The Lake Michigan Ozone Study: An Application of Satellite-Based Land-Use and Land-Cover Mapping to Large-Area Emissions Inventory Analysis

Donald E. Luman and Minhe Ji

Abstract

Recent research has revealed that emissions arising from both man-made (anthropogenic) and vegetative (biogenic) sources contribute to ground-level ozone problems on a regional scale. To properly estimate emissions for the Lake Michigan Air Quality Region, a photochemical reactive grid model has been implemented based upon a comprehensive land-use and land-cover inventory derived from the classification of 11 scenes of Landsat-5 Thematic Mapper data. The characterization of land use and land cover for large geographic areas presents unique problems, and this paper discusses the development of the inventory, the relationships between biogenic/anthropogenic sources, as well as the specialized procedures devised to successfully complete the remote sensing analysis. Of particular interest is the use of U.S. Census TIGER data to stratify the classification of urban versus rural areas, a variation of the traditional principal components transformation approach to reduce the amount of waveband data, and the strategies employed to conduct accuracy assessment.

Introduction

The Clean Air Act Ammendments of 1990, which ended nearly a decade of political stalemate over the nation's air quality laws, imposed new federal standards regarding urban smog, toxic air pollution, and acid rain. Based upon existing air quality levels, various political jurisdictions within the United States have been segregated into designated nonattainment areas. There are several such nonattainment areas pertaining to ground-level ozone in the vicinity of Lake Michigan, and they are collectively referred to as the Lake Michigan Air Quality Region (LMAQR), comprising nearly 120,000 square kilometres (Figure 1). Despite the adoption of emission control measures within the LMAQR during the 1980s, the states of Illinois, Indiana, Michigan, and Wisconsin have continued to experience numerous violations of the ozone National Ambient Air Quality Standard (NAAQS). The Lake Michigan Ozone Study[1] (LMOS) was initiated by the Lake Michigan States to develop a regional solution to the ground-level ozone problem. The goal of the study is to prepare a photochemical reactive grid model of the formation and transport of ozone that will be used to support the development of a regional control plan.

Both man-made (anthropogenic) and natural (biogenic) precursor emissions (e.g., hydrocarbons and oxides of nitrogen) contribute to ozone air quality problems. Research extending back to the mid-1960s has established that biogenic emissions arising from organic compounds may be comparable to, or even exceed, the non-methane hydrocarbon-based emissions from anthropogenic sources on a regional scale (Table 1). Biogenic emissions have been found to be highly reactive compounds which can contribute to the formation of tropospheric ozone (Winer, 1989) and may perform an important role in the production of ground-level ozone. While Winer acknowledges the fact that many vegetative species emit significant amounts of hydrocarbons is well-documented, there is essentially little or no quantitative information on emission factors for agricultural crops. The dearth of information is attributed to the emphasis given to assessing biogenic emissions for selected vegetative species including forests, scrublands, and grasslands, as well as those which are prevalent within urban airsheds.

The total assigned plant emissions (TAPE) provides an approximation of the relative emissions factors for differing plant species, and the various agricultural crops examined in the Winer study can be qualitatively grouped by orders of magnitude using TAPE emission rates (Table 2). Pasture/wheat and alfalfa/cherry/sorghum are shown to be included in the low- and medium-emissions categories, respectively, which are agricultural crops of particular importance within the LMAQR. In a related study, it was revealed that there was

[1]The Lake Michigan Ozone Study is jointly funded by the U.S. Environmental Protection Agency and the states within the Lake Michigan Air Quality Region, and is administered by the Lake Michigan Air Directors Consortium.

D.E. Luman was with the Department of Geography and Minhe Ji is with the Center for Governmental Studies, both at Northern Illinois University, Dekalb, IL 60115-2854. He is presently with the Illinois State Geological Survey, 615 E. Peabody Drive, Champaign, IL 61820.

Photogrammetric Engineering & Remote Sensing, Vol. 61, No. 8, August 1995, pp. 1021–1032.

0099-1112/95/6108-1021$3.00/0

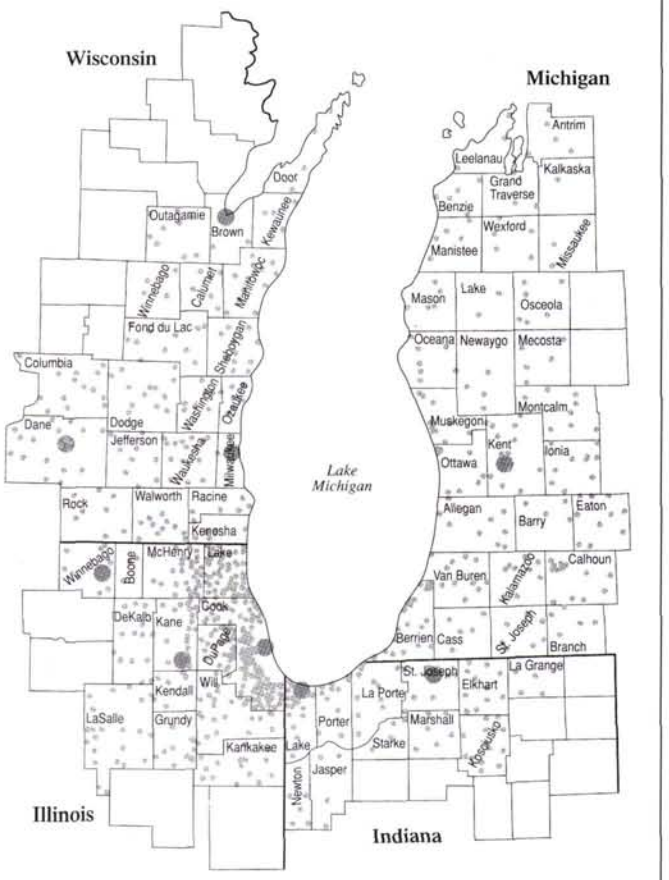

Figure 1. Lake Michigan Air Quality Region. Each dot represents an urban or built-up area, with the larger dots delineating the largest urbanized areas within the LMAQR.

at least two orders of magnitude difference in the non-methane hydrocarbons (NMHC) emissions between alfalfa/wheat and corn. Furthermore, the emission rates for deciduous forest species are comparable to or exceed that of corn, with one coniferous species (loblolly pine) exhibiting NMHC emissions several orders of magnitude higher than those of agricultural crops or deciduous forest species (Table 3). In airsheds such as the LMAQR where both forest and agricultural plant species are abundant, and whose biogenic emissions can surpass anthropogenic sources, these sources must be accounted for in any regional emissions inventory.

When combined with land-use data or biomass distribution maps, data from emissions studies permit calculation of an explicit gridded emissions inventory (Figure 2). In the past, such inventories have customarily been based upon generalized land-use and land-cover information, including U.S. Geological Survey Land Use and Land Cover and Associated Map data (USGS LU/LC) (Loelkes *et al.*, 1983), county-level comprehensive plans, and state-level land-use and land-cover inventories, among others. Such data sources lack sufficient taxonomic resolution to account for the significant differences in biogenic emissions acquired from recent research. For example, the USGS LU/LC classification system (Anderson *et al.*, 1976) aggregates row crops and pasture at Level II (e.g., category 021, Cropland and Pasture), two agricultural crops which possess differences in emissions rates of several orders of magnitudes. Furthermore, the USGS LU/LC data are seriously outdated for much of the LMOS project area, with the dates of the source materials extending from 1971 through 1982. Given the factors of temporal and spatial variability in biogenic and anthropogenic emission rates across the LMAQR, a current inventory of land-use and land-cover databases needed to be developed which exceeded the spatial and taxonomic resolution of previous emissions inventories. After much discussion and debate, a compromise classification scheme incorporating 18 land-use and land-

TABLE 1. NON-METHANE HYDROCARBON EMISSION TOTALS FOR THE LAKE MICHIGAN AIR QUALITY REGION (SOURCE: LMADCO).

Source Category	*Emissions (kg/day)*
Motor Vehicles	570,000
Point Sources (Industrial Operations)	630,000
Area Sources (Commercial/Consumer Operations)	750,000
Biogenics	1,400,000
Total Emissions	**3,350,000**

TABLE 2. QUALITATIVE GROUPING OF AGRICULTURAL CROPS BY RATES (μG $HR^{-1}GM^{-1}$) OF TOTAL ASSIGNED PLANT EMISSIONS (FROM WINER, 1989).

Low *<1*	*Middle* *1 - 10*	*High* *>10*
Bean	Alfalfa	Pistachio
Nectarine	Almond	Rice
Olive	Apricot	Tomato
Orange	Carrot	
Pasture	Cherry	
Wheat	Cotton	
	Grape	
	Lemon	
	Orange	
	Peach	
	Plum	
	Safflower	
	Sorghum	
	Walnut	

TABLE 3. EMISSION RATE ESTIMATES (30°C)(μG CM^{-1} HR^{-1}) (FROM WINER, 1989)

Type	*Examples*	*Isoprene*	*α-Pinene*	*Other NMHC*
High isoprene	Oak	22.9	0	1.8
Low isoprene	Sycamore	8.4	0	2.3
Deciduous, no isoprene	Maple	0	1.4	4.3
Coniferous	Loblolly Pine	0	2.8	8.0
Agriculture	Alfalfa, Wheat	------	------	0.015
	Tobacco	------	------	0.35
	Corn	------	------	2.0
	Other	------	------	0.015
Water	-------	------	------	145 μg m^{-2} hr^{-1}

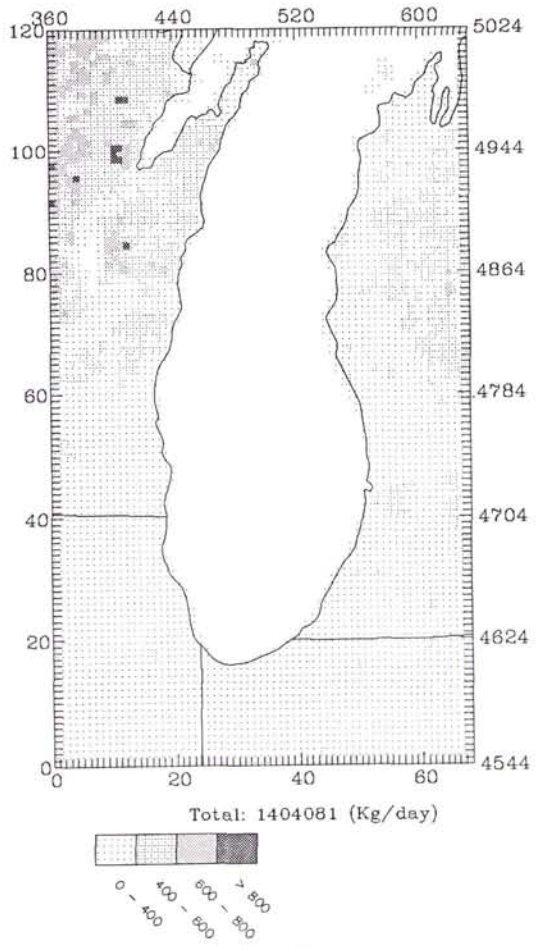

Figure 2. Non-methane hydrocarbon emissions from biogenic sources in the Lake Michigan Air Quality Region.

cover categories was adopted (Table 4). These categories were selected both for their compatibility with satellite-based sensor systems as well as their importance in maintaining the principal objectives of the LMOS project.

Data Sources

Selection of an Appropriate Primary Data Source

The LMOS photochemical reactive grid model uses a three-dimensional grid structure to cover the region of interest. The horizontal dimension of the grid cell model is typically on the order of 4 to 5 kilometres, and emissions information for this spatial scale, or smaller, is anticipated for the proposed modeling. The need for land-use and land-cover data on this scale over the entire LMAQR therefore imposes special limitations on potential data sources. Landsat-5 Thematic Mapper (TM) satellite data were selected as the primary data source for development of the land-use and land-cover database, appropriate given the potential range in spatial resolution parameters expected for the grid model. To ensure that important contrasts in vegetative phenology were captured, multitemporal data sets were initially considered; however, budgetary constraints resulted in the acquisition of a single coverage,

TABLE 4. LAKE MICHIGAN OZONE STUDY CLASSIFICATION SCHEME.

Information Class	*Comments*	*Abbrev.*
High Density Urban	Commercial/industrial; >90% impervious surface	**HD**
Medium Density Urban	Commercial/residential; 50-90% impervious surface	**MD**
Low Density Urban	Residential/open space;<50% impervious surface	**LD**
Transportation	Major roadways derived from USGS DLG Transportation	**TR**
Row Crops	Major Commodity crops, e.g., corn/soybeans	**RC**
Small Grains	Major Commodity crops, e.g., wheat/oats/rye	**SG**
Grassland & Other Ag.	Predominately open space, scrubland, hay	**GR**
Orchards	Major orchards derived from USGS LU/LC	**OR**
Upland Deciduous Forest		**DU**
Lowland Deciduous Forest		**DL**
Upland Coniferous Forest		**CU**
Lowland Coniferous Forest		**CL**
Mixed Upland Forest		**MU**
Mixed Lowland Forest		**ML**
Undifferentiated Water	Permanant open water; lakes and ponds, rivers, etc.	**WA**
Undifferentiated Wetland	Persistent, mostly palustrine wetland habitat	**WT**
Rangeland		**RA**
Barren Land	Quarries, ruderal land, etc.	**BL**

late spring/early summer "snapshot" of the LMOS study area. Eleven TM scenes were selected with acquisition dates incorporating three calendar years and spanning almost two months in vegetative phenology (Figure 3). Given that important phenological changes occur in the upper Midwest approximately at the rate of 7 to 10 days for every degree of geographic latitude, the 5° latitudinal extent of the LMOS pro-

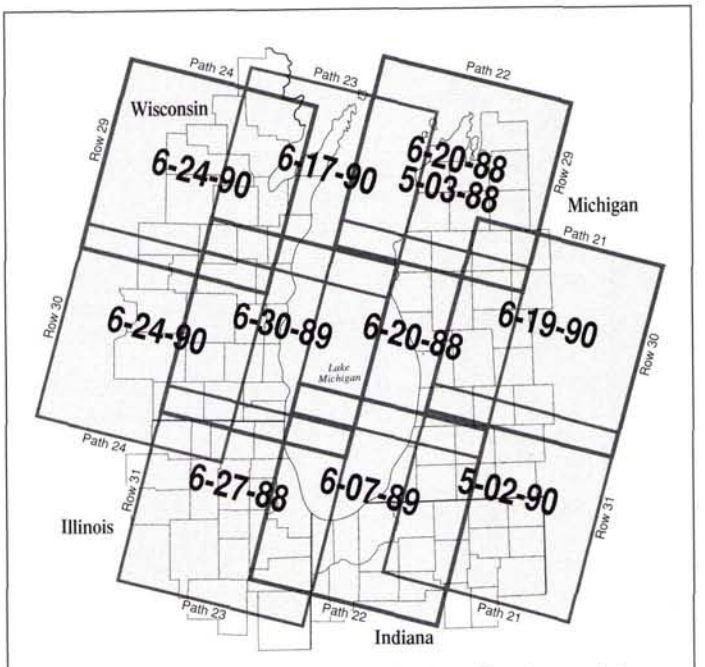

Figure 3. Landsat 5 TM scene selection. Due to persistent cloud cover over Path 22/Row 29, portions of two TM scenes were used.

ject area constituted an additional, important factor in the selection of the satellite data.

All TM data sets had pre-processing procedures applied and included (1) geocoding of the image data to the Universal Transverse Mercator Projection, UTM Grid Zone 16, and resampling to a 25- by 25-metre spatial resolution, and (2) terrain correction. Even though the local relief within any single Landsat TM data set does not exceed 150 metres, the terrain correction procedure performed by STX Corporation also ensures the mosaicking ability of adjacent TM scenes. To facilitate the proposed grid modeling, a convenient means for subsetting the completed land-use and land-cover map data was required. Accordingly, each TM data set was subsequently trimmed to conform to the USGS 1:100,000-scale quadrangle areas contained within the respective TM scene area (Figure 4).

Selection of Ancillary Data Sources

The physical dimensions of the LMOS project area precluded any systematic field verification. Two additional sources of image data were acquired to complement the TM data sets and provide the higher resolution necessary to conduct the classification and post-classification accuracy assessment procedures. A 15 percent sample, or approximately 700 individual frames of color-infrared photography from the National Aerial Photography Program (NAPP) and National High Altitude Photography (NHAP-2), was selected using a systematic sampling scheme. The LMOS project area manifests no pronounced topographic lineaments or other alignments; therefore, this sampling approach ensured a proper level of landscape stratification while also adhering to budgetary constraints. While much of the photography differed by 1 to 2 years and some as much as 4 years in calendar date from the TM scene data, with several frames differing only by several days, the majority of the photography spanned only 1 month in phenology. This meant that persistent landscape features such as forested land, wetland, water, etc., could be accurately interpreted. Because the nominal photographic scales, orientation, and location of all flightlines are known, systematic variables for the NAPP/NHAP-2 programs, an ARC/INFO GIS map coverage was created which defined the areal extent of each frame of photography.

A second photographic data source involved the acquisition of USDA-ASCS 35-mm Color Compliance photography. Each slide is acquired in a near-vertical orientation and portrays approximately one PLSS section (259 hectares) of land area. Typically acquired in the upper Midwest during the summer months, the photography portrays an unusually high level of feature detail (Figure 5). While the photography is itself valuable, it is the ASCS-578 and ASCS-156EZ crop report information developed from a combination of photointerpretation and farm manager data at the local ASCS office that is of even more interest. For a modest charge, the individual county ASCS office can compile a lithographed "photo-map" delineating each field and specific crop type for a desired crop year (Figure 6). The ASCS Compliance information was acquired for a sample of 26 counties which were determined to be representative of the diversity and prevalence for agricultural crops within the LMOS study area. The UTM coordinates for the centroids of all fields identified within these data were digitized for each TM scene area and encoded with the crop type. This approach significantly improved the ability to characterize agricultural lands with a measured level of accuracy, a landscape which combines spatial diversity with a high degree of spectral and temporal variation.

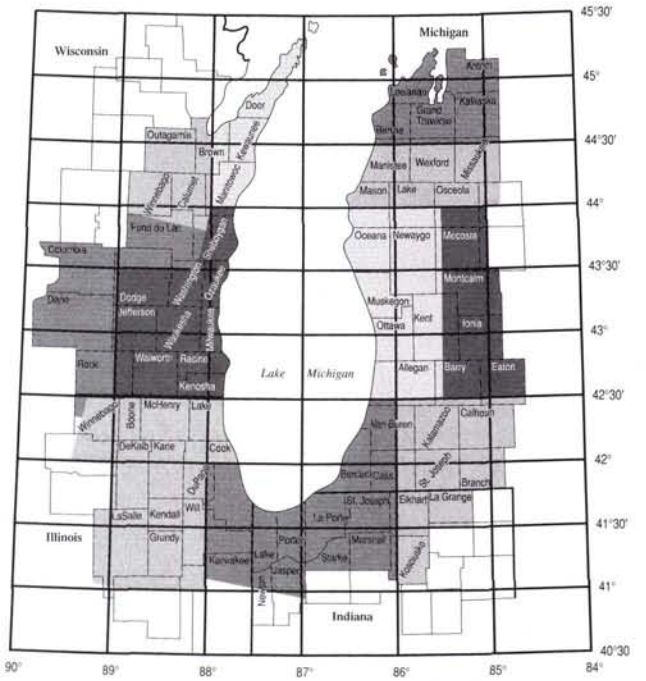

Figure 4. Landsat TM scene areas as subsetted by USGS 1:100,000-scale map coverage. Compare with Figure 2.

Analysis

Data Reduction by Principal Components Transformation

Because of the large-scale nature of the study, a method of reducing the substantial amount of image data was deemed

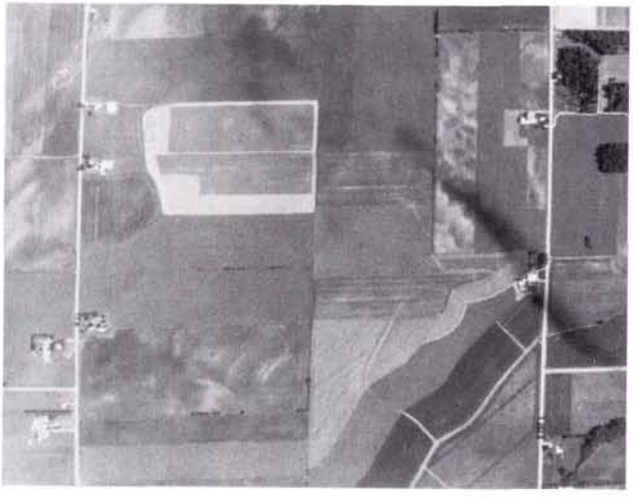

Figure 5. USDA-ASCS 35-mm color compliance photography acquired over Sun Prairie Township in Dane County, Wisconsin during June, 1990. Compare with ASCS Crop Report information shown on Figure 6.

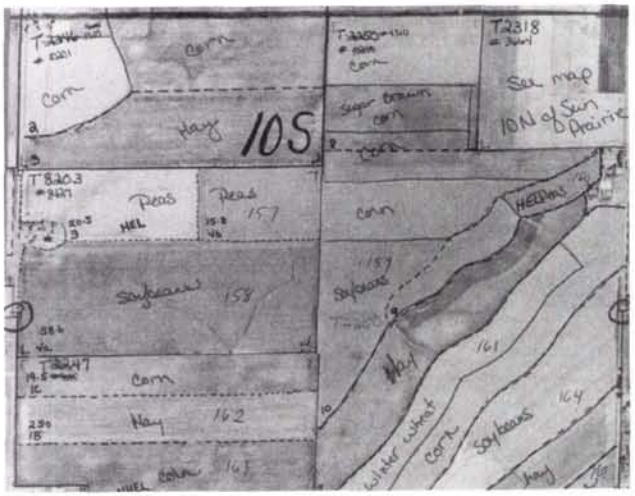

Figure 6. ASCS Crop Report "photo-map" corresponding to Figure 5. Older aerial photography is typically utilized as the base for annotating the individual fields. Approximately one PLSS section of land is enclosed.

necessary, and a principal components transformation (PCA) was selected because it is a widely recognized and efficient procedure for eliminating the redundancies inherent in multispectral data (Table 5a) (Showengerdt, 1983; Richards, 1986; Gong and Howarth, 1990). Table 5b presents the factor loading matrix for the Path 23/Row 30 scene data after the initial stage of the PCA model, with the first three components accounting for 98.97 percent of the total variance (Table 5c). Because the greatest drop-off in eigenvalues occurs between Component 3 and Component 4 and this has traditionally been used as a cut-off criterion (Rummel, 1970), components beyond this level were eliminated from any further consideration. This procedure effectively reduced the amount of data which needed to be processed to only 43 percent of the original waveband information.

Component 1 is actually what is referred to as a "total brightness image" (Richards, 1986) because most of the TM bands load high on that component. This factor pattern is typical of remote sensing applications which employ the standard (initial) principal components transformation, actually an *unrotated* factor method. The authors elected instead to utilize a *varimax* rotation method for the PCA transformation, a standard PCA option in most statistical program packages such as SAS, SPSS, and BIOMED, among others. The advantage of this approach is that, while the amount of total variance is unchanged, the interpretability of the output component images is greatly enhanced. After varimax rotation, the factor loading matrix exhibits a strong pattern of correlated bands for the first three components (Table 5d) and can be visually interpreted much like the original Landsat TM data set (Figure 7). This is an important consideration because many decisions concerning the identification of landscape elements are ultimately still dependent upon the analyst's direct interpretation of the color and/or tone.

Pre-Classification Procedures

ERDAS image processing software was utilized for the majority of the image analysis and raster GIS operations. Prior to the classification stage, spectral signatures were extracted from each PCA data set utilizing an Isodata *K*-means clustering procedure (Duda and Hart, 1973), and experimentation indicated that 200 to 250 spectral clusters should be derived for each data set. The separability of all possible pairwise combinations was assessed utilizing the transformed divergence statistic (Whitsitt and Landgrebe, 1977), and the authors discovered that equal or greater separability is maintained at lower transformed divergence values when utilizing principal component feature bands as compared to untransformed image data sets, due presumably to the independent nature of the component data. A transformed value of 1600 for all TM scene areas was utilized as the threshold of separability, reducing the original 200 to 250 clusters to approximately 100 merged clusters.

To ensure that the spectral signatures for urban and

TABLE 5a. CORRELATION/COVARIANCE MATRIX OF TM SCENE PATH 23/ROW 30.

	BAND 1	BAND 2	BAND 3	BAND 4	BAND 5	BAND 6
BAND 1	223.88[1]	0.9053[2]	0.8624	0.0311	0.3017	0.4709
BAND 2	130.42[3]	83.92	0.9356	0.0044	0.4135	0.5935
BAND 3	235.57	150.22	287.42	0.0200	0.5545	0.7318
BAND 4	-70.88	-16.30	-64.47	721.74	0.0599	0.0036
BAND 5	271.99	194.94	417.77	217.66	1095.20	0.8434
BAND 6	233.31	160.37	329.56	-36.87	690.61	516.36

[1]Variance; [2]Pearson Product Moment Correlation; [3]Covariance

TABLE 5b. INITIAL FACTOR METHOD: PRINCIPAL COMPONENTS FACTOR LOADING PATTERN FOR TM SCENE 23/30.

	Component 1	Component 2	Component 3
BAND 5	0.97323	0.14250	-0.16697
BAND 6	0.96364	-0.18443	-0.10566
BAND 3	0.87119	-0.32041	0.35161
BAND 2	0.79518	-0.26284	0.52610
BAND 1	0.71278	-0.36798	0.57888
BAND 4	0.13648	0.97841	0.15248

TABLE 5c. EIGENVALUES AND PROPORTION OF VARIANCE FOR TM SCENE 23/30 DATA SET.

	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6
Eigenvalue	1,915.24	796.33	186.86	20.59	8.59	0.91
Difference	1,118.91	609.47	166.27	12.00	7.67	
Variance	65.40%	27.19%	6.38%	0.70%	0.29%	0.03%
Cumulative	65.40%	92.59%	98.97%	99.68%	99.97%	100.0%

TABLE 5d. PCA ROTATED FACTOR LOADING PATTERN MATRIX ROTATION METHOD: VARIMAX.

	Component 1 'Visible'	Component 2 'Mid-IR'	Component 3 'Near-IR'
BAND 1	***0.94430***	0.27455	-0.10724
BAND 2	***0.91463***	0.37619	-0.01018
BAND 3	***0.83177***	0.53206	-0.10168
BAND 5	0.33549	***0.91430***	0.21648
BAND 6	0.47490	***0.86122***	-0.08097
BAND 4	-0.09195	0.06855	***0.99298***

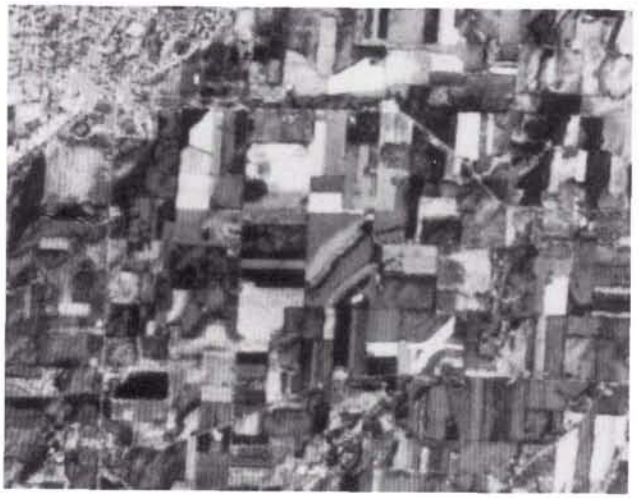

Figure 7. Landsat TM image acquired on 24 June 1990, and displayed using Principal Components 1, 2, and 3. The field patterns shown in Figures 5 and 6 are at the scene center.

built-up lands were properly characterized during the unsupervised training stage, the Place Boundaries for all incorporated areas (Figure 1) were extracted from the 1990 post-Census TIGER database and used to separate urban areas from the PCA data sets (Figure 8). Subsequently, 100 to 125 spectral clusters for each TIGER-based PCA data set were derived and, in contrast to the non-urban PCA data sets, no signature merging was performed in order that subtle urban landscape elements were not inadvertently combined. This innovative use of TIGER data was demonstrated throughout the LMOS project to be very effective, especially in a large-scale study where urban and built-up lands constitute a small but important percentage of the overall landscape.

Classification Procedures

The performance of several standard classifiers, including minimum-Euclidean distance, minimum-Mahalanbois distance, and non-thresholding maximum-likelihood, were evaluated for a test area prior to the classification of the PCA data sets. The results indicated greater improvement in classification accuracy would be achieved from the use of a maximum-likelihood classifier, and this is supported by recent research (Gong and Howarth, 1990).

Post-Classification Procedures

The result of the classification stage were two classification maps (urban and non-urban) for each TM scene area, each comprised of approximately 100 spectral classes. Several post-classification procedures were imposed upon the classification maps:

- Because a fundamentally unsupervised approach was utilized, each spectral class was assigned to the most appropriate LMOS category through careful comparison to all available ancillary information.
- The final classification maps often delineated the transportation network inconsistently. Therefore, major roadways, railroads, and airports derived from the USGS 1:100,000-scale Digital Line Graphs were imposed upon the final classification maps by a raster GIS model.
- It was impossible to delineate the Orchards category through spectral classification. This category was delineated principally by imposing the Orchards category derived from Level II, USGS LU/LC data onto the final classification. A raster GIS model was used to preserve certain landscape elements (e.g., forested and grassland areas) within the generalized orchard areas in order to impart a more "natural" appearance.
- The compromise classification scheme required that forested tracts within the LMOS study area be separated into upland and lowland phases. The use of the 1:250,000-scale DEMs proved to be of little use due to the three-arc-second resolution, especially within the southern half of the project area where the occurrences of forest were primarily restricted to riverine areas. Buffers created along the perennial streams derived from the USGS 1:100,000-scale DLG Hydrography were used to reclassify upland forest to their respective lowland counterparts. This methodology was refined in the northern portion of the project area where extensive tracts of lowland forest do exist. The Marsh category from the DLG Hydrography was used as a Boolean mask to reclassify not only lowland forests, but also to segregate wetland habitat from the Grassland & Other Agriculture category.
- The urban and non-urban classification maps were subsequently combined using a raster GIS model and recoded to create a single 18-category classification map for each TM area (Figure 9). Every effort was also made to ensure the transition between classification maps derived from differing dates of imagery was as consistent as possible (Figure 10). A constrained logical filter was applied to the completed maps to remove isolated artifacts while preserving linear features.

Accuracy Assessment Analysis

Sampling Scheme and Sample Size

While sufficient research has been conducted on the topic of sample size determination to establish general guidelines, there have been few systematic investigations critically evaluating the sampling schemes used to assess the accuracy of

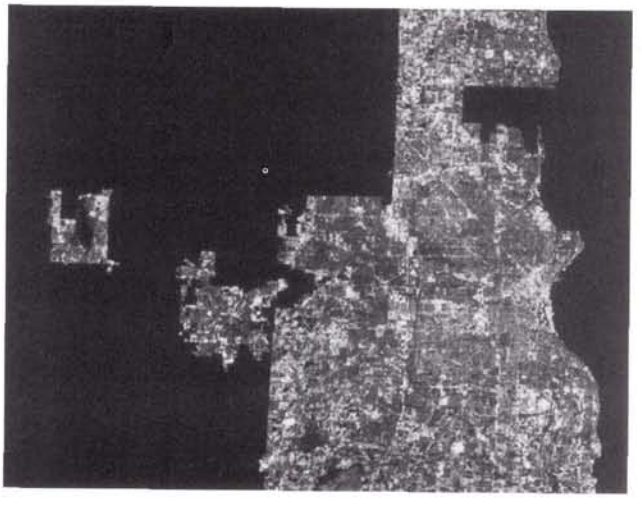

Figure 8. Landsat TM data set acquired on 30 June 1989, which has been clipped for the Milwaukee, Wisconsin area using the 1990 TIGER Incorporated Place Boundaries. Principal Components 1, 2, and 3 are displayed.

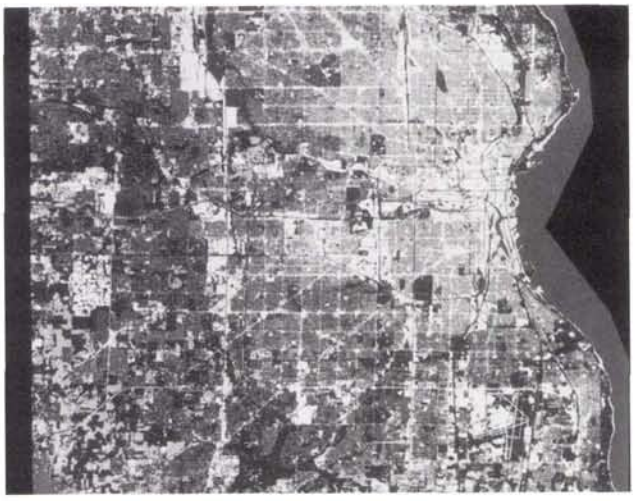

Figure 9. Final classification of the Milwaukee, TIGER-based PCA data set. White delineates high-density urban land cover, while the medium-gray and dark-gray toned areas represent medium and low-density urban land cover, respectively.

land-use and land-cover maps derived from remote sensor data. Congalton (1988) used three generalized landscape classes (e.g., forested, agricultural, and rangeland) to evaluate five most commonly used sampling strategies, including simple random sampling, stratified random sampling, cluster sampling, systematic sampling, and stratified systematic unaligned sampling. He concluded that systematic and stratified systematic unaligned sampling should not be used within predominantly agricultural landscapes, because these two sampling schemes consistently overestimate the population mean. Simple random sampling performs adequately for all three landscapes, and stratified random sampling is advised to ensure that small but important landscape areas are sampled.

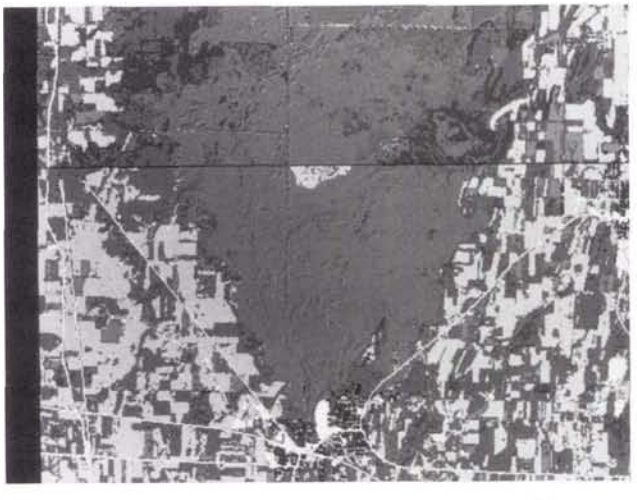

Figure 10. Juncture of classification maps derived from two separate Landsat TM scene areas. The join line has been purposely enhanced to depict the transition area.

The LMOS project area contains significant tracts of both forested and agricultural land cover. Furthermore, the agriculture category is subdivided into three subclasses, creating numerous small areas and increasing the spatial complexity of this landscape. A random sampling scheme stratified by landscape category was therefore deemed most appropriate. A second level of stratification was also emplaced upon the sampling by utilizing the ARC/INFO coverage created for the reference aerial photography as a binary mask to stratify sample point selection, such that all map areas outside of the photo areas were disallowed for sampling.

Because of the diversity in scene acquisition dates, variation in TM scene quality, and changes in phenological conditions across the LMOS project area induced both by latitude and the year-to-year differences in ground conditions, the computer classification was conducted on a scene-by-scene basis. This required that the accuracy assessment procedures, and therefore the point sampling, also be carried out on the same basis. The stipulated level of mapping accuracy for the project was stated such that only the overall accuracy must attain at least 85 percent. Each classification map can be considered as a separate statistical population possessing a binomial distribution, and the appropriate sample is derived as follows (from Jensen, 1986):

$$N = \frac{4 \cdot p \cdot \tilde{q}}{E^2}$$

where p = expected percent accuracy (e.g., 85 percent), $\tilde{q} = 100 - p$, and E = allowable error (e.g., 5 percent).

Because the accuracy assessment phase included no actual field work, lowering the allowable error assists in offsetting procedural errors (Fitzpatrick-Lins, 1981), thereby decreasing the probability of chance error. Reducing the allowable error slightly to a value of 4 produces a minimum sample size (N) of 319 points for each TM scene area. Table 6 is the contingency matrix showing the disposition of all sample pixels derived from the sampling of classification maps for the 11 TM scene areas. Although the accuracy assessment procedures were carried out on a scene-by-scene basis, the accuracy/error statistics for all 11 scene areas have been composited into a single contingency matrix for the purpose of discussion. This was accomplished by weighting the individual categorical accuracy/error data contributed from each TM scene area utilizing the respective ground areas for each mapping category.

Classification Accuracy versus Classification Error

Traditionally, the omission and commission errors are presented in conjunction with the contingency matrix. The former represent those sample pixels pertaining to an actual class on the landscape which the computer classification has failed to recognize (off-diagonal column elements), while the latter refers to those sample pixels from other landscape classes which the computer classification has incorrectly assigned as belonging to the particular landscape class of interest (off-diagonal row elements). The interpretation of omission and commission errors is often a source of confusion to the user, and it is sometimes more expedient to have a simple measure of categorical accuracy. The terms "producer's accuracy" and "user's accuracy" have therefore been

TABLE 6. CONTINGENCY MATRIX AND WEIGHTED SUMMARY STATISTICS FOR LMOS PROJECT AREA.

Classified Data \ Reference Data	HD	MD	LD	TR	RC	SG	GR	OR	DU	DL	CU	CL	MU	ML	WA	WT	BL	Total Samples	Total Km^2	UA(%)*	CE(%)*	$\hat{\kappa}$(%)*
HD	**90**	4	2		3		2									1	1	103	2,174.5	83.6	20.9	82.9
MD	5	**118**	4				2		1							1	1	132	3,265.5	90.7	8.6	89.9
LD	4	10	**63**		2		5		1		1						1	87	929.6	71.3	37.6	70.7
TR				**37**			2											39	781.8	91.7	10.7	91.6
RC	1	3			**902**	13	65	1										983	35,464.7	91.8	8.4	88.0
SG					5	**92**	13		1									111	3,427.3	79.2	16.9	77.9
GR	1	6	3		76	50	**796**		40	6	3		2			5		988	37,860.1	79.1	23.2	73.0
OR								**25**			1							26	435.2	97.9	4.1	97.9
DU		1			2	6	37		**634**	14	2	2	9	1		3		714	18,733.6	90.0	10.8	87.5
DL					5	3	9		4	**146**			2	3		8		180	3,720.1	84.2	16.1	83.2
CU						1			1	1	**83**		8		1	1		96	1,759.6	82.7	20.7	82.0
CL										1		**17**	1	1				20	248.1	89.1	0.9	89.1
MU									7	3	6	1	**57**	1		4		79	1,637.5	75.1	32.0	73.7
ML							1		1	3				**15**		1		21	496.4	75.5	38.2	75.3
WA					1		1		1						**161**			164	7,506.9	98.3	1.7	98.3
WT					2		2			5	2			1	2	**66**		80	1,605.2	72.3	17.7	71.7
BL	3						1										**38**	42	246.8	92.6	17.8	92.6
Total	104	142	72	37	998	165	936	26	691	179	98	20	79	22	164	90	41	3,864	120,292.8	$\bar{x}$=85.0	$\bar{x}$=15.7	$\bar{x}$=83.8
PA (%)*	89.7	80.3	88.5	99.8	89.0	60.4	81.9	93.3	92.3	79.7	80.0	82.3	80.8	76.8	99.2	71.2	96.1	$\bar{x}$=84.8				
OE (%)*	10.3	19.7	11.5	0.2	11.0	39.6	18.1	6.7	7.7	20.3	20.0	17.7	19.2	23.2	0.8	28.8	3.9	$\bar{x}$=15.2				

* User's Accuracy (UA), Commission Error (CE), Kappa Coefficient ($\hat{\kappa}$), Producer's Accuracy (PA), and Omission Error (OE) values have been composited from the eleven TM scene areas and areally weighted. Information classes are in same order as Table 4.

advanced (Story and Congalton, 1986), and these measures have been included as part of the contingency matrix for this paper. The producer's accuracy is so-called because the producer (or originator) of the classified map is principally interested in how well sample pixels from the reference data can be mapped using the remote sensing data. In contrast, the user's accuracy is an indication of the probability or *reliability* (Congalton and Rekas, 1985) that a sample from the classification map actually represents that category on the actual landscape.

Comparison of the omission error and producer's accuracies in Table 6 shows they are associated in a simple, inverse manner (e.g., PA% = 100 − OE% or OE% = 100 − PA%). Unfortunately for the user, although authors have associated the user's accuracy with commission error, these two measures are not similarly related. The commission errors and user's accuracy are computed using the column and row marginal totals, respectively, an important distinction because these totals can be quite different. In contrast to omission error and producer's accuracy, there may be little actual correlation between commission error and user's accuracy (Figures 11 and 12). Nevertheless, because it is the user who must ultimately be reassured as to the integrity of the maps interpreted from the remote sensing data, the categorical data for Figures 11 through 15 (*x*-axis) have been ordered according to the user's accuracy.

From the viewpoint of the user, when the classified maps for the entire LMOS project area are summarized, the weighted overall mean for user's accuracy is 85.01 percent. Furthermore, the overall weighted mean for commission error is 15.7 percent. From the standpoint of the producer, the weighted overall mean for producer's accuracy is 84.8 percent with a corresponding weighted overall omission error of 15.2 percent (Table 6). In the absence of any universal standard regarding what level of error should be considered significant, it is suggested that the widely accepted standard of 85 percent classification accuracy (Anderson, 1976; Ginevan, 1979; Hay, 1979; Aronoff, 1982; Rosenfield and Fitzpatrick-Lins, 1982; Congalton, 1983) could be utilized to establish an *ad hoc* criterion for classification error. Consequently, classified maps or individual map categories which exceed 15 percent error should be identified and the patterns of error should be examined more closely. Utilizing these two crite-

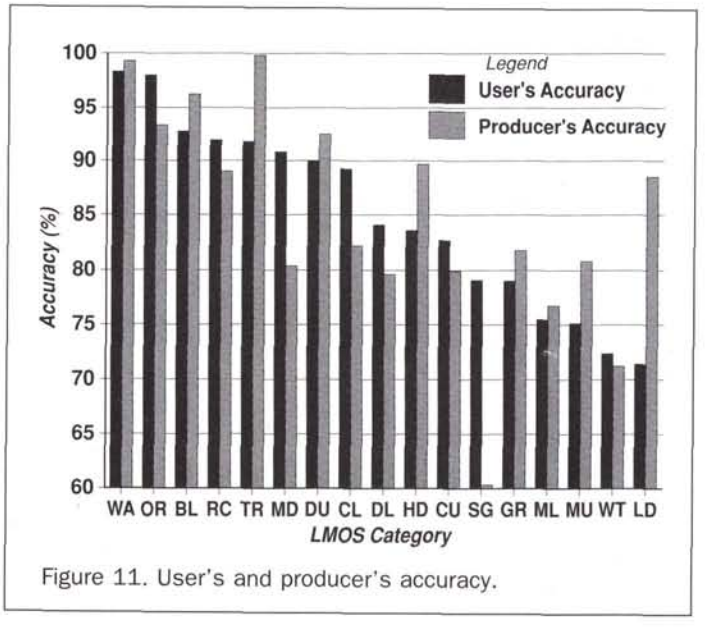

Figure 11. User's and producer's accuracy.

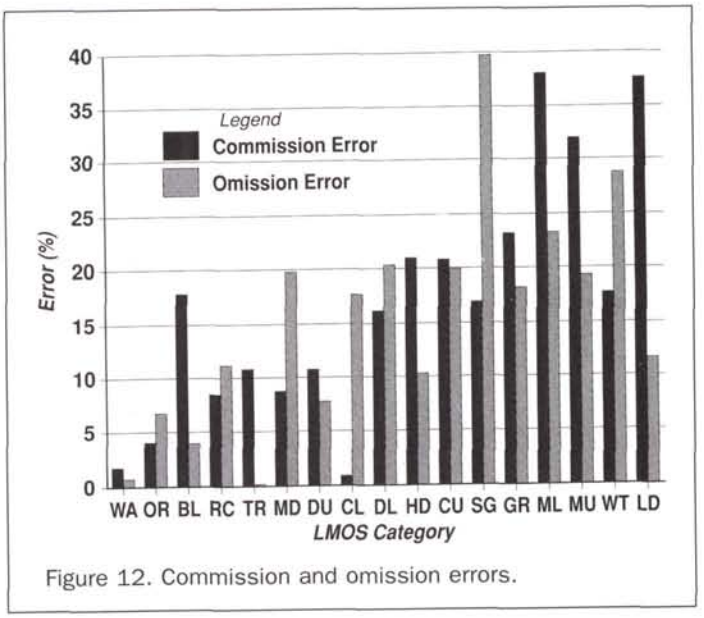

Figure 12. Commission and omission errors.

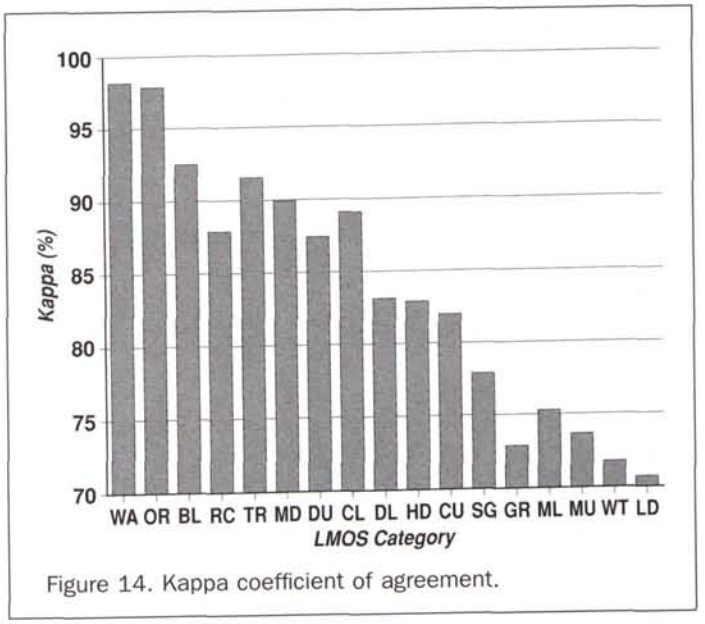

Figure 14. Kappa coefficient of agreement.

ria, the overall interpretation of the remote sensing data for the LMOS project is shown to be at the threshold of acceptability.

A cursory inspection of Figure 11 reveals that several of the individual map categories exceed a user's accuracy of 85 percent or higher, indicative of good reliability. The map class which registered the lowest user's accuracy is LD urban land cover (71.3 percent), perhaps not surprising given that this landscape class tends to be more contextual in nature and not spectrally distinct. The associated commission error is significant (37.6 percent) (Figure 12) and is comprised of sample pixels from seven other landscape classes. Conversely, the LD urban class achieved a high producer's accuracy (88.5 percent) and therefore low omission errors (11.5 percent), indicative that comparatively little of the reference data was misclassified.

Inspection of Figures 11 and 12 also shows that the SG class is distinctive in terms of the disparity between user's/producer's accuracies and the associated commission/omission errors. The success in characterizing this highly spatially and temporally variable cover type lies in judicious TM scene selection, and while the May-June period should capture much of the winter wheat in the LMOS project area, the many other small grains (e.g., barley, oats, spring wheat, etc.) are extremely difficult to consistently discriminate. The differences in farm management practices, most importantly those arising from participation in state/federal conservation reserve programs, cause as much local variation in reflectance patterns as does the TM scene acquisition date (Figures

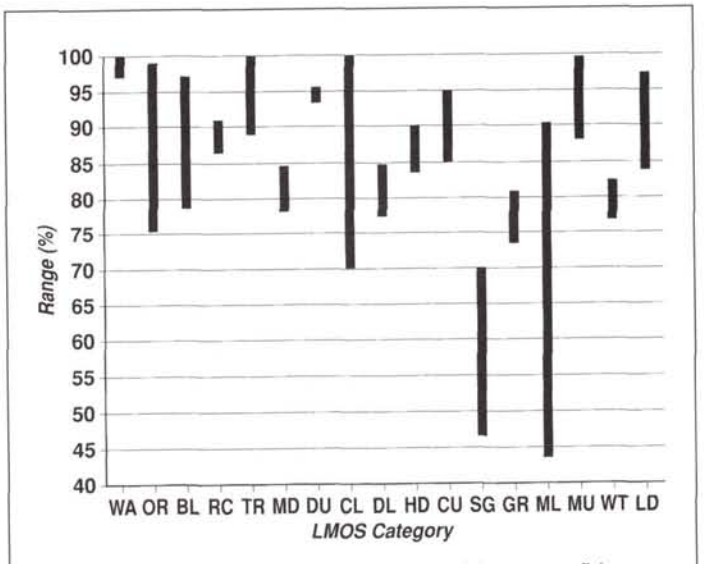

Figure 13. Ninety-five percent upper and lower confidence limits.

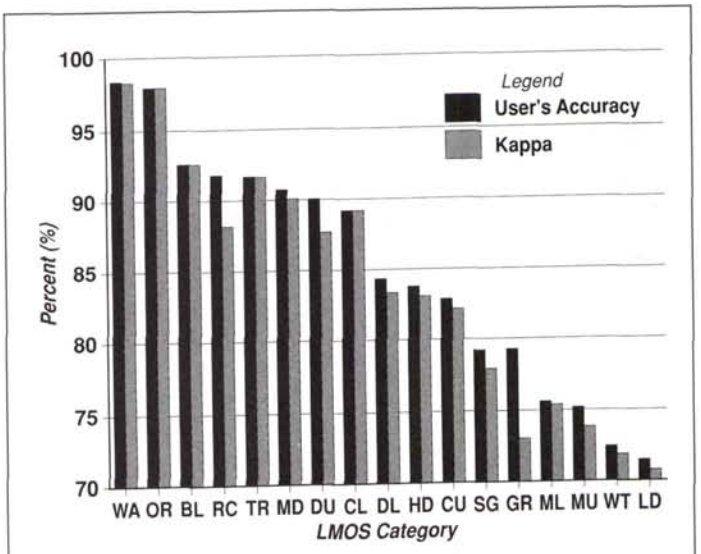

Figure 15. User's accuracy versus kappa.

5 through 7). While the acceptable user's accuracy (79.2 percent) may give the impression that the category has been mapped reliably, the low producer's accuracy (60.4 percent) is indicative that a significant amount of the reference information has been omitted from that landscape class.

The above discussion can easily be extended to all map categories, but the fundamental concern to the users of the classified map data is how the patterns of accuracy/error relate to differences in the spatial allocation of anthropogenic and biogenic emission rates. For example, commission errors for the GR category, comprising 31.5 percent of the total LMOS area, are conspicuous in that they are distributed across ten landscape classes. Unfortunately, the majority of the omission and commission errors associated with the GR class occur with RC agriculture and DU forest, two categories which possess biogenic emission rates several orders of magnitude higher than any vegetative category. What is fortuitous is that the emission rate for grassland is more comparable to that of small grains and therefore does not significantly affect the proposed modeling for that landscape class. It is, therefore, important to recognize the pattern of such errors in order that they may be compensated for in the photochemical grid modeling. One method of increasing the reliability of the proposed model is to merge selected categories which are compatible in terms of biogenic/anthropogenic emission rates (for example, SG→GR), thereby minimizing the errors of commission and omission.

Confidence Level Testing

The estimated mean probability of correct classification for each land-use and land-cover category is a random variable with a binomial probability distribution. The associated statistical error of an accuracy estimate is computed based upon estimates of the mean of each binomial distribution, and it is useful to express this error in the form of an interval within which the true accuracy of each map category lies. As such, it provides an additional, important source of information to the map user. The two-tailed 95 percent confidence limits for a binomial distribution are determined as follows (from Richards, 1986):

$$\frac{x + 1.921 + 1.960 \cdot \sqrt{\frac{x(n-x)}{n} + 0.960}}{n + 3.842}$$

where n is the number of samples for a category or entire map and x is the number of correct sample pixels.

The great disparity in the confidence limits among the various mapping categories is evident in Figure 13, with the ML and CL forest, OR, and SG categories having the greatest uncertainty attached to them. With the exception of the SG class for the reasons explained above, this uncertainty can be directly attributed to the inability of the stratified sampling procedure to provide sufficient numbers of test pixels.

The confidence limits computed for the overall classification yielded acceptable results, with a lower limit of 84.95 percent and an upper limit of 87.62 percent, and the large number of test samples ensured the narrow range of uncertainty of 2.67 percent. Because it is of principal interest to assess how well the overall classification met the minimum standard of 85 percent accuracy, a one-tailed lower confidence limit is more appropriate and is determined as follows (from Jensen, 1986):

$$p = \tilde{p} - 1.645 \cdot \sqrt{\frac{\tilde{p} \cdot \tilde{q}}{n} + \frac{50}{n}}$$

where p is the accuracy of the classification map expressed as a percent, $\tilde{p}$ is the overall percentage correct from the original error matrix, $\tilde{q} = 100 - \tilde{p}$, and n is the sample size.

The lower confidence limit for the overall classification is computed to be 85.20 percent, and thus the user can be confident that the analysis exceeds the 85 percent overall accuracy criterion.

Kappa Coefficient of Agreement

The individual measures discussed above each estimate error and assess accuracy utilizing only a portion of the contingency matrix, and the differing interpretations which result can be a source of unnecessary confusion to the map user. The Kappa coefficient of agreement is both gaining increasing acceptance in the remote sensing community (Hudson and Ramm, 1987) and is attractive in that it effectively summarizes the entire error matrix to a single statistic. Originally devised by Cohen (1960) and Bishop (1975), and recommended for use in remote sensing applications by Congalton *et al.* (1983) and Rosenfield and Fitzpatrick-Lins (1986), the Kappa coefficient is a quantitative measure of the difference between the observed agreement between two images/maps (the traditional "overall percentage correct") and the agreement that may be contributed solely by the chance matching of the two images/maps (derived principally from the off-diagonal elements). In effect, the Kappa coefficient adjusts the *overall* percentage correct measure by subtracting the estimated contribution of chance agreement, which is to infer that the agreement between the two images/maps cannot be attributed exclusively to the "success" of the computer classification. An excellent, general discussion of the Kappa coefficient of agreement is contained in Campbell (1986) and is derived as follows (from Hudson and Ramm, 1987):

$$\hat{\kappa} = \frac{(N \cdot \sum_{i=1}^{r} X_{ii} - \sum_{i=1}^{r} X_{i+} \cdot X_{+i})}{(N^2 - \sum_{i=1}^{r} X_{i+} \cdot X_{+i})}$$

where N = total number of sample points,
X_{ii} = cell value of ith row and ith column,
X_{+i} = sum of row values in ith column, and
X_{i+} = sum of column values in ith row.

Table 6 shows that the estimate of the Kappa coefficient for the overall classification performed 83.8 percent better than if the sample pixels were assigned to random map categories. This is strong evidence that chance occurrence is not an important factor in the performance of the overall classification. In order to assess the accuracy of individual map categories in comparison to the reference data, the conditional Kappa coefficient must be utilized and is derived as follows (from Gong and Howarth, 1992):

$$\hat{\kappa}_i = \frac{N \cdot x_{ii} - x_{i+} \cdot x_{+i}}{N \cdot x_{i+} - x_{i+} \, X_{+i}}$$

The estimated conditional Kappa coefficients for individual classes indicate that chance occurrence is potentially significant in several map categories, primarily those falling below $\hat{\kappa} = 80$ percent (Figure 14). Examination of Figures 4 to 6 suggests a variety of reasons for the lowered conditional

Kappa coefficient values, including unsatisfactory confidence ranges, a high proportion of errors of omission or commission, or decreases in user/producer accuracy percentages. Figure 15 reveals that there is a strong positive relationship (Pearson's $r = 0.9841$, $r^2 = 96.85$ percent) between the conditional Kappa coefficient values and the user's accuracies for the weighted categorical data contained in Table 6. Furthermore, when the original, *unweighted* categorical data for the 11 TM scene areas are utilized instead of the summarized data shown in Table 6, this strong relationship is maintained (Pearson's $r = 0.9934$, $r^2 = 98.68$ percent). This high correlation may in part be obvious, because the determinant of the conditional Kappa coefficient incorporates the summation of classified data, which is the basis for calculating the user's accuracy.

The relationship between user's accuracy and the conditional Kappa coefficient was also determined for all LMOS categories on a scene-by-scene basis. Despite the significant correlation (Pearson's $r^2 \geq 99$ percent) between the conditional Kappa coefficient and user's accuracy for the majority of individual categories, Figure 15 shows small disparities for a few categories (GR, RC, and DU classes, respectively), though the correlation is still quite high (Pearsons $r^2 = 94.1$ percent, 97.1 percent, and 98.1 percent, respectively). These disparities are associated with map classes that have the largest numbers of test samples (see Table 6), and are related to the fact that the conditional Kappa coefficient uses the total number of sample points (N) as a weight. The conditional Kappa coefficient is therefore sensitive not only to the number of sample points in each map class, but also to the distribution of off-diagonal elements within the contingency matrix. Because the user's accuracy does not incorporate N in its derivation, the largest disparities between these two statistical measures will occur within individual map classes with larger numbers of test samples, in which the potential for higher proportions of off-diagonal elements is greater. Coincidentally, the Grassland & Other Agriculture, Row Crops, and Deciduous Upland forest categories were the most highly-sampled map classes in the Lake Michigan Ozone Study, indicative of their respective dominance within the study area. Furthermore, these land-cover categories also required substantially more effort in their interpretation from the remote sensor data (as is manifested in the larger proportion of off-diagonal errors).

Conclusions

The creation of a comprehensive land-use and land-cover inventory utilizing Landsat TM data has been completed for a first-time application of such data with air quality modeling. The anthropogenic and biogenic emission rates for each of the map categories has been established, and the increased spatial and taxonomic resolution of the completed database will facilitate photochemical grid modeling at an order of magnitude higher than heretofore has been accomplished with land-use and land-cover information.

During the year-long extent of the project, the authors experienced many procedural and conceptual problems which are further magnified when attempting such a large-scale mapping project as the Lake Michigan Ozone Study. Even though much research has been conducted on classification and accuracy assessment techniques as they apply to map products derived from remote sensing data, after more than two decades of land-based remote sensing, there are as yet no universally accepted strategies. For example, the authors suggest that further investigation is warranted to clarify the many, subtle interrelationships existing among the various accuracy/error measures currently utilized with maps derived from remote sensing data. The question is not whether one particular statistical measure should be utilized in place of another, but rather, how can the comparability be improved regarding map products produced from remote sensing data. Much more fundamental research is needed to establish standard approaches, and, until then, the producers of remote sensing data products must carefully document the procedures employed for the users of such information.

References

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data, U.S. Geological Survey Professional Paper 964*, U.S. Government Printing Office, Washington, D.C., 28 p.

Aronoff, S., 1982. Classification accuracy: A user approach, *Photogrammetric Engineering & Remote Sensing*, 48(11):1299-1307.

Bishop, Y.M.M., P.W. Holland, and F.E. Fienberg, 1975. *Discrete Multivariate Analysis and Practice*, MIT Press, Cambridge, Massachusetts, 557 p.

Campbell, J.B., 1987. *Introduction to Remote Sensing*, Guilford Press, 551 p.

Cohen, J., 1960. A coefficient of agreement for nominal scales, *Educational and Psychological Measurement*, 20(1):37-46.

Congalton, R.G., 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data, *Photogrammetric Engineering & Remote Sensing*, 54(5):593-600.

Congalton, R.G., and A. Rekas, 1985. COMPAR: A computerized technique for the in-depth comparison of remotely sensed data, *Proceedings of the 51st Annual Meeting of the American Society of Photogrammetry*, Washington, D.C., pp. 98-106.

Congalton, R.G., R. Olderwald, and R. Mead, 1983. Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques, *Photogrammetric Engineering & Remote Sensing*, 49(12):1671-1678.

Duda, R.D., and P.E. Hart, 1973. *Pattern Recognition and Scene Analysis*, John Wiley and Sons, New York, 482 p.

Fitzpatrick-Lins, K., 1981. Comparison of sampling procedures and data analysis for a land-use and land-cover map, *Photogrammetric Engineering & Remote Sensing*, 47(3):343-351.

Gong, Peng, and P.J. Howarth, 1990. An assessment of some factors influencing multispectral land-cover classification, *Photogrammetric Engineering & Remote Sensing*, 56(5):597-603.

———, 1992. Frequency-based contextual classification and gray-level vector reduction for land-use identification, *Photogrammetric Engineering & Remote Sensing*, 58(4):423-437.

Ginevan, M.E., 1979. Testing land-use map accuracy: another look, *Photogrammetric Engineering & Remote Sensing*, 45(10):1371-1377.

Hay, A.M., 1979. Sampling design to test land use map accuracy, *Photogrammetric Engineering & Remote Sensing*, 45(4):529-533.

Hudson, W.D., and R. Ramm, 1987. Correct formulation of the kappa coefficient of agreement, *Photogrammetric Engineering & Remote Sensing*, 53(4):421-422.

Jensen, J.R., 1986. *Introductory Digital Image Processing*, Prentice-Hall, 379 p.

Loelkes, G.L., G.E. Howard, Jr., E.L. Schwertz, Jr., T.D. Lampert, and F.W. Miller, 1983. *Land Use/Land Cover and Environmental Photointerpretation Keys*, U.S. Geological Survey Bulletin 1600, 142 p.

Richards, J.A., 1986. *Remote Sensing Digital Image Analysis: An Introduction*, Springer-Verlag, 281 p.

Rosenfield, G.H., and K. Fitzpatrick-Lins, 1986. A coefficient of agreement as a measure of thematic classification accuracy, *Photogrammetric Engineering & Remote Sensing*, 52(2):223-227.

Rosenfield, G.H., K. Fitzpatrick-Lins, and H.S. Ling, 1982. Sampling for thematic map accuracy testing, *Photogrammetric Engineering & Remote Sensing*, 48(1):131-137.

Rummel, R.J., 1970. *Applied Factor Analysis*, Northwestern University Press, Evanston, Illinois, 617 p.

Schowengerdt, R.A., 1983. *Techniques for Image Processing and Classification*, Academic Press, 249 p.

Story, M., and R. G. Congalton, 1986. Accuracy assessment: A user's perspective, *Photogrammetric Engineering & Remote Sensing*, 52(3):397-399.

U.S. Geological Survey, 1985. *Digital Line Graphs from 1:100,000-Scale Maps: Data Users Guide 2*, Reston, Virginia, 74 p.

Van Genderen, J.L., B.F. Lock, and T.A. Vass, 1978. Remote sensing: statistical testing of thematic map accuracy, *Remote Sensing of Environment*, 7(1):3-14.

Whitsitt, S.J., and D.A. Landgrebe, 1977. *Error Estimation and Separability Measures in Feature Selection for Multiclass Pattern Recognition*, LARS Publication No. 082377, Purdue University, West Lafayette, Indiana, 396 p.

Winer, A M., 1989. *Hydrocarbon Emissions from Vegetation Found in California's Central Valley*, Final report, California Air Resources Board, Statewide Air Pollution Research Center, University of California-Riverside, 23 p.

(Received 13 October 1992; revised and accepted 7 July 1993; revised 21 December 1993)

Donald E. Luman
Donald E. Luman is a Senior Professional Scientist at the Illinois State Geological Survey in Urbana-Champaign. Dr. Luman's research interests include the development of digital image processing/GIS strategies for conducting large-area land-use/land-cover inventories, the application of remote sensing for monitoring inland wetland habitat, and the characterization of agricultural landscapes. He has conducted several workshops for governmental agencies focusing upon remote sensing and GIS in Africa, Thailand, China, and the United States.

Minhe Ji
Minhe Ji is a doctoral student in the Department of Geography at the University of South Carolina. He graduated from Wuhan Technical University of Surveying and Mapping (China) in 1977 and received an M.S. in Geography from Northern Illinois University in 1989, and was employed as a Research Associate in the Center for Governmental Studies. His research interests include digital image analysis for land-cover characterization, GIS database development for socio-economic and environmental policies, and analytic and computer cartography.