INTRODUCTION

Creating maps, cartography, pre-dates most other forms of human communication or expression (Clarke 1995). The technical use of maps by professionals is well known but use in everyday analysis is beginning to grow rapidly. This growth is due to the power, widespread availability, and speed of modern computers, specifically geographic information systems (hereinafter GIS).

Although there are many (sometimes conflicting) definitions of GIS, one of the more useful statements was put forth by Cowen (1988) who stated, "...GIS is best defined as a decision support system involving the integration of spatially referenced data in a problem solving environment". GIS are powerful in natural resource decision making because of their ability to store, retrieve, update, and create an infinite number of thematic maps. Key to this process is creating, aggregating, and assimilating spatial information. Essential to this system are software, hardware, personnel, and space (Koeln et al. 1994). The availability of these components, especially personal computers, has made GIS very commonplace.

GIS excel at facilitating the examination of interactions. The spatial interactions of many abiotic and biotic factors that occur in ecosystems, at a variety of scales, are very important. GIS are a critical tool for making analyses and presentations in ecosystem science and management. This tool has been, and is being, used in a variety of disciplines where spatial planning and decisions are involved (Butler et al. 1987, Harris et al. 1978, FAO 1988, Moreno and Heyerdahl 1992). Koeln (1994) presented an excellent overview of geographic information systems for wildlife science.

One basic thematic map or layer, in GIS terms, is land cover. According to Burley (1961), land cover is all of the vegetation and human constructs covering the surface of the land (e.g., open water, shrubs, developed, etc.). "Land cover" and "land use" are terms that are often used together or synonymously. Land use differs from land cover in that the former implies a human use related to the land (Anderson et al. 1976). “Land use” can be thought of as a sub-class of “land cover”.

Current land cover data are vital to many applications including: making basic habitat assessments, constructing statewide wildlife distribution models, delineating specific vegetative communities, calculating soil loss, evaluating water quality within and between watersheds, and monitoring wilderness management. The list of categories to be mapped is determined through the objectives of the mapping effort. This thesis describes such a mapping effort, one of presenting a basic land cover map, along with methods useful for natural resource applications in the Commonwealth of Virginia. Emphasis was placed on general vegetation types, as opposed to more detailed categories (e.g., oak dominated montane forest, industrial development, row crop, etc.).

Objectives

The specific objectives of this project were:

Primary:

To produce a relatively current land cover map of Virginia using Landsat Thematic Mapper (hereinafter TM)

Secondary:

A. To demonstrate the feasibility of producing a fine-scale map covering a relatively large area, in a reasonable amount of time using a personal computer.
B. To provide working solutions to the common problems of haze/clouds, topographic shadows, and obtaining ground points in reflectance-based remote sensing.
C. To present baseline data on the areas and proportions of Virginia in major land cover types.

Justification

To approach the growing problems of natural resource management, spatially explicit information about physical, biotic, and human factors must be available in a variety of geographic and temporal scales (USFWS 1976). Local planners and managers require detailed knowledge of the region for which they have responsibility as well as information on the regional, state, and national levels (Aldrich 1979). Land cover data are essential at several scales.

Knowing the spatial area, amount, and arrangement of forests, water, wetlands, shrubs, agriculture, and development can be invaluable in a variety of wildlife and natural resource applications. These applications range from managing individual species, to accessing the quality and health of watersheds, to assessing biodiversity. Wilson (1985) wrote that the worst disaster of the present day is the loss of genetic and species diversity via habitat destruction. He emphasized the need for large-scale habitat inventories. Inventories can provide information about each cover-type, including relative type abundance, average or range of area of stands, and where in the state the most area of each vegetation type occurs. A map showing the spatial arrangements is crucial. Accurate land cover maps that concentrate on vegetation in all areas are desperately needed to protect the rich natural heritage of the nation (Scott et al. 1987). The maps produced through this effort could be a piece of the information needed for a variety of management activities in Virginia including the promotion of faunal diversity. To my knowledge no such map is complete or publicly available for the entire state of Virginia.

There are maps showing distributions of individual species of vegetation throughout Virginia (Harvill et al. 1977). There are also large-scale maps showing potential vegetative communities, such as Bailey's ecoregions (1976), and Kuchler's potential natural vegetation (1964). The difference between the above listed maps and the map developed herein is its timeliness, its sampling intensity, its use at several scales, its resolution, and its basis for classification. This map uses the most recent data available about the land cover of Virginia. I based classification on actual land cover, concentrating on the broad structural vegetation types.

Remote sensing analysis of the land cover of the Commonwealth is likely to fulfill the needs of agencies and landowners for knowing accurate and current information about the quantity, quality, and spatial distribution of the natural resource base (Lachowski et al. 1992). Before agency researchers can meaningfully appraise environmental conditions, estimate change, or conceive of future needs, they must first know what is present, its quality, and where it is located.

Scope

This thesis presents a statewide land cover map for the entire commonwealth of Virginia. A 1 kilometer buffer around Virginia has been used to provide information just beyond the state-line. Within this statewide context, it is desirable to provide the finest spatial data possible.   Satellite imagery made this feasible.

Satellite remote sensing instruments are an excellent base from which to observe large-scale ecosystems systematically, consistently, and synoptically (Wickland 1991). Landsat satellite instruments provide information from wide views with relatively fine spatial detail (Campbell 1996). Of the Landsat instruments, the Thematic Mapper (TM) has several advantages: (1) a high signal-to-noise ratio, (2) narrow bands of radiometric data, (3) high geographic accuracy, and (4) many spectral bands (Scott et al. 1993). TM also offers high spatial resolution. TM has a pixel size of 30m x 30m (Campbell 1996), a size which will be retained as the minimum mapping unit (this resolution will give a detailed picture of the state of basic habitat in Virginia).

It is hoped that this thesis will provide insight into the current condition and status of the land cover of Virginia. The dates of the data used range from 1991 to 1993 with most of the imagery from 1992. These were the most current available data. I believe this information will serve scientist and resource managers of the Commonwealth well for the immediate future, but new versions are likely to be needed as land-use changes.

LITERATURE REVIEW

Habitat

The meaning of habitat, as the word is used herein, is fundamental to the concept of, need for, and procedures used to create a wildlife-oriented land cover map for Virginia. Habitat has been defined many ways. Several current writers define habitat simply as the area where an animal lives (Gilbert and Dodds 1992, Morrison et al. 1992). Upon closer examination the idea of habitat can become very complex.

Hutchinson (1965) promoted the concept of *niche*, or multi-dimensional hypervolume, where each dimension represents an environmental factor influencing an organism. Klopfer (1969) who used the term umwelt, which means the totality of an organism's surroundings, to describe habitat, expanded this idea. Giles (pers. comm.) argued for a similar standard of completeness with the term "faunal space." Faunal space includes abiotic factors, landscape factors, plants and their conditions, animals themselves, space (volume), humans along with human values, and the past -- all of which make up the environment in which a population fulfills its life requisites or has its expansion limited -- over a stated period. Land cover, previously defined as the vegetation or human-made constructs covering the Earth's surface, is one component of an organism's niche, umwelt, faunal space, or habitat.

Leopold (1933), in his classic work *Game Management*, did not defined habitat but stated that an area of land is habitable by an individual animal if it offers suitable places for hiding, feeding, resting, playing, sleeping, and raising young. Leopold did define "environmental types" as areas of food and cover. As an example, he established that there could be between 4 and 40 environmental types for northern bobwhite quail (*Colinus virginianus*), depending on the refinement needed. The 4 basic types required by quail were revealed as woodland, brushland, grassland, and cultivation. These, I believe, are equivalent to land cover types, land units that are readily perceived, distinctively different, large, with easily agreed upon names.

Land cover (or Leopold's environmental types) is just one component of an animal's niche or faunal space. Other biotic and abiotic information layers are essential (e.g., elevation, aspect, potential evapotraspiration, solar radiation, land use, soils, disturbance) along with landscape variables (e.g., fragmentation, patch size, distance to water, distance to roads), all within an information system, to get an approximate understanding of habitat.

Another idea that is critical when discussing habitat is scale (Wiens 1989). Wildlife responds to habitat on a variety of levels. Micro, macro, and landscape scale are 3 broad habitat scale levels. I shall attempt to delimit each level because each has been used and defined in several ways. Microhabitat variables include such things as stem density, vertical cover, and soil moisture. For example, a bird may roost on a particular branch to avoid predators and to adjust its body temperature, i.e., exist within a microhabitat. Land cover, soil type, and aspect are considered herein as macrohabitat variables. The same bird may be interpreted to have chosen branches in a coniferous forest on a south-facing slope, i.e., habitat at a larger scale, probably to achieve desirable thermoregulation, thus macrohabitat. Landscape habitat factors might include nearness to water, distance to forest edge, volume of habitat patch, and migratory regions. Continuing the example, the bird might choose an area within that coniferous stand away from the edge of the forest, thereby avoiding nest predators or parasites, and select an area close to sources of life requisites such as water, i.e., landscape-scale habitat preference.

Land cover can also be delineated on several scales. The land cover map in this thesis, derived from statewide satellite imagery, represents macro and landscape factors of habitat in Virginia. An example of the use of both scales to analyze habitat, through land cover data, is the work of Palmeirim (1988). He used both cover types and landscape characteristics (derived from Landsat TM satellite imagery) to map the distribution, suitability of habitat, and density of bird species in Northeast Kansas.

Similar to Palmeirim’s (1988), the results presented herein can be used at both (macro and landscape) habitat scales. Also similar to Palmeirim (1988) and others (e.g., Scott et al. 1993 and Rohde 1978) this thesis demonstrates that an efficient and consistent way to get habitat information at several scales over a large area is through satellite remote sensing (e.g., Landsat TM).

Remote Sensing and Habitat

Remote sensing means gaining useful information about the Earth's surface via instruments at a distance, generally through emitted or reflected electromagnetic energy (Campbell 1996). According to Best (1982), Dr. Paul Dalke in 1937 first applied remote sensing to wildlife management, using simple aerial photography. Since then, both fields have expanded. Now resource professionals, practicing ecosystem and multi-use management use various methods of remote sensing, from airborne videography to satellite hyperspectral sensors.

Information on quality and distribution of habitat or the status of current wildlife populations are the usual objectives of wildlife-oriented remote sensing endeavors (Best 1982). Research, such as Scheierl and Meyer's (1976) study of waterfowl habitat and Hagen and Meyer's (1977) investigation into black bear habitat, use remote sensing to quantify and evaluate overall habitat. Other researchers have used remote sensing to find areas that are critical to certain species. One example of such a study is Aronoff's (1982) search for moose winter habitat. Zack (1983) presented a review of the early use of satellite remote sensing to analyze habitat.

Another wildlife application of remote sensing is direct population estimation (Lancia et al. 1994). This has been done via spectrometers (Payne 1984), infrared scanners (Brooks unpublished), airborne videography (Anthony et al. 1995), or visual counting (Overton 1969).

Of all of these uses, information about land cover (a large component of faunal space) is one the most important output of remote sensing.   Wildlife managers can use mapped information and data along with species-habitat association models to evaluate current habitat conditions, plan future objectives, and model the effects of specific management activities (Giles 1990). One such book of habitat-species relationships is Hamel's (1992) *Land Manager's Guide to the Birds of the South*. This management aid relies mostly on land cover, i.e., forest type, to describe habitat suitability for bird species.

Other analyses of bird habitat have used both on-site land cover and landscape variables (Flather and Sauer 1996, Robbins et al. 1989, Lynch and Whigham 1984, Whitcomb et al. 1981). The ability to develop landscape metrics adds the power of knowing the status of an area, utilizing information on its surroundings, and making site evaluations to plan for wildlife. Landscape metrics may include such variables as fragmentation, edge volume or area, patchiness, interspersion, juxtaposition, and patch size (Trani 1996).   These measurements represent some of the spatial relationships of a site in comparison to other sites. The typical land cover map is a two-dimensional representation of macro-habitat. It seems impossible to represent accurately and precisely an infinitely complex multi-dimensional world as a two-dimensional land cover map or as a series of one-dimensional landscape metrics. An alternative is to attempt to classify and group this information and present the classes spatially in maps.

Classification and Classification Systems

Classification, according to Platts (1980), is the process of "...identifying, labeling, ordering, or grouping a disaggregated set of objects based on predetermined similarities or differences". This action renders order from chaos and separates the world into workable units (Bailey et al.1978). "Workable units" means significant groups that can be remembered and easily analyzed with other interrelated grouped variables. The complex structure, relationships, and dynamics of faunal space make classification a necessity. Pfister and Arno (1980) stated that vegetation classification is artificial, determined through objectives. The needs for developing classes or groups of information are dependent upon the objectives, the uses to which the classes may be put. An objective in wildlife habitat classification is to discriminate and categorize lands having similar value for each wildlife population (Williamson 1981). This thesis separates Virginia into land cover types, each with potentially similar value to many wild animal species.

There are several dangers in classification. Williamson (1981) discussed these in depth. Two problems about which Williamson warned were the Boundary Problem and the Obsolescence Problem. These are very relevant to land cover classification and mapping. The Boundary Problem is the dilemma of trying to place continuous data into discrete categories. In land cover mapping, this problem is apparent when trying to put the continuous successional stages of herbaceous, shrubby, coniferous forest, mixed coniferous/deciduous forest, and deciduous forest into their respective classes. There are seldom definite borders between these vegetational stages (except under extreme management). There may be differences in ages as well as in rates of succession on a site. Yet, each may offer different use or value as wildlife habitat for different species. All may be worthwhile to map.

Time further complicates classification of an area. The Obsolescence Problem is the issue of ever-changing categories and items within these categories. One of the largest problems in land cover mapping is not that the classes will change but that the landscape is continuously changing. By the time an area is classified and labeled, its land cover may have been altered by people or by normal ecological succession. Most maps are obsolete by the time they are printed (Giles pers. comm.). Such obsolescence does not mean that classifying and mapping land cover is not a worthwhile task. Classifying complex systems is a prerequisite of concept formation (Gilmour 1951). Classifying, mapping, and quantifying ecosystems are required processes for making informed policy decision and improving management (Gimbarzevsky 1978, Udo de Haes and Klijn 1994). Classification, at least forming temporary classes, is necessary, yet difficult.

It is impossible to get one ideal land classification (Anderson et al. 1976). The classes and their structure are determined by the perceived objectives for the final product. Land classifications in the United States have been created for wetlands and open water (Cowardin et al. 1979), potential natural vegetation (Kuchler 1964), ecoregions (Bailey 1976, Omernik 1986), land use/land cover (Anderson et al. 1976), and forested lands (Erye 1980). Comprehensive terrestrial land cover schemes are being developed continually (TNC 1994, Driscoll et al. 1984, Jennings 1993). The recently formed Federal Geographic Data Committee (FGDC) is attempting to link these systems by identifying national standards for a total vegetational land cover classification (FGDC 1996). Other attempts have been made at regional ecosystem (or habitat) classification (Mueggler and Stewart 1980, Smalley 1982, Alexander 1985, DeVelice et al. 1986, Cooper et al. 1991). Land cover, especially vegetative cover, is an important cog in these comprehensive systems. Other factors typically included are climate, topography, disturbance (or human influence), and hydrology.

Two of the above classification systems relate directly to statewide land cover mapping for wildlife. These are the FGDC's vegetation classification and Anderson et al. (1976) land use and land cover classification. Both of these systems are hierarchical, providing several levels of detail or scales (e.g., global to a local watershed). Areas can be separated into categories of increasingly finer detail or aggregated into general classes. This design makes finding equivalent categories in different classification schemes possible by increasing the potential for finding equal or analogous classes. The above mentioned systems have a few general upper-level categories, several intermediate tiers, and open-ended detailed lower levels. These open-ended categories allow some freedom in mapping areas to meet objectives of local projects.

The FGDC classification serves primarily for vegetation communities. Consistent mapping between natural resource agencies is its objective. The proposed scheme integrates the vegetation systems of UNESCO (1973), Driscoll et al. (1984), TNC (1994), and Jennings (1993) into a national mapping scheme. At the time of this writing the FGDC system was still under review. It is a potential national standard for natural resource mapping.

Anderson et al. (1979) land use/land cover classification (hereinafter Anderson's) has been in use for several decades. This scheme covers all types of land cover, not just vegetation. The significance of this system, besides its applicability to all land cover and wide use, is that it was designed specifically for remote sensing. It is used as easily with 1 km resolution satellites as it is with fine-scale low altitude aerial photography. Since the work presented in this thesis is based on satellite remote sensing, this general land cover scheme will be referenced.

Satellite Remote Sensing and Landsat Thematic Mapper

Remote sensing is a proven means of mapping Earth's land cover (Roughgarden et al. 1991). Satellite-based land observation instruments offer consistent, synoptic, digital data with a range of spatial detail. Resolutions vary with each sensor from 1 km pixels in the Advanced Very High Resolution Radiometer (AVHRR) to the fine 10 m pixels in panchromatic SPOT (Le Systeme Pour l'Observation de la Terre) imagery (Campbell 1996). Currently under development are several systems capable of 1 m spatial resolution (Wynne and Carter 1997).

Spectral resolution or the area of the electro-magnetic spectrum being measured is another characteristic of each satellite instrument. Sections of the spectrum are recorded in bands. Each pixel has a digital number (DN) or a brightness value (BV) for each band. The DN is the relative reflectance of that area of land in that part of the electromagnetic spectrum (Jensen 1986). The greater the number of bands and the area of the spectrum covered, the greater is the likely discriminatory power. SPOT 10 m panchromatic data contain 1 broad band covering 0.51 to 0.73 *u*m. Landsat MSS (multispectral scanner) has 4 bands representing the green, red, and 2 near-infrared segments. One of the most spectrally robust instruments is the Landsat TM. This system records 7 bands (Table 1). TM's broad spectral range and 30 m resolution make it an ideal instrument for statewide land cover mapping (Scott et al. 1993). Specifics about Landsat TM images used in this project will be discussed in the later chapters.

Image Processing

Image processing (in the context of this thesis) means taking un-processed, remotely sensed digital data, manipulating it on a computer, and creating a product, typically a map. Specific to this thesis, the operation consists of taking Landsat TM imagery and producing a statewide land cover map. There have been many texts written about this topic, including those by Campbell (1996), Lillesand and Kiefer (1994), Jensen (1996), and Cracknell and Hayes (1991). While all of the above works differ slightly, there is basic agreement on the major steps of image processing. The following 4 major components, preprocessing, classifying, post processing, and assessing accuracy will be introduced.

Preprocessing is preparing raw digital data for the main analysis, usually classification. According to Campbell (1996), preprocessing can be separated into 3 functions: feature extraction, radiometric corrections, and geometric corrections.

Feature extraction is the process of determining which part of the raw imagery to utilize. Landsat TM and other reflectance-based Earth observation data contain several bands. Each band takes up computer space and increases processing time. Feature extraction determines which combinations of bands are to be processed. Feature extraction may be determined by the objectives of the classification. Table 1 lists the intended uses of each Landsat TM band. If the objectives are fairly broad in scope, all of

Table 1. Spectral ranges and representative uses of the 7 Landsat Thematic Mapper bands (adapted from Campbell, 1996).

Band	Spectral Range		Intended Purpose
1	0.45 - 0.52 *u*m	blue-green	soil and vegetation separation
2	0.52 - 0.60	green	reflection from vegetation
3	0.63 - 0.69	red	chlorophyll absorption
4	0.76 - 0.90	near IR	water body delineation
5	1.55 - 1.75	mid IR	vegetative moisture
6	10.40 -12.50	far IR	hydrothermal mapping
7	2.08 - 2.35	mid IR	plant heat stress

the bands may be kept or all may be assimilated. A common strategy in feature extraction is principal component analysis (PCA). PCA creates a few new assimilated bands with most of the variation of the original data, but without the redundancy and noise. PCA achieves this data reduction by looking for correlations between bands and uses these correlations to reduce any redundancy. Campbell (1996) recommended Davis (1986) and Gould (1967) for a detailed discussion of principle component analysis. Another strategy in feature extraction is arithmetic operations. These operations can take the form of simple band ratios, vegetation indices, the tasseled cap transformation (Crist 1983), and even multitemporal combinations. For a detailed review of arithmetic operations see Mather (1987). The objective of feature extraction is to reduce data storage space and computation time required to process remotely sensed images. If concerns about space and time are not significant limiting factors and/or the full radiometric pixel range is beneficial, then this data reduction may be eliminated.

Once the bands to be interpreted are identified, they must be radiometrically corrected. Radiometric correction means altering the brightness values (or digital numbers) to fix sensor specific-noise or to compensate for atmospheric effects such as haze and topographic shadows. Sensor-specific noise can show up as striping, which can be attributed to a number of possible errors in data transmission or collection. It can also be attributed to sensor malfunction (Lillesand and Kiefer 1994). This form of radiometric correction can be performed relatively easily with filters or arithmetic operations. A much more difficult problem is overcoming atmospheric interference.

Radiation passes from the sun through the atmosphere to the Earth’s surface and then back through the atmosphere to be recorded on the Landsat instrument. Particulates in the atmosphere can cause major problems due to total concealment of ground features (e.g., clouds or smoke), or by altering, to varying degrees, the spectral value (Lavreau 1991). There has been much research in this area (Chavez 1996, de Hann et al. 1991, Hill and Sturm 1991, Lavreau 1991, Chavez 1989). Cracknell and Hayes (1991) provided an excellent review of the principles behind atmospheric correction.

Other than atmospheric interactions, the factors that influence the recorded BV for a land cell include reflectance of the target, angle of the sensor, solar elevation angle, and slope and aspect of the target in relation to the solar direction (Mather 1987). Reflectance of the target is dependent on land cover and is therefore the sought-after variable. The angle of the sensor and solar elevation angle (within the 24-hour cycle) are essentially constant and are not factors in interpreting images due to the sun-synchronous orbits of Earth observation satellites (Campbell 1996). The influence of the slope and aspect of the land on reflectance values is called the topographic effect. This effect is an important factor, especially in mountainous environments. Slopes with South-facing aspects reflect more energy than north-facing slopes (in the Northern Hemisphere) due to the angle of the Earth surface relative to the sun. The magnitude of this problem changes with the yearly tilt cycle of the Earth (i.e., the Northern Hemisphere is farther from the sun in winter and closer in summer). The effect of the tilt of the Earth is greatest during the winter solstice. The steepness of the slope determines the angle of reflection of radiation from the sun to the sensor, and therefore the steeper the slope; the more the effects of aspect are compounded. Steep south-facing slopes will reflect the most radiation while steep north-facing slopes will reflect the least. There have been several studies seeking solutions to this radiometric problem (Civco 1989, Hall-Konyves 1987, Jones et al. 1988, Leprieur et al. 1988, and Teillet et al. 1982).

Several of the research papers listed above reporting methods to correct both atmospheric and topographic radiometric problems require the use of additional data sets. These might include digital elevation models (DEMs) to calculate slope and aspect or actual ground reflectance data to compare with the atmospherically altered satellite imagery. Using these ancillary data in performing radiometric correction and in classifying sites requires that the satellite data be registered with a consistent georeferencing system. Such referencing allows all geospatial data to be used readily in a GIS. Therefore, the geometric correction is very important. The objective of this process is to give the data the geometric integrity of a map (Lillesand and Kiefer 1994). First, a standard correction is made to adjust the effects of the Earth's rotation on the satellite image. Figure 1 shows the parallelogram pattern that results from this rectification. The next step is to georeference the image to a map projection and to a coordinate system. Lastly, the image must be georectifed to the terrain, i.e., adjustments must be made to account for land with steeper slopes having more area per 2-dimensional cell than flat areas. These 3 steps can be accomplished through resampling. Geometric corrections can be done at any stage in image processing. It is necessary to have it complete, however, before ancillary data are brought in for radiometric correction, classification, or accuracy assessment.

Geometric corrections as well as the other steps in preprocessing must be done with caution. Most of these processes alter the pixel reflectance values in some way. Alteration of these values may affect classification accuracy. It is usually best to alter these BV as little as possible (Campbell 1996).

The act of labeling each pixel or cell into a category, within a GIS theme (or layer), is image classification. There are 2 primary types of classification (using single date imagery): spectral pattern and spatial pattern recognition (Lillesand and Kiefer 1994). Spectral pattern classification relies on statistics of the BVs of each individual cell for each band. Spatial pattern classification is based on statistics about each cell in relation to its neighboring cells. Spectral pattern classification is the most traditional approach. Spectral methods are available on most software and are easily understood. These advantages make spectral pattern classification adequate for simple land cover mapping.

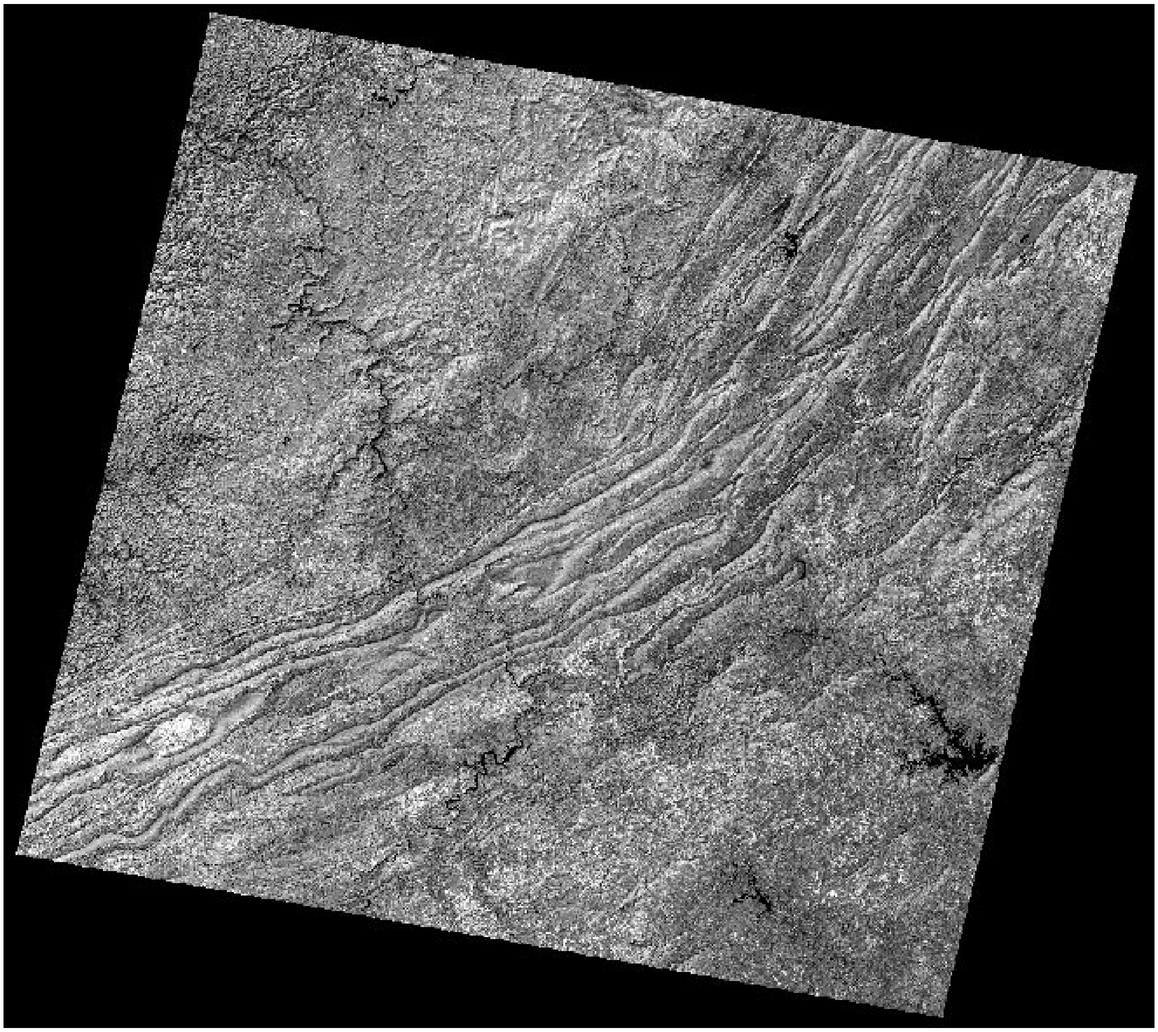

Figure 1. Landsat Thematic Mapper scene of Southwest Virginia (Band 4 of scene 1734).

There are 3 types of spectral pattern classification: supervised, unsupervised, and hybrid classification. Under each of these categories there are many specific classifiers, or computerized algorithms for performing image classification (Campbell 1996).

Supervised classifiers require the input of known ground locations with known cover types. These are referred to as “training areas.” The computer then analyzes these sites and computes statistics about the reflectance values of each category, according to a specific algorithm. Next all of the pixels are examined and, based on their BVs, placed in the category with the closest (stated statistical bounds) training values. This method requires a large amount of ancillary data (ground data) for training sites, prior to classification.

For example, cover types such as forest, water, agriculture, and development can be identified on an aerial photograph. The same large, homogeneous, and contiguous blocks of the above cover type can then be located on the corresponding satellite imagery. On-screen digitizing (or other methods) can then be used to designate these areas as training sites. Statistics are calculated for the pixels within the collection of training sites of each land cover type. The classifier then looks at each pixel in the image and places it in one of the above categories based on statistics of the training areas.

Problems that could reduce thematic accuracy in supervised classification include: lack of training data for every land cover type, poor quality of ground data, low quantity of training sites, poor distribution of training sites. In the first case, a gap in training data could cause large areas not to be classified or to be labeled as an incorrect class, dependent on the algorithm. Poor (i.e., inaccurate or imprecise) training data will adversely affect the training statistics, thus reduce classification accuracy. If there is a poor distribution of training data, i.e., the sites for each category are not fairly evenly spread across the area to be classified and training sites are clustered in a small area of the image, then spatial autocorrelation problems may occur (Campbell 1981). If the number of training areas for each category is low, classification statistics may be skewed, depending on the classifier. If there is a chance that any of the above problems may be encountered, then unsupervised classification should be considered (Mather 1987).

Unsupervised classification, sometimes referred to as clustering, does not require large amounts of initial input (Jensen 1996). In this method, the computer looks for "natural" groupings of reflectance values. The analyst may enter a few initial starting values such as number of desired thematic or information classes, threshold standard deviations for separating clusters, and number of passes the algorithm informs the computer to make across the image (PCI 1994). The classifier then groups the pixels into clusters based on similarity of BVs. It is then up to the analyst to translate these clusters into informational classes.

According to Campbell (1996) there are several advantages and disadvantages to this type of classification. The advantages are: low amount of *a priori* knowledge of the area is required, potential for human error is reduced, additional relevant classes may be identified which may have otherwise been overlooked, and the output classes are spectrally consistent. Disadvantages include: lack of analyst control and the possibility of the spectrally-clustered output being inconsistent with the desired informational classes. There are other disadvantages. Spectral relationships change over seasons and years. These changes limit applications to a single time period. There is also a need for detailed ground data when translating the clustered output from spectral classes to information classes (*a posteriori*). The timing of required ground data is just the opposite of supervised classification (Lillesand and Kiefer 1994).

To minimize the drawbacks and maximize the advantages of supervised and unsupervised operations, a hybrid method can be performed. A hybrid is any combination of supervised and unsupervised classification. One example of a using a hybrid approach to obtain homogeneous training areas is illustrated in Lillesand and Kiefer (1994). By clustering a small area of an image, checking the clusters, and then using the clusters as training sites for a supervised classification, better classification accuracy of the desired land cover categories will result. Bauer et al. (1994) showed that a similar hybrid approach, called "guided clustering", was very successful in separating forest types in Minnesota. The disadvantages to using this category of classifiers are the long processing time required and the lack of algorithms available on software packages.

Once the imagery is classified, the next step is post-processing. During this phase the thematic layer is prepared for accuracy assessment, map output, and/or further analysis. One common procedure in this step is post-classification smoothing. Smoothing removes the "salt and pepper" appearance on the map caused by spectral variability in pixel-based classification algorithms (Lillesand and Kiefer 1994). Smoothing can be accomplished through various filters. Other items that may be completed in this step include: conversion from raster to vector data types, changing the pixel size, creating a color palette, and putting together mosaics of separate images to cover an entire study area.

The last step in image processing is to access the accuracy of the output. Accuracy in this phase refers to how close the classified pixels are to the actual land cells. This is also referred to as thematic accuracy. Spatial accuracy, or how close a land feature is geographically to its actual place on the Earth's surface, is not directly considered. Spatial accuracy, however, must be high in order to compare the classification to reality (i.e., complete the thematic accuracy assessment).

Land cover mapping is not complete unless accuracy is assessed (Congalton 1996). Due to the arithmetic nature of GIS overlays, errors are propagated through each step of a geographic analysis. It is essential to keep track of error in individual map layers (Janssen and van der Wel 1994). There has been much attention given to the accuracy assessment of land cover maps in the last decade (Verbyla 1995).

Typical accuracy reports include overall accuracy, user's and producer's accuracy for each category, an error matrix, and the Kappa statistic. The *overall accuracy* (this and the subsequent 2 phrases, defined here, will be repeatedly used throughout the remainder of this thesis) is the number of total reference cells divided by the number of correctly classified cells. The *user's accuracy*, which reflects errors of commission, is the percentage of reference points within a class that were actually classified as that class (reference points within a class labeled another cover type would be viewed as commission error). *Producer's accuracy*, which reflects omission errors, is the percentage of pixels that were labeled a certain class that were verified by the reference data. Both of these statistics can be calculated using the error matrix. The error matrix is a table with the reference data on the horizontal axis and the classified data on vertical axis (Table 10). The Kappa statistic is a measure of how much better the results of classification are versus random pixel assignment (Congalton and Mead 1983). The Kappa unit expresses the percent improvement of the classification results over random classification. This statistic can also be calculated from the error matrix. Congalton (1991) and Congalton (1996) provided thorough reviews of this accuracy assessment research and of the current standards.