satellite Imagery

Joe Fortier, John Rogan, Curtis E. Woodcock, and Daniel Miller Runfola

Abstract

Mapping past time periods (retrospective mapping) using remotely sensed data is hindered by a lack of coincident calibration and validation information. The identification of features of same ground cover invariant across time and their use as calibration and validation data addresses this challenge by: (a) streamlining the process of image calibration for multiple dates, and (b) allowing each image to generate its own spectral signature. This study investigates the use of temporally invariant calibration and validation data to map land-cover in Massachusetts, employing five satellite images collected from five separate dates and different sensors. The results indicate that this technique can be used to produce land cover classifications of similar overall map accuracy to published mapping studies. Classification accuracy using this method is highly dependent on the characteristics (radiometric, spectral, and spatial) of the satellite imagery.

Introduction

Land-cover mapping is essential for monitoring the biosphere and contemporary global change (Turner *et al.* 1995). The application of remote sensing technology to inventory forest cover and productivity (Wulder, 1998), map post-wildfire burn severity (Miller and Yool, 2002), identify urban expansion (Ji *et al.*, 2001; Weng, 2001), and monitor desertification (Tripathy *et al.*, 1996) has proven the worth of satellite imagery as an effective data source to monitor past and future landscape conditions (Turner *et al.*, 1995). Further, satellite records extend back over 30 years, with data sources ranging from MSS to Landsat ETM+. However, one of the limitations to utilizing this dataset in land-cover mapping is the lack of ground reference data from past date(s) of imagery to be used in calibration and validation of mapping products (Rogan and Chen, 2004). This paper proposes a new methodological approach to overcome this limitation.

With an increase in availability of fine spatial resolution and hyperspectral sensors, the quantity of remotely sensed data is continuously increasing (Rogan and Chen, 2004). These new data sources are building on the wealth of current and historical information gathered from longer running medium spatial resolution (20 to 30 m) sensors that provide the means for numerous regional operational mapping and monitoring programs (see Franklin and Wulder, 2002). Remote sensing image libraries, such as those acquired by the Landsat mission over the past 30+ years, allow practitioners to create regional scale map products with a variety of temporal resolutions (Goward and Williams, 1997), an attribute which is critical for large scale monitoring efforts (Hill *et al.*, 1995). Many of these libraries have also recently adopted policies which allow free access to satellite imagery, greatly enhancing the general availability of imagery for analysis all around the globe (e.g., the US Landsat archive[1], China and Brazil's China-Brazil Earth Resources Satellite (CBERS)[2], and planned free data for the European Space Agency's Sentinel-2 program[3]). Categorical land-cover maps which can be produced from satellite imagery from sources such as CBERS, Landsat, and Sentinel are useful to identify land change either directly through post-classification comparison (Jensen *et al.*, 1995), or indirectly using spectral-based change mapping techniques (Cohen and Fiorella, 1998; Singh, 1989). The creation of land-cover maps for land change mapping is often hindered by a lack of ground reference information and/or aerial photography that is temporally explicit to the date of remotely sensed imagery (Washington-Allen *et al.*, 2006). Therefore, a dearth in temporally pertinent calibration and validation data is a constraint when any classification approach is undertaken. To compensate for the scarcity of calibration and/or validation data, methodologies such as signature extension, whereby algorithmic calibration is performed on one date or location of imagery, and the resultant signatures are extended to a second date or location of imagery, have been documented and to some degree explored (Olthof *et al.*, 2005).

Signature extension methods, also called generalization, have been primarily adopted to classify across image scenes/locations (Minter, 1978; Iverson *et al.*, 1994; Fazakas and Nilsson, 1996; Olthof *et al.*, 2005). Within-scene,

[1] *http://glovis.usgs.gov/*
[2] *http://www.cbers.inpe.br/en/index_en.htm*
[3] *http://www.esa.int/esaEO/SEMXK570A2G_environment_0.html*

Joe Fortier, John Rogan, and Daniel Miller Runfola are with Clark University, Graduate School of Geography, 950 Main Street, Worcester, MA 01610 (jrogan@clarku.edu).

Curtis E. Woodcock is at Boston University, Department of Geography, One Silber Way, Boston, MA 02215.

Photogrammetric Engineering & Remote Sensing
Vol. 77, No. 2, February 2011, pp. 181–189.

0099-1112/11/7702–0181/$3.00/0

temporal generalization methods have been explored to a lesser extent but have been shown to produce mapping products of comparable accuracy to those created through traditional classification techniques (Woodcock *et al.*, 2001). However, signature extension has been proven problematic when presented with differences in image radiometric properties, variable atmospheric effects across images, and spatial/temporal variation in vegetative phenology (Pax-Lenney *et al.*, 2001; Olthof *et al.*, 2005; Rogan *et al.*, 2002).

In an attempt to overcome these problems, this paper proposes the identification and use of temporally invariant ground features as calibration and validation data for the classification of past dates of imagery (retrospective mapping) for which ground data are not available. It is illustrated that this approach addresses many of the challenges associated with retrospective mapping, while overcoming the limitations of signature extension by: (a) only requiring one calibration and validation dataset for all images, and (b) allowing each image to generate its own signature characteristics specific to its inherent spectral and spatial properties. Further, the proposed methodology provides a cost and time efficient method to classify images of identical spatial extents from multiple sensors and points in time.

To investigate the utility of this new methodology, images from five separate satellite platforms and five separate dates spanning a 33 year period between 1973 and 2006 are selected for land-cover classification. Using the imagery, temporally invariant samples are identified to create an invariant data set. This invariant data set is then applied to classify land-cover in each image to test the plausibility of using invariant calibration data in long term retrospective and cross-satellite sensor land-cover mapping.

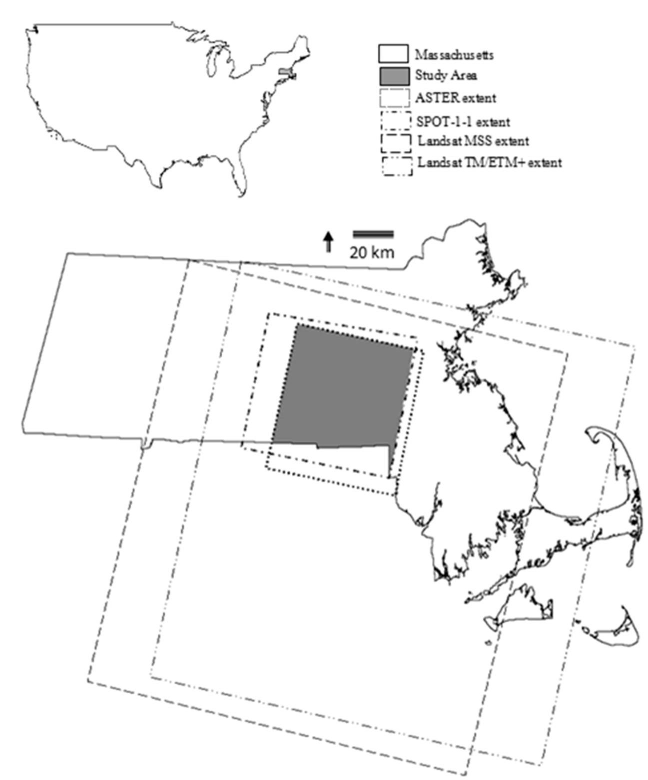

Figure 1. Map of the study area and rationale for study area selection.

Study Area and Data

The study area encompasses a 2,800 km^2 portion of south-central Massachusetts containing the metropolitan city of Worcester and surrounding towns. The region is dominated by heterogeneous mixture of needleleaved and broadleaved forests; herbaceous lands used for horse pasture, crop based agriculture, urban parkland, golf courses, and recreational fields; sparsely vegetated land including bare soil, sand quarries, and impervious surfaces from residential and commercial developed areas; and numerous standing water features such as lakes, rivers and reservoirs. Land change in this region over the last 30 years has been predominately suburban expansion from urban centers such as the cities of Boston and Worcester into areas previously occupied by forest and agriculture (DeNormandie, 2009; Breunig, 2003). The study region's boundaries represent the intersecting footprints of all five remotely sensed datasets used in this study (Figure 1).

Five satellite images were acquired for the study area including three Landsat scenes from 1973, 1987, and 1999 representing Landsat-Multispectral Scanner (MSS), Thematic Mapper (TM), and Enhanced Thematic Mapper Plus (ETM+) data, respectively, as well as a Satellite Pour l'Observation de la Terre-1 (SPOT-1) scene from 2001 and an Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) scene from 2006. All imagery was captured during the month of July with the exception of the 1987-TM image, which was captured during the early fall. Figure 2 highlights the differences in spectral resolution between each remote sensing platform in the visible through shortwave infrared spectrum. Acquisition resolution varied across the sensors, including 60 m (1973-MSS image), 20 m (SPOT-1 image), and 15-m (bands 1 to 3 of ASTER); the rest of the data were acquired at 30 m resolution. All data were resampled to 30 m resolution using a nearest neighbor resampling algorithm to facilitate their compatibility with all other data sets (Table 1).

Methods

Data Preprocessing

All data were projected into the Massachusetts State Plane coordinate system to minimize the degree of positional distortion resultant from projection into one of the two UTM zones (18 and 19 north) spanned by the study area. Atmospheric correction was performed on all spectral images using the COS(T) dark object subtraction method (Chavez, 1996). Ground location comparison was performed visually between the satellite data and a Massachusetts roads layer (1:100 000 scale) to determine if geometric correction was necessary. The 1973-MSS and 1999-ETM+ images were found to match well with the roads. Geometric correction was performed on the 1987-TM image using 39 ground control points with a global Root Mean Square Error (RMSE) of 11 m; on the 2006-ASTER image using 23 ground control points with a RMSE of 8 m; and to the 2001-SPOT-1 image using 23 ground control points with a RMSE of 9 m. These RMSEs were deemed tolerable for analysis on the basis of existing literature (Rogan and Chen, 2004; Jensen, 1996), although there is some discussion as to ideal tolerances (Townshend *et al.*, 1992).

Classification Scheme

Table 2 describes the land-cover categories examined in this study. Classes include needleleaf forest, broadleaf forest, herbaceous, non-forested wetlands, low intensity impervious (between 50 and 75 percent), high intensity impervious (>75 percent), and water. These seven categories are chosen based on prior knowledge of the study area, selecting the most common and dominant land-cover types which could be easily identified in both orthophotography and spectral

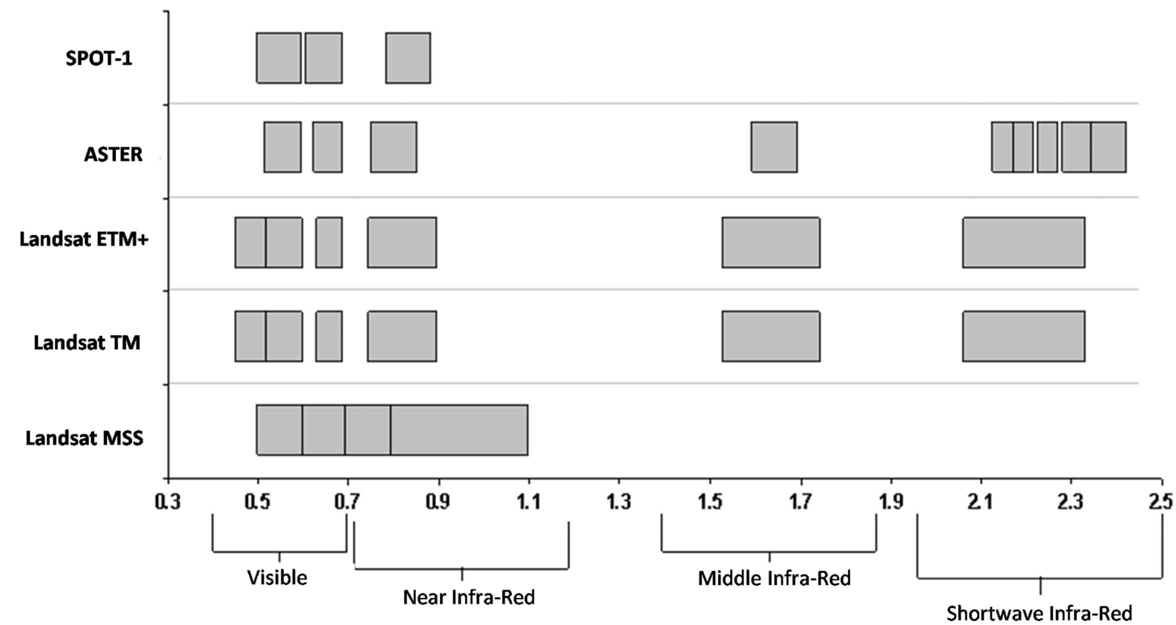

Figure 2. Bandwidths and electromagnetic wavelengths collected by all sensors used in this study.

TABLE 1. SPECTRAL AND ANCILLARY DATA USED IN CLASSIFICATION

Spectral Data	Resolution	Platform
23/07/1973: Bands 4–7	60 m	Landsat-1 MSS
16/09/1987: Bands 1–5 and 7	30 m	Landsat-5 TM
07/07/1999: Bands 1–5 and 7	30 m	Landsat-7 ETM+
20/07/2001: Bands 1–3	20 m	SPOT-1
10/07/2006: Bands 1–9	15/30 m	ASTER

TABLE 2. DEFINITIONS USED FOR MAP CLASSIFICATIONS

Category	Description
1. Needleleaf Forest	Forested land ≥50% needleleaf conifer canopy cover
2. Broadleaf Forest	Forested land ≥50% broadleaf deciduous canopy cover
3. Herbaceous	Managed or open grass, pasture, and cultivated lands
4. Wetland	Vegetated lands (sparse forest, shrub, or herbaceous) with water table present at or near surface
5. 50 to 75% Impervious	>50% and <75% presence of impervious surfaces (build-up)
6. >75% Impervious	≥75% presence of impervious surfaces (build-up)
7. Standing Water	Standing water present >11 months

color composites. For example, the split between low intensity and high intensity impervious was chosen to facilitate the distinction between urban core and suburban areas throughout the study area, while the wetlands category is included in this scheme based on its ecological importance to the landscape of Massachusetts (Brooks and Hayashi, 2002).

Ground Reference Data

The following steps were taken to identify invariant features using the 1973-MSS (earliest image) and 2006-ASTER (latest image) data:

1. The images from 1973 (MSS) and 2006 (ASTER) were overlaid, both displayed in a near-infrared false color composite for visual interpretation purposes.
2. Visual assessment was performed to identify static land-cover by looking for areas that exhibit generally static reflectance properties as well as static feature shapes. Good examples were found in established urban areas, large homogeneous patches, and conservation areas (wetlands and forest). Invariant reflectance values ranged between 1 to 5 percent in general.
3. Once an invariant location was indentified, a cover category was attributed based on 1-meter color orthophotographs collected by the Massachusetts Office of GIS (MassGIS; **http://www.mass.gov/mgis/**) in April 2000.
4. For further verification that a land-cover is truly invariant over time a visual inspection of imagery from the intermediate time steps (1987-TM and 1999-ETM+) was used to show no temporary changes in state occurred. Additionally where available a comparison of GIS land-use information from 1971 and 1999, acquired from the Massachusetts Department of GIS, was performed (Figure 3).

Invariant site calibration was performed by digitizing 3,600 m^2 polygons on the photographs, each representing at least four pixels at 30 m spatial resolution. These polygons were identified in center regions of large, contiguous invariant features (e.g., inside large forest stands, the middle of agricultural fields, and high density urban centers) to avoid edge conditions where land change is most likely to occur (Rogan and Miller, 2006). A total of 269 calibration samples were collected for the study area, representing 42 sites for needleleaf forest, 40 sites for broadleaf forest, 36 sites for herbaceous, 34 sites for non-forested wetland, 41 sites for low intensity impervious, 39 sites for high intensity impervious, and 37 sites for water. Figure 4 presents the spectral response curves of the seven land-cover categories.

An independent validation dataset was produced based on a stratified random sampling design using individual 30 m × 30 m pixels as the sampling unit and the seven

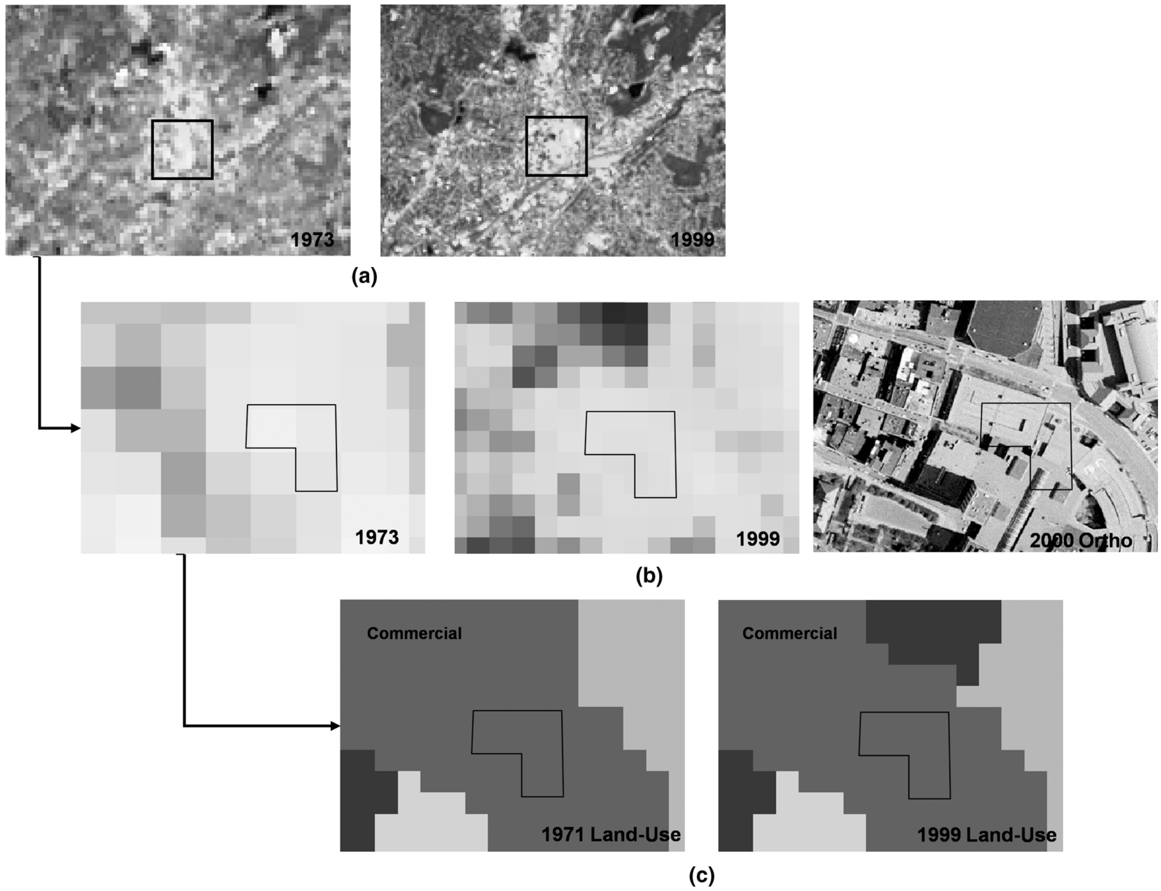

Figure 3. Steps to identify pseudo-invariant features for retrospective mapping:. (a) identifying invariant features, e.g., Worcester business district, (b) draw reference polygon and assigning a land-cover identification, e.g., bare land, presence of build up >60 percent, and (c) verification of feature invariance using land-use data.

land-cover categories as strata. Strata locations were derived from a circa 2000 land-use/land-cover map with an overall map accuracy of 87 percent (MaFOMP, 2009). Sixty-five random points were generated in each stratum as potential validation locations. Each point location was explored to verify: (a) correct class definition using the 1 m orthophotography, and (b) that the land-cover at the site did not change over the 33 year study period so that it could be applied to correctly validate all five years of classified imagery. Sample points located in areas where land-cover change had occurred were subsequently removed from the validation set. After removing points that were identified as variant across images, 358 calibration/validation samples remained, ranging from 42 to 59 pixels per class (Table 3).

Classification
A See5 classification tree algorithm (CTA) was used for image classification (Quinlan, 1996). CTA uses binary splitting to develop rules which maximize homogeneity at each split and continues recursively until all data are categorized into homogenous classes (Friedl and Brodley, 1997). The CTA was chosen due to its non-parametric nature and proven robustness in land-cover/land-use classification over traditional classifiers such as Maximum Likelihood, as well as for its speed, transparency of data partitioning decisions, and ease of use (DeFries and Chan, 2000; Pal and Mather, 2003).

Classification Evaluation
Error matrices (see Tables 5 through 9) were generated to calculate overall accuracy measures that provide evaluations of map quality. Individual class error (omission and commission) was explored to identify areas of classification difficulty in the individual map products.

Since a stratified sampling strategy was used in this analysis, it was necessary to weight the proportion of correctly classified pixels per stratum according to the areal proportion of the landscape encompassed by each stratum. This will minimize the effect of high map accuracies caused by homogeneous classes (e.g., water) which encompass only small portions of the study area. Proportion-weighted overall accuracy was calculated:

$$\sum\left(\frac{X_k}{Z_k} * \frac{P_k}{P}\right) \tag{6}$$

where X_k is the number of correctly classified samples in strata k, Z_k is the total number of samples in strata k, P_k is the total study area covered by strata k, and P is the total area of the landscape. The inverse student-t was used to determine the confidence interval of the overall accuracy values using the proportion of the study area covered by the validation plots. For example, the study area is 100,000

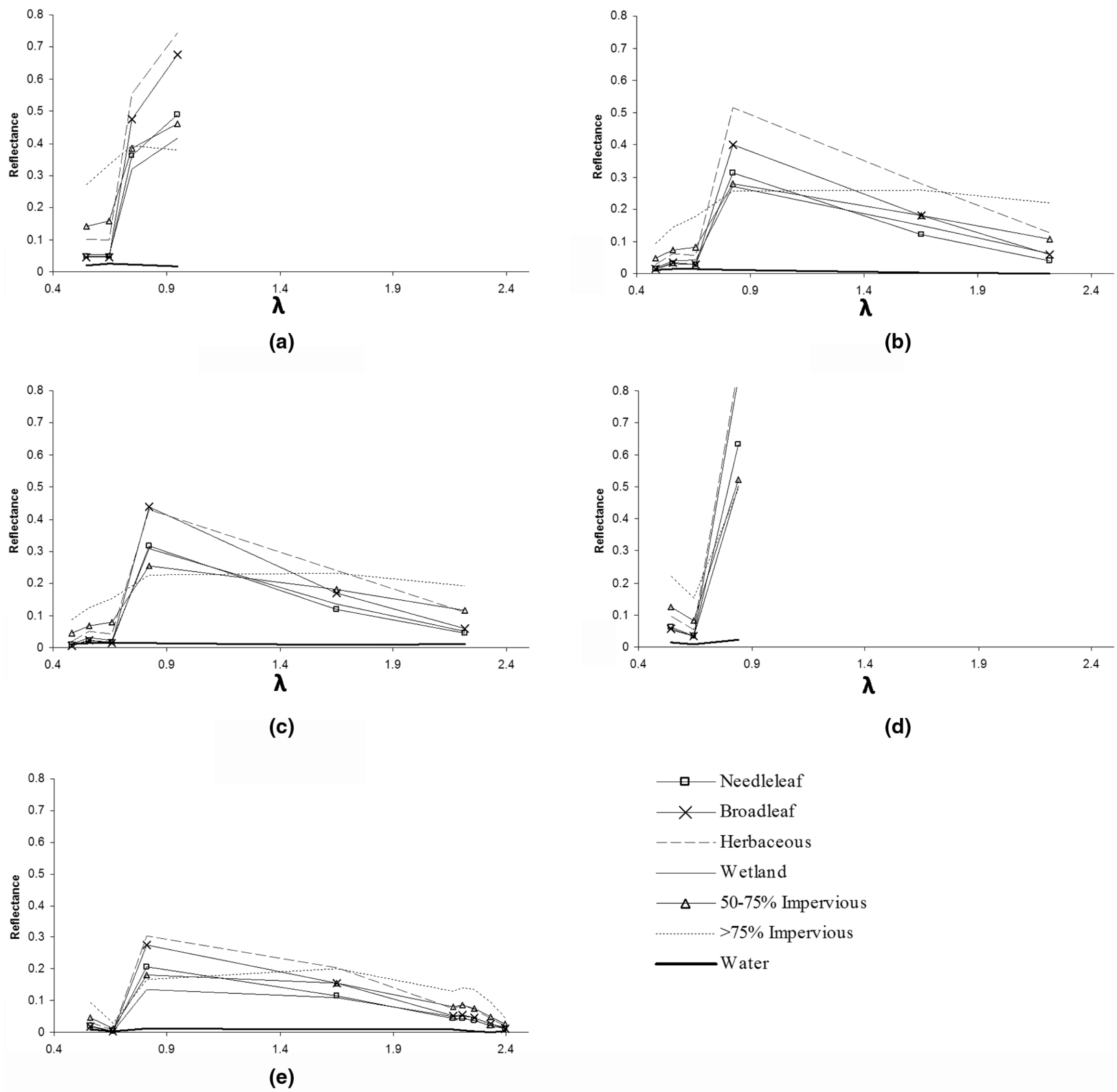

Figure 4. Mean spectral response curves of the seven land-cover categories: (a) 1973-MSS, (b) 1987-TM, (c) 1999-ETM+, (d) 2001-SPOT, and (e) 2006-ASTER.

TABLE 3. NUMBER OF CALIBRATION AND VALIDATION SAMPLES ACQUIRED FOR EACH CLASS, AND THE ESTIMATED PROPORTION OF STUDY AREA PER CLASS

Class	Estimated Proportion of study area	Stratified Random Calibration Samples	Stratified Random Validation Samples
Needleleaf	4%	168	42
Broadleaf	37%	160	57
Herbaceous	14%	144	44
Wetland	8%	136	51
Impervious (50 to 75%)	24%	164	59
Impervious (>75%)	7%	157	58
Water	6%	148	47

TABLE 4. OVERALL ACCURACY AT THE 95 PERCENT CONFIDENCE LEVEL FOR MAPS CREATED USING INVARIANT SAMPLE SITES

Classification Year	Overall Accuracy (95% confidence)	Proportion-weighted Overall Accuracy (95% confidence)	Overall Kappa
1973	62 to 72%	59 to 82%	.62
1987	83 to 90%	81 to 97%	.94
1999	81 to 88%	80 to 95%	.82
2001	75 to 83%	71 to 91%	.75
2006	83 to 90%	75 to 94%	.84

square kilometers and only 3,000 square kilometers were sampled and validated. Instead of assuming the accuracy reported by those 3,000 square kilometers is exactly the same over the rest of the study area we determined a confidence interval at ±5 percent and the reported the accuracy (in the case of the 1973 classification) as a range of 62 to 73 percent (where the derived overall accuracy was 67 percent). Following much of the existing literature the overall kappa values for each map were calculated (see Congalton and Mead, 1983; Congalton, 1991; Congalton and Green, 2009) (see Table 4), although use of the kappa statistic has recently

Table 5. Error Matrix for the 1973 (MSS) Classification (Validation Data in Columns)

		Validation							
		Need.	Broad.	Herb.	Wetl.	Imp (L)	Imp (H)	Water	Total
	Needleleaf	**30**	9	1	5	0	0	0	45
	Broadleaf	2	**42**	3	11	1	0	0	59
M	Herbaceous	0	2	**35**	24	4	0	0	65
a	Wetland	10	4	1	**9**	6	0	0	30
p	Impervious (L)	0	0	4	1	**43**	24	0	72
	Impervious (H)	0	0	0	1	5	**34**	0	40
	Water	0	0	0	0	0	0	**47**	47
	Total	42	57	44	51	59	58	47	358

Table 6. Error Matrix for the 1987 (TM) Classification (Validation Data in Columns)

		Validation							
		Need.	Broad.	Herb.	Wetl.	Imp (L)	Imp (H)	Water	Total
	Needleleaf	**38**	3	0	1	0	0	0	42
	Broadleaf	2	**52**	1	6	0	0	0	61
M	Herbaceous	0	0	**38**	11	0	0	0	49
a	Wetland	2	1	4	**33**	1	0	4	45
p	Impervious (L)	0	1	0	0	**57**	10	0	68
	Impervious (H)	0	0	1	0	1	**48**	0	50
	Water	0	0	0	0	0	0	**43**	43
	Total	42	57	44	51	59	58	47	358

Table 7. Error Matrix for the 1999 (ETM+) Classification (Validation Data in Columns)

		Validation							
		Need.	Broad.	Herb.	Wetl.	Imp (L)	Imp (H)	Water	Total
	Needleleaf	**42**	3	0	4	0	0	0	49
	Broadleaf	0	**50**	0	8	0	0	0	58
M	Herbaceous	0	2	**41**	16	1	0	0	60
a	Wetland	0	2	0	**23**	0	0	5	30
p	Impervious (L)	0	0	3	0	**57**	11	0	71
	Impervious (H)	0	0	0	0	1	**47**	0	48
	Water	0	0	0	0	0	0	**42**	42
	Total	42	57	44	51	59	58	47	358

Table 8. Error Matrix for the 2001 (SPOT) Classification (Validation Data in Columns)

		Validation							
		Need.	Broad.	Herb.	Wetl.	Imp (L)	Imp (H)	Water	Total
	Needleleaf	**34**	6	0	9	0	0	0	49
	Broadleaf	3	**45**	1	3	0	0	0	52
M	Herbaceous	0	1	**32**	5	0	0	0	38
a	Wetland	5	3	1	**29**	0	0	4	42
p	Impervious (L)	0	2	10	3	**56**	14	0	85
	Impervious (H)	0	0	0	2	3	**44**	0	49
	Water	0	0	0	0	0	0	**43**	43
	Total	42	57	44	51	59	58	47	358

Table 9. Error Matrix for the 2006 (ASTER) Classification (Validation Data in Columns)

		Validation							
		Need.	Broad.	Herb.	Wetl.	Imp (L)	Imp (H)	Water	Total
	Needleleaf	**38**	8	0	1	0	0	0	47
	Broadleaf	1	**44**	2	7	0	0	0	54
M	Herbaceous	0	2	**34**	3	0	0	0	39
a	Wetland	3	0	0	**39**	0	0	3	45
p	Impervious (L)	0	3	3	1	**57**	5	0	69
	Impervious (H)	0	0	5	0	2	**53**	0	60
	Water	0	0	0	0	0	0	**44**	44
	Total	42	57	44	51	59	58	47	358

been the topic of some debate (Pontius and Millones, 2008). In addition to examining these quantitative results, a qualitative visual review of the map products was utilized to identify spatial errors quantitative analysis may not reflect. This qualitative examination is particularly important for this methodology, as temporally invariant pixels are frequently "pure" cases, the exclusive selection of which can artificially inflate map accuracy (Rogan *et al.*, 2008).

Results

Table 4 presents both the unweighted and proportionally weighted overall map accuracy results for each year of classified imagery using a 95 percent confidence level. Proportionally weighted overall accuracy between the five maps ranged from 59 percent to 97 percent, with the lowest overall accuracies produced by the 1973-MSS data (59 percent to 83 percent, at a 95 percent confidence level) and highest by the 1987-TM data (81 to 97 percent, at a 95 percent confidence level). Figure 5 presents the Producer's Accuracy for each map and class, which ranged between 70 to 100 percent for all classes with the exception of the wetlands category in the majority of cases and >75 percent impervious in the 1973 map only. The wetlands category produced the lowest class accuracies and had the highest range of variability in accuracy between the maps ranging from 18 percent (1973-MSS) to 76 percent (2006-ASTER). The water and 50 to 75 percent impervious categories produced the highest accuracies ranging from 89 to 100 percent (water) and 73 to 97 percent (50 to 75 percent impervious).

Even though the statistical accuracy is similar between some maps, in particular 1987-TM, 1999-ETM+, and 2006-ASTER, qualitative analysis shows clear difference in the spatial configuration of the maps. The major visual inaccuracies on all five maps can be attributed to the over classification and/or poor placement of the wetlands category. Most prominently in the 2006-ASTER image wetlands are over-classified. The 1973-MSS and 1999-ETM+ over-classified wetlands in smaller patches spread across the image, and also over-classified small patches of herbaceous into forested areas giving the illusion of a more heterogeneous landscape than actually exists. The 2001-SPOT image is less affected by over-classification of wetlands and herbaceous, but tended to over-classify "50 to 75 percent impervious" in the southwest portion of the image.

Discussion and Conclusions

Retrospective mapping using remotely sensed data is currently hindered by a lack of temporally explicit calibration and validation information. The purpose of this paper is to propose a solution to the lack of coincident calibration and validation sites in retrospective mapping: the identification and use of features of "same ground cover" which are invariant across time. This paper has addressed the challenges of retrospective mapping by: (a) streamlining the process of image calibration for multiple dates, and (b) allowing each image to generate its own signature.

The lowest overall map accuracies resulted from the two sensors most limited in radiometric, spectral, and spatial resolution (MSS and SPOT-1). The results from the MSS image, which were lower than all other classifications, are most likely hindered by spatial and radiometric resolution of the sensor. The coarser 60 m spatial resolution of MSS causes mixing of land-cover types within each pixel, making it harder to identify pure examples of each class, particularly within heterogeneous classes such as herbaceous and 50 to 75 percent impervious. The 6-bit radiometric resolution of MSS likely introduced confusion between similar classes for which seperability is important for accuracy (e.g., 50 to 75 percent impervious and >75 percent impervious). Both the MSS and SPOT sensors are likewise hindered by the lack of shortwave infrared bands, which provide information on moisture content in vegetation and could enhance separability between the vegetated classes that account for approximately 63 percent of the study area. A review of classification tree splitting rules from the TM, ETM+, and ASTER classification indicate that the shortwave infrared (TM/ETM+ bands 5 and 7 and ASTER bands 4 through 9) aided in the distinction between needleleaf forest and wetlands, herbaceous and 50 to 75 percent impervious, and between 50 to 75 percent impervious

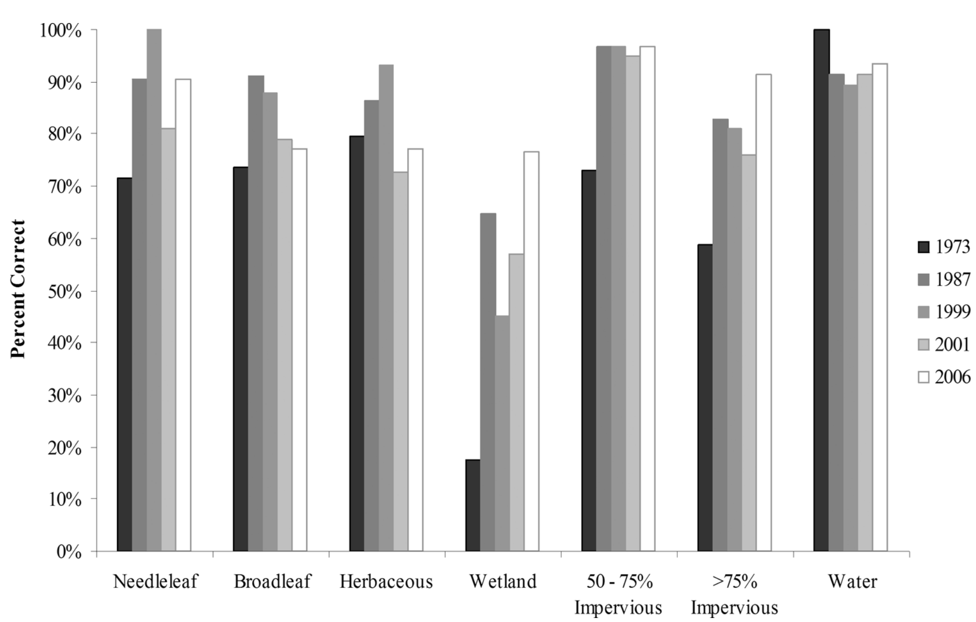

Figure 5. Per-class Producer's Accuracy for all five map products.

and >75 percent impervious (Figure 6). Seperability of classes, assessed by reviewing the spectral response curves for each class, shows a high similarity between the needleleaf and wetlands classes in all bands except for the SWIR. This suggests that accuracy in sensors omitting SWIR bands is driven by sensor quality, rather than classification methodology. Conversely, the success of the 1987-TM image is most likely due to the date of image capture, as previous research indicates that imagery showing early to mid-senescent stages of vegetation phenology (e.g., September/October in Northeast United States) can increase the separability of vegetation classes and improve map accuracy (Kalensky and Scherk, 1975; Schriever and Congalton, 1995). All other images were captured in the month of July, which represents the height of vegetation phenology in this region.

The per-class accuracy results show that the classified maps which produced the highest and lowest class accuracies are highly variable, and no single set of imagery outperformed the others for every class. The 1973-MSS data clearly produced the lowest per-class accuracies in the majority of categories, which is concurrent with the poor results of the 1973-MSS imagery in the overall accuracy comparison. Per-class accuracies achieved by the other four dates of imagery fell between 73 percent to 100 percent. The exception to this case is the wetlands category which saw accuracies as low as 18 percent (1973-MSS) and only as high as 76 percent (2006-ASTER). The poor accuracy of the wetlands class could be primarily attributed to the fact that the wetland class refers to a state of land-cover and is not itself a land-cover type, as these areas frequently transition between dry and wet states. There is high variability within this class (e.g., how wet the land is in any particular area corresponds to a high variability in spectral response), leading to confusion with all three of the other major vegetation categories: needleleaf forests, broadleaf forests, and herbaceous. Other than wetlands, the majority of class confusion in the classified products occurred predominantly between like classes such as: needleleaf and broadleaf forests, 50 to 75 percent impervious and >75 percent impervious, herbaceous and broadleaf forests, and water with wetlands, all classes with semantic similarities.

While selecting temporally invariant calibration sites to classify multiple images can be applied to both retrospective and current datasets, the results of each map will be restricted by the spatial and spectral properties of the data from which they are made. Further, for larger time series assessing the invariant nature of each sample site may be impossible due to resource restraints. This limitation can be particularly problematic when attempting to map land covers which change over non-regular intervals (e.g., commercially managed forests). While choosing sample sites well away from edges (where change is more likely to occur) can help to mitigate this issue, it can introduce challenges in classifying "transitional" or "mixed" habitat types. Acknowledging these limitations, the results show that this method of classification based on temporally invariant calibration sites can: (a) expedite the process of ground data collection by requiring only one calibration set for multiple maps, (b) overcome the limitation of ground reference unavailability, (c) be applied across multiple sensors of differing spatial and spectral characteristics, and (d) produce land-cover maps of overall statistical accuracy >76 percent.

Acknowledgments

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. SES-0849985 (REU Site) and by the Clark University O'Connor '78 Endowment. Special thanks to Dr. Trevor Jones (University of British Columbia) for his assistance in preparing this manuscript.

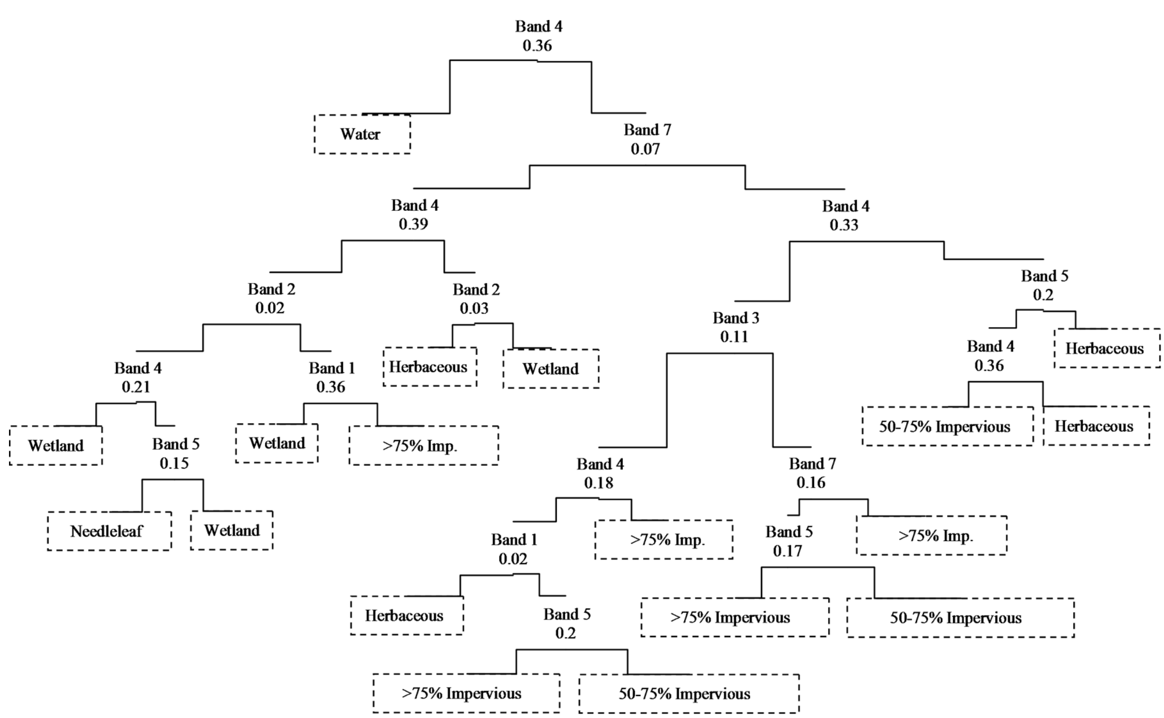

Figure 6. Example classification tree produced using 1999-ETM+ spectral bands 1 through 5, 7, and ancillary variables.

References

Breunig, J., 2003. *Losing Ground: At What Cost? Changes in Land Use and Their Impact on Habitat, Biodiversity, and Ecosystem Services in Massachusetts*, Third edition, Massachusetts Audubon Society, Summary Report 1–24

Brooks, R.T., and M. Hayashi, 2002. Depth-area-volume and hydroperiod relationships of ephemeral (vernal) forest pools in southern new England, *Wetlands.* 22(2):247–255.

Chavez, P.S., 1996. Image-based atmospheric corrections revisited and improved, *Photogrammetric Engineering & Remote Sensing*, 62(10):1025–1036.

Cohen, W.B., and M. Fiorella, 1998. Comparison of methods for detecting conifer forest change with Thematic Mapper imagery, *Remote Sensing Change Detection, Environmental Monitoring Methods and Applications* (R.S. Lunetta and C.D. Elvidge, editors), Ann Arbor Press, Chelsea, Michigan, pp. 89–102.

Congalton, R., and R. Mead, 1983. A quantitative method to test for consistency and correctness in photointerpretation, *Photogrammetric Engineering & Remote Sensing*, 49: 69–74.

Congalton, R., 1991. A review of assessing the accuracy of classifications of remotely sensed data, *Remote Sensing of Environment*, 37:35–46.

Congalton, R., and K. Green, 2009. *Assessing the Accuracy of Remotely Sensed Data – Principles and Practices*, CRC Press, Boca Raton, Florida, Taylor & Francis Group.

DeFries, R.S. and J.C.W. Chan, 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data, *Remote Sensing of Environment*, 74:503–515.

DeNormandie, J., 2009. *Losing Ground: Beyond the Footprint: Patterns of Development and Their Impact on the Nature of Massachusetts*, Fourth edition, Massachusetts Audubon Society, pp. 1–32.

Fazakas, Z., and M. Nilsson, 1996. Volume and forest cover estimation over southern Sweden using AVHRR data calibrated with TM data, *International Journal of Remote Sensing,* 17(9):1701–1709.

Franklin, S.E., and M.A. Wulder, 2000. Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas, *Progress in Physical Geography*, 26(2):173–205.

Friedl, M.A., and C.E. Brodley. 1997. Decisions tree classification of land cover from remotely sensed data, *Remote Sensing of Environment*, 61:399–409.

Goward, S.N., and D.L. Williams, 1997. Landsat and Earth systems science: Development of terrestrial monitoring, *Photogrammetric Engineering & Remote Sensing*, 63(7):887–900.

Hill, J., J. Megier, and W. Mehl, 1995. Land degredation, soil erosion and desertification monitoring in Mediterranean ecosystems, *Remote Sensing Reviews*, 12:107–130.

Iverson, L.R., E.A. Cook, and R.L. Graham, 1994. Regional forest cover estimation via remote sensing: The calibration center concept, *Landscape Ecology*, 9:159–174.

Jensen, J.R., K. Rutchey, M.S. Koch, and S. Narumalani, 1995. Inland wetland change detection in the Everglades Water Conservation Area 2A using a time series of normalized remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 61(2):199–209.

Jensen, J.R., 1996, *Introductory Digital Image Processing: A Remote Sensing Perspective*, Second edition, Prentice Hall, Upper Saddle River, New Jersey, 316 p.

Ji, C.Y., Q. Liu, D. Sun, S. Wang, P. Lin, and X. Li, 2001. Monitoring urban expansion with remote sensing in China, *International Journal of Remote Sensing*, 22(8):1441–1455.

Kalensky, Z., and L.R. Scherk, 1975. Accuracy of forest mapping from Landsat computer compatible tapes, *Proceedings of the 10th International Symposium on Remote Sensing of Environment*, 06–10 October, Ann Arbor, Michigan, pp. 1159–1167.

MaFOMP, 2009. Massachusetts Forest Monitoring Program, URL: *http://www.clarku.edu/departments/hero/forestchange.cfm*, Clark University, Worcester, Massachusetts (last date accessed: 01 December 2010).

Miller, J.D., and S.R. Yool, 2002. Mapping forest post-fire canopy consumption in several overstory types using multi-temporal Landsat TM and ETM data, *Remote Sensing of Environment*, 82:481–496.

Minter, T.C., 1978. Methods of extending crop signatures from one area to another, *Proceedings of The LACIE Symposium - A Technical Description of Large Area Crop Inventory Experiment (LACIE)*, 23–26 October, Houston, Texas.

Olthof, I., C. Butson, and R. Fraser, 2005. Signature extension through space for northern landcover classification: A comparison of radiometric correction methods, *Remote Sensing of Environment*, 95(3):290–302.

Pal, M., and P.M. Mather, 2003. An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment*, 86:554–565.

Pax-Lenney, M., C.E. Woodcock, S, Macomber, S. Gopal, and C. Song, 2001. Forest mapping with a generalized classifier and Landsat TM data, *Remote Sensing of Environment*, 77:241–250.

Pontius, R.G., and M. Millones, 2008. Problems and solutions for kappa-based indices of Agreement, *Proceedings of Studying, Modeling and Sense Making of Planet Earth*, Mytilene, Greece, 8 p.

Quinlan J.R., 1996. Improved use of continuous attributes in c4.5, *Journal of Artificial Intelligence Research*, 4:77–90.

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

Rogan, J., and D. Chen, 2004. Remote sensing technology for mapping and monitoring land- cover and land-use change, *Progress in Planning*, 61(4):301–325.

Rogan, J., and J. Miller, 2006. Integrating GIS and remotely sensed data for mapping forest disturbance and change, *Understanding Forest Disturbance and Spatial Pattern: Remote Sensing and GIS Approaches* (M. Wulder and S. Franklin, editors), Taylor and Francis, CRC Press.

Rogan, J., J. Franklin, D. Stow, J. Miller, D.A. Roberts, and C. Woodcock, 2008. Mapping land cover modifications over large areas: A comparison of machine learning techniques, *Remote Sensing of Environment*, 112(5):2272–2283.

Schriever, J.R., and R.G. Congalton, 1995. Evaluating seasonal variability as an aid to cover-type mapping from Landsat Thematic Mapper data in the Northeast, *Photogrammetric Engineering & Remote Sensing.* 61(3):321–327.

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

Townshend, J.R.G., C.O. Justice, C. Gurney, and J. McManus, 1992. The effect of misregistration on the detection of vegetation change, *IEEE Transactions on Geosciences and Remote Sensing*, 30:1054–1060.

Tripathy, G.K., T.K. Ghosh, and S.D. Shah, 1996. Monitoring of desertification process in Karnataka state of India using multi-temporal remote sensing and ancillary information using GIS, *International Journal of Remote Sensing*, 17(12):2243–2257.

Turner, B.L. II, D. Skole, S. Sanderson, G. Fischer, L. Fresco, and R. Leemans, 1995. Land-use and land-cover change: Science/research plan, *International Geosphere-Biosphere Programme Report No. 35, Human Dimensions of Global Environmental Change Programme Report No. 7*, Stockholm and Geneva, 96 p.

Washington-Allen, R., N. West, R.D. Ramsey, R.A. Efroymson, 2006. A protocol for retrospective remote sensing-based ecological monitoring of rangelands, *Rangeland Ecology & Management.* 59(1):19–29.

Weng, Q., 2001. A remote sensing-GIS evaluation of urban expansion and its impact on surface temperature in the Zhujiang Delta, China, *International Journal of Remote Sensing*, 22(10):1999–2014.

Woodcock, C.E., S.A. Macomber, M. Pax-Lenney, W.B. Cohen, 2001. Monitoring large areas for forest change using Landsat: Generalization across space, time and Landsat sensors, *Remote Sensing of Environment*, 78(1):194–203.

Wulder, M., 1998. Optical remote-sensing techniques for the assessment of forest inventory and biophysical parameters, *Progress in Physical Geography*, 22(4):449–476.

(Received 24 May 2010; accepted 15 September 2010; final version 23 October 2010)