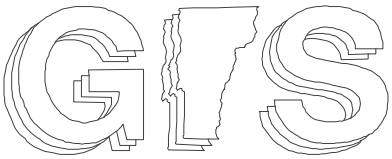

Final Report: Development of Land Cover/Land Use Geographic Information System Data Layer for the Lake Champlain Basin and Vermont

Vermont Center for Geographic Information, Inc.
206 Morrill Hall
Burlington, Vermont 05405-0106

September 30, 1997

I. Executive Summary

A new digital land cover database is now available, covering Vermont and the Lake Champlain Basin. The data was developed from a supervised classification of Landsat Thematic Mapper satellite data, enriched by GIS land use layers developed by local cooperating organizations. The principal contractor was Dr. Thomas Millette of Mount Holyoke College GeoProcessing Laboratory (GPL). The overall accuracy of the land cover classes has been estimated to be 85.9%, meeting the 85% standard proposed by the contractor.

The final land cover data has been delivered to the Vermont Center for Geographic Information (VCGI), and is available for distribution. Call VCGI at (802) 656-4277, or visit our Web site at *http://geo-vt.uvm.edu/* for information on how to obtain this data.

II. Introduction

A. *Purpose:* The purpose of the project was to create a digital database which would be useful for land use planning, land cover change detection and nonpoint source water pollution studies. For many of these purposes, a seamless data layer covering large areas is essential.

B. *Role of Mt. Holyoke College and VCGI:* Dr. Millette's laboratory performed the image classification and GIS enhancement. VCGI provided technical and administrative oversight of the contractor, and access to any Vermont GIS data layers needed by the contractor. The data points used to determine the classification accuracy were collected by a subcontractor, Robert Turner.

C. Land cover classes and codes are as follows:

		Cover Class	Code
Urban Built-up Land	I	Residential	11
	I	Commercial	12
	I	Industrial	13
	I	Other Urban	17
	I	Transportation/Utilities	14
Agricultural	I	Row Crop	211
	I	Hay/Permanent Pasture	212
	I	Orchards	22
	I	Other Agricultural/Mixed Open Grassy	24
Brush		Brush or Transitional	3
Forest	I	Deciduous	41
	I	Coniferous	42
	I	Mixed Forest	43
Water		Water	5
Wetlands	I	Forested Wetlands	61
	I	Non Forested Wetlands	62
Barren Land		Barren Land	7

Figure 1 - Land Cover Classes & Codes

III. Data Development and Integration

The land cover database was developed in a two step process that integrated data from remote sensing and GIS sources. In the first step, Landsat Thematic Mapper data was classified using the supervised approach to training and a maximum likelihood classifier. Data used for this classification include:

Scene Path/Row	Acquisition Date	Sensor
14/29	05/09/93	Thematic Mapper
14/30	05/09/93	Thematic Mapper
13/29	04/27/91	Thematic Mapper
13/30	05/02/93	Thematic Mapper

Imagery Data Sources

Land cover classes developed from this process were subsequently enhanced by integration with locally developed GIS data layers. Categories updated with GIS include the urban and built-up land, the water and wetlands, orchards and barren land. A complete listing of GIS data and metadata used for this second process can be found in Appendix B of the final report. Final integration of all thematic land cover layers was done in a raster GIS. The processing sequence of GIS Boolean overlays was as follows:

1. Landsat spectral land cover classifications
   GIS Data:
   2. Orchards
   3. Barren
   4. Surface Waters
   5. Wetlands
   6. Urban
   7. Roads/Utilities

Processing Sequence

Thus, the initial image classification was updated by a GIS layer of orchard locations, followed by barren land, surface waters, etc. The significance of the order of overlay is that where the same area is contained in several different GIS overlay classes, the last class overlaid will override other classes. For instance, a road through barren land would be classified as roads/utilities, not barren land.

It is important to understand that although the same land cover codes are used throughout the study area, different GIS layers were used to update different portions of the area. The data

collection methods and quality of these GIS data layers vary greatly. The contractor warranties the spectral classifications, the accuracy assessment data, and the overall accuracy assessment estimate, however it does not warranty any of the GIS data or metadata provided by cooperating organizations.

IV. Accuracy Assessment

The Accuracy Assessment database for the land cover was developed from GPS-based field inspections. For each site location, the classification shown in the land cover data layer was compared to the true land cover as recorded in the field. The contractor designed a data acquisition system using a GPS, digital camera, and laptop computer. Data collection was in digital form, thus eliminating possibility of error in changing the data from one format to another such as paper notes to digital tables.

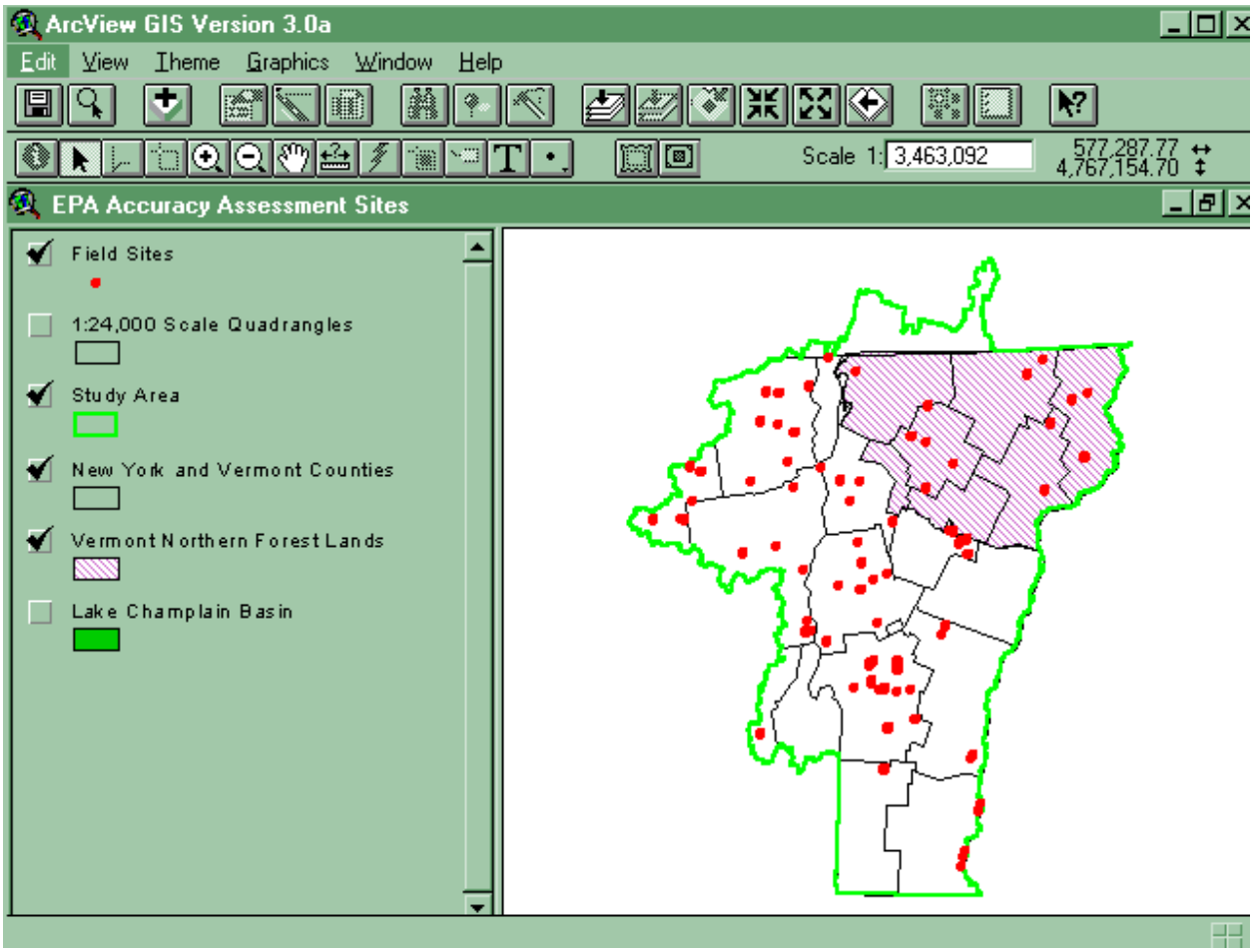

Figure 2 - Accuracy Assesment Field Sites

The sampling methodology was studied and approved. It consisted of a stratified random cluster procedure where 20 sample points were established at 50-meter intervals in a 200x225-meter site. The sites were selected from random points throughout the project area with measures included to insure adequate number of samples in each land cover class. A landowner notification system was established and approved. Most landowners gave permission to collect data on their land. In the few cases where permission was not granted, the project team selected a new random point to collect data.

A total of 1732 individual sites were visited. At each site, the land cover class was recorded and a digital photo was taken. Thirty-four sites were eliminated due to GPS receiver problems. A total of 1458 of the remaining 1698 sites were correctly classified, resulting in an overall accuracy estimate of 85.9%. A confusion matrix for the classification illustrating the errors of omission and commission is shown below. A complete listing of accuracy assessment sites can be found in Appendix A of the contractor's final report.

V. Confusion Matrix

Classified Data down the SIDE **Field Data** across TOP

	3	5	7	11	12	13	14	17	22	24	41	42	43	61	62	211	212
3	16					5	5**B**			1	1	3				3	1
5		92		5	6	1								2	2		
7			94				17**B**				5	6	1	1			1
11				119	1		58**B**										5
12		6**D**		1	81	1	36**B**										
13				1	1	83	7**B**									2	
14	1	1**D**		1		1	96				15		2			3	8
17				5			1**B**	13			2						
22							2**B**		71								
24		1**D**							4**C**	7							
41		11**D**				1	1**B**				116	1	13				
42							2**B**				2	84	4	1	1		
43		10**D**					1**B**	1		1	7	35 **A**	101	3	2		
61										1	6	4		64			
62		1**D**									1	4	8	66	37		
211										7**C**			1			98	2
212	1		2				1**B**				1	1	4	1	1	8	89
corr ect																	
sites	18	122	94	127	83	92	227	14	75	17	156	138	134	138	43	114	106

Total number of accuracy assessment sites = 1698
Total number of sites classified correctly = 1458
(1261 raw accurate sites and 197 corrections detailed in **A,B,C,D**)

Accuracy assessment estimate = 85.9%

To interpret the confusion matrix use the following steps:

1. The intersection of field class attributes (plotted on the x-axis) and the classified class attributes (plotted on the y-axis) illustrates the nominal correctly classified pixels.

2. Cells with **A** indicate cells mis-classified by the field crew as mixed forest when subsequent analysis indicated they were in fact coniferous forest [mis-classification applies to 25 of the 35 sites].

3. Cells with **B** indicate correctly classified pixels that have been overlain by the transportation/utilities GIS layer (a total of 131 pixels).

4. Cells with the **C** indicate pixels that fall into an agricultural catchall category that are considered to be correct if classified into a more detailed agricultural category (total of 11 pixels).

5. Cells with **D** indicate correctly classified pixels that have been overlain by the surface water GIS layer (total of 30 pixels).

When interpreting the accuracy assessment data, the 86% accuracy estimate should be considered conservative, and is probably an underestimate by as much as 2-3%. This possible underestimate stems from protocols used in building the GPS-based field data base and the subsequent use of the field data in the accuracy assessment. At the time of field data acquisition, in addition to collecting an attribute for land cover (*CVR_ALPHA*), a second land cover attribute (*CVR_BETA*) was collected when a site might be classified as something other than the alpha attribute. At the same time a third attribute indicating the possibility of a mixed pixel (*MIX_PXL*) was collected and noted. It was originally intended that mixed pixels (totaling 142) would be removed from the accuracy assessment population since they would be at a higher risk of mis-classification than pure pixels. It was also originally planned that beta attributes of land cover (totaling 32 pixels) could also be used as substitutes for alpha attributes when it improved the accuracy assessment rating. However, since the initial accuracy assessment of 86% exceeded the 85% standard without dropping mixed pixels and making use of any beta attributes, neither protocol was used. Were these protocols to be used, accuracy would certainly improve (perhaps as much as 3%).

The original proposal for developing the accuracy assessment database called for a minimum of 75 samples for each of the 17 information class, totaling 1275 samples. Although the project captured 1698 useable points on the ground (423 more than required), several categories were undersampled. The scrub/shrub category (code 3) has 18 samples, 57 short of the required 75. The other urban category (code 17) has 14 samples, 61 short of the required 75. The mixed agriculture/ open category (code 24) has 17 samples, 58 short of the required 75.

The undersampling in these categories is the result of the infrequency of their occurrence on the ground and the fact that seed clusters were generated with a stratified random approach. The only approach to filling these categories would be to abandon random seeding, which would have tainted the accuracy estimate. However, because all these categories are catchall classes (code 3, code 24 for agriculture and code 17 for urban), their undersampling is not considered to be a problem.

A digital database of the accuracy assessment data has been delivered to VCGI, containing:

1. GPS coordinates of field verification sites
2. Land cover attributes and field notes from field sites
3. Digital photographs of the field sites
4. Digital images of map location of field sites

An *ArcView* project has been developed to manage the accuracy assessment data and allow efficient access to it. The GPS field sites have been built into an *ARC/INFO* point coverage, with land cover codes stored as the CVR_ALPHA and CVR_BETA items. The field site photographs are managed by the *ArcView* project as BMP files, while the map images are stored as TIFF files. The entire accuracy assessment database comprises 600 megabytes of data including all images.

VI. Metadata

All metadata for the database has been provided in a format compliant with the Federal Geographic Data Committee (FGDC) standards using the *ARC/INFO* utility DOCUMENT.AML. GIS Data developed by GPL has metadata fully compliant with FGDC standards. Data provided by cooperators has been documented to the extent that was provided to GPL.

VII. Data Products

The following data products were created in this study, and are available from VCGI:

A. Land cover database in *ARC/INFO* raster (GRID) format (35 megabytes)
B. Land cover database in *ARC/INFO* vector format (1.3 gigabytes, or 3.4 gigabytes as export/.e00 files)
C. Accuracy assessment database and *ArcView* project (600 megabytes)
D. FGDC-compliant metadata for land cover database
E. Unprocessed Landsat satellite data (for Vermont public sector use only, subject to EOSAT license agreement)

VCGI may also distribute statistical tables showing percentages of land cover by

watershed, county, or other unit of interest, as warranted by user demand.

VIII. Further Information

For further information about this project, please contact VCGI:

Vermont Center for Geographic Information, Inc.
206 Morrill Hall
Burlington, Vermont 05405-0106
Telephone: (802) 656-4277 — FAX: 656-0776 — Email: *geowiz@vcgi.uvm.edu*
http://geo-vt.uvm.edu