Classification and Accuracy Assessment for Coarse Resolution Mapping within the Great Lakes Basin, USA

John S. Iiames and Ross S. Lunetta

Abstract

This study investigated elements important to regional landscape assessments: (a) appropriate mapping spatial resolutions (regional versus subregional), and (b) accuracy assessment procedures (point-based versus area-based). The study used MODIS NDVI time-series data to derive landcover products (2007) in a study area within the Laurentian Great Lakes Basin (GLB). Area-based reference data (i.e., "maplets") were varied in size and number to assess landcover proportionality agreement and to provide accuracy assessment metrics generated by point-based methodology. High spatial resolution Landsat ETM data was used to assess pixel purity (PP) for the MODIS 250 m pixels imbedded within the maplets (i.e., percent homogeneous) for the dominant cover type. Comparisons between the maplet reference data found a 21.7 percent variation in accuracy values between PP50 percent (67.9 percent accuracy) and PP100 percent (89.6 percent accuracy). Point-based accuracy assessments typically use 100 percent homogeneous reference pixels to assess landcover products, positively biasing the accuracy values. Our area-based methodology allows for the assessment at varying reference pixel homogeneity.

Introduction

Numerous challenges are encountered in designing mapping methods and accuracy assessment procedures for medium-to-coarse spatial resolution imagery for heterogeneous landscapes. To date, most of these products have been developed at the global scale and are assessed for accuracy using techniques suitable for finer resolution imagery. Here, we investigated (a) appropriate mapping scale resolutions (regional versus subregional), and (b) accuracy assessment procedures (point-based versus area-based) for phenology-based landcover classification using year 2007 time series of the Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) for our study area within the Laurentian Great Lakes Basin (GLB), USA. The area-based method (i.e., "maplets") was modified to address the positive bias inherent with assessments that only include 100 percent homogeneous reference pixels. In our method we are able to assess these coarser spatial resolution data at varying pixel purity (PP) levels.

Mapping Scale Issues

Regional to global scale landcover maps (i.e., 250 m to 1.0 km) have been derived from numerous satellite remote sensing systems including the MODIS (Giri *et al.*, 2005), SPOT Vegetation (Global Landcover 2000) (Bartholomé and Belward, 2005), MERIS (GLOBcover) (Defourny *et al.*, 2006), and the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR) (IGBP-DISCover) (Loveland *et al.*, 1999). Landcover classification algorithms used at the global scale have been limited in capturing the local and regional variations in landcover, due in part to limitations in the number of training sites available to accurately represent regional areas. For example, the MODIS classification algorithm uses a database of cover types (n = 2000) to represent the entire globe. The MODIS land team has established these training sites to be geographically and ecologically comprehensive (Muchoney *et al.*, 1999).

An earlier global product developed from 1.0 km AVHRR NDVI composites (IGBP-DISCover) (1992 to 1993) addressed the large geographic extent issue by defining pseudo ecoregions using an unsupervised classification clustering of the NDVI data to identify areas of spectral similarity (Loveland *et al.*, 2000). A total of 961 clusters were identified globally with 205 located in North America. Friedl *et al.* (2000) suggested that subregional imagery differences between areas of similar vegetation composition may be responsible for inducing a unique spectral signature. This effect seems to preclude the use of smaller areas of interest when classifying large geographic regions. It was posited that clouds may obscure similar sites, creating a low NDVI signature in the shadowed area. Cover type confusion also has been documented at higher latitudes for phenology-based NDVI classification (Loveland *et al.*, 2000; Friedl *et al.*, 2000). It should be noted that geographic stratification may not yield significant classification accuracy differences based on the classification algorithm employed. Shao and Lunetta (2011) found that there were no advantages to stratification of the entire GLB to a regional level using a neural network (NN) classifier. However, in that study, the limiting factor seemed to be the small percentage of training pixels.

Accuracy Assessment Issues

Assessing the accuracy of these coarser spatial scale resolution maps requires a deviation from the normal one-to-one (pixel-wise) assessment process where one homogeneous reference pixel, typically derived from higher resolution data are compared to the similar pixel with associated thematic label.

United States Environmental Protection Agency, 109 T.W. Alexander Dr., MD E243-05, Research Triangle Park, NC, 27711 (iiames.john@epa.gov).

Photogrammetric Engineering & Remote Sensing
Vol. 79, No. 11, November 2013, pp. 1015–1026.

0099-1112/13/7911–1015/$3.00/0

doi: 10.14358/PERS.79.11.1015

At issue is the dominance of non-homogeneous reference data in moderate-to-coarse spatial resolution imagery, where data resolution range from 10^2 to 10^3 m multiple landcover types dominate (Cihlar *et al.*, 2000). For example, a study in the Albemarle-Pamlico Watershed of North Carolina (NC) and Virginia (VA) found that only 6.0 percent of all the 250 m pixels were composed of a single landcover type (Knight *et al.*, 2006). The low proportion of homogeneous reference pixels within a classification scene will affect the application of a standard confusion matrix-based accuracy assessment approach for coarser spatial resolution mapping products. Statistics generated from the confusion matrix are statistically valid based on the assumption that samples are derived from relatively pure pixels of discreet cover classes (Foody, 2002). That is, the Kappa coefficient implicitly assumes that the testing sample is homogeneous. With finer resolution imagery (e.g., 10 to 30 m), reference samples are constrained to homogeneous areas with respect to one cover type. Additionally, to ensure that homogeneous pixels are not contaminated with spectral bleeding from adjacent pixels, reference pixels are usually selected within a cluster of pixels of the same cover type. Accuracy statements made from contingency tables generated from these pure reference pixels tend to be optimistically biased (Plourde and Congalton, 2003). The lack of pure reference data, typical with coarser resolution data, also affects selecting a sample size capable of generating statistically valid accuracy statements across all cover classes, where a standard sample size of $n = 50$ per landcover class has been suggested (Congalton and Green, 2009).

Some have suggested that the more reasonable assessment process for moderate-to-coarse resolution landcover is to derive areal sampling documenting the fractions of cover types present (Knight *et al.*, 2006; Latifovic and Olthof, 2004). One method, referred to as the maplet method (i.e., "area-based", "non-site specific"), allows the level of "correctness" to be assessed based on the agreement between the maplet reference cover proportions and the classification cover proportions of the same maplet areas (Latifovic and Olthof, 2004).

Maplets are higher spatial resolution maps of small geographic areas used to assess the accuracy of coarser resolution maps (Chrisman, 1991). Maplets were developed initially as a validation approach for large area datasets to deal with the issue of assessing class accuracies across a large number of classes. This methodology was first posited by Chrisman (1991) and further elucidated in practice by Stoms (1996). Lioubimtseva and Defourny (1999) compared the total area of cover types throughout three large maplet areas ranging from approximately 5,137 to 6,225 km^2. Beyond the comparison of landscape proportions, they also assigned dominant cover type labels to each pixel (30 m^2) within the maplet areas to generate contingency tables to compare total, user's and producer's accuracies between areas. Stoms (1996) used only one large maplet (2,240 km^2) for San Diego County, California Cihlar *et al.* (2000) used a tiling design to refine mapping cover type proportions from 1.0 km AVHRR data in order to compare proportions derived from the coarser AVHRR data resampled to a 30 m spatial resolution. Though not strictly defined as "maplets," this research investigated scaling issues associated with area-based assessments at three spatial scales, 30 m, 480 m, and 1.0 km. Schneider *et al.* (2003) implemented three maplet methods to supplement traditional accuracy assessment procedures in urban areas by fusing multiple sources of coarser resolution imagery. This research illustrated the benefit of areal comparisons to better understanding the nature and quantity of errors. For example, a comparison of reference maplets derived from the National Land Cover Database (NLCD), provided locational information leading to the identification of error type that revealed registration errors as the primary error component associated with urban cover extent. They also cautioned that the maplet aggregation method may introduce additional error sources (Schneider *et al.*, 2003).

Study Objectives

The objectives of this study were (a) to investigate scaling (regional to subregional) impacts on classification accuracies using 250 m multi-temporal NDVI imagery for 2007, and (b) to compare two accuracy assessment approaches: area-based and point-based. Here, we define regional (Omernik Level III) and subregional (Omernik Level IV) scales based on Omernik's classification of ecoregions within the contiguous United States (Omernik, 1987). We employed a novel methodology of distributing multiple smaller maplets throughout the classified image and determine the optimal maplet resolution and maplet numbers for classification assessment. First we classified the larger regional scale (115,934 km^2) Omernik Level III (OL3) ecoregion ("Northern Hardwood Forest") using ENVI's Spectral Angle Mapper (SAM), a hyperspectral image classification technique applied to continuous time-series NDVI for four cover types (woody deciduous and coniferous vegetation, barren, and grass). Then, we applied the same classification algorithm across 30 smaller subregional scale Omernik Level IV (OL4) ecoregions nested within the larger OL3 ecoregion. To test regional/subregional classification impacts we compared both OL3 and OL4 classifications against a reference dataset derived from the 2006 NLCD. Finally, both classifications (regional and subregional) were assessed over one OL4 ecoregion extent (Toimi Drumlins) using point-based and area-based accuracy assessment procedures.

Study Area

We performed classifications within an ecoregion sub-basin structure for the United States portion of the GLB corresponding to the Omernik Ecoregion Classification System. Omernik developed the ecoregions for the conterminous US at four levels, with subdivisions predicated on "perceived patterns of a combination of causal and integrative factors including land use, land surface form, potential natural vegetation, and soils" (Omernik, 1987). The US portion of the GLB is composed of 12 OL3 ecoregions covering 328,128 km^2 with over one-third of the area comprising the Northern Lakes and Forests Ecoregion (Omernik Code = 50). OL3 designations were designed to address regional analysis, whereas OL4 designations provide useful information at the local level of analysis. The OL3 Northern Lakes and Forests Ecoregion is further segmented into 30 distinct OL4 ecoregions ranging between 1 to 7 percent of the OL4 parent region (Figure 1).

The area-based versus point-based accuracy assessment comparisons were focused within the OL4 Toimi Drumlins ecoregion (5,473 km^2) nested within the larger OL3 ecoregion, or 4.4 percent of this area. The Northern Lakes and Forests ecoregion is characterized by nutrient-poor glacial soils dominated by coniferous and northern hardwood forests. The glacial processes on this ecoregion have produced undulating till plains, morainal hills, broad lacustrine basins, and sandy outwash plains. The Toimi Drumlins, located north by north-east of Duluth, Minnesota, are described by a rolling topography of ridge and troughs where drumlins are typically 1.6 km long, 0.4 km wide, 9 to 16 m high, and oriented in a southwest-northeast direction. Soils are medium to coarse-textures of Superior and Rainy Lobe glacial till. Inter-drumlin areas are poorly and very poorly drained and vegetation is dominated by aspen-birch, spruce-fir, white-red-jack pine, and oak-hickory cover types.

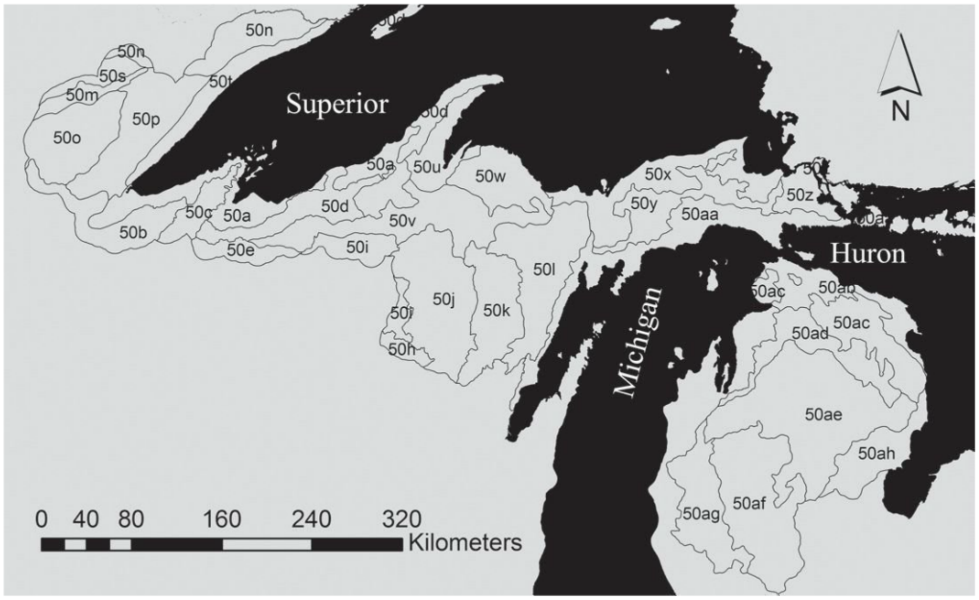

Figure 1. The OL4 ecoregions (n = 30) within the OL3 Northern Lakes and Forests Ecoregion of the GLB.

Methodology

Overview

Classifications were performed using biweekly time-series MODIS NDVI (2007) at two ecoregional levels (OL3 and OL4) with training sites selected specific to those two ecoregions (Figure 2). To allow for direct comparison between the two classifications, the OL3 classification was subset to the 30 OL4 ecoregions geographic extents (Figure 1). Traditional point-based accuracy metrics (i.e., percent correct, commission and omission errors, Kappa analysis, etc.) were generated for the 30 OL3 and OL4 classifications using a reference dataset developed from the 2006 NLCD (Figure 2). Site-specific "traditional" point-based accuracy methods were then compared with non-site specific "maplets" area-based procedures in assessing accuracy metrics for both OL3 and OL4 classifications over the Toimi Drumlins OL4 Omernik ecoregion (Figure 2). Data used with this experimental design is summarized in Table 1.

MODIS NDVI Preprocessing

The MODIS 250 m NDVI product (MOD13Q1) was downloaded for a seven-year period (2000 to 2007) from the USGS Land Processes Distributed Active Archive Center (*https://lpdaac.usgs.gov/*) to support phenology-based classifications across the GLB. The MOD13Q1 product consisted of 23 scenes developed from 16-day composites over the one calendar year. Though data for all seven years was collected in order to provide the necessary inputs for a missing data/cleaning algorithm developed internally at the Environmental Protection Agency (EPA) (Knight *et al.*, 2006), only the 2007 (n = 23) was used for classification purposes. Data were reprojected from the native sinusoidal projection to the Albers-equal area conic projection using a nearest-neighbor operator. Next, each individual scene was clipped to the GLB boundary layer and sequentially stacked. A series of filtering and cleaning steps were applied to the NDVI data stack based on the filtering and cleaning algorithm detailed in Lunetta *et al.* (2006). The resulting filtered and cleaned 2007 NDVI datastack for the GLB was then temporally subset to 12-bands corresponding only to the March through October growing season (determined by NDVI deciduous leaf growth/senescence curves), thereby reducing the contamination of snow and ice existent over a significant portion of the calendar year.

Classifications

The GLB landcover classification included seven classes (water, urban, barren, deciduous woody vegetation, coniferous woody vegetation, grass, and agriculture), however only four (woody deciduous and coniferous vegetation, barren, and grass) classes were used in this study. Water pixels were excluded because they were not pertinent to the study and agricultural pixels were previously assessed by Shao *et al.* (2010). The urban component of the classification was not included in this study based on the failure of applying a previously successful methodology using the Sequential Maximum Angle Convex Cone (SMACC) endmember model (Gruninger *et al.*, 2004) to identify urban endmembers from the temporal data.

For the OL3 classification (regional extent), we used the SAM hyperspectral classifying algorithm to classify multi-temporal NDVI data across the GLB. This SAM method was implemented in the GLB due to the high mapping accuracies derived from the Albemarle-Pamlico watershed mapping project (Knight *et al.*, 2006). Training data was visually identified for all four classes using (a) Landsat-7 SLC-on (1999 to 2003) leaf-off imagery including NDVI (used to distinguish conifer/deciduous differences), (b) USDA 2007 digital orthophoto quarter quadrangles (DOQQs), (c) forest cover inventory data from the Minnesota Department of Natural Resources (DNR) (Forest Inventory Management (FIM)), Wisconsin DNR (Wisconsin landcover data (WISCland)), and the Michigan DNR (Integrated Forest Monitoring, Assessment, and Prescription (IFMAP)), and

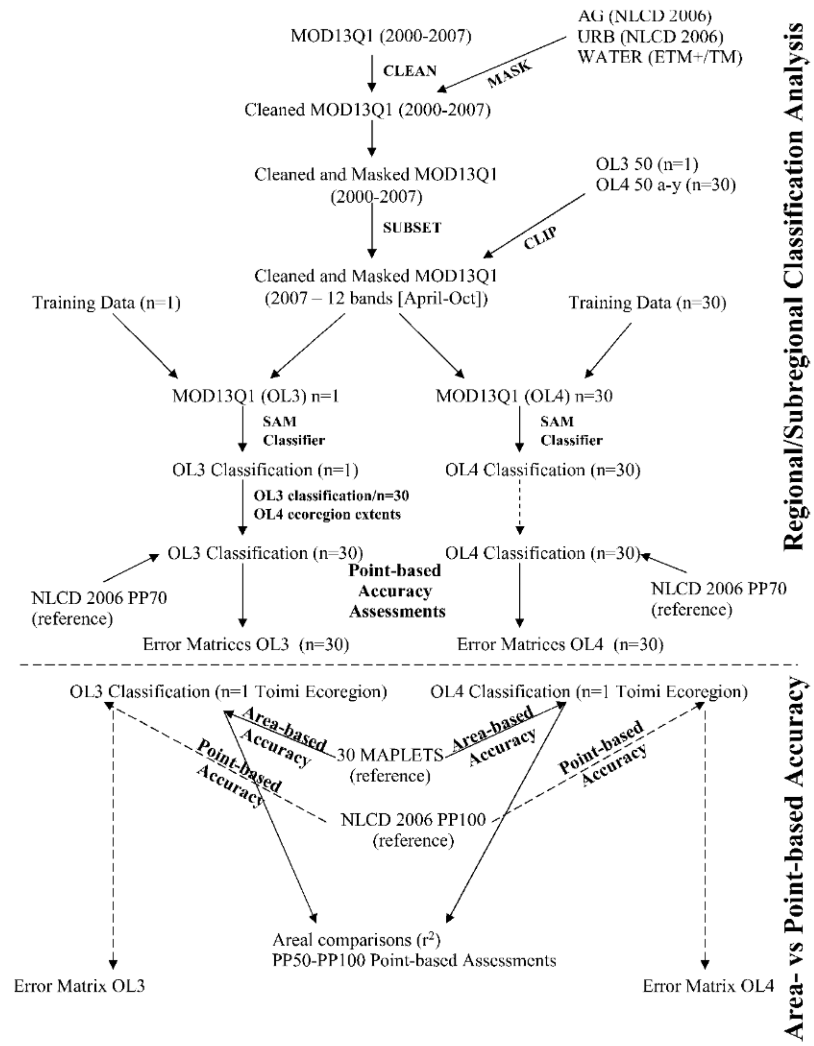

Figure 2. Tier (1) **Regional/Subregional Classification Analysis:** MODIS 250 m NDVI data (MOD13Q1) is cleaned, masked, and subset to the GLB boundary. Next, the MOD13Q1 data are temporally subset to year 2007 (April to October) and then clipped to the Omernik OL3 Ecoregion (Northern Lakes and Forests); then that region is clipped into the 30 finer OL4 Omernik Ecoregions. SAM classifiers map the one OL3 and 30 OL4 ecoregions and point-based accuracy assessments using the NLCD 2006 are applied to test regional versus sub-regional mapping. Tier (2) **Area: versus Point–based Accuracy (Maplets):** Multiple maplets (30) are generated to test both the OL3 and OL4 classification in one OL4 ecoregion (Toimi). Maplets are created from ISODATA unsupervised classification of Landsat TM data. Point-based assessments are generated for both classifications using the NLCD 2006 dataset as reference. Maplets provide areal comparisons and also allow for the varying of reference pixel homogeneous levels to create accuracy values at varying pixel purity levels.

(d) 100 percent homogeneous 250 m pixels derived from the 2006 NLCD. These training data at the OL3 level were distributed across the entire region, typically acquiring three to four signatures per class; increasing the number of signatures per class beyond this range contributed to more confusion in the resulting classification. Temporal training signatures, defined as endmember spectra in ENVI, were retrieved using ERDAS Imagine®, then saved as a text file and later imported into ENVI. The Spectral Angle Mapper (SAM) algorithm uses an *n*-dimensional angle to match unclassified pixels to a reference signature. Here, temporal NDVI value similarity between the training data and the unclassified pixel is determined by comparing the angle between the two values, treating these values as vectors in a space with dimensionality equal to the

Table 1. Data Summary. T1 - Tier 1 (Regional versus Subregional Mapping Comparison) T2 - Tier 2 (Area-based Maplets versus Point-based Accuracy Assessment Procedures): DNR – Department of Natural Resources; MI – Michigan; MN – Minnesota; WI – Wisconsin; USDA – United States Department of Agriculture;*R – raster; ** P – polygon; *** path/row

Data	Type/Scale	Data Origin	Data Origin Type/Scale	# Bands	Dates	Purpose	Study
Water Mask	R/250 m	ETM+/TM ***(21/29–30; 22/28–30;23/28–29; 24/27–28; 25/27–28; 26/27–28; 27/27–28)	*R/30 m	4		Mask GLB	T1/T2
Urban Mask	R/250 m	NLCD 2006	R/30 m	1		Mask GLB	T1/T2
AG Mask	R/250 m	NLCD 2006	R/30 m	1		Mask GLB	T1/T2
MODIS NDVI	R/250 m	MOD13Q1	R/250 m	189	2000–2007	Data clean	T1/T2
MODIS NDVI	R/250 m	MOD13Q1	R/250 m	12	2007	Classify	T1/T2
Landsat 7	R/30 m	ETM+ (21/29–30; 22/28–30;23/28–29; 24/27–28; 25/27–28; 26/27–28; 27/27–28)	R/30 m	4	1999–2003 (leaf-off)	Train	T1/T2
USDA 2007 DOQQs	R/0.5 m	USDA	R/0.5 m	4	2007 (leaf-on)	Train/Assess	T1/T2
MN FIM	P/NA	MN DNR	**P/NA	NA		Train/Assess	T1/T2
WISCland	R/30 m	WI DNR	R/30 m	1		Train/Assess	T1
MI IFMAP	P/NA	MI DNR	R/30 m	1		Train/Assess	T1
PP100% LC	R/250 m	NLCD 2006	R/30 m	1		Train	T1/T2
PP70% LC	R/250 m	NLCD 2006	R/30 m	1		Assess (point)	T1
PP100% LC	R/250 m	NLCD 2006	R/30 m	1		Assess (point)	T2
OL3	R/250 m	OL 3	P/NA	1		Define region	T1/T2
OL4	R/250 m	OL 4	P/NA	1		Define subregion	T1/T2
Maplets	R/250 m	ETM+ (26/27–28)	R/30 m	4	10/5/2002	Assess (area)	T2

number of bands (Kruse *et al.*, 1993). Finally, the completed OL3 classification was subset to the OL4 (n = 30) ecoregion boundaries to facilitate direct comparisons. This same classification process was repeated for each OL4 ecoregion (n = 30), however training data was specific to individual OL4 ecoregions.

MODIS NDVI Analysis

To first address the regional versus subregional classification issue of coarser spatial resolution imagery we attempted to geolocate accuracy assessment point-sample locations that were 100 percent homogeneous with respect to pixel purity (i.e., PP100). To achieve the minimum number of samples per class (n = 50) (Fry *et al.*, 2011), ⩾6,000 PP100 pixels (i.e., 30 OL4 ecoregions × 4 classes × 50 points/class = 6,000) were needed based on the 30 OL4 regions for four cover types. To ensure pixel purity, areas containing numerous contiguous PP100 pixels are commonly used for sub-sampling to offset any geometric registration issues and minimize spectral contamination from adjacent pixels. However, only 750 pixels across all 30 OL4 ecoregions met these criteria. Also, a majority of the available reference pixels were predominantly deciduous and coniferous. To compare classifications across the OL4 ecoregions, we relaxed the pixel purity requirement to PP70 (i.e., 70 percent one cover type) and utilized isolated pixels. We used the NLCD 2006 to create a majority reference map identifying all 250 m pixels dominated (>70 percent) by one cover type (n = 611,636) (Wickham *et al.*, 2013). To identify PP70 pixels, NLCD cover type proportions were calculated using Matlab software for every 250 m pixel location within the US portion of the GLB. Each NLCD cover class was converted to an ERDAS Imagine® IMG file and stacked to provide all 15 NLCD classes in one IMG file using ERDAS Model Maker. This datastack was consolidated into a 1-band reference image file where cells were populated by PP70 NLCD landcover for all 15 classes. Reference grid structure adhered to the MODIS 250 m grid structure to allow for direct comparison. Regional and Subregional (OL3 and OL4) classifications were assessed across the OL4 subregional extents for basic correspondence to the selected reference dataset using the GIS Analysis Summary Module in ERDAS Imagine®. Results were transferred to error matrices and accuracy statistics were generated for overall accuracy, commission and omission errors, Kappa and Z-statistics.

To compare area-based maplets versus point-based accuracy assessment procedures, point-based and area-based reference datasets were developed for the Toimi Drumlin OL4 ecoregion (Figure 1, [50p]). For the point-based dataset, a total of 127 PP100 percent pixels completely contained within similar landcover pixels were identified within this OL4 ecoregion. To ensure correct labeling of the reference pixels, ancillary datasets were visually compared to the 127 reference pixels (These datasets were explained earlier in the "Classifications" section of this paper). The area-based maplet reference dataset was developed by creating a 25 × 25 grid (cell size = 5 × 5 km, n = 625) which was superimposed over the OL4 Toimi Ecoregion boundary. This grid was developed using the "create fishnet" tool under the X Tools dialogue in Esri ArcMap®. We selected all 5 × 5 km cells (n = 173) that were completely contained within this Toimi Drumlin OL4 ecoregion and randomly selected 30 of these cells (i.e., maplets) for processing (Figure 3). A sample of n = 30 was chosen based on the results from Cihlar *et al.* (2000) to provide the minimum sample size required to retain cover composition. They found that ±7 percent area sampled was required for 30 m data and ±15 percent for 500 m, where gains

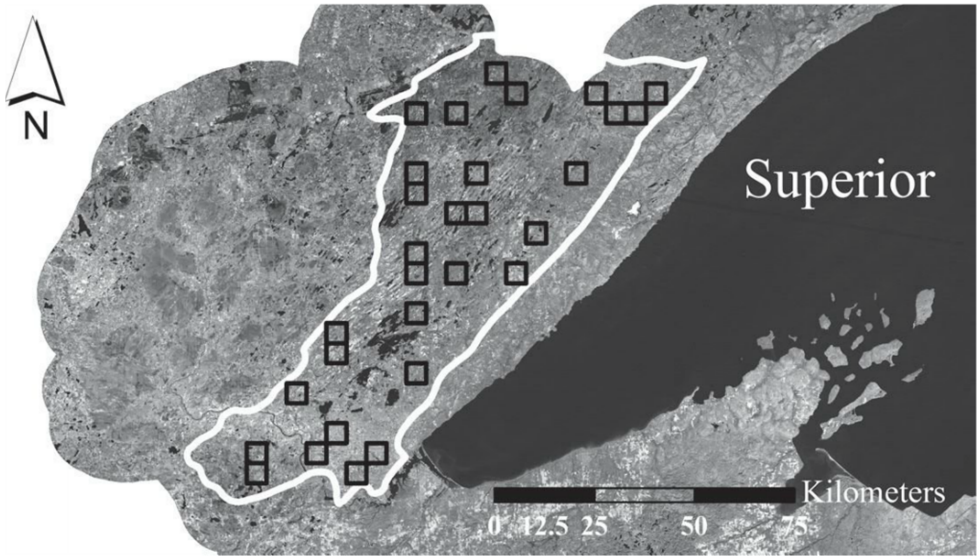

Figure 3. The Toimi Drumlin OL4 ecoregion [50p] and the distribution of randomly selected 5 × 5 km maplets (n = 30).

of increased sampling thereafter increased at a decreasing rate. By sampling 17 percent of the Toimi Drumlin Ecoregion, we exceeded this convergence point.

Next, we downloaded two 05 October 2002, Landsat ETM+ scenes for image processing (path/rows: 26/27 and 26/28). These scenes met the requirements for spectral similarity, low cloud cover (<10 percent), leaf-off/snow-free landscape conditions, and temporal target window (2007). To ensure that cover composition did not change within the 30 maplet areas, the imagery was checked against the 2007 leaf-on DOQQs. All maplets within the ecoregion showed no significant change compared to the 2007 DOQQs and the Landsat ETM+ imagery and thus were appropriate to support the analysis. We also used the NLCD 2006 landcover change product to confirm our visual inspection where <2.1 percent change was detected within the 30 maplets between 2001 and 2006 (Fry *et al.*, 2011). This dataset was developed supplementary to the 2006 NLCD where an algorithm isolates spectrally changed pixels along with the change trajectory (increase or decrease in biomass).

The 30 maplet areas were independently classified using the ISODATA algorithm, where spectrally similar clusters were later manually labeled as (a) water, (b) urban, (c) barren, (d) deciduous woody vegetation, (e) coniferous woody vegetation, and (f) grassland. This maplet classification approach was also implemented by Lioubimtseva and Defourny (1999) where they combined a maximum likelihood supervised classification with an unsupervised algorithm (ISODATA) to produce maplets with four to seven cover types. The dominant and the percent cover by class for each 250 m pixel per maplet area were calculated. Each 5 × 5 km maplet was also reduced to four additional resolutions (1 × 1 km, 2 × 2 km, 3 × 3 km, and 4 × 4 km) to test the appropriate maplet resolution for assessments. These maplet reductions were all centered about the center 30 m pixel of the original 5 × 5 km maplet. The same supplemental datasets used to confirm the cover types for the 127 pixels in the point-based reference dataset were also used to ensure label accuracy with the 30 selected maplets.

Both a point and area-based analysis of classification accuracies for OL3 (regional) and OL4 (subregional) products were compared across the Toimi Drumlins OL4 ecoregion extent. The barren class was eliminated from the assessment process due to the insignificant representation resulting in a three class assessment. A point-based accuracy assessment was first applied to both the OL3 and OL4 classification results using randomly selected reference points (n = 127; PP100 percent). Overall map and per-class accuracy were calculated through map-reference comparisons using contingency tables (Congalton, 1991). Errors of omission and commission were ascertained though the calculation of user's and producer's accuracies. Kappa statistics were also generated to determine if the values contained in an error matrix represented a result significantly better than random (Congalton and Green, 2009). A Z-statistic was generated for both error matrices using a pair-wise comparison (Jensen, 1996) to test the independence of Kappa values. Proportional cover type values were compared across the 30 maplet areas (25 km^2 or 5 × 5 km) within this same OL4 ecoregion and point-based assessments (PP $\geqslant$50 to 100 percent) were generated for only the 250 m pixels within the 30 maplet areas to observe the effects of pixel heterogeneity on overall accuracy. Finally, we investigated the impact of maplet size classes versus accuracy results for five resolutions (1 × 1 km, 2 × 2 km, 3 × 3 km, 4 × 4 km, and 5 × 5 km) and we also tested the optimal number of maplets using only the 5 × 5 km maplet size.

Results and Discussion

Regional/Subregional Land Cover Comparison Analysis

Landcover across the Northern Lakes and Forests Ecoregion extent was completed at the OL3 (regional) level, then stratified into 30 OL4 (subregional) ecoregional extents to allow for direct comparison with the 30 OL4 derived landcover products (Figure 1 and Plate 1). Overall classification (i.e., both combined 30 OL4 ecoregions and single OL3 ecoregion) accuracies

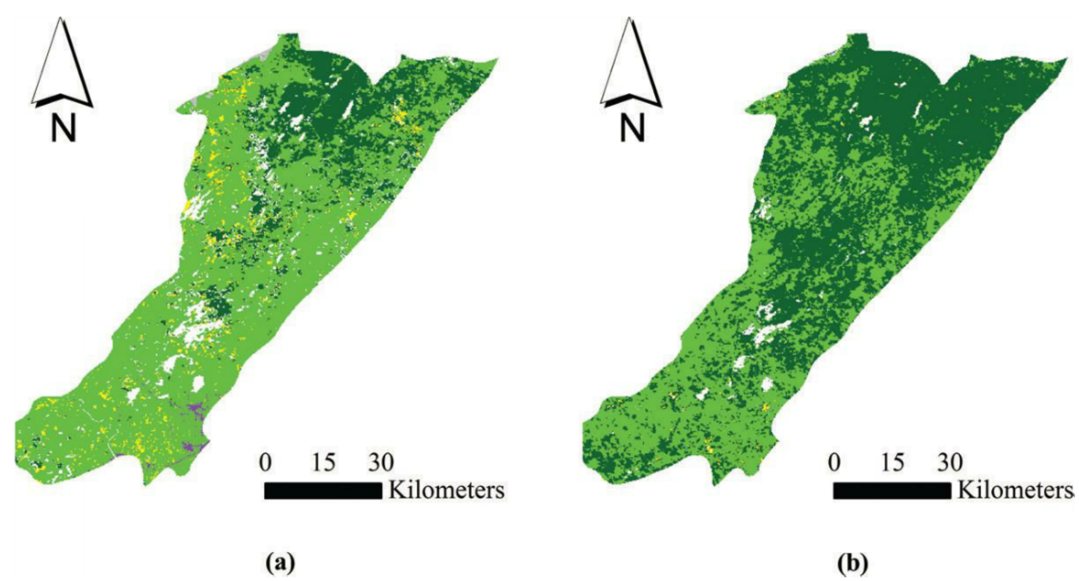

Plate 1. A comparison classification results for (a) OL3, and (b) OL4 classifications across the OL4 Toimi ecoregional extent (dark green: coniferous forest and light green: deciduous forest).

were similar for both the OL3 (83.3 percent) and OL4 (85.8 percent) products (Figure 4); however a pairwise Z-statistic test indicated that they were significantly different (Z = 22.55; p = 0.05). Comparing OL4 and OL3 classifications across all 30 ecoregional extents indicated that the subregional OL4 accuracies were superior to the regional OL3. Pairwise comparisons showed that 19 of 30 OL4 classifications had higher accuracies with nine of 19 OL4 classifications exhibiting a >5 percent accuracy differential and four of 19 exceeding the 10 percent differential (Table 2). In Figure 5, only two of the 11 classification comparisons where the OL3 accuracy value exceeded the OL4 classification resulted in a >5 percent accuracy differential. OL3 versus OL4 classification Z-statistic differences (p = 0.05) were observed in 25 of the 30 comparisons (Table 2). Kappa coefficients for the OL3 classifications showed moderate agreement across most of the 30 sites (68 percent), similar to those achieved for OL4 classifications (73 percent). Also, commission and omission errors were lowest for the deciduous and coniferous classes for both classifications, with the deciduous being lowest (Figure 6). This may have been a result of wetter soil conditions within the coniferous areas because wetland categories were not considered (Watt and Heinselman, 1965). A majority of the wetlands within the study area are spruce dominated.

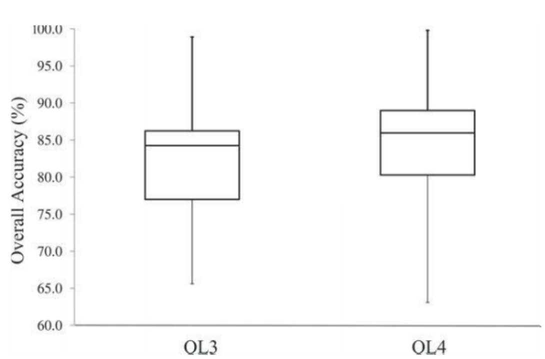

Figure 4. Point-based overall accuracy box and whisker plots ±1.0 standard deviations about the mean (solid line), median (dashed line), minimum, and maximum value outliers for OL3 and OL4 classifications (n = 30).

In summary, OL4 classifications performed consistently better across all 30 OL4 ecoregional extents when compared to the corresponding OL3 classification. This was attributed to a number of underlying issues specific to the GLB area that tended to increase the variability of temporal NDVI signatures. These included climatic variability due to lake influences, snow cover periodicity, and data quality issues associate with high latitude areas. Also, the wide MODIS scan angle can cause regional variation in NDVI values. As the view angle increases beyond nadir the sensor field of view includes fewer shadowed components and more illumination of the canopy elements (Gupta, 1992). Statistically significant differences were among 25 of the 30 OL3/OL4 classification comparisons. The five locations that resulted in no significant differences were attributed to the similarity between training signatures and NDVI values of a particular cover type across multiple OL4 ecoregions. Also, if we assume that accuracy differences of <5 percent between OL3 versus OL4 comparisons were a function of classification noise and intrinsic reference database errors (Lunetta *et al.*, 2001), then the OL4 classifications soundly outperformed OL3. Although an overall accuracy difference of 2.5 percent was observed between the two classifications, a majority of subregional differences occurred in the pair-wise comparisons at all 30 OL4 ecoregional extents.

Table 2. Accuracy (ACC), Kappa, and Pairwise Z-statistic Results for OL3 and OL4 Classifications across the 30 OL4 Ecoregions. (*Percentage of PP70% NLCD 2006 Training Data per OL4 Ecoregion)

OL4	Area (km²)	Reference Area (%)*	ACC (OL3)%	ACC (OL4)%	Diff (OL3-OL4)%	Kappa (OL3)	Kappa (OL4)	Z Stat
50a	6449.7	41.6	84.6	77.0	−7.6	0.52	0.38	19.86
50aa	6540.8	17.9	73.5	79.6	6.1	0.47	0.54	8.03
50ab	3304.2	11.8	66.3	70.4	4.1	0.40	0.40	0.13
50ac	4616.8	22.4	76.4	76.1	−0.3	0.37	0.32	4.74
50ad	3620.4	44.7	86.8	88.8	2.0	0.42	0.45	2.56
50ae	11656.1	30.4	85.7	84.5	−1.2	0.73	0.71	4.29
50af	7864.6	22.0	78.9	86.2	7.3	0.47	0.53	5.75
50ag	4323.1	32.0	86.2	88.2	2.0	0.66	0.67	0.88
50ah	3109.9	10.8	82.8	77.5	−5.3	0.59	0.54	3.21
50b	3435.5	38.9	94.4	95.4	1.0	0.48	0.45	1.48
50c	1589.1	61.0	85.1	83.6	−1.5	0.64	0.58	6.66
50d	5150.5	51.0	74.1	85.7	11.6	0.40	0.46	9.03
50e	1373.8	29.5	88.7	86.7	−2.0	0.43	0.37	2.10
50h	651.0	46.2	98.9	96.2	−2.7	0.43	0.17	3.41
50i	2280.6	11.7	82.7	93.8	11.1	0.56	0.79	12.56
50j	6760.1	41.8	89.5	94.5	5.0	0.30	0.47	13.39
50k	4300.7	27.7	93.6	93.5	−0.1	0.58	0.56	1.03
50l	6124.1	13.9	87.3	88.5	1.2	0.36	0.36	0.12
50m	1092.8	32.2	65.6	82.6	17.0	0.37	0.71	24.20
50n	3377.6	42.3	74.4	75.9	1.5	0.47	0.51	5.59
50o	4633.1	25.0	92.8	89.7	−3.1	0.48	0.41	4.54
50p	5472.7	44.4	85.8	86.7	0.9	0.72	0.73	3.19
50s	1084.7	32.8	66.8	74.9	8.1	0.37	0.49	7.11
50t	2876.5	56.7	87.2	85.6	−1.6	0.67	0.63	5.17
50u	2409.3	32.1	86.9	99.8	12.9	0.46	0.99	44.88
50v	6395.7	25.3	80.9	88.8	7.9	0.46	0.59	14.19
50w	3212.4	29.6	83.9	93.8	9.9	0.40	0.66	18.87
50x	5344.7	45.3	76.6	81.9	5.3	0.52	0.58	9.93
50y	3765.4	8.2	65.6	70.1	4.5	0.40	0.46	3.58
50z	1673.0	8.7	66.4	63.1	−3.3	0.33	0.30	2.14

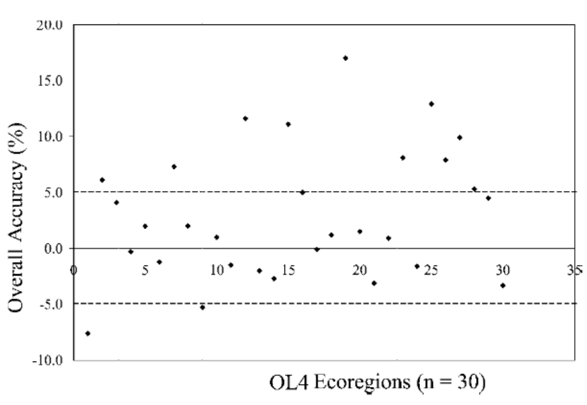

Figure 5. Accuracy differentials across thirty OL4 ecoregions = (OL4 accuracy) - (OL3 accuracy). Dash lines: ±5.0 percent difference levels.

Area-based (Maplet) versus Point-based Accuracy Assessments

The OL3 and OL4 classifications were assessed using both point-based and area-based methods over the Toimi Drumlins Ecoregion (Figure 1, [50p]). A traditional point-based accuracy assessment was performed using PP100 reference pixels (n = 127) existent within this ecoregion extent. Point-based accuracy metrics indicate that there was a significant difference (Z = 2.03; p = 0.05) between the OL3 classification accuracy (87.9 percent; Kappa = 0.79) and the OL4 classification accuracy (95.3 percent; Kappa = 0.91). Producer's and user's accuracies were high for both classifications within all three cover types (deciduous, coniferous, and grass). One exception was the producer's accuracy (14.3 percent; n = 7) for the OL4 grass cover type (Table 3). However, the very few reference locations skewed any attempt to explain error issues. Also, classes with low PP100 sample numbers were indicative of highly heterogeneous areas or under-representative classes.

The area-based assessment design incorporated 30 maplets randomly distributed throughout the one OL4 ecoregion (Toimi Drumlins). We found that the OL4 cover type proportions were better correlated than OL3 classifications,

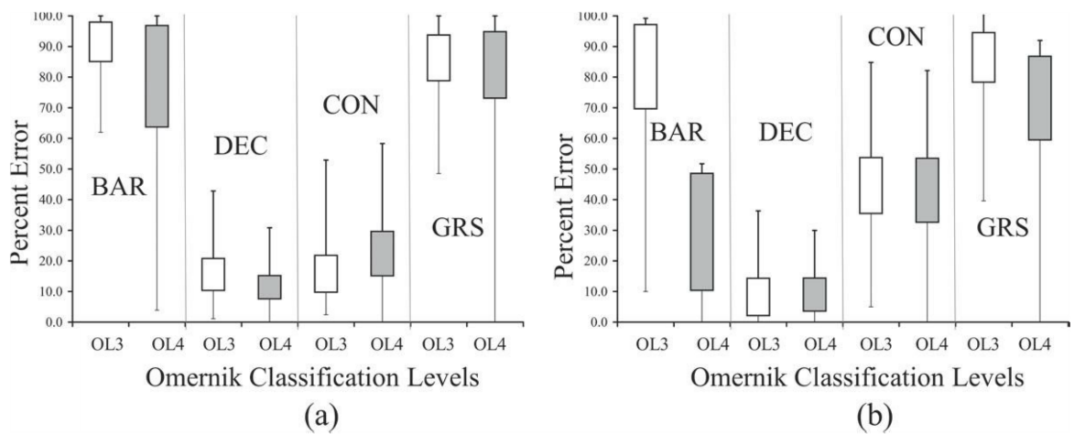

Figure 6. (a) Omission, and (b) Commission error boxes and whisker plots bracketing ±1.0 standard deviation about the mean, and minimum and maximum value outliers for two classification scales (OL3 and OL4) across the OL4 (n = 30).

Table 3. Classification Metrics for Point-based Accuracy Assessment of the Toimi Drumlins (OL4) Ecoregion for Deciduous (DEC), Coniferous (CON), and Grassland (GRS) Cover Types; Accuracy Metrics Include both Producers (P) and Users (U) Accuracies (%), Overall Accuracy (%), and Kappa Coefficients

CLASSIFICATION	n	DEC (P/U)	CON (P/U)	GRS (P/U)	Accuracy	Kappa
OL3	127	100.0 / 80.6	74.1 / 100.0	100.0 / 88.9	87.9	0.78
OL4	127	100.0 / 98.3	100.0 / 92.4	14.3 / 100.0	95.3	0.91

especially with respect to the deciduous and coniferous classes. The reference deciduous and coniferous proportions across all 30 maplets were 51.7 percent and 40.7 percent, respectively. The OL4 deciduous and coniferous proportions of 44.7 percent and 54.8 percent can be compared to the OL3 proportion of 75.8 percent and 19.6 percent across all 30 maplets. This extreme OL3 deciduous overestimation was also apparent by visual comparison of both classifications (Plate 1). A simple correspondence plot was used to compare the deciduous and coniferous maplet areas comparing reference data and classification results for the OL4 (Figure 7). This graph illustrates that the OL4 classification (a) overestimated areas of high (>30 percent) coniferous content, (b) underestimated areas with low (<50 percent) deciduous cover, and (c) overestimated areas with high (>50 percent) deciduous cover. Maplet regression analysis comparing OL3 deciduous (r^2 = 0.34) and coniferous (r^2 = 0.37) classes with OL4 deciduous (r^2 = 0.49) and coniferous (r^2 = 0.65) classes indicated that the OL4 results were superior to OL3 (Figure 8). The grass class accounted for 3.5 percent of the total maplet area. Regression coefficients showed that the OL4 classification had moderate correlation with the reference dataset (r^2 = 0.58, SE = 6.2 ha), whereas the OL3 had no agreement (r^2 = 0.01, SE = 118.2 ha).

We explored the effect of varying pixel purity for maplet reference pixels on overall accuracy using the point-based procedure within the 30 (5 × 5 km) maplet sites. Maplet pixels were identified for PP with respect to one cover type across six PP levels corresponding to ⩾50 percent, ⩾60 percent, ⩾70 percent, ⩾80 percent, and ⩾90 percent and 100 percent. Results showed that accuracy values varied by 21 percent

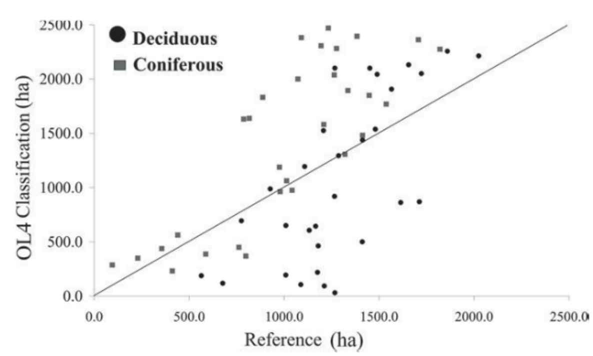

Figure 7. The correspondence between deciduous and coniferous for reference maplets data (n = 30) across the OL4. The line indicates a 1:1 correspondence.

with a minimum overall accuracy of 67.9 percent (⩾50 PP) and a maximum of 89.6 percent (PP100 percent) (Table 4). The PP100 class represented 7.4 percent of pixels within the study area.

A research objective was to determine optimal maplet resolutions and numbers (n) for classification assessments. Resulting OL4 regression coefficients for deciduous (r^2 = 0.34 to 0.48) and coniferous (r^2 = 0.44 to 0.65) across the five

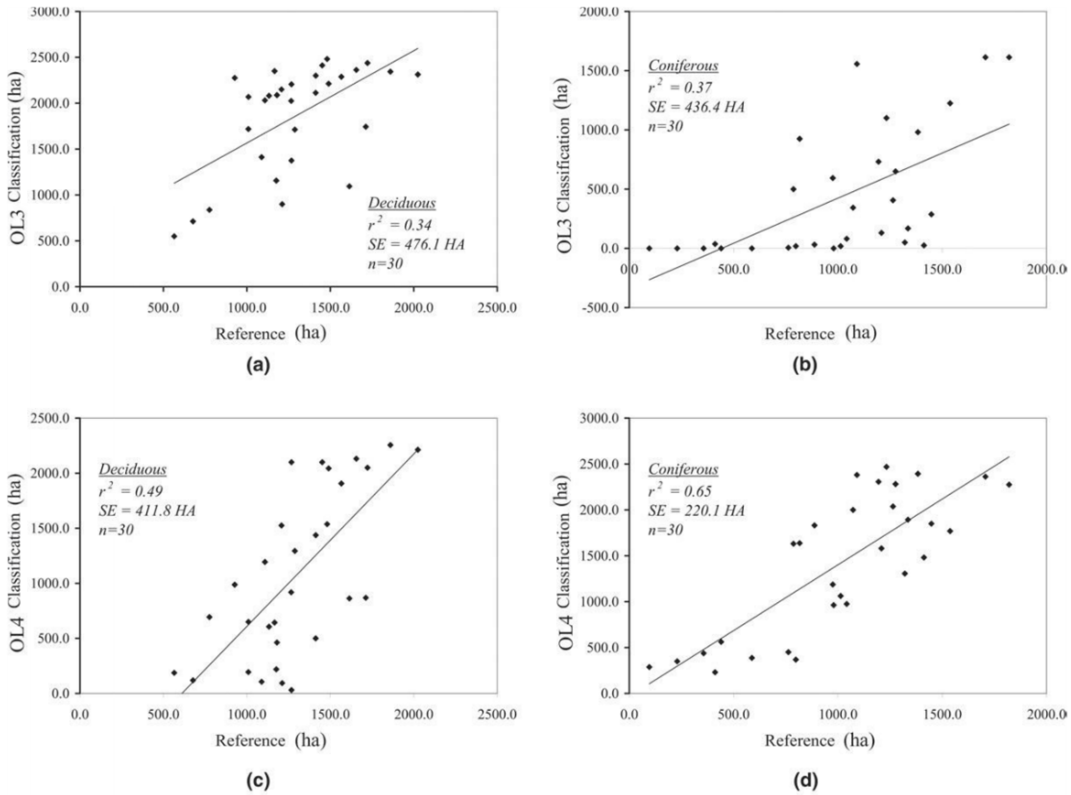

Figure 8. The regression coefficient results for (a and b) OL3 and (c and d) OL4 scale classifications for deciduous and coniferous forests across the OL4.

Table 4. Accuracy Metrics for Multiple Reference Pixel Purity (PP) Levels for OL4 Classification across 30 (5 × 5 km) maplet sites. Note that only 7.4 Percent of Pixels were 100 Percent Homogeneous for One Cover Type. The Majority Mlass Represents the Standard "Majority Call" Commonly Used to Label Moderate-to-Coarse Resolution Pixels

MATRIX Class	Kappa	LL Kappa	UL Kappa	Accuracy (%)	Total Area (%)
Majority	0.40	0.39	0.42	67.9	100.0
PP50%	0.42	0.40	0.43	69.1	94.8
PP60%	0.49	0.47	0.51	72.9	75.4
PP70%	0.57	0.55	0.58	77.3	56.0
PP80%	0.64	0.62	0.66	81.6	38.3
PP90%	0.72	0.69	0.74	85.7	21.6
PP100%	0.79	0.75	0.83	89.6	7.4

Table 5. Maplet Regression Coefficients Mean Standard Error (SE), and Root Mean Square Error (RMSE) Comparing OL4 (n = 30) Classification Results across the OL3 Toimi Drumlin Ecoregion (Note the Consistently Better Results for Coniferous versus Deciduous Forest)

Resolution	Cover Type	n	r^2	SE (ha)	RMSE (ha)
1 × 1 km	Deciduous	30	0.34	15.7	28.6
2 × 2 km	Deciduous	30	0.48	58.4	100.3
3 × 3 km	Deciduous	30	0.44	139.4	214.3
4 × 4 km	Deciduous	30	0.45	263.6	373.7
5 × 5 km	Deciduous	30	0.48	411.8	560.5
1 × 1 km	Coniferous	30	0.44	10.9	26.7
2 × 2 km	Coniferous	30	0.61	38.2	86.8
3 × 3 km	Coniferous	30	0.60	85.3	181.9
4 × 4 km	Coniferous	30	0.63	145.9	306.5
5 × 5 km	Coniferous	30	0.65	220.1	461.8

maplet grid resolutions (1 × 1 to 5 × 5 km) are listed in Table 5. For both cover types regression coefficients increased significantly between 1.0 to 2.0 km resolutions and thereafter stabilized, suggesting maplet resolutions >1.0 km would produce the highest correlation values. Using pixels coded to a simple majority within all 30 maplets, we compared the OL4 classifications using accuracy metrics generated from point-based accuracy assessments for all five maplet resolutions.

The results indicated that both accuracies (67.9 to 70.2 percent) and Kappa (0.40 to 0.44) statistics remained relatively stable across all resolutions (Table 6). The cover type proportions remained constant except at finer resolutions (<4 × 4 km), where some lesser represented cover types (bare and urban) dropped out completely (Table 7). The Producer's and

Table 6. Accuracy Metrics Including Kappa Coefficient (KHAT) Lower Limit (LL) and Upper Limit (UL), Over Accuracy (ACC), and Cover Type Percentages for Deciduous (DEC), Coniferous (CON), and Grasslands (GRS) across Maplets (n = 30) within OL4 for Five Maplet Resolutions

Maplet Size	KHAT	LL KHAT	UL KHAT	ACC (%)	DEC (%)	CON (%)	GRS (%)
5 × 5 km	0.40	0.39	0.42	67.9	54.7	40.4	3.5
4 × 4 km	0.40	0.38	0.42	68.3	54.0	41.7	3.1
3 × 3 km	0.41	0.38	0.44	68.7	53.9	42.4	2.7
2 × 2 km	0.42	0.38	0.46	69.8	52.8	44.2	2.6
1 × 1 km	0.44	0.36	0.51	70.2	55.6	42.1	2.3

Table 7. Mean Cover Type Percentages Generated from 20 Iterations each of Randomly Selected 5 × 5 km Maplets from the Original 30 Maplets across the OL4 Toimi Drumlin Ecoregion to Assess the Impact on Maplet Numbers on Classification Results; the Range of 15 to 30 Maplets had little Impact on Classification Outcomes.

	30 Maplets	25 Maplets	20 Maplets	15 Maplets
DEC (%)	51.72	51.67	51.42	51.82
CON (%)	40.73	40.68	40.92	41.01
GRS (%)	4.56	4.69	4.62	4.20

User's accuracies remained relatively unchanged across all resolutions. Results also indicated no benefit associated with maplet numbers >15 (Table 7). The spatial resolution of the maplet had more significance as to the representation of proportionally minor cover types when compared to the actual number of maplets required to make statistically relevant statements. The 5 × 5 km maplet with a count of ⩾15 maplets proved most relevant to the assessment OL4 classifications.

Conclusions

The issue of applying moderate-to-coarse spatial scale remote sensor data for regional-local scale classifications has been well documented in the literature. Overall, reported accuracies have been quite variable compared to that achieved from finer spatial scale data. Herold *et al.* (2008) cited three global landcover products that ranged from 66.9 percent to 78.3 percent in overall accuracy. To complicate matters, confidence intervals may have previously been overestimated due to the low number of reference samples and inherent positive bias by ignoring spatial autocorrelation impacts in the reference data sampling design. Also, accuracy values calculated at the global scale are frequently not applicable at continental scales. With this in mind, it was our intent to investigate the mapping of coarser spatial resolution time-series imagery at regional-local scales. Our findings indicate that classification products generated from training sites at the local level resulted in higher accuracy values across the majority of the broader regional area when compared to those derived from regional level training data. Our results include the caveat that the reference data derived from the NLCD 2006 had inherent error (not 100 percent accurate) and was highly spatially auto-correlated. This same issue exists for the maplet dataset where point-based accuracy metrics were generated for comparisons.

Many global classification products employ accuracy statements that are vague and non-site specific (Herold *et al.*, 2008). Employing the traditional point-based assessment on these coarser data types to produce accuracy metrics has numerous limitations. The assumption of "pure pixels" that underlies the standard approach of assessing error through a contingency matrix approach is often invalid. In this study, approximately 7 percent of the 250 m pixels across the study area were homogenous with respect to one cover type. Due to the limited number of 250 m homogenous pixels available (n = 750) obtaining the minimum number of reference pixels (i.e., 50 per class) for all four cover classes was not possible. Also, error assessments based on homogeneous pixels make no statements concerning the accuracy of the vast majority of the pixels being evaluated. Supplementary information can be obtained through the incorporation of area-based assessment procedure to determine the goodness-of-fit. In this study, the random distribution of these maplets allowed us to determine the correlation of cover classes with the reference data. Also, accuracy patterns were evident as cover types proportions changed. Finally, we calculated pixel heterogeneity which allowed us to create point-based error matrices that could account for pixel purities ranging from 50 to 100 percent. Although point-based accuracy methods may be adequate at finer spatial data (i.e., 30 m Landsat) where the mixed pixel issue is of lesser relevance, our results suggest the implementation of the maplet design for assessment of medium-to-coarse resolution landcover over large regional extents. Within maplet areas both site- and non-site-specific accuracy metrics can be evaluated. Identification of all levels of reference pixel purity within these maplet areas allows the user to better understand areas of confusion across a heterogeneous landscape. For example, an area dominated with PP70 over the majority of the map could be assessed using PP70 reference pixels to establish how well the analyst has mapped the dominant PP level. Further research is needed to investigate maplet reference data error and its affect on the assessment of global landcover products.

Acknowledgments

The authors would like to thank the three anonymous reviewers for their input into this work. Also, the authors would like to thank Joseph Knight, Jayantha Ediriwickrema, Drew Pilant, Russ Congalton, and Yang Shao for their feedback at various points along this process. The US 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

Bartholomé, E., and A. Belward, 2005. GLC2000: A new approach to global landcover mapping from Earth observation data, *International Journal of Remote Sensing*, 6(9):1959–1977.

Cihlar, J., R. Latifovic, J. Chen, J. Beaubien, Z. Li, and S. Magnussen, 2000. Selecting representative high resolution sample images for landcover studies, Part 2: Application to estimating landcover composition, *Remote Sensing of Environment*, (72):127–138.

Chrisman, N.R., 1991. The error component in spatial data, *Geographical Information Systems - Volume 1: Principles* (D.J. Maguire, M.F. Goodchild, and D.W. Rhind, editors), John Wiley and Sons, New York, pp. 165–174.

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

Congalton, R.G., and K. Green, 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, Second Edition, CRC/Taylor & Francis, Boca Raton, Florida, 183 p.

Defourny, P., C. Vancutsem, P. Bicheron, C. Brockmann, F. Nino, L. Schouten, and M. Leroy, 2006. GLOBCOVER: A 300 m global landcover product for 2005 using ENVISAT MERIS time series, *Proceedings of ISPRS Commission VII Mid-Term Symposium: Remote Sensing: from Pixels to Processes*, Enschede, The Netherlands, 08–11 May.

Foody, G.M., 2002. Status of landcover classification accuracy assessment, *Remote Sensing of Environment*, (80):185–201.

Friedl, M.A., C. Woodcock, S. Gopal, D. Muchoney, A.H. Strahler, and C. Barker-Schaaf, 2000. A note on procedures used for accuracy assessment in landcover maps derived from AVHRR data, *International Journal of Remote Sensing*, 21(5):1073–1077.

Fry, J., G. Xian, S. Jin, J. Dewitz, C. Homer, L. Yang, C. Barnes, N. Herold and J. Wickham, 2011. Completion of the 2006 National Land Cover Database for the Conterminous United States, *Photogrammetric Engineering & Remote Sensing*, 77(9):858–864.

Giri, C., Z. Zhu, and B. Reed, 2005. A comparative analysis of the Global Landcover 2000 and MODIS landcover data sets, *Remote Sensing of Environment*, 94:123–132.

Gruninger, J., A.J. Ratkowski, and M.L. Hoke, 2004. The Sequential Maximum Angle Convex Cone (SMACC) endmember model, *Proceedings of SPIE, Algorithms for Multispectral and Hyperspectral and Ultraspectral Imagery*, 12 April 2004, Orlando, Florida, Vol. 5425-1.

Gupta, R.K., 1992. District level NOAA/AVHRR vegetation indices and scan angle effects, *International Journal of Remote Sensing*, 13(4):715–735.

Herold, M., P. Mayaux, C.E. Woodcock, A. Baccini, and C. Schmullis, 2008. Some challenges in global landcover mapping: An assessment of agreement and accuracy in existing 1 km datasets, *Remote Sensing of Environment*, (112):2538–1226.

Jensen, J.R., 1996. *Introductory Digital Image Processing: A Remote Sensing Perspective*, Second edition, Prentice Hall, Inc., Upper Saddle River, New Jersey, 315 p.

Knight, J.F., R.S. Lunetta, J. Ediriwickrema, and S. Khorram, 2006. Regional scale landcover characterization using MODIS-NDVI 250 m multi-temporal imagery: A phenology-based approach, *GIScience and Remote Sensing*, 43(1):1–23.

Kruse, F.A., A.B. Lefkoff, J.B. Boardman, K.B. Heidebrecht, A.T. Shapiro, P.J. Barloon, and A.F.H. Goetz, 1993. The spectral image processing system (SIPS) - Interactive visualization and analysis of imaging spectrometer data, *Remote Sensing of Environment*, 44:145–163.

Latifovic, R., and I. Olthof, 2004. Accuracy assessment using sub-pixel fractional error matrices of global landcover products derived from satellite data, *Remote Sensing of Environment*, (90):153–165.

Lioubimtseva, E., and P. Defourny, 1999. GIS-based landscape classification and mapping of European Russia, *Landscape and Urban Planning*, (44):63–75.

Loveland, T.R., Z. Zhu, D.O. Ohlen, J.F. Brown, B.C. Reed and L. Yang, 1999. An analysis of the IGBP global land cover characterization process, *Photogrammetric Engineering & Remote Sensing*, 65(10): 1021–1031.

Loveland, T.R., B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L. Yang, and J.W. Merchant, 2000. Development of a global landcover characteristics database and IGBP DISCover from 1 km AVHRR data, *International Journal of Remote Sensing*, 21(6–7):1303–1365.

Lunetta, R.S., J. Iiames, J. Knight, R.G. Congalton, and T.H. Mace, 2001. An assessment of reference data variability using a virtual field reference database, *Photogrammetric Engineering & Remote Sensing*, 67(6):707–715.

Lunetta, R.S., J.F. Knight, J. Ediriwickrema, J.G. Lyon, and L.D. Worthy, 2006. Landcover change detection using multi-temporal MODIS NDVI data, *Remote Sensing of Environment*, 105:142–154.

Muchoney, D., A. Strahler, J. Hodges, and J. LoCastro, 1999. The IGBP DISCover confidence sites and the system for terrestrial ecosystem parameterization: Tools for validating global land cover data, *Photogrammetric Engineering & Remote Sensing*, 65(10):1061–1067.

Omernik, J.M., 1987. Ecoregions of the conterminous United States, *Annals of the Association of American Geographers*, 77:118–125.

Plourde, L.C., and R.G. Congalton, 2003. Sampling method and sample placement: How do they affect the accuracy of remotely sensed maps?, *Photogrammetric Engineering & Remote Sensing*, 69(3):289–297.

Schneider, A., M.A. Friedl, D.K. McIver, and C.E. Woodcock, 2003. Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 69(12):1377–1386.

Shao, Y., R.S. Lunetta, J. Ediriwickrema, and J. Iiames, 2010. The use of MODIS-NDVI data for mapping cropland across the Great Lakes basin, *Photogrammetric Engineering & Remote Sensing*, 75(1):73–84.

Shao, Y., and R.S. Lunetta, 2011. Sub-pixel mapping of tree canopy, impervious surfaces, and cropland in the Laurentian Great Lakes Basin using MODIS time-series data, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2): 336–347.

Stoms, D.M., 1996. Validating large-area landcover databases with maplets, *Geocarto International*, 11(2):87–95.

Watt, R.F., and M.L. Heinselman, 1965. Foliar Nitrogen and Phosphorus level related to site quality in a northern Minnesota spruce bog, *Ecology*, 46(3):357–361.

Wickham, J.D., S.V. Stehman, L. Gass, J. Dewitz, J.A. Fry, and T.G. Wade, 2013. Accuracy assessment of NLCD 2006 landcover and impervious surface, *Remote Sensing of Environment*, 130:294–304.

(Received 06 February 2013; accepted 03 May 2013; final version 30 May 2013)