Land Cover Mapping of North and Central America - Global Land Cover 2000

Rasim Latifovic[1], Zhi-Liang Zhu[2], Josef Cihlar[1], Chandra Giri[2], Ian Olthof[3]

1) Natural Resources Canada, Canada Centre for Remote Sensing (CCRS), Ottawa, ON, Canada

e-mail: Rasim.Latifovic@geocan.nrcan.gc.ca

2) US Geological Survey EROS Data Centre (EDC), Sioux Falls, SD, USA

3) Noetix Research Inc., Ottawa, Canada

Corresponding author
Rasim Latifovic
Application Division
Canada Centre for Remote Sensing
Natural Resources Canada
588 Booth Street
Ottawa, Ontario K1A 0Y7
Canada
Phone (613) 947-1816
Fax: (613) 947=1383

Abstract

The Land Cover Map of North and Central America for the year 2000 (GLC 2000-NCA), prepared by NRCan/CCRS and USGS/EROS Data Centre as a regional component of the Global Land Cover 2000 project, is the subject of this paper. A new mapping approach for transforming satellite observations acquired by the SPOT4/VGTETATION (VGT) sensor into land cover information is outlined. The procedure includes: 1) conversion of daily data into ten-day composite; 2) post-seasonal correction and refinement of apparent surface reflectance in ten day composite images; and 3) extraction of land cover information from the composite images. The pre-processing and mosaicking techniques developed and used in this study proved to be very effective in removing cloud contamination, BRDF effects, and noise in SWIR. The GLC 2000–NCA land cover map is provided as a regional product with 28 land cover classes based on modified Federal Geographic Data Committee/Vegetation Classification Standard (FGDC NVCS) classification system, and as part of a global product with 22 land cover classes based on Land Cover Classification System (LCCS) of the Food and Agriculture Organisation. The map was compared on both areal and per-pixel bases over North and Central America to the International Geosphere-Biosphere Programme (IGBP) global land cover classification, the University of Maryland global land cover classification (UMd) and the MODIS Global land cover classification produced by Boston University (BU). There was good agreement (79%) on the spatial distribution and areal extent of forest between GLC 2000–NCA and the other maps, however, GLC 2000–NCA provides additional information on the spatial distribution of forest types. The GLC 2000–NCA map was produced at the continental level incorporating specific needs of the region.

1. Introduction

Land cover mapping at coarse spatial resoluton provides key environmental information needed for scientific analyses, resource management and policy development at regional, continental and global levels. Coarse resolution satellite data have often been used to map global or regional land cover, particularly those acquired by the Advanced Very High Resolution Radiometer (AVHRR) (Townshend, 1994; Loveland and Belward, 1997; DeFries et al., 1998; DeFries and Townshend, 1999; Cihlar, 2000). Along with the emerging needs for new land cover information from diverse user communities, the development of sensing technologies and information extraction methodologies from remotely sensed data has also advanced. A variety of remotely sensed data and approaches adapted to the specific application requirements have resulted in products with thematic resolution ranging from a few land cover types to those with a considerable level of detail.

Currently available datasets acquired by the SPOT/VEGETATION (VGT) and the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors provide opportunities for generating new land cover products at global and regional scales. The Global Land Cover 2000 project (GLC 2000), lead by the Global Vegetation Monitoring Unit (GVMU) of the Joint Research Centre of the European Commission (EC/JRC), was launched in recognition of emerging opportunities for producing better land cover products. The GVMU is coordinating and implementing the GLC 2000 in collaboration with a network of 30 partners around the world. The main goal of the project is to produce a harmonized land cover database of the world for the year 2000. The year 2000 is considered as a reference year for environmental assessment in relation to various activities; the millennium ecosystem assessment in particular.

In producing a global land cover map at 1 km spatial resolution, the GLC 2000 project has taken advantage of SPOT/VGTATION satellite data provided under the auspices of the Vegetation Data for the Millennium Ecosystem Assessment, the VEGA 2000 initiative. The scope of this initiative is to offer the scientific community the opportunity to make use of SPOT/VEGETATION data to contribute to the objectives of the Millennium Ecosystem Assessment Initiative.

The GLC 2000 project has adopted a multiple thematic resolution approach based on Land Cover Classification System (LCCS) of the Food and Agriculture Organisation (Di Gregorio and Jansen, 2000). The LCCS is designed to accommodate more detailed land cover

characterisation at national or regional scales, but also through its hierarchical structure to accommodate coarser thematic resolution at the global scale.

To ensure compatibility among regional products, the dichotomous modular-hierarchical approach is maintained at the global level by using a set of independent common diagnostic criteria. These criteria are used to translate the regional legends to the global legend accepted at the project level. More details on the classification system, concepts, and diagnostic criteria for each dichotomous and hierarchical level are provided in (Di Gregorio and Jansen, 2000).

The U.S. Geological Survey's EROS Data Center and Canada Centre for Remote Sensing (CCRS) of Natural Resources Canada (NRCan) were participants in the GLC 2000 and contributed to the project by producing a land cover map of North and Central America. (GLC 2000-NCA) In this paper, we present results of GLC 2000-NCA initiative including results of a comparative analysis between GLC 2000-NCA and previously available global land cover products over North and Central America.

2. Methods

2.1. Data and pre-processing

The GLC 2000 project employs the VEGA 2000 dataset consisting of pre-processed daily global composites acquired by the VEGETATION instrument on board the SPOT 4 satellite. The VEGA 2000 dataset includes daily global products S01 at full resolution (1km) for the period from November 1, 1999 to December 31, 2000. Each composite contains four at-surface apparent reflectance bands: blue (0.43-0.47 μm), red (0.61-0.68 μm), near infrared (0.78-0.89 μm), and shortwave infrared (1.58-1.75μm) and six pseudo bands including composite time specifying the acquisition time for each pixel, solar azimuth (SA), solar zenith (SZ), view zenith (VZ), and view azimuth (VA) angles.

The North and Central American data subsets were enclosed in two separate geographical windows: 1) a North American window and 2) a Central American window with geographical extents specified in (Table 1).

Table 1. VEGA 2000 data windows

Geographic Window	Upper left corner		Low right corner	
	Longitude [°]	Latitude [°]	Longitude [°]	Latitude [°]
North America	–180.00	75.00	–13.00	40.00
Central America	–125.00°	50.00°	-50.00	00.00

The first step in pre-processing was to compositing single day images using maximal (Normalized Difference Vegetation Index (NDVI) criteria. Overall, 36 ten-day composites were generated for the period from January to December 2000. Visual assessment of 10 days composites revealed a significant amount of noise caused by the pixel and sub-pixel cloud contamination, strong Bidirectional Reflectance Distribution Function (BRDF) effects due to different viewing geometry and noise in Short Wave Infra-Red (SWIR) band due to sensors malfunctions. To reduce data noise, 10-day composites were corrected for atmospheric effects and normalized to a common viewing geometry. The correction procedure was applied separately on the Central and North American data subsets. After correction, data subsets were combined into a single data set and used for mapping. Data processing was performed using VGT Manager; a software package specially developed to work with SPOT/VEGETATION seasonal satellite observations. The VGT Manager includes several processing modules, each of them addressing a different source of noise. Fig 1 shows the internal structure of the VGT Manager, while methodological bases on which modules were developed are outlined in the following paragraphs:

Identification of contaminated pixels

The CECANT (Cloud Elimination from Composites using Albedo and Normalized difference vegetation index Trend) (Cihlar, 1996; Cihlar et al., 1999) algorithm was used to identify satellite observations obscured by the atmospherics effects. The pixel-specific algorithm is based on a high contrast between land and clouds or snow/ice, especially for land covered by green

vegetation, and on the difference between measured and expected NDVI values for a given pixel and composite period.
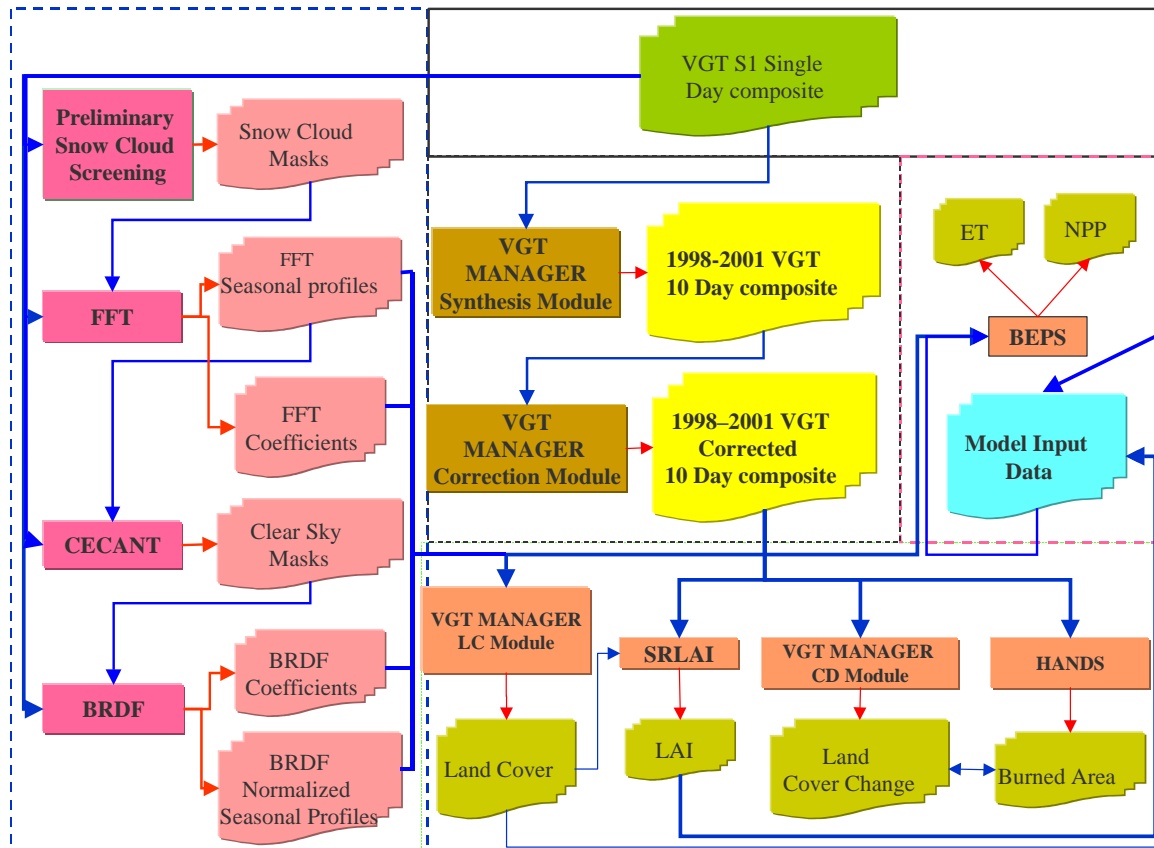
Figure 1. VGT Manager processing modules with structural organisation

Normalization to a common viewing geometry

The BRDF normalization module of the VGT Manager was built using the NTAM (Non-linear Temporal Angular Model) bi-directional reflectance model (Latifovic et al., 2001). The module enables derivation of spectral band-specific NTAM coefficients on a pixel-by-pixel basis (using the VGT S01) or on a land cover type basis (using VGT S10). In this study, the BRDF model coefficients for each VGT band were derived from a set of samples stratified by land cover. The IGBP land cover map was used to stratify and randomly select sample sets for individual land cover types during the entire growing season. For any time period, only cloud-free

pixels (as indicated by CECANT) were included in the derivation of the model coefficients. The pixel land cover information and band- specific model coefficients were then used to normalize the reflectance of clear sky pixels to a common viewing geometry (view zenith angle of 0° and solar zenith angle of 45°). A similar approach to normalize satellite measurements to common viewing geometry has been employed by various investigators in the past (e.g. Gutman 1994; Li et al., 1996; Hu et al., 1996, and Roujean 1992)

A sampling design that uses land cover information from another source may introduce some land cover information into the normalized data, but only on a very general level and not spatially explicit. Thus, we assume a negligible influence on the spectral clustering due to reflectance normalization based on land cover dependent BRDF model coefficients.

Temporal interpolation for contaminated pixels.

Missing and contaminated values in the pixels' seasonal profile are derived by temporal interpolation (Cihlar et al., 1997). The interpolation procedure is designed around the assumption that 10 days composites provide enough measurements needed for interpolation of missing points on the seasonal profile. For pixels obscured during the beginning and the end of the growing season, the apparent surface reflectances were estimated using a polynomial interpolation, while linear interpolation between adjacent non-contaminated values was used during the growing season.

2.2. Classification methodology

Land cover mapping was accomplished in two phases. The first phase included data correction, initial clustering and cluster agglomeration. This phase was accomplished by NRCan/CCRS and resulted with an image of 78 clusters. Phase two included agreement between USGS/EROS Data Centre and NRCan/CCRS on a common classification legend and labelling spectral clusters into land cover types. The second phase also included intensive after labelling consultations to confirm agreement on thematic meaning for each cluster.

Initial clustering and agglomeration was carried out by classification procedure that combines unsupervised and supervised classification approaches. Enhancement-Classification Method (ECM) (Beaubien et al., 1999) and Classification by Progressive Generalization (CPGcs) (Cihlar et al., 1998b; Latifovic et al., 1999) were combined to capture most of the land cover information visible in enhanced images. The applied procedure includes the following three steps:

- Initial clustering
- Cluster agglomeration based on spectral similarity
- Cluster agglomeration/splitting based on phenological similarity

Initial clustering

Input channels into clustering algorithm were averaged values of the corrected surface reflectance for the peak of greenness period from July 1 to August 31, 2000. Mean values were computed for the blue, red, Near Infra-Red (NIR), and SWIR bands, and then linearly scaled from 16 to 8 bits to facilitate processing and visualization. Our decision to use reflectance measurements for initial clustering instead of NDVI was based on several considerations: 1) the SWIR band available in SPOT/VGT data offers additional information that can improve discrimination of vegetation and other land cover types; 2) our goal was high thematic resolution (28 land cover classes), the NDVI has been successfully used only for discriminating very broad land cover types (DeFries and Townshend, 1999); 3) our low confidence in corrected data from the beginning and the end of the growing season when measurements are strongly affected by the presence of clouds, haze snow and ice.

A large initial cluster set was generated using the Iterative Self-Organizing Data Analysis Technique (ISODATA) (Bezdek, 1973) unsupervised clustering algorithm. The output image with 150 clusters was visually assessed against the three input bands to evaluate the initial clustering result. For this purpose, the cluster centre coordinates in red-green-blue space were used to build clusters' pseudo colour table. Visual comparison between the cluster image and three input bands showed very close resemblance, indicating that most of the information in the original spectral data was captured.

A high number of initial clusters, in our case 150, is required to ensure extraction of all land cover information. Starting with a small number of initial clusters often results in a mixed land cover content of these clusters and the loss of small, but potentially important land cover types. This is because clustering procedures are designed to partition a feature's spectral space by optimising the information preservation criteria, regardless of the thematic meaning or cluster importance to the classification. Similar concerns explain our decision for using only monthly means for initial clustering instead of the entire growing season. Ideally, each clusters represents a statistically similar population of pixels with a unique spectral characteristic, however, a pixel's

spectral characteristic changes during the growing season, thereby increasing the overlap between clusters in multi-spectral and multi-temporal space. Many studies have shown that class separability and ultimately classification accuracy rise initially with an increase in the number of discriminating bands used to a point beyond which the addition of data acquired in other bands has either no significant effect or results in a decrease in classification accuracy (Foody, G., and M. Arora 1997; Nelson et al. 1984; Piper 1992; Shahshahani and Landgrebe 1994).

The approach used in this study attempted to utilize seasonal temporal data in a more controlled way. The seasonal profiles were used for assessing cluster quality and similarity between clusters to aid cluster agglomeration.

Cluster agglomeration based on spectral similarity

First generalization step of the initially clustered image was based on the cluster pair's spectral similarity. The clusters pairs with low separability measured by the Jeffries Mastusuta's distance were combined into a single cluster. In this manner, the initial 150 cluster image was agglomerated into an image with 100 clusters. The agglomerated image was than visually assessed by comparing its information content against the enhanced three-band image, ascertaining that all appropriate signatures were still present in classified image. New signature characteristics for all the combined clusters were recomputed and used in the next generalization step.

Clustering agglomeration/splitting based on temporal data

A further generalization step applied to all remaining clusters was based on the similarity of their seasonal profiles. The information available in the clusters' average seasonal profile was used for cluster agglomeration in the following way.

- A clusters' average seasonal profile was derived from twenty 10-day composites (from April to October) by averaging all pixels that belonged to a given cluster in each 10-day composite.
- The absolute difference between profiles was computed for all clusters pairs and employed as a merging criterion in the interactive agglomeration phase.

Phenological information contained in seasonal profiles was also exploited for cluster splitting. A cluster with an inappropriate land cover mixture was reclassified using composites that represented different pheonological events during the growing season. For example, clusters

distributed in the Central Lowland area of US where initial clusters contained a mixture of broadleaf and high biomass crops were reclassified using composites from September when broadleaf and harvested crops were more readily separable.

The 100-cluster image was further combined into spectrally distinct classes using clusters' average seasonal profile matching criteria. In this generalization step, 100 clusters were merged into 78 clusters. Six of the initial clusters were re-classified by employing the cluster splitting approach.

Cluster agglomeration based on thematic similarity – cluster labelling

The labelling procedure employed in this project included independent labelling performed separately by two teams. USGS/EROS Data Center and NRCan/CCRS teams labelled the 78-cluster image separately using a common classification legend. USGS/EROS Data Center assigned the preliminary land cover type to all clusters over the region that included the U.S., Mexico and Central America, and NRCan/CCRS to all clusters over the region that included Canada and Alaska.

The iterative labelling of clusteres were carried out using available digital and map-based information. Overall, more than 200 reference data souces including National Land Cover Data (NLCD) of US, Land Use/Land Cover map of Mexico, Land Cover data of Canada, forest classification maps, ecoregions, vegetation distribution, soils, land use/land cover, and elevation data were used in the process. In case of uncertainty, Landsat TM/ ETM+ data was used to verify the results. Similar labelling principles were employed earlier in producing other global land cover products (Belward et .al 1999; Lovelend et al., 1999). The preliminary assigned labels by both teams were then combined into 28 land cover types using the following set of rules:

1. A cluster distributed in both regions and assigned to the same land cover type was accepted without further examination,
2. A cluster distributed only over one region and confirmed by the other team was accepted,
3. A cluster distributed in both regions and assigned to different land cover types was reconsidered with two possible outcomes:
   - A common label was found
   - The cluster was split because it really presented different land cover types in different labelling regions.

The spatial and temporal components that existed in the image data were employed in interpretative agglomeration, together with analyst knowledge and other ancillary data. The decisions regarding which clusters to combine were influenced by the classification legend, requirements of the intended application, the confidence that a cluster represented a specific land cover type, and the availability of ancillary data or expert knowledge of surface properties for the mapped area. The final interactive phase of agglomeration process allowed iterative adjustments of classification results until the satisfactory accuracy was reached. An alternative to an interpretive agglomeration would be an automated cluster agglomeration procedure, which can make use of spectral, temporal and spatial similarity between clusters or pixels. However, a fully automated agglomeration approach offers limited capability because it usually includes only image information.

Post-classification operations

Post classification operations included some additional image processing performed in cases of known confusion. Spectrally similar classes such as low biomass cropland and grassland were confused with each other or with tundra. Therefore, they were separated from spectrally similar pixels in the northern forest and treeless regions by masking the area south of the boreal forest and changing the class labels for the appropriate clusters. The water and urban areas land cover classes were imported from the updated Digital Chart of the World (DCW) database available at EROS Data Center (EDC).

3. Results

The GLC2000-NCA (Fig. 2.) provides several levels of thematic detail based on three land cover legends: the LCCS based GLC 2000 global level legend with 23 classes (Table 2), the FGDC NVCS based legend with 28 classes (Tables 3a and 3b), and the IGBP legend with 17 classes. A translation table of LCCS, FGDC NVCS, and IGBP legends are summarized in Tables 3a and 3b.

A modified FGDC NVCS classification system was used to design the classification legend for the regional GLC2000 - NCA product. A modification of the FGDC NVCS was considered necessary because the FGDC NVCS (i) was developed mainly for vegetation mapping
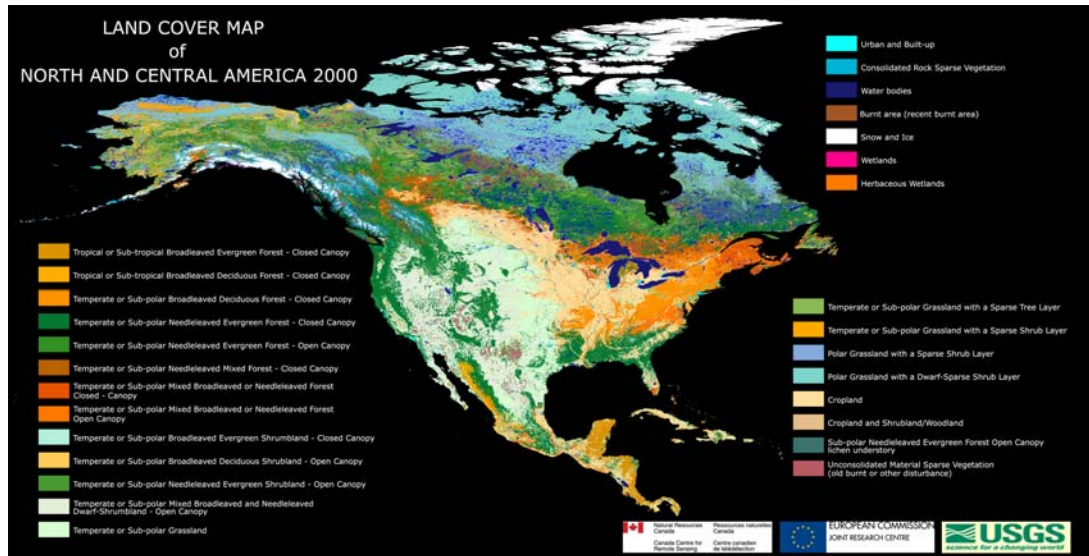
Figure 2. GLC 2000 - North America regional land cover product derived from VGT 2000 data

and does not cover all aspects of land cover characterization, (ii) the system was developed mainly for field applications and may not necessarily be suitable for remote sensing applications, and (iii) it was developed primarily for the U.S., and therefore some land cover classes found in Canada, Mexico or Caribbean are not included in the classification scheme. CCRS has developed a classification scheme that is compatible to FGDC NVCS up to the hierarchical level of subgroup and suitable for Canada (Cihlar et al., 2002).

The modified FGDC NVCS, developed for North and Central America, is a hierarchical classification scheme based on land cover, climate, life form, leaf type, seasonality and canopy cover. Definitions and detailed descriptions of these parameters can be found at http://www.fgdc.gov/standards/status/sub2_1.html. A total of 112 classes can be accommodated in this classification system, out of which 28 land cover classes were observed or selected for regional purposes. These 28 classes, presented in (Tables 3a and 3b), were later regrouped into 23 classes equivalent to the LCCS classification legend for integration of the North and Central American land cover classification into the global GLC 2000 product.

Table 2. GLC 2000 Global Land Cover Legend

Nr.	Decription
1	Tree Cover, broadleaved, evergreen
2	Tree Cover, broadleaved, deciduous, closed
3	Tree Cover, broadleaved, deciduous, open
4	Tree Cover, needle-leaved, evergreen
5	Tree Cover, needle-leaved, deciduous
6	Tree Cover, mixed leaf type
7	Tree Cover, regularly flooded, fresh
8	Tree Cover, regularly flooded, saline, (daily variation)
9	Mosaic: Tree cover / Other natural vegetation
10	Tree Cover, burnt
11	Shrub Cover, closed-open, evergreen (with or without sparse tree layer)
12	Shrub Cover, closed-open, deciduous (with or without sparse tree layer)
13	Herbaceous Cover, closed-open
14	Sparse Herbaceous or sparse shrub cover
15	Regularly flooded shrub and/or herbaceous cover
16	Cultivated and managed areas
17	Mosaic: Cropland / Tree Cover
18	Mosaic: Cropland / Other natural vegetation
19	Bare Areas
20	Water Bodies (natural & artificial)
21	Snow and Ice (natural & artificial)
22	Artificial surfaces and associated areas
23	No Data

Table 3a. Legend based on FGDC NVCS and LCCS classification system

USER_CLASS_NAME	LCCS Level (code meaning provided in LCCS Classification Concept and User Manual FAO UN 2000)	NVCS FGDC	Global Legend	IGBP
Tropical or Sub-tropical Broadleaved Evergreen Forest - Closed Canopy	A3 A10 B2 XX D1 E1-O1 / A3 A10 B2 XX D1 E1-O2	1	1	2
Tropical or Sub-tropical Broadleaved Deciduous Forest - Closed Canopy	A3 A10 B2 XX D1 E2-O1 / A3 A10 B2 XX D1 E2-O2	2	2	2
Temperate or Sub-polar Broadleaved Deciduous Forest - Closed Canopy	A3 A10 B2 XX D1E2-O5 / A3 A10 B2 XX D1 E2-O7	3	5	4
Temperate or Sub-polar Needleleaved Evergreen Forest - Closed Canopy	A3 A10 B2 XX D2 E1-O5 / A3 A10 B2 XX D2 E1-O7	4	4	1
Temperate or Sub-polar Needleleaved Evergreen Forest - Open Canopy	A3 A11 B2 XX D2 E1-A12-O5 / A3 A11 B2 XX D2 E1-A12-O7	5	4	1
Temperate or Sub-polar Needleleaved Mixed Forest - Closed Canopy	A3 A10 B2 XX D2 E1-O5 / A3 A10 B2 XX D2 E2-O5 / A3 A10 B2 XX D2 E2-O7	6	6	5
Temperate or Sub-polar Mixed Broadleaved or Needleleaved Forest - Closed Canopy	A3 A10 B2 XX D1-O5 / A3 A10 B2 XX D1-O7 / A3 A10 B2 XX D2-O5	7	6	5
Temperate or Sub-polar Mixed Broaddleleaved or Needleleaved Forest - Open Canopy	A3 A11 B2 XX D1-A12-O5 / A3 A11 B2 XX D1-A12-O7 / A3 A11 B2 XX D2-A12-O5	8	6	5
Temperate or Subpolar Broadleaved Evergreen Shrubland - Closed Canopy	A4 A10 B3 XX D1 E1-O5 / A4 A10 B3 XX D1 E1-O7	9	12	7
Temperate or Subpolar Broadleaved Deciduous Shrubland - Open Canopy	A4 A11 B3 XX D1E2-A12-O5 / A4 A11 B3 XX D1E2-A12-O7	10	12	6
Temperate or Subpolar Needleleaved Evergreen Shrubland - Open Canopy	A4 A11 B3 XX D2 E1-A12-O5 / A4 A11 B3 XX D2 E1-A12-O7	11	11	7
Temperate or Sub-polar Mixed Broadleaved and Needleleaved Dwarf-Shrubland - Open Canopy	A4 A11 B3 XX D1-A12 B10-O5 / A4 A11 B3 XX D1-A12 B10-O7 / A4 A11 B3 XX D2-A12 B10-O5	12	12	7
Temperate or Subpolar Grassland	A6 A10 -05/A6 A10-07	13	13	10
Temperate or Subpolar Grassland with a Sparse Tree Layer	A6 A10-05/A6 A10-07/A3 A14-A16	14	9	8
Temperate or Subpolar Grassland with a Sparse Shrub Layer	A6 A10-05/A6 A10-07/A4 A14-A16	15	13	8
Polar Grassland with a Sparse Shrub Layer	A6 A10-08/A4A14-A16	16	14	16
Polar Grassland with a Dwarf-Sparse Shrub Layer	A6 A10-08/A4 A14 B3-A16 B10	17	14	16
Cropland	A3	18	16	12
Cropland and Shrubland/woodland	A3/A4 A14	19	18	12
Subpolar Needleleaved Evergreen Forest Open Canopy - lichen understory	A8 A10-O5 / A8 A10-O7	20	4	8
Unconsolidated Material Sparse Vegetation (old burnt or other disturbance)	A2/A4 A14	21	14	7
Urban and Built-up	A1	22	22	13
Consolidated Rock Sparse Vegetation	A1/A2	23	14	16
Water bodies	A1	24	20	17
Burnt area (recent burnt area)	A2/A4 A14	25	10	16
Snow and Ice	A2/A3	26	21	15
Wetlands	A3/A13	27	15	11
Herbaceous Wetlands	A2/A13	28	15	16

(SEMI)NATURAL TERR. VEGETATION	Code
I . A. Life form of the Main strata	
Woody	A1
Trees	A3
Shrubs	A4
Herbaceous	A2
Forbs	A5
Graminoids	A6
Lichens/Mosses	A7
Lichen	A8
Mosses	A9
A. Cover	
Closed (>70-60%)	A10
Open (70 - 60 – 20-10%)	A11
(70 - 60 – 40%)	A12
(40 – 20 – 10%)	A13
Sparse (20-10 – 10%)	A14
Scattered (4-10%)	A15
B. Height	
7-2 m (for Woody)	B1
>30 – 3m (for Trees)	B2
>14	B5
14-7m	B6
7-3	B7
5-0.3m	B3
5-0.5m	B14
5-3	B8
3-0.5m	B9
<0.5m	B10
3-0.03m	B4
3-0.3m	B15
0.8-0.3m	B11
0.8-0.3	B12
0.3-0.03	B13
C. Spatial Distribution/Macropattern	
Continuous	C1
Fragmented	C2
Striped	C4
Cellular	C5
Parklike Patches	C3
II. D. Leaf Type	
Broadleaved	D1
Needleleaved	D2
Aphyllous	D3
E. Leaf Phenology	
Evergreen	E1
Semi-Evergreen	E4
Deciduous	E2
Semi-Deciduous	E4
Mixed	E3
Mixed (for Forbs/Graminoids	E5
Annual	E6
Perennial	E7
III. F. Stratification – Second Layer	
Secon Layer Absent	F1
Second Layer Present	F2

Table 3b Overview of Environmental attributes of each major land cover type of the LCCS classification system (Di Gregorio and Jansen, 2000) for decoding Table 3

3.1. Comparative analysis of North and Central America to other global land cover maps

In order to evaluate continental land cover composition, the product was compared with other global land cover classifications, in particular:

- The IGBP Global land cover classification produced by the U.S. Geological Survey for the International Geosphere-Biosphere Programme [IGBP (Loveland et al., 2000)];
- The University of Maryland global land cover classification [UMd (Hansen et al., 2000)];
- The MODIS Global land cover classification produced by Boston University [BU, (Hodges et al., 2001)].

Table 4. Relationships among classification legends used in the four land cover products

IGBP	14 common class for all maps in bold	GLC - 2000 NCA	BU and IGBP	UMd
1	**Evergreen Needleleaf Forest**	4, 5	1	1
2	**Evergreen Broadleaf Forest**	1, 2	2	2
3	**Deciduous Needleleaf Forest**	NA	3	3
4	**Deciduous Broadleaf Forest**	3	4	4
5	**Mixed Forest**	6, 7, 8	5	5
6	**Closed Shrublands**	10	6	8
7	**Open Shrublands**	9, 11, 12, 21	7	9
8	**Woody Savannas**	14, 15, 20	8, 9	6, 8
9	Savannas			
10	**Grasslands**	13	10	10
11	**Permanent Wetlands**	27	11	NA
12	**Croplands**	18, 19	12, 14	11
13	**Urban and Built-Up**	22	13	13
14	Cropland and Other Vegetation Mosaic			
15	**Snow and Ice**	26	15	NA
16	**Barren or Sparsely Vegetated**	16, 17, 23, 25, 28	16	12
17	Water			

These land cover products were created using different classification methods and input data, but with the same fundamental objective of providing improved land cover information for diverse applications. The IGBP classification was created on a continental basis using NOAA/AVHRR acquired maximum NDVI values for 1992-1993 as input into an unsupervised clustering algorithm. The resulting clusters were then merged and labelled through intensive post-classification refinement (Loveland et al., 2000). The UMd classification was created using a supervised classification tree, based on 41 temporal metrics derived from NDVI and all five AVHRR bands. Training sets were derived from a large number of LANDSAT scenes. Selected areas for training were interpreted according to the adopted classification scheme (Hansen et al., 2000). The BU classification uses year 2000 growing season MODIS data at 1km spatial resolution, and fuzzy ARTMAP neural network and decision tree classifiers (Strahler et al., 1999).

Both area and per-pixel basis comparison of land cover types over North and Central America was preformed using a common legend for all four maps, which included the 14 IGBP classes presented in Table 3a and 3b. Lookup tables used to convert the original legends to the common legend are shown in Table 4. All four maps were transformed into common Lambert Azimuth Equal Area (LAEA) projection and co-registered to each other to allow per- pixel comparison. The Root Mean Square (RMS) error accepted was ± 1 pixel.

3.3. Visual, area and per pixel comparison

GLC2000-NCA land cover classification was visually assessed through a systematic comparison to the available fine resolution reference data following procedures suggested by Mayaux (2002). The Satellite SILC data set over Canada and NLCD over U.S. both at ~30 m spatial resolution were used for visual assessment of the GLC2000-NCA map. Assessment of the areas for which fine resolution land cover data were not available was performed through simultaneous interpretation and comparison of LANDSAT satellite images.

Based on visual inspection and comparison, the GLC2000-NCA land cover classification depicts the continental and regional-to-landscape land cover composition well. At the continental level, the land cover map clearly delineates major ecological regions such as Arctic, Tundra, Taiga, Boreal Forest, Plains, Marine Forest and Temperate Forest. At the landscape level, visual assessment was in good agreement with reference data. Figure 3 shows two examples of such comparisons. The first comparison is between the 14 classes GLC2000-NCA map and a

LANDSAT ETM+ scene acquired over Nevada, USA. The second is a comparison with a LANDSAT 7 ETM+ classification for an area west of Lake Winnipeg, Canada using the same colour scheme for both classifications. The agreement between these two classifications appears to be quite high. However, the assumption of high agreement at the pixel level based only on good visual agreement between classifications at different spatial resolutions needs to be confirmed by a rigorous quantitative accuracy assessment.
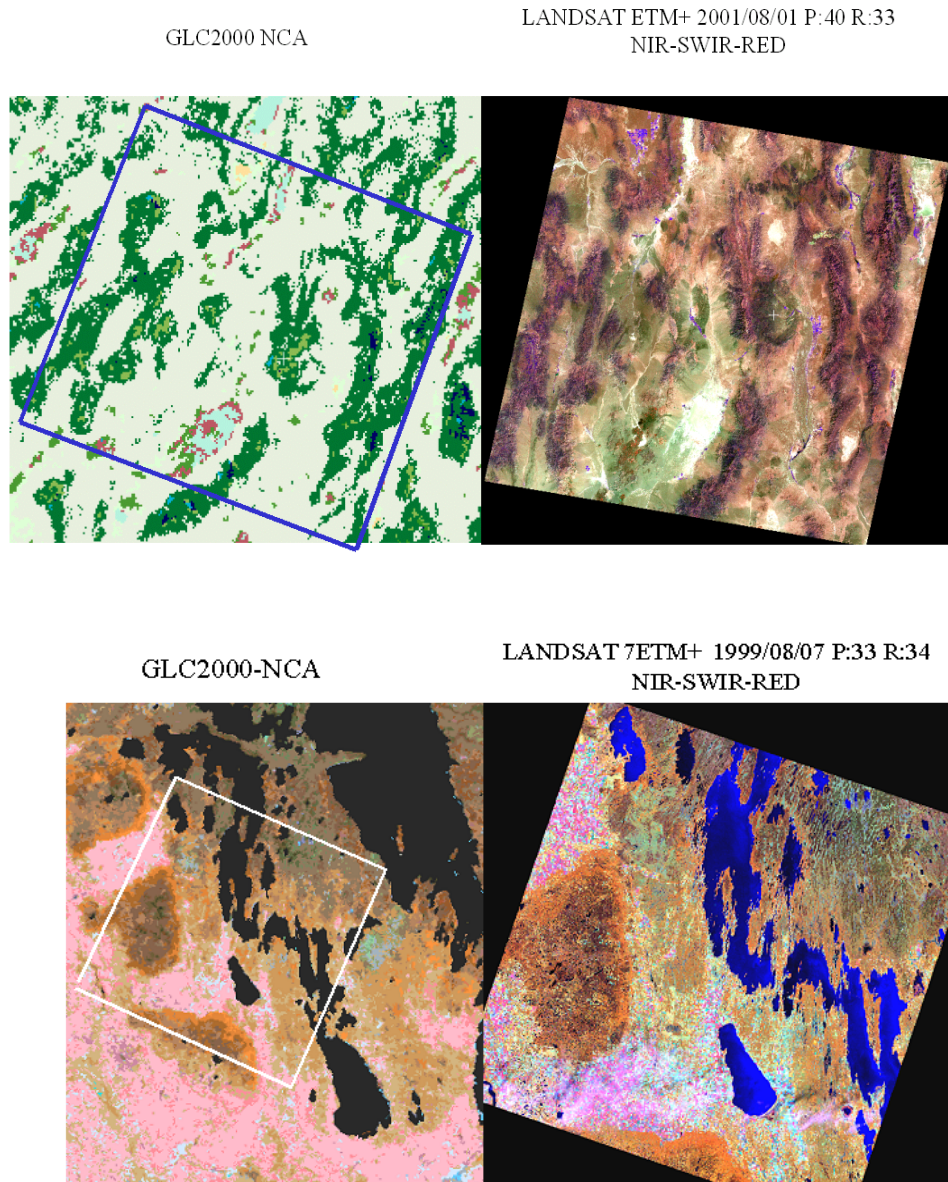
Figure 3. VGT land cover map evaluation by visual comparison with a LANDSAT based classification

Further evaluation of GLC2000-NCA map was carried out by comparison GLC2000-NCA based land cover area estimates per class against estimates derived from other land cover classifications at different levels of detail. Table 4 shows land cover area distributions over North and Central America derived from IGBP, GLC2000-NCA, BU and UMd classifications. The IGBP, GLC2000-NCA and BU classifications had relatively good overall agreement. The average percentage of absolute difference in area per class, estimated relative to GLC2000-NCA were 2.45% for IGBP, 2.72% for BU and 5.06% for UMd. The source of difference between UMd and other classifications was partly due to differences between classification legends; the UMd legend is without snow/ice, permanent wetland classes and cropland/natural vegetation mosaic. However, it appears that a major difference came from the use of different separation criteria between woody savannas, evergreen needleleaf forest and closed shrubland classes. Additional evaluation and comparison on area per land cover class was performed for three most similar classifications i.e. GLC2000-NCA, IGBP and BU.

At the highest aggregation level with only two classes, i.e. forest including woody savannas and non-forest, the area per class differences were in the range of +-4%. However, when looking at the class area distribution inside the forest represented by six forest classes, the differences were greater, ranging from 0.03% in case of evergreen broadleaf forest to 7 % in case of evergreen needleleaf forest. The greatest differences among the products occurred in the areas of mixed and evergreen needleaved forests. These differences might have attributed due to various interpretation criteria and thresholds applied for delineating these two classes.

Agreement was high among maps in depicting the composition of non-forest classes (Table 5), as IGBP, GLC2000-NCA and BU had similar class area distributions. However, differences among non-forest classes were more randomly distributed than among forest classes. The results indicate that classes with better spectral discrimination and less ambiguous class definitions such as forest, cropland, barren land, ice and snow have higher area agreement than classes with less detailed class definitions and more spectral confusion, such as open /close scrublands and wetlands.

Table 5. Percent area of each land cover map occupied by land cover class

Land cover type	GLC2000 NCA	IGBP	BU	UMd
Forest classes	43.5%	47.8%	42.2%	54.0%
Evergreen Needleleaf Forest	21.8%	17.4%	14.4%	10.3%
Evergreen Broadleaf Forest	3.5%	2.6%	4.9%	2.5%
Deciduous Needleleaf Forest	0.0%	0.0%	0.1%	0.0%
Deciduous Broadleaf Forest	5.4%	7.3%	4.5%	4.2%
Mixed Forest	10.4%	13.0%	10.4%	5.6%
Woody Savannas	2.5%	7.5%	8.0%	31.4%
Non-forest classes	54.3%	51.9%	57.8%	46.1%
Closed Shrublands	0.9%	2.4%	0.8%	7.8%
Open Shrublands	16.2%	11.4%	23.6%	13.4%
Grasslands	8.0%	8.7%	10.7%	9.8%
Permanent Wetlands	0.3%	1.5%	0.5%	0.0%
Croplands	14.3%	17.5%	15.7%	9.2%
Urban and Built-Up	0.5%	0.4%	0.4%	0.4%
Snow and Ice	3.0%	1.7%	1.6%	0.0%
Barren or Sparsely Vegetated	13.4%	8.1%	4.4%	5.5%

The results of a pixel-by-pixel comparative analysis among land cover product showed somewhat less agreement than in the case of total land cover proportions (Table 6.) (Fig. 4.) The BU and GLC2000-NCA classifications had the highest agreement in depicting forested areas. The total land cover proportions were, however, balanced by differences in land cover distribution between northern and southern areas. For example, the GLC2000-NCA map depicted significantly less forest in the northern transition zone compared to the other land cover classifications, while the other classifications showed less forest in the southern transition zone compared to GLC2000-NCA (Fig. 4). The Figure 4 also shows that central forested areas are in a good agreement, while disagreement occurs mostly along the edges. Change in forested area, mostly caused by forest fires occurred between 1992 and 2000 accounted for a significant part of the differences between IGBP and UMd classifications produced from 1992 AVHRR data, and GLC2000-NCA and BU classification produced from 2000 data. For illustration, the Canadian Interagency Forest Fire Centre (CIFFC) report on the total area of Canadian forest burnt shows

that 230,000 km^2 forest area burned cumulatively from 1992 to 2000, or approximately 5% of forestland. These areas were mapped in the GLC2000-NCA as recent burns or as open or closed shrubland in regenerating areas of older fires (Fig. 5.), while these same areas were classified as forest in the other maps.
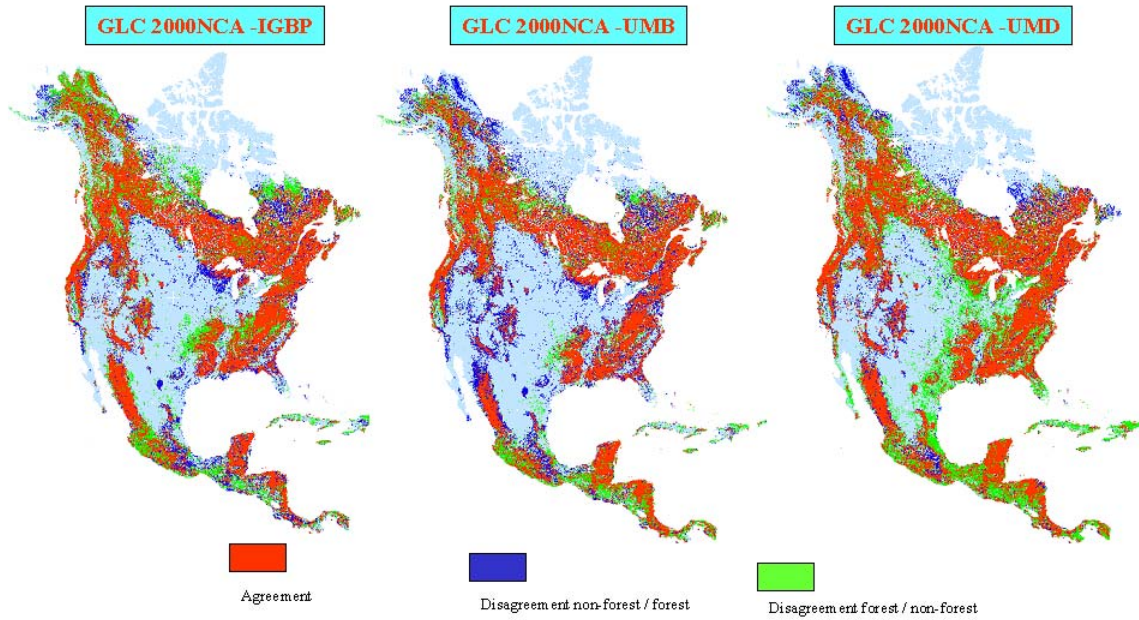

Figure 4. Forest area agreement for the GLC2000-NCA versus IGBP, UMD and BU maps.

Table 6. Agreement in depicting the forest area over North America among analysed land cover maps

	Agreement [%]	Kappa
GLC2000-NCA - IGBP	78.85%	0.554
GLC2000-NCA - BU	79.03%	0.548
GLC2000-NCA - UMd	78.17%	0.490

It appears that the primary reason for remaining differences among maps is the lack of a standardized approach for interpreting remote sensing data, especially in cases off transitional forest – non-forest and other classes. Regardless of the unique class definition, the applied thresholds in interpreting transitional forest from non-forest classes often leads to large differences among land cover classifications. The comparative analysis showed that

discrimination of sparse forest from open shrubland in sub-arctic regions is a major source of difference among maps. Such comparative analyses emphasise the need for international projects where collaborative work between different national agencies carried out by local experts can resolve questions concerning implementation of internationally accepted standards for land cover products.
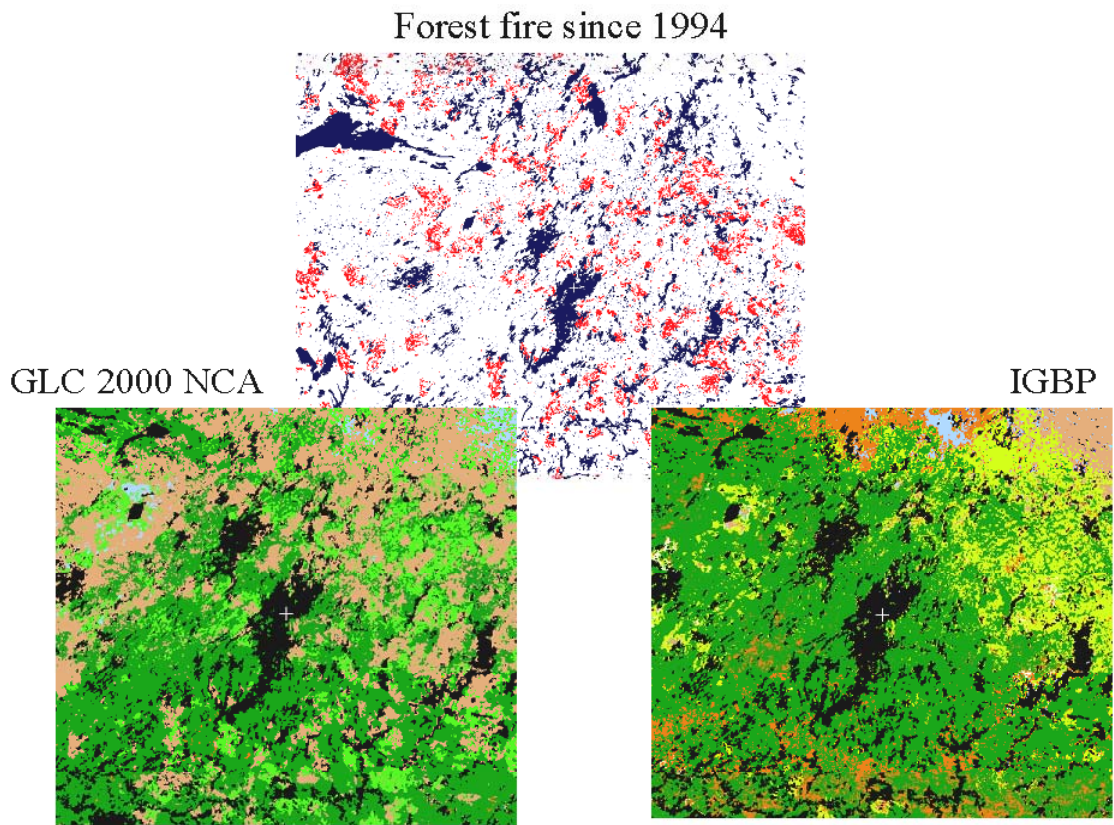
Figure 5 Forest fire occurrence in Canadian boreal forest between 1994 and 2000 (Li et al. 2000)

A detailed quantitative accuracy assessment is being carried out separately for US, Canada, Mexico and the Central American countries using land cover data derived from fine resolution satellite data (Landsat).

4. CONCLUSIONS

A new regional land cover product of North and Central America for the year 2000 has been produced. This product served as an input to GLC-2000 for North and Central America. The land cover classes were, however, aggregated to a higher level following a global legend. Several conclusions may be drawn from the current work.

The VGT data set used for land cover mapping at regional and global scales requires careful planning of all the data processing and analysis steps that demand significant processing time. Correction and normalization of satellite data over full season is a prerequisite for improved land cover classification.

The pre-processing and mosaicking techniques developed and used in this study proved to be very effective in removing cloud contamination, BRDF effects, and noise in SWIR.

Comparisons among four land cover maps over North and Central America showed good agreement in land cover proportions. The average difference on a per-pixel basis was 20%. The spatial distribution of per-pixel differences occurred mostly along class edges and transition zones. Analysis concluded that most of the disagreement at a rather general level resulted from the application of different criteria when delineating classes rather than data limitations.

This paper provides a complete information extraction procedure from raw data to final land cover product including post-seasonal correction before product generation. The classification approach is unique in that it employs seasonal temporal profiles at the spectral cluster level instead of at the pixel level. This approach allows more controlled use of temporal information available in remote sensing data, which decreases the amount of noise in the classified image. The paper also suggests methods to control, assess and improve each generalization step, ultimately leading to a better final product. The paper provides an opportunity to evaluate a new sensor for global and region mapping since the GLC 2000-NCA is the first land cover product over North and Central America that utilizes VGT data. Perhaps most importantly, this paper demonstrates that regional mapping can be performed with success as a collaborative effort in which each side brings local expertise

Future work on GLC2000-NCA will include a full comparative accuracy assessment with the other three land cover products in this paper, and a quantitative analysis of factors affecting map accuracy.

5. Acknowledgements

6. References

Beaubien, J., Cihlar, J., Simard, G., & Latifovic, R. (1999). Land cover from multiple Thematic Mapper scenes using a new enhancement - classification methodology. *Journal of Geophysical Research, 104,* 27909-27920.

Bezdek, J.C. (1973). Fuzzy mathematics in pattern classification. Ph.D. dissertation, Cornell University, Ithaca, NY.

Cihlar, J. (1996). Identification of Contaminated Pixels in AVHRR Composite Images for Studies of Land Biosphere. *Remote Sensing of Environment, 56,* 149-163.

Cihlar, J., Ly, H., Li, Z., Chen, J., Pokrant, H. & Huang, F. (1997). Multitemporal, multichannel AVHRR data sets for for land biosphere studies – artifacts and corrections. *Remote Sensing of Environment, 60*, 35-57.

Cihlar, J., Chen, J., Li, Z., Huang, F., Latifovic, R., & Dixon, R. (1998a). Can interannual land surface signal be discerned in composite AVHRR data? *Journal of Geophysical Research, 10,* 23163-23172.

Cihlar, J., Xiao, Q., Chen, J., Beaubien, J., Fung, K., & Latifovic, R. (1998b). Classification by progressive generalization: a new automated methodology for remote sensing multichannel data. *International Journal of Remote Sensing, 19,* 2685-2704.

Cihlar, J. Latifovic, R., Chen, J., & Li., Z. (1999). Testing Near-Real Time Detection of Contaminated Pixels in AVHRR Composites. *Canadian Journal of Remote Sensing, 25,* 160-170.

Cihlar, J. (2000). Land cover mapping of large areas from satellites: status and research priorities. *International Journal of Remote Sensing,* 21: 1093-1114.

Cihlar, J., Latifovic, R., Chen, J. M., Trishchenko, A., Du, Y., G. Fedosejevs, & Guindon, B. (2003). Systematic corrections of AVHRR image composites for temporal studies. *Remote Sensing of Environment,* (in press).

Cihlar, J., Latifovic, R., Beaubien, J., Guindon, B., & Palmer, M. (2002). TM-based accuracy assessment of a land cover product for Canada derived from SPOT VEGETATION data. *Canadian Journal of Remote Sensing,* (submitted).

DeFries, R.S., Hansen, M., Townshend, J.R.G., & Sohlberg, R. (1998). Global land cover classification at 8km spatial resolution. Part 1: Training and validation data derived from Landsat imagery. *International Journal of Remote Sensing, 19,* 3141-3168.

Danko, D. M., 1992 The Digital Chart of the Word. GeoInfo Systems, 2, 29-36.

DeFries, R.S., & Townshend, J.R.G. (1999). Global land cover characterization from satellite data: from research to operational implementation? *Global Ecology and Biogeography, 8,* 367-379.

Di Gregorio, A., & Jansen, L.J.M. (2000). Land cover classification system (LCCS): classification concepts and user manual. Environment and Natural Resources Service,

GCP/RAF/287/ITA Africover – East Africa Project and Soil Resources, Management and Conservation Service, Food and Agriculture Organization.

Fernandes, R.A., Pavlic, G., Chen, W., & Fraser, R. (2001). Canada-wide 1-km water fraction derived from National Topographic Data Base maps. Natural Resources Canada.

Foody, G., and M. Arora. (1997). An evaluation of some factors affecting the accuracy of classification by an artificial neural network. *International Journal of Remote Sensing,18 ,* (4): 799-810.

Gutman, G. G. (1994). Global data on land surface parameters from NOAA AVHRR for use in numerical climate models. *Journal of Climate, 7,* 699-703.

Hansen, M.C., DeFries, R.S., Townshend, J.R.G., & Sohlberg, R. (2000) Global land cover classification at 1km spatial resolution using a classification tree approach. *International Journal of Remote Sensing, 21,* 1331-1364.

Hodges, J. C. F., Friedl, M. A., & Strahler, A. H. (2001). The MODIS Global Land Cover Product: New Data Sets for Global Land Surface Parameterization, Proceedings of the Global Change Open Science Conference, Amsterdam, 2001.

Hu, B., W. Lucht, A.H. Strahler, C.B. Schaaf and M. Smith. 200 Surface albedo and Angle-Corrected NDVI from AVHRR Observation of South America. *Remote Sensing of Environment, 71*, 119-132

Li, Z., J.Cihlar, X. Zheng. L. Moreau, and H.Ly, 1996. Bidirectional effects of AVHRR measurements over boreal regions. *IEEE Transactions on Geoscience and Remote Sensing, 34,* 1308-1322

Latifovic, R., J. Cihlar, & Chen, J. M. (2001). A comparison of BRDF models for the normalisation of satellite optical data to a standard sun-target-sensor geometry. *I.E.E.E. Transactions on Geoscience and Remote Sensing,* (in press).

Latifovic, R., & Olthof, I. (2003). Accuracy assessment of global land cover products derived from satellite data. *Remote Sensing of Environment,* (submitted).

Latifovic, R., Cihlar, J. & Beaubien, J. (1999). Clustering methods for unsupervised classification. *Proceedings of the 21st Canadian Remote Sensing Symposium*, II-509-II-515, June 1999, Ottawa, Canada.

Latifovic, R., Beaubien, J., & Cihlar, J. (2002). Land cover classification of LANDSAT ETM+ scene 26 July 1999 Path 09/Row 23. Satellite Information for Land cover of Canada (SILC) Project, Natural Resources Canada Centre for Remote Sensing, Ottawa, Ontario.

Li, Z., S. Nadon, J. Cihlar, 2000. Satellite detection of forest fires: Development and application of an algorithm, *International Journal of Remote Sensing*, *21*, 3057-3069.

Loveland, T.R., & Belward, A.S. (1997). The IGBP-DIS global 1 km land cover data set, DISCover: first result. *International Journal of Remote Sensing, 18*, 3289-3296.

Loveland, T.R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of global land cover characteristics database and IGBP DISCover from 1km AVHRR data. *International Journal of Remote Sensing, 21*, 1303-1330.

Mayaux, P. (2002). GLC 2000 project – guidelines for quality control. GLC 2000 meeting, Ispra, Italy, March 2002. 5pp.

Nelson, R. E., Latty, R.S., and Mott, G. (1984). Classifying Northern forests using Thematic Mapper simulator data. *Photogrammetric Engineering and Remote Sensing, 50,* 3263-3270

Pavlic, G., Fernandes, R., Chen, W., Fraser, R., & Leblanc, S. G. (2002). Methods for Deriving Canada Wide Geo-Spatial Datasets in Support of Environmental Monitoring and Modelling. The Joint International Symposium on GeoSpatial Theory, Processing and Applications, Ottawa, Canada, 8-12 July 2002.

Piper, J., (1992). Variability and bias in experimentally measured classifier error rates. *Pattern Recognition Letters, 13,* 685-692

Roujean, J. L., M.J. Leroy and Deschamps P. Y., (1992) A bi-directional reflectance model of the earth's surface for the correction of remote sensing data. *Journal of Geophisical Research, 9* (D18), 20455-20468.

Shahshahani, B. M. and Landgrebe , D.A., (1994). The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. *I.E.E.E. Transaction on Geoscience and Remote Sensing*, *32*, 1087-1095

Strahler, A., Muchoney, D., Borak, J., Friedl, M., Gopal, S., Lambin, E. & Moody, A. (1999). *MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD)*, Version 5.0, NASA EOS-MTPE Documentation, NASA, Washington, DC, 66pp.

Townshend, J.G.R. (1994). Global data sets for land cover applications from the Advanced Very High Resolution Radiometer: an introduction. *International Journal of Remote Sensing, 15,* 3319-3332.