

**A CART Based Sub-pixel Method to Map Spatial and Temporal
Patterns of Prairie Pothole Lakes with Climatic Variability**

Bo Zhang

School of Earth Science
The Ohio State University
125 S. Oval Mall, 275
Columbus, OH 43210,
USA.

(Email: zhang.423@osu.edu)

Abstract:

The Prairie Pothole Region in the United States contains millions of seasonal, semi-permanent, or permanent lakes and wetlands that typically range in size from 0.1 to 10 hectares. These lakes and wetlands are vulnerable to climate change, especially in our study area in South Dakota, in which a period of deluge following a sharp drought considerably expanded the areal extent of prairie pothole lakes during the last decade of the twentieth century. Preliminary estimates of lake areas, determined using Landsat 5 and 7 images, had appreciable errors especially for the smallest of these lakes. This paper describes a new sub-pixel approach integrated with a CART (Classification and Regression Tree) model using GIS (Geographical Information System) to quantify mixed water pixels along lake boundaries to improve area estimations for pothole lakes. Errors in estimated area were typically 10 percent or less for lakes greater than 1 hectare in size. An analysis of lakes in our study area demonstrates how lake area changed with the transition from drought to deluge. Small lakes exhibited a distinct seasonal variation in contrast to large lakes that tended to follow precipitation trends more broadly. The total water area of lakes is consistent with broad variation in rainfall.

1. Introduction:

The Prairie Pothole Region (PPR) of North America has millions of lakes and wetlands. These surface-water bodies exhibit tremendous hydrological variability on annual and interannual time scales because of marked variability in climate. Understanding the regional behavior of these lakes and wetlands is important because their hydrologic and ecologic functions are significant at regional and even continental scales. Area-wise, the PPR constitutes only 10% of the total waterfowl breeding area in North America but produces more than 50% of all breeding ducks and more than 60% of continental mallards (Guntenspergen, et

al., 2000; Batt, et al., 1989). Moreover, wetlands in the PPR serve to reduce the severity of floods (McAllister et al., 2000), to attenuate contaminant loading and transport from fertilizer and pesticide, and to sequester carbon.

Our broad interest with the lakes is in understanding their response to episodes of drought and deluge. The severe drought from 1988-92 in South Dakota was the second worst of 20th century, and led to significant hydrologic impacts. This drought was followed by the most significant wet period of the century (Winter and Rosenberry, 1998). Lake and wetland flooding occurred to an extent not seen perhaps since the turn of 1900.

However, the large extent of the PPR and its geological complexity limit the possibility for detailed hydrological surveys. Fortunately GIS and remote sensing approaches may be useful for studies of wetland hydrology since 1970s (Lunetta and Balogh, 1999; Ozesmi and Bauer1, 2002; Pietroniro and Prowse, 2002; Sawaya et al., 2003; Schmugge et al., 2002; Stewart et al., 1980). Satellite sensors significantly improve the observations of hydrological conditions, by providing long term data records over large areas. Hence remotely sensed data have been broadly applied in hydrological studies of the PPR (e.g. Todhunter and Rundquist, 2004; Work and Gilmer, 1976).

Satellite-based approaches are useful in the study of prairie pothole lakes to estimate spatial changes as a function of climatic variability. Not surprisingly, the archive of Landsat data from 1985 to 2002 is particularly useful. However, the moderate spatial resolution (30m) impacts the assessment of lakes and wetlands, especially when small pothole lakes are represented by just a few pixels. The problems of resolution are of particular concern in the PPR where more than 80% wetlands are smaller than 0.8 ha (Johnson and Higgins, 1997). Those small ponds are manifested by a few mixed water/land pixels and a few pure water pixels. It is apparent that traditional classification methods using moderate Landsat data underestimated the areal extent of water bodies because mixed water/land pixels along

boundaries are not properly included in the classification. For our applications then, approaches are required capable of interpreting the composition of different classes within a single pixel.

The goal of this paper is to describe a new sub-pixel method to estimate the water area of small water bodies integrated with GIS, represented by perhaps several Landsat pixels. We illustrate this approach in an assessment of the hydrologic behavior of lakes and wetlands on the Prairie Coteau of South Dakota.

2. Methodology

2.1 Description of the Study Area

The study area is located on the northern tip of the Prairie Coteau area of northeastern South Dakota. During Wisconsin deglaciation, stagnation of the ice on this upland area produced an extremely hummocky topography and ultimately thousands of close-basin lakes and potholes. The map of the study area (Figure 1) shows the large recreational lakes also found there. The most important of these large lakes are those of the Waubay Lakes chain. Available stage hydrographs for several of these large lakes show the dramatic change in water levels through the drought-deluge transition. No similar data exist for the smaller pothole lakes. Most of the lakes and wetlands occur in closed basins essentially disconnected from river systems. With the exception of interlake transfers, the Waubay Lakes system is also closed hydrologically.

This region has a subhumid to subarid continental climate with short hot summers, long cold winters, low levels of precipitation, and high evaporation. The mean annual precipitation ranges from 50 to 56 cm (20-22 in) and temperature ranges from 1F ° in Jan to 85F ° in Jul. Winter precipitation data are shown in Figure 2 for the WAUBAY NATL WILD LIFE Station. The data summary clearly shows the swing from extreme drought conditions in 1988-92 to

extremely wet conditions, 1995-1998.

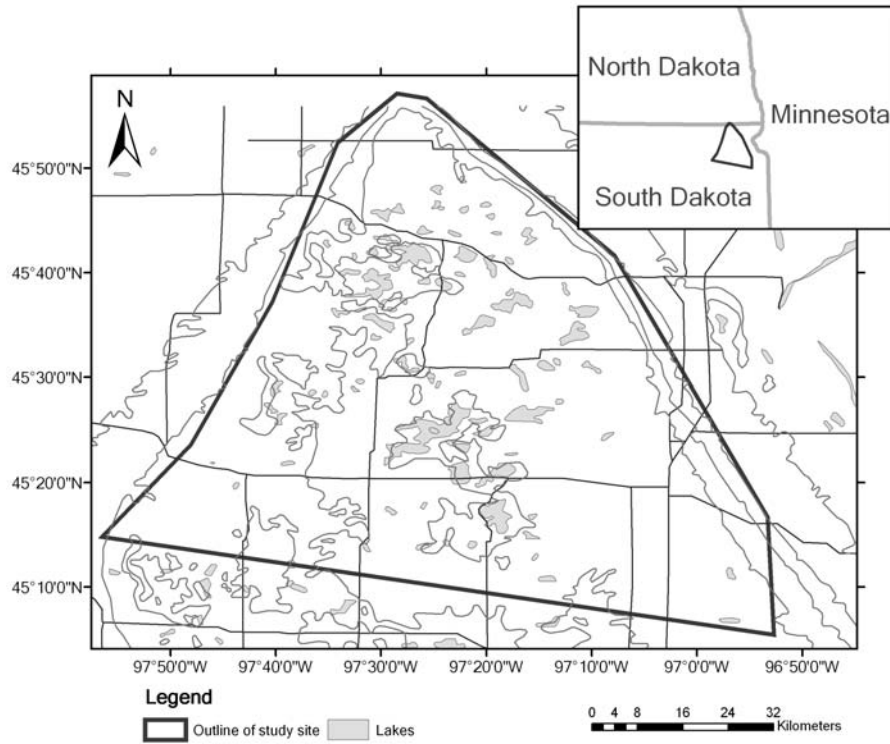


Figure 1 Study area of the Prairie Coteau Region of South Dakota.

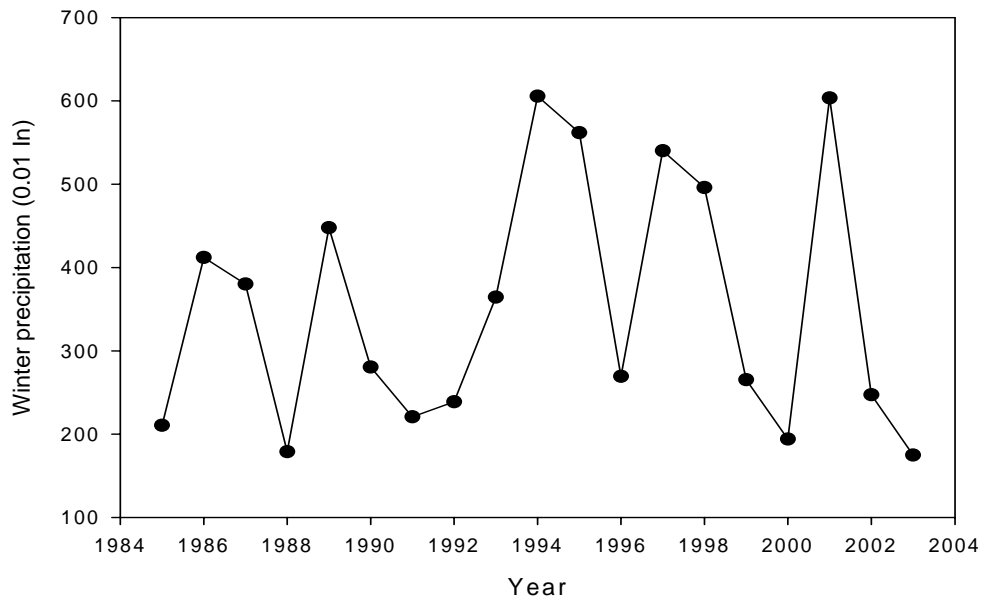


Figure 2 Annual Winter Precipitation within the Study Site.

Note: Winter season include the November and December of the previous year and the January and February of the current year.

2.2 Image Processing and Classification

There is sufficient Landsat TM and Landsat ETM+ coverage of the Prairie Coteau Region to examine the expansion of wetlands and lakes in this region following the drought of 1988-1992. Fourteen Landsat scenes were selected (Table 1) to provide areal information of wetlands and lakes from 1985 to 2003. A Landsat image scene, which was acquired on 7/24/2003 together with a DOQQ (Digital Orthophoto Quarter-Quadrangle) created on 7/04/2003 were used as data samples or learning datasets to construct the CART (Classification and Regression Tree) scheme. Images were calibrated and processed as the first step in an overall assessment. Processing covert Digital Value to absolute Reflectance of each pixel to diminish variability in solar irradiance (Chander and Markham, 2003; Landsat Project Science Office website). Furthermore, general atmospheric condition is assumed for each image based on the meteorological conditions on the dates when images were obtained.

Table 1 List of Landsat Images

Satellite Sensor	Image Date	Image Quality
Landsat TM 5	5/19/1985	9
Landsat TM 5	4/23/1987	9
Landsat TM 5	4/15/1990	9
Landsat TM 5	8/5/1990	9
Landsat TM 5	5/6/1992	9
Landsat TM 5	9/17/1994	9
Landsat TM 5	8/19/1995	9
Landsat ETM+	5/4/1997	9
Landsat ETM+	9/25/2000	9
Landsat ETM+	6/16/2001	9
Landsat ETM+	8/27/2001	9
Landsat ETM+	5/18/2002	9
Landsat ETM+	7/29/2002	9
Landsat ETM+	9/15/2002	9

Note: Value of Image Quality equals to 9 indicating that there was no errors detected, and the image is a perfect scene, as defined by the Landsat 7 Processing System (LPS).

In our study, the first procedure is classification to identify water pixels according to their spectral responses in remotely sensed images, and then, summing all connected water pixels

provides area estimation for a lake. A variety of classification approaches exists such as unsupervised classification and supervised classification (Jensen, 1996). These methods work well for small homogenous areas mainly with uniform ground characteristics. Several approaches have been implemented in delineating water boundaries on Landsat images, for example, density slicing of Band 5 and supervised maximum likelihood classification (Frazier and Page, 2000; Manavalan et al., 1993). These popular hard classifiers however produce poor estimates for small ponds (Frazier and Page, 2000), because they overlooked partial water pixels mixed with other land features along the boundaries of small ponds.

In recent years, methods such as sub-pixel classification and fuzzy classification have improved the classification accuracy in extracting certain ground features. Typically, such approaches are based on complicated models and machine learning technologies, including linear or nonlinear spectral mixture models, neural networks and decision trees etc. For example, neural network and decision trees have been applied to forest, vegetation, and urban area classification with somewhat improved accuracies as compared to traditional methods (e.g. Atkinson and Tate, 2000; Brown de Colstouna et al., 2003; DeFries and Chan, 2000; Pal and Mather, 2003; Tatem, et al., 2002). In particular, decision tree approaches have the capability of increasing classification accuracy significantly without invoking a prerequisite Gaussian assumption by maximum likelihood method. Decision tree approaches are also easier to interpret, as compared to black-box model approaches like neural networks. Reported comparison among different classification methods including decision tree, maximum likelihood and fuzzy c-means clustering indicates that maximum likelihood produces the highest accuracy for pure pixels, while decision tree methods provide the greatest accuracy for mixed pixels (Xu et al., 2005).

In our study, two classification steps were implemented for pure water pixels and mixed water pixels respectively. First, each calibrated image was classified using an unsupervised

ISODATA (Iterative Self-Organizing Data Analysis Technique) clustering algorithm (Tou and Gonzalez, 1974) to provide seven classes – Pure water, Wetlands, Agricultural land, Range land, Bare land, Forest and Build-up land according to existing investigations of land cover and the scheme recommended by USGS. The ISODATA algorithm was chosen because of its simplicity and robustness as well as its ease of incorporation into other software packages. Water pixels classified were treated as pure water pixels and were preserved from other pixels by a convolution process. Then mixed water pixels were identified as pixels adjacent to pure water pixels, assuming that lake boundaries are located within 1 pixel of pure water pixels, based on the common shapes of pothole lakes within the study area. Finally, a CART (Classification and Regression Tree) based method to quantify the extent to which water is represented in mixed pixels along the boundary of each lake is adopted. The method we developed here incorporated both pure water pixels and mixed water pixels in area estimation is called Sub-pixel method in contrast to the unsupervised method counting only pure water pixels.

2.3 Water Fraction Classification by CART

To construct the CART scheme, a total of 488 pixels were randomly selected (299 pixels for data training and 189 pixels for validating) from the sample Landsat image using ArcGIS software. The DOQQ within a 20 days time window was used as the ground truth reference. In effect, for the lake-boundary pixels, the training data link the Landsat spectral data for each pixel to an exact estimate of how much water is present in each pixel, as indicated from the DOQQ. The validating data are represented as a set of new pixels used to verify and refine the CART model by comparison with independent estimates from the DOQQ.

One of the considerations in using a tree-based approach is to determine how many nodes should be included in the tree. A training data set was classified to determine the

relative error associated with different numbers of nodes. After splitting the training dataset in various trials, a tree was selected with the smallest relative error (the ratio of the variance from the fitted tree model to the original sample variance) of 0.45 and only eight terminal nodes (Figure 3). The eight terminal nodes indicate that eight classes of water fraction values assigned to the training dataset are appropriate. In addition, the small size of the tree also avoids potential overfitting, which in larger trees is usually minimized by a pruning process (Pal and Mather, 2003).

Figure 3 shows the structure of the CART model for mixed water pixel classification. In Figure 3, the 299 mixed pixels in the training dataset are firstly separated by their reflectance values in band 4. Pixels with a reflectance less than or equal to 0.2378 are assigned to node 2 and others go to node 3. Then each node is continuously split until the terminal nodes are reached. The terminal nodes provide the final classes for each of the pixels.

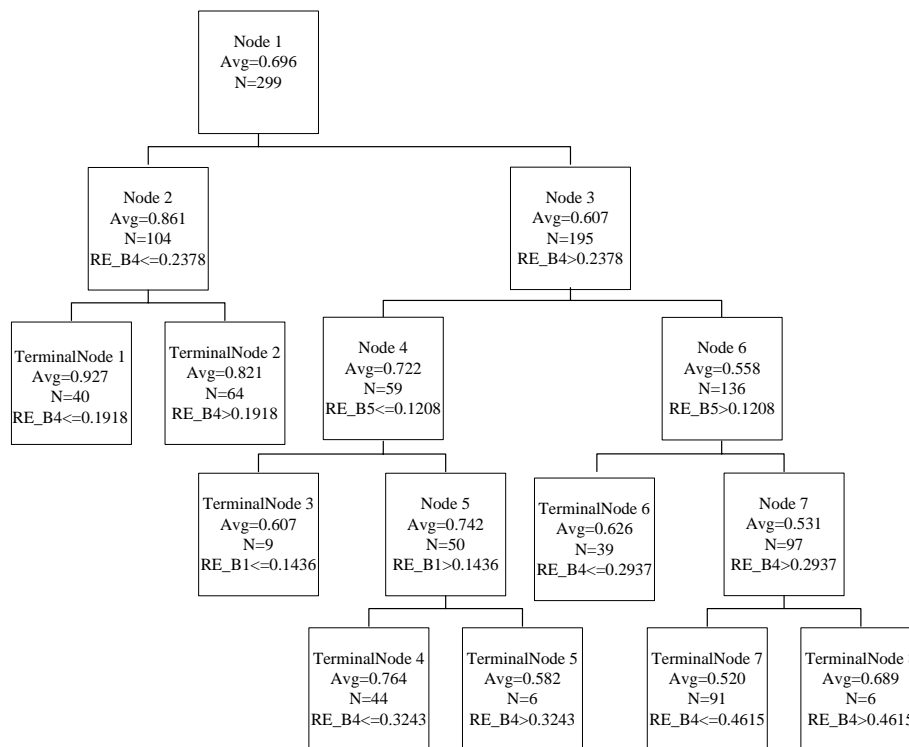


Figure 3 Tree Structure of CART Model.

Note: Avg - the average water fraction of all the pixels within this node; N - the total number of pixels assigned to this node; RE_B4 - the reflectance of band 4.

2.4 Validation of CART model

The splitting procedures for constructing the CART model also include a feature selection process, which provides measures of relative importance (Table 2). Relative importance is calculated as G^2 statistics and quantifies the weights of the attributes (bands of Landsat images) of the data in the classification. Table 2 shows that Bands 4, 5, 7, 1 are sequentially the most important in classifying mixed water pixels with relative importance values 100.0, 72.3, 59.5, and 48.1, respectively. This result further indicates that bands 4, 5 and 1 are sufficient to classify the training data set. Because band 7 is highly correlated with bands 4 and 5, band 7 was not included in the model. Thus, the CART model involves bands 4, 5 and 1. The relative importance values are in the same order as an earlier study (Xu, et al., 2005) showing that band 4 is the most important in classifying mixed pixels. It also follows then that band 4 is most useful to delineate water boundaries in land cover classification (Jensen, 1996).

Table 2 Relative Importance Values of CART Model

Band	Relative Importance
Band 4	100.000
Band 5	72.308
Band 7	59.478
Band 1	48.149
Band 2	14.602
Band 3	3.112

The performance of this CART model is evaluated by applying the tree to the validating dataset. Pixels in the validating dataset are independently sampled from the training dataset. The goodness of fit of the model to the testing dataset validates the CART model in the classification and implies no severe overfitting problems needed to be concerned. Table 3 presents the classification results for both training and validating datasets. For each terminal node in the table, the mean water fraction values between the training data and the validating data are very close, with the largest difference of 0.138 in terminal node 5. The standard

errors are relatively small and stable, ranging from 0.144 to 0.178. The standard deviations of both training and validating datasets are reasonably constant. Table 3 shows that there is a good fit with the validating dataset and validates the tree model in our mixed water pixel classification. However, the errors evidently grow as the mean water fraction values in the terminal nodes decrease. In other words, smaller quantities of water within a pixel are associated with greater uncertainties in the estimate. Therefore, pixels with water fraction values less than 0.5 are not considered in following analysis to avoid additional errors.

Table 3 Results of Splitting Training and Validating Dataset by CART Model

Terminal Node	Dataset	Count	Mean	StdDev	StdError
1	Training	40	0.927	0.074	
	Validating	24	0.861	0.135	0.150
2	Training	64	0.821	0.135	
	Validating	56	0.832	0.140	0.141
3	Training	9	0.607	0.117	
	Validating	13	0.555	0.162	0.170
4	Training	44	0.764	0.140	
	Validating	19	0.711	0.154	0.163
5	Training	6	0.582	0.090	
	Validating	3	0.444	0.113	0.178
6	Training	39	0.626	0.170	
	Validating	17	0.536	0.142	0.168
7	Training	91	0.520	0.146	
	Validating	57	0.434	0.127	0.154
8	Training	6	0.689	0.218	
	Validating	0	*	*	*

Note: * no data.

Notice also that no validating data were assigned to terminal node 8. Actually, terminal node 8 is abnormal compared with terminal node 7. Terminal node 8 has a higher average value of water content and a greater reflectance in band 4 than those in terminal node 7. This result contradicts the general principle that the lower reflectance in band 4 implies the higher water fraction. This unusual result is caused by specific errors in training dataset sampling.

The standard error associated with terminal node 2 has the smallest value of 0.141 with the second largest number of pixels of 56 in the validating data set. Terminal node 5 has the largest standard error with the smallest number of pixels of 3. This result implies that large sample size of pixels could reduce the standard error and increasing the sample size would improve the classification model.

According to these analyses, we improved the CART model (shown in Figure 4) to refine scheme and to correct potential errors and applied it to all images. We redefined water fraction of all mixed pixels into 6 classes (less than 0.5, 0.5, 0.6, 0.75, 0.82, and 0.90) related to the eight terminal nodes of the original CART model. The nodes of the previous CART model having similar mean values are regrouped into one class of water fraction. Moreover, pixels with a water fraction value smaller than 0.5 are ignored to reduce errors and uncertainties in classification. To fix the issue of terminal node 8 in the previous CART model, terminal node 7 and terminal node 8 are merged as one class with the same attributes as their parent node- the internal node 7. These modifications simplify the structure of the classifier and make it more reasonable and robust than the original CART model.

The rules of the modified CART model in Figure 4 can be illustrated by an example of water fraction class of 0.6: if a pixel has a reflectance value of band 4 larger than 0.2378, a reflectance value of band 5 less than or equal to 0.1208, and a reflectance value of band 1 less than or equal to 0.1436, this pixel will be assigned to the class with a water fraction value equals to 0.6.

Figure 5 shows a sample of images classified by water fraction values. It is clear that most of areas of lakes and wetlands are covered if both pure and mixed water pixels are counted in. Otherwise, severe biases would be introduced with only pure water pixels.

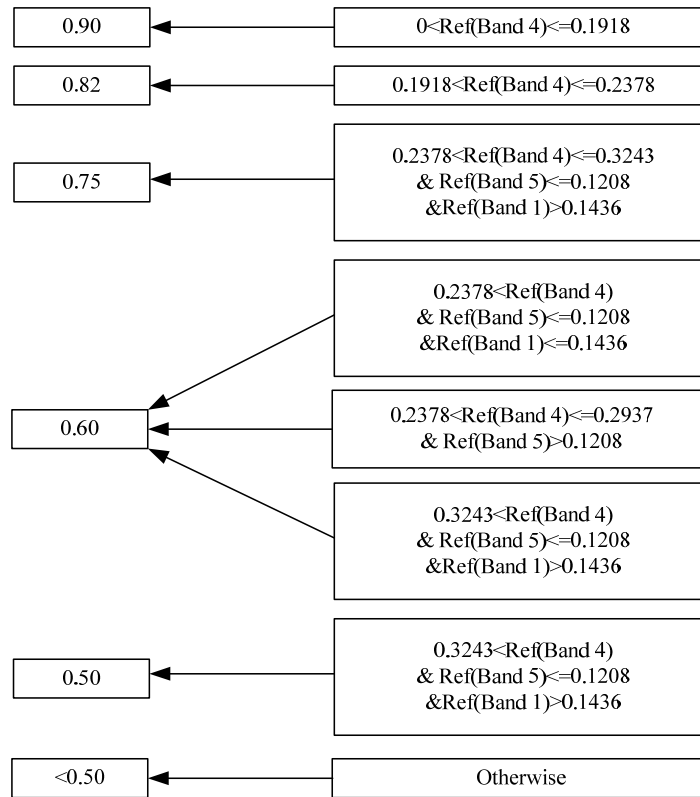


Figure 4 Rules of Sub-Pixel Classifier for Classifying Mixed Water Pixels.

Note: Ref (*) represents the reflectance value of the band.

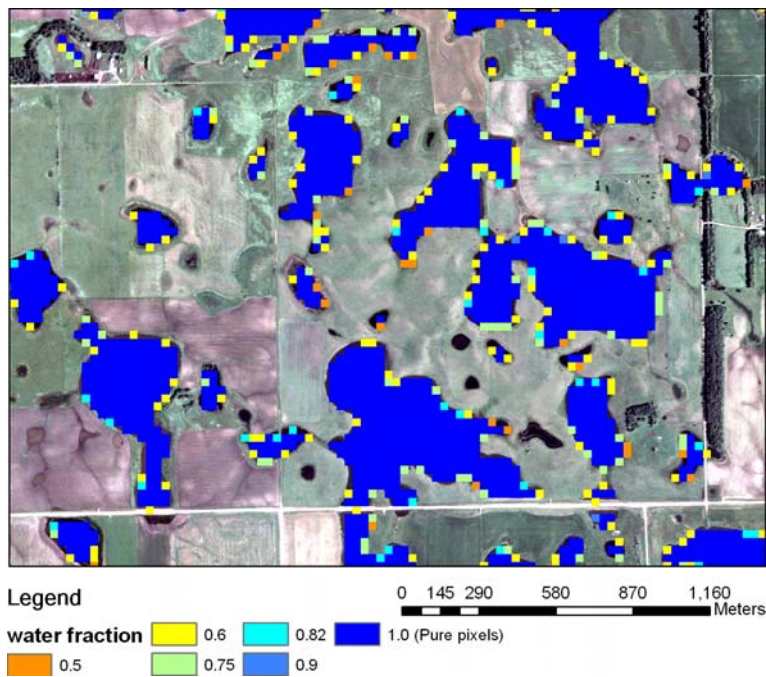


Figure 5 Sample Image Classified by Water Fraction Values.

3. Results and Discussion

3.1 Assessment of Unsupervised Classification

Because a land-cover classification is integral to our approach, an assessment is needed of the accuracy of the ISODATA algorithm in unsupervised classification. Accuracy in the assignment of each class was determined by evaluating the classification of 71 randomly created control points. Table 4 shows that the simple unsupervised classification method is successful in identifying pure water pixels with approximate 100% accuracy. Among all the other land-cover classes, Forest has the lowest accuracy because forests are sparsely scattered within this region in relatively small patches, which are commonly mixed together within other classes such as built-up land, range land and wetlands. Range land and agricultural land are both major classes that are also difficult to differentiate from each other in Landsat images. Because the water class is of primary concern, the spotty performance in other classification is not a problem to our study.

Table 4 Accuracy of Unsupervised Classification by ISODATA Algorithm

Class Name	Reference Totals	Classified Totals	Correct Number	Producers Accuracy	Users Accuracy
Agricultural land	10	22	9	90.0%	40.9%
Bare land	7	9	5	71.4%	55.6%
Build-up land	1	1	1	100.0%	100.0%
Forest	3	5	1	33.3%	20.0%
Range land	31	17	14	45.2%	82.4%
Water	13	13	13	100.0%	100.0%
Wetland	6	4	3	50.0%	75.0%
Total	71	71	46	64.8%	64.8%

3.2 Accuracy of area estimation

This section presents an analysis of the improvement in lake-area estimates from Landsat images provided by the sub-pixel method outlined above. The basis for comparison is a

conventional unsupervised classification (ISODATA) approach that is capable of identifying pure water pixels. This analysis involves 42 lakes of different sizes that are randomly selected. For each lake, a highly accurate measurement of area is determined from the high-resolution DOQQ, which provides the ground truth reference. For each of the test lakes selected from the study area in South Dakota, areas are estimated using (1) the Landsat image with only the unsupervised classification method and (2) the classification scheme improved by the sub-pixel approach. Relative errors of area estimation in the two different Landsat-based methods are calculated as ratios by the following equation, for each lake, relative errors provide a measure of accuracy of the methods.

$$\text{Relative error} = \left(\frac{\text{Area estimated by Landsat} - \text{Area estimated by DOQQ}}{\text{Area estimated by DOQQ}} \right) \quad (1)$$

The results of this analysis are shown in Figure 6 (a), where the relative errors are plotted versus actual lake areas estimated by DOQQ. As expected, errors of estimation increase with a decrease in the lake areas. However, area estimates based on the sub-pixel method have remarkably smaller errors as compared to the unsupervised classification. Indeed, the sub-pixel approach works well for quite small lakes. For example, even lakes about 0.5 ha (or about 5 Landsat pixels) can be estimated with errors of about 20% or less. The estimates of area for lakes larger than 1 ha (or about 11 Landsat pixels) are less than 10%.

Another advantage of sub-pixel method is that the errors are generally unbiased with errors almost symmetrically distributed around 0, as shown by the histogram plots in Figure 6 (b). The relative errors from the sub-pixel method have an approximately normal distribution with a mean of -0.04 and a standard deviation of 0.12, while the distribution of the relative errors from the unsupervised method has a mean of -0.31 and a standard deviation of 0.23. This result indicates an unbiased estimation of lake areas from sub-pixel method as compared to the strongly skewed distribution from the unsupervised method. In addition, the standard

deviation for the new method is only about half of that from the unsupervised method, indicating that the method is robust even when lake sizes are quite small.

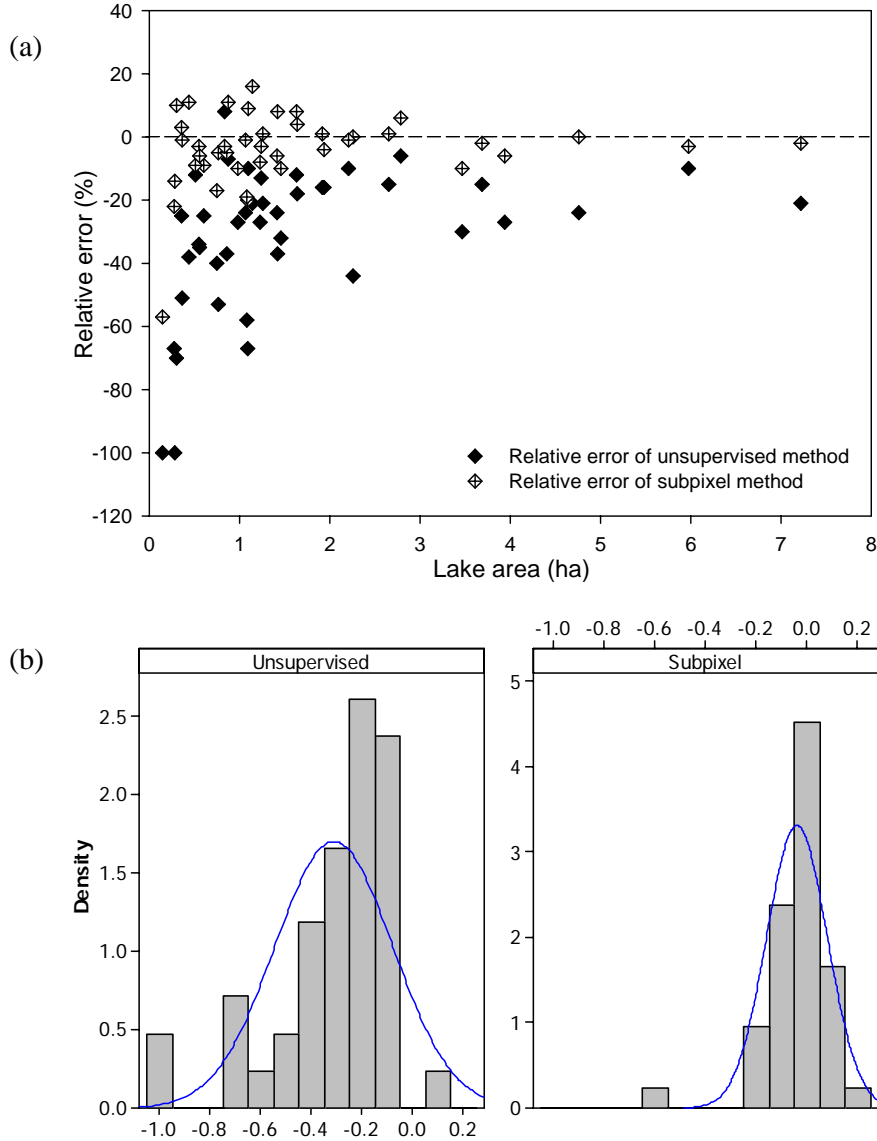


Figure 6 (a) Relative Errors of Lake Area Estimation vs. Actual Areas. (b) Histograms of Relative Errors of Lake Area Estimation from Landsat Images by the Two Methods with Fitted Curves.

The skewness of underestimation from the unsupervised method is produced by mixture of water and sub-emergent plants or grass within the pixels along the boundaries of prairie lakes and wetlands. Because vegetation has a considerably high reflectance in Band 4 and

Band 5 of Landsat images, even a small ratio of vegetation would create a high DN in these bands, which cause the mixed water pixels misclassified into other categories.

Some general errors of both methods remain, which can be attributed one or several of the following possibilities. There exists a 20-day time difference between when the DOQQ and the Landsat image were collected. Small lakes, in particular, might exhibit some changes over this period. A second possible source of error might be related to errors in the sub-pixel classification because of the limited sampling size of pixels and uncertainties in the sampling process. The third possibility is the registration of the DOQQ and Landsat images. A slight geometric distortion could also create small errors in estimates of water fraction. On balance, the Landsat images can provide much better area estimation if the sub-pixel approach is adopted.

3.3 Spatial and Temporal Patterns of Prairie Pothole lakes in South Dakota

The purpose of this section is to demonstrate how this sub-pixel approach can be used to characterize the change in the area of lakes in a study area in South Dakota. The period (1985-2002) covered by the Landsat imagery is interesting in that it captures a swing from extreme drought to deluge.

Total lake area, as a function of time, is determined through this interacting period of eliminate variability to demonstrate a potential application of the sub-pixel approach. For the purposes of this analysis, lakes and wetlands are classified according to their sizes as follows: small lakes having areas less than 1 ha (12 pixels), large lakes having areas greater than 9 ha (100 pixels), and medium lakes whose sizes are between small and large lakes. The parameters are estimated for each of the groups as well as the overall and are presented in Figure 7.

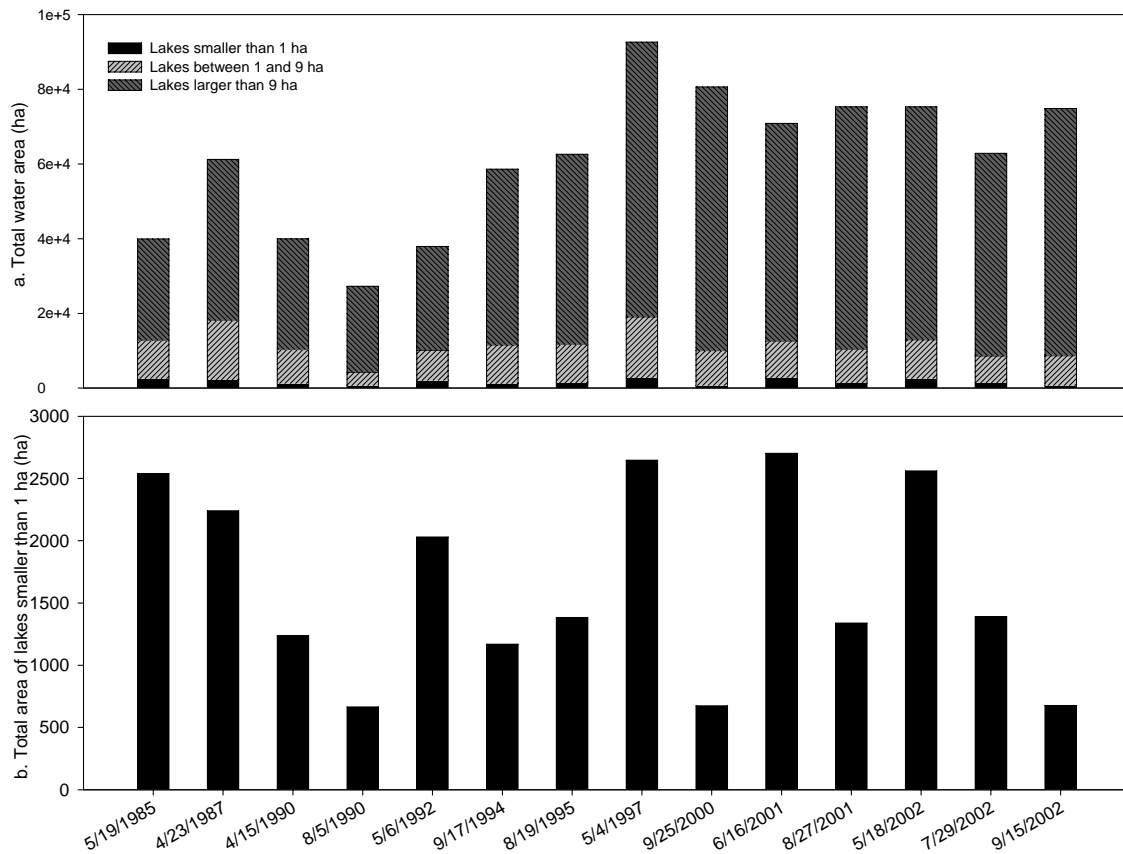


Figure 7 Total Areas of Lakes in Different Sizes (Small, Medium, and Large) vs. Dates

Figure 7 is a summary diagram showing the interannual variation of total lake area. Panel (a) of Figure 7 shows a major decline in total water area from 1987 to the low in 1990, where minimum total area is 27237.7 ha on 8/5/1990. This decline was followed by a steeply increasing trend from 1992 to 1997 where the total water area expanded to a maximum 92625.4 ha on 5/4/1997. In effect, the total lake area more than tripled during this five years period. These large changes in area are clearly related to the contemporaneous climatic variation from drought to deluge. The plot indicates that the year 1985 and the years from 1990 to 1992 were recorded as the severe droughts in contrast to the extreme by wet years from 1997 to 2000. Panel (a) of Figure 7 also shows that more than 80% of the overall area is provided by large lakes (larger than 9 ha). Hence it was variations in the areas of these lakes that impacted the overall pattern. Because the contributions of small lakes (less than 1 ha)

were small and difficult to recognize in Panel (a), the time variation of the area of small lakes are displayed in Panel (b) (Figure 7). Clearly, these small lakes behave differently in that there are three minima represented by a total area of around 670 ha happened on 8/5/1990, 9/25/2000 and 9/15/2002. The largest total areas of about 2600 ha for small lakes were observed on 5/19/1985, 5/4/1997 and 6/16/2001. This result implies that the areas of the small lakes were controlled by seasonal climate patterns.

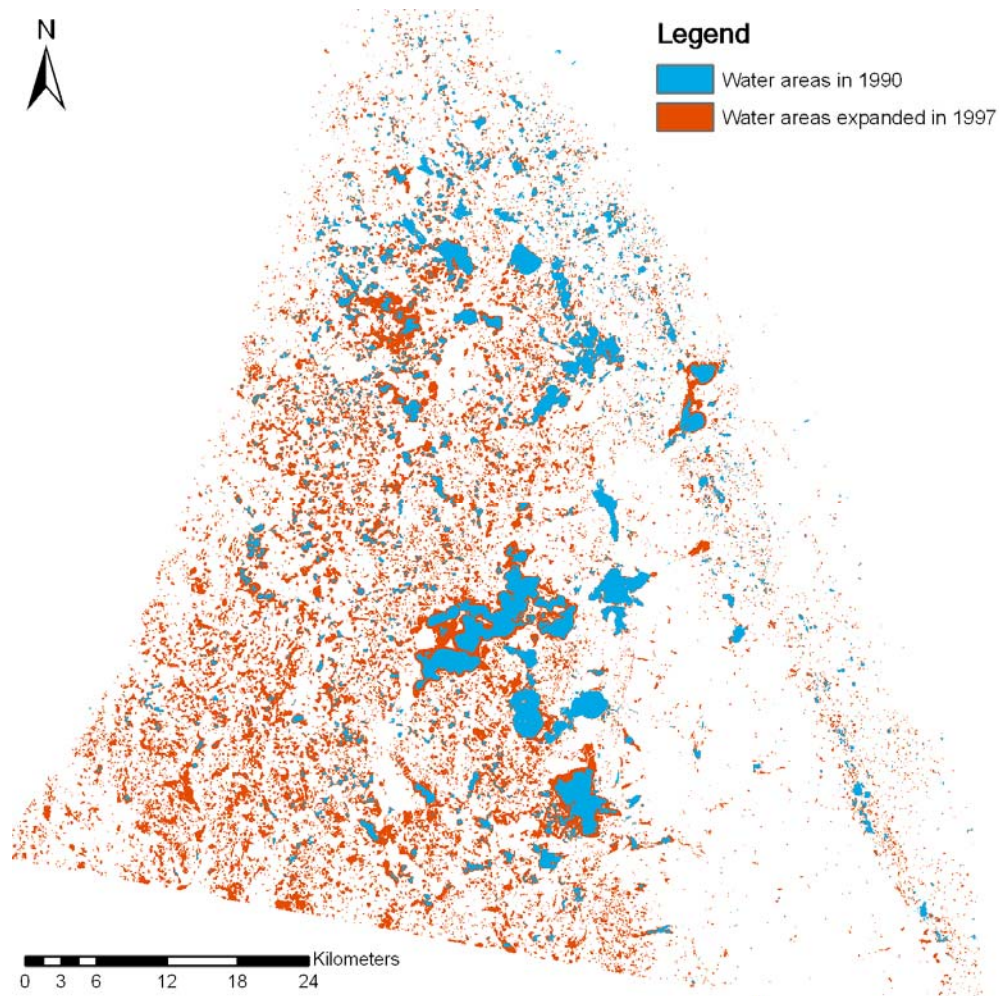


Figure 8 Comparison of Lake Occurrences between 1990 (dry year) and 1997 (wet year)

Figure 8 shows a regional comparison of water areas for two extreme hydrological situations. In contrast to the effects of severe drought on 8/5/1990, both the number and the

areas of lakes expanded significantly in response to much wetter conditions, for example, on 5/4/1997. Some big lakes like the lakes in Waubay Lake chain coalesced. Moreover, the majority of new emergent lakes were concentrated within the southwestern region in contrast to the relatively much smaller changes along eastern edge and northeastern portion of the region. A more detailed analysis of DEMs (Digital Elevation Models) indicates that the elevation of southwestern region is much lower with much less steeper local relief than other areas. This setting facilitated the expansion of water area when deluge occurred.

4. Conclusion:

A sub-pixel based scheme for analyzing Landsat data showed good promise in substantially improving estimates of lake areas, especially for lakes as small as within 1 ha. This improvement in accuracy comes about by including mixed water pixels along lake boundaries.

The usefulness of this approach was demonstrated for a study area in South Dakota, which was substantially impacted by the combination of drought and deluge. Changes in total water area from 1985 – 2002 using Landsat images showed systematic changes controlled by climate. The total water area is essentially tracking the behavior of large lakes (>9ha), which appears to be tracking the broad climatic behavior (drought to deluge). The smallest lakes appear to be controlled by seasonal variability in climate. The emergence of a large number of lakes and wetlands in spring implies the importance of the snow-melt at this time of the year. Areas of small lakes and wetlands appear to be dramatically reduced in late autumn due to the evaporative losses through summer. More work is presently under ways using this sub-pixel approach to examine the broad statistical structure of the surface water resource.

Acknowledgements:

The authors would like to thank the Ohio View and the South Dakota View for providing Landsat data. DOQQ datasets have been provided by the USGS through the EROS data center. The project is supported by the Nation Science Foundation, NSF Award EAR-0440007.

Reference:

- Atkinson, P. M. and N. J. Tate, 1999. *Advances in Remote Sensing and GIS analysis*, John Wiley & Sons, England, 288 p.
- Batt, B.D.J, M. G. Anderson, C. D. Anderson, and F. D. Caswell, 1989. The use of prairie potholes by north American ducks, *Northern Prairie Wetlands*, (A. van der Walk, editor) Iowa State Univ. Press, AmesIowa, pp. 204-227.
- Brown de Colstoun, E. C., M. H. Story, C. Thompson, K. Commisso, T. G. Smith, and J. R. Irons, 2003. National park vegetation mapping using multitemporal Landsat 7 data and a decision tree classifier, *Remote Sensing of Environment*, 85 (3): 316-327.
- Chander, G. and B. Markham, 2003. Revised Landsat-5 TM radiometric calibration procedures and postcalibration dynamic ranges, *IEEE Transactions on Geoscience and Remote Sensing*, 41 (11): 2674-2677.
- DeFries, R. S. and J. C.-W. Chan, 2000. Multiple criteria for evaluating machine learning algorithms for land cover classification from satellite data, *Remote Sensing of Environment*, 74 (3): 503-515.
- Frazier, P. S. and K. J. Page, 2000. Water body detection and delineation with Landsat TM data. *Photogrammetric Engineering and Remote Sensing*, 66 (12): 1461-1468.
- Guntenspergen, G. R., S. A. Peterson, S. G. Leibowitz, and L. M. Cowardin, 2002. Indicators of wetland condition for the Prairie Pothole Region of the United States, *Environmental Monitoring and Assessment*, 78 (3): 229-252.
- Johnson, R. R., K. F. Higgins, M. L. Kjellsen, and C. R. Elliott., 1997. Eastern South Dakota wetlands, Northern Prairie Wildlife Research Center Online, URL: <http://www.npwrc.usgs.gov/resource/wetlands/eastwet/index.htm> Brookings: South Dakota State University, Jamestown, ND (Version 02 July 1998).
- Jensen, J. R., 1996. *Introductory Digital Image Processing, a Remote Sensing Perspective*,

second ed. Prentice Hall, Upper Saddle River, N.J., 316 p.

Landsat Project Science Office at NASA's Goddard Space Flight Center Online, URL:

<http://ftpwww.gsfc.nasa.gov/IAS/handbook/handbook.htmls/chapter11/chapter11.html>,

Greenbelt, Maryland (last date accessed: 15 May 2006).

Lunetta, R. S. and M. E. Balogh, 1999. Application of multi-temporal Landsat 5 TM imagery for wetland identification, *Photogrammetric Engineering and Remote Sensing*, 65 (11): 1303-1310

Manavalan, P., Sathyanath P. and Rajegowda G. L., 1993. Digital image analysis techniques to estimate waterspread for capacity evaluations of reservoirs, *Photogrammetric Engineering and Remote Sensing*, 59 (9): 1389-1395

McAllister, L. S., B. E. Peniston, S. G. Leibowitz, B. Abbruzzese, and J. B. Hyman, 2000. A synoptic assessment for prioritizing wetland restoration efforts to optimize flood attenuation, *Wetlands*, 20 (1): 70-83.

Ozesmi, S. L. and M. E. Bauer, 2002. Satellite remote sensing of wetlands, *Wetlands Ecology and Management*, 10 (5): 381-402.

Pal, M. and P. M. Mather, 2003. An assessment of the effectiveness of decision tree methods for land cover classification, *Remote Sensing of Environment*, 86 (4): 554-565.

Pietroniro, A. and T. D. Prowse, 2002. Applications of remote sensing in hydrology. *Hydrological Process*, 16 (8): 1537-1541

Sawaya, K. E., L. G. Olmanson, N. J. Heinert, P. L. Brezonik, and M. E. Bauer, 2003. Extending satellite remote sensing to local scales: land and water resource monitoring using high-resolution imagery, *Remote Sensing of Environment*, 88 (1-2): 144-156.

Schmugge, T. J., W. P. Kustas, J. C. Ritchie, T. J. Jackson, and A. Rango, 2002. Remote sensing in hydrology, *Advances in Water Resources*, 25: 1367-1385.

Stewart, W., V. Carter, and P. D. Brooks, 1980. Inland (non-tidal) wetland mapping,

- Photogrammetric Engineering and Remote Sensing*, 46 (5): 617-628.
- Tatem, A. J., H. G. Lewis, P. M. Atkinson, and M. S. Nixon, 2002. Super-resolution land cover pattern prediction using a Hopfield neural network, *Remote Sensing of Environment*, 79 (1): 1-14.
- Todhunter, P. E. and B. C. Rundquist, 2004. Terminal lake flooding and wetland expansion in Nelson County, North Dakota, *Physical Geography*, 25 (1): 65-68.
- Tou, J.T., Gonzalez, R.C., 1974. *Pattern recognition principles*, Addison-Wesley, MA, 377 p.
- Winter, T. C., and D. O. Rosenberry, 1998. Hydrology of Prairie Pothole wetlands during drought and deluge: a 17-year study of the cottonwood lake wetland complex in North Dakota in the Perspective of longer term measured and proxy hydrological records, *Climate Change*, 40 (2): 189-209.
- Work, E. A. and D. S. Gilmer, 1976. Utilization of satellite data for inventorying prairie ponds and lakes. *Photogrammetric Engineering and Remote Sensing*, 42 (5): 685-694.
- Xu, M., P. Watanachaturaporn, P. K. Varshney, and M. K. Arora, 2005. Decision tree regression for soft classification of remote sensing data, *Remote Sensing of Environment*, 97 (3): 322 - 336.