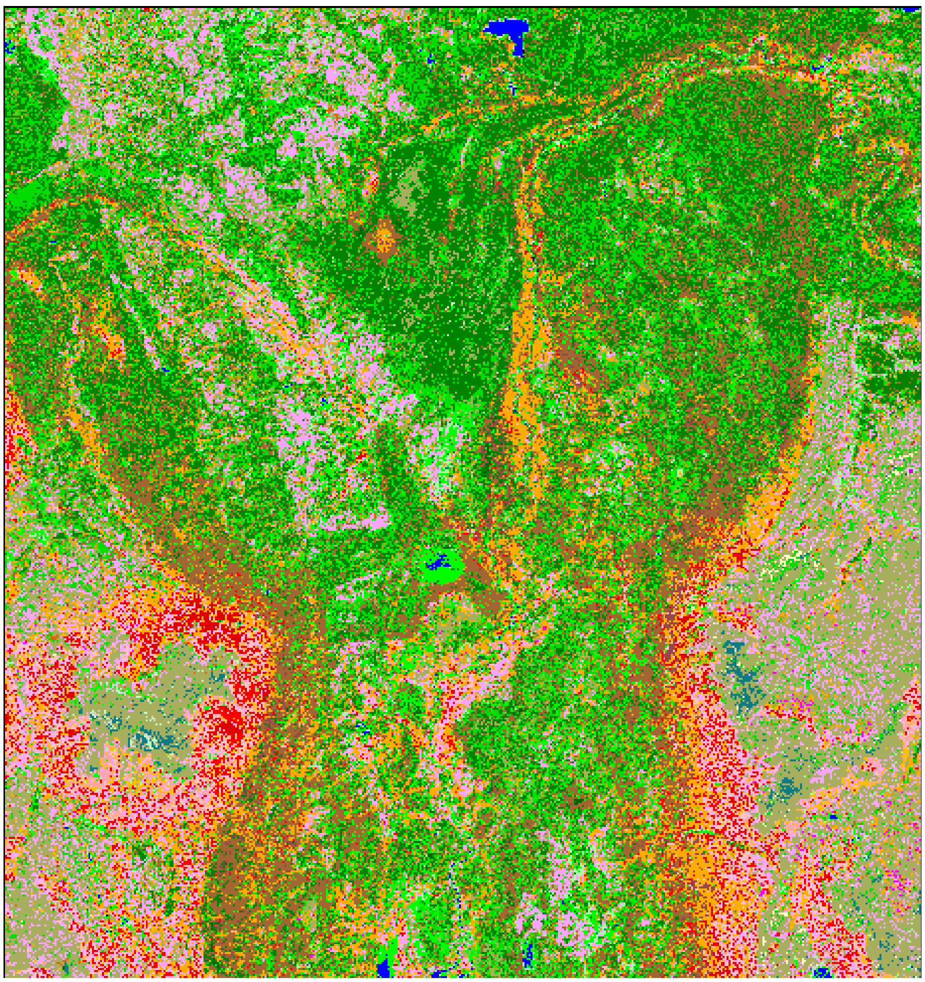
National Park Service
U.S. Department of the Interior

Natural Resource Program Center

Katmai National Park and Preserve Landcover Mapping Project
Final Report

Natural Resource Technical Report NPS/KATM/NRTR—2003/002



ON THE COVER
Subeset of Katmai classification.
Graphic by Geographic Resource Solutions.

Katmai National Park and Preserve Landcover Mapping Project Final Report

Natural Resource Technical Report NPS/KATM/NRTR—2003/002

Ken Stumpf
Geographic Resource Solutions
1125 16th Street, Suite 213
Arcata, CA 95521

June 2003

U.S. Department of the Interior
National Park Service
Natural Resource Program Center
Fort Collins, Colorado

The Natural Resource Publication series addresses natural resource topics that are of interest and applicability to a broad readership in the National Park Service and to others in the management of natural resources, including the scientific community, the public, and the NPS conservation and environmental constituencies. Manuscripts are peer-reviewed to ensure that the information is scientifically credible, technically accurate, appropriately written for the intended audience, and is designed and published in a professional manner.

The Natural Resources Technical Reports series is used to disseminate the peer-reviewed results of scientific studies in the physical, biological, and social sciences for both the advancement of science and the achievement of the National Park Service's mission. The reports provide contributors with a forum for displaying comprehensive data that are often deleted from journals because of page limitations. Current examples of such reports include the results of research that addresses natural resource management issues; natural resource inventory and monitoring activities; resource assessment reports; scientific literature reviews; and peer reviewed proceedings of technical workshops, conferences, or symposia.

Printed copies of reports in these series may be produced in a limited quantity and they are only available as long as the supply lasts. This report is also available from the Natural Resource Program Center website (http://www.nature.nps.gov/publications/NRPM) on the internet, or by sending a request to the address on the back cover.

Please cite this publication as:

Stumpf, K. 2003. Katmai National Park and Preserve landcover mapping project final report. Natural Resource Technical Report NPS/KATM/NRTR—2003/002. National Park Service, Fort Collins, Colorado.

NPS D-118, June 2003

Table of Contents

Introduction

The Katmai Land Cover Mapping Project was a National Park Service project conducted under Task Order #99003 under Contract #1443CX991097005. This Task Order's project background and objectives included:

"The Inventory and Monitoring (I&M) Program of the National Park Service (NPS) seeks … to provide reliable and consistent scientific information that may be used to assess the status and trends of national park ecosystems. Land cover mapping is conducted under the inventory portion of I&M Program with the goal of providing basic vegetation information that is useful for making resource management decisions on a Park-wide basis. I&M Program mapping is intended to provide data that will conform to NPS standards, be compatible with other I&M Program mapping efforts, and provide the information required for the design of monitoring programs within the National Park System. Resulting map data and descriptions will provide valuable baseline information about vegetation/land cover status and are also expected to provide input to resource assessment efforts, such as habitat evaluation and fire fuels modeling. The Alaska Support Field Office (AKSO) recognizes that the size and remote nature of Alaska's National Parks require that land cover mapping efforts are conducted at a smaller scale with coarser resolution than other non-Alaskan I&M Program land cover mapping efforts within the NPS system."

The Katmai National Park and Preserve (KATM) Land Cover Mapping Project was initiated in July, 2000 in an effort to upgrade and improve vegetation/land cover information for this region. Projects of this nature covering millions of acres are typically based on satellite imagery and provide generalized categorical values and class information. The AKSO also recognizes that the KATM Land Cover Mapping Project has to be compatible with similar efforts, but also wishes to determine if the KATM mapping effort could possibly form the basis of more detailed mapping efforts of the future. Land cover mapping efforts involving the use of traditional remote sensing methodologies generally produce static categorical data products that lose all detailed information at the pixel level. This loss of information has seriously limited the usefulness and accuracy of previous image classification mapping efforts, making it difficult to truly verify existing classification results or to derive new map data using reclassification logic or rule-based aggregation processes. As a direct result, such map products have limited adaptability to diverse resource management needs, which may require different, more detailed information than that delivered in a single categorical land cover map.

Seeking a more powerful and versatile product, the AKSO issued a Work Order to Geographic Resource Solutions (**GRS**) to develop a land cover classification map of Katmai National Park and Preserve. The project was to benefit from **GRS**'s land cover mapping techniques, which enable the development and maintenance of detailed original land cover data throughout map development processes and delivery of a final map data set that retains this highly detailed "bird's-eye view" level of information in the form of a land cover database. Generalized land cover maps may then be derived from this detailed land cover database. In addition, alternative more directed detailed maps may be derived from the land cover database as needed.

The KATM Land Cover Mapping Project was undertaken by **GRS**, in association with the Alaska Natural Heritage Program (AKNHP), and with support from the AKSO. **GRS** directed training site selection, navigation, and verification components of field data collection efforts, as well as all image data acquisition, preparation, training, evaluation, and classification efforts that led to the creation of the delivered land cover data sets. The AKNHP provided expertise regarding vegetation classification, land cover type description, and plant species identification during field data collection efforts, as well as input and feedback regarding post-field class descriptions and classification-related issues. AKSO personnel carried out ground vegetation transects concurrent with aerial surveys, managed field data collection logistics, and provided project coordination and review.

Project Area

Katmai National Park and Preserve is approximately 4-million acres in size, spanning the Alaska Range on the Alaskan Peninsula. KATM encompasses a diversity of vegetation and landforms in a landscape that is largely unchanged by human development. The southern and eastern portions of the Park border the coastline of Cook Inlet and Shelikof Strait with a climate moderated by the ocean. The Alaska Range parallels the coast in this region and includes a number of glaciers and peaks above 7000' elevation, as well as dramatic volcanoes. The northern and western portions of KATM are of much more moderate relief, contain many large lakes, and experience the colder climate of the Bristol Bay region.

The project area included all portions of Katmai National Park and Preserve, including a 10-mile buffer adjacent to KATM, where possible (Figure 1). Field data collection and verification was conducted within the actual KATM boundaries; however, the mapping effort was performed on the entire project area.

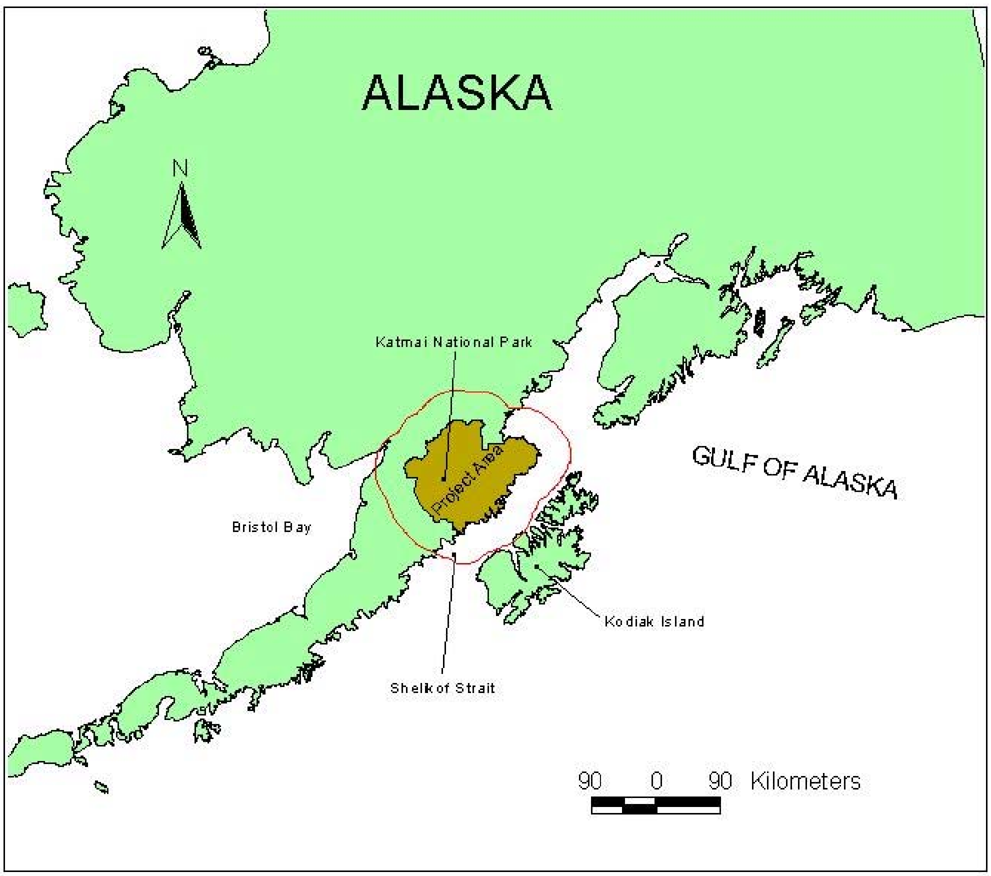


Figure 1: Katmai Land Cover Mapping Project Area

Project Methodology

The methodology **GRS** used during this mapping project employed many of the basic components of typical image processing projects. Satellite imagery of the subject area was acquired and prepared for classification; "ground truth" information describing the variety of land cover types present in the Katmai National Park & Preserve were acquired; image classification training sets representative of the different land cover components were developed and tested; supervised and unsupervised classification techniques were applied; and resulting classification pixel data sets were merged to form a detailed and accurate pixel map of the project area. However, these workflow processes were not applied in what might be considered a traditional or textbook sense during this project; the development of highly detailed and accurate pixel data, as undertaken in this project, requires a different approach to image processing. In order to provide a thorough understanding of the processes used during this project, each component of **GRS**'s land cover mapping methodology will be described. Major phases of this methodology include Data Acquisition and Preparation, Field Training Site Data Collection, Image Classification Training Set Development, Image Classification, and Land Cover Pixel Map Development.

Data Acquisition and Preparation

The initial phase of this project started in July, 2000 and involved acquiring and preprocessing of the imagery prior to the collection of any field data or "ground truth." This setup phase of the project included three significant tasks, which are a) acquisition and normalization of the satellite imagery, b) development of ecoregion areas, and c) identification of candidate image classification field training sites.

Completion of these three tasks was integral to the successful development of comprehensive and accurate training data sets and subsequent image classification efforts. Completion of these tasks greatly enhances field data collection efforts by reducing the total number of sites required for the classification effort, as well as the number of sites subsequently rejected as unsuitable due to confusion or suspect data. Reduction of field data collection needs is a significant concern in a project such as the Katmai Land Cover Mapping Project, due to the KATM project area's limited access, large size, high cost of data acquisition, and short ten-day window of field data collection activities. Data preparation is a key factor in eliminating potentially invalid, confused, and redundant training sites and developing the comprehensive and representative project training data set necessary to develop an accurate and detailed land cover map.

Image Acquisition and Normalization

A total of five Landsat scenes were processed during this project – three LANDSAT TM-5 scenes and two LANDSAT TM-7 scenes. At the start of the project, three LANDSAT TM-5 scenes were selected by AKSO and provided to **GRS** on July 12th, 2000. These three scenes were acquired by the same instrument, but at different times. Table 1 lists specific information about these three scenes.

Table 1: LANDSAT imagery available for KATM Land Cover Mapping Project.

SCENE	PATH	ROW	DATE	GMT	SUN	SUN	PIXEL
					ELEVATION	AZIMUTH	SIZE(m)
TM7219	72	19	08/30/91	2029	38.09	152.48	30.0
TM7119	71	19	09/03/95	2029	35.31	146.83	30.0
TM7018	70	18	08/21/87	2047	41.04	151.81	30.0
TM7019	70	19	08/16/00	2112	43.21	157.58	30.0
TM7020	70	20	08/16/00	2112	44.33	156.10	30.0

These three initial scenes were delivered to **GRS** in ERDAS IMAGINE format, and contained all seven spectral bands (1, 2, 3, 4, 5, 6, and 7). Each original image was translated into an Intergraph/MGE-compatible TIFF format bands for review and evaluation of actual coverage and quality. No digital elevation data were delivered with these three initial scenes.

Evaluation of the supplied imagery confirmed the need to utilize all three images in order to cover the entire KATM project area. In addition, review indicated the need to generate a cloud cover mask for scene TM7119, as significant portions of the image were effected by the presence of cloud cover. Fortunately, TM7219 had no such problem and full coverage was provided by the combination of these three scenes. Evaluation also confirmed the presence of striping in portions of scenes TM7219 and TM7018. Striping was particularly conspicuous over open water surfaces, but also appeared to extend somewhat into small land areas. De-striping of the scenes was not attempted, since it required scenes in their raw uncorrected format. As these were the only scenes that provided full coverage of KATM project area, the apparent striping was recognized as an artifact of the imagery that might cause classification problems.

An additional concern with respect to these three scenes was the differences caused by senescence of vegetation, particularly in scenes 7018 and 7119. Due to these concerns, **GRS** understood that these scenes might be replaced by the AKSO with more current imagery (in the latter stages of the project), should the AKSO and **GRS** determine acquisition of this imagery would be beneficial to the Project. Replacement of some of the original imagery did in fact occur, as scene TM7018 and most of TM7119 were replaced by two scenes of LANDSAT TM-7 imagery acquired during August, 2000. The cloud-free coverage of nearly half the KATM project area, more recent acquisition date, higher radiometric quality, and replacement of the earliest and latest acquired scenes were strong arguments for replacing the TM7018 and TM7119 scenes in this project. While initial efforts were applied to the three original scenes, the additional 2000 scenes were also processed using the same methodology to consistently preprocess the imagery prior to developing training data and classification information from this new imagery. Information describing the acquisition specifics of these two “replacement” scenes is also shown in Table 1. All project images and data layers/sets developed throughout the project were created in or re-projected to match the original imagery projection parameters shown in Table 2.

After reviewing all imagery and extracting image-specific information regarding sun elevation and azimuth angle, **GRS** proceeded to process the imagery in preparation for training site development and classification applications. This preparation involves the normalization of the images for differential illumination due to topographic influences of slope and aspect.

Table 2: Projection Parameters of the Katmai Land Cover Mapping Project.

Parameter	Value
Projection	Albers
Units	Meters
Spheroid	Clarke1866
X shift	0.0000000
Y shift	0.0000000
First Standard Parallel	55 00 0.000
Second Standard Parallel	65 00 0.000
Central Meridian	-154 0 0.000
Central Parallel	50 00 0.000
False Easting	0.00000
False Northing	0.00000

Integral to the **GRS** workflow is the use of imagery that has been normalized with respect to differential illumination resulting from the influences of slope and aspect. During past projects **GRS** has evaluated several methods of image normalization and presently uses the Backwards Radiance Transformation Correction (BRTC) based on a non-Lambertian assumption and a Minnaert constant (Civco 1989) (Colby 1991) to normalize the imagery. This technique uses estimates of slope and aspect from a co-registered digital elevation model (DEM), and image acquisition sun-angle and azimuth parameters to normalize the images for differential illumination caused by terrain relative to the sensor. This correction method minimizes the effect of aspect and slope, but maintains the signatures of the different land cover types, thereby resulting in more accurate classification information (Geographic Resource Solutions 1993). In addition to improving the consistency of the classification, **GRS** has determined that normalization often results in the reduction of the number of training sets required to classify a given land cover condition. This is a significant factor when undertaking data collection efforts in very remote, rugged, and inaccessible areas, such as Katmai National Park & Preserve.

DEMs used during the image geo-correction process were acquired after delivery of the original imagery. The original DEMs used were a series of USGS 15 minute Alaska DEMs that covered the entire project area. These were re-projected and re-sampled to 30 m pixel size using bilinear convolution to match the KATM Project (AKSO) projection parameters (Table 2) and image specifications (Table 1). Individual DEMs were subsequently consolidated into a single project area DEM, which was used to generate slope, aspect, and incidence angle data sets that corresponded to pixels in the raw imagery (Note: DEM, slope, and aspect map data used in this project for this purpose are included on the delivery CD-ROM as ArcInfo grid data sets).

GRS then used a proprietary process called ***GRS_illumcor*** to normalize each band of the imagery. Normalization coefficients were based on processing each scene, exclusive of large bodies of water, as these have been shown to negatively effect normalization results. Normalized imagery were checked relative to the raw imagery to determine overall correctness of the process. Cross-tabulation matrices were generated and checked to determine the presence/absence of 'data shadows' in the normalized bands. These would have indicated extreme and inappropriate adjustments during the normalization process that typically result for differences in the registration of the DEM data and the satellite imagery. In this project, these type data might also result from the resampling (bilinear convolution) of the original DEM data that was performed to reproject the imagery to the proper Albers projection – the resampled data might appear smoothed and less rugged than DEM data acquired at the same 30 meter resolution. These possibly smoothed data might also cause some alignment issues with the original satellite imagery. No such artifacts were found in the data during the examination of the cross tabulation reports. The normalized imagery were also examined for "hot spots", which are also indicative of areas of misalignment between the DEMs relative to the imagery. “Hot-spots” are small areas of pixels that have approached the maximum digital value (255) of the 8-bit data. They occur most frequently in areas of extreme slope facing away from the sun where lighter digital values are present. They result from misalignment of the imagery and the DEMs (it is really an illuminated area in the imagery overlaying a dark slope in the DEM data). In some cases, the DEM data is simply wrong, leading to the over-correction (brightening) of the imagery thereby forming the “hot-spot” (see Figure 2). Imagery and DEMs were processed and results were reviewed until the most suitable normalization results were obtained. While some minor “hot-spots” were found, no major misalignment problems between the DEM data and imagery were identified that would have seriously impacted the use of these data in this project.

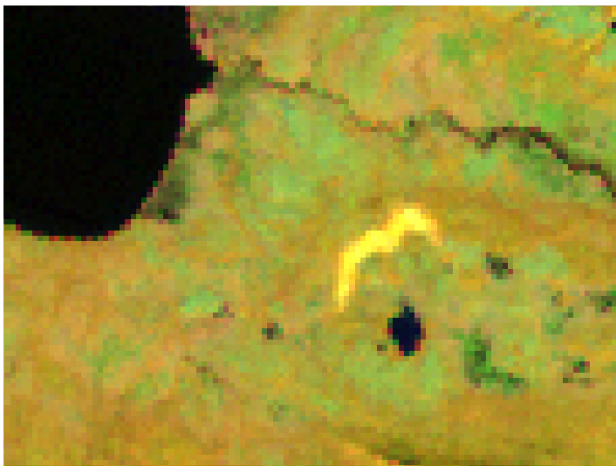


Figure 2: “Hot-spot” in the imagery

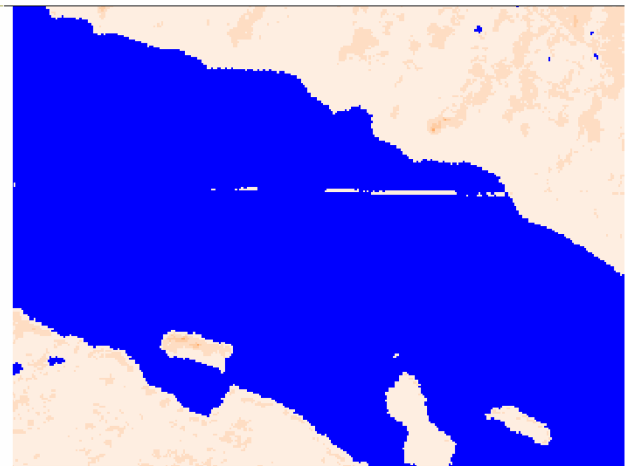


Figure 3: DEM cliff (edge mismatch) in a lake

Some elevation data mismatches were observed along the edges of DEM quadrangles after merging the individual DEMs into a project-wide DEM. This edge mismatch tended to happen along a few DEMs that lacked proper edge matching of elevations (see Figure 3). This mismatch was particularly evident in some lakes where the elevations of the lakes appeared to have been ‘burned’ into the DEMs, but with different elevations, thereby causing the mismatch (an elevation “cliff” in a lake). Misalignments resulted in areas of high slope/aspect and resulted in “hot-streaks” or narrow areas of abnormally bright pixels along seams between a very few DEM quadrangles. “Hot-streaks” varied between one and two pixels in width, and sometimes ran the entire height or width of the quadrangles involved. In some areas the terrain was so steep that even normalization could not bring any data out of these areas (terrain shadows). These types of DEM artifacts affected extremely small extents of the project area and did not affect the development of signatures for valid vegetation cover classes, since no training areas were developed in areas with these “hot” artifacts or terrain shadows. Some of these very dark pixel areas were subsequently classed as “water” due to the similarity of these areas of dark data and “water” signatures. These pixels were subsequently modeled to “terrain shadow” in the final pixel classification map.

Examples of image normalization are shown in Figures 4 and 5. These figures represent two subsets of raw and normalized imagery, along with the associated DEMs used while processing Katmai National Park. Both figures demonstrate how the normalization process removes the effects of differential illumination from the imagery. In Figure 4 the effects of normalization can be seen along the opposite aspects along both the east/west and north/south oriented ridges. The normalized image shows data of similar spectral content on both sides of the opposite slopes.

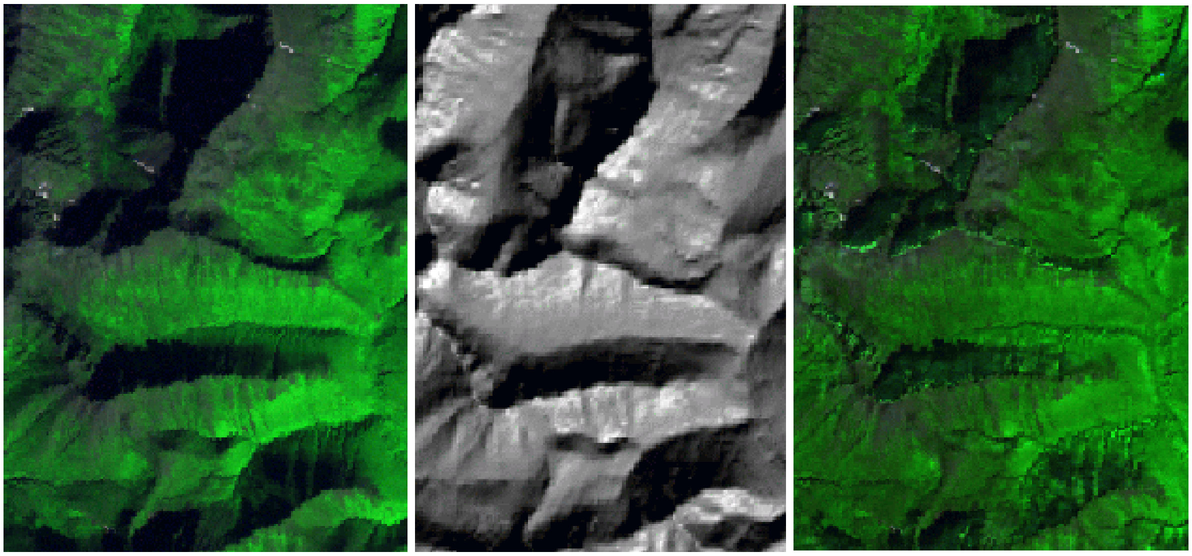


Figure 4: Illumination Normalization - Raw Image / DEM / Normalized Image

In Figure 5 this can be seen in the left portion of the image along the eastern shore of the lake. A significant dark area that appears to be a part of the lake is actually a shadow that has been normalized to show that it has spectral data. The overall impact of this process is that there is less confusion amongst different land cover types and greater agreement of spectral signatures between similar land cover types resident on different slopes and aspects. The end result is that fewer training sites are needed to develop the image training sets to classify the spectrum of land cover types.

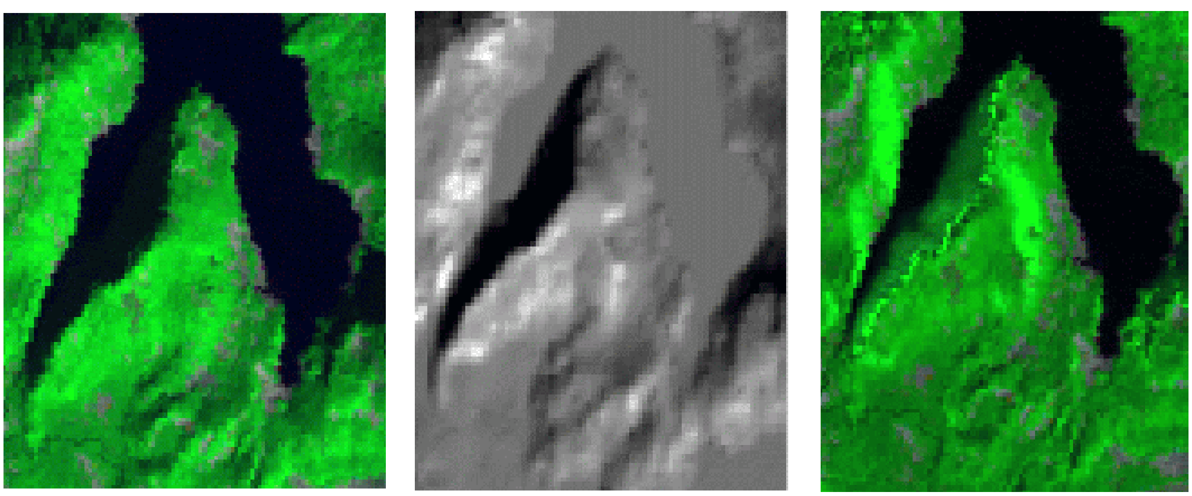


Figure 5: Illumination Normalization - Raw image / DEM / Normalized image

Development of Ecoregion Areas

A land cover mapping project of this nature requires the development of training data sets that are representative of the entire range of land cover types that are present in the project area. Ideally, all of the training sites in a project area would have distinct spectral signatures corresponding to different land cover characteristics. **GRS** stratifies the project area prior to field data collection efforts in an effort to recognize the diversity and relative magnitude of land cover types present in the project area. Unfortunately, stratification of the entire project area using an isodata classifier does not often provide for suitable representation of the many land cover types of interest. Some types, such as Coniferous Forest, may be clustered into a very few isodata classes while other classes such as Water or Barren may be represented in too many different isodata classes (size of a cluster is important in this process and small, less frequently occurring types may be clustered into a single larger type during the process). The resulting stratification may be more generalized for some, often important, land cover types and excessive for other types – and therefore may not be a representative grouping of the range of potential land cover types of interest that might be used for field data sampling. In order to develop additional strata for the limited types, an initial higher-level stratification by major ecotype regions is applied to each scene. Isodata classes are developed for each ecotype region to increase the number of land cover type strata that may subsequently be sampled and used in the land cover classification.

Ecotype boundaries for the KATM project area were provided by the AKSO and made available to **GRS** in digital format for use in this project. These ecotype boundaries represented over twenty sub-regions of the KATM project area. After consulting AKSO and AKNHP vegetation experts to determine the most appropriate use of these ecotype areas, some geographically distinct areas of similar land cover characteristics were generalized into larger ecotypes, resulting in eleven consolidated ecotype regions (see Figure 6). These generalized ecoregions were used for the initial stratification of the image data set that would be used in subsequent training data development efforts.

Identification of Candidate Image Classification Training Sites

The principal objective of this task was to develop a set of candidate training areas that would maximize field data collection efforts during the ten-day period allotted for aerial survey of field sites. This would be accomplished by eliminating the collection of data from invalid or heterogeneous spectral sites, redundant sites, and sub-minimum size sites while at the same time developing a training data set representative of the area being mapped, both in terms of the diversity of land cover types present and the geographic distribution of those types. Areas suitable for aerial survey would be identified prior to field data collection efforts to maximize efforts during the limited time available for field data collection.

Image Stratification

The creation of a pool of potential/candidate training sites relies on stratification of the project area. The project area is stratified into many different classes, where each class represents a grouping of spectrally similar pixels; through stratification, areas of spectral homogeneity that represent different land cover characteristics are identified.

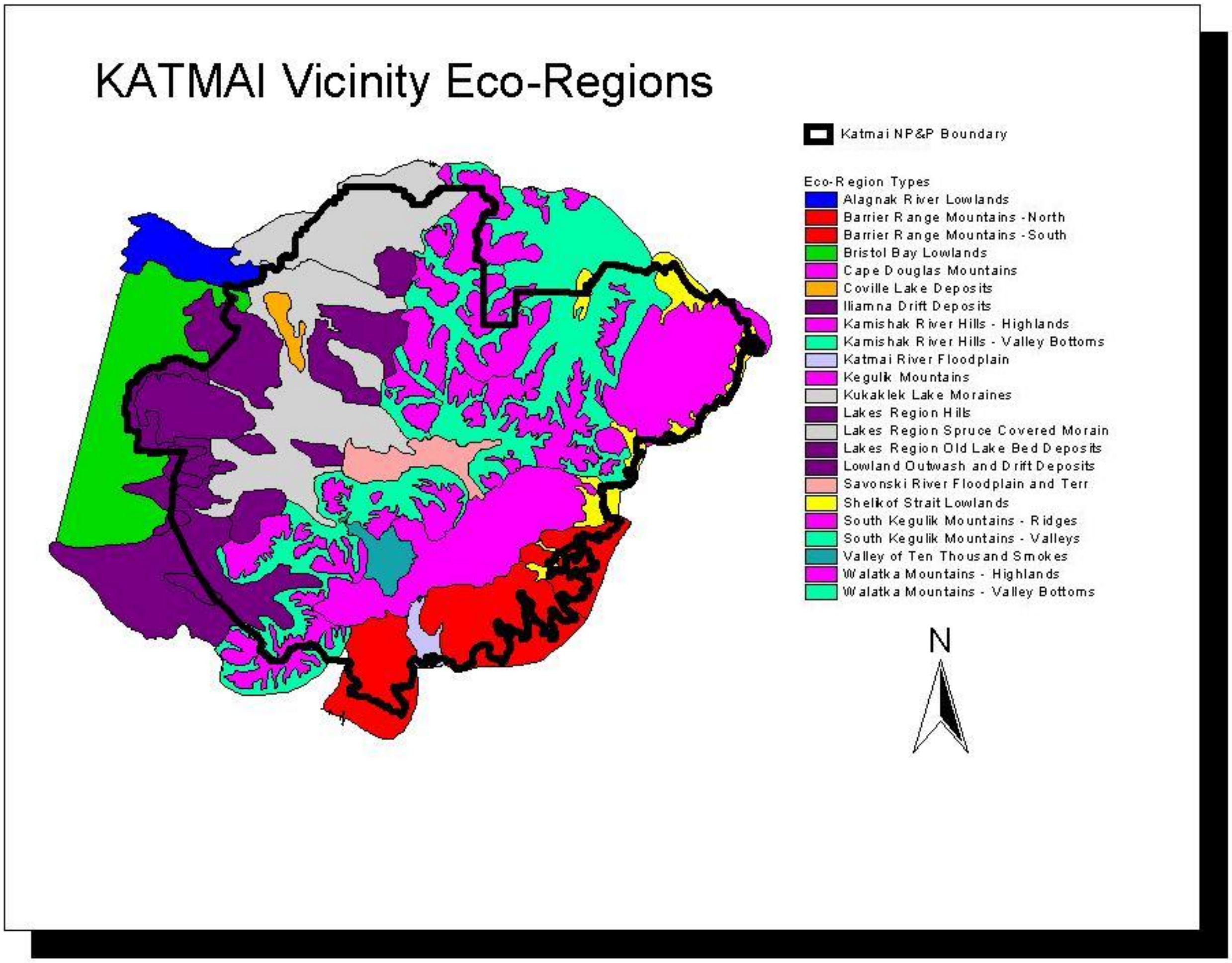


Figure 6: Katmai Eco-Region Boundaries

The goal of image stratification is to group the project area data into a large number of different classes, each representative of a somewhat unique set of land cover features that need to be sampled and represented in the final data set. Stratification yields two significant results. One result is that the diversity of the sample area is represented in the stratified image. A second result is that the frequency or magnitude of each class is estimated by the number of pixels that are assigned to each individual class. After stratification the potential diversity of the area and the relative significance, in terms of size, of each class may be determined. This information enables the development of training data representative of all the diverse land cover data present in the area, while not under sampling some classes and over sampling others.

During stratification, much of the same data is processed that will also be processed during the subsequent classification efforts. The data foundation for this effort is the imagery and ecoregion area data used to differentiate different portions of the project area. Ecotype regions are introduced to form an initial level of stratification, while the classification within the ecotype region is a second level of stratification that assures that many strata representative of the diversity of land cover types throughout the project area will be developed.
A large number of classes are developed during the isodata classification processing within each area of interest (ecotype region). Between 30-40 training classes per isodata classification are typically developed, and sometimes as many as training 50-60 training classes

are developed in each scene/ecoregion data set. If initial classification efforts result in too few or too many classes, the isodata classification statistical parameters are reset and the imagery is reclassified to develop the desired number of isodata classes. Too few classes will yield sample areas with too much variability, whereas too many classes may result in a very heterogeneous data set with areas too small to realistically identify and sample in the field.

Development of unsupervised classification data was completed for each of the three original scenes using ZI Imaging's Image Analyst software in a Bentley Systems' MicroStation environment. **GRS** used an ISODATA clustering algorithm based on a minimum Jeffries-Matusita (J-M) distance of 1.4 and an initial random seeding of 230 classes to generate unsupervised classes within the 11 ecotype regions present within each LANDSAT image. For each image, original spectral bands 1,2,3,4,5, 7, plus a NDVI band were classified. Each NDVI band was the result of the (B4 – B3)/(B4 + B3) ratio and was utilized in lieu of normalized imagery. (Note: Normal **GRS** procedures call for processing normalized imagery. Stratification works best when processing normalized imagery as normalization will typically result in fewer, more distinct classes, rather than the myriad of 'shades of gray' present when processing the non-normalized imagery. During this project, this was not possible, as there was insufficient time to finish the normalization processing prior to the initiation of the field data collection efforts. The NDVI band was included in an attempt to enable further differentiation of the imagery that might normally be accomplished through use of the normalized imagery. Normalization processes were completed and applied during the field data collection efforts in King Salmon, Alaska).

The application of ecotype region masks limited the spectral variability found within any particular region and enhanced the identification of separable classes throughout the project area. Clustering parameters used in unsupervised classifications ensured high homogeneity of resulting classes. Careful review of the J-M distance reports generated for each scene confirmed sufficient separability of the unsupervised classes.

Candidate Site Selection

The end result of the stratification process is a set of isodata class maps that contain pixels that have been stratified into the many different spectrally homogeneous classes. This map was then processed and pertinent database information was developed to guide the selection of potential candidate training site areas. Processing in grid format yielded a data set of isodata type polygons, each having a unique identifier and pixel frequency. These data were loaded into a table (isodata) for processing. A total of 8.6 million unique areas were identified. A small selection of data from the isodata table is shown in Table 3. Pixel frequency values may be used to filter areas too small to sample, as well as describe the distribution of the project area, or ecoregion area, by isodata class. The distribution of pixels by isodata class was generated by simply summing the pixel frequency by isodata value. Pixel frequency by isodata class for one of the ecoregions is shown in Table 4. This information was useful in the identification of the relative abundance of the different classes and the identification of both common and scarce isodata classes. Frequently occurring isodata classes are readily distinguished from scarce isodata classes. Scarce isodata classes were identified and targeted for sampling.

The next step in the sample area selection process was to filter out the areas thought to be too small to sample. The minimum size (number of pixels) sample site will obviously be related to the resolution of the imagery being processed. The most important aspects of this minimum size limit are that the sample site is large enough to use as a supervised image classification training site, the site can be easily located in the field by the field crew, and the site can be distinguished from the surrounding land cover types. There is no point in selecting a sample site that is too small to use as part of the supervised classification training set, or that cannot be found or identified in the field when performing an aerial survey from a helicopter. For this project, a minimum size training site area of 60 pixels, or approximately 13.0 acres was selected. Areas meeting this minimum size were readily identified by performing a simple query of the isodata class database based on the size attribute of each contiguous area. This query had the following construct:

select id, iso_class from isodata where pix_count >= 60

Several records that met these criteria are shown as **bold** type in Table 3.

Table 3: Isodata Table Data

id	iso_class	#pixels
...		
24971	16024	14
24972	16003	1
24973	16020	1
24974	16021	1
24975	16003	3
24976	**16021**	**78**
24977	16024	3
24978	16003	9
24979	16007	2
24980	16010	1
24981	16010	12
24982	16024	1
24983	16019	5
24984	16010	3
24985	16010	1
24986	**16024**	**70**
24987	16027	1
24988	16011	1
24989	16027	3
24990	16024	1
24991	16011	2
24992	16021	1
24993	16010	1
24994	16021	2
24995	16011	1
24996	16007	1
24997	16027	1
24998	16011	3
24999	16021	1
25000	16024	2
...		

Identification of areas larger in size than this minimum size reduced the number of isodata class areas from over 8.6 million areas to only 36,833 potential sample areas that were all at least 60 pixels in size. These database table rows were inserted into a new database table called *candidate_trsite* in which candidate training data sample site data were stored. This reduced set of candidate areas was checked to determine that it was still representative of all the isodata classes present in the stratified isodata data and to assure that all isodata classes were represented in the candidate_trsite table. The database check that was performed to determine that all isodata classes were represented in the candidate sample data set was performed using a database query of the following construct:

select distinct iso_class from isodata
where iso_class not in (select distinct iso_class from candidate_trsite)

This query did not identify any missing isodata classes. All iso_classes had at least one sample area that met the minimum size limit.

The next step was to determine if any isodata classes were extremely rare or too few in number, such that sampling these areas might be difficult. It was important to identify and include scarce classes in the candidate database so there would be sufficient coverage of the many land cover types present in the project area. Scarce classes were identified using the following query:

select iso_class, count(*) from candidate_trsite
group by iso_class
order by iso_class having count(*) < 5

This query identified 42 iso_classes that were considered scarce iso_classes - those having less than 5 candidate sample sites of the minimum 60 pixel size throughout the entire project area. Additional candidate areas were generated for these isodata classes by decreasing the minimum size limit from 60 pixels to 45 pixels, or approximately 10.0 acres. 305 of these additional sample areas for scarce isodata classes were added to the candidate sample site database, bringing the total number of candidate sample area sites to 37,138. Sites in scarce iso_classes were flagged in the database, to identify their presence in subsequent database query operations. Scarce sites were flagged by setting a database table column value to reflect the scarcity of the class using the following SQL statement:

Table 4: Pixel Frequency by Isodata Class

Iso Class	Frequency	Pct(%)	Cumul. Pct(%)
....			
16001	458483	14.0%	14.0%
16002	72529	2.2%	16.2%
16003	239658	7.3%	23.5%
16004	127876	3.9%	27.4%
16005	38189	1.2%	28.6%
16006	97988	3.0%	31.6%
16007	205247	6.3%	37.8%
16008	201082	6.1%	44.0%
16009	16901	0.5%	44.5%
16010	253098	7.7%	52.2%
16011	79024	2.4%	54.6%
16012	219726	6.7%	61.3%
16013	21269	0.6%	62.0%
16014	119732	3.7%	65.6%
16015	11098	0.3%	66.0%
16016	117028	3.6%	69.6%
16017	86893	2.7%	72.2%
16018	15197	0.5%	72.7%
16019	31445	1.0%	73.6%
16020	143917	4.4%	78.0%
16021	128875	3.9%	82.0%
16022	8862	0.3%	82.2%
16023	29193	0.9%	83.1%
16024	245038	7.5%	90.6%
16025	11469	0.4%	91.0%
16026	50330	1.5%	92.5%
16027	68518	2.1%	94.6%
16028	21670	0.7%	95.2%
16029	155979	4.8%	100.0%
...			
Total	3276314	100.0%	

update candidate_trsite set tr_status = 1 where iso_class in (select distinct iso_class from candidate_trsite group by iso_class having count(*) < 5)

Information regarding scarcity of iso_class sites was very useful in prioritizing sample sites when actually selecting the specific training data sample sites that would comprise the sampling plan during the field data collection sampling efforts.

The next step in this process was to move from the grid world into the vector world, in order to integrate and manipulate the candidate sample unit data in a graphic context. To accomplish this conversion process, two steps were necessary. First, the isodata class map was reclassed to form a candidate area grid map - all pixels in areas that were not candidate sample units were reset to a value of 0 (NODATA), while all other pixel values remained the same. The resulting grid map represented only those areas that had been determined to be candidate training data collection sites. A portion of the resulting grid is shown in the left side of Figure 7. This candidate area grid was then vectorized to form a vector database representing the candidate sample areas. Other data, such as isodata class labels, were developed to enhance the training site information. These area boundaries and corresponding labels could then be overlaid on the imagery, as shown in the right side of Figure 7. The resulting database represented the initial set of candidate training sites.

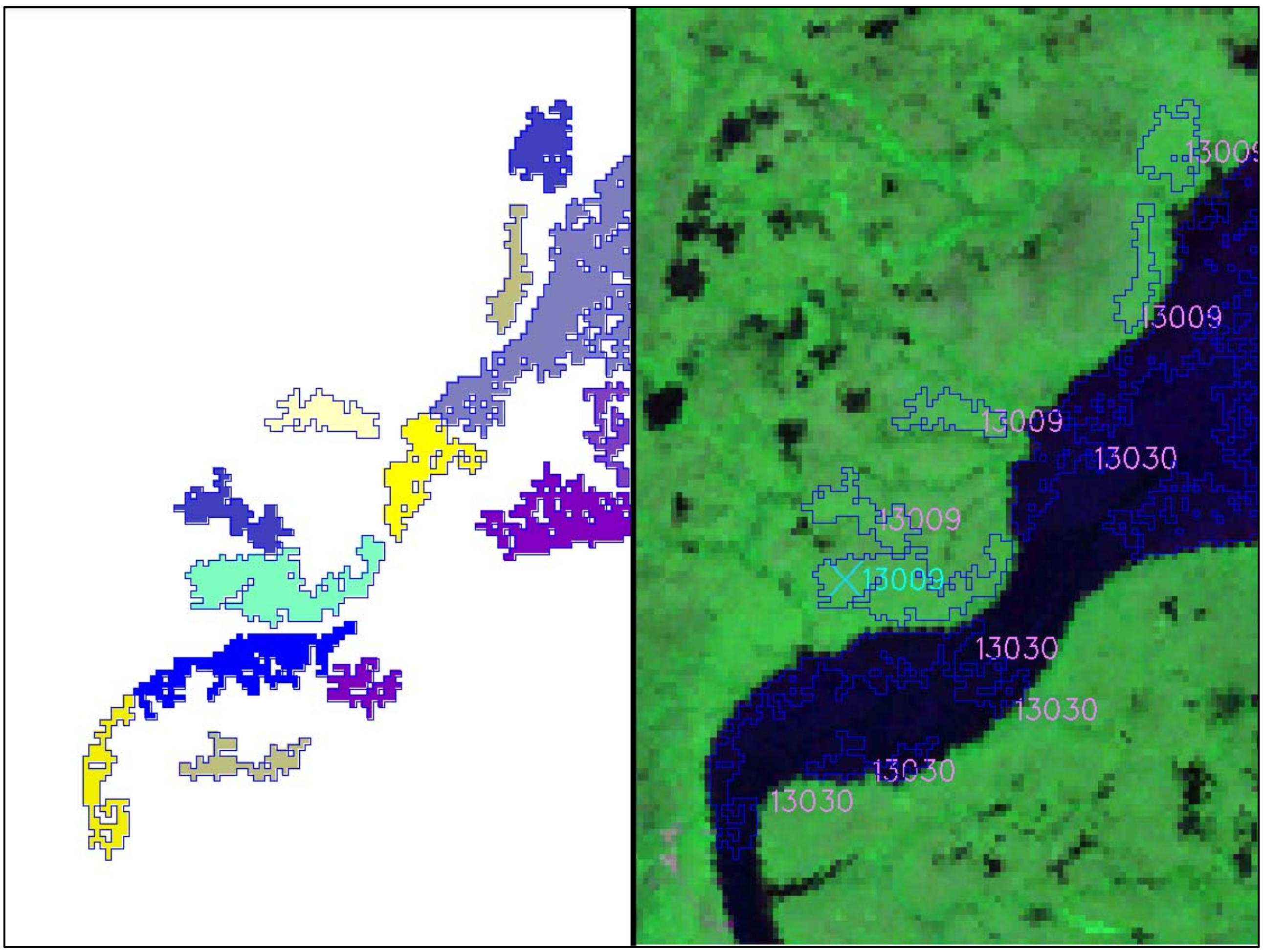


Figure 7: Candidate Training Sites

Candidate Site Field Maps and Data

A series of 1:47,500 scale (20” X 15”) color maps were developed to enable training area location and navigation in the field. Each map showed candidate training area polygons, their unique isoclass numbers, and 400’ elevation contour lines, over a 4,5,3 RGB composite built from LANDSAT imagery. A total of forty field maps were generated and labeled using a row-column code with origin on the upper left corner of the project area grid. Two copies of each map were produced, one for the aerial survey crew and the other for the ground survey crew. All maps were laminated for protection and to enable flight planning and notations directly on the map with erasable color pens. A portion of one of these field maps is shown in Figure 8.

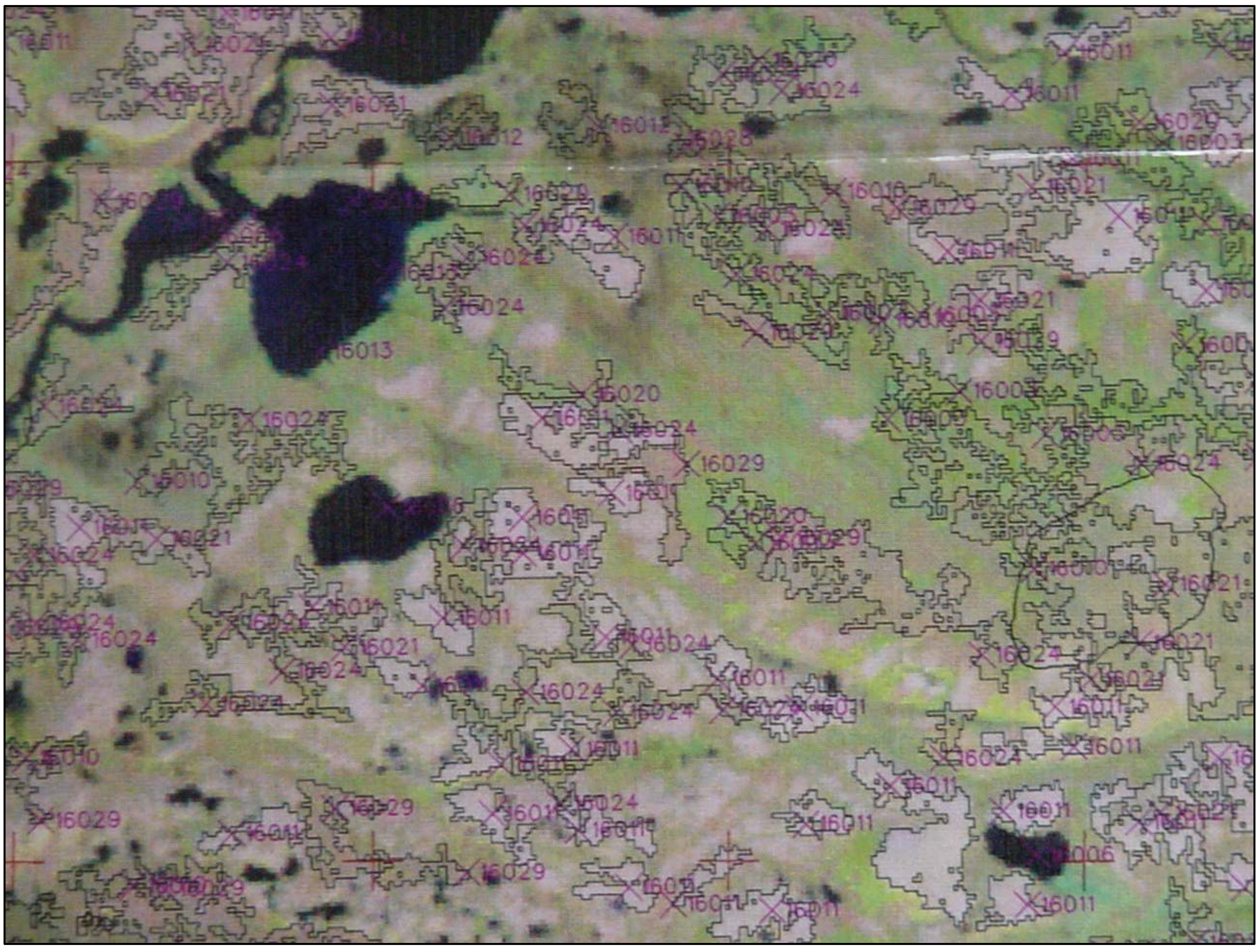


Figure 8: Field Map with Candidate Training Sites

Efforts were made to pre-determine potential training site locations that covered the spectrum of significant land cover types within the project area. However, all unsupervised areas meeting homogeneity and size requirements were kept as part of the pool of candidate training sites, in order to enable changes in the field (i.e., opportunistic sampling) due to weather or field conditions. Additional training site characteristics were also generated; overlaying the candidate training sites with the slope, aspect, and elevation themes enabled the development of site specific estimates of these characteristics for each candidate site. X,Y coordinate values representing each training site were also loaded into the database table. These training area coordinates were used for navigation and positional confirmation using GPS receivers in the field. Location of training areas in this manner allowed for careful flight planning and more efficient and safer access to field plots. The pool of candidate sites was completed and mapped.

Field Sampling Restrictions

Aerial sampling from a Bell-Jet Ranger IV helicopter was used to maximize the number and diversity of sampled training sites. This approach would yield the land cover components as viewed from above (“bird’s-eye view”) and enabled rapid (and safer) on-demand access to target training areas. However, constraints to sampling efforts existed due to fuel supply and d availability, aircraft range, crew-ferrying time, no-flight zones, and weather conditions. No training areas were targeted in the AKSO-defined no-flight zones shown in Figure 9. In addition, training sites where identified and targeted in areas of image overlap, as these sites could be used for training in all overlapping scenes. Large groups of candidate training site areas incorporating as many ecoregions and isodata classes as possible were identified for sampling.

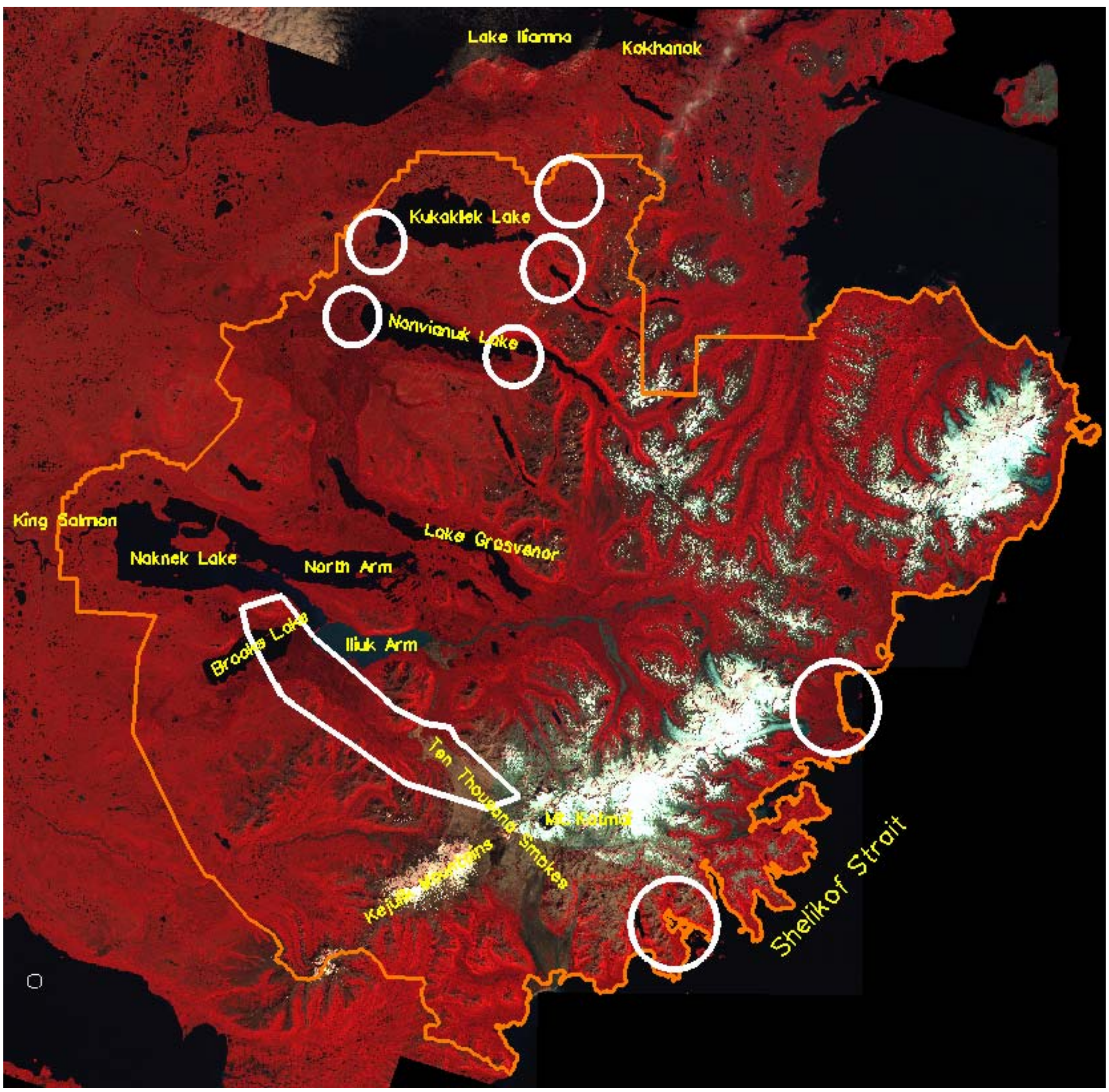


Figure 9: AKSO-designated no-flight zones over Katmai project area

Aware of these constraints, and at the request of the AKSO, efforts were made to maximize field data collection efficiency by reducing travel time between training collection sites and the number of collection sites required to develop comprehensive supervised classification training data. Thus, potential training sites were organized as groups of sites, as much as possible, to minimize distance traveled between sample areas, while at the same time facilitate access by both ground and helicopter crews and adequately sample the geographic diversity of the Katmai Project Area. Sample areas that fulfilled the sampling constraints, as well as the apparent project training data needs were identified. These proposed sample areas are shown in Figure 10.

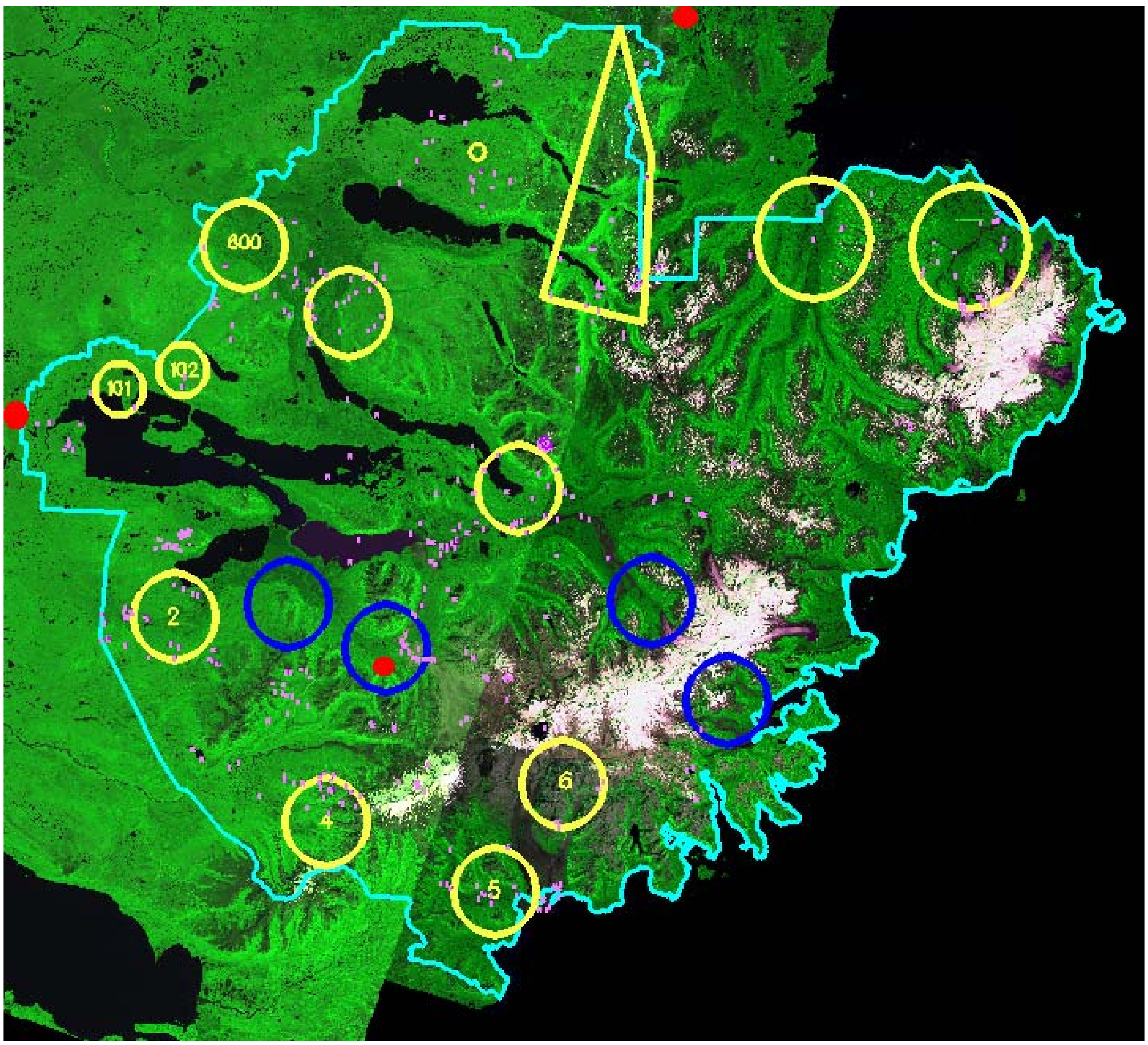


Figure 10: Sampling Areas-of-Interest in the Katmai Project Area

Field Training Site Data Collection

Field data or “ground truth” information is one of the critical components of a land cover mapping effort like this project. The field data provide the land cover or “bird’s-eye view” descriptions that will both guide the classification efforts as well as describe the resulting land cover features in the final classification map. Accurate and detailed field data are essential for the successful completion of such an effort. Any misinformation collected during this stage and later applied during classification efforts will be embedded in the final map.

Field Data Collection Goals

Image classification training data sets must encompass the complete range of significant land cover characteristics present in the project area. To this end, **GRS** worked with AKSO staff and AKNHP botanical and ecological experts to identify a matrix of the land cover characteristics and conditions thought to exist in the project area. This matrix represented the variety of land cover characteristics, including species/cover components and vegetation density classes. The land cover class matrix was composed of a combination of the target (known) land cover classification system, as well as other recognized (potential) land cover types that might also be found within the project area. This land cover type matrix formed the basis of the land cover classification system that would be applied during the mapping project. This matrix also formed the basis for field data collection efforts, as classes that would ultimately be recognized and mapped would require training site locations associated with the many different land cover attribute descriptions (“ground truth”). The range of possible land cover types must be adequately sampled within the subject imagery to develop suitable classification training sets that will yield accurate and reliable image classification pixel maps. The basic matrix/listing of land cover types used to guide field data collection efforts is shown in Table 5.

At each field data collection site information would be developed that would describe the cover characteristics of the site. This information amounted to a “bird’s-eye view” from above of the land cover components that summed to a total of 100 percent cover. A field form was developed by AKSO, **GRS**, and AKNHP staff to record this information. A copy of one of these field forms is shown in Appendix A. The cover characteristics recorded at each site amount to a “bird’s-eye view” of the site, such that the site would be describes by cover characteristics that would sum to 100% cover. Additional cover (understory or overtopped) would also be noted as part of the information. In additional any specific characteristics regarding moisture regime and environment would also be recorded. Lastly, a land cover type would be assigned based on the cover composition of the site relative to the land cover classification rules being used in this project.

The land cover classification system used for KATM was a modification of the Alaska vegetation classification developed by Viereck (Viereck et al 1992), and was designed by AKNHP vegetation experts in conjunction with AKSO. The system defined cover classes in terms of vegetation communities named after dominant species existing in principal layers. Schematic descriptions of included cover classes can be reviewed in the diagram in Appendix B. A snippet of the C code used to estimate the cover class is also included in Appendix B.

Table 5: Data Source Summary: Frequency by Type and Source

Type	Aerial Sampled	Aerial Used	TSmith Used	AKSO Used	Suppl. Used	Total Used
Spruce:Woodland	12	12	0	0	1	13
Spruce:Open	10	9	3	0	2	14
Spruce:Closed	1	1	0	0	1	2
Spruce Stunted:Woodland	0	0	0	0	1	1
Spruce Stunted:Open	0	0	1	0	0	1
Spruce/Deciduous:Woodland	2	2	0	0	0	2
Spruce/Deciduous:Open	7	7	0	1	1	9
Spruce/Deciduous:Closed	2	2	0	0	0	2
Balsam Poplar:Woodland	1	1	3	0	0	4
Balsam Poplar:Open	9	7	3	2	0	12
Balsam Poplar:Closed	1	1	1	0	0	2
Birch:Woodland	5	4	2	0	0	6
Birch:Open	5	5	0	0	0	5
Birch:Closed	6	6	2	0	0	8
Deciduous:Woodland	0	0	1	0	0	1
Deciduous:Open	2	2	1	0	0	3
Deciduous:Closed	3	3	1	0	0	4
Alder-Willow:Tall:Open	2	1	0	1	1	3
Alder:Tall:Open	9	9	0	1	1	11
Birch-Willow:Tall:Open	1	1	0	0	0	1
Willow:Tall:Open	5	5	3	3	0	11
Alder-Willow:Tall:Closed	4	4	0	0	2	6
Alder:Tall:Closed	18	18	5	2	5	30
Mixed Shrub:Tall:Closed	0	0	0	0	1	1
Willow:Tall:Closed	2	2	4	2	0	8
Alder-Willow:Low:Open	0	0	0	0	0	0
Birch-Willow:Low:Open	1	1	0	0	0	1
Mixed Shrub:Low:Open	7	7	10	1	1	19
Willow:Low:Open	18	16	0	0	5	21
Mixed Shrub:Low:Closed	1	1	2	0	1	4
Dwarf Shrub:Mixed	54	50	9	3	6	68
Forb	0	0	0	1	1	2
Graminoid	0	0	0	1	0	1
Graminoid:Dry/Mesic	4	4	7	1	3	15
Herbaceous:Marsh	13	12	11	1	4	28
Lichen	2	2	0	0	0	2
Moss/Lichen	1	1	1	0	0	2
Sparse Vegetation	5	5	0	0	13	18
Barren	9	9	0	0	5	14
Snow/Ice	0	0	0	0	2	2
Terrain Shadow	0	0	0	0	0	0
Water	0	0	0	0	3	3
Totals	**222**	**210**	**70**	**20**	**60**	**360**

Field Data Collection Operations

Field data collection efforts were scheduled to occur during the first two weeks of August, 2000. Data collection efforts were initiated from King Salmon, AK, located approximately 5 miles west of the northwesterly boundary of the Park. The collection of field data was a joint effort by personnel from AKSO, **GRS**, and the AKNHP. All field data efforts were coordinated with AKSO personnel and guided by the project data collection needs, as identified by **GRS**, AKNHP, and AKSO. **GRS** was responsible for the development of daily aerial field survey data collection plans that included the selection of potential candidate field training sites and the information necessary to locate these locations. The AKNHP was responsible for the ecological characterization of the different land cover types present in the KATM project area. This included the development of detailed land cover descriptions for the field training sites actually visited during aerial surveys. The AKSO was responsible for planning and coordinating all field data collection logistics over the two-week field data collection effort.

GRS staff flew in and set up a computer system in the project's headquarters in the Katmai National Park & Preserve offices at King Salmon. This system contained all project data, as well as GIS and Image Processing software. **GRS** personnel used this system to develop daily field survey plans and schedules, query and review candidate training site locations for alternative plans, and monitor data collection progress relative to project land cover information needs.

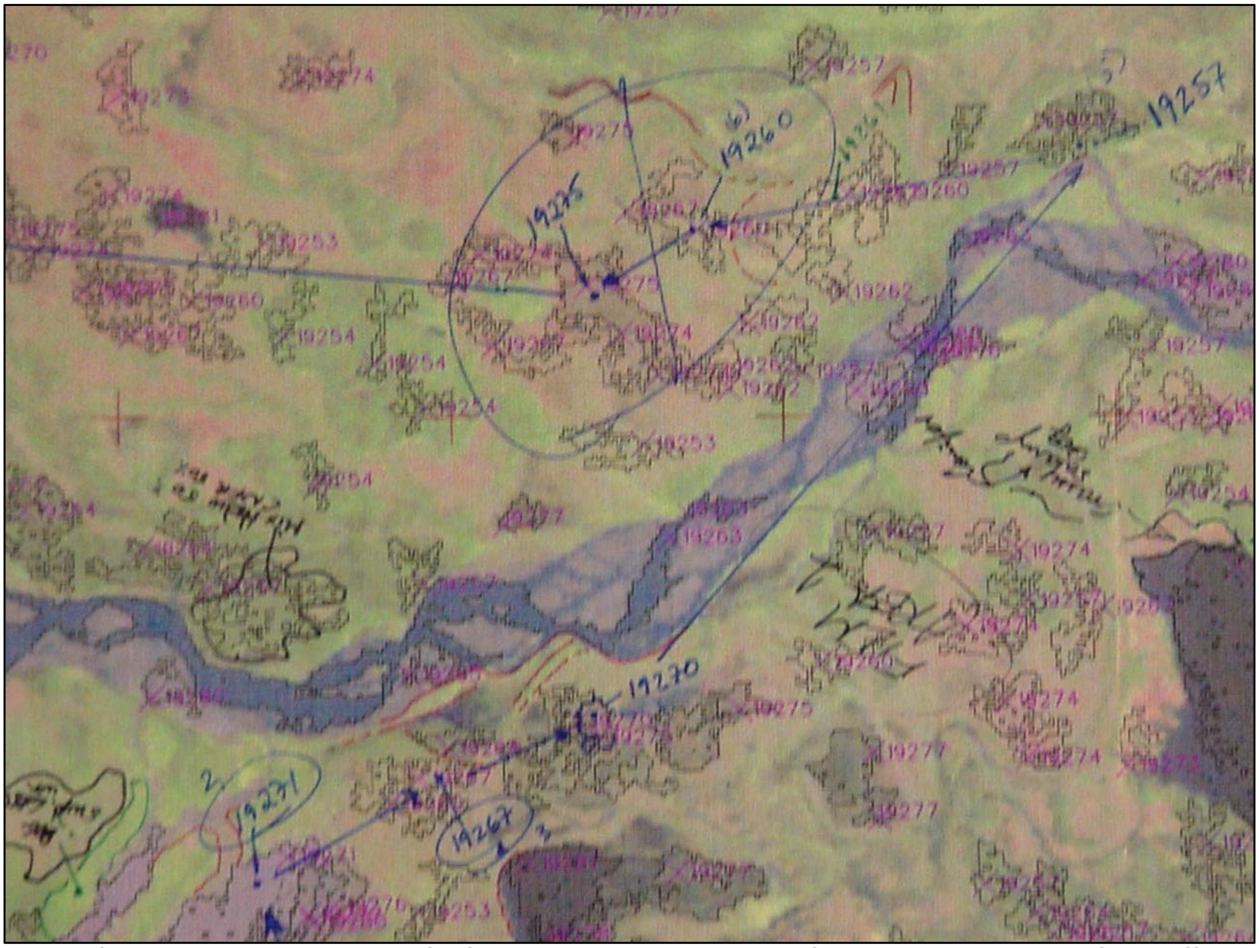
Field crews arrived at King Salmon on August 1st, 2000 to setup the project headquarters. Ground reconnaissance, as well as the development of proposed flight plans, started on August 2nd, in preparation for the start of data collection efforts on August 3rd. Bad weather forced cancellation of fieldwork on the 3rd and this time was directed towards additional planning of field data collection and discussion of data collection goals and objectives. After the weather improved, daily aerial surveys based from King Salmon were accomplished from August 4th through 8th. On August 9th data collection operations were moved from the project's headquarters in King Salmon to Kakhonak, 88 miles to the northeast along the south shore of Lake Illiamna. Daily aerial surveys of the northern and northeastern portions of the KATM project area were initiated from Kakhonak on August 10th through 12th. Field sampling concluded on August 12th and field crews returned to the Anchorage area the evening of the 12th. In all, field data was collected on eight days during the nearly two week field data collection efforts. During those days, approximately 32 hours of flight time or approximately 4 hours/day of actual aerial survey time were logged surveying aerial sites. Much of the remaining time was spent transporting field data collection crews to and from the field and refueling the helicopter.

Aerial Survey Logistics

GRS staff generated a daily flight plan consisting of a list of scheduled training sites, reviewed and selected the day before, to fulfill data collection needs. This flight plan guided the aerial survey data collection efforts using the Bell helicopter. This tabular report formed the basis of the field sampling form that included the sampling area number (a regional number assigned to different areas of the project area), candidate training site *iso_*class number, positional coordinates (lat/longs), aspect, slope, and elevation of the sample site locations. A unique training site identification number was assigned to each aerial survey site as it was visited. This *trsite_id* was composed of concatenating the date with a sequential visit numbers for each location. For example, training area code "081022" would designate the 22nd area visited on August 10th. Training area IDs were generated by **GRS** personnel for each training area visited. This report was ultimately generated using the GIS by performing a formatted query of the

candidate sites sequenced for any particular day. Field maps, imagery, and aerial photos were reviewed and sampling sequences were developed to minimize travel time.
Field maps were annotated with additional information as an aid in navigation to the selected areas. A tentative sampling sequence was delineated on appropriate field maps indicating selected flight paths, and used as a flight plan for the aerial survey efforts. Figure 11 shows a portion of one of these annotated field maps.

Figure 11: Annotated Field Map with Selected Training Sites



Actual field locations, based on GPS receivers, were planned for collection as part of this effort. The original work plan called for the use of a Precision Lightweight GPS Receiver (PLGR), to locate and “mark” surveyed training area locations from the air. However, aerial survey crews were quickly forced to rely on the helicopter’s installed Trimble GPS navigation unit instead, when the handheld instrument failed to function properly while in flight. This equipment failure necessitated a daily transformation of coordinate values from the project’s projection parameters, to the aircraft’s GPS navigation parameters, which differed in projection units and datum. **GRS** automated this process in order to generate reliable navigational coordinate positions for daily sampling schedule reports. The reverse process was also automated in order to facilitate the input of training site field coordinates in the project’s database and maps. Unfortunately, this unexpected development created uncertainty regarding the true location of some surveyed sites, in particular those sampled the first two days of aerial survey efforts. Ground crews were able to successfully use the PLGR units to locate their field plots.

GRS made efforts to adhere and exceed daily aerial survey schedules. However, weather conditions and other restrictions did not always allow for complete sampling of the scheduled areas or sites. Opportunistic sampling was implemented whenever feasible in order to supplement scheduled sampling efforts and prevent shortage of sampled areas due to changing weather conditions. Opportunistic sampling was implemented in two ways. The first way entailed locating alternative *iso_class* training areas printed on field maps that were spectrally equivalent to and nearby scheduled areas (i.e., same *iso_class* code) that could not be sampled. The second type of opportunistic sampling involved spotting and sampling of promising areas while en route to other locations. These areas were not necessarily plotted on the field maps and may not have had assigned *iso_class* codes. However, they were visually estimated to be large and appeared homogeneous enough to qualify as valid sampling sites. These sites also tended to be land cover types known to be ‘rare’ or missing in the land cover matrix.

The aerial survey crew consisted of the helicopter pilot, one **GRS** image processing specialist and one AKNHP botanical specialist. The **GRS** specialist’s responsibilities included navigating to and locating scheduled training sites using coordinates listed in the daily sample schedule form and on the field maps. Target coordinates were read to the pilot, who would enter them into the aircraft’s navigation system and navigate to the desired training site location. As the helicopter approached a target location, the **GRS** specialist used field maps to confirm arrival to the target site. Once arrival to a target sample site was visually confirmed and its extent described to the entire field crew, the pilot began to slowly circle the target site and the AKNHP botanist proceeded with the ecological characterization of the land cover present within the training site. Meanwhile, the **GRS** specialist photographed and videotaped the site, recorded coordinate readouts from the aircraft’s navigation console, and recorded pertinent comments. Once the AKNHP botanist and **GRS** specialist declared completion of their tasks, the **GRS** specialist provided navigational information and the pilot began flying to the next training site.

Each day after aerial survey site data were collected, the data were processed and the number of sites that had been sampled were tallied in the land cover matrix thereby indicating the frequency of samples by land cover type. This information was used to identify ‘scarce’ types still lacking samples as opposed to common types with many samples. Subsequent data collection efforts were adjusted daily in an effort to completely cover/fill the land cover type matrix. In addition, remaining unsampled types were identified so that they could be sampled as opportunistic training sites, if the survey crew found suitable sites of these types.

<u>Aerial Survey Results</u>

An average sampling day consisted of approximately four hours for actual training area aerial surveying. The total of 222 training sites were described by the aerial survey crew, while 86 ground sites were described by the AKSO ground crew. An average of 27 sites were surveyed from the air per sampling day. At the end of each sampling day, each **GRS** image-processing specialist imported coordinate values for sampled training sites into the project database and maps. Each sampled site field-recorded position was converted into a point feature attributed with its unique training site ID. This ID was later used to relate each point to its respective land cover description data. All recorded digital imagery was also saved to disk. These efforts resulted in plotted flight paths, where each aerial site was a waypoint on the flight path. Locations were reviewed relative to the candidate site locations and the imagery, and adjusted for any inaccuracies in the GPS locations collected from the helicopter’s GPS unit. Unfortunately, this device was no better than ± 100-200 meters, often indicating significant differences from sample areas and indicated ground position. Being able to review the data as it was collected was a significant advantage as positions were adjusted, when possible, to conform to known positions relative to indicated positions.

As data collection efforts progressed, field data were reviewed on a daily basis to assess the variability and number of land cover classes visited to-date. These reviews enabled **GRS** personnel to guide data collection efforts to include classes that appeared under-represented in the current sample set. A summary of the number of sites surveyed from the air by land cover type during the August, 2000 aerial survey is shown in Table 5. These data collection sites and classes were distributed over the entire project extent visited, spanning all three LANDSAT scenes. Those types shaded in light green were not sampled during the aerial survey and data was estimated from one of the other available sources.

Cover data for surveyed sites were imported into the project database tables and made available for database query and report tools used in the subsequent resolution of class confusion, assessment of training class performance, and characterization of each site's land cover attributes. SQL queries were performed to check for invalid data, such as sites for which the total "bird's-eye" view cover did not equal 100 percent. SQL statements were also used to generate the modified Viereck land cover types, to check for incorrect or inconsistent field calls. Data inconsistencies were identified and corrected. Data were summarized using ***GRS_transum_tr*** to develop estimates of cover by individual specie, as well as groups of species, and were loaded into the relational database tables as attributes of each site. In addition, the predominant cover component and percent cover were also loaded into the database tables.

Supplemental Data Sources

In addition to the aerial survey data, additional data sources were also used during this project. These sources included AKSO ground survey site data, Dr. Thomas Smith plot data, and photo-interpreted data. Land cover data for all three of these sources were processed to develop the 100% "bird's-eye" view land cover descriptions. These data were also loaded into the relational database tables and processed to develop information comparable to the information developed from the aerial survey sites. Data inconsistencies were identified by **GRS** and corrected by AKSO and AKNHP as best as possible to develop usable training data. Summary data were generated and loaded into the database. Locations were input into the GIS and all sites were spatially referenced to the project database, in the event that they might be used as supplemental training sites. Areas of questionable location, as identified by AKSO, were identified and withheld from use as potential supplemental training sites.

In addition, supplemental data were provided from an AKSO fly-over of portions of the Katmai project area during August, 2001. This fly-over was designed to provide additional information for supplemental sites identified by **GRS**. Digital photography and verbal descriptions were acquired in an effort to gather additional information about these few questionable areas. These data did not have "bird's-eye" view cover descriptions, but rather general type descriptions. Cover descriptions were developed for a few of these sites (those that were used as training areas), but only to a very general level based on photo-interpretative efforts performed in the office.

Image Classification Training Set Development

GRS uses both supervised and unsupervised classification to develop land cover pixel classification maps. Both these methods are dependent on training sets. Training set development is a critical component of a classification effort. "Ground truth" or land cover attribute data must be properly correlated to spectral data to develop accurate and detailed land cover map data. If the attribute descriptions are not properly associated with the spectral data, then the resulting land cover map will be an inaccurate representation of the project area. Care must be exercised to properly train the spectral data with the "ground truth."

The goal of this phase of the project is to build image training data sets that most accurately reflect the land cover types present in the project area, while at the same time minimizing confusion and uncertainty within the training data. Training areas should reflect the areas described during data collection efforts and have small statistical variation within the spectral data of each training site. Training areas of this sort tend to minimize overlap of spectral statistics and confusion between training data. Proper spatial registration, as well as the appropriate size, are requirements of accurate training data. As training sites are developed they are reviewed to determine their validity. Reviews include evaluations of confusion and fidelity/self-classification. Valid sites are used in classification efforts and invalid sites are withdrawn and discarded. A key component in **GRS**' supervised methodology is that an individual training class is developed for each survey site, rather than for groups of sites that are thought to represent the range of a training class. This different approach tends to keep spectral statistical ranges small relative to the statistics for classes formed from groups of sites and results in reduced statistical overlap and confusion of data for data of different land cover characteristics.

Supervised Training Set Development Strategy

The first sites used to develop image training data were the aerial sites that were surveyed during the August, 2000 field data collection effort. These were the sites that were selected for sampling based on their spectral homogeneity and size. Several obvious land cover classes such as water, snow/ice, and barren, were developed from training areas selected with the aid of aerial photography and satellite imagery. Where possible, the same obvious land cover training sites were used in each scene in the areas of scene overlap, to minimize the number of training sites needed to classify the imagery.

After review of the resulting frequency of sites by class (Table 5), certain land cover types still were lacking representative training sites. For these missing or under-represented land cover types, additional sites were added using data from the alternative sources of training site information. Training sites were first added from those areas visited by AKSO ground crews, as these sites were comprised of data most compatible with the aerial sites in terms of the data collection time period and descriptive information. Additional sites, if still needed, were then added from the Dr. Thomas Smith survey sites. Lastly, after review of all these additional sites, supplemental sites were added from photo-interpreted areas to finalize the classification training data sets.

Supervised Training Site Development Techniques

Training areas were developed in sets that would be applied to the specific images (scene) during classification efforts. Three sets were initially developed, one for each original scene – 7219, 7119, and 7018. As scenes were added during the course of the project, additional training sets were developed for these newly acquired data (TM7). Sets of training areas were initially developed scene by scene attempting to sample the same areas within overlapping scenes. Each scene specific training set was later copied and modified (deletions and additions) to represent specific classification efforts (e.g. “spruce” versus “no spruce”, “water” versus “no water”, and so forth) for that specific scene. Training areas within each training set were created individually using region growing techniques, a neighborhood seeding method that used a 3x3 pixel initial seed window surrounding a point interactively selected on the image to grow a homogeneous spectral area around the seed window. Location of the initial point in each scene was determined by the approximate coordinate location of the surveyed site, as indicated by a labeled point in the GIS survey site coverage overlaid on the imagery. Training area pixel inclusion was determined specific to each scene using a statistical threshold of 2 standard deviations from the seed area mean spectral data. The maximum search area extent was limited to a 41x41 pixel area.

These parameters were the basic parameters applied while developing training sites at all training site locations. However, if a sufficiently large enough area could not be grown using these exact parameters, then the standards were relaxed as follows:

Option 1: The search area was expanded to 81x81 pixels. Initial seeding points were moved slightly, but were still quite close to the location of the training site point, so that the sample area represented the surveyed area. If this approach did not produce a sufficiently large enough training area, then option 2 was applied. Option 1 was the most desirable, since it would still keep training area spectral statistics as tight as possible, thereby reducing potential confusion during classification.

Option 2: The seed area would be enlarged to an initial 5x5-pixel window. If this approach did not produce a satisfactory training area, then option 3 was applied.

Option 3: The seed area would be set back to an initial 3x3-pixel window and the pixel inclusion parameters would be relaxed to a threshold of 2.5 standard deviations. If this approach did not produce a satisfactory training area, then option 4 was applied.

Option 4: The seed area would be set back to an initial 3x3-pixel window and the pixel inclusion parameters would be relaxed to a threshold of 3.0 standard deviations. This last choice usually generated areas of sufficient size if the prior options did not.

Although the ideal training area size was thought to be greater than 60 pixels, a minimum of 15 pixels was used in an effort to develop very tight spectral statistics. Sites this small most commonly occurred when training sites were developed using sites that were not pre-qualified as part of the candidate training site selection process. These smaller training areas were included in the initial classification in order to discard as few training sites as possible.

<u>Supervised Training Set Development</u>

Each of the aerial survey sites visited during the 2000 field season was included in the initial image classification training sets developed for this project. As land cover characteristics for these sites were evaluated and tabulated, missing or underrepresented types were then added to the training data sets. The following sources of additional training site data were reviewed and used when possible.

The second source of training sites were those sites visited by the AKSO ground crew during the same 2000 field season. Although the cover for these sites was verified on the ground, there was no assurance that every site would exhibit the necessary spectral homogeneity and extent required for a training site. In some cases, it was possible to verify suitability for inclusion of these sites as training classes from digital photographs taken by the AKSO ground crew.

The third source for training sites were those visited by Dr. Thomas Smith's ground crew in past years. Dr. Smith's database was comprised of over 400 sites, of which over 170 were considered as potential training sites after being verified for positional accuracy by AKSO and ANHP personnel. Land cover data often did not total to 100% cover, but these site values were reviewed and adjusted by AKSO and AKNHP staff to develop corrected "bird's-eye" view land cover summaries for each site to be used in the training data sets.

Finally, a fourth source for training site locations were those sites developed by **GRS** personnel, to represent obvious types and remaining void areas in the classification. In some cases these areas represented the obvious types like water, barren, and glacier/snow types that were easily identified from aerial photography. However, some of these additional sites were added in stubborn gaps that remained in the classification maps after initial classification efforts. After growing training polygons in these remaining unclassified areas, a combination of aerial photo-interpretation and comparison of statistical similarity with existing classes were used to define preliminary cover type descriptions for these supplemental classes. Many of these training sites, and/or spectrally similar areas, were visited by AKSO personnel during a supplemental aerial survey undertaken during September of 2001. Although this fixed-wing survey did not produce detailed "bird's-eye" view cover component data for the visited sites, it provided **GRS** with general categorical descriptions of existent land cover types and digital photographs of these supplemental sites.

Supervised Training Site Review and Evaluation

The appropriateness of the training data was determined in two ways. One way involved a review of classification confusion reports that described confusion between training signatures within each training data set. The second involved a review of classification fidelity to determine if classified pixels within training sites yield the same land cover data descriptions (attributes) as the field survey data descriptions. If these evaluations yielded results indicating problem situations or data inconsistencies regarding a site, then the specific site was reevaluated and the problem identified. If the problem could not be resolved, the site was discarded from the training data set.

Unfortunately, both of these evaluations are only useful in the determination of whether or not the training data that have been developed provide results that are consistent with other training data. Neither of these measures necessarily indicates whether or not the training data are an accurate representation of what the true land cover conditions are. For instance, if the vegetation in a specific site is incorrectly identified, or assigned incorrect cover estimates, these data will not necessarily be identified as incorrect or inconsistent unless there are other similar, slightly overlapping training sites that indicate an inconsistency amongst the land cover data. Therefore, sites that are totally non-overlapping will tend to have high pixel fidelity, but could still be wrong, if the underlying "ground truth" are incorrect. Therefore, data lacking confusion and exhibiting high self-classification may still be wrong. Large training data sets with multiple sample sites are desirable as they may help confirm the 'goodness' of the data.

Class Confusion Evaluation

Training classes were evaluated for confusion in terms of their Jeffries-Matusita (J-M) distance, which is a reliable indicator of class separability (Swain and Davis 1978). A maximum J-M distance of 2.0 between two classes means that they are perfectly distinguishable from each other. A minimum J-M distance of 0.0 indicates that two classes are, as far as the classifier is concerned, spectrally undistinguishable from each other. In general, the probability of error in distinguishing two classes drops as their J-M distance increases. At a J-M distance of 1.4, this probability is approximately 1%. Figure 12 shows the relationship of the J-M distance to the upper and lower limits of the probability of error.

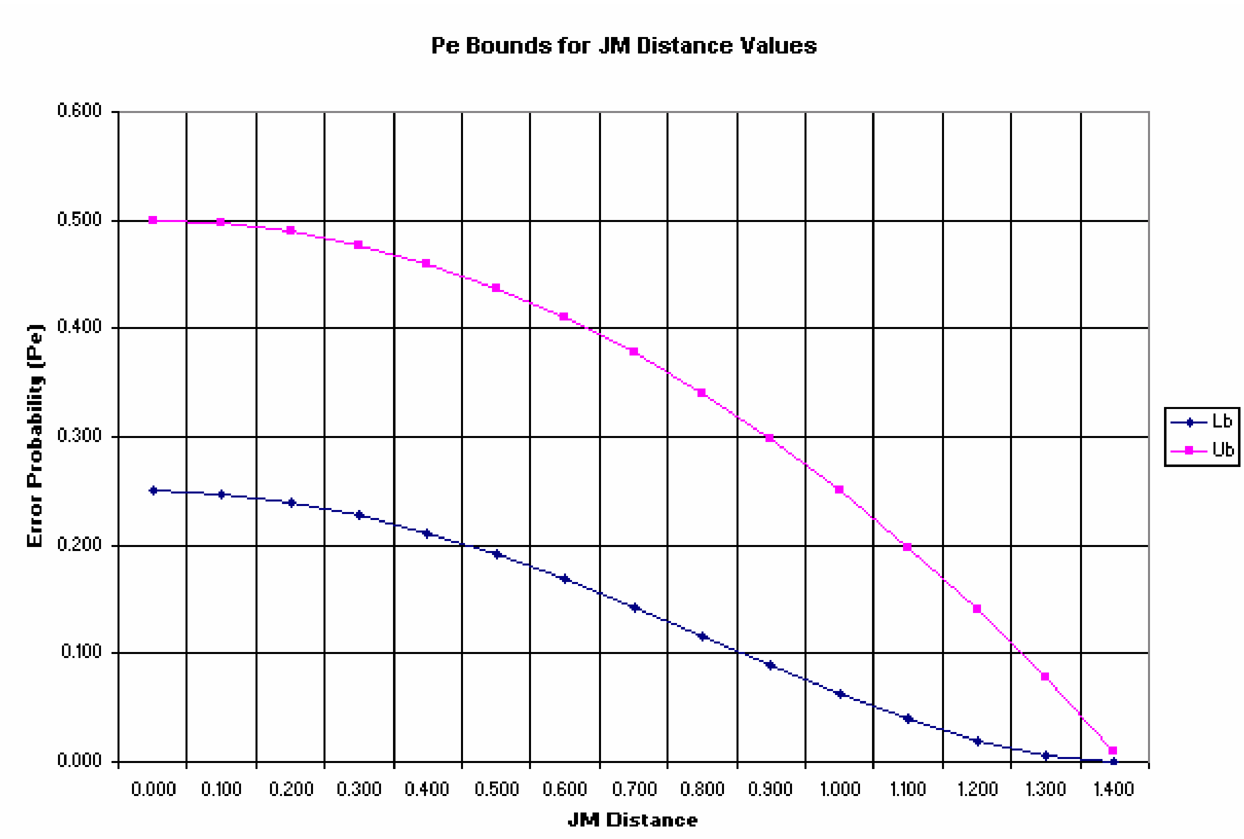


Figure 12: Probability of Error and J-M Distance

The Confusion Report used to evaluate class confusion is generated for all training sites used in all of the training sets. The target minimum J-M distance value for inclusion in the Confusion report was set at a 1.4 threshold. Training sites related by J-M distance values below this threshold were reviewed and evaluated.

Sample portions of the Confusion Report are shown in Table 6. The data in this report are organized by training site and the training sites are listed in order of the Calculated Class. For each site, all potentially confused training sites are listed in order of the magnitude of the J-M distance, along with the land cover characteristics of these sites. The training set in which the confusion exists is also listed for each confused site. The training set name is a combination of

::

Bad Confusion:

::

Trsite_id:	80816		33	Willow:Low:Open	2	0	68	30	0	0	0	0	148

Training Set	Confused_id	J-M Dist	Cover Type	Calculated Class	Tree Cover	Pct Conifer	Shrub Cover	Forb Cover	Barren Cover	Other Cover	Aspect	Slope	Elev
7119_sup	10158	0.669	11	Balsam Poplar:Open	30	0	5	60	5	0	117	1	23
7119_sup	1062	1.001	11	Balsam Poplar:Open	46	0	0	75	0	0	307	1	184
7119_sup	80501	1.058	13	Birch:Open	75	20	20	5	0	0	255	9	581
7119_sup	80820	1.104	13	Birch:Open	35	0	43	17	5	0	315	1	187
7119_sup	80802	1.137	13	Birch:Open	35	0	50	15	0	0	68	2	331
7018_sup	80820	1.150	13	Birch:Open	35	0	43	17	5	0	315	1	187
7119_sup	81008	1.178	21	Alder:Tall:Closed	2	0	93	5	0	0	0	0	3
7018_sup	80821	1.341	12	Deciduous:Mixed:Closed	70	0	15	15	0	0	135	1	207
7119_sup	80706	1.359	68	Wet:Sedge	0	0	0	28	0	0	293	2	102
7018_sup	81008	1.392	21	Alder:Tall:Closed	2	0	93	5	0	0	0	0	3

::

Good/Acceptable Confusion:

::

Trsite_id:	80713		Hwd	Balsam Poplar:Open	50	0	25	25	0	0	63	1	197

Training Set	Confused_id	J-M Dist	Cover Type	Calculated Class	Tree Cover	Pct Conifer	Shrub Cover	Forb Cover	Barren Cover	Other Cover	Aspect	Slope	Elev
7119_npgW	80707	1.013	Hwd	Balsam Poplar:Open	41	0	9	50	0	0	207	1	102
7119_ypgW	80707	1.013	Hwd	Balsam Poplar:Open	41	0	9	50	0	0	207	1	102
7119_ypgW	80819	1.309	Hwd	Balsam Poplar:Open	55	0	40	5	0	0	207	1	259
7119_npgW	80819	1.309	Hwd	Balsam Poplar:Open	55	0	40	5	0	0	207	1	259
7119_ypgW	1076	1.336	Hwd	Balsam Poplar:Open	35	0	45	20	0	0	0	0	256
7119_npgW	1076	1.336	Hwd	Balsam Poplar:Open	35	0	45	20	0	0	0	0	256

Trsite_id:	80815		Hwd	Birch:Closed	70	0	5	25	0	0	212	4	197

Training Set	Confused_id	J-M Dist	Cover Type	Calculated Class	Tree Cover	Pct Conifer	Shrub Cover	Forb Cover	Barren Cover	Other Cover	Aspect	Slope	Elev
7119_npgW	80821	0.557	Hwd	Birch:Closed	70	0	15	15	0	0	135	1	207
7119_ypgW	80821	0.557	Hwd	Birch:Closed	70	0	15	15	0	0	135	1	207
TM7_ypgW	80821	1.075	Hwd	Birch:Closed	70	0	15	15	0	0	135	1	207
TM7_npgW	80821	1.075	Hwd	Birch:Closed	70	0	15	15	0	0	135	1	207
TM7_ypgW	80825	1.145	Hwd	Balsam Poplar:Open	50	0	45	5	0	0	304	1	374
TM7_npgW	80825	1.145	Hwd	Balsam Poplar:Open	50	0	45	5	0	0	304	1	374
7119_npgW	80814	1.288	Hwd	Birch:Closed	75	0	5	20	0	0	17	8	180
7119_ypgW	80814	1.288	Hwd	Birch:Closed	75	0	5	20	0	0	17	8	180
7219_ypg	80501	1.364	Hwd	Birch:Closed	75	20	20	5	0	0	255	9	581
7219_ypgW	80501	1.364	Hwd	Birch:Closed	75	20	20	5	0	0	255	9	581
7219_ypg	80503	1.375	PHw	Mixed Deciduous-Conifer:Closed	75	32	10	15	0	0	280	9	626
7219_ypgW	80503	1.375	PHw	Mixed Deciduous-Conifer:Closed	75	32	10	15	0	0	280	9	626
TM7_ypgW	80814	1.394	Hwd	Birch:Closed	75	0	5	20	0	0	17	8	180
TM7_npgW	80814	1.394	Hwd	Birch:Closed	75	0	5	20	0	0	17	8	180

Table 6: Confusion Report (portion)

the characteristics of the training set that distinguish it from other training sets. These characteristics include the scene, set/mask (all supervised [sup], spruce [ypg], and no spruce [npg]), and subset (water [W] versus no water). 7219_ypgW represents the training set with Water sites for the Spruce masked area of scene 7219.

Confusion data were reviewed to identify potential sources of confusion. Review involved identification of nonconforming sites and was not based on any quantitative data limits, as there is tremendous potential overlap of land cover characteristics that have different land cover types. In general, those sites that were nonconforming are those that were not expected to be confused when one considered their descriptive data. Confusion between sites of similar or equal land cover types or cover characteristics was considered in agreement and acceptable. This type confusion was “good confusion,” as this sort of comparison confirms that the overlapping signatures are classifying pixels of similar land cover characteristics. On the other hand, confusion between sites of dissimilar or completely different land cover types or cover characteristics was considered unacceptable. This confusion was “bad confusion,” as it indicates similar or overlapping spectral data for dissimilar land cover type characteristics. The magnitude of the confusion, as indicated by the J-M distance, was also considered when evaluating training classes. J-M distance values near or above 1.3 were considered still viable for class separation, given a less than 10% chance of misclassification. In reality, class separation is still possible at lower J-M distances, but the probability of classification error is greater. In these cases, the nature of what classes were confused with each other became most important in evaluating the confusion. For the Katmai Land Cover Mapping Project, if the confused classes were of fairly similar land cover characteristics, **GRS** accepted the overlap of the spectral training data, since the resulting class would likely still fall within the similar land cover type. However, if land cover values were sufficiently confused and the types were dissimilar, then the confusion had to be resolved.

The first step in resolving the confusion was to recheck the location, shape, and size of the training areas involved to be certain that the training data had been collected in the appropriate locations. If the training polygons were correct, then resolution required either removal of the badly confused class or segregation of the confused classes into training data sets that would be applied separately to the imagery. The segregation of training data into sets that had “spruce” data and “no spruce” data is an example of how confusion between some of the low-density spruce types and non-spruce types could be resolved.

After resolving as much class confusion as was possible, a certain degree of confusion remained. The confusion that occurred between Spruce and Wet Herbaceous classes could not be resolved by means of training area editing. In order to resolve these inconsistencies, a spruce mask was developed based on knowledge of spruce distribution within the project area. The extent of this mask was delineated by the AKSO and applied by **GRS**. The spruce mask was implemented by creating a second training set (for each scene) in which all classes containing more than 2.5% spruce cover were eliminated. The “no spruce” training set was then applied to those portions of the imagery that were thought to be void of spruce (as indicated by the mask), while the training sets that included spruce were applied to those portions of the imagery where spruce presence was acceptable. The development and application of different training sets was instrumental in minimizing the confusion between a few Wet Herbaceous and White Spruce training classes, that otherwise were virtually inseparable. In another instance of confusion, Water classes were being misapplied to pixels in barren or sparsely vegetated mountainous areas. Development of training sets with “water” and with “no water” classes resolved this classification problem. After review and evaluation of the training data within the specific training data sets, all invalid or confused classes that could not be resolved were removed. The complete Confusion Report generated after review and resolution of all confusion is included as a project deliverable along with all the other delivered data.

Classification Fidelity Evaluation

A test of how the confusion affects the classification output may be determined by review of the classification fidelity data. All training classes were evaluated for classification fidelity. Each training site was evaluated to determine how well the individual training area was classified by comparing the characteristics of the pixels that formed each training class with the observed field characteristics of the training area. Fidelity is viewed in two ways. The first compares the number of pixels of the training area that were actually classed as that training class value. The second compares the land cover characteristics developed by summarizing the classified pixel attributes of the training site with the observed land cover characteristics.

The Fidelity Report (see Table 7) is used to evaluate classification fidelity. Formatted as a spreadsheet, this report shows fidelity data for different versions of the participating training sets, as applied to each scene (the training set names follow the same conventions as in the Confusion Report). Report data were developed by computing for each training area the stratified averages of the different classified pixels' land cover values based on a Maximum Likelihood, 90% threshold, supervised classification of these areas. Fidelity data included the scene ID, training site ID, percent pure (self-classification), area pixel count, calculated cover-type code, predominant cover component, and predominant component's cover, as well as a break-down of training area cover composition based on the classification of the training area pixels. These data were then compared to the "ground truth", a more generalized calculated field call based on the field data, and the field described cover components of each site (also included in the table) to evaluate the fidelity of each training class.

The following criteria were used in evaluating pixel fidelity for each training class:

- Self-Match -- the level of self-classification (percent pure) of an area – how many pixels of the training area were classified as that same class. The following percent self-classification limits were used to describe the degree of self-matching:

Degree of Match	Self-Classification Limits
Strong Match (**M**)	>= 75%
Slight Match (**m**)	>= 50% and < 75%
Slight Mismatch (**n**)	>= 25% and < 50%
Strong Mismatch (**N**)	< 25%

- Type-Match -- the degree of matching land cover attributes.
    - Sites that matched land cover classes of type, predominant species, and density for the land cover components.
    - Sites that showed land cover class component values that differed from calculated ones by no more than approximately ten percent – these areas might not exactly match class values, but still matched land cover attributes. These cases typically concerned comparison of values that are similar, but which may in fact span type or class thresholds or limits.

Scene/Training Set	trsite id	Type Match	Self-Match	percent pure	pixel count	type	pr_comp	pr_comp cover	cover class	tree cover	veg cover	conf cover	hdwd cover	shr cover	tsh cover	lsh cover	dsh cover	hrb cover	bar cover	oth cover	same pixels
7119_npgW	81105			81.3%	32	SVg	Dwarf shrub	10.3		0.0	29.9	0.0	0.0	11.2	0.0	0.8	10.3	18.7	70.0	0.1	26
7119_ypgW	81105			81.3%	32	SVg	Dwarf shrub	10.3		0.0	29.9	0.0	0.0	11.2	0.0	0.8	10.3	18.7	70.0	0.1	26
7219_npgW	81105			98.1%	52	SVg	Dwarf shrub	10.3		0.0	25.1	0.0	0.0	11.3	0.0	1.0	10.3	13.8	74.9	0.0	51
7219_ypgW	81105			98.1%	52	SVg	Dwarf shrub	10.3		0.0	25.1	0.0	0.0	11.3	0.0	1.0	10.3	13.8	74.9	0.0	51
tm7_npg	81105			100.0%	11	SVg	Dwarf shrub	10.0		0.0	25.0	0.0	0.0	11.0	0.0	1.0	10.0	14.0	75.0	0.0	11
tm7_npgW	81105			100.0%	11	SVg	Dwarf shrub	10.0		0.0	25.0	0.0	0.0	11.0	0.0	1.0	10.0	14.0	75.0	0.0	11
tm7_ypgW	81105			100.0%	11	SVg	Dwarf shrub	10.0		0.0	25.0	0.0	0.0	11.0	0.0	1.0	10.0	14.0	75.0	0.0	11
TrainCalc_Data_Match	81105	M	M	93.0%		Sparse Vegetation															
TrainCalc'ed_032602	81105	M	M		1	SVg	Dwarf shrub	10.0		0.0	25.0	0.0	0.0	11.0	0.0	1.0	10.0	14.0	75.0	0.0	1
TrainCalc'ed_072302	81105	M	M		1	SVg	Dwarf shrub	10.0		0.0	25.0	0.0	0.0	11.0	0.0	1.0	10.0	14.0	75.0	0.0	1
TrainCall	81105	M	M		1	Sparse Vegetation	Dwarf shrub	10.0		0.0	25.0	0.0	0.0	11.0	0.0	1.0	10.0	14.0	75.0	0.0	1
7119_npgW	81107			100.0%	24	TSh	Tall shrub aln	40.0	Open	0.0	97.0	0.0	0.0	65.0	65.0	0.0	0.0	32.0	2.0	1.0	24
7119_ypgW	81107			100.0%	24	TSh	Tall shrub aln	40.0	Open	0.0	97.0	0.0	0.0	65.0	65.0	0.0	0.0	32.0	2.0	1.0	24
7219_npgW	81107			79.0%	138	TSh	Tall shrub aln	41.5	Open	0.1	96.8	0.0	0.0	69.8	67.8	1.7	0.4	26.9	2.4	0.8	109
7219_ypgW	81107			79.0%	138	TSh	Tall shrub aln	41.5	Open	0.1	96.8	0.0	0.0	69.8	67.8	1.7	0.4	26.9	2.4	0.8	109
tm7_npg	81107			26.9%	130	TSh	Tall shrub aln	37.1	Open	0.2	98.8	0.0	0.0	75.9	65.5	6.9	3.5	23.0	0.7	0.3	35
tm7_npgW	81107			26.9%	130	TSh	Tall shrub aln	37.1	Open	0.2	98.8	0.0	0.0	75.9	65.5	6.9	3.5	23.0	0.7	0.3	35
tm7_ypgW	81107			26.9%	130	TSh	Tall shrub aln	37.1	Open	0.2	98.8	0.0	0.0	75.9	65.5	6.9	3.5	23.0	0.7	0.3	35
TrainCalc_Data_Match	81107	M	m	52.0%		Tall shrub:Open:Alder-Willow															
TrainCalc'ed_032602	81107	M	m		1	TSh	Tall shrub aln	40.0	Open	0.0	97.0	0.0	0.0	65.0	65.0	0.0	0.0	32.0	2.0	1.0	1
TrainCalc'ed_072302	81107	M	m		1	TSh	Tall shrub aln	40.0	Open	0.0	97.0	0.0	0.0	65.0	65.0	0.0	0.0	32.0	2.0	1.0	1
TrainCall	81107	M	m		1	Alder:Tall:Open	Tall shrub aln	40.0	Open	0.0	97.0	0.0	0.0	65.0	65.0	0.0	0.0	32.0	2.0	1.0	1
7119_ypgW	81223			100.0%	28	PGl	White Spruce	25.0	Open	27.0	70.0	92.6	7.4	65.0	10.0	20.0	35.0	5.0	3.0	0.0	28
7219_ypg	81223			82.5%	40	PGl	White Spruce	23.3	Open	26.4	70.9	88.3	11.7	64.0	8.5	18.4	37.1	6.9	2.7	0.0	33
7219_ypgW	81223			82.5%	40	PGl	White Spruce	23.3	Open	26.4	70.9	88.3	11.7	64.0	8.5	18.4	37.1	6.9	2.7	0.0	33
TrainCalc_Data_Match	81223	M	M	87.0%		White Spruce:Open															
TrainCalc'ed_032602	81223	M	M		1	PGl	White Spruce	25.0	Open	27.0	70.0	92.6	7.4	65.0	10.0	20.0	35.0	5.0	3.0	0.0	1
TrainCalc'ed_072302	81223	M	M		1	PGl	White Spruce	25.0	Open	27.0	70.0	92.6	7.4	65.0	10.0	20.0	35.0	5.0	3.0	0.0	1
TrainCall	81223	M	M		1	Spruce:Open	White Spruce	25.0	Open	27.0	70.0	92.6	7.4	65.0	10.0	20.0	35.0	5.0	3.0	0.0	1
7119_npgW	81224			71.1%	128	TSh	Tall shrub aln	15.0	Open	0.7	99.1	0.0	0.0	54.3	28.6	22.5	3.2	44.8	0.2	0.0	91
7119_ypgW	81224			71.1%	128	TSh	Tall shrub aln	15.0	Open	0.7	99.1	0.0	0.0	54.3	28.6	22.5	3.2	44.8	0.2	0.0	91
7219_npg	81224			66.7%	45	TSh	Tall shrub sal	25.0	Open	0.2	98.9	0.0	0.0	62.3	42.7	19.0	0.7	36.6	0.9	0.0	30
7219_npgW	81224			66.7%	45	TSh	Tall shrub sal	25.0	Open	0.2	98.9	0.0	0.0	62.3	42.7	19.0	0.7	36.6	0.9	0.0	30
7219_ypg	81224			66.7%	45	TSh	Tall shrub sal	25.0	Open	0.2	98.9	0.0	0.0	62.3	42.7	19.0	0.7	36.6	0.9	0.0	30
7219_ypgW	81224			66.7%	45	TSh	Tall shrub sal	25.0	Open	0.2	98.9	0.0	0.0	62.3	42.7	19.0	0.7	36.6	0.9	0.0	30
TrainCalc_Data_Match	81224	M	m	69.3%		Tall shrub:Open:Alder-Willow															
TrainCalc'ed_032602	81224	M	m		1	TSh	Tall shrub aln	20.0	Open	0.0	100.0	0.0	0.0	53.0	33.0	20.0	0.0	47.0	0.0	0.0	1
TrainCalc'ed_072302	81224	M	m		1	TSh	Tall shrub aln	20.0	Open	0.0	100.0	0.0	0.0	53.0	33.0	20.0	0.0	47.0	0.0	0.0	1
TrainCall	81224	M	m		1	Tall shrub:Open:Alder-Willow	Tall shrub aln	20.0	Open	0.0	100.0	0.0	0.0	53.0	33.0	20.0	0.0	47.0	0.0	0.0	1

Table 7: Detail of Pixel Fidelity Report for Katmai training classes.

- Sites that showed land cover class component values that differed from calculated ones by no more than approximately twenty-five percent – these areas typically did not match class values in one category of land cover, but still matched the general land cover attributes.

Type match comparison typically concerned comparison of values that are similar in most cases, but not all cases, sometimes resulting in one different categorical value because the field attributes and weighted averages spanned the threshold of a class decision rule value.

Sites that met the first two type-match criteria for similarity (<= 10 percent difference) of the major type characteristics (classes and /or values) were considered strong type matches (**M**). Sites showing class and attribute values that differed by more than 10 percent but less than 25 percent were not considered as strong a match, but still considered a slight match (**m**). Sites that differed by more than 25 percent were considered mismatches (**m**) and were deemed unacceptable if causes of the mismatch could not be resolved and eliminated.

Matches were determined for each training set in which a training site participated. A separate line in the report describes the fidelity results of a training class within a training set. The degree of self-match was developed for each site by training set based on the percent self-classification. In addition, a degree of self-match was determined for the site as a whole, based on the average of the percent self-classification values by training set. The report also includes a description of the training site field description (TrainCall), as well as the calculated summary and type data (TrainCalc'ed[date]) to verify that the calculated data based the project classification logic and lookup tables agree with the field description (different dates reflect values based on corrections and modifications made to the field data descriptions to correct missing and erroneous data). Table 8 shows results of the evaluation of the Fidelity report data, for both self matches and type matches for all the training sites involved in the project.

It is apparent in these data that a low "percent pure" self-match value did not necessarily imply poor fidelity of the class. While only 65 percent of the sites were classified as strong matches based on self-classification, 99.7 percent of the sites were considered matches by type characteristics. Nearly all of the sites that did not self-match ("purity") did match by type characteristics. Sites confused with other sites of very similar land cover attributes show these low "percent pure" values, but still yield very similar land cover characteristics resulting in a match. This situation indicates that the "good confusion" of training site data helps develop similar land cover characteristics, as opposed to different values. In earlier reports, truly confused sites showed mismatches by self-match and type-match. These mismatched sites typically confirmed confusion of the subject site with another site. This confusion was addressed by reviewing the aforementioned J-M Distance Confusion Report, assessing which sites participated in the confusion, and resolving or removing the truly confused ("bad") training sites from the same training set, thereby avoiding the inclusion of confused training data in the classification effort. While no accuracy assessment was performed on this data set, the resolution of class confusion, as determined from fidelity and J-M distance reports, builds confidence in these classification data.

Supervised Training Set Development Results

The complete training data set used for this project was comprised of 360 training sites. Of the original 222 aerial survey sites, 210 were found suitable for inclusion in the final training data sets. These sites are identifiable in the final pixel map, as they have trsite_id values that range between 80401 and 91226.

Table 8: Fidelity Report Match by Purity and Type

All Scenes	------- Self-Classification --------				------------ Same Type -------------			
Type	**N < 25%**	**n 25-49%**	**m 50-74%**	**M >= 75%**	**N**	**n**	**m slight**	**M same**
Spruce:Woodland	0	1	1	11	0	0	0	13
Spruce:Open	0	1	5	8	0	0	0	14
Spruce:Closed	0	0	0	2	0	0	0	2
Spruce Stunted:Woodland	0	0	0	1	0	0	0	1
Spruce Stunted:Open	0	0	1	0	0	0	0	1
Spruce/Deciduous:Woodland	0	1	0	1	0	0	0	2
Spruce/Deciduous:Open	0	2	5	2	0	0	0	9
Spruce/Deciduous:Closed	0	0	1	1	0	0	0	2
Balsam Poplar:Woodland	0	1	1	2	0	0	0	4
Balsam Poplar:Open	1	2	6	3	0	0	0	12
Balsam Poplar:Closed	0	1	1	0	0	0	0	2
Birch:Woodland	1	0	1	4	0	0	0	6
Birch:Open	0	1	2	2	0	0	0	5
Birch:Closed	0	0	1	7	0	0	0	8
Deciduous:Woodland	0	0	0	1	0	0	0	1
Deciduous:Open	0	0	3	0	0	0	0	3
Deciduous:Closed	0	0	1	3	0	0	0	4
Alder-Willow:Tall:Open	0	0	2	1	0	0	0	3
Alder:Tall:Open	0	0	4	7	0	0	0	11
Birch-Willow:Tall:Open	0	0	0	1	0	0	0	1
Willow:Tall:Open	0	0	5	6	0	0	0	11
Alder-Willow:Tall:Closed	0	1	3	2	0	0	0	6
Alder:Tall:Closed	0	5	11	14	0	0	1	29
Mixed Shrub:Tall:Closed	0	0	0	1	0	0	0	1
Willow:Tall:Closed	0	1	4	3	0	0	0	8
Birch-Willow:Low:Open	0	0	0	1	0	0	0	1
Mixed Shrub:Low:Open	0	2	8	9	0	0	0	19
Willow:Low:Open	0	1	5	15	0	0	0	21
Mixed Shrub:Low:Closed	0	1	0	3	0	0	0	4
Dwarf Shrub:Mixed	1	4	14	49	0	0	0	68
Forb	0	0	1	1	0	0	0	2
Graminoid	0	0	0	1	0	0	0	1
Graminoid:Dry/Mesic	0	0	4	11	0	0	0	15
Herbaceous:Marsh	0	0	5	23	0	0	0	28
Lichen	0	0	0	2	0	0	0	2
Moss/Lichen	0	0	0	2	0	0	0	2
Sparse Vegetation	0	0	0	18	0	0	0	18
Barren	0	0	3	11	0	0	0	14
Snow/Ice	0	0	0	2	0	0	0	2
Water	0	0	0	3	0	0	0	3
Totals	**3**	**25**	**98**	**234**	**0**	**0**	**1**	**359**
Percent Match				**65.0%**				**99.7%**

A total of 20 of the AKSO sites participated in final training sets. Classes of this type are identified by their trsite_id value, which is between 1000 and 1900. These trsite_id values were developed by adding a constant value of 1000 to the AKSO site number.

A total of 70 of the Dr. Thomas Smith ground sites were added to the composite training data set. These sites are identified by trsite_id values that ranged from 10,000 to 10999. These values were developed by adding a constant of 10,000 to the Dr. Smith site id values.

GRS added 60 additional sites to the final training data sets. 10 of these sites were in obvious classes of water, barren, and snow/glacier. These sites are identified by trsite_id values between 1 and 10. The other 50 were in land cover classes representing vegetation types not included in any of the field data training sets. 13 of these sites represented Sparse Vegetation types, 8 sites represented Forb types, 6 represented Dwarf Shrub types, 17 represented shrub types and 6 represented tree types. The sites have trsite_id values that range from 1901- 4999.

Of all the 360 training sites developed, some were used in the classification of more than one scene, as they were located in areas covered by overlapping imagery, while others were used in only one scene. Of the 360 training sites, 264 were used for more than one scene. However, 84 sites were used only for scene 7219, 11 sites were used only for scene 7119, and one site was used for only the TM7 replacement imagery. A breakdown of the number of training sites by scene and source is shown in Table 9. This indicates that the training sets for the TM7 imagery were an improvement, in terms of number of sites, over the original 7018 scene, there were still nearly 100 fewer sites in the TM7 training sets than were present in the 7219 or 7119 training data sets.

Table 9: Supervised Training Sites by Scene and Source

Source	Aerial Survey	AKSO	TSmith	Supplemental	Total
Scene					
7219	168	7	34	36	245
7119	157	11	39	43	250
7018	41	5	33	12	91
TM7	68	12	37	32	149
Total	**210**	**20**	**70**	**60**	**360**

All of the 360 training sites comprised the different training sets used to classify the imagery. As there were issues involving the misapplication of Water classes, four training sets were developed for each scene to guide the application of Spruce and Water training data. A total of 9 different training sets were developed. The following training sets were developed for each scene that had Water and/or Spruce misclassification issues – Spruce without Water, Spruce with Water, No Spruce without Water, and No Spruce with Water. The number of training sites by scene and training set is shown below in Table 10.

Table 10: Supervised Training Sites by Training Set

Scene	7219	7219	7119	7119	TM7	TM7
Training set	No Water	Water	No Water	Water	No Water	Water
Spruce	241	245	N/A	250	N/A	149
No Spruce	185	189	N/A	229	144	147

Unsupervised Training Set Development

Unsupervised training data sets were also developed to perform an ISODATA classification of the imagery. These unsupervised data sets were created using the same statistical parameters and techniques that were applied during the initial image pre-processing classification efforts during the development of the candidate training site locations with a few exceptions. While the initial unsupervised classification efforts were used to develop classes within different strata of the project area, these unsupervised efforts were designed to develop classification maps that could be used to supplement supervised classification data and fill in remaining gaps and voids in the supervised pixel maps.

GRS generated unsupervised training sets for each normalized LANDSAT image set. The major ecoregions that formed the basis for stratification and unsupervised classification during the training site selection process were not involved in this mapping effort. Instead, unsupervised training data were developed for the cloud free portions of each scene. As many as 200 training classes were targeted for inclusion in each training set. If too few classes were developed, then the statistical parameters were relaxed and initial seed values altered to generate more classes. Resulting signature files were reviewed for valid (non-singular) covariance matrices prior to acceptance of an unsupervised training set.

As the project progressed and initial results were reviewed, it became apparent that there would be a need to segregate the imagery into two major areas – those areas where spruce was present and those areas were it was not present. After scene replacement with the newer imagery, the spruce mask used during supervised classification efforts was applied during the development of unsupervised class maps, resulting in both "spruce" and "no spruce" unsupervised class maps for the two major scenes that now covered the project area – scenes 7219 and TM7. Scene 7119 was not reprocessed in this manner, as it contributed so little area to the final map within the park boundaries.

A total of 110 classes were developed in the 'spruce' unsupervised training set for scene 7219 and were numbered from 51001-51117. 206 classes were developed in the "no spruce" unsupervised training set for 7219 and were numbered from 55101 to 55212. 95 classes were developed in the 'spruce' unsupervised training set for scene TM7 and were numbered from 53101-53097. 182 classes were developed in the "no spruce" unsupervised training set for scene TM7 and were numbered from 57101 to 57186.

Image Classification

GRS applied a hybrid classification methodology that uses both supervised and unsupervised techniques. Multiple classification maps are developed using supervised and unsupervised classification techniques with different statistical thresholds and training sets. The resulting different pixel class maps are then systematically merged to form a composite or final land cover pixel classification map. Classification data of the highest level of confidence are used first to form the initial classification map. Classification data from progressively lower levels of statistical confidence are then added to this initial map, filling in voids and gaps in the prior version of the map, until the final pixel map has been developed. Supervised classification techniques typically result in the classification of approximately 90-95% of the mapped area, while unsupervised techniques typically classify the remaining 5-10% of the imagery, depending on how thoroughly the training data represent the variety of land cover characteristics found in the project area.

Supervised Classification

Supervised classifications efforts were carried out for all participating scenes using bands 1, 2, 3, 4, 5, and 7 and the Maximum Likelihood classifier. Initial efforts involved the three initial scenes 7219, 7119, and 7018. Classification efforts were performed once again after replacement of scene 7018 and the majority of scene 7119 with the new TM7 data. In this second effort, a (B5 – B4)/(B5 + B4) ratio band was added to the classification. This pseudo-NDVI band was introduced in an attempt to increase separability between Spruce and Wet Herbaceous classes.

Two masks were developed to limit the application of the training data. A cloud mask was developed to limit the application of all training data to only those portions of the scene that were cloud-free and remove clouded portions of the project areas from the final pixel classification map. A spruce mask was also used in this classification effort in an effort to reduce confusion and misapplication of spruce in areas/types in which it did not occur. Some confusion existed between some Spruce and Wet Herbaceous types, while spruce was also occurring in low densities in types dominated by other land cover attributes, such as Dwarf Shrub. The spruce mask was developed based on knowledge of spruce distribution within the project area by the AKSO and applied by **GRS** in the individual scene classifications. The spruce mask was used to limit the assignment of spruce characteristics to those areas were spruce presence was acceptable.

Each scene was classified using the Maximum Likelihood classifier with different combinations of mask values, training sets, and statistical (confidence) thresholds. Supervised classifications were performed on each scene using increasingly relaxed confidence thresholds of 90%, 97.5%, and 100%. With a maximum of four training sets and three statistical thresholds, it was possible to develop a total of 12 (4X3) supervised classification maps for each scene. However, not all combinations of mask/area and training set were run, if the resulting classification map would make an insignificant contribution to the final map. For example, the water misclassification issue primarily involved scene 7219. Therefore, it was not necessary to split the training sites into separate ”water” and “no water” training sets for scenes TM7 or 7119. In all, a total of 10 supervised classification maps were developed for scene 7219, 7 classification maps were developed for scene TM7, and 4 classification maps were developed for scene 7119.

Unsupervised Classification

Unsupervised classification maps were generated using bands 1, 2, 3, 4, 5, and 7. The iso-clustering method was used to develop unsupervised class maps for all participating scenes. Parameters for this process were selected to ensure the maximum number of separable classes within each data set. The resulting unsupervised signature files were applied using the Maximum Likelihood classifier and 97.5% threshold to classify each scene. Two unsupervised classifications were performed for scenes 7219 and 7119, one for the “spruce” region and one for the “no spruce” region of each scene. Scene 7119 was not reclassified using masked or split (water/non-water) training classes due to the very small area this scene contributed to the project area.

In all, a total of 12 classification maps were developed for scene 7219, 9 classification maps were developed for scene TM7, and 4 classification maps were developed for scene 7119.

Land Cover Database Development

The resulting classification maps of each scene were then used to create one complete classification map. This map is a composite of portions of all the supervised and unsupervised classification maps. Land cover attribute descriptions were developed for each pixel class in the final map, based on the “ground truth” data used to estimate the attributes of each training class. Unsupervised class attributes were developed by relating each unsupervised class to supervised classes that occupied the same locations, reviewing these data for satisfactory data content, and then calculating land cover attributes based on the specific related supervised pixel classes. Land cover components of the “ground truth” information were included in the final database tables.

Final Pixel Map Development

The final pixel classification map was a composite pixel map built from the supervised and unsupervised classification maps developed for each participating scene and training set. All raw classification grids were standardized to common training class values by means of class lookup tables that matched each raw-grid value to its particular class id through each class’ unique *gridval* value. This was achieved using custom **GRS** grid reclassification processing software. The final map’s grid values were composed of the standardized *gridval* values each representing a specific supervised or unsupervised class with unique land cover attributes.

Pixels for this final composite map were assembled or included from the supervised and unsupervised classifications by overlaying them in a hierarchical order based on the specific scene, the training set, and the statistical threshold of the classification. As each subsequent pixel classification map was overlaid on the composite map, pixel values were transferred into the composite map when the pixel value filled a VOID location in the composite map. In this way, subsequent map data never replaced data in the composite map, but instead contributed new data in this map. All map data were mosaiced by training set for a given threshold level and scene before data from another threshold level or scene were added to the final map. The source of each pixel location is represented in the source grid as a numeric value that represents the scene, threshold, and training set. The mosaicing hierarchy was based on the following priorities in Table 11.

Table 11: Mosaicing Priorities for Pixel Map Development

- **Scene Priority** – the scene value is the 3rd (leftmost) character of the source value. the order in which the three scenes were mosaiced was:

Order	Scene	Source value (X––)
1	7219	2––
2	TM7	––
3	7119	1––

- **Statistical Threshold** – the threshold level is represented by the 2nd digit (–X–) of the source value. The order in which classifications of different thresholds were mosaiced was:

Order	Threshold	Source value (–X–)
1	supervised 90% threshold	–1–
2	supervised 97.5% threshold	–2–
3	**unsupervised** 97.5% threshold	–3–
4	supervised 100% threshold	–4–

- **Training Set** – the training set is represented by the 1st (rightmost) digit (––X) of the source value. The order by scene in which the different training set classifications were mosaiced was:

Scene	Order	Training Set	Source value (––X)
7219	1	"spruce" and "no water"	––4
7219	2	"spruce" and "water"	––3
7219	3	"no spruce" and "no water"	––2
7219	4	"no spruce" and "water"	––1
TM7	1	"spruce" and "water"	––3
TM7	2	"no spruce" and "water"	––1
TM7	3	"no spruce" and "no water"	––2
7119	1	"spruce" and "water"	––3
7119	2	"no spruce" and "water"	––1

The scene order affected which pixels would be used in the final map in areas of overlap where coverage was provided by multiple scenes. Scene 7219 took precedence over scene TM7, which took precedence over 7119. The final supervised 100% threshold classification was used to force every pixel location to have a class value in the final pixel map. These pixels can be distinguished from other pixels as they have values between 26001 and 50001. These values were determined by taking the original pixel values (1001-25001) and adding an offset of 25000 to each value.

Attribute Development

The Katmai land cover type database consists of two database tables. These two tables contain all the information that describes each class of the final pixel classification map. The main table is the value attribute table or .VAT table. The VALUE item (the column that contains the actual numeric pixel values) represents the grid values (*gridval*) of the grid data. This table contains all the land cover class summary information such as the trsite_id and calc_class, as well as percentage cover of trees, conifers and hardwood cover composition, and percent cover of all other general land cover components including, tall and low shrubs, dwarf shrub, and herbaceous forbs. These cover components comprise the “bird’s-eye view” description of each land cover type. Each table record also contains information regarding scene participation, hydric regime, and modifier. The second table (*cover_info.dat*) contains the individual species records (species, species code, cover, strata, and so forth) for each supervised training site used in the classification. This table, when joined to the .VAT table using the trsite_id columns of the .VAT table and cover_info.dat tables. A sample listing of a database record is shown in Table 12. Listings and definitions of the attributes of the katmai_lcover.vat and the cover_info.dat tables are included in Appendix D.

VALUE(gridval)	=	4011
COUNT	=	68972
SOURCE	=	AIR_00
SPEC_CLASS	=	Spruce:Open
GRIDCLVAL	=	4
MAJOR_CLASS	=	Spruce:Open
REGIME	=	mesic
MODIFIER	=	boreal
REMAP_VAR	=	0
TYPE	=	PGl
PREDOMINANT_COMP	=	White Spruce
PR_COMP_COVER	=	0.0
COVER_CLASS	=	Open
TREE_COVER	=	25.0
OTHER_VEG_COVER	=	70.0
CONF_COVER	=	25.0
HDWD_COVER	=	0.0
SHR_COVER	=	15.0
TSH_COVER	=	0.0
LSH_COVER	=	15.0
DSH_COVER	=	45.0
HRB_COVER	=	10.0
BAR_COVER	=	5.0
OTH_COVER	=	0.0
TRSITE_ID	=	80401
CALC_CLASS	=	White Spruce:Open
ALNUS_COVER	=	0.00
SALIX_COVER	=	13.00

Table 12: Sample Land Cover Type Attribute Listing (katmai_lcover.vat)

The land cover characteristics or attributes in the final pixel classification map database, as represented in the *katmai_lcover.vat* table, are derived from two sources – supervised classification data and unsupervised classification data.

The first source of information is the entire set of land cover information (field observations) associated with each training site used to form a training class in the supervised training sets. Throughout this entire classification effort a link has been retained that associates the original training site description with the *gridval* of the final pixel map. This link is used to relate or join the land cover attributes of the description with the mapped pixel elements. Using this link aerial survey data, Thomas Smith data, AKSO ground data, or supplemental data were associated with the proper pixel values in the final map.

Table 13: CrossTab Results

```
…
53054,10034,388
53057,10003,6
53057,10034,5533
53057,10036,2635
53058,10003,245
53058,10035,3920
53058,10080,169
53058,10081,1628
53058,10082,931
…
```

The second source of data describes the unsupervised pixel classes used to supplement the supervised classification data. All the unsupervised data were developed and applied without knowledge of any land cover characteristics. In order to estimate the land cover characteristics of unsupervised classes, the *isodata* classes developed in unsupervised classifications were related to supervised classes of the merged 90% threshold Maximum Likelihood classification using cross tabulation reports of the appropriate grids (the cross tabulation report describes a pixel-by-pixel overlay of the supervised and unsupervised pixel maps). Unsupervised map data were related to supervised map data that reflected similar masking of the spectral data (e.g. unsupervised “no spruce” classification map data were related to supervised “no spruce” map data). Each cross tabulation report described the frequency of each supervised class pixel (grid) value that comprised each unsupervised class of that map, as shown in the sample data in Table 13. In this example, unsupervised class 53057 corresponds to three supervised classes with grid values 10003, 10034, and 10036 – all different Open Broadleaf classes - with differing numbers of pixels (6, 5533, and 2635). Land cover attributes were then estimated by processing these data with ***GRS_polysum***. This process estimates land cover attributes of each unsupervised class by calculating the average attributes of all the different class attributes of the supervised pixels that are related to that unsupervised class weighting each contributing class by the number of pixels of that class. Using this approach, an unsupervised class’ characteristics are based solely on the attributes of all the supervised class attributes that occupy the same pixel locations in the supervised pixel map. An unsupervised class comprised solely of one supervised class (see 53054) would be assigned the attributes of that class, as the weighted average would be the same as the supervised class values. GRS_polysum rejected unsupervised classes lacking sufficient correlation to supervised classes during processing of these data. Sufficient correlation was set at a minimum of 50 percent coverage of the unsupervised class with supervised pixel values. Cross tabulation data were also reviewed to remove any water pixels from classes that did not appear to be “water” types. Land cover attributes were estimated in this manner for the classes in each unsupervised classification map that was used in the development of the final pixel classification map.

The land cover components estimated at each supervised training site were also included in a database table. These table records may be joined to the land cover attributes table using the trsite_id columns as the key field. These data were not generated for unsupervised classes. A sample listing of a record from this table is shown in Table 14.

```
Record 1:

  TRSITE_ID                    =          80404
  SOURCE                       =         AIR_00
  SPECIES                      =   Picea glauca
  SPCODENUM                    =             45
  SPCODE                       =           PIGL
  COVER                        =           15.0
  STRATA                       =        TR_CNFR
  COMMENT                      =
  GRIDVAL                      =           6001


Record 2:

  TRSITE_ID                    =          80404
  SOURCE                       =        AIR_00
  SPECIES                      =        Salix L
  SPCODENUM                    =            138
  SPCODE                       =          SALIX
  COVER                        =            5.0
  STRATA                       =             TS
  COMMENT                      =
  GRIDVAL                      =          31001
```

Table 14: Sample Cover Data Attribute Listing (cover_info.dat)

Results

The results of this land cover mapping project are contained in the composite classification pixel map and the associated data tables that describe the land cover characteristics of each pixel type. A total of 1,302 pixel classes are represented in the final map. Final map frequency of pixels by source, listed in the order in which the pixels were added to the final pixel classification map is shown in Table 15. A summary of pixels by scene and classification effort is shown in Table 16.

Table 15: Final Pixel Map Frequency by Source

Classification	Source Value	Frequency	Percent of Total
7219/"spruce-no water"/90% ML	214	4,094,263	13.6%
7219/"spruce-water"/90% ML	213	1,136,223	3.8%
7219/"no spruce-no water"/90% ML	212	4,368,705	14.5%
7219/"no spruce-water"/90% ML	211	367,358	1.2%
TM7/"spruce-water"/90% ML	13	1,904,169	6.3%
TM7/"no spruce-no water"/90% ML	12	1,109	0.0%
TM7/"no spruce-water"/90% ML	11	9,111,461	30.3%
7219/"spruce-no water"/97.5% ML	224	994,093	3.3%
7219/"spruce-water"/97.5% ML	223	16,572	0.1%
7219/"no spruce-no water"/97.5% ML	222	1,295,317	4.3%
7219/"no spruce-water"/97.5% ML	221	14,706	0.0%
TM7/"spruce-water"/97.5% ML	23	57,249	0.2%
TM7/"no spruce-no water"/97.5% ML	22	4,565	0.0%
TM7/"no spruce-water"/97.5% ML	21	1,937,149	6.4%
7119/"spruce-water"/97.5% ML	123	197,380	0.7%
7119/"no spruce-water"/97.5% ML	121	64,093	0.2%
7219/"spruce-water"/97.5% UnSup	233	771,317	2.6%
7219/"no spruce-water"/97.5% UnSup	231	1,104,106	3.7%
TM7/"spruce-water"/97.5% UnSup	33	29,449	0.1%
TM7/"no spruce-water"/97.5% UnSup	31	2,248,643	7.5%
7219/"spruce-water"/100% ML	243	102,190	0.3%
7219/"no spruce-water"/100% ML	241	55,294	0.2%
TM7/"no spruce-water"/100% ML	41	90,326	0.3%
7119/"spruce-water"/100% ML	143	131,305	0.4%
7119/"no spruce-water"/100% ML	141	18,810	0.1%
Total		**30,115,852**	**100.0%**
7219		**14,320,144**	**47.6%**
TM7		**15,384,120**	**51.1%**
7119		**411,588**	**1.4%**

Table 16: Number of Pixels by Classification Effort

Scene	Training Set	90% Threshold	97.5% Threshold	Isodata Classification	100% Threshold	Total	Percent of Project	Percent by Scene
7219	spruce – water	1,136,223	16,572	771,317	102,190	2,026,302	6.7%	
7219	spruce - no water	4,094,263	994,093	0	0	5,088,356	16.9%	
7219	No spruce – water	367,358	14,706	1,104,106	55,294	1,541,464	5.1%	
7219	No spruce - no water	4,368,705	1,295,317	0	0	5,664,022	18.8%	47.6%
TM7	spruce – water	1,904,169	57,249	29,449	0	1,990,867	6.6%	
TM7	spruce - no water	0	0	0	0	0	0.0%	
TM7	No spruce – water	9,111,461	1,937,149	2,248,643	90,326	13,387,579	44.5%	
TM7	No spruce - no water	1,109	4,565	0	0	5,674	0.0%	51.1%
7119	spruce – water	0	197,380	0	131,305	328,685	1.1%	
7119	No spruce – water	0	64,093	0	18,810	82,903	0.3%	1.4%
Total		**20,983,288**	**4,581,124**	**4,153,515**	**397,925**	**30,115,852**	**100.0%**	**100.0%**
Percent of Total		**69.7%**	**15.2%**	**13.8%**	**1.3%**	**100.0%**		

Data Modeling

All of the pixel classes in the final pixel map were derived from the image classification processes, with one exception. This one exception in the delivered data set involves one class developed using post-classification modeling of the data. After review of these classification data, there was an indication that there were some “water” pixels in locations or situations where they should not exist. These situations were typically areas of north or northwesterly aspect on steeper aspects. These water pixels were very likely the result of very dark (low digital value) pixels in areas of very low illumination that are most appropriately described as “terrain shadows.” After all the bands were normalized for differential illumination, these very dark pixels were still very dark. Unfortunately, very dark pixels tend to be confused with “water” pixels, as water reflects virtually no energy. These dark pixels were subsequently classified in error as “water” pixels, when they should be labeled as “terrain shadow” pixels. Post processing or modeling was applied to the final pixel classification map, in an effort to identify, and reclass as “terrain shadow” as many of these misclassified “water” pixels as possible.

This modeling effort entailed identifying suspect “water” locations in the final map and altering their values to indicate “terrain shadow” as opposed to “water.” Reclassification to “terrain shadow” was accomplished using the following rules:

1. The pixel was classified as a “water” type
2. The slope of the pixel was greater than 4 degrees or 9% slope
3. The aspect of the pixel was between an azimuth of 258 and 23 degrees (WSW to NNE aspects)

A total of **38,532** pixels that meet these criteria were reclassified as “terrain shadow.” Unfortunately, further investigation revealed that some valid “water” pixels were now misclassified as “terrain shadow.” This misclassification resulted along the edges of lakes where the slope map may have been slightly misaligned with the satellite imagery, resulting in areas of high slope along the edges of large bodies of water that overlap water pixels. This misalignment of the slope and “water” data caused some valid pixels to be changed to “terrain shadow” pixels. As further modeling refinements and pixel aggregation were not in the scope of this mapping effort, these data remained in the “terrain shadow” class in the final map.

Data Display and Representation

The resulting pixel map represents the 1,302 pixel classes used in the classification of the spectral data. These data values alone are not of much use, as they simply represent a value from 0 to 59,999. However, when these values are joined to the land cover information data associated with each training site, the resulting database includes many different levels of information. These levels range from major land cover types to percent composition of individual specie groups, such as alder and willow. All of these data may be used to develop different maps and information. As the terrain shadowing modeling effort indicates, these data may be used to develop any number of different maps based on the detailed information present in this data base, as well as ancillary data (spatially co-registered) that may enhance modeling efforts. Figure 13 demonstrates this capability, as it represents the field estimated types ("calculated") on the left in an area dominated by shrub types, but shows on the right a map of the herbaceous cover for the same area. While the herbaceous cover (or any other cover component) is not a part of the type map, each may be mapped separately using the database attributes. Many of the different attributes in Table 13 may be represented in similar maps, thereby resulting in many maps of the Katmai project area, rather than one land cover type map. Similarly, these same attributes can be used to model the data and enhance or modify the pixel classification map into other maps representing related data, such as wildlife habitat suitability or fire fuel class.

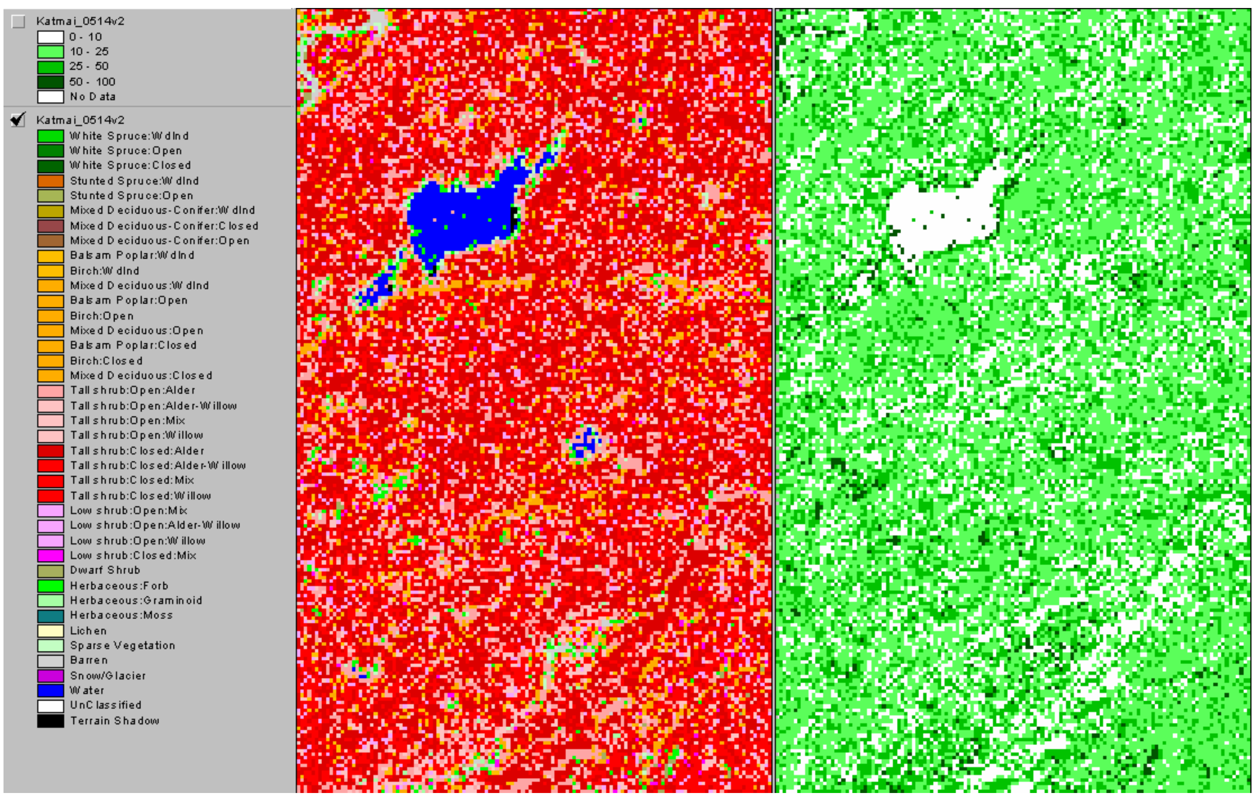


Figure 13: Type Map and Cover Component Map

Land Cover Type Area Distribution and Source

Table 17 represents the distribution of pixels and area by the forty (40) "calculated" land cover classes used in this project. These are the classes that were used when the land cover distributions were processed using the various land cover component thresholds used to make the field calls for each training site. Pixels values reflect only those pixels within the Katmai National Park and Preserve and do not include pixels within the ten-mile buffer around the park.

Table 17: Area and Frequency by 'Calculated' Land Cover Class

"Calculated" Class	Pixel Frequency	Acreage	Percent of Total	Type Total
Stunted Spruce:Wdlnd	3,016	670.6	0.0%	
White Spruce:Wdlnd	280,766	62,426.9	1.5%	
Stunted Spruce:Open	13,633	3,031.2	0.1%	
White Spruce:Open	282,833	62,886.5	1.5%	
White Spruce:Closed	9,882	2,197.2	0.1%	**3.2%**
Mixed Deciduous-Conifer:Wdlnd	81,161	18,045.7	0.4%	
Mixed Deciduous-Conifer:Open	294,354	65,448.1	1.6%	
Mixed Deciduous-Conifer:Closed	39,813	8,852.2	0.2%	**2.3%**
Balsam Poplar:Wdlnd	141,496	31,460.9	0.8%	
Birch:Wdlnd	205,832	45,765.7	1.1%	
Mixed Deciduous:Wdlnd	109,769	24,406.6	0.6%	
Balsam Poplar:Open	327,959	72,920.0	1.8%	
Birch:Open	524,826	116,692.4	2.9%	
Mixed Deciduous:Open	247,475	55,024.8	1.3%	
Balsam Poplar:Closed	92,468	20,559.8	0.5%	
Birch:Closed	221,558	49,262.3	1.2%	
Mixed Deciduous:Closed	52,978	11,779.4	0.3%	**10.5%**
Tall shrub:Open:Alder	547,386	121,708.5	3.0%	
Tall shrub:Open:Alder-Willow	253,202	56,298.2	1.4%	
Tall shrub:Open:Mix	37,914	8,430.0	0.2%	
Tall shrub:Open:Willow	780,693	173,583.2	4.3%	
Tall shrub:Closed:Alder	1,102,781	245,197.8	6.0%	
Tall shrub:Closed:Alder-Willow	461,069	102,516.4	2.5%	
Tall shrub:Closed:Mix	140,543	31,249.0	0.8%	
Tall shrub:Closed:Willow	297,044	66,046.2	1.6%	**19.7%**
Low shrub:Open:Mix	532,783	118,461.6	2.9%	
Low shrub:Open:Alder-Willow	18,206	4,048.0	0.1%	
Low shrub:Open:Willow	822,019	182,771.8	4.5%	
Low shrub:Closed:Mix	89,657	19,934.8	0.5%	**8.0%**
Dwarf Shrub	2,541,267	565,038.0	13.8%	**13.8%**
Herbaceous:Forb	826,521	183,772.8	4.5%	
Herbaceous:Graminoid	197,483	43,909.4	1.1%	
Herbaceous:Moss	199,309	44,315.4	1.1%	
Herbaceous:Lichen	24,789	5,511.7	0.1%	**6.8%**
Sparse Vegetation	966,119	214,811.7	5.3%	**5.3%**
Barren	3,026,054	672,828.0	16.5%	**16.5%**
Snow/Glacier	1,139,822	253,433.7	6.2%	**6.2%**
Water	1,384,385	307,811.1	7.5%	**7.5%**
Terrain Shadow/Unclassified	38,532	8567.4	0.2%	
Unclassified	11,636	2,587.2	0.1%	**0.3%**
Total	**18,369,033**	**4,084,262.6**	**100.0%**	**100.0%**

Table 18 represents the distribution of pixels and area by the major land cover classes that were derived by collapsing the forty "calculated" land cover classes in Table 17 into twenty-five (25) land cover classes that tend to reflect more generalized vegetation and density classes.

Table 18: Area and Frequency by Major Land Cover Class

Major Class	Pixel Frequency	Acreage	Percent of Total
Spruce Stunted:Woodland	3,016	670.6	0.0%
Spruce:Woodland	280,766	62,426.9	1.5%
Spruce Stunted:Open	13,633	3,031.2	0.1%
Spruce:Open	282,833	62,886.5	1.5%
Spruce:Closed	9,882	2,197.2	0.1%
Spruce/Deciduous:Woodland	81,161	18,045.7	0.4%
Spruce/Deciduous:Open	294,354	65,448.1	1.6%
Spruce/Deciduous:Closed	39,813	8,852.2	0.2%
Deciduous:Woodland	457,097	101,633.2	2.5%
Deciduous:Open	1,100,260	244,637.3	6.0%
Deciduous:Closed	367,004	81,601.5	2.0%
Tall Shrub:Open	1,619,195	360,019.9	8.8%
Tall Shrub:Closed	2,001,437	445,009.5	10.9%
Low Shrub:Open	1,373,008	305,281.5	7.5%
Low Shrub:Closed	89,657	19,934.8	0.5%
Dwarf Shrub	2,541,267	565,038.0	13.8%
Herbaceous - Forb	1,024,004	227,682.2	5.6%
Herbaceous - Moss	199,309	44,315.4	1.1%
Herbaceous - Lichen	24,789	5,511.7	0.1%
Sparse Vegetation	966,119	214,811.7	5.3%
Barren	3,026,054	672,828.0	16.5%
Snow/Ice	1,139,822	253,433.7	6.2%
Water	1,384,385	307,811.1	7.5%
Terrain Shadow	38,532	8,567.4	0.2%
Unknown	11,636	2,587.2	0.1%
Total	**18,369,033**	**4,084,262.6**	**100.0%**

Table 19 represents the "calculated" land cover type by their source, in terms of the classification threshold. The four sources represent the three thresholds of supervised classifications (90%, 97.5%, and 100%) and the unsupervised classification. These values indicate the source of the different land cover type data, which can be interpreted to indicate different levels of confidence in the pixel data. For instance, pixels derived from the 90% threshold classification efforts were classified with tighter statistics than pixels derived from the 97.5% threshold classification efforts and would be expected to be more a more accurate assignment of pixel classes than those pixels derived from classification efforts with looser statistics. Note that some "mixed" types were derived solely from the unsupervised classification efforts. These "mixed" types arise from the computation of the weighted average characteristics of each unsupervised class based on the overlay of the supervised pixel data with the unsupervised data. The primary source(s) for each land cover type is indicated in **boldface** type. In a few cases, two sources contributed nearly equally and both are highlighted.

Table 19: Area by Cover Type and Source

Class	90% Threshold	97.5% Threshold	Iso Classif.	100% Threshold	Total	90% Threshold	97.5% Threshold	Iso Classif.	100% Threshold
Stunted Spruce:Wdlnd	1,919	1,091	0	6	3,016	**63.6%**	36.2%	0.0%	0.2%
White Spruce:Wdlnd	191,940	52,060	36,586	180	280,766	**68.4%**	18.5%	13.0%	0.1%
Stunted Spruce:Open	10,494	3,135	0	4	13,633	**77.0%**	23.0%	0.0%	0.0%
White Spruce:Open	214,339	55,059	13,386	49	282,833	**75.8%**	19.5%	4.7%	0.0%
White Spruce:Closed	7,425	2,443	0	14	9,882	**75.1%**	24.7%	0.0%	0.1%
Mixed Deciduous-Conifer:Wdlnd	35,792	7,286	38,083	0	81,161	**44.1%**	9.0%	**46.9%**	0.0%
Mixed Deciduous-Conifer:Open	188,628	48,258	57,392	76	294,354	**64.1%**	16.4%	19.5%	0.0%
Mixed Deciduous-Conifer:Closed	32,141	7,670	0	2	39,813	**80.7%**	19.3%	0.0%	0.0%
Balsam Poplar:Wdlnd	71,002	17,123	53,266	105	141,496	**50.2%**	12.1%	**37.6%**	0.1%
Birch:Wdlnd	111,919	35,343	58,356	214	205,832	**54.4%**	17.2%	28.4%	0.1%
Mixed Deciduous:Wdlnd	14,460	3,794	91,495	20	109,769	13.2%	3.5%	**83.4%**	0.0%
Balsam Poplar:Open	264,475	63,272	16	196	327,959	**80.6%**	19.3%	0.0%	0.1%
Birch:Open	385,298	99,223	39,925	380	524,826	**73.4%**	18.9%	7.6%	0.1%
Mixed Deciduous:Open	158,885	36,941	50,797	852	247,475	**64.2%**	14.9%	20.5%	0.3%
Balsam Poplar:Closed	75,477	16,793	0	198	92,468	**81.6%**	18.2%	0.0%	0.2%
Birch:Closed	173,138	48,132	0	288	221,558	**78.1%**	21.7%	0.0%	0.1%
Mixed Deciduous:Closed	41,345	11,356	0	277	52,978	**78.0%**	21.4%	0.0%	0.5%
Tall shrub:Open:Alder	277,774	62,792	204,173	2,647	547,386	**50.7%**	11.5%	**37.3%**	0.5%
Tall shrub:Open:Alder-Willow	114,432	18,307	120,380	83	253,202	**45.2%**	7.2%	**47.5%**	0.0%
Tall shrub:Open:Mix	0	0	37,914	0	37,914	0.0%	0.0%	**100.0%**	0.0%
Tall shrub:Open:Willow	591,153	158,351	26,864	4,325	780,693	**75.7%**	20.3%	3.4%	0.6%
Tall shrub:Closed:Alder	827,863	174,139	76,584	24,195	1,102,781	**75.1%**	15.8%	6.9%	2.2%
Tall shrub:Closed:Alder-Willow	370,168	80,983	8,689	1,229	461,069	**80.3%**	17.6%	1.9%	0.3%
Tall shrub:Closed:Mix	94,046	41,371	0	5,126	140,543	**66.9%**	29.4%	0.0%	3.6%
Tall shrub:Closed:Willow	249,085	47,534	0	425	297,044	**83.9%**	16.0%	0.0%	0.1%
Low shrub:Open:Mix	401,713	104,609	16,643	9,818	532,783	**75.4%**	19.6%	3.1%	1.8%
Low shrub:Open:Alder-Willow	0	0	18,206	0	18,206	0.0%	0.0%	**100.0%**	0.0%
Low shrub:Open:Willow	585,361	163,003	68,802	4,853	822,019	**71.2%**	19.8%	8.4%	0.6%
Low shrub:Closed:Mix	69,413	19,946	0	298	89,657	**77.4%**	22.2%	0.0%	0.3%
Dwarf Shrub	1,456,350	426,870	638,704	19,343	2,541,267	**57.3%**	16.8%	25.1%	0.8%
Herbaceous:Forb	417,052	220,282	181,357	7,830	826,521	**50.5%**	26.7%	21.9%	0.9%
Herbaceous:Graminoid	139,309	55,261	0	2,913	197,483	**70.5%**	28.0%	0.0%	1.5%
Herbaceous:Moss	121,516	48,694	28,641	458	199,309	**61.0%**	24.4%	14.4%	0.2%
Herbaceous:Lichen	19,877	4,889	0	23	24,789	**80.2%**	19.7%	0.0%	0.1%
Sparse Vegetation	389,843	242,490	325,481	8,305	966,119	**40.4%**	25.1%	**33.7%**	0.9%
Barren	1,780,561	513,993	665,303	66,197	3,026,054	**58.8%**	17.0%	22.0%	2.2%
Snow/Glacier	931,450	77,337	114,694	16,341	1,139,822	**81.7%**	6.8%	10.1%	1.4%
Water	1,350,176	25,755	6,343	2,111	1,384,385	**97.5%**	1.9%	0.5%	0.2%
UnClassified/Terrain Shadow	20,580	7,848	21,237	503	50,168	**41.0%**	15.6%	**42.3%**	1.0%
Total	**12,186,399**	**3,003,433**	**2,999,317**	**179,884**	**18,369,033**	**66.3%**	**16.4%**	**16.3%**	**1.0%**

Deliverables

Deliverables for this mapping project included the following data themes and tables:

1. **Katmai_lcover** – an ArcInfo grid coverage and associated attribute data that represents the final composite pixel map representing the results of the supervised and unsupervised classification efforts implemented during this project.

2. **Cover_info.dat** – an ArcInfo table that contains the species-specific cover descriptions for each supervised training site associated with a pixel class in the Katmai_landcover grid.

3. **Source –** an ArcInfo grid coverage that corresponds to the Katmai_landcover grid and represents the source, in terms of the scene and classification, of each pixel's classification information.

4. **Slope -** an ArcInfo grid coverage that corresponds to the Katmai_landcover grid and represents the slope, in terms of degrees, as developed from the DEMs.

5. **Aspect -** an ArcInfo grid coverage that corresponds to the Katmai_landcover grid and represents the aspect, in terms of degrees counter clockwise from North, as developed from the DEMs.

6. **BTMGain -** an ArcInfo grid coverage that corresponds to the Katmai_landcover grid and represents the change in pixel value in Band 4 resulting from implementation of the illumination normalization process.

7. **Trsites** – an ArcInfo point coverage that represents the location and associated information describing each supervised training site used during the supervised classification efforts.

8. **Ecoregions** – an ArcInfo area coverage representing the original set of ecoregions used during the initial candidate site selection processing. Stratification was based on the value in the *ecoregion* column of the ecoregions.pat table.

9. **Katmai_confusion.rpt –** a text file that contains the confusion data summary for the supervised classification training data sets.

10. **Katmai_fidelity_review.xls** – an Excel spreadsheet that contains the fidelity data summary by site, as well as summaries in total and by scene.

11. Sample ArcView v3.2a legend files (*.avl) to assist in viewing these data.

All of these deliverables were referenced to the Project Alber's coordinate system, as defined in Table 2. All data were delivered on CD-ROM in both an ArcInfo grid/cover format as well as an ArcInfo interchange format. All deliverables are located under the 'Katmai_Mapping_2000' directory on the CD-ROM. Coverages were delivered under the 'Katmai_Maps' directory, interchange files were delivered in zipped format under the 'exports' directory, and reports were delivered under the 'reports' directory. This information is contained in a readme.txt file found under the main Katmai_Mapping_2000 directory.

Conclusions and Discussion

Scene Replacement

The two 2000 TM7 scenes used to replace the original Landsat TM5 data were initially acquired from the EROS Data Center in a geo-corrected format. This imagery was rejected due to unacceptable preprocessing of the imagery – Band 4 digital data had been adjusted differently in each scene using a cubic convolution filter, thereby preventing the mosaicing and subsequent classification of these new scenes. As this was the only delivery format available from the EROS Data Center, alternative means of orthorectification were acquired through Resource GIS and Imaging (RGI) of Vancouver, Canada. The scenes provided by RGI were then mosaiced and used for classification efforts in this project.

Class Representation and Training Site Availability

Replacement of scenes 7119 and 7018 with newer TM7 scenes 7019 and 7020 generated a change in the distribution and availability of training areas. The summary list of classes trained within each classified scene reported in Table 9 shows a net gain of 58 training sites in scene TM7 when compared with scene 7018. However, there was a net loss of 101 training sites (40%) when compared with scene 7119. While the newer imagery improved estimation of land cover type classes in the area previously covered by 7018, it resulted in a decreased number of classes in that portion of the project area previously covered by 7119. This was due to the decreased area of overlap between 7219 and the TM7 data, as opposed to the original area of overlap between 7219 and 7119, in which a significant number of training sites were located (see Figure 14).

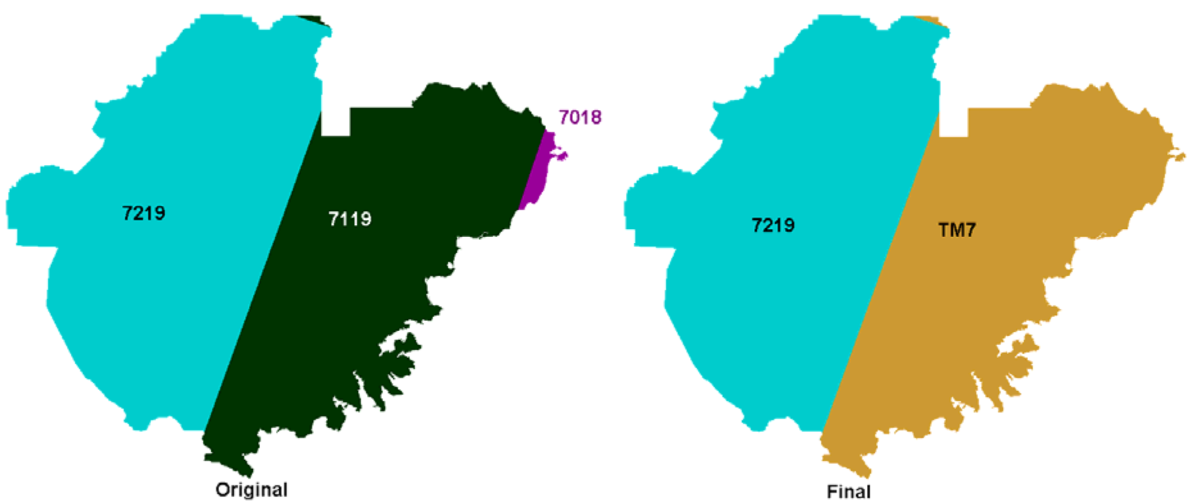


Figure 14: Original and Final Scene Extent

Unfortunately, the portion of the project area covered by scene 7018 was only about 1% of the total area, while scene 7119 covered approximately 52% of the project area. Scene replacement translated into a reduced variety (from 60 to 49) of the available surveyed training areas/classes. Particularly affected classes were “Open Spruce” and “Spruce/Deciduous,” whose available training areas were reduced from 6 to 2 and 9 to 0, respectively. Scene TM7 had 86 fewer training sites/classes used in final classifications than scene 7219. The overall result is that the pixel classification in the TM7 coverage may be more generalized due to fewer classes. In addition, the unsupervised classification also contributed a slightly larger percentage of pixels in this area than in 7219 – 15% of the pixels as opposed to 13%.

However, a clear benefit from scene replacement was the reduction in the confusion between "Spruce" and "Wet-Herbaceous" classes. The newer imagery allowed a cleaner separation of these classes during classification. However, this may have been caused by the reduction in available "Spruce" class training sites in the new imagery as there were only 2 "Spruce" classes in the replacement training data set, as opposed to 6 in the original training data set. Whatever confusion remained between these classes was controlled by means of the spruce mask, as described earlier. The 2000 imagery may also have enabled the removal of differences in senescence that may have affected the classifications, thereby ensuring a better classification performance of the training areas surveyed in August of 2000. Unfortunately, time did not allow for a pixel-by-pixel comparison between the original classification data and the final classification data to quantify these differences.

<u>Positional Accuracy of Results</u>

The replacement data was offset significantly from the original imagery. The apparent shift was approximately 3 pixels in an easterly direction and 5 pixels in a northerly direction. All replacement data were shifted to coincide, as best as possible, with the location of the original pixel data. Of particular interest was that the replacement imagery, in terms of positional accuracy and alignment to USGS DEM data, did not exhibit the same alignment that either the replacement imagery acquired from the EROS Data Center, or the original imagery reflected. Conversations with RGI indicate that it may not be possible for a third-party to achieve the same alignment without obtaining the same control data used by EROS Data Center for their internal processing of the imagery. This lack of consistency potentially compromises the spatial location of the image data and the resulting pixel classification maps. Without a conflicting data set that enabled the prompt identification of a sizable data offset, the user of such image data must be careful to check the location of the imagery with respect to its true position on the earth. Such discrepancies could dramatically effect the development of training classes; if training sites are accurately located with a GPS and then referenced to an inaccurately located image data set, the training data might reflect spectral data that was not really present at the training site. Ideally, EROS Data Center would change its internal image processing policy to use nearest neighbor instead of cubic convolution for re-sampling of LANDSAT scenes, so that a third-party firm would not be needed to orthorectify the imagery. However, until such time, future projects may have to either accept the 'convoluted' data or turn to private sources (IKONOS) and third-party companies for comparable or improved image products. Care will need to be exercised to assure that future data are accurately located.

Training Data Collection Efforts

<u>Uncertainty of Training Site Locations</u>

One of the major issues potentially impacting the results of this mapping effort is the reduced confidence in the true ground location of aerial survey training sites. The failure of the PLGR GPS units to receive and record spatial coordinates made the location of training sites much more difficult and potentially inaccurate. The need to "make do" with the aircraft's GPS unit for determining survey locations (as opposed to the more accurate PLGR unit) introduced an unknown amount of error in the assessment of the latter during image training efforts. The difference in format and datum between the two devices led to uncertainty during the first three days of data collection efforts, until the coordinate differences had been properly verified. This uncertainty was compounded by the fact that helicopter GPS values were shown to often be 100 meters, or further, from the true location of the training site in any direction. The potential impact of this uncertainty is that training data may have been collected in the wrong locations. The potential inclusion of erroneous nearby pixels during region growing operations, due to erroneous placement of inaccurate training site locations (shifted toward the edge of target areas) might have introduced unwanted confusion between apparently unrelated classes. This additional confusion required multiple reviews to resolve, and in some cases the problems never could be resolved. Table 20 lists those training areas excluded from final classification processes due to confusion that was likely due to the positional uncertainty of the training site location.

Although for different reasons, the same uncertainty of training site locations applied to some of the Dr. Thomas Smith (TSMITH) data collection sites that were included in the project. In these cases, uncertainty existed as to the actual extent (and location) of the training areas, particularly for those sites for which no photographic record was available. AKSO staff reviewed all TSMITH data and locations and those sites that were deemed inaccurately located were withheld from use as training data classes.

Table 20: Untrained Sites

trsite_id	cover_type_des
80726	Alder-Willow:Tall:Open
80803	Balsam Poplar:Open
80711	Balsam Poplar:Open
80703	Birch:Woodland
80402	Dwarf Shrub-Bare Soil
80634	Dwarf Shrub-Sedge
80636	Dwarf Shrub:Mixed
80722	Dwarf Shrub:Mixed
80810	Herbaceous:Freshwater
80601	Spruce:Open
80720	Willow:Low:Open
80502	Willow:Low:Open

In future projects of this nature, care must be taken to perform an actual test of GPS equipment under conditions identical to actual field sampling (as opposed to hearsay accounts of its use) to ensure proper functioning of such a critical instrument. While the PLGR units were used successfully on the ground, and even in a moving truck, during the days immediately prior to the aerial survey efforts, these conditions were not the same as those in the helicopter. Prior knowledge of this problem would have eliminated some of the problems and uncertainty of locations visited during the first few days of the project.

<u>Land Cover Data Estimates</u>

Another issue regarding the collection of field data regards the estimation of land cover component that formed the basis of the Land Cover Type Classification. The approach that was applied during this project was that the cover components would be estimated separately on the field data collection forms and the Land Cover Type class would be assigned on the basis of the type components. The land cover estimates were recorded on the aerial field survey sheets from a helicopter, as it slowly circled the area of interest. Two issues arose with this method of data collection. The first issue concerns verification of the cover estimates. None of the aerial estimates were field verified to determine their accuracy. While it is entirely possible to develop accurate estimates in this manner, these estimates should be verified to determine that they are accurate. It is doubtful that estimates of this nature can be much more accurate than $\pm$ 10 percent of the true value (on the average), as this is the typical level of error for a ground cover survey using 100 point transects. There were instances where either too much or too little cover may have been estimated for a particular species, possibly resulting in mistyping of the site. An illustrative example of this scenario was training site 80607. The cover components of this site included an estimate of 10% spruce cover, resulting in a type assignment of “Spruce:Woodland.” However, this site was confused with a number of Dwarf Shrub sites. A review of the photo record for this site indicated that there was very likely much less than 10% spruce cover. This training site, with less spruce cover would have been typed as a “Dwarf Shrub” type, a type which would have been quite consistent with its confused types.

While there will always be cases such as this one, where estimates span thresholds of very different types (tree type versus a shrub type), care must be exercised to accurately estimate the key cover components that determine the eventual type designation of an area.

Other concerns involved the estimation of field data without verification for consistency of the estimates. In several situations, all data were collected in the field only to find out several weeks later that the cover data did not add up to 100% cover, as was our stated goal. The plot was then reviewed with respect to all other field notes and photos to estimate the missing cover. Errors may have been introduced by fixing these anomalies in the office, rather than in the field as the data were collected. Fixing these errors in the office not only introduced possible error, but also took a substantial amount of time as all data were subjected to rigorous checks and editing. In addition, some types were assigned to data that were inconsistent with the cover components that had been estimated. This was evident, for example, in training site number 80401, where the assigned cover class was “Spruce:Woodland;” even though it was attributed with 25% of Spruce cover, which made it an “Open” type. These discrepancies nearly always involved type, species, or density class thresholds. These discrepancies are significant, as the field personnel use their calls or type assignments to categorize different land cover types with their visual observations – they train their eye to recognize the different types. If the assignment that has been made is wrong, and not corrected until after leaving the field, then the field observer may be associating field observations with incorrect types or density classes, thereby introducing potential errors into the data. For example, if a field observer estimates tree cover at 20 percent, and the observer is consistent in calling other areas with similar tree cover as 20 percent cover, that doesn’t make them 20 percent cover trees when they are really 30 percent cover. Being consistently inaccurate is not the same as being right. Similar errors can occur with species composition, size/height, and specie designation (aspen versus balsam poplar).

Although these situations and other obvious ones were detected, their existence indicated that methods used to estimate and record land cover data for each survey site may have allowed less obviously inaccurate data into the project’s database from which subsequent cover data were developed. Uncertainty regarding cover component data and class calls should be avoided through the incorporation of procedures to verify land cover data estimates on site in future projects. Possible solutions could include: ground verification of a certain percentage of aerial survey sites to provide verification of estimates and help the field observers calibrate their ‘eyes’; hovering directly over the center of each training site, and/or mounting a remote-controlled high-resolution stereo-capable film camera on the aircraft (which would obtain high resolution, stereo pairs for subsequent photo-interpretation to verify observations). In addition, field survey personnel should be trained using paired ground/air plots to recognize differences in cover, species, and size/height immediately prior to making the actual training site estimates. Problems concerning data records are another matter than could be avoided through the used of hand-held data recorders. Rather than log field estimates on field forms with written notes and numbers the use of data recorders is strongly recommended. These data recorders may be programmed to perform data validity checks to be certain that not only are all the data collected, but that the data are consistent with each other. All instances where the “bird’s-eye” cover estimates did not total 100% would be resolved before leaving the site. Inconsistent type calls would be flagged for review and correction in the field. Erroneous data would be avoided eliminating potential data inaccuracies as well as the time required to review and resolve these problems. Some of these problems were not uncovered until an inconsistency in the fidelity test or confusion report identified the problem. Elimination of erroneous data would have saved considerable efforts as field data tables were reviewed, corrected, and exchanged at least four or five times prior to resolving every last problem. These problems amounted to several weeks worth of effort over the course of the project in additional data review, verification, and translation of the data, as well as redoing all associated reports and analyses.

Aerial Survey Data Collection

A total of nine days were dedicated to collection aerial survey data. The first possible day was lost to weather, resulting in eight days of extensive aerial survey work. During these eight days a total of 32 hours of actual aerial survey time were logged, or an average of four (4) hours/day. Other flight time was spent (lost) due to refueling flights, transportation of ground personnel, and down time due to limited refueling opportunities. The average number of aerial survey training sites visited per day was 27. Of the total of 222 sites surveyed from the air, 210 (95%) were ultimately used in the supervised classification efforts.

Reduced aerial survey time resulted in fewer surveyed areas than was originally expected. This limitation in training area number came to bear in particular when 2000 imagery replaced the original scenes around which data collection efforts was planned and implemented. Training sites in the replacement imagery were limited due to virtually no scene overlap with scene 7219 and the fact that only two days of aerial survey efforts were conducted in the eastern portions of the project area. On the basis of these field data collection efforts and taking into account such factors as weather, safety, and fuel availability, as many as 40 sites could be visited during a full aerial sampling day (i.e., nearly 75% flight time dedicated to aerial training site surveying instead of 50%). One could conservatively estimate a 50% increase in the total number of aerial survey training areas, if such a schedule could have been followed, but this would require sacrificing other uses of the helicopter. This estimated level of productivity should be kept in mind for future projects for which a minimum number of training areas would be necessary and other data sources, such as the TSMITH ground plots, are not available to supplement data collection efforts.

Image Processing Issues

Water and Non-Water Confusion in Mountainous Regions

Image classification efforts were based on a number of training sites in order to reduce confusion and misclassification of data. Particularly noteworthy was the need to partition training sets into “water” and “non-water” training data sets to reduce the assignment of “Water” pixels in higher elevation, flat to gently sloping, non-water situations, as occurred in scene 7219 (this was a situation not related to the “terrain shadowing” dark pixels assigned Water values). While this situation was resolved in this manner, further analysis revealed the cause of this problem. A few “Water” training areas were collected in areas that inadvertently included very shallow water. These shallow-water training areas often ran up to the edge of the water bodies, and appear to have mistakenly absorbed training pixels from adjacent “barren” pixels along the shoreline of the water bodies during the region growing process. These training classes actually had a mix of water and barren spectral data that led to the classification of some valid “Barren” areas as “Water” type pixels. While removal of the water training areas resolved this situation in this case, care should be exercised to not include these very shallow water types if the training data comes ashore onto the neighboring “barren” pixels. In particular, in large bodies of shallow water this may be a serious concern, as the region growing processes more easily absorb these different edge pixels when the area size is larger (the larger size creates a large sample size and may easily absorb variant pixels without exceeding variance thresholds). Building training areas in riverine situations with combinations of shallow water and gravel is another situation to avoid. Splitting the training sets solved the problem this time, but less work is involved if greater care is taken in building the water training classes.

Water and Isolated Herbaceous Type Pixels

In certain water situations "Herbaceous" or "Sparse Vegetation" classes were assigned resulting in pixilated water bodies. In this situation, the vegetation cover attributed to the pixel data was emergent grasses and other aquatic vegetation in types that were dominated by water. Unfortunately, in the classification rules, "Herbaceous" or "Sparse Vegetation" types were designated if either the herbaceous cover was greater than 25% cover, or if the sum of the vegetation cover was greater than 15% cover. Therefore, a "Herbaceous" type could be assigned if the class had 25% herbaceous cover and 75% water, and a "Sparse Vegetation" class with only 15% vegetation cover and 85% water. These classification rules resulted in a few training classes that looked spectrally similar to water, but were classified as a non-water pixel class. In these situations, the classification thresholds may warrant adjustment if the assignment of these non-water classes is too much of a problem. As with some of the other pixelation issues, aggregation or modeling are other ways to address resolution of these issues if the end user finds the results distracting or confusing. The presence of the cover by component in the database record is a tremendous help in both identifying and resolving this situation.

Spruce/Wet Herbaceous Confusion

Throughout the image training process, another stubborn confusion between a few "Wet Herbaceous" and "Spruce" types became apparent. After undertaking exhaustive spectral training efforts to separate these types some confusion still remained. In order to resolve this issue spruce masks were developed, one for each scene. These masks were used to identify where spruce was thought to exist, as opposed to where it was not present. In this way, training areas containing spruce cover were excluded from areas known to lack any significant spruce cover. However, confusion remained between these two types within the spruce-inclusive areas of each scene. This was particularly evident in scenes 7219 and 7119, which contained larger numbers of spruce and wet herbaceous training areas (41 versus 13 and 20 versus 24 respectively). The newer TM7 mosaiced imagery did not show much confusion between these types. While this may have been due to the earlier acquisition date and lack of senescence in the "Wet Herbaceous" type pixel, this reduction may also have been due to the reduced number of available spruce training areas in the replacement imagery as compared to the original 7119 scene. Scene 7219 retained some of the Spruce/Wet Herbaceous confusion that prompted the use of spruce masks to isolate areas unlikely to contain spruce. "Spruce" inclusions within "Wet Herbaceous" areas should definitely be viewed with suspicion. Where "Spruce" pixels are isolated and surrounded by "Wet Herbaceous" types, an aggregation or modeling process could be used to identify and resolve these inconsistencies.

Bibliography

Civco, D. 1989. "Topographic Normalization of Landsat Thematic Mapper Digital Imagery," Photogrammetric Engineering and Remote Sensing, Vol. 55(9), pp. 1303-1309.

Colby, J., 1991. "Topographic Normalization in Rugged Terrain," Photogrammetric Engineering and Remote Sensing, Vol. 57(5), pp. 531-537.

Geographic Resource Solutions, 1993. Klamath Province Mapping Pilot Study Report. Arcata, CA. Prepared for the California Department of Forestry and Fire Protection. 78 pp.

Swain, P. and S. Davis, 1978. Remote Sensing: The Quantitative Approach. McGraw-Hill Book Company, New York, NY.

Viereck, L.A. et al., 1992. The Alaska Vegetation Classification. U.S. Department of Agriculture. Forest Service. Pacific Northwest Research Station General Technical Report: PNW-GTR-286.

Appendices

Appendix A: Katmai Mapping Project Aerial Survey Form

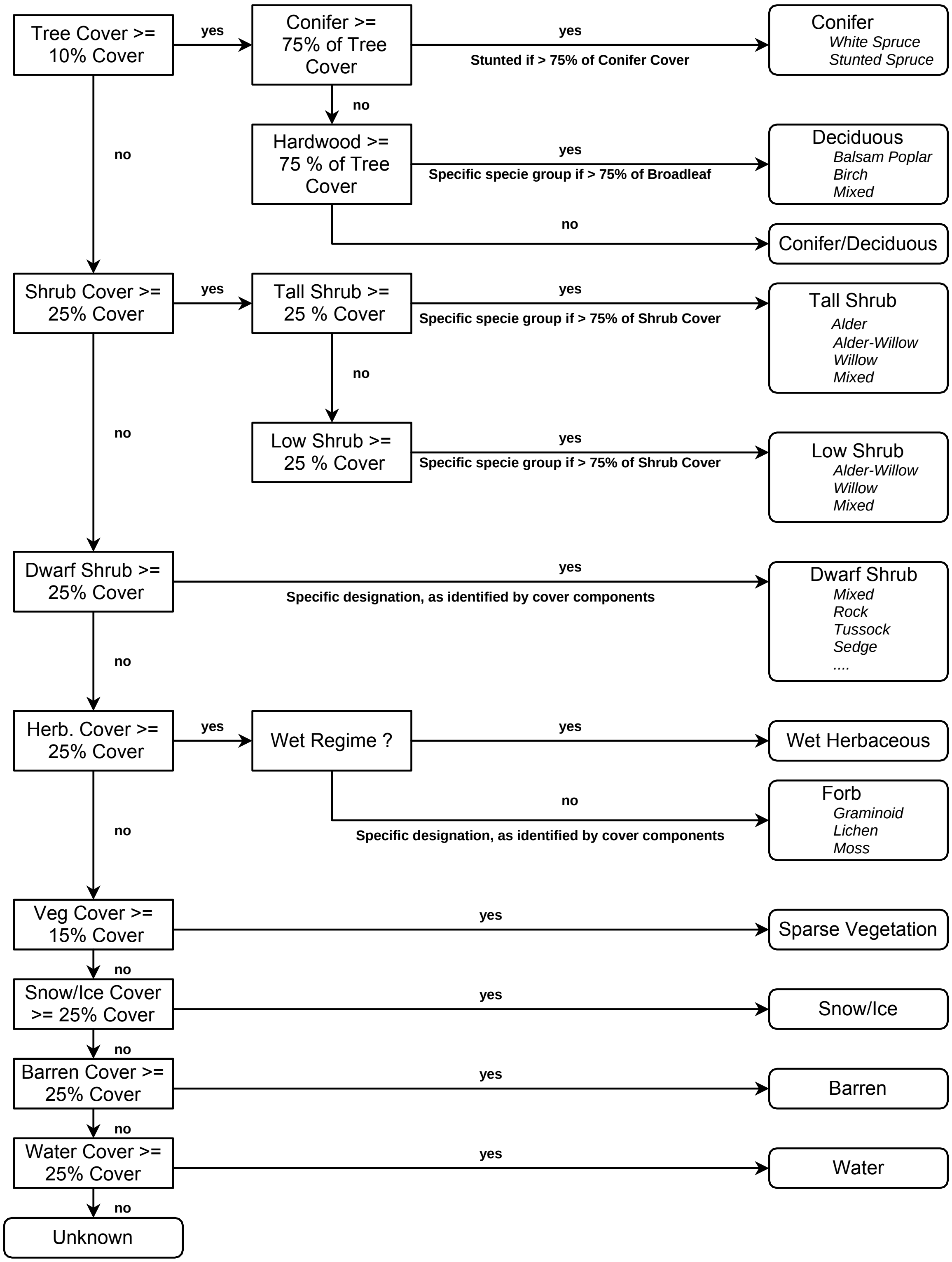
Air Photo#: **081010**	Map: **C11**	Polygon code: 081010	Lat/Long: 8/10/00
Landcover code: **15**	Landcover class name:_______ **Open Balsm Poplar**	Azimuth:	Slope (0-100 °):
Surveyors: **S.Klein K.Stumpf** S.Henricy J.Grunblatt K.Boggs B.Koltun		Photos: Roll #:	Frame #s: YES
Dig photo: sess. #:	photo #:	azimuth:	
Modifier: Riparian Alpine Subalpine Boreal Tussock Bog Meadow			Video: YES
Hydrologic Regime (circle one): Dry - **Mesic** - Wet - Aquatic		% Tussocks:	Landform:

Hgt (m)	Tree	%	Tall shrub (>1.5m)	%	Low shrub (1.5m >20cm)	%	Dwarf shrub (<20 cm)	%	Herbaceous	%
	Aspen	10	Alder	25	Devil's Club	10			Fern	5
	Poplar	45							Other	5
									Lichen	
									Moss	
									Sand/Gravel	
									Mud	
									Bedrock	
									Rock (Talus)	
									Water	
	Total	**55**	**Total**	**25**	**Total**	**10**	**Total**	**0**	**Total**	**10**
									GRAND TOTAL	**100**

	Spruce		**Mixed forest**		**Low shrub**		**Herbaceous**
1	Closed White Spruce	17	Closed Mixed	25	Closed Low Alder-Willow	40	Dry/Mesic Graminoid
2	Open White Spruce	18	Open Mixed	26	Closed Low Shrub Birch-Willow	41	Dry/Mesic Graminoid Forb
3	Open Black Spruce	19	Woodland Mixed	27	Closed Low Willow	42	Tussock
4	Open Stunted Spruce (black spruce)			28	Closed Low Shrub Birch-Ericaceous Shrub-Willow	43	Wet Salt Marsh Herbaceous
5	Open Spruce Lichen			29	Closed Low Shrub Birch	44	Wet Freshwater Herbaceous
6	Woodland Spruce			30	Open Low Alder-Willow	45	Aquatic
7	Woodland Stunted Spruce			31	Open Low Alder		
8	Woodland Spruce Lichen		**Tall shrub**	32	Open Low Shrub Birch-Ericaceous Shrub-Willow		**Other**
		20	Closed Tall Alder	33	Open Low Willow	46	Sparse Vegetation
			-Willow	34	Open Low Shrub Birch-Ericaceous Shrub	47	Marine Mudflat
	Deciduous forest	21	Closed Tall Alder			48	Sand/Gravel
9	Closed Birch	22	Closed Tall Willow			49	Bedrock
10	Closed Aspen	23	Open Tall Alder			50	Snow/Ice
11	Closed Balsam Poplar	24	Open Tall Willow		**Dwarf shrub**	51	Shadow/Indeterminate
12	Closed Mixed Deciduous			35	Dryas Dwarf Shrub	52	Silty Water
13	Open Birch			36	Mixed Dwarf Shrub	53	Clear Water
14	Open Aspen			37	Mixed Dwarf Shrub-Sedge	54	Marine Water
15	**Open Balsam Poplar**			38	Mixed Dwarf Shrub-Lichen	55	Human Development
16	Open Mixed Deciduous			39	Mixed Dwarf Shrub-Rock	56	Burn Site

INCLUSIONS/STRINGERS:	**COMMENTS:**

Appendix B: Katmai Mapping Project Land Cover Type Classification



Type Assignment Fortran Source Code:

```
c
c           Katmai type rules
```

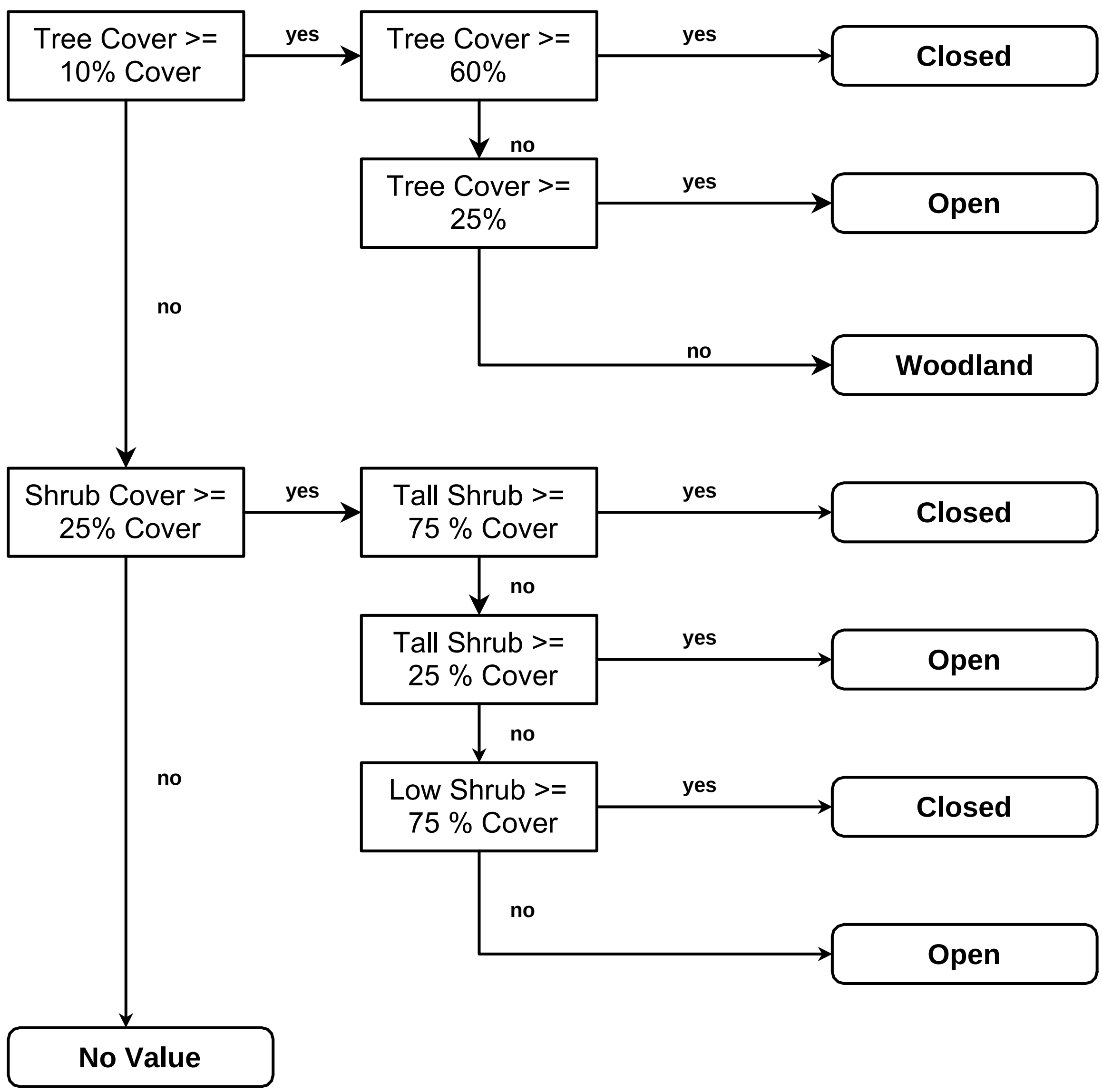
```
c
       type='Unk'
       maxden=0
c
c          tree ??
c
       if ( (cfsum + hwdsum) .ge. 25.0 )then
c
c          have a tree type
c
        maxden=0
        densmx=0
c
        do i = 1,maxtree
         if(sumcov(i,m1).gt.densmx)then
          densmx=sumcov(i,m1)
          maxden=i
         end if
        end do
c
c          conifer ??  set default to mixed spruce type
c
        if( denscf .gt. 75. ) then
         type='UnP'
c
c   white spruce
c
         if(sumcov(1,m1)/denscf .ge. 0.667)then
            type='PGl'
          else
c
c   black/stunted spruce
c
          if(sumcov(2,m1)/denscf .ge. 0.667)then
            type='pgs'
c
c   uspecified spruce - leave default as is
c
          endif
         endif
        else
c
c          hardwood
c
         if( denscf .lt. 25.)then
          type='Hwd'
         else
c
c          hardwood/conifer mix
c
           type='PHw'
c
         endif
        endif
c
c          end of tree
c
       else
```

```
c
c          non-tree type
c
c          shrub
c
        if ( shrsum .ge. 25. )then
         maxden=0
         densmx=0
c
c set default to low shrub for katmai
c
         type='LSh'
         ttlsum=sumcov(26,m3)+sumcov(27,m3)+sumcov(28,m3)
         tlwsum=sumcov(29,m3)+sumcov(30,m3)+sumcov(31,m3)+sumcov(32,m3)

         if(ttlsum .ge. 25.)then
          type='TSh'
          mmnshr=26
          mmxshr=28
         else
          if( tlwsum .ge. 25.)then
            type='LSh'
            mmnshr=29
            mmxshr=32
          else
           if( (tlwsum+ttlsum) .ge. 25. .and. sumcov(33,m3).lt.25.)then
             type='LSh'
             mmnshr=26
             mmxshr=32
           else
             type='DSh'
             mmnshr=33
             mmxshr=33
           end if
          end if
         end if

         do i = mmnshr,mmxshr
          if(sumcov(i,m3).gt.densmx)then
           densmx=sumcov(i,m3)
           maxden=i
          end if
         end do
c
        else
         if(hrbsum .ge. 25. )then
           maxden=0
           densmx=0
c
c set default to forb for katmai
c
           type='Frb'
           do i = minherb,maxherb
            if(sumcov(i,m3).gt.densmx)then
             densmx=sumcov(i,m3)
             maxden=i
            end if
           end do
```

```
         if(maxden .eq. 36)type='Grm'
         if(maxden .eq. 37)type='Frb'
         if(maxden .eq. 38)type='WMo'
         if(maxden .eq. 39)type='Lch'
c
        else
c
c           Sparse Vegetation
c
         if( (shrsum+hrbsum+cfsum+hwdsum+aqusum) .ge. 15. )then
          densmx=0
          maxden=0
          type='SVg'
          do i = mincon,maxaqu
           if(sumcov(i,m3).gt.densmx)then
            densmx=sumcov(i,m3)
            maxden=i
           end if
          end do
c
c major component > 10 percent ?
c
          if(densmx .lt. 10.)then
c              default to misc forb
           maxden=40
c
           if(shrsum .ge. hrbsum.and.
     *        shrsum .ge. aqusum )maxden=32
c
           if( (hwdsum+cfsum).ge.shrsum.and.
     *         (hwdsum+cfsum).ge.aqusum.and.
     *         (hwdsum+cfsum).ge.hrbsum )maxden=24
c
           if(hwdsum .ge. cfsum .and.
     *        hwdsum .ge .aqusum .and.
     *        hwdsum .ge. shrsum .and.
     *        hwdsum .ge. hrbsum )maxden=25
c
           if(cfsum .ge. hwdsum .and.
     *        cfsum .ge .aqusum .and.
     *        cfsum .ge. shrsum .and.
     *        cfsum .ge. hrbsum )maxden=20
          endif
c
         else
c
c        non-veg type
c
         densmx=0.
         type='???'
c
c

         if(barsum.gt.densmx.and.barsum.ge.25.)then
           densmx=barsum
           type='Bar'
           typmax=0.
           do i = minbar,maxbar
```

```
            if(sumcov(i,m3).gt.typmax)then
              typmax=sumcov(i,m3)
              idxtyp=i
            end if
          if(idxtyp .eq. 57)type='SnG'
          if(idxtyp .eq. 58)type='Bar'
          enddo
        endif
c
        if(watsum.gt.densmx.and.watsum.ge.25.)then
          densmx=watsum
          type='H2O'
          typmax=0.
          do i = minwat,maxwat
            if(sumcov(i,m3).gt.typmax)then
              typmax=sumcov(i,m3)
              idxtyp=i
            end if
          enddo
        endif
c
c          undefined - blank out
c
        if(othsum.gt.33.33)then
          densmx=0
          type='???'
          typmax=0.
          do i = minundf,maxundf
            if(sumcov(i,m3).gt.typmax)then
              typmax=sumcov(i,m3)
              idxtyp=i
            end if
          if(idxtyp .eq. 59)type='ClS'
          enddo
        endif
c
        endif
       endif
      endif
     endif
c
       endif
     endif
c
c      end of rules
c
```

Appendix C: Katmai Mapping Project Land Cover Density Classification



Appendix D: Table Item Listings and Definitions

Items for table katmai_lcover.vat:

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC	INDEXED?
1	VALUE	4	10	B	-	Indexed
5	COUNT	4	10	B	-	-
9	SOURCE	20	20	C	-	-
29	SPEC_CLASS	50	50	C	-	-
79	GRIDCLVAL	11	11	I	-	-
90	MAJOR_CLASS	50	50	C	-	-
140	REGIME	20	20	C	-	-
160	MODIFIER	20	20	C	-	-
180	REMAP_VAR	6	6	I	-	-
186	TYPE	6	6	C	-	-
192	PREDOMINANT_COMP	20	20	C	-	-
212	PR_COMP_COVER	8	23	F	4	-
220	COVER_CLASS	10	10	C	-	-
230	TREE_COVER	8	23	F	4	-
238	OTHER_VEG_COVER	8	23	F	4	-
246	CONF_COVER	8	23	F	4	-
254	HDWD_COVER	8	23	F	4	-
262	SHR_COVER	8	23	F	4	-
270	TSH_COVER	8	23	F	4	-
278	LSH_COVER	8	23	F	4	-
286	DSH_COVER	8	23	F	4	-
294	HRB_COVER	8	23	F	4	-
302	BAR_COVER	8	23	F	4	-
310	OTH_COVER	8	23	F	4	-
318	TRSITE_ID	11	11	I	-	-
329	CALC_CLASS	50	50	C	-	-
379	ALNUS_COVER	8	23	F	2	-
387	SALIX_COVER	8	23	F	2	-

Items for table cover_info.dat:

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC	INDEXED?
1	TRSITE_ID	11	11	I	-	-
12	SOURCE	20	20	C	-	-
32	SPECIES	50	50	C	-	-
82	SPCODENUM	6	6	I	-	-
88	SPCODE	10	10	C	-	-
98	COVER	8	23	F	4	-
106	STRATA	10	10	C	-	-
116	COMMENT	50	50	C	-	-
166	IMA_ID	6	6	I	-	-
172	GRIDVAL	11	11	I	-	-

Items for the table trsites.pat

COLUMN	ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC	INDEXED?
1	AREA	8	18	F	5	-
9	PERIMETER	8	18	F	5	-
17	TRSITES#	4	5	B	-	-
21	TRSITES-ID	4	11	B	-	Indexed
25	TRSITE_ID	11	11	I	-	-
36	SOURCE	20	20	C	-	-
56	ISO_CLASS	11	11	I	-	-
67	ELEVATION_FT	6	6	I	-	-
73	SLOPE_DEGREES	6	6	I	-	-
79	ASPECT	6	6	I	-	-
85	X_COORD	8	23	F	4	-
93	Y_COORD	8	23	F	4	-
101	PHOTO_ID	8	8	C	-	-
109	SCENE_IND	4	4	C	-	-
113	IMA_ID	11	11	I	-	-
124	COVER_TYPE_DES	40	40	C	-	-
164	GRIDVAL	11	11	I	-	-
175	GRIDCLVAL	6	6	I	-	-
181	HABITAT	55	55	C	-	-
236	REGIME	20	20	C	-	-
256	MODIFIER	20	20	C	-	-
276	MAJOR_CLASS_DES	40	40	C	-	-
316	SCN_7219	1	1	C	-	-
317	SCN_7119	1	1	C	-	-
318	SCN_7018	1	1	C	-	-
319	SCN_TM7	1	1	C	-	-

Definitions:

ALNUS_COVER	The estimated cover of Alnus sp.
ASPECT	The estimated aspect(azimuth) at the training site location
BAR_COVER	The estimated cover of barren area
CALC_CLASS	The Land Cover Class estimated through application of the Land Cover Classification rules
COMMENT	Comment field
CONF_COVER	The estimated percent of tree cover comprised by conifers
COUNT	The frequency of the pixel value in the map
COVER	The estimated percent land cover of a component
COVER_CLASS	The density class of the Land Cover Type: Woodland = 10 - 24.9% tree cover Open = 25 - 74.9% shrub cover = 25 - 59.9% tree cover Closed >= 75% shrub cover >= 60% tree cover
COVER_TYPE_DES	The field estimated Land Cover Type
DSH_COVER	The estimated cover of Dwarf Shrubs
ELEVATION_FT	The estimated elevation at the training site location
GRIDCLVAL	Major class grid value - numeric code for MAJOR_CLASS
GRIDVAL (VALUE)	The pixel value of this class in the final pixel map
HDWD_COVER	The estimated percent of tree cover comprised by hardwoods
HRB_COVER	The estimated cover of herbaceous plants
IMA_ID	For TSMITH sites - the original IMA ID number
ISO_CLASS	The original iso_class number used during candidate site development and selection
LSH_COVER	The estimated cover of Low Shrubs
MAJOR_CLASS	The major land cover class descriptive name - a generalized Land Cover Type
MODIFIER	Modifier (morphologic) Code
OTHER_VEG_COVER	The sum of all non-woody vegetation cover (shrub + forb)

OTH_COVER	The estimated other cover, including water
PHOTO_ID	The digital photo name for the image of the training site
PREDOMINANT_COMP	The predominant component of the land cover - specific to species or group of similar species
PR_COMP_COVER	The estimated percent land cover of the
REGIME	Hydrologic Regime
REMAP_VAR	ReMap Value - scratch column used to store sql derived values and generate different map representations
SALIX_COVER	The estimated cover of Salix species
SCENE_IND	Indicator value to identify in which scenes a training site was trained and used: 0 = not trained 1 = trained Scenes are indicated by the column of the value - From left to right: 7219, 7119, 7018, and TM7.
SCN_7219	The indicator that the site was used in this scene (0=No/1=Yes)
SCN_7119	The indicator that the site was used in this scene (0=No/1=Yes)
SCN_7018	The indicator that the site was used in this scene (0=No/1=Yes)
SCN_TM7	The indicator that the site was used in this scene (0=No/1=Yes)
SHR_COVER	The sum of Tall and Low Shrub cover
SLOPE_DEGREES	The estimated slope(degrees) at the training site location
SOURCE	Source of the training data: GRS_00 = Aerial Survey AKSO_00 = AKSO ground survey TSMITH = Dr. Thomas Smith IMA data GRS/AKSO_01 = supplemental sites GRS added UNSUP = unsupervised classes
SPECIES	The species/name of the land cover component
SPEC_CLASS	Specific class descriptive name assigned in the field

SPCODENUM	The species code number of the land cover component
SPCODE	The species code symbol of the land cover component
STRATA	The vegetation profile of the land cover component: TR_CNFR = Conifer TR_BDLF = Broadleaf TS = Tall Shrub LS = Low Shrub DS = Dwarf Shrub H = Herbaceous M = Moss O = Other
TREE_COVER	The estimated percent tree cover (density)
TRSITE_ID or TRSITE-ID	Training site id number - the unique identifier of the training site
TSH_COVER	The estimated cover of Tall Shrubs
TYPE	Land Cover Type Code: PGl = Spruce Pgs = Stunted Spruce PHw = Conifer/Broadleaf Hwd = Broadleaf/Hardwood TSh = Tall Shrub LSh = Low Shrub DSh = Dwarf Shrub Hrb = Forb:Herbaceous Grm = Graminoid Lch = Lichen WMo = Wet Moss SVg = Sparse Vegetation Bar = Barren SGl = Snow/Glacier H2O = Water Unk = Unknown
VALUE (GRIDVAL)	Pixel value used to associate grid to attribute tables
X_COORD	The X coordinate value (Albers Projection) at the training site location
Y_COORD	The Y coordinate value (Albers Projection) at the training site location

NPS D-118, June 2003

National Park Service
U.S. Department of the Interior

Natural Resource Program Center
1201 Oakridge Drive, Suite 150
Fort Collins, CO 80525

www.nature.nps.gov

EXPERIENCE YOUR AMERICA ™