A Methodology for Developing Land Cover Data for the Houston-Galveston Area Council and Clean Rivers Program Region

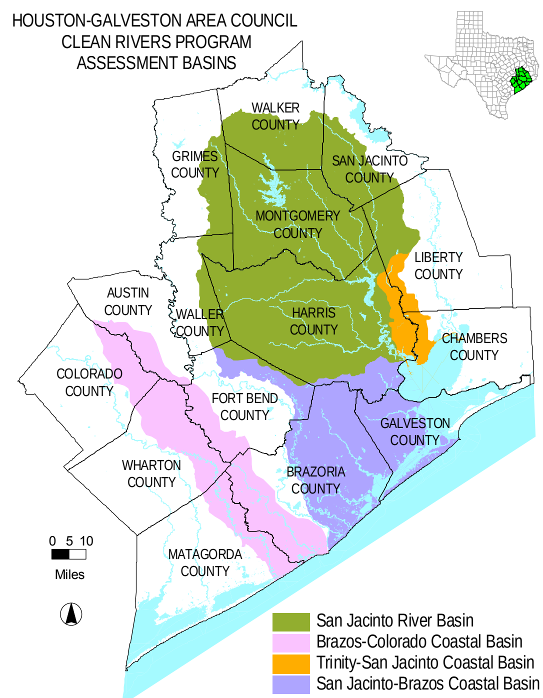
1. Purpose and Goal

Land cover data are an important element in understanding the dynamics of water quality throughout individual watersheds. Knowing the land cover within a watershed, especially that adjacent to the stream channel, provides for a better distinction between environmental and anthropogenic influences affecting waterways.

The purpose of this work is to assist regional and state programs in summarizing water quality for the region. An important component of the summary, is low cost, accurate land cover data that is current and easily integrated with resident data management and analytical tools. This effort will determine suitable land cover categories that most accurately represent the diverse nature of the H-GAC Clean Rivers Program (CRP) Region. The categories developed will be based on concerns, issues, and activities that face the integrity of the area waterbodies and the general environment. The end product will be a regional data set showing the general distribution of land cover throughout the region.

2. Geographic Area of Interest

The Area of Interest for this project includes the 13 county region of H-GAC and the additional areas (of San Jacinto and Grimes Counties) that comprise the four assessment basins for the CRP Region. The four basins are the Trinity-San Jacinto Coastal Basin, San Jacinto River Basin, San Jacinto-Brazos Coastal Basin and the San Bernard segments of the Brazos-Colorado Coastal Basin. The total area to be included in this project is roughly 12,500 square miles or 8,000,000 acres.



3. Expected Uses

The demand for current land cover data on a recurring basis is growing, especially in the areas of the region that are experiencing rapid change in land cover characteristics. Applications for the data include:

- Locating various land cover types within a watershed (*where?*)
- Determining acreage and/or percentage of each land cover type within a watershed *(how much?)*
- Associating water quality characteristics with land cover types and spatial patterns. *(Does water quality change with land cover?)*
- Assessing historical water quality data with changes in land cover temporally, identifying portions of the watershed that may be sensitive or vulnerable to water quality impairments due to surrounding land cover.

4. Definition of Land Cover types to be classified

The exact list of cover types to be defined will not be determined until an unsupervised classification and field training have been completed. Based on program goals and resources, data analysis needs, and size and diversity of the area to be classified, the intent is to define the classes listed below.

1.0 Upland
- 1.11 High-Intensity Developed
- 1.12 Low-Intensity Developed

1.2 Cultivated Land
1.3 Grassland
1.4 Woody Land
1.5 Bare Land

2.0 Wetland
3.0 Water and Submerged Land

This classification system is based on the National Oceanic and Atmospheric Administration (NOAA) Coastal Ocean Program's Coastal Change Analysis Program (C-CAP). The C-CAP classification system is hierarchical, reflects ecological relationships, and focuses on land cover classes that can be discriminated primarily from satellite remote sensor data. It was adapted and designed to be compatible with other nationally standardized classification systems, especially the US Geological Survey, Environmental Protection Agency and the Fish & Wildlife Service (CCAP Guidance 1995). Each of the classes and subclasses listed are described in greater detail below.

Developed Land (Derived from the Anderson et al. [1976] Urban or Built-up class) characterizes constructed surfaces comprised of concrete, asphalt, roofing, and other building materials with or without vegetation. This class has been divided into two subclasses based on the amount of constructed surface relative to the amount of vegetated surface present. High Intensity Developed Land contains little or no vegetation. This subclass includes heavily built-up urban centers as well as large constructed surfaces in suburban and rural areas. Large buildings (such as multiple family housing, hangars, and large barns), interstate highways, and runways typically fall into this subclass. Low Intensity Developed Land contains substantial amounts of constructed surface mixed with substantial amounts of vegetated surface. Small buildings (such as single family housing, farm outbuildings, and sheds), streets, roads, and cemeteries with associated grasses and trees typically fall into this subclass.

Cultivated Land (Agricultural Land in Anderson et al. 1976) includes herbaceous (cropland) and woody (orchards, nurseries, vineyards, etc.) cultivated lands. Seasonal spectral signatures, geometric field patterns and road network patterns may help identify this land cover type. Always associated with agricultural land use, cultivated land is used for the production of food and fiber.

Grassland differs from Rangeland in Anderson et al. (1976) by excluding shrub-brushlands. Unmanaged Grasslands are dominated by naturally occurring grasses and forbs which are not fertilized, cut, tilled or planted regularly. Managed Grasslands are maintained by human activity such as fertilization and irrigation, are distinguished by enhanced biomass productivity, and can be recognized through vegetative indices based on spectral characteristics. Examples of such areas include lawns, golf courses, forest or shrub areas converted to grassland, or areas of permanent grassland with altered species composition. This category includes managed pastures and pastures with vegetation that grows vigorously as fallow. Managed Grasslands are used for grazing or for growing and harvesting hay and straw for animal feed.

Woody Land includes non-agricultural trees and shrubs. The category alleviates the problem of separating various sizes of trees and shrubs using satellite remote sensor data but allows a height-based separation if high-resolution aerial photography are available. The class may be partitioned into three subclasses: Deciduous, Evergreen, and Mixed. These three subclasses generally can be discriminated with satellite remote sensing systems.

Bare Land (derived from Barren Land in Anderson et al. 1976) is composed of bare soil, rock, sand, silt, gravel, or other earthen material with little or no vegetation. Anderson's Barren Land was defined as having limited ability to support life; C-CAP's Bare Land is defined by the absence of vegetation without regard to inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the vegetated classes. Unusual conditions such as a heavy rainfall may occasionally result in growth of a short-lived, luxuriant plant cover. Wet, nonvegetated exposed lands are included in the Wetland categories. Bare Land may be bare temporarily because of human activities. The transition from Woody Land, Grassland, or Cultivated Land to Developed Land, for example, usually involves a Bare Land phase. Developed Land also may have temporary waste and tailing piles. Woody Land may be clearcut producing a temporary Bare Land phase. When it may be inferred from the data that the lack of vegetation is due to an annual cycle of cultivation (eg. plowing), the land is not included in the Bare Land class. Land temporarily without vegetative cover because of cropping or tillage, is classified as Cultivated Land, not Bare Land.

Wetlands are lands where saturation with water is the dominant factor determining soil development and the types of plant and animal communities living in the soil and on its surface (Cowardin et al. 1979). A characteristic feature shared by all wetlands is soil or substrate that is at least periodically saturated with or covered by water. The upland limit of wetlands is designated as (1) the boundary between land with predominantly hydrophytic cover and land with predominantly mesophytic or xerophytic cover; (2) the boundary between soil that is predominantly hydric and soil that is predominantly nonhydric; or (3) in the case of wetlands without vegetation or soil, the boundary between land that is flooded or saturated at some time during the growing season each year and land that is not (Cowardin et al. 1979). The majority of all wetlands are vegetated and are found on soil.

Water and Submerged Land: All areas of open water with < 30% cover of trees, shrubs, persistent emergent plants, emergent mosses, or lichens are assigned to Water and Submerged Land, regardless of whether the area is considered wetland or deepwater habitat under the Cowardin et al. (1979) classification. The Water class includes Cowardin et al.'s (1979) Rock Bottom and Unconsolidated Bottom, and Nonpersistent Emergent Wetlands, as well as Reefs and Aquatic Beds that are not identified as such. Most C-CAP products will display water as a single class.

The CCAP Classification System does provide for a more detailed list of subclasses from what is shown above, but it is not known if those land cover subclasses will be classified during this project.

5. Sensor and Imagery Requirements

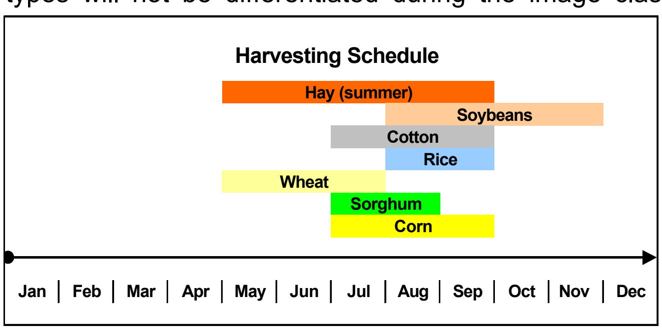
The preferred C-CAP satellite sensor system is Landsat Thematic Mapper (TM). Although the spatial resolution of TM is not as good as that of other satellite or aircraft systems, it is generally less expensive to acquire and process for large-area coverage (CCAP Guidance 1995). Spectrally, the TM sensor captures seven bands of the electromagnetic spectrum. Only six will be used (Two Middle Infra-red, Near Infrared, Red, Green and Blue Visible bands) during the classification process. TM imagery have a spatial resolution or minimum measurement unit of 30 square meters. The acquired imagery will possess little or no cloud cover, with a maximum percent of cloud cover of 20%. A TM image scene covers 185 x 172 km on the ground, or 115 x 106 miles. Our geographic area of interest will require four Landsat scenes for complete spatial coverage.

To enhance the classification of an impervious class such as Developed, the temporal period of acquisition should be in the Winter or “Leaf-off” months. February 2002 is considered a suitable timeframe for meeting our broad classification needs. However, to thoroughly delineate each of the listed land cover type, specifically Cultivated Land and Wetland, a second set of imagery will be acquired.

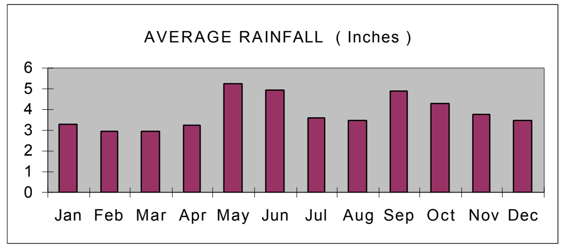
The Texas Agricultural Census (TAC, 1997) along with other meteorological information for the region indicate early July 2001 as a suitable second acquisition date for the classification of Cultivated Land and Water classes. According to the TAC, the three dominant crop types for the counties in this region, are Hay, Rice (irrigated), and Sorghum (refer to table on the right).

MAJOR CROP	TOTAL ACREAGE
HAY	252304
RICE	212123
SORGHUM	190793
COTTON	133156
SOYBEANS	105028
CORN	58853
WHEAT	2544

The goal of acquiring a second temporal set of imagery is to increase the confidence in our classification methodology (and subsequently a higher overall accuracy assessment). The second set of imagery will capture data prior to dates of peak crop harvesting. Individual crop types will not be differentiated during the image classification. The table to the left lists the dominant crop types for the region and their period of harvesting. Although Hay is harvested all summer, other dominant crop types are harvested during a three-month period starting in July.



Available crop information for our region indicates that all of the dominant crops shown use some type of chemical application (commercial fertilizers and pesticides). Commercial fertilizers include nitrogen, phosphorous, and potassium. Rice is fertilized with nitrogen and phosphorous while all crops utilize some form of disease or insect control using the insecticides and fungicides. Although fertilizers and pesticides may be designed to “breakdown” rapidly, and their impact on the Region’s water quality is not known, determining the distribution of cultivated land is an integral part to assessing a watershed’s water quality.



The National Weather Service Rainfall measurements (taken from both Hobby and Bush Airports from 1927 to 2001) indicate the wettest monthly averages are May, followed by June then September. Rainfall is consistent throughout the year with less than 2.25 inches separating the highest and lowest monthly averages. The month of July is below the annual average in rainfall. In addition, it is a summer month with a greater potential for possible sunshine making it a suitable timeframe for the second image acquisition date.

6. Image Processing Techniques and Accuracy Assessment

The most significant effort of this project will be devoted to image classification and the accuracy assessment of the classification. Prior to the classification process, several preprocessing steps will be completed. Four Landsat 7 scenes will be required to cover the study area. Two dates for each Landsat scene, representing both the summer and winter seasons, will be purchased. An attempt will be made to obtain cloud free imagery, however, if clouds are present they will be masked to prevent confusion during the image classification. Each of the scenes will be registered to a USGS 7.5 minute quadrangle base map with an allowable error of 0.5 pixels. Multi-date imagery from both the winter season and summer season will be incorporated into the image classification to provide additional spectral reflectance data on phenological changes in vegetation growth and moisture variations that occur from season to season. Multi-temporal information on the spectral reflectance of our land cover classes should allow us to separate land cover such as coniferous and deciduous forest cover and capture spectral reflectance of residential areas that are normally hidden under deciduous canopy cover during the summer.

Reflectance values of dark and bright areas in the images, such as water bodies and high intensity developed areas will be used in a regression to calibrate the overall brightness of the image scenes. This method will preserve variation resulting from seasonal changes in phenological growth of vegetation while eliminating differences in spectral reflectance that may result from sensor offset and gain and variation in scene illumination. Each scene will then be input into a tasseled cap transformation algorithm to produce an image of greenness, brightness, and moisture. The tasseled cap transformation is a data reduction technique that isolates variation in spectral reflectance that is the result of physical structures in the image scene. The resulting greenness, brightness, and moisture bands for the two corresponding scenes will be combined to form a single multi-temporal dataset for input into the unsupervised and supervised image classification.

The Image Processing Techniques (see flowchart on p. 7) will involve an iterative process, using a hybrid, unsupervised and supervised classification methodology. Initiated with an Unsupervised Classification (UC), a common image processing routine (ISODATA) will serve primarily as an exploratory procedure to determine homogenous areas for the location of training sites for the supervised classification and to uncover dominant spectral signatures in the imagery. A maximum likelihood classification algorithm incorporating prior probability of evidence images for the land cover classes will be used for the supervised classification. Examples of the prior probability images include National Wetland Inventory (NWI) maps, a thresholded and reclassed NDVI image to isolate areas of vegetated and non-vegetated surfaces, and buffered TIGER road layers to isolate high and low intensity developed areas.

Following the initial supervised classification the result will be evaluated visually and through an in process classification assessment. The in process assessment will utilize the soft classification tools in Idrisi to identify the presence of any non-singleton classes, areas of uncertainty in the classification and determine which of our training sites need to be adjusted to eliminate confusion in the classification. The soft classifier is based on Dempster-Schafer theory and produces a set of images that provide statements of the degree of membership of each pixel to the designated land cover classes, and also generates an image that quantifies the uncertainty in the classification. A "cluster busting" stratification methodology as utilized in the NOAA C-CAP program may also be explored to stratify and isolate those land cover classes that are providing the most confusion in the image classification.

After a satisfactory result is achieved, an accuracy assessment will be performed using a combination of field data and high-resolution aerial orthophotography. A stratified random sampling methodology consisting of 30–100 sample points per land cover class will be used to determine the accuracy of the classification. Sample points within a reasonable distance of roads will be verified. The other inaccessible sample points will be evaluated using high-resolution aerial orthophotography. The classification methodology will be repeated until a Kappa Index of Agreement (KIA) of 70% or greater per land cover class and an overall KIA of 70% or greater has been achieved. Each classified scene will then be mosaiced and edge matched to form the final classified image of the region.

Accuracy Assessment
Data of unknown accuracy are of little value to the end user. The purpose of the Accuracy Assessment is to compile the data necessary to support credible statements regarding data accuracy. The recommended accuracy assessment for this classification is a test based on comparison with both independent field samples and corresponding sets of high-resolution aerial imagery.

The source of primary assessment data will be collected through field sampling. Field sampling data will be used to verify the classification of stratified, random, geographic locations generated by the image processing software. Industry literature recommends a minimum of 30 random locations to verify each land cover type (van Genderen and Lock 1977). If any of the field verification sites are inaccessible or located within a training site, the software will generate additional sites to verify. Inaccessible sites may be verified "virtually" using high-resolution digital orthophotography or other ancillary data sets. Field verification work will be conducted by full and part-time staff, with most of the field work completed by those involved directly with the image classification.

In addition, any available habitat data collected through the CRP Regional Monitoring effort will be included. For example, we would compare a significant number of monitoring locations with the same locations on the image to determine how many are alike. Other data sets (such as appraisal parcel data, orthophotography) will be used to strengthen field verification and the accuracy assessment. To improve the final accuracy of the image classification, we will submit watershed maps for the entire study to appropriate local area experts who are knowledgeable in the setting of a particular watershed.

Based on the general use of the data, availability of project resources, our goal is to develop data that meets or exceeds an overall accuracy of 70%. That is, 70% or more of the random locations from the classified image match what is found through the direct field verification and ancillary data sets.

Software and Hardware Used
The software and hardware used to support the classification and assessment tasks include the Idrisi Digital Image Processing & Geographic Information System (GIS) Software (Clark Labs, Worcester, MA.), on a Dell Pentium 4 Personal Computer with Windows 2000, and Trimble Pathfinder Pro XRS Global Positioning System (GPS) Unit. When possible, project staff will utilize ERDAS' Imagine Image Processing Software to enhance and compare image classification efforts completed with Idrisi.

7. Results

All procedures (including successes and failures), accuracy assessment, applicability of data, metadata, and instructions on how to use the final data will be documented and made available.

8. Project Timeline

January	Receive, install Hardware and Software Work with historical imagery, other raster data sets for research and preparation
February	Receive Final Approval for Project Methodology
March	Acquire and preprocess Imagery (July 2001 and February 2002) Conduct Unsupervised Classification of Imagery
April	Conduct Pre-Supervised Classification Field work (determine training sites) Conduct Initial Supervised Classification Revisit those training sites with uncertain classifications Repeat Supervised Classification Conduct Field Verification (visit 30-100 per class) Conduct Preliminary Accuracy Assessment
May	Repeat Supervised Classification Repeat Accuracy Assessment
June	Compare Supervised Classification with ancillary data
July	Complete Final Accuracy Assessment Submit Draft Project Results.
August	Incorporate any comments, Submit Final data, metadata, and results. Make all data and information available to the Internet.

9. Reference Information

Accuracy Assessment: To ensure an adequate number of accuracy assessment samples per land cover type from the stratified random sampling of the final classified image, a minimum of 30 sample points per type will be required; which allows for a 15% error in accuracy (van Genderen and Lock 1977).

Classification: The process of assigning individual pixels of a digital image to categories, generally on the basis of spectral reflectance or radiometric characteristics (REMOTE SENSING Handbook for Tropical Coastal Management).

Clean Rivers Program: The Texas Clean Rivers Program (CRP) was implemented to maintain and improve the quality of surface water resources within each river basin in Texas. The CRP is a partnership involving the Texas Natural Resource Conservation Commission (TNRCC), other state agencies, river authorities, local governments, industry, and citizens. Using a watershed management approach, CRP partner agencies work with the TNRCC to identify and evaluate surface water quality issues and to establish priorities for corrective action. For more information visit http://www.hgac.cog.tx.us/intro/introwater.html

Coastal Change Analysis Program: or C-CAP, is a cooperative interagency and state/federal effort to detect coastal upland and wetland land cover and submersed vegetation and to monitor change in the coastal region of the United States (Cross and Thomas 1992; Haddad 1992). The project utilizes digital remote sensor data, in situ measurement in conjunction with global positioning system systems, and geographic information system (GIS) technology to monitor changes in coastal wetland habitats and adjacent uplands. For more information visit http://www.csc.noaa.gov/products/sccoasts/html/ccapguid.htm

Electromagnetic Spectrum: Electromagnetic radiation is energy propagated through space between electric and magnetic fields. The electromagnetic spectrum is the extent of that energy ranging from cosmic rays, gamma rays, X-rays to ultraviolet, visible and infrared radiation including microwave energy.

Kappa Statistic or Coefficient: proposed by (Cohen, 1960) expresses the agreement between the reference points collected in the field and the categories on the map if as the condition of random agreement is removed. The Kappa coefficient is highly recommended for accuracy assessment of remotely sensed data (Congalton, 1991). The Kappa range is 0.00 to 1.00, where 0.00 is a classification no better than randomly assigning pixels and 1.00 is a perfect classification.

Land Cover/Use: A term that includes categories of land cover and categories of land use. Land cover is the vegetation or other kind of material that covers the land surface. Land use is the purpose of human activity on the land; it is usually but not always related to the land cover.

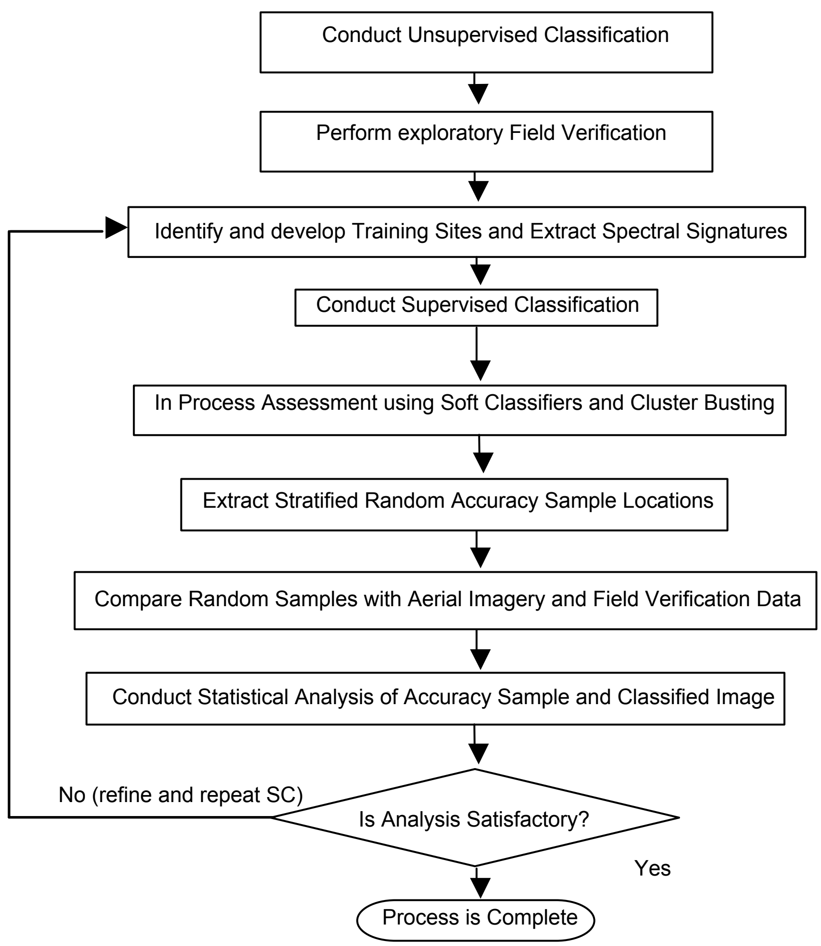
Pixel: A contraction of the words picture element, it is the smallest unit of information in an image or raster map. Referred to as a cell in an image or grid.

Polygon: A multi-sided geographic feature that represents an area on a map.

Real-time Differential Global Positioning System (DGPS): Real-time DGPS is the process of correcting GPS positions at an unknown location with a nearby radio beacon broadcasting a known location simultaneously. This process enables accurate (sub-meter) geographic data collection to occur.

Resolution: A measure of the amount of detail that can be seen in an image; the size of the smallest object recognizable using the detector.

Stratified random sampling: A sampling method in which the elements of the population are allocated into sub-populations (e.g. strata) before the sample is taken, and then each stratum is randomly sampled (Lund eReports in Physical Geography, No. 1, Oct. 1997).



Flowchart for the Image Processing and Accuracy Assessment Task