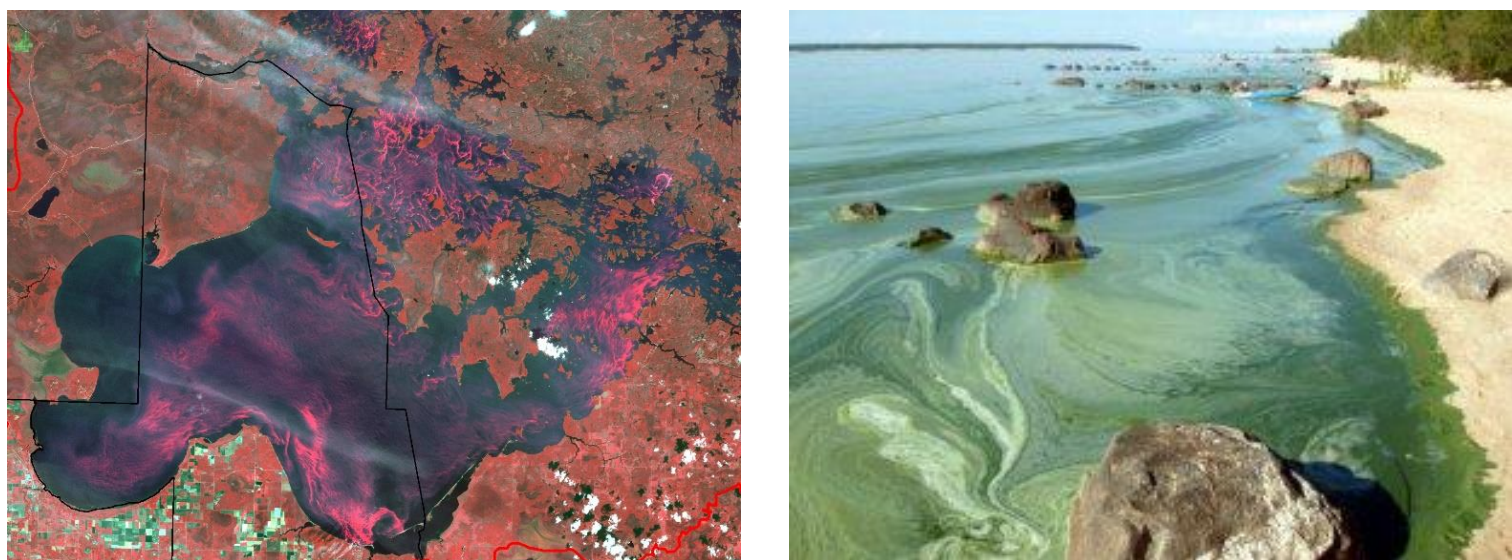# Improved Land Cover Classification by Integrating Landsat Imagery with Lidar and Object-based Image Analysis for Land Cover Classification of the International Lake of the Woods/Rainy River Basin

**Leif G. Olmanson & Marvin E. Bauer** — UNIVERSITY OF MINNESOTA

Landsat August 28, 2001

Rationale and Background

With a perceived increase in the frequency and intensity of cyanobacterial algal blooms in Lake of the Woods (LOW), there has been an increased effort to collect information about the nature of algal blooms, nutrient concentrations and sources of nutrients to the LOW. As part of this effort, land cover maps of the LOW Watershed are needed as inputs to hydrology models and analyses of land use and land use change.

Site Location and Size

The international Lake of the Woods/Rainy River Basin is an important natural, economic and recreational resource straddling the border between Canada and the United States. The Basin is approximately 400 km (240 mi) east-to-west and 260 km (156 mi) north-to-south. Nearly 60% of its drainage area of about 70,000 km² (27,000 mi²) is located in Ontario, with about 40% in Minnesota and a small portion in Manitoba.
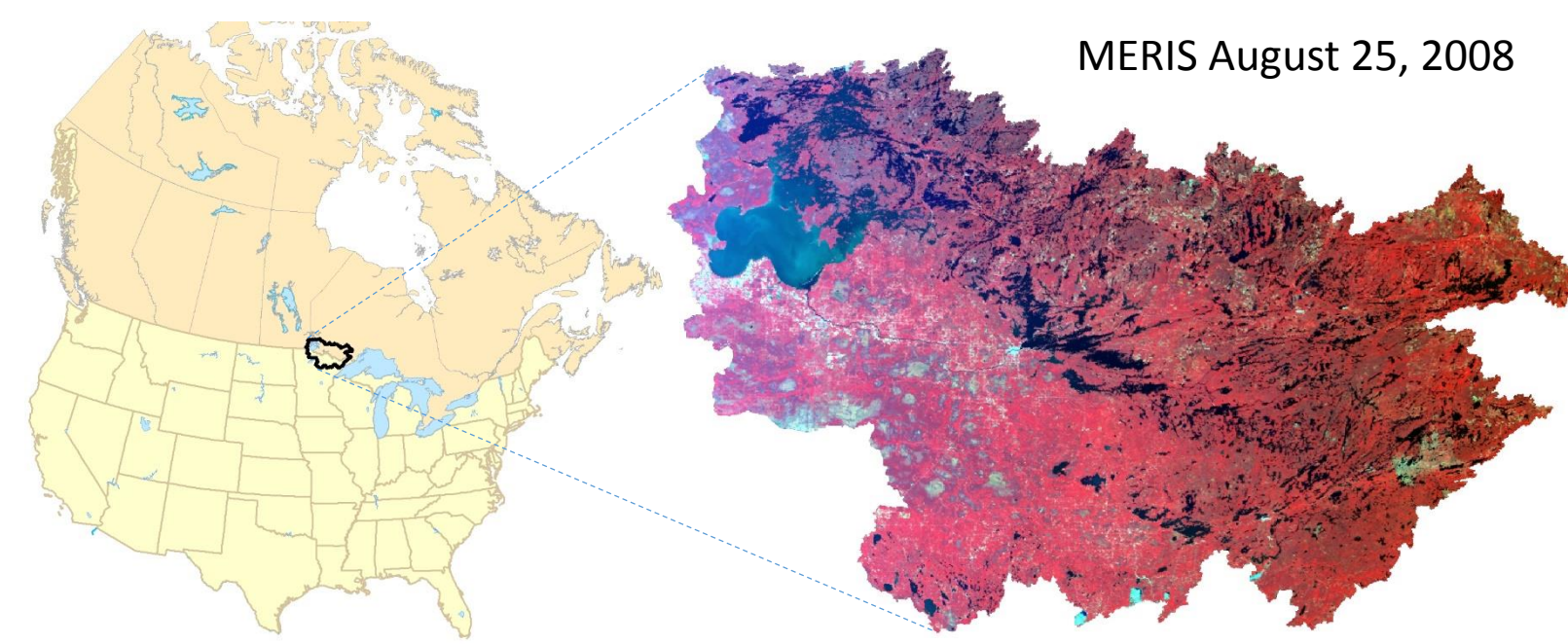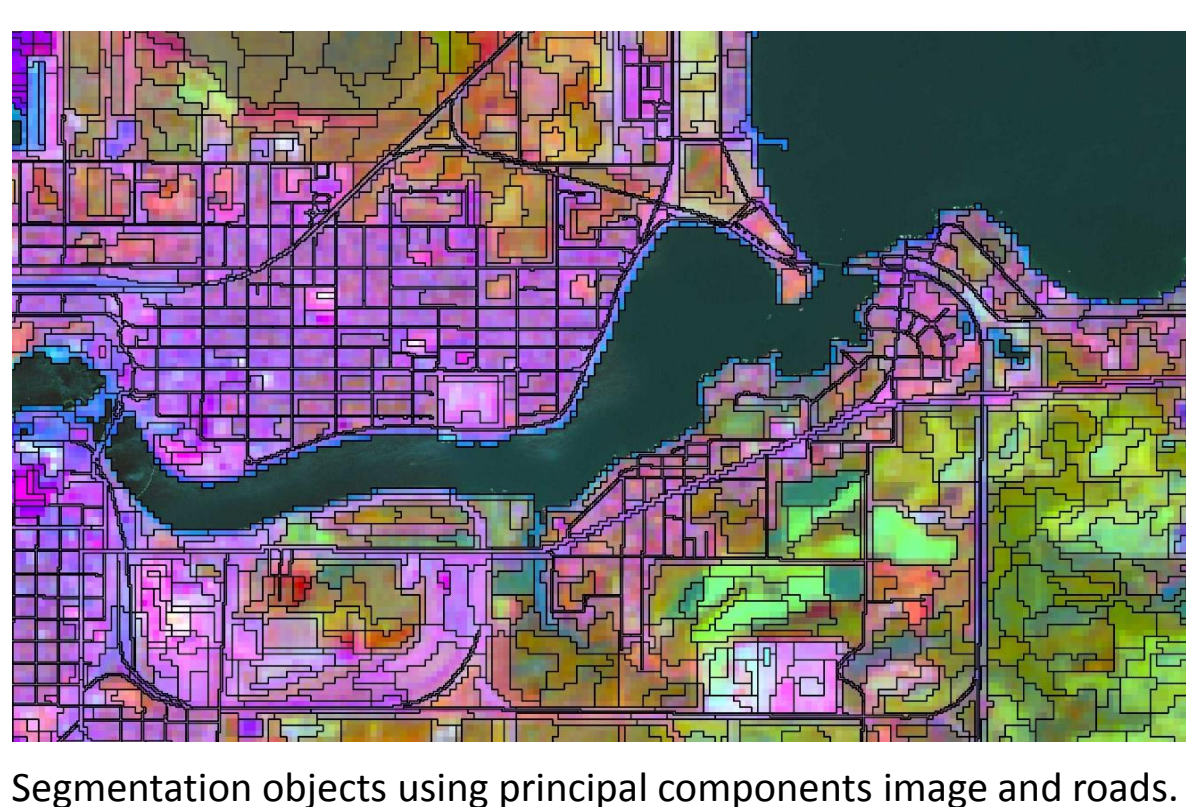

MERIS August 25, 2008
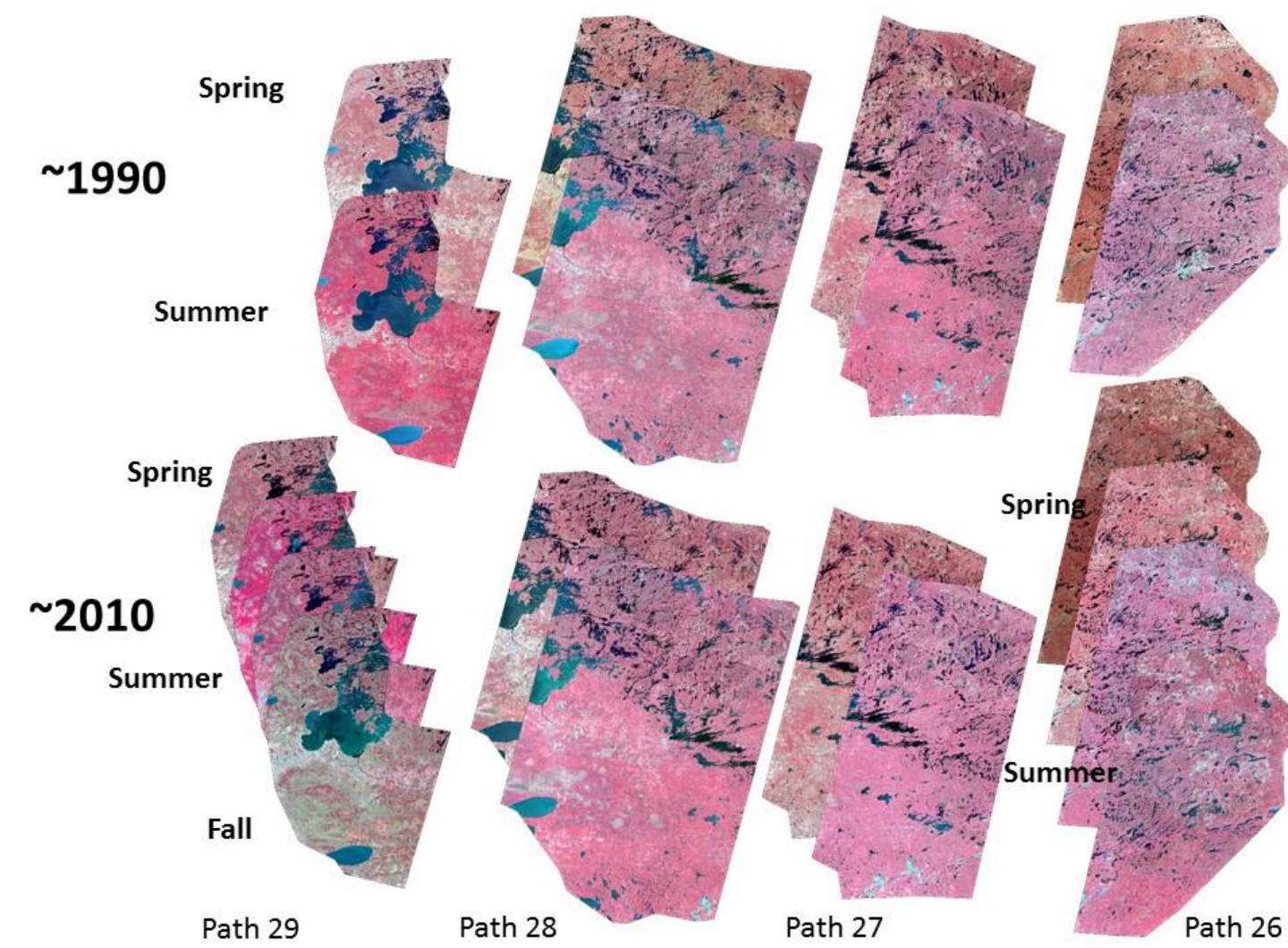
Methods

This project used a combination of multitemporal Landsat imagery, lidar data (Minnesota) and object-based image analysis to cover the entire extend of the Lake of the Woods/Rainy River Basin using a uniform method for the ~1990 and ~2010 time periods so that land cover and changes over that time period can be quantified and used for hydrologic modeling.

To complete the land cover classification for 1990 and 2010 time periods for the entire 70,000 km² Lake of the Woods/Rainy River Basin 44 Landsat images were utilized. These images included Landsat Thematic Mapper (TM) and Operational Land Imager (OLI) images from different seasons and vegetation development stages to distinguish different kinds of vegetation and other cover.
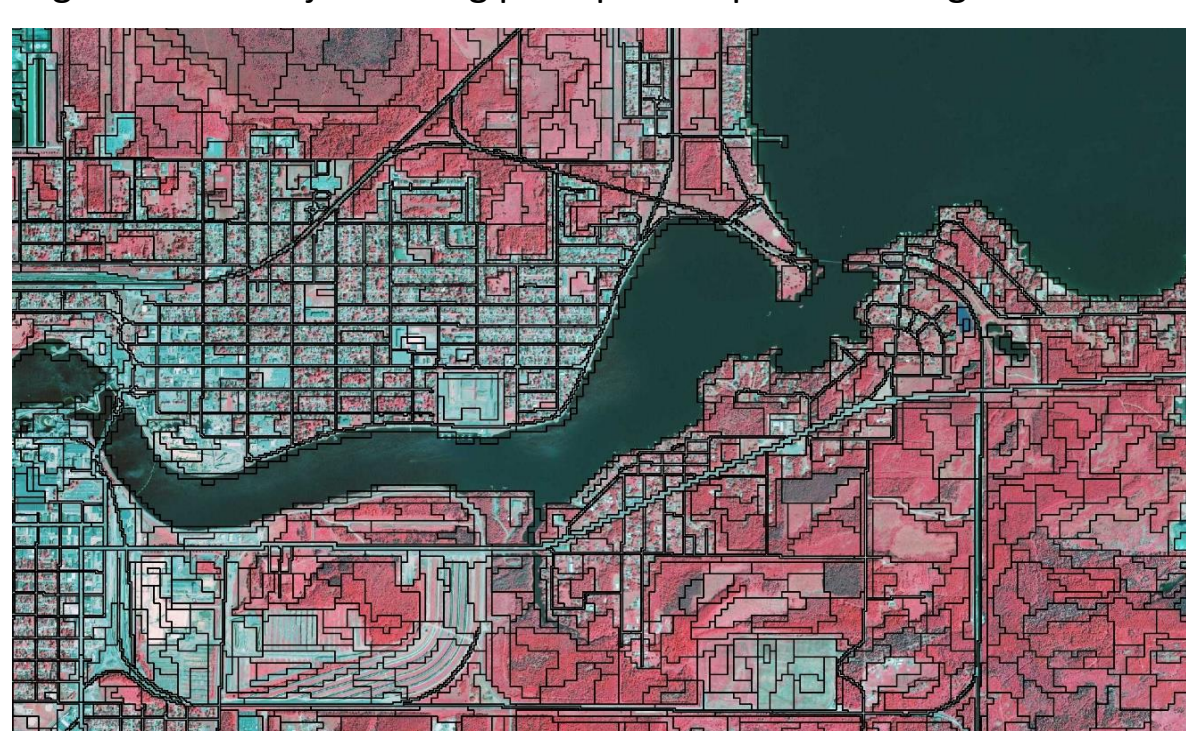
With multiple dates of imagery per Landsat path, each with seven to ten spectral bands it is useful to compress the images using the principal components. Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Additional transformations of the Landsat data used were the "tasseled cap" greenness, brightness and wetness and the normalized difference vegetation index (NDVI).

Lidar data acquired provides additional information on height and elevation. Lidar LAS files were acquired from the Minnesota DNR for the tiles within the areas covering the Minnesota portion of the Lake of the Woods/Rainy River Basin. The LAS tiles were used to generate mean and maximum vegetation height rasters at 20-meter spatial resolution. The DNR-provided 1-meter bare earth DEM was also used to create additional lidar-derivative layers at 10-meter spatial resolution, such as Compound Topographic Index (CTI), slope, and dissection. These DEM derived variables are especially useful for wetland identification. More information about the lidar data are available at: http://www.mngeo.state.mn.us/chouse/elevation/Lidar.html.

Additional GIS layers used for image segmentation included major roads, other roads and railroads were acquired from the Minnesota Department of Transportation, Ontario Ministry of Natural Resources and Forestry and Manitoba Infrastructure and Transportation. Once the classification was finalized and accuracies were calculated these layers were also overlaid on the maps so that roads narrower than the 30 m Landsat data could be included in the 2010 map. For the 1990 map an edited version of the roads layer was used which excluded roads that were identified as being built since 1990.

Segmentation objects using principal components image and roads.

Since we wanted to utilize the same objects for 1990 and 2010 we determined that using the concentrated spectral data in the PCA image for each Landsat path along with the roads and railway layer produced the best segmentations. The objects created using eCognition were suitable for both time periods and could be used to classify change between 1990 and 2010.

Extraction of Object Features: Once image objects were created, a large number (>200) of features could be derived and potentially used for classification. The primary features included: spectral-radiometric data, including means, modes, quantiles and standard deviations of individual bands and several transformations of the Landsat and lidar data; geometry, including asymmetry, compactness, density, rectangular fit, roundness, shape and texture.

Classification: To decrease the error between the 1990 and 2010 classification we classified both time periods together and added classes at Level 3 to identify change. These included areas that were developed, changes in agriculture, barrier and shore land areas that eroded away or were created, and forested areas that were harvested or regenerated. Since we had four Landsat paths of imagery each had to be classified separately since images from different dates would have different vegetation phenology which would increase classification error.
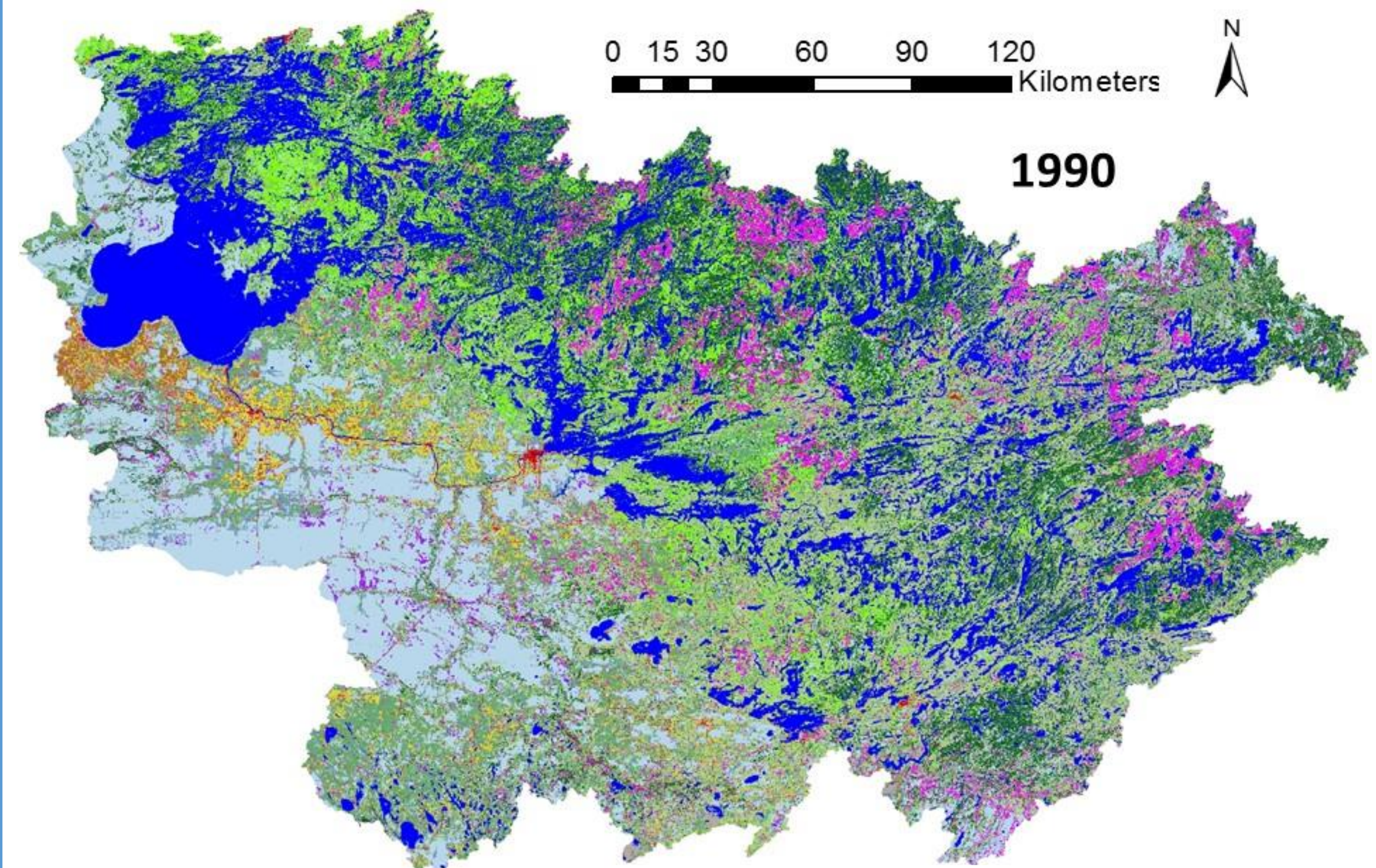

Objects over 2008 NAIP false color imagery for the same area.

To take advantage of the lidar data that were available in Minnesota, but not Ontario or Manitoba, the Minnesota sections of each Landsat path were classified separately. Therefore to cover the entire Lake of the Woods/Rainy River Basin using the best available data sources eight separate classifications were completed and mosaicked to create the final land cover maps and change maps.

Random forest, a state-of-the-art approach which could handle and take advantage of the large number of features, was used for the classification of the objects. It is an ensemble learning method for classification that operates by constructing multiple decision trees. Each tree is grown on different random subsamples of the training data and during the split selection process by using a subsample of the available features. It allows for the use of a large number of features or variables and identifies the important predictors.
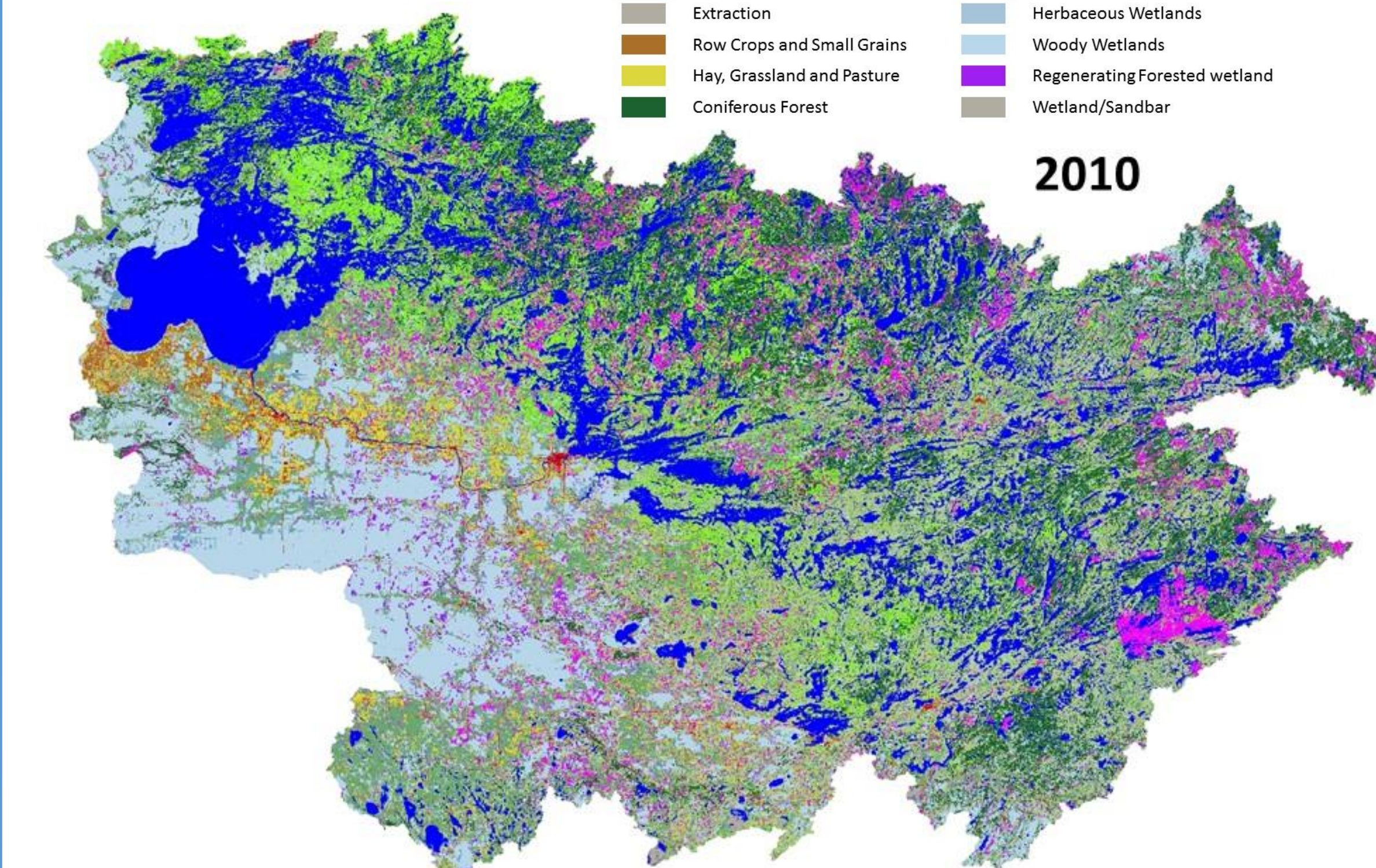
Acknowledgments

~1990
Spring / Summer
Spring / Summer / Fall
Path 29   Path 28   Path 27   Path 26

~2010
Spring / Summer
Spring / Summer / Fall

PCA 1   PCA 2
PCA 3

Lidar DEM

Slope   Compound Topographic Index (CTI)   Dissection

Vegetation Height
Point Cloud   Maximum   Mean

1990

Land Cover Maps

Legend

Developed High density	Deciduous Forest
Developed Medium density	Mixed Forest
Developed Low density	Sparse Forest
Developed Managed Grass	Regenerating Forest
Developed Roads	Lakes and Ponds
Extraction	Herbaceous Wetlands
Row Crops and Small Grains	Woody Wetlands
Hay, Grassland and Pasture	Regenerating Forested wetland
Coniferous Forest	Wetland/Sandbar

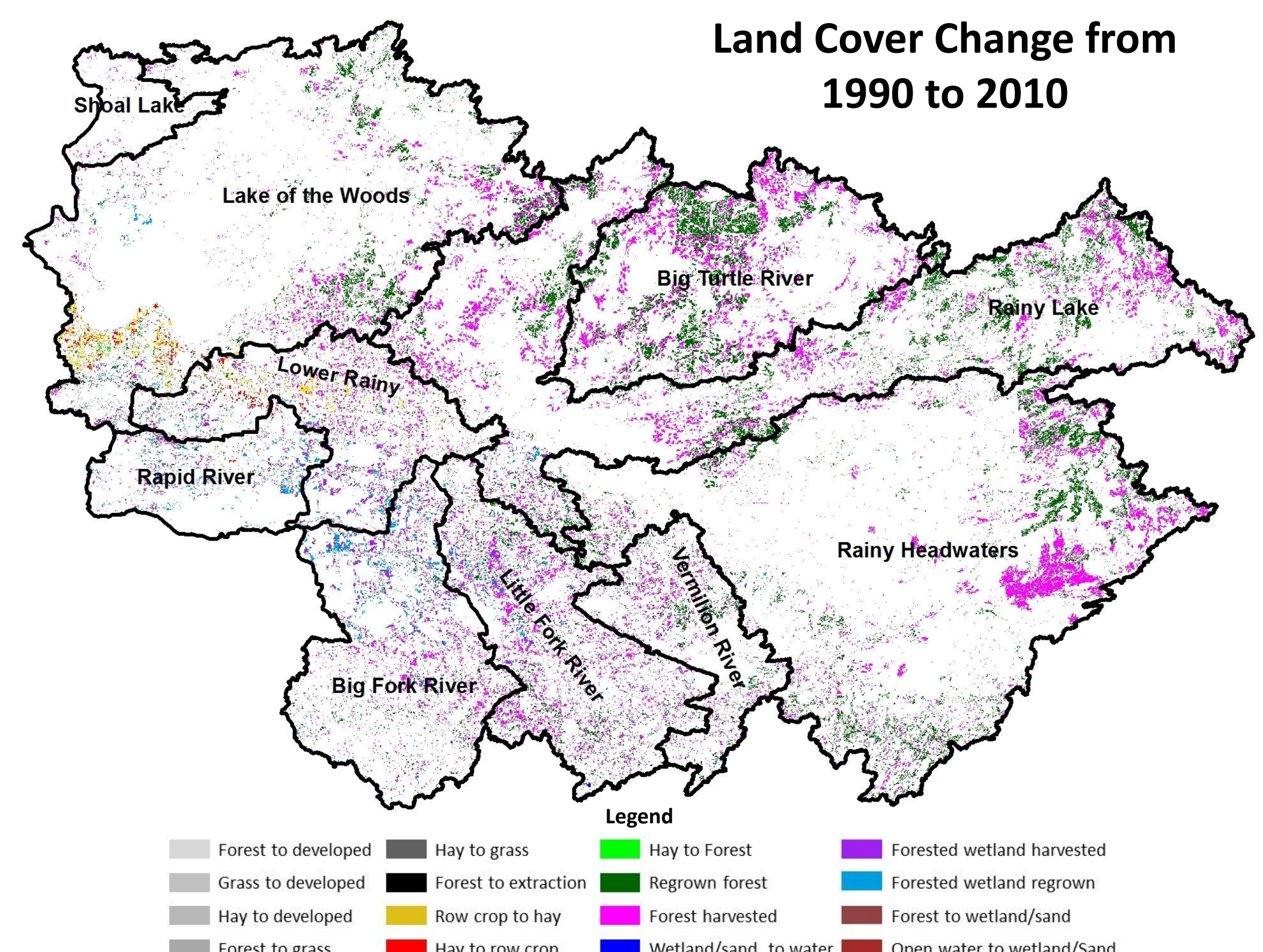### 2010

With the finalized maps we tabulated the area and calculated percent of area by class at the sub-basin and total Lake of the Woods/Rainy River Basin for 1990 and 2010. The Lake of the Woods/Rainy River Basin is dominated by forest, wetlands and lakes which comprises 96.3 percent of the basin. Developed areas increased in the basin from 1.30 percent in 1990 to 1.33 percent in 2010. The most obvious changes are due to forest harvesting which is most apparent on the land cover maps or the land change map. There were no changes from 1990 to 2010 for approximately 88% of the basin. While forests that have been harvested move around between the time periods there was also an increase to 5.6% of the basin in 2010 from 4.6% in 1990 that was detected.

Land Cover Change from 1990 to 2010

Shoal Lake / Lake of the Woods / Big Turtle River / Rainy Lake / Lower Rainy / Rapid River / Little Fork River / Vermillion River / Rainy Headwaters / Big Fork River

Legend

Forest to developed	Hay to grass	Hay to Forest	Forested wetland harvested
Grass to developed	Forest to extraction	Regrown forest	Forested wetland regrown
Hay to developed	Row crop to hay	Forest harvested	Forest to wetland/sand
Forest to grass	Hay to row crop	Wetland/sand to water	Open water to wetland/Sand

Accuracy assessment: We evaluated classification accuracy by comparing the classification results to an independent stratified (by class) random reference sample of 6,610 objects (20 percent of the reference data that were withheld from classifier training). The data set has an overall average classification accuracy of 96.4% for level 1 and 92.2% for level 2 land cover classifications. The use of lidar improved the accuracy of the classification; for example in 2010 the overall accuracy for level 1 with lidar data was 98.0% vs. 96.1% without lidar.

Conclusion: A combination of multitemporal Landsat data and lidar data for the Minnesota portion along with object-based image analysis and the random forest classifier enabled accurate level 1 and 2 land cover classifications for the Lake of the Woods/Rainy River Basin for 1990 and 2010. The digital format of the classifications makes it possible to easily include them with other digital maps and data in a GIS for further analysis and modeling. The classification maps and metadata can be viewed or downloaded from a Google Maps based web application at http://portal.gis.umn.edu.

0 15 30 60 90 120 Kilometers