Classification and Error Assessment Scale Issues for MODIS Time-series Classifications across two Ecoregion Extents within the Laurentian Great Lakes Basin

John S. Iiames and Ross S. Lunetta

John S. Iiames and Ross S. Lunetta are with the U.S. Environmental Protection Agency, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, N.C. 27711 (iiames.john@epa.gov).

ABSTRACT

This study applied a phenology-based land-cover classification approach across the Laurentian Great Lakes Basin (GLB) using time-series data consisting of 23 Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) composite images (250 m) from calendar year 2007. Two classification products (levels III and IV) were evaluated within one level IV Omernik Ecoregion in the GLB using both point-based and area-based "maplets" assessment methodologies. Classification accuracies were assessed at both stratification levels using 127 homogeneous reference pixels (single cover type) for three cover types (deciduous, coniferous, and grass). Error matrices indicated an overall level III classification accuracy of 87.9% (KHAT = 0.78) compared to 95.3% (KHAT = 0.9109) at level IV. Also, there was a statistically significant difference between the two matrices ($Z = 2.03$; $p=0.05$).

Level IV classification extent proportions performed better than level III when compared with reference maplets, especially with respect to the deciduous and coniferous cover classes. The reference deciduous and coniferous proportions were 51.7% and 40.7%, respectively. The level IV classification comparisons were deciduous 44.7% ($r^2 = 0.49$) and coniferous 54.8% ($r^2 = 0.65$) compared to 75.8% ($r^2 = 0.34$) and 19.6% ($r^2 = 0.37$) at level III. Error matrices generated at 6 pixel-purity (PP) levels ($\geq$ 50%, $\geq$ 60%, $\geq$ 70%, $\geq$ 80%, $\geq$ 90% and 100%) within the maplet areas resulted in an overall minimum accuracy of 67.9% ($\geq$50% PP) and a maximum accuracy of 89.6% (100% PP). Only 7.4% of the (250 m) pixels had 100% PP.

I. INTRODUCTION

For regional scale mapping with moderate-to-coarse spatial resolution satellite imagery, a number of primary issues have to be addressed prior to project initiation. Satellite sensor selection is based on optimizing spatial, spectral, temporal, and radiometric resolutions appropriate to capture the vegetation and anthropogenic variations seen across the landscape. Next, image classification (supervised, unsupervised, neural network, etc.) and accuracy assessment methodology are determined. Should training data extend over a regional or a sub-regional level? Will the traditional point-based accuracy assessment method suffice in informing the categorical confusion of the map? Variability is introduced across large geographical extents due to sub-regional differences in vegetation types, climate, geology/soil types, etc. Accordingly, across what geographical extent should the classification algorithms be applied to capture the local variation of land-cover? The issue of cover type heterogeneity existent across the landscape within the grid framework also has a basis in deciding how to assess the accuracy of the mapping project. In this study we have applied a phenology-based land-cover classification across the Laurentian Great Lakes Basin (GLB) at two ecoregion scales using a 2007 time series of 23 Moderate Resolution Imaging Spectroradiometer (MODIS) Normalized Difference Vegetation Index (NDVI) composite images (250 m pixel size). Land-cover products were evaluated using both the traditional point-based accuracy methodology and area-based (maplets) comparisons.

Regional to global scale land-cover maps have been derived from numerous satellite remote sensing systems including the MODIS, SPOT Vegetation (VGT), ENVISAT, and the National Oceanic and Atmospheric Administration (NOAA) Advanced Very High Resolution Radiometer (AVHRR). Land-cover classification algorithms used at the global scale have been limited in capturing the local and regional variations in land-cover, partially 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 team has established these training sites to be geographically and ecologically comprehensive [1].

An earlier global product developed from 1 km AVHRR NDVI composites (1992–1993) addressed the large geographic extent issue by defining pseudo ecoregions via an unsupervised classification clustering of the

NDVI data to identify areas of spectral similarity [2]. A total of 961 clusters were identified globally with 205 located in North America. Friedl et al. [3] suggested that subregional imagery differences between areas of similar vegetation composition on the ground may be responsible for inducing a unique spectral signature. This effect seems to preclude the use of smaller areas of interest when classifying large regional imagery datasets. It was posited that clouds may obscure two 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 [2], [3]. It should be noted that geographic stratification may not yield significant differences based on the classification algorithm employed when comparing classification accuracies from the whole to the part. Shao and Lunetta [4] 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 pixels chosen for NN training.

Assessing the accuracy of these moderate-to-coarse resolution maps requires a deviation from the normal one-to-one assessment process where one homogeneous reference pixel, typically derived from higher resolution data (e.g., aerial photography, Landsat ETM+, etc.), is compared to the similar pixel with associated thematic label. At issue is the dominance of non-homogeneous reference data, where data ranging spatially from 10^2 –10^3 m has been shown to contain multiple land-cover types [5]. A study in the Albemarle-Pamlico Watershed of NC and VA found that only 6.0% of all the 250 m pixels were composed of a single landcover type [6]. Some have suggested that the more reasonable assessment process for moderate-to-coarse resolution land-cover is to derive areal sampling documenting the fractions of cover types present [6, 7]. One method, referred to as the maplet method, 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 [7].

Maplets are higher resolution maps of small geographic areas used to assess the accuracy of coarser resolution maps [8]. Maplets were developed 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 [8] and further elucidated in practice by Stoms [9]. Lioubimtseva and Defourny [10] compared the total area of cover types throughout three large maplet areas ranging from approximately 5,137 – 6,225 km^2. Beyond the comparison of landscape proportions, they also assigned dominant cover type labels to each 30 m^2 pixel within the maplet areas to generate contingency tables to compare total, user's and producer's accuracies between areas. Stoms [9] used only one large maplet (2,240 km^2) for San Diego County, CA. Cihlar [5] used a tiling design (i.e. maplets) to refine mapping cover type proportions from 1.0 km AVHRR data to compare proportions derived from the coarser AVHRR data resampled to 30 m spatial resolution. Comparative Landsat ETM+ data was classified in order to directly correlate areal estimates of the two mapping products. Schneider et al. [11] implemented three maplet methods to supplement traditional accuracy assessment procedures in urban areas by fusing multiple sources of coarser resolution imagery. They indicated that the benefit of the areal comparisons allowed a better understanding of the nature and quantity of errors. For example, a comparison of reference maplets derived from the National Landcover Dataset (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 errors [11].

The objectives of this study were (1) to investigate the regional to sub-regional effect on GLB classification of 250 m multi-temporal NDVI imagery for 2007 and (2) to compare two accuracy assessment processes: non-site specific (area-based) and site-specific (point-based). We first classified the larger (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 six cover types. We then applied the same classification algorithm across the thirty smaller Omernik Level IV (OL4) ecoregions nested within this larger OL3. We compared both OL3 and OL4 classifications against a reference dataset derived from the 2006 NLCD. Finally, both classifications were assessed over one OL4 extent (Toimi Drumlins) using point-based and area-based accuracy assessment procedures.

II. STUDY AREA

We applied our classification to an ecoregion sub-basin for the United States section of the GLB corresponding to the Omernik Ecoregion Classification System. Omernik developed the ecoregions for the conterminous United States 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' [12]. The United States portion of the GLB is composed of 12 OL3 ecoregions covering 328,128 km^2 (Table 1), with over one-third of the area composing the Northern Lakes and Forests Ecoregion. OL3 designations are 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 from 1–7% of the OL4 parent region (Table 2: Figure 1).

For the area-based versus point-based accuracy assessment comparisons we concentrated our research within the OL4 Toimi Drumlins ecoregion (5,472.7 km^2) nested within the larger OL3 ecoregion. 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, 9–16 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.

II. METHODOLOGY

A. Overview

A two-tiered experimental design was developed (1) to assess the impacts of mapping at regional and sub-regional scales, and (2) to evaluate accuracy assessment information derived from both non-site specific area-based 'maplets' and traditional site-specific point-based approaches. 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. In order to allow for direct comparison between the two classifications, the OL3 classification was masked to the 30 OL4 ecoregions existent within the OL3 geographic extent (Figure 1). Traditional point-based accuracy metrics were generated for the 30 OL3 and OL4 classifications using a reference dataset developed from the 2001 NLCD. The second-tier of this study focused on comparing site-specific 'traditional' point-based accuracy methods with non-site specific 'maplets' area-based procedures in assessing accuracy metrics over one OL4 Omernik ecoregion (Toimi Drumlins).

B. Reference Data

To first address the scale issue with classification of medium-to-coarse resolution imagery we attempted to geolocate point-sample locations that were 100% homogeneous with respect to pixel purity (i.e. PP100). To achieve the minimum number of samples per class (n = 50), ≥ 9,000 PP100 pixels would be needed based on the 30 OL4 regions for the six cover types [13]. To ensure pixel purity, areas containing numerous PP100 pixels are commonly used for sub-sampling to offset any geometric registration issues and minimize spectral contamination from adjacent pixels. 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 and utilized isolated pixels. We used the NLCD 2001 to create a majority reference map identifying all 250 m pixels dominated (> 70%) by one cover type (n = 611,636). To identify PP70 pixels, NLCD cover type proportions were calculated using Matlab software for every 250 m pixel location within the U.S 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.

Point-based and area-based reference datasets were developed for the Toimi Drumlin OL4 ecoregion. For the point-based dataset, a total of 127 PP100 pixels completely contained within similar land-cover pixels were identified within this OL4 ecoregion. To ensure correct labeling of the reference pixels, ancillary datasets were compared to the 127 reference pixels. These datasets included (1) Landsat 7 SLC-on (1999-2003) leaf-off imagery including NDVI, (2) USDA 2007 digital orthophoto quarter quadrangles (DOQQs), and (3) Minnesota Department of Natural Resources forest cover inventory data (Forest Inventory Management (FIM)). The area-based maplet reference dataset was developed by creating a 25 x 25 grid with each cell 5 x5 km (n = 625). This grid was developed using the 'create fishnet' tool under the X Tools dialogue in ESRI ArcMap. We selected all 5 x 5 km cells (n = 173) that were completely contained within this OL4 ecoregion and randomly selected 30 of these cells (i.e. maplets) for processing (Figure 2). Next, we downloaded two October 5, 2002, Landsat ETM+ scenes for image processing (Path/Rows: 26/27 and 26/28). These scenes met the requirements of spectral similarity, low cloud cover (< 10%), and leaf-off/snow-free landscape conditions. To ensure that cover composition did not change within the 30 maplet areas, the imagery was checked against the 2006 leaf-on DOQQs. All maplets within the ecoregion showed no significant change compared to the 2006 DOQQs and the Landsat ETM+ imagery and thus were appropriate to support the analysis. However, visual inspection was limited due to the difficulty in distinguishing deciduous from coniferous cover types. Therefore, a secondary dataset was used to check general cover pattern similarity by comparing the NLCD 2006 data with Landsat ETM+ imagery from 2002. This confirmed that there were no significant changes between the dates — change within the 30 selected maplets was < 2.1%.

The Landsat imagery was subset to the 30 maplet areas and each area was independently classified using the unsupervised Iterative Self-Organizing Data Analysis Technique (ISODATA), where spectrally similar clusters were later 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 [10] where they combined a maximum likelihood supervised classification with an unsupervised algorithm (ISODATA) to produce maplets with 4–7 cover types. The dominant and the percent cover by class for each 250 m pixel per maplet area were calculated. Each 5 x 5 km maplet was also reduced to four additional resolutions (1 x 1 km, 2 x 2 km, 3 x 3 km, and 4 x 4 km) to test the appropriate maplet resolution for assessments. 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.

B. MODIS NDVI Preprocessing

The MODIS 250 m NDVI product (MOD13Q1) was downloaded for a 7-year period (2000–2007) from the USGS Land Processes Distributed Active Archive Center (https://lpdaac.usgs.gov/) to support phenology-based classifications across the GLB. The MOD13Q1 product consists 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 EPA [6], only the 2007 (n=23) was used in this study. 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. [14]. The resulting filtered and cleaned 2007 NDVI datastack for the GLB was then temporally subset to 12-bands corresponding only to the growing season, thereby reducing the contamination of snow and ice existent over a significant portion of the calendar year.

C. Water and Agricultural Masks

Both the water and agricultural classes were excluded from the classification and accuracy assessment procedures. Water pixels were excluded because they were not pertinent to the study and agricultural pixels were previously assessed by Shao et. al. [15]. A water mask was created from cloud-free Landsat ETM+ and TM imagery collected close as possible to the target 2007 year. Then, an ISODATA unsupervised classification algorithm was implemented in ERDAS Imagine to cluster the imagery into 20 distinct classes. The water and non-water classes were relabeled per class as “1’s” and “0’s” respectively. The resulting image was then resampled from 30 m to 10 m pixels to match the 250 m MODIS NDVI grid format then subsequently degraded to the 250 m resolution. Using the relational operator within ERDAS Imagine Model Builder, 250 m pixels greater than 50% water were identified and included as the water mask. An agricultural mask was created using the 2001 NLCD to identify 250 m pixels >50% agriculture cover.

D. Classifications

The complete GLB land-cover classification system includes seven classes: (1) water, (2) urban, (3) barren, (4) deciduous woody vegetation, (5) coniferous woody vegetation, (6) grass, and (7) agriculture. The water and agricultural classes were first masked out of the NDVI datastack. Next, we applied the Sequential Maximum Angle Convex Cone (SMACC) endmember model [16] to identify urban endmembers from the time-temporal data. The SMACC algorithm, initially designed for multi-spectral imagery analysis, uses an iterative process that identifies spectral similarity based on bright and dark pixel differences. This technique was previously used to identify urban pixels within the Albemarle-Pamlico (NC and VA) watershed [6], but proved ineffectual in the GLB due to the large geographic extent of this basin. Therefore, our mapping would be constrained only to the other four cover classes (barren, deciduous, coniferous, and grass). For these remaining classes, we used a hyperspectral classifying algorithm (Spectral Angle Mapper) to classify the 12 time-series NDVI images across the GLB. Training data was identified using 2007 DOQQs, forest inventory data (previously discussed), and 100% homogeneous 250 m pixels as determined by the 2001 NLCD. 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 number of bands [17]. Finally, the completed OL3 classification was subset to the OL4 (n=30) ecoregion boundaries to facilitate direct comparisons.

E. Accuracy Assessments

Regional and Sub-regional (OL3 and OL4) classifications were assessed across the OL4 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. Both a point and area-based analysis of classification accuracies were compared across the Toimi Drumlins OL4 ecoregion. 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). Overall map and per-class accuracy was calculated through map-reference comparisons using contingency tables [18]. 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 [19]. A z-statistic was generated for both error matrices using a pair-wise comparison [13] to test the independent of Khat values. Proportional cover type values were compared across the 30 maplet areas (25 km^2 or 5 x 5 km) within this same OL4 ecoregion and point-based assessments (PP $\geq$ 50% – 100%) were only 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 x 1 km, 2 x 2 km, 3 x 3 km, 4 x 4 km, and 5 x 5 km per side of pixel).

IV. RESULTS AND DISCUSSION

A. Classification Accuracy

Overall classification accuracies were similar for both the OL3 (83.3%) and OL4 (85.8%) products (Figure 3). A pairwise Z-statistic test indicated that both overall classifications were significantly different (z-statistic = 22.55; p=0.05). A comparison of the OL4 sites (n=30) for both classifications indicated that at the finer OL4 results were superior to the OL3. The pairwise comparisons showed that 19 of 30 OL4 sites had higher accuracies with > 50% of these 19 OL4 sites exhibiting a > 5% accuracy differential and > 21% exceeding the 10% differential (Table 3; Figure 4). Eleven OL3 classifications exhibited a > 5% accuracy differential compared to the OL4 classification. The mean accuracy improvement across the OL4 sites was 3.0% compared to OL3 classifications (84.6% vs. 81.6%). OL3 versus OL4 classification z-statistic differences (p=0.05) were observed in 25 of the 30 ecoregions (Table 3). Kappa coefficients for the OL3 classifications showed moderate agreement across most ecoregions (68%), similar to that achieved for OL4 classifications (73%). Also, commission and omission errors were lowest for the deciduous and coniferous classes for both classifications, with the deciduous being lowest (Figure 5). This may have been a result of wetter soil conditions within the coniferous areas because wetland categories were not considered. A majority of the wetlands within the study area are spruce dominated.

OL4 classifications performed consistently better across all 30 sites when compared to the OL3. 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. It has been shown that as the view angle increases beyond nadir the sensor field of view includes fewer shadowed components and more illumination of the canopy elements [20]. Approximately 83% of OL3 showed significant differences between the two classification levels. Locations with no significant differences (n = 5) can be attributed to robust 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% between sites were a function of classification noise and intrinsic reference database errors [21]; the OL4 classifications soundly outperformed OL3. Results demonstrated that 19 OL4 classifications outperformed the classification; of which 83% resulted in > 5% accuracy differences. Only 18% of the 11 OL3 accuracy exceedance sites exceeded the >5% differential. Although an overall accuracy difference of 2.5% was seen when combining all 30 sites, subregional differences can be seen when the coarser geographic extent is classified at the fine resolution.

B. Maplet versus Point-Based Assessments

The low proportion of homogeneous reference pixels within a classification scene affects applying the standard confusion matrix-based accuracy assessment approach for medium-to-coarse resolution mapping products. One issue associated with assessing the accuracy of moderate-to-coarse spatial resolution map products by applying a standard confusion matrix-based approach is the low proportion of homogeneous pixels. Statistics generated from the confusion matrix are statistically valid based on the assumption that samples are derived from pure pixels of discreet cover classes [22]. For example, the Kappa coefficient implicitly assumes that the testing sample is homogeneous. Typically, reference samples are constrained to areas that are homogeneous with respect to one cover class. Additionally, to ensure that homogeneous pixels are not contaminated with spectral bleeding from adjacent pixels of different land-cover, reference pixels are usually selected embedded 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 [23]. The lack of pure reference pixels also affects selecting a sample size capable of generating statistically valid accuracy statements across all cover classes.

A traditional point-based accuracy assessment was performed for the OL3 and OL4 classifications for the one OL4 Toimi Drumlins ecoregion extent using a randomly selected subset (n=127) of the PP100 reference pixels (n=750) previously identified across the entire OL3 ecoregion. Point-based accuracy metrics indicate that there was a significant difference ($Z = 2.03$; $p = 0.05$) between the OL3 classification (overall accuracy = 87.9%; KHAT = 0.79) and the OL4 classification (overall accuracy = 95.3%; KHAT = 0.91). Producer's and user's accuracies were high for both classifications within all three cover types (deciduous, coniferous, and grass). The only exception was the producer's accuracy of only 14.3% for the OL4 classification for grass (Table 4).

The proportional-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, especially with respect to the deciduous and coniferous classes. The reference deciduous and coniferous proportions were 51.7% and 40.7%, respectively. The OL4 deciduous and coniferous proportions of 44.7% and 54.8% can be compared to the OL3 proportion of 75.8%and 19.6%. This extreme OL3 deciduous overestimation is also apparent by visual comparison of both classifications (Figure 6). A simple correspondence plot was used to compare the deciduous and coniferous maplet areas comparing reference data and classification results for the OL4. This graph illustrates that the OL4 classification (i) overestimated conifer in areas of high (> 30%) coniferous content, and (ii) underestimated areas with low (< 50%) deciduous cover while overestimating areas with high (> 50%) deciduous cover (Figure 8). Regression analysis using maplet reference data for deciduous ($r^2 = 0.49$) and coniferous ($r^2 = 0.65$) classes for OL3 and

OL4 indicated that the OL4 results were superior to OL3 (r^2 =0.37 coniferous; r^2 =0.34 deciduous) (Figure 8). The grass class accounted for 3.5% 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 affect of varying pixel purity for maplet reference pixels on overall accuracy using the point-based procedure within the 30 (5 x 5 km) maplet sites. Maplet pixels were identified for percent purity with respect to one cover type across six PP levels corresponding to ≥ 50%, ≥ 60%, ≥ 70%, ≥ 80%, and ≥ 90% and 100%. Results showed that accuracy values varied by 21% with a minimum overall accuracy of 67.9% (≥ 50% PP) and a maximum of 89.6% (PP100) (Table 8). The PP100 class represented 7.4% 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 = 0.34–0.48) and coniferous (r^2 = 0.44–0.65) across the five maplet grid resolutions (1 x 1 km –5 x 5 km) are listed in Table 5. For both cover types regression coefficients increased significantly between 1 x 1 and 2 x 2 km resolutions and thereafter stabilized, suggesting maplet resolutions > 1 x 1 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 (67.9–70.2%) and Khat (0.40–0.44) statistics remained relatively stable across all resolutions (Table 6). The cover type proportions remained constant except for the finer resolutions (< 4 x 4 km), where some lesser represented cover types (bare and urban) dropped out completely (Table 6). The Producer’s and User’s accuracies remained relatively unchanged across all resolutions. Results also indicated no benefit associated with maplet numbers > 15 (Table 7). In summary, 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 x 5 km maplet with a count of 15 or more maplets proved relevant to the assessment OL4 classifications.

V. 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. [24] cited three global land-cover products that ranged from 66.9% to 78.3% 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 moderate-to-coarse 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% 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 [24]. Employing the traditional point-based assessment on these medium-to-coarse 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% 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 six cover classes was not possible. Also, error assessments based on homogeneous pixels makes no statements concerning the accuracy of the vast majority of the pixels being evaluated. Supplementary information can be obtained through the incorporation of proportional 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 50–100%. In summary, point-based accuracy methods are valid on finer spatial data (i.e., 30 m Landsat) where the mixed pixel issue is of lesser relevance. We encourage the implementation of the maplet design for assessment of medium-to-coarse resolution land-cover 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 understand areas of confusion over a heterogeneous landscape.

REFERENCES

[1] D. A. Muchoney, A. Strahler, J. Hodges, and J. LoCastro, “The IGBP DISCover confidence sites and the system for terrestrial ecosystem parameterization: Tools for validating global land-cover data,” *Photogramm. Eng. Remote Sens.*, vol. 65, pp. 1061-1067, 1999.

[2] T. R. Loveland, B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L. Yang, and J.W. Merchant, “Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data,” *Int. J. Remote Sens.*, 21(6-7):1303-1365. 2000.

[3] M. A. Friedl, C. Woodcock, S. Gopal, D. Muchoney, A.H. Strahler, and C. Barker-Schaaf, “A note on procedures used for accuracy assessment in land cover maps derived from AVHRR data,” *Int. J. Remote Sens.*, vol. 21, no. 5, pp. 1073-1077, 2000.

[4] Y. Shao and R. S. Lunetta, “Sub-pixel mapping of tree canopy, impervious surfaces, and cropland in the Laurentian Great Lakes Basin using MODIS time-series data”, *IEEE J. of Selected Topics in Applied Earth Observations and Remote Sensing*, vol 4, no 2, pp. 336-347, June 2011.

[5] J. Cihlar, R. Latifovic, J. Chen, J. Beaubien, Z. Li, and S. Magnussen, “Selecting representative high resolution sample images for land cover studies. Part 2: Application to estimating land cover composition,” *Remote Sens. Environ.*, vol. 72, pp. 127–138, 2000.

[6] J. F. Knight, R. S. Lunetta, J. Ediriwickrema, and S. Khorram, “Regional scale land cover characterization using MODIS-NDVI 250 m multi-temporal imagery: A phenology-based approach,” *GIScience and Remote Sensing*, vol. 43, no. 1, pp. 1-23, 2006.

[7] R. Latifovic and I. Olthof, “Accuracy assessment using sub-pixel fractional error matrices of global land cover products derived from satellite data,” *Remote Sens. Environ.*, vol. 90, pp. 153-165, 2004.

[8] N. R. Chrisman, “The error component in spatial data,” In: D.J. Maguire, M. F. Goodchild, D. W. Rhind, (Eds.), Geographical Information Systems – volume 1: Principles. Longman Scientific and Technical, pp. 165–174, 1991.

[9] D. M. Stoms, “Validating Large-Area land Cover Databases with Maplets,” *Geocarto International*, vol. 11, No. 2, pp. 87–95, June 1996.

[10] E. Lioubimtseva and P. Defourny, “GIS-based landscape classification and mapping of European Russia,” *Landscape and Urban Planning*, vol. 44, pp. 63-75, 1999.

[11] A. Schneider, M. A. Friedl, D. K. McIver, and C. E. Woodcock, “Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data,” *Photogramm. Eng. Remote Sens.*, vol. 69, no. 12, pp. 1377–1386, Dec. 2003.

[12] J.M. Omernik, “Ecoregions of the conterminous United States. *Annals of the Association of American Geographers*,” vol. 77, pp. 118-125, 1987.

[13] R. Congalton and K. Green, *Assessing the accuracy of remotely sensed data: principles and practices (2nd Edition)*. Boca Raton, FL: CRC/Taylor & Francis, 183pp., 2009.

[14] R.S. Lunetta, J.F. Knight, J. Ediriwickrema, J.G. Lyon, and L.D. Worthy, “Land-cover change detection using multi-temporal MODIS NDVI data”. *Remote Sens. Environ.*, 105, 142-154, 2006.

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

[16] J. Gruninger, A. J. Ratkowski, and M. L. Hoke, "The Sequential Maximum Angle Convex Cone (SMACC) Endmember Model," in *Proceedings SPIE, Algorithms for Multispectral and Hyperspectral and Ultraspectral Imagery*, vol 5425-1, Orlando, Fl., 2004.

[17] F. A. Kruse, 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 Sens. Environ.*, v. 44, p. 145 - 163. 1993

[18] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sens. Environ.*, 37:35-46. 1991.

[19] J. R. Jensen, Introductory *digital image processing: A remote sensing perspective (2nd Edition)*. Upper Saddle River, New Jersey, USA: Prentice Hall, Inc., 1996.

[20] R. K. Gupta, "District level NOAA/AVHRR vegetation indices and scan angle effects," *Int. J. Remote Sens.*, vol. 13, no. 4, pp. 715-735, 1992.

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

[22] G. M. Foody, “Status of land cover classification accuracy assessment,” *Remote Sens. Environ.*, vol. 80, pp. 185-201, 2002.

[23] R. G. Congalton and L. C. Plourde, “Sampling methodology, sample placement, and other important factors in assessing the accuracy of remotely sensed maps.” In: G. B. M. Heuvelink, M. J. P. M. Lemmens (Eds.), Accuracy 2000. Delft University Press, Amsterdam, pp. 117–124. 2000.

[24] M. Herold, P. Mayaux, C.E. Woodcock, A. Baccini, C. Schmullis, “Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets,” *Remote Sens. Environ.*, vol. 112, pp. 2538-1226, 2008.

Table 1. A listing of 12 OL3 Ecoregions that comprise the United States portion of the GLB including code name, identification numbers, area and representative percentages of total area. The Northern Lakes and Forests (35.3%) and S. Michigan/N. Indiana Drift Planes (20.3%) represented 55.6% of the total US OL3 region.

OL3 (Name)	**OL3 (Code Id)**	**Area (km^2)**	**Percent**
Northern Lakes and Forests	50	115,935	35.3
North Central Hardwood Forests	51	20,480	6.2
Southeastern Wisconsin Till Plains	53	15,631	4.8
Central Corn Belt Plains	54	3,787	1.2
Eastern Corn Belt Plains	55	16,922	5.2
S. Michigan/N. Indiana Drift Plains	56	66,529	20.3
Huron/Erie Lake Plains	57	24,860	7.6
Northeastern Highlands	58	8,061	2.5
N. Appalachian Plateau and Uplands	60	10,350	3.2
Erie Drift Plain	61	3,631	4.2
North Central Appalachians	62	928	0.3
E. Great Lakes and Hudson Lowlands	83	31,015	9.5
TOTAL		**328,128**	**100.0**

Table 2. OL4 Ecoregional area distribution within the OL3 Northern Lakes and Forests (#50) including code identifications and area percent of OL3 by code (n=30). Area percents vary by a factor of approximately 19x ranging from 9.4% (50ae) to 0.5% (50h).

OL4 (Name)	OL3 (Code)	Percent
Lake Superior Clay Plain	50a	5.2
Menominee-Drummond Lakeshore	50aa	5.2
Cheboygan Lake Plain	50ab	2.6
Onaway Moraines	50ac	3.7
Vanderbilt Moraines	50ad	2.9
Mio Plateau	50ae	9.4
Cadillac Hummocky Moraines	50af	6.3
Newaygo Barrens	50ag	3.5
Tawas Lake Plain	50ah	2.5
Minnesota/Wisconsin Upland Till Plain	50b	2.8
St. Croix Pine Barrens	50c	1.3
Superior Mineral Ranges	50d	4.1
Chequamegon Moraines & Outwash Plain	50e	1.1
Perkinstown End Moraines	50h	0.5
N. Wisconsin Highlands Lakes Country	50i	1.8
Brule and Paint River Drumlins	50j	5.4
Wisconsin/Michigan Pine Barrens	50k	3.5
Menominee Drumlins & Ground Moraine	50l	4.9
Mesabi Range	50m	0.9
Boundary Lakes and Hills	50n	2.7
Glacial Lakes Upham and Aitken	50o	3.7
Toimi Drumlins	50p	4.4
Nashwauk/Marcell Moraines and Uplands	50s	0.9
North Shore Highlands	50t	2.3
Keweenaw-Baraga Moraines	50u	1.9
Winegar Dead Ice Moraine	50v	5.1
Michigamme Highland	50w	2.6
Grand Marais Lakeshore	50x	4.5
Seney-Tahquamenon Sand Plain	50y	3.0
Rudyard Clay Plain	50z	1.3
TOTAL		**100.0**

Table 3. Accuracy (ACC), Kappa (KHAT) and Pairwise Z-statistic results for OL3 and OL4 classifications across the OL4 ecoregions (n=30).

OL3 (Code)	ACC (OL3)	ACC (OL4)	KHAT (OL3)	KHAT (OL4)	Z-statistic
50a	84.6	77.0	0.52	0.38	19.86
50aa	73.5	79.6	0.47	0.54	8.03
50ab	66.3	70.4	0.40	0.40	0.13
50ac	76.4	76.1	0.37	0.32	4.74
50ad	86.8	88.8	0.42	0.45	2.56
50ae	85.7	84.5	0.73	0.71	4.29
50af	78.9	86.2	0.47	0.53	5.75
50ag	86.2	88.2	0.66	0.67	0.88
50ah	82.8	77.5	0.59	0.54	3.21
50b	94.4	95.4	0.48	0.45	1.48
50c	85.1	83.6	0.64	0.58	6.66
50d	74.1	85.7	0.40	0.46	9.03
50e	88.7	86.7	0.43	0.37	2.10
50h	98.9	96.2	0.43	0.17	3.41
50i	82.7	93.8	0.56	0.79	12.56
50j	89.5	94.5	0.30	0.47	13.39
50k	93.6	93.5	0.58	0.56	1.03
50l	87.3	88.5	0.36	0.36	0.12
50m	65.6	82.6	0.37	0.71	24.20
50n	74.4	75.9	0.47	0.51	5.59
50o	92.8	89.7	0.48	0.41	4.54
50p	85.8	86.7	0.72	0.73	3.19
50s	66.8	74.9	0.37	0.49	7.11
50t	87.2	85.6	0.67	0.63	5.17
50u	86.9	99.8	0.46	0.99	44.88
50v	80.9	88.8	0.46	0.59	14.19
50w	83.9	93.8	0.40	0.66	18.87
50x	76.6	81.9	0.52	0.58	9.93
50y	65.6	70.1	0.33	0.30	2.14

Table 4. 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 (KHAT) coefficients.

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

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)**
1x1 km	Deciduous	30	0.34	15.7	28.6
2x2 km	Deciduous	30	0.48	58.4	100.3
3x3 km	Deciduous	30	0.44	139.4	214.3
4x4 km	Deciduous	30	0.45	263.6	373.7
5x5 km	Deciduous	30	0.48	411.8	560.5
1x1 km	Coniferous	30	0.44	10.9	26.7
2x2 km	Coniferous	30	0.61	38.2	86.8
3x3 km	Coniferous	30	0.60	85.3	181.9
4x4 km	Coniferous	30	0.63	145.9	306.5
5x5 km	Coniferous	30	0.65	220.1	461.8

Table 6. Accuracy metrics including Kappa coefficient (KHAT) lower limit (LL) and upper limit (UL), over accuracy (ACC), and cover type percentages for water (WAT), urban (URB), barren (BAR), 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 (%)**	**WAT (%)**	**URB (%)**	**BAR (%)**	**DEC (%)**	**CON (%)**	**GRS (%)**
5x5 km	0.40	0.39	0.42	67.9	1.2	0.2	0.0	54.7	40.4	3.5
4x4 km	0.40	0.38	0.42	68.3	1.1	0.1	0.0	54.0	41.7	3.1
3x3 km	0.41	0.38	0.44	68.7	0.9	0.1	0.0	53.9	42.4	2.7
2x2 km	0.42	0.38	0.46	69.8	0.4	0.0	0.0	52.8	44.2	2.6
1x1 km	0.44	0.36	0.51	70.2	0.0	0.0	0.0	55.6	42.1	2.3

Table 7. Mean cover type percentages generated from 20 iterations each of randomly selected 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**
WAT (%)	2.03	1.99	2.09	2.07
URB (%)	0.94	0.95	0.93	0.89
BAR (%)	0.02	0.02	0.02	0.02
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

Table 8. Accuracy metrics for multiple reference pixel purity (PP) levels for OL4 classification across 30 maplet sites. Note that only 7.4% of pixels were 100% homogeneous for one cover type. The majority class represents the standard "majority call" commonly used to label moderate-to-coarse resolution pixels.

MATRIX Class	KHAT	LL KHAT	UL KHAT	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

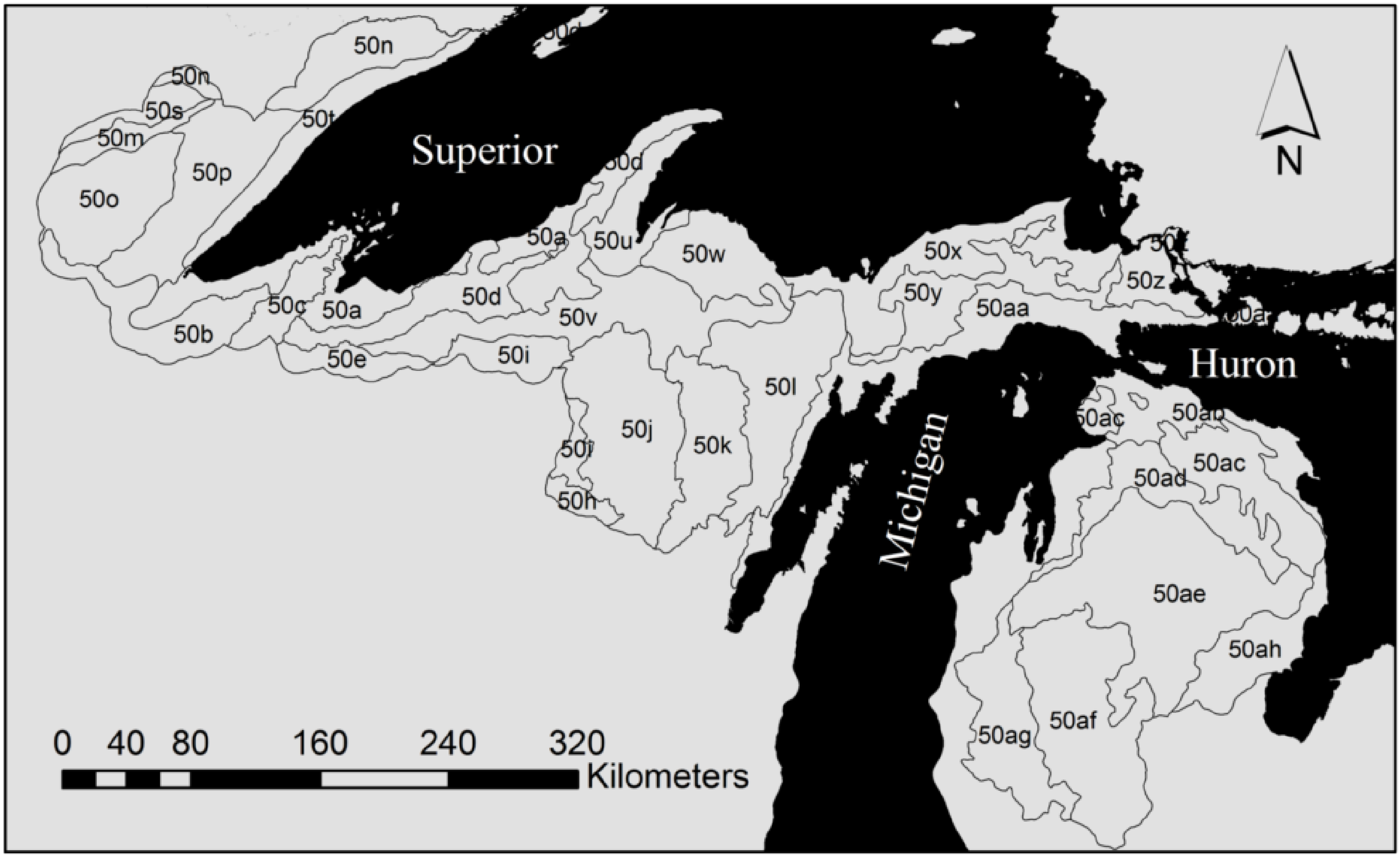

Figure 1. The OL4 ecoregions (n=30) within the OL3 Northern Lakes and Forests Ecoregion of the GLB.

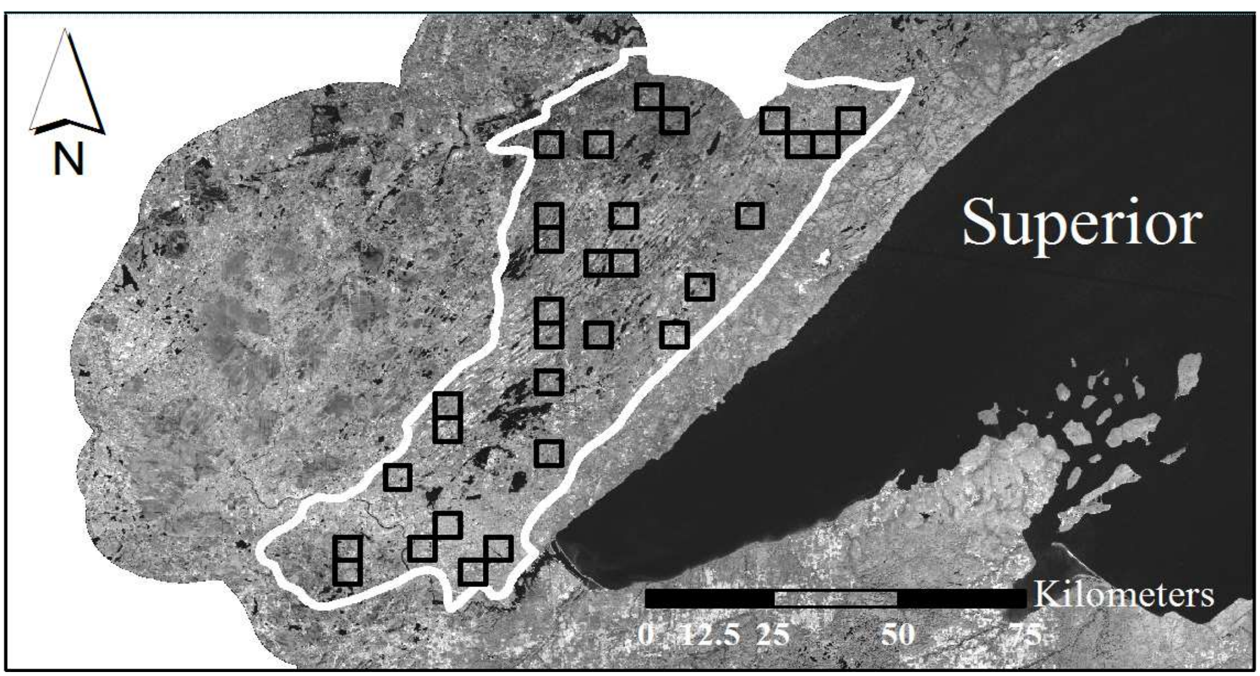

Figure 2. The Toimi Drumlin OL4 ecoregion (50p) and the distribution of randomly selected 5 x 5 km maplets (n=30).

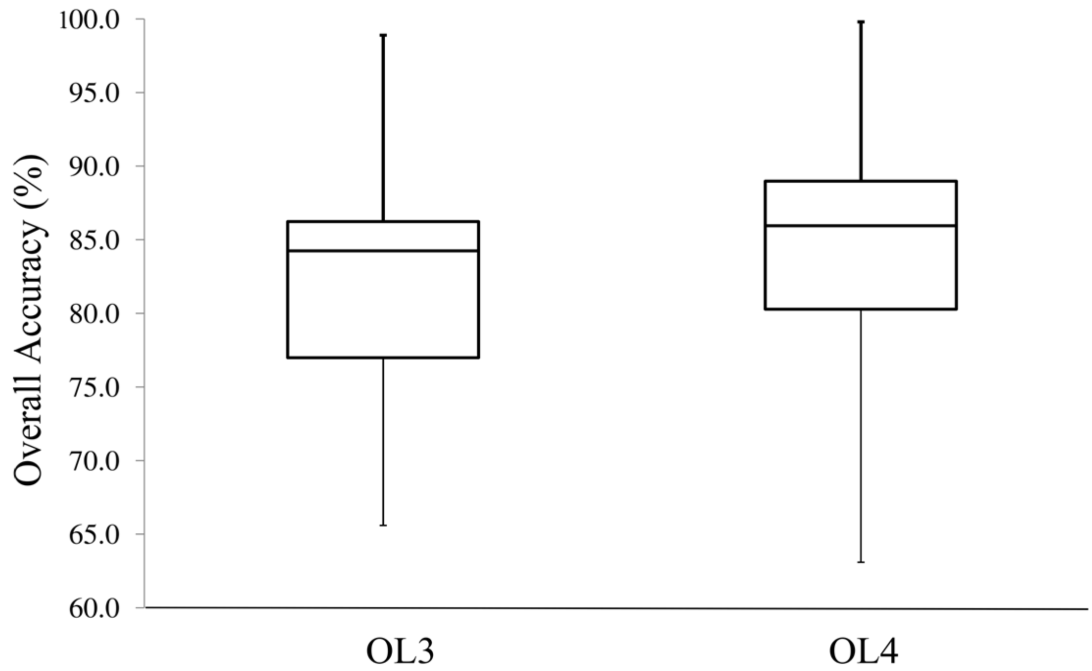

Figure 3. Overall accuracy box and whisker plots ± 1.0 standard deviations about the mean, median, minimum, and maximum value outliers for OL3 and OL4 classifications (n=30).

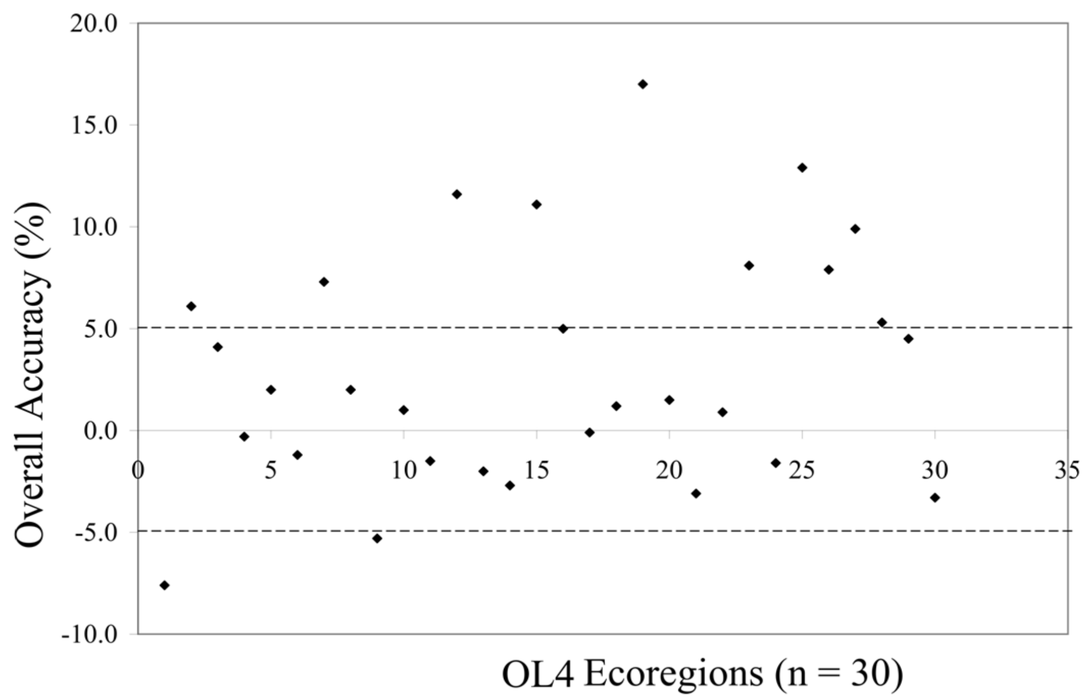

Figure 4. Accuracy differentials across thirty OL4 ecoregions = (OL4 accuracy) - (OL3 accuracy). Dash lines: ±5.0% difference levels.

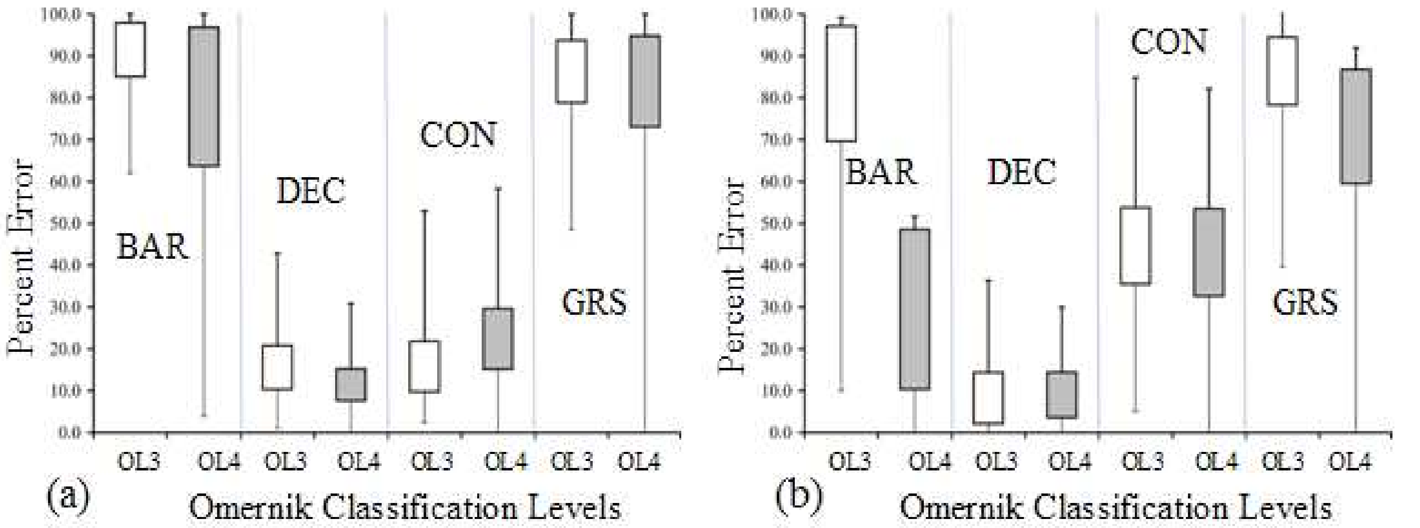

Figure 5. Omission (a) and Commission (b) error boxes and whisker plots bracketing ± 1.0 standard deviation about the mean, and minimum and maximum value outliers for two classification scales (OL3 and OL4) across the OL4 (n=30).

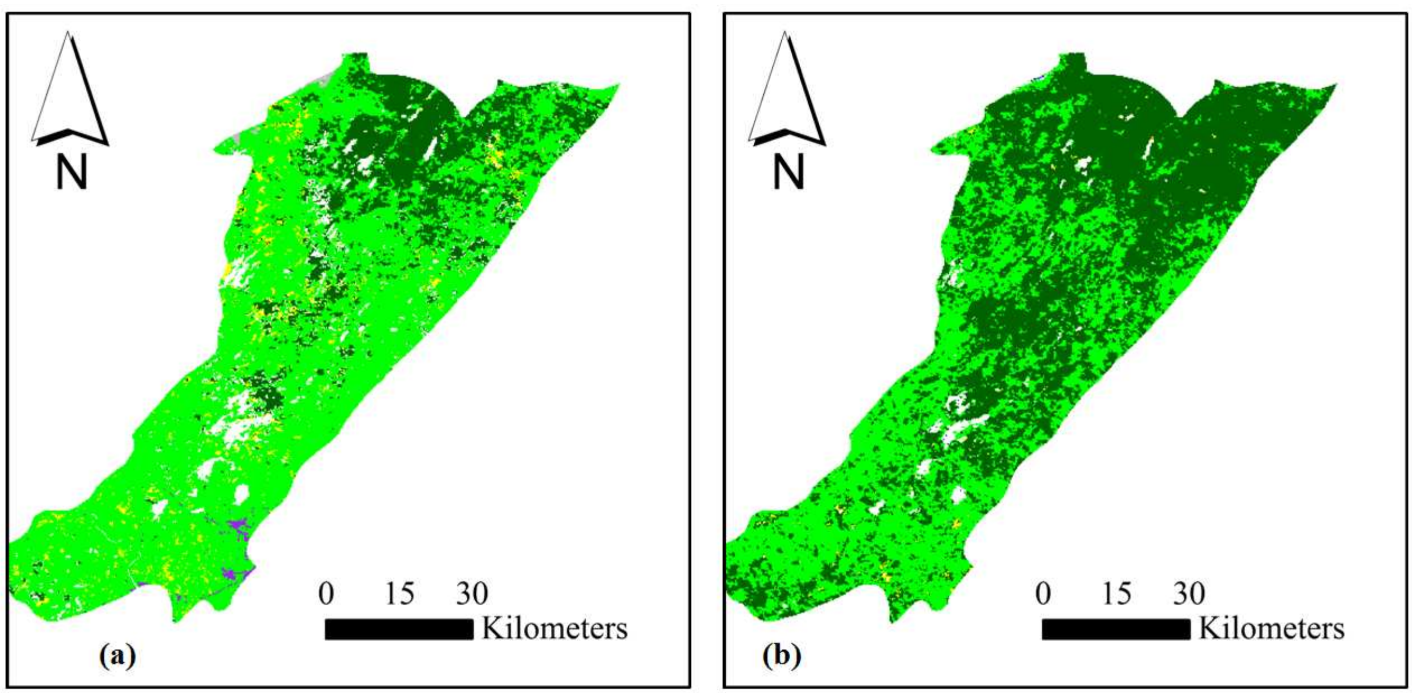

Figure 6. A comparison classification scale results for OL3 (a) and OL4 (b) classifications across the OL4 (dark green – coniferous forest and light green – deciduous forest).

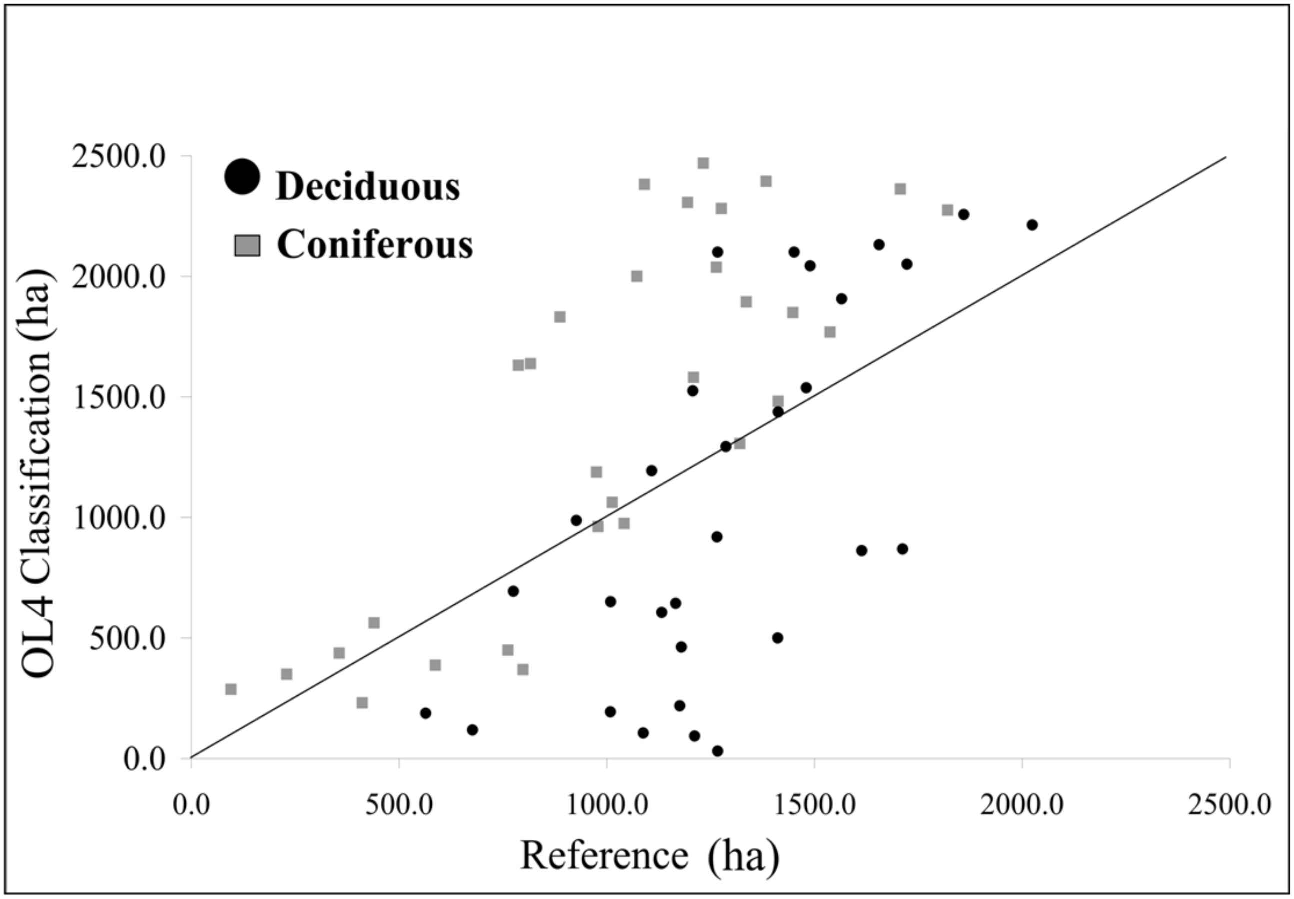

Figure 7. The correspondence between deciduous and coniferous for reference maplets data (n = 30) across the OL4. The line indicates a 1:1 correspondence.

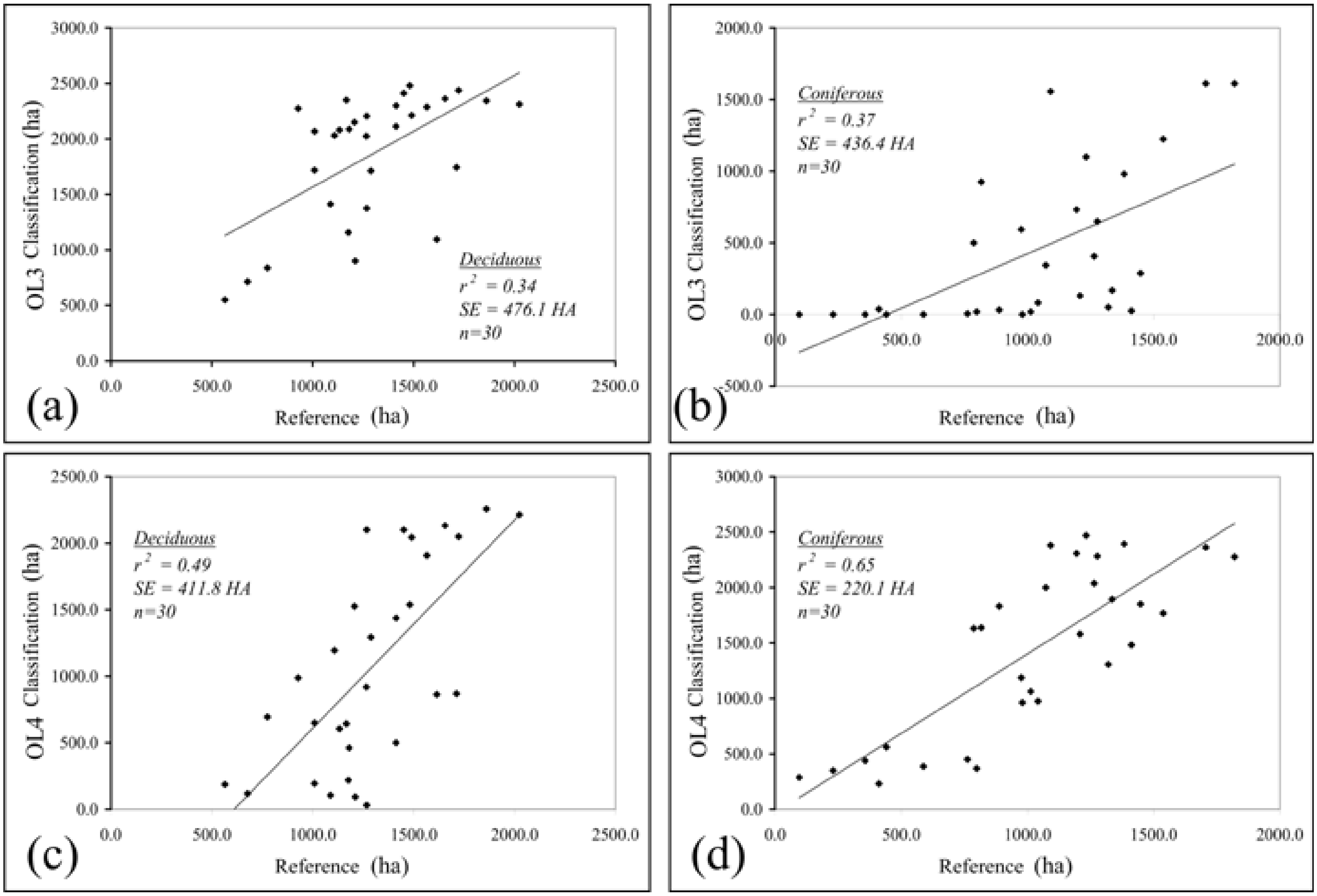

Figure 8. The regression coefficient results for OL3 (a, b) and OL4 (c, d) scale classifications for deciduous and coniferous forests across the OL4.