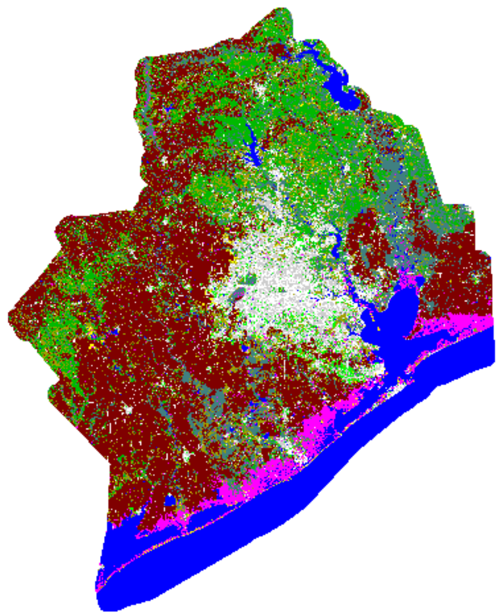
Houston-Galveston Area Council
2008 Land Cover Image Processing Protocol



Prepared by:
Roger Meyer
811 Nth 9th Ave East
Duluth, MN 55805

Prepared for:
Texas Commission on Environmental Quality
Texas Clean Rivers Program
12100 Park 35 Circle
Austin, Texas 78753
December 2008

Table of Contents

Tables

Figures

1. Introduction

The objective for this project was to develop an updated, 2008 land cover data set that could be utilized as a component in watershed and water quality analysis within H-GAC's Clean Rivers Program (CRP) assessment basins. The targeted area for land cover mapping included the 15 counties encompassing the four CRP assessment basins, approximately 17,814 square miles (Figure 1). A low cost processing protocol that could

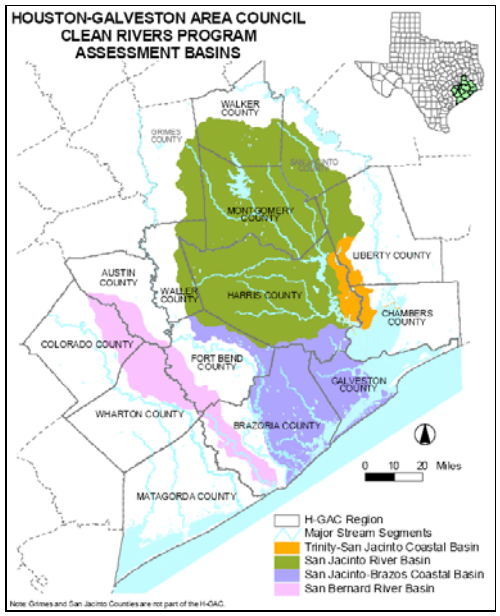


Figure 1. Map showing the 15 counties encompassing the study area and H-GAC's assessment basins

produce a land cover product with an acceptable level of accuracy was required. An acceptable level of accuracy was determined to be per category and overall Kappa values greater than 70%. The Multi-Resolution Land Characteristics Consortium's (MRLC) image processing procedures used for creation of the 2001 National Land Cover Database (NLCD) were leveraged to improve classification accuracy and produce land cover data comparable to NLCD and NOAA Coastal Change Analysis Program (C-CAP) products at the Anderson I level (Anderson et al., 1976). Specific technical procedures used for image preprocessing, training site development, classification and post processing are reviewed in the following sections of this document.

2. Classification Scheme

The land cover classification scheme for this mapping effort was designed to be comparable to the standardized and hierarchical classification schemes used for the NLCD and NOAA C-CAP land cover products. NOAA is a participating partner in the MRLC and is responsible for providing land cover for the coastal regions of the NLCD. The following is a list of the targeted land cover class descriptions adapted from the NOAA Coastal NLCD classification schemes (Dobson et. al., 1995):

Developed, Higher Intensity - Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial.

Developed, Lower Intensity - Includes areas with a mixture of constructed materials and vegetation. These areas most commonly include single-family housing units.

Developed, Open Space - Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes.

Cultivated - Areas characterized by herbaceous vegetation that has been planted or is intensively managed for the production of food, feed, or fiber; or is maintained in developed settings for specific purposes. Herbaceous vegetation accounts for 75-100 percent of the cover. Also includes pasture/hay areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation.

Grassland/Shrub -Areas characterized by non-cultivated grassland, shrublands and transitional woody land. Shrubland includes natural or semi-natural woody vegetation with aerial stems, generally less than 6 meters tall, with individuals or clumps not touching to interlocking. Both evergreen and deciduous species of true shrubs, young trees, and trees or shrubs that are small or stunted because of environmental conditions are included. Grasslands include mixture of areas characterized by natural or semi-natural herbaceous vegetation; herbaceous vegetation accounts for 75-100 percent of the cover. Also includes areas dominated by upland grasses and forbs. These areas are not subject to intensive management, but they are often utilized for grazing.

Forest - Areas characterized by tree cover (natural or semi-natural woody vegetation, generally greater than 6 meters tall); tree canopy accounts for 25-100 percent of the cover.

Woody Wetland - Areas where forest or shrubland vegetation accounts for greater than 20 percent of vegetative cover and the soil or substrate is periodically saturated with or covered with water.

Herbaceous Wetland - Areas where perennial herbaceous vegetation accounts for 75-100 percent of the cover and the soil or substrate is periodically saturated with or covered with water.

Barren - Areas characterized by bare rock, gravel, sand, silt, clay, or other earthen material, with little or no "green" vegetation present regardless of its inherent ability to support life. Vegetation, if present, is more widely spaced and scrubby than that in the "green" vegetated categories; lichen may be extensive.

Water - All areas of open water, generally with less than 25% cover of vegetation or soil.

3. Satellite Imagery

Landsat 5 Thematic Mapper (TM) scenes for 2008 were purchased from the USGS EROS Data Center and preprocessed using MRLC procedures. Standardization and registration with MRLC imagery provided access to a pool of mult-temporal Landsat imagery from the MRLC scene library, and allowed NOAA C-CAP land cover data to be utilized as the framework for establishing spatial prior probabilities for target land cover classes.

3.1. Image Acquisition

Portions of four Landsat TM scenes were required to cover the study area. The original intention was to purchase multi-temporal imagery to facilitate the capture of land cover signatures resulting from seasonal changes in vegetation phenologies. However, only imagery from the March/April time period was obtained due to the limited availability of cloud free imagery in 2008 (table 1). All 4 scenes were purchased from the USGS EROS Data Center in NLAPS and FAST-L7A format and imported into IDRISI software for preprocessing. The scenes were radiometrically and geometrically corrected by USGS (Level 1G systematically corrected) and rotated and aligned to the UTM coordinate system using the WGS84 datum and a 30 meter resolution. Purchased imagery was cloud free with the exception of scene 2639 which required some cloud and cloud shadow masking prior to classification.

Table 1. 2008 satellite imagery purchased from the USGS EROS Data Center

Path	Row	Satellite	Acquisition Date	Scene ID
26	39	Landsat 5	2008/03/19	LT50260392008079EDC00
26	40	Landsat 5	2008/03/19	LT50260402008079EDC00
25	39	Landsat 5	2008/04/13	LT50250392008104EDC00
25	40	Landsat 5	2008/04/13	LT50250402008104EDC00

3.2. Preprocessing

The four 2008 Landsat TM scenes purchased from USGS were resampled to an Albers equal area projection using image to image registration with terrain corrected satellite imagery from the MRLC scene library. The 2008 imagery was resampled to within 6 meters using a quadratic mapping function and nearest neighbor resampling with a minimum of 20 control points per scene. Resampled scenes were then clipped to a 6 km buffer of the 15 counties comprising the study area.

The georeferenced imagery was standardized using MRLC processing procedures which convert DN to at satellite reflectance and transform the image bands using a regionalized tasseled cap transformation (Huang et.al, 2002). Custom Python scripting was used to automate standardization of the imagery including conversion to radiance, at satellite reflectance, and tasseled cap transformation (Figure 2).

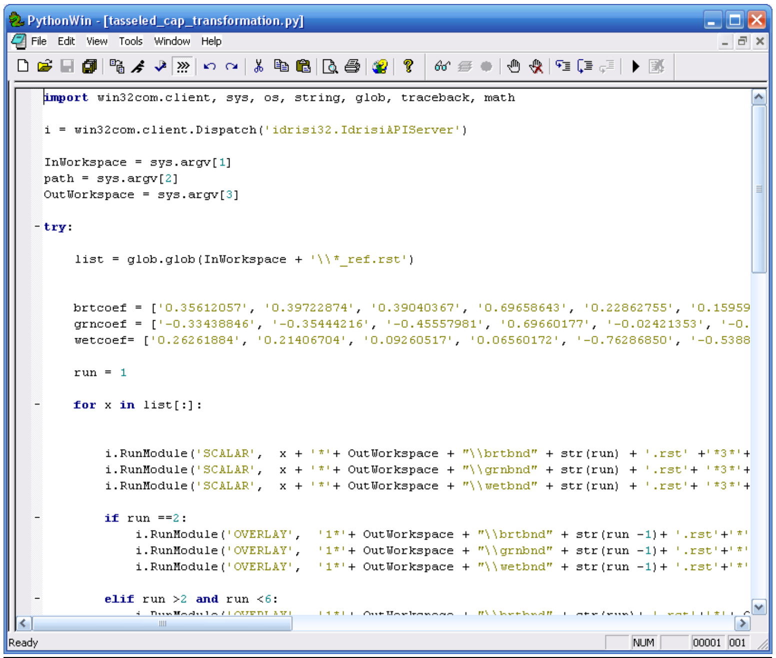


Figure 2. Custom python script for automating Tasseled Cap transformations

4. Training Site Development

As seen in figure 3, a geographically stratified random sample of USGS quarter quads was extracted for collection of training data for the target land cover classes (Lillisand et al, 1998). Color infrared 2006 and 2008 NAIP imagery corresponding to the randomly selected quarter quads were acquired and used to digitize training sites for the target land cover classes. The sampling scheme provided an even distribution of training sites across the region. Concurrent examination of aerial photography with the satellite imagery ensured representation of the spectral variation of target land cover classes within each scene, as well as determination of land cover change when digitizing training sites.

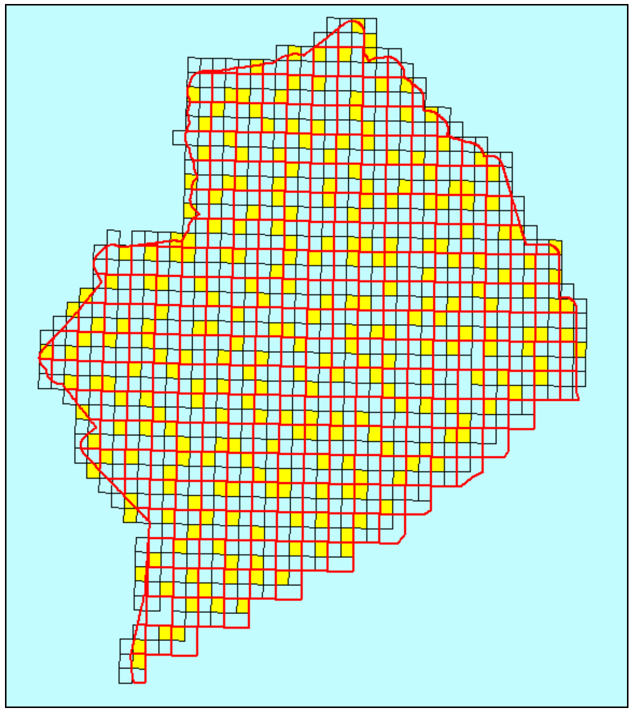


Figure 3. Stratified random sample of targeted USGS DOQQs (yellow quads) for training site development

5. Classification Methods

Landsat scenes were classified individually and merged after classification. A hybrid supervised/ unsupervised method that leveraged image segmentation and spatial prior probabilities for target land cover classes was used for classification of the Tasseled Cap bands for each Landsat scene.   Figure 3 provides a graphical representation of the work flow for classifying each Landsat scene.

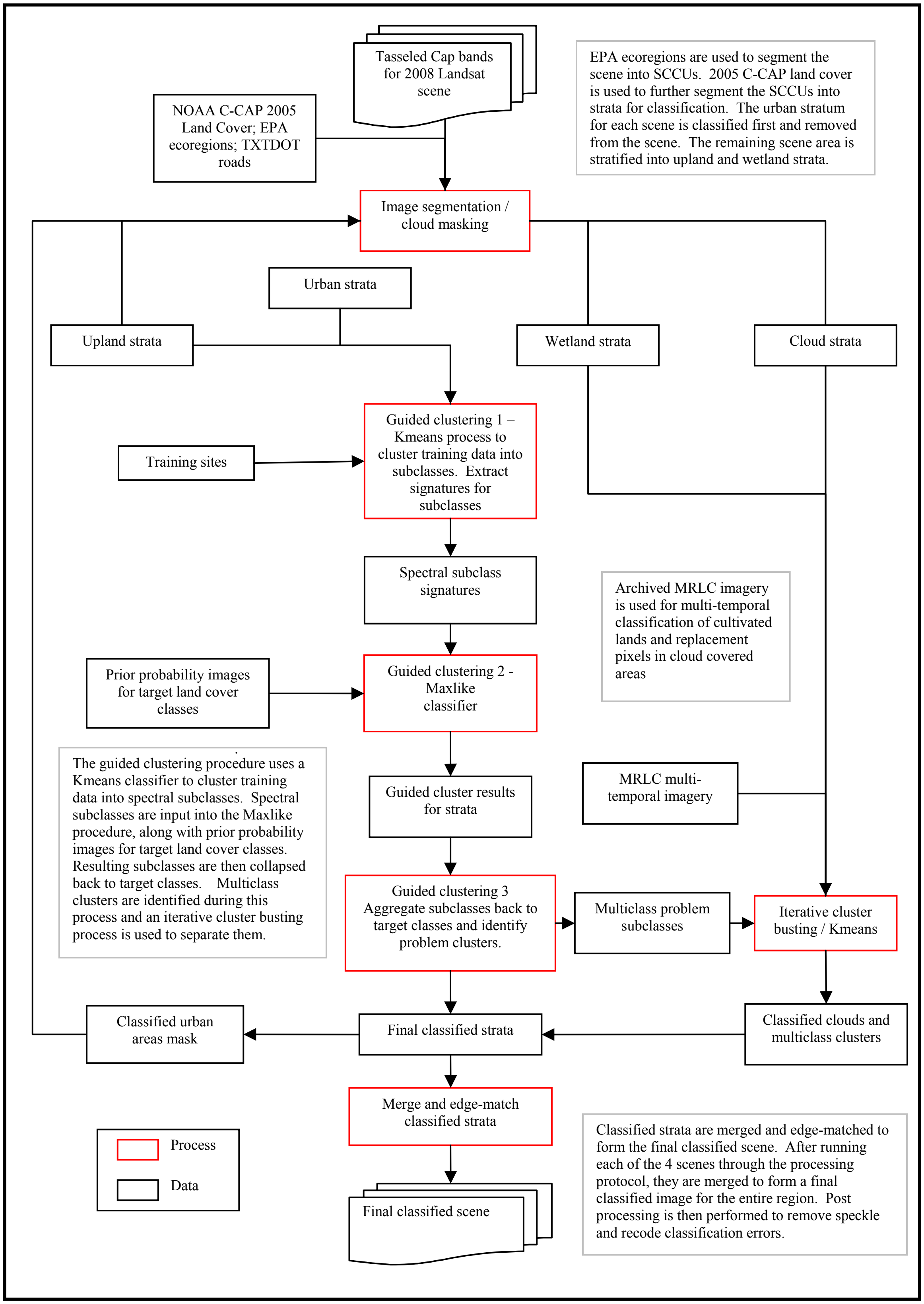


Figure 4. Classification workflow for processing Landsat TM scenes

5.1. Image Segmentation

The 2008 Landsat scenes were segmented into spectrally consistent classification units (SCCU) prior to classification to reduce spectral mixing of target land cover classes (Lillisand et al, 1998). First, a mask of potential urban areas is created by merging urban classes extracted from the 2005 C-CAP land cover data with a buffered (30m) TXTDOT roads layer (Texas Dept. of Transportation, 2000). Some manual modifications of the urban mask were required to capture new urban and residential developments. The urban stratum is classified and then removed from the Landsat scene. The remaining scene area was stratified into SCCUs using EPA level 3 ecoregions. Only those ecoregions occupying a large area of the Landsat scene were used for stratification. C-CAP 2005 land cover data served as the spatial framework for further stratifying each ecoregion SCCU into wetland and upland strata. Each stratum was classified separately and merged after classification.

Some cloud and cloud shadow masking was required for scene 2639. An automated process which differences the first two components of PCA457 and PCA123 band combinations was used to generate the initial cloud and cloud shadow masks (Varlyguin, et. al. 2001). The cloud and cloud shadow masks were then manually edited, combined and buffered by 30 meters (figure 4). An iterative cluster busting process using cloud free imagery from the MRLC image archive was used for classification of cloud masked areas.

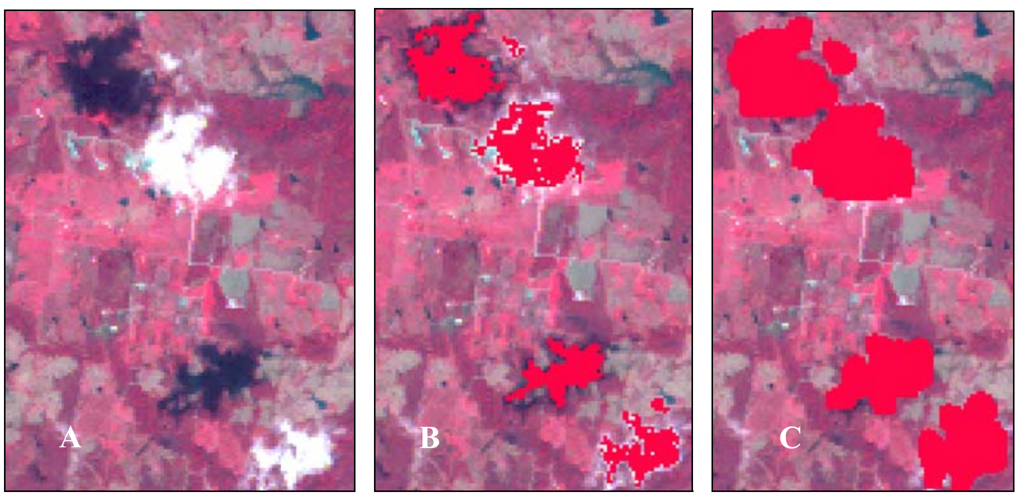


Figure 4. Cloud / shadow masking process: (A) cloud and shadows, (B) PCA mask result, C) and final mask with 30m buffer

5.2. Image Classifiers

Guided clustering was the primary method used for classification of the upland and urban strata. The method is ideal for large area classification, can be automated, and training sites are not required to be homogenous (Bauer et. al., 1994). Guided clustering involves the following processing steps:

1. Isolate image training pixels for target class.
2. Cluster resulting pixels into sub-classes.
3. Repeat 1 and 2 for all target classes.
4. Perform maximum likelihood classification using all sub-classes.
5. Collapse subclasses back to original target classes.
6. Perform any post-processing procedures.

The Kmeans classifier in IDRISI was used to cluster the training site data into subclasses. Subclass signatures are then generated and input into the Maxlike classifier along with spatial prior probability images for the target classes. Transitional probability images for land cover classes generated by the MARKOV change analysis module in IDRISI were used to establish spatial prior probabilities for target land cover classes. The Markov change analysis module uses two land cover datasets from different time periods, in this case 2001 and 2005 C-CAP land cover datasets, to produce a probability image for each land cover class that reports the probability that each land cover type would be found at each pixel after the specified number of time units have elapsed (Eastman, 2006). After running the Maxlike classifier, the resulting subclasses are collapsed back to the original target classes.

An iterative cluster busting process is used to resolve any problem multi-class subclasses that are identified during the guided clustering process. The cluster busting process is an unsupervised classification method that requires the interpreter to assign clusters to the target land cover classes. Clusters that can be confidently assigned to target classes are omitted from the next iteration of the clustering algorithm. Another iteration of the clustering algorithm is then performed on the remaining clusters that cannot be identified. This iterative process is continued until the interpreter can no longer confidently assign clusters to the target classes (Jensen, 1987).

5.4. Cultivated Classification

In some cases, guided clustering did not produce adequate results for classification of cultivated areas. Cultivated areas were often underestimated or mixed with other target land cover classes. Multi-temporal Tasseled Cap imagery from the MRLC scene library was incorporated into the iterative cluster busting process and used to assist in extracting spectral signatures for cultivated areas. An effort was made to obtain temporal imagery representative of crop phenologies during the early, peak, and late growing seasons (table 3). Selection of imagery was limited, however, by percent cloud cover for individual scenes. The main objective was to capture multi-temporal spectral signatures indicative of tillage, planting, growth and harvest.

Table 2. Ancillary MRLC archive imagery used to assist in agriculture classification

Path	Row	Satellite	Acquisition Date	Scene ID
26	39	Landsat 7	2002/02/23	LE7026039000205450
26	39	Landsat 5	2004/07/14	LT5026039000419650
25	39	Landsat 5	2006/12/04	LT5025039000633850
25	39	Landsat 5	2000/09/30	LT5025039000027410
25	40	Landsat 5	2004/12/14	LT5025040000434910
25	40	Landsat 5	2000/09/30	LT5025040000027450

26	40	Landsat 7	2002/02/23	LE7026040000205450
26	40	Landsat 7	2000/09/29	LE7026040000027351

5.5. Wetland Classification

Accurate classification of wetland habitats is difficult due to spectral mixing, especially when classifying a large region. Also, it can be difficult to identify wetland training sites using aerial photography. Most land cover mapping efforts utilize extensive field work for collection of training data and verification. Due to the absence of field verification data for this mapping effort, USFWS National Wetland Inventory (NWI) and C-CAP 2005 land cover data were used for determining the potential spatial location of wetlands during classification. Unsupervised classification and iterative cluster busting was used for classification of wetland strata. The primary rules for interpreting clusters and assigning them to target land cover classes were as follows:

- The combined NWI and C-CAP wetland data was considered to be accurate unless examination of the aerial photography and satellite imagery indicated the change had occurred (e.g., wetland has been converted to developed land or cultivated land, or woody wetland has been clear cut and is now in a transitional stage). If change was verified then the cluster was assigned to the suggested land cover class.
- If no change was indicated and the cluster consisted primarily of persistent woody vegetation then the cluster was assigned to the woody wetland category.
- If no change was indicated and the wetland cluster did not indicate the presence of woody vegetation than the cluster was assigned to the wetland category (non-woody).
- If no change was indicated and the cluster consisted primarily of water reflectance then the cluster was assigned to the open water class.

This method limited spectral mixing associated with classifying wetlands over a large region and spectrally updated the NWI and C-CAP wetland data if major land change had occurred.

6. Post Classification Processing

Post processing of classified imagery consisted of recoding classification errors and filtering to remove speckle.

6.1. Recoding

The Idrisi recode model was developed to semi-automate the recoding process (figure 6). The model requires the input of vector masks delineating the area to be recoded and an .avl file designating the “from” and “to” classes. The model then extracts the area to be recoded, recodes the pixels, and burns them back into the classification. To date, two passes through the 2008 land cover dataset have been made to recode classification errors.

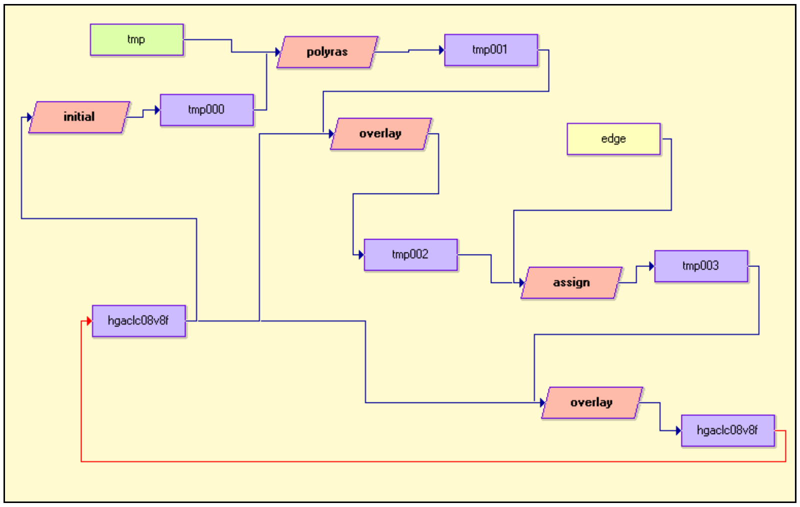


Figure 5. Custom Idrisi recode model

6.2 Filtering

The classification process can produce areas of isolated pixels that differ from the majority class. These isolated pixels are a result of the complexity of separating land cover signatures in a satellite image, or can reflect the actual heterogeneity of land cover. For mapping purposes it is a common procedure to generalize an image to reduce the occurrence of these areas of isolated pixels. The filter model uses the group and area modules in IDRISI to extract single pixels (figure 7). Single pixels are recoded to the mode pixel in a 3x3 neighborhood. These recoded pixels are then burned back into the classification.

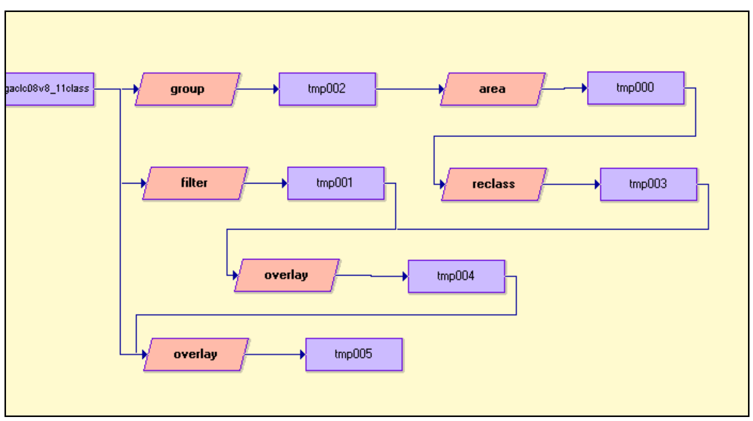


Figure 6. Custom Idrisi filter model for removing speckle

7. Results

Accuracy assessment is a critical component of land cover mapping and allows users to determine appropriate uses for the data (Congalton, 1996). A detailed accuracy assessment for the H-GAC 2008 land cover dataset should be produced at different levels of post-processing, sampling restrictions, and class aggregation to better enable users to determine the efficacy of the dataset.

A pixel-to-pixel accuracy assessment conducted on the unprocessed, raw classified image incorporates conservative bias because of the difficulty in separating true classification error from error that may result due to misregistration, temporal differences between the collection of the ground truth data and satellite imagery, or the inability of the examiner to confidently interpret individual pixels on aerial photography (Verbyla and Hammond, 1995). Thus, a conservative assessment may actually underestimate the true accuracy of the classification.

An optimistic assessment of the unprocessed classification restricts sampling to areas in the image, which contain only one land cover class in a 3x3 pixel neighborhood. This type of assessment incorporates optimistic bias because sampling is limited to homogenous areas where land cover is more easily identified by the examiner (Hammond and Verbyla, 1996). Examples of the final map products produced for accuracy assessment of the H-GAC 2008 land cover dataset can be found in Appendix A and include:

Table 3. Final map products produced for accuracy assessment by H-GAC

Product	Description
hgaclc08v2_filter_10class.rst	10 class land cover map, filtered to remove speckle, recoding of classification errors
hgaclc08v2_filter_3x3hom_10class.rst	For optimistic accuracy assessment, 10 class post processed land cover map restricted to 3x3 homogenous pixel blocks.

**v2 represents the version number for the classification*

8. Conclusion

It is important for individual users of the data to determine appropriate use of the H-GAC land cover data set. Users of the data are encouraged to review and comprehend the image processing and accuracy assessments protocols used in the creation of the data set in order to determine if the data would be suitable for their needs. The data set is not intended to serve as the primary tool for regulatory or jurisdictional decision-making. Regulatory applications using the land cover data should involve rigorous field verification before any decisions or conclusions are made. Specifically, the data set was created for broad-scale planning and research applications at the county and regional level. Some general examples of appropriate and inappropriate uses would include:

Appropriate Uses

- Regional and county planning.
- Large area resource management planning.
- Educational purposes for students and citizens.
- Regional or county level water quality and watershed analysis.
- Basic research on county or regional distribution of land cover to determine specific areas for monitoring or management focus.

- Broad-scale evaluation of the environmental impact or benefits of a major project.
- Change detection and time series analysis.

Inappropriate Uses

- Determining the accuracy of other data using the H-GAC data set.
- Determining the location of jurisdictional wetlands.
- Determining exact area coverage of land cover without consideration of the overall accuracy and per category accuracy of the data.
- Establishing exact boundaries for regulatory enforcement.
- Mapping areas finer than the original resolution of the data.
- Combining or altering the data set and redistributing them.
- Establishing definite occurrence or nonoccurrence of a feature without consideration of probabilities determined by the accuracy assessments.

References

Anderson J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geol. Surv. Prof. Pap. 964.

Bauer, M.E., T.E. Burk, A.R. Ek, P.R. Coppin, S.D. Lime, T.A. Walsh, D.K. Walters, W. Befort, and D.F. Heinzen. 1994. Satellite inventory of Minnesota forest resources. *Photogrammetric Engineering andRemote Sensing* 60(3):287-298.

Congalton, R.G. 1996. Accuracy Assessment: A Critical Component of Land Cover Mapping. Gap Analysis. ISBN-1-57083-03603 *American Society for Photogrammetry and Remote Sensing.* 1996. p. 119 – 131.

Dobson, J.E., Bright, E.A., Ferguson, R.L., Field, D.W., Wood, L.L., Haddad, K.D., Ireland III, H., Jensen, J.R., Klemas, V.V., Orth, R.J. and Thomas, J.P. (1995) *NOAA* Coastal change analysis program (C-CAP): guidance for regional implementation, National Marine Fisheries Service, Seattle, Washington, NOAA Technical Report NMFS 123

Hammond, T.O. and D.L. Verbyla. 1996. Optimistic bias in classification accuracy assessment. International Journal of Remote Sensing, 17: 1261-1266.

Huang, C., L. Yang, C. Homer, B. Wylie, J. Vogelman and T. DeFelice, *"At-Satellite Reflectance: A First Order Normalization of Landsat & ETM+ Images"*.

Huang, C., Wylie, B., Homer, C., Yang, L., and Zylstra, G., 2002. *Derivation of a Tasseled cap transformation based on Landsat 7 at-satellite reflectance: International Journal of Remote Sensing*, v. 23, no. 8, p. 1741-1748.

Eastman JR, 2006, 'IDRISI Andes: Guide to GIS and Image Processing' Clark Laboratories, Clark University, Worcester, USA.

Jensen, J.R., Ramsey, E.W., Mackey, H.E., Christensen, E.J., and Sharitz, R.R. 1987. Inland Wetland Change Detection Using Aircraft MSS Data. *Photogrammetric Engineering and Remote Sensing* 53(5): 521-529.

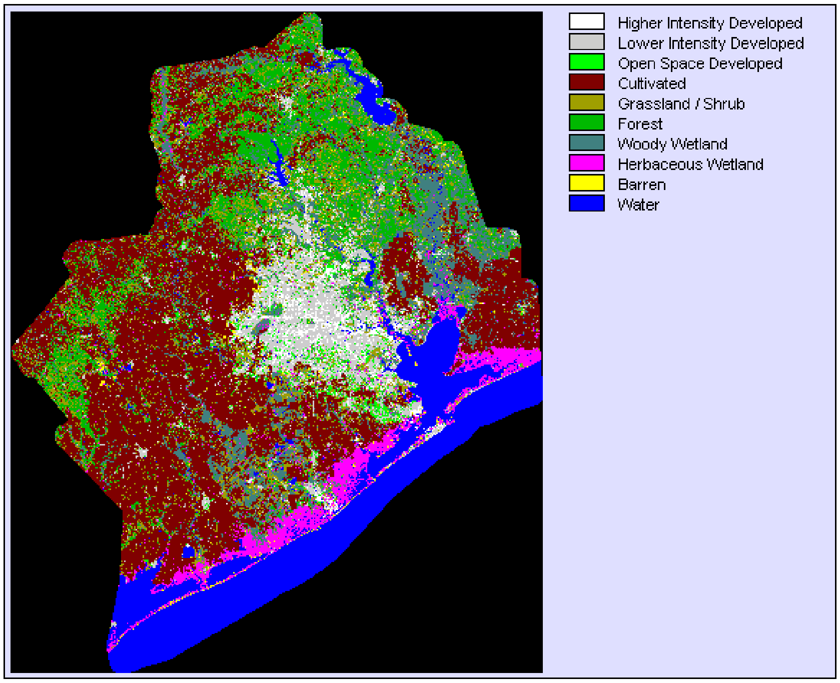
Lillesand, T., J. Chipman, D. Nagel, H. Reese, M. Bobo, and R. Goldmann. 1998. Upper Midwest Gap analysis program image processing protocol. Report prepared for the U.S. Geological Survey, Environmental Management Technical Center, Onalaska, Wisconsin, June 1998. EMTC 98-G001. 25 pp. + Appendixes A-C. (NTIS PB98-154560)

Varlyguin, D., R K Wright, S J Goetz, S D Prince (2001). Advances in land cover classification from applications research: a case study from the mid-Atlantic RESAC. Available at www.geog.umd.edu/resac and on ASPRS CD-ROM *in* American Society for Photogrammetry and Remote Sensing (ASPRS) Conference Proceedings, Washington DC.

Verbyla, D.L. and T.O. Hammond. 1995. Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. International Journal of Remote Sensing, 16: 581-587.

APPENDIX A: Final map products for accuracy assessment

10 class land cover map:



10 class land cover map restricted to 3x3 homogenous pixel blocks: