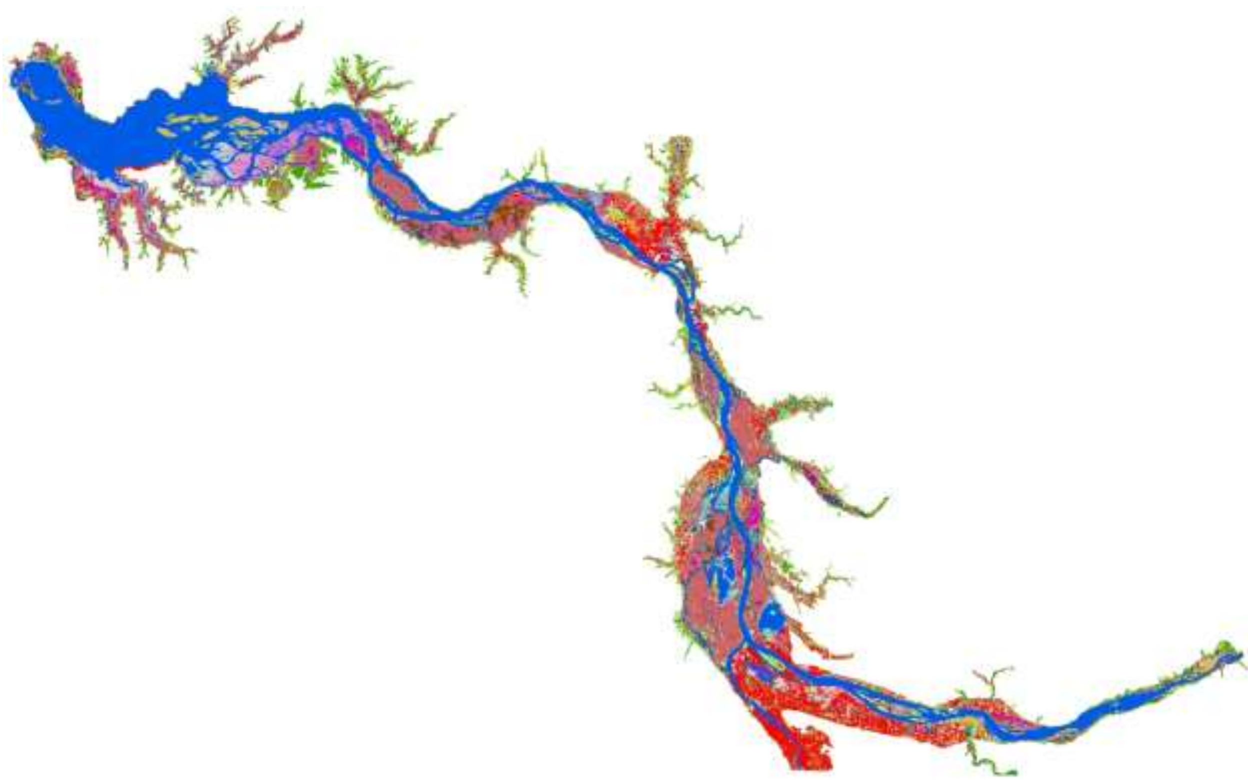
High Resolution Land Cover Mapping in the Lower Columbia River Estuary



Prepared by:

Sanborn Map Company

Lower Columbia River Estuary Partnership

May, 2011

Table of Contents

Abstract

The Lower Columbia River Estuary Partnership recently completed a high resolution land cover classification map for the Lower Columbia River and Estuary. The classification was done by the Sanborn Map Company, and incorporated recent (2009) aerial imagery, LiDAR, LandSAT and other supporting data sets. Additional support for this project was provided by the NOAA Coastal Services Center Coastal Change Analysis Program (C-CAP). Image classification utilized an innovative high resolution, image segmentation and object based approach that has been refined by Sanborn. The chosen classification scheme was based on existing, standard classifications, with additional focus placed on identifying estuarine and tidal freshwater habitats. Training data used for the classification consisted of existing data collected for previous mapping efforts, as well as an extensive set of new field data which was collected specifically for this effort. A separate set of field data was also collected to serve as accuracy assessment data. Mapping results were within expected accuracy ranges for all nearly all land cover classes. With the exception of 3 classes, all user's and producer's accuracies were calculated at greater than 75%. The NOAA C-CAP program incorporated this data set into their coastal land cover mapping inventory. A version of this, with the land cover classes rolled into the standard C-CAP classification scheme, is available from the NOAA Digital Coast geospatial data portal.

Introduction

The Lower Columbia River Estuary Partnership (EP) works to protect and restore the lower Columbia River and Estuary, one of 28 estuaries in the nation designated as an 'Estuary of National Significance'. The EP works in three areas to achieve this goal: habitat restoration, ecosystem monitoring, and educational programs. Up to date, detailed, land cover information that accurately identifies wetlands vegetation and other estuarine features, is vital to support many of these activities. This data is intended to serve many purposes including the following: identification and quantification of types and extent of habitat change relative to historic conditions; assistance with habitat restoration and habitat monitoring site locations; as a stand-alone Level 6 in the Columbia River Estuarine Ecosystem Classification. This is a new hierarchical mapping framework that is currently being developed at the University of Washington and USGS, with EP funding. It describes the landscape at six different spatial scales, and incorporates biological, geological, and hydrologic processes which shape the landscape.

In January 2010, recognizing a need for current and accurate data, the EP contracted the Sanborn Map Company to generate a land cover map of the Lower Columbia River historical floodplain. In developing its mapping strategy, the EP searched for solutions that would overcome limitations of previous LandSAT based classifications of the study area (i.e. limited selection of images, pixilation, poor separation of certain classes), within its budget. During the contractor selection period, the EP began discussions with NOAA's Coastal Change Analysis Program (C-CAP) and, by partnering with NOAA, were able to capitalize on the mapping framework they developed with private industry partner The Sanborn Map Company. This framework uses a high resolution image segmentation approach to image processing, rather than the traditional pixel based approach typically used for land cover classification. The end result of this process is a map which conforms more closely to natural breaks in land cover classes, compared to a traditional pixel based classification. In addition, the data can easily be made available in both vector and raster digital formats. By partnering with NOAA and Sanborn, the EP was able to take advantage of this innovative approach to land cover mapping, within a limited budget.

The map data was derived primarily from recent high resolution airborne imagery (2009 NAIP), LandSAT imagery (2007 and 2008), and LiDAR (2010) data. All data layers were carefully selected to coincide with the most opportune on-the-ground conditions. For example, NAIP images taken at the lowest possible tides were selected, in order to maximize the extent of exposed ground in the map. This was of particular importance in the lower estuary, where large tidal ranges and extensive tidal flats result in significant areas of wetting and drying throughout a tidal cycle. National Wetlands Inventory (NWI) data was incorporated as an additional source of wetlands information. However, use of this data was limited, as much of it throughout the region is out-of-date. The EP was also interested in distinguishing tidally influenced areas from areas where tidal influence is prevented due to the existence of anthropogenic barriers (levees, tidegates, culverts, etc). A data layer describing these conditions was developed by the EP, and provided to Sanborn for incorporation in the final map product. Development of this data set is described in a separate document.

The general approach for developing the high resolution, segment based map produced here is similar to more traditional methods. The general approach consisted of training data collection during two separate field campaigns undertaken at different points in the vegetation growth cycle, as well as an image processing phase wherein the imagery was classified into the selected land cover classes, using spectral and other modeling techniques. The classification process was refined several times throughout the project, until an acceptable map was produced which accurately identified the desired classes. Finally, the map was assessed for accuracy, using a separate set of on the ground data collected during the field campaigns.

This paper discusses the specifics related to development of the land cover map data. Discussion includes the use of image segmentation and object based classification techniques to improve high resolution classification, challenges associated with tide and river conditions, and the selection of multiple resolution imagery. In addition, the use and impact of LiDAR data, the classification scheme used and associated challenges, as well as the classification and accuracy assessment are also discussed.

Classification Scheme

The backbone of any remotely sensed land cover product is the development of a totally exhausted, mutually exclusive, and hierarchical classification scheme. Before any work could be conducted for field data collection or mapping, LCREP and stakeholders conducted a kickoff meeting identifying critical land cover types and developed a plan to complete the classification scheme. Table 1 lists the classes used for this project. Appendix A includes the dichotomous land cover key, further defining the classes.

Table 1: LCREP Classification Scheme

Class #	Class
10	Coniferous Upland Forest
11	Deciduous Upland Forest
12	Mixed Upland Forest
21	Coniferous Wetland Forest
22	Deciduous Wetland Forest
23	Mixed Wetland Forest
40	Upland Shrub/Scrub
41	Wetland Shrub/Scrub
50	Upland Herbaceous
51	Wetland Herbaceous
60	Aquatic Beds
70	Agriculture
71	Tree Farm

80	Barren
81	Mud
82	Sand
83	Bare
84	Rock
85	Cobble
90	Urban/Developed/Impervious
91	Water
93	Open Space Developed

Field Data Collection

The field data collection consists of adherence to safety protocols and reporting, equipment inventorying and testing before and after field use, workflows, quality control, and access protocols. These topics are briefly covered below and are subject to change based on field conditions, weather, and access.

Field Equipment

Field equipment will be inventoried for each day of use in the field. All equipment will be tested prior to use and issued with contingency plans for quick replacement in the event of failure or damage. Field equipment consisted of the following:

1) Field Laptop
    a. Imagery
    b. Field segments
    c. Ancillary data layers

2) USB GPS for laptop

3) USB jump drives for backup

4) Camera GPS
    a. GPS should be time-synced to camera every day
    b. GPS should be turned on at the beginning of each field day

   c. GPS should be carried with camera

5) Paper maps

6) Contact information list

7) Job hazard analysis (JHA) forms (Appendix C)

Daily Work Flow

The data collection work flow accommodated deviations in the schedule but is generally followed the following format:

1) Daily routes were predetermined to maximize the numbers and variety of sites visited daily.

2) Job Hazard Analysis forms signed daily prior to starting fieldwork.

3) Camera GPS synced and turned on to provide accurate geo-location.

4) Current location in relation to segments known at all times using issued field maps and real-time GPS location

5) In ArcGIS ArcMap, segments being attributed selected using the field tool (Figure 1) based on current location and aerial imagery on laptop.

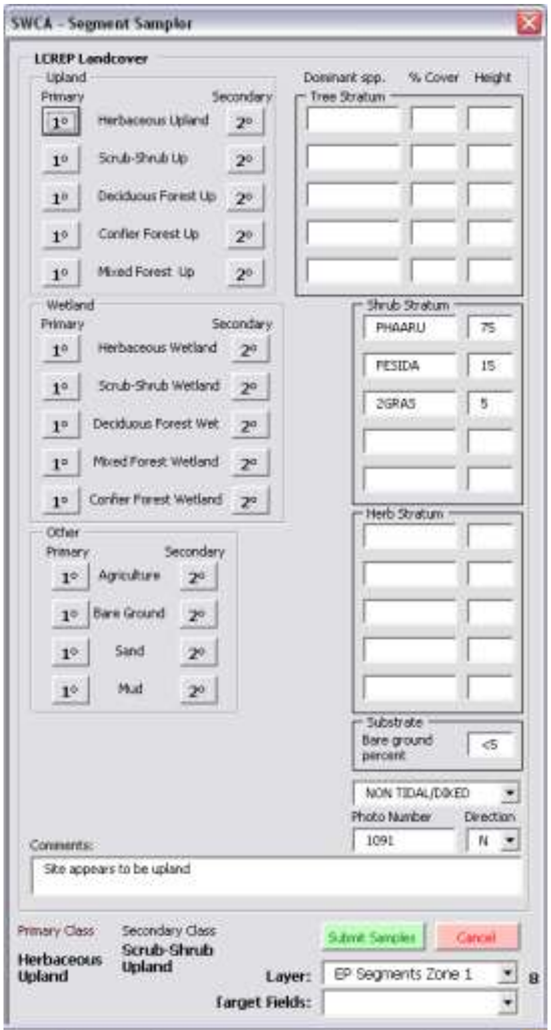


Figure 1. ArcGIS field form for recording segment data

6) Primary land cover elements identified based on elements identified by the classification dichotomous key (Appendix A) and percent cover visually estimated based on field crew perspective (oblique view) and aerial imagery.

7) Types and percentages of vegetated cover types (>20%) filled out and heights noted for tree classes. Shrub and herb stratum percent cover noted.

8) Percentage of bare ground noted in the field form.

9) Modifiers such as tidal, non-tidal, and diked noted out for each layer where evidence of modifiers exist.

10) Photographs of the area taken and later synced to camera GPS and used for further field verification.

11) Any comments relevant to defining the segment commented with the segment tool and hard copy field notes.

12) Samples committed to the database (Figure 2).

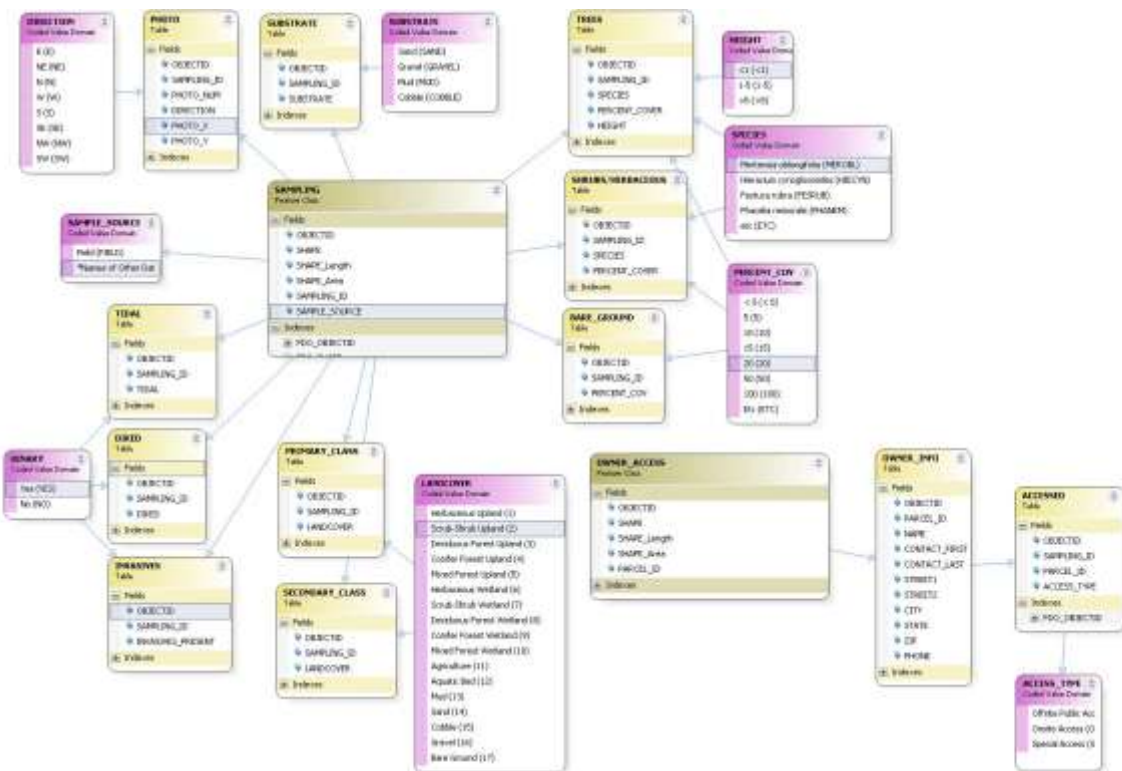


Figure 2. Data schema detailing the information collected during field work and regarding ownership and access

13) Map refreshed and the extent of the classified polygons QC for logical class call and consistency.

    a. Data layers such as SSURGO soils, NWI polygons, and existing land cover referenced to check the relative accuracy of the land cover primary class.

14) The next polygons selected for classification selected and attributed for the area following Step 1.

Quality Control

Quality control is the critical step for evaluating the accuracy and integrity of field collected data. The following protocols were used to maintain quality control throughout the field data collection phase:

1) Field collected data evaluated daily while collecting and nightly (when possible) for consistency.
2) Entire sample database backed up daily.
3) Field data checked visually for logical consistency.
4) Field data checked visually for logical consistency based on NWI, SSURGO, and existing landcover datasets.
5) Data found to be erroneous based on visual or ancillary data were flagged for review by field teams and analysts.

Access Issues

Access can be a limiting factor if not planned appropriately. Described below is the approach to maintaining the field session moving at a reasonable pace and maximizing class captures.

1) Primarily, field crews used public right-of-way, county roads, and public access sites for sampling
2) Where applicable, field crews secured access for sampling on specific areas.
3) If approached by concerned members of the public, field crew personnel will explain that they are conducting a survey of potential areas to sample from and inquire if the persons would like to participate in the project.
4) If not, field teams continued without sampling in the area to avoid any issues.

Field Sampling Results

In an effort to collect the maximum numbers of sample points and achieve the goal of 100 sample points per class, the following sampling scheme was proposed. The area was divided into nine sampling zones (Figure 3). Each sampling zone was assessed in approximately one day. Primary access to sampling was by car surveys conducted from public roads coupled with public boating access. Each zone had around 100,000 image segment polygons to be sampled from.

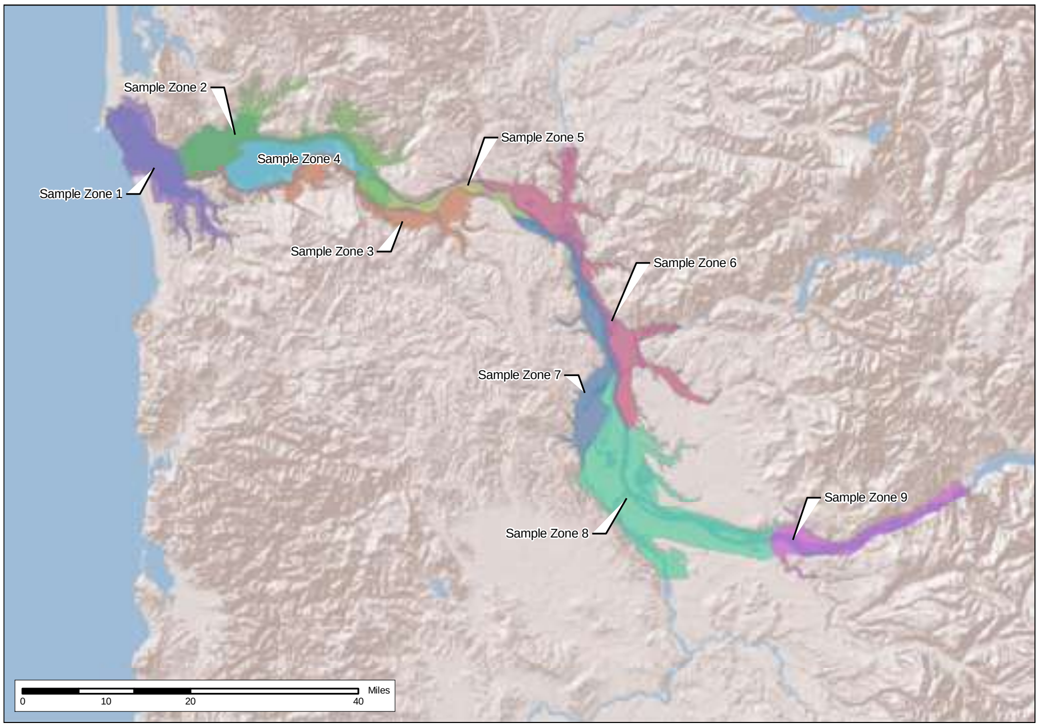


Figure 3. EP sampling zones.

Each of the sampling zones had existing data that were used as samples. The existing data were evaluated for currentness and accuracy. The usable data was matched to applicable land cover categories and the categorical information was applied to polygons and reviewed. Next, these data was augmented with newly collected field samples and photo interpreted segments. The most notable existing data was provided by the Port of Portland's Natural Resource Inventory (NRI) data for the Port properties and CASI datasets provided by the Estuary Partnership. The numbers of samples by data source are detailed in Appendix B. The numbers of samples per zone are detailed in Appendix C.

Sampling Results

Field samples totaled 7154 with 6130 recorded during the first field session and 1024 collected during the second field season. NRI data accounted for 10639 samples. These samples were primarily located in the Portland metropolitan area although Port properties extend beyond Portland city limits. 6404 segment samples were collected from the original four CASI classifications across six different sample zones.

350 photo-interpreted (PI) points were collected. PI points were chosen in areas where calls were less confusing or potentially erroneous calls less likely. Google Earth Street View, field photographs, and aerial imagery contributed to the PI calls. Additionally, NWI, soils data, and past land cover datasets such as NOAA CCAP and USGS GAP were used to verify and inform decisions.

726 individual photos were taken and geolocated for the collection effort. The photos are a record of numerous field sampled segments as well as the segments surrounding landscape features. Photos were used to double check sample calls and to photo interpret additional sample points.

The sampling distribution (Figure 4) was dependant upon access constraints and public roads. The overwhelming majority (>95%) of field sample data were collected from public roadways and State of Oregon, State of Washington and Federal lands.

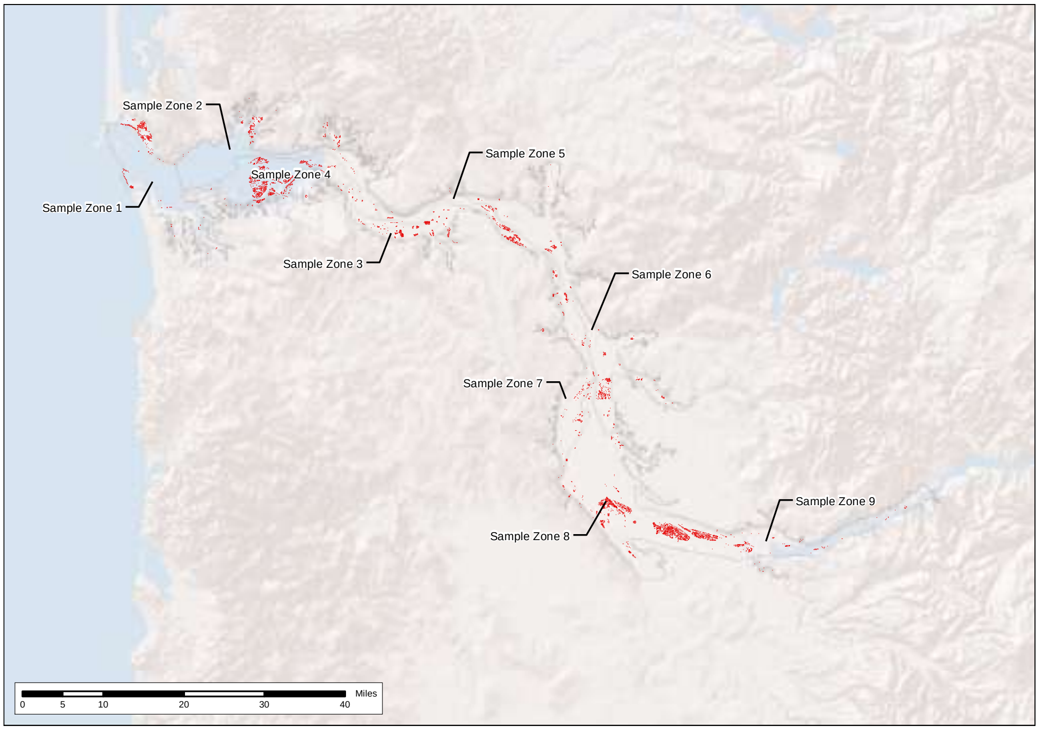


Figure 4. Distribution of final samples (red).

In an effort to obtain field data in sites accessible only by water (Figure 5), three days of boat sampling were conducted. Much of these boat-accessible areas are within US Fish & Wildlife properties although samples along the Columbia River shoreline were also collected.

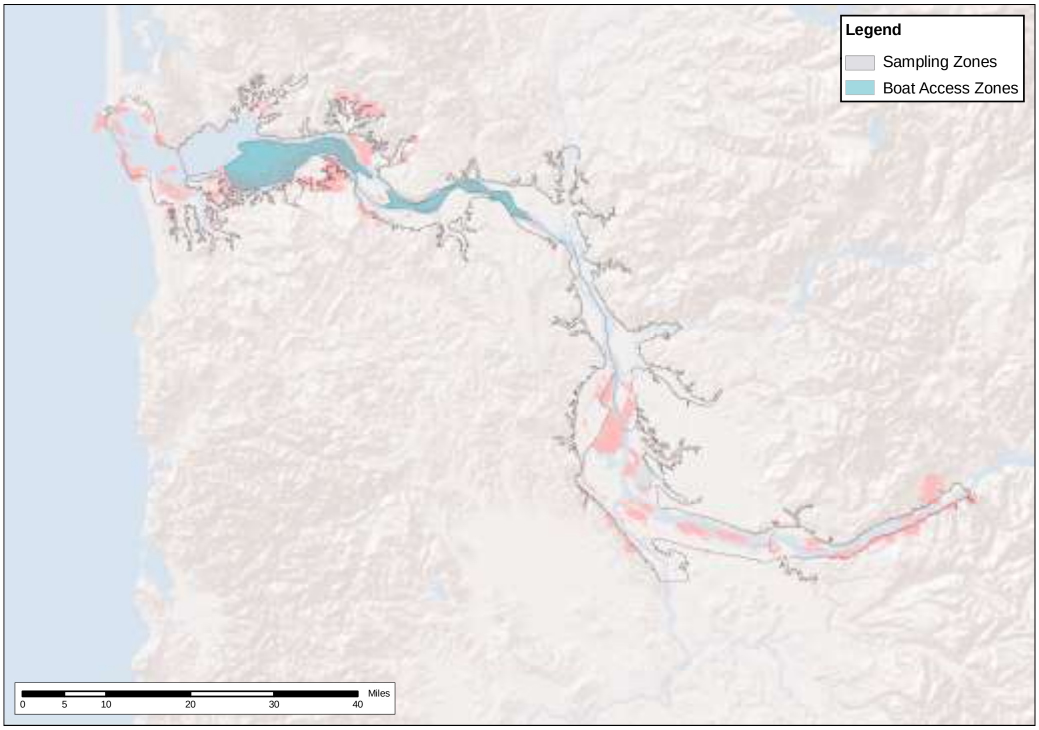


Figure 5. Boat access zones (blue) and state and federal owned properties (pink).

Accuracy assessment segment collection goals were met in all assessed classes. Appendix D details the numbers of segments per class used for determining accuracy of the final map.

Conclusions

The overall data collection effort proved to be efficient and relatively easy, based on samples collected per day, using the methods proposed in the field sampling plan. The GIS based collection tool proved to be useful in collection of the various land cover components. Access issues proved to be less important than previously anticipated primarily because of the biologist's familiarity with local vegetation and the ease of data collection. This made accessing sites on-foot to identify plants less of an issue.

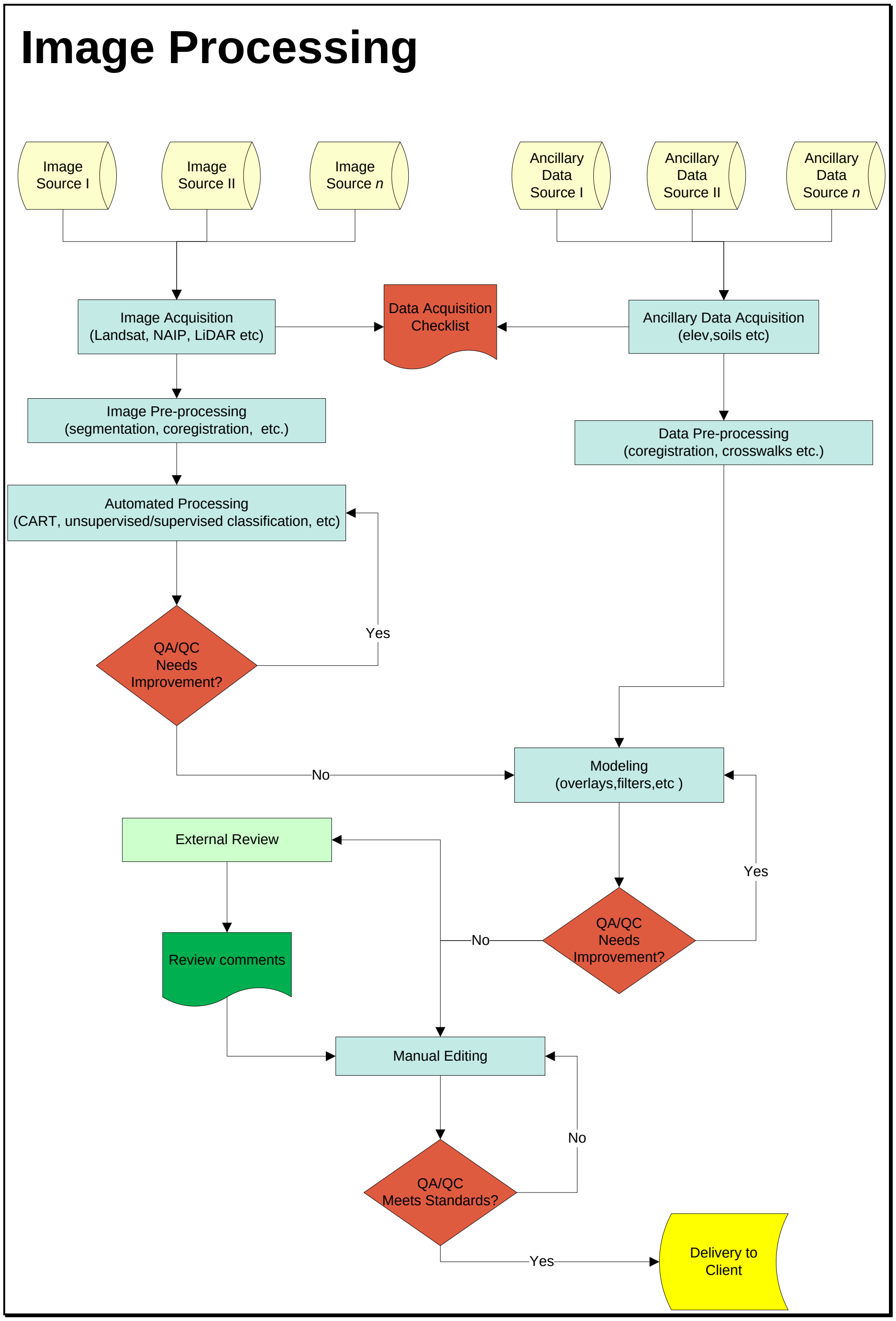
Process Introduction

This section will introduce the techniques used to develop a high resolution land cover dataset from high resolution imagery. This approach uses the spatial advantages of high resolution imagery in combination with the spectral advantages of multi-date Landsat. This technique moves from the traditional pixel based classification to an object orientated classification for high resolution imagery, so that the classification uses the information about spectral brightness and texture but also context and shape of segments. The object orientated approach produces a map that is more similar to that that would be developed by hand delineation which often makes the map easier to interpret for the user, since humans generally prefer elements that have defined boundaries than continuous surfaces.

The approach developed uses image segmentation, classification and regression tree analysis, accurate field data, photo interpretation and modeling with ancillary datasets to produce the final land cover map.

Fortunately the LCREP project will has freely available imagery from the National Agricultural Imagery Program (NAIP). In addition there are LiDAR data available for the project area. These data were obtained by the project team in addition to multi-date Landsat scenes. The high resolution datasets are used to create segments, and LiDAR, Landsat and high resolution imagery will be used as the independent variables for the regression tree approach described here.

Figure 6. Image Processing Overview



Imagery Selection Summary

In order to develop an up-to-date land cover dataset for the LCREP, Sanborn selected three sets of Landsat imagery which cover the project area and were collected from the summer of 2006 to the fall of 2009. Sanborn also used high resolution 4 band digital imagery that was collected by the National

Agriculture Imagery Program (NAIP) to generate the land cover dataset. This section provides technical specifications for the sensors used to collect these data sources and summarizes the process Sanborn used for selecting the images to be used for classification.

Landsat

The Landsat satellite program is the longest continuously acquired collection of space born moderate resolution land remote sensing data. Currently, Landsat 5 and Landsat 7 sensors are operational and data collected by these sensors are available free of charge from the United States Geological Survey (USGS) (http://glovis.usgs.gov). Most images acquired from the USGS have been processed to the Standard Terrain Correction which provides systematic radiometric and geometric accuracy by incorporating ground control points from a Digital Elevation Model for topographic accuracy. Although imagery from both Landsat 5 and Landsat 7 were initially considered for selection, Landsat 7 acquisitions since May 2003 are collected and archived without scan line correction, due to the Scan Line Corrector failure. This accounts for an approximate loss of 22 percent of the entire scene and severely limited the use of these data sources for the LCREP project. As a result, all Landsat images selected for the project were acquired from the Landsat 5 TM satellite.

Landsat 5 TM imagery has seven spectral bands, including a thermal band, with the following spectral and spatial resolution specifications:

- Band 1 Visible (0.45 – 0.52 µm) 30 m
- Band 2 Visible (0.52 – 0.60 µm) 30 m
- Band 3 Visible (0.63 – 0.69 µm) 30 m
- Band 4 Near-Infrared (0.76 – 0.90 µm) 30 m
- Band 5 Near-Infrared (1.55 – 1.75 µm) 30 m
- Band 6 Thermal (10.40 – 12.50 µm) 120 m
- Band 7 Mid-Infrared (2.08 – 2.35 µm) 30 m

In order to select the best possible combination of Landsat imagery for the project Sanborn reviewed all archived Landsat 5 imagery that has been collected since 2006 for numerous variables of interest. Variables of particular importance included cloud coverage, seasonality, scene pair date, river water level, and any sensor processing errors such as banding or striping of the data.

The first step in the selection process was to visually inspect all post winter 2006 images through the USGS online viewing application and all images that were below 10% cloud cover over the study area were downloaded for further inspection. Next, using information river water level information compiled by LCREP the images were assessed for the water level at the time of data capture. Given the importance of identifying estuarine and tidal freshwater habitats in the project area, selection of images collected at lower water levels were considered critical to the success of mapping these habitats of interest. In addition to this low water level criteria, image collection date between image pairs also needed to be taken into consideration because the study area is not covered by a single Landsat image. Two Landsat images cover the study area and obtaining images that are close in calendar date was considered beneficial because images that are close in date capture more consistent phenological conditions and reduce the variation in solar angle and incidence between each set of images. Given this set of guiding criteria Sanborn selected 3 sets of Landsat imagery from row 28 and paths 46 (east) and 47 (west) (Table 2).

Table 2.

Most suitable based on tide/cloud combination
Suitable based on tide/cloud combination

Year	Date	Scene (W=47, E=46)	Sensor	Astoria Tide Level (ft, mlw)	Longivew Tide Level (ft, mlw)	Max Cloud Cover (%)
2009	21-May	46	TM5		4	3
2009	28-May	47	TM5	-1		3
2007	1-Jun	46	TM5		1.7	OK over AOI
2007	7-May	47	TM5	-0.8		OK over AOI
2006	16-Jul	46	TM5		1.8	3
2009	29-Jun	47	TM5	1		9
2008	7-Sep	46	TM5		0.5	3
2006	25-Sep	47	TM5	4		3

Sanborn selected 6 images that fall within 2 feet of the mean water level which were identified as suitable or most suitable based on tide and cloud combinations by LCREP. Two of the images Sanborn selected fall outside this water level criteria, but were chosen because they were the cloud free image that best matched the phenology of the seasonal set (September 2006) or were collected close to the day of their respective seasonal image (7 days apart for the May 2009 set). Thus, this combination of images incorporates images that are close in date while also maximizing low river conditions in the imagery stack.

In order to further minimize the influence of high river levels on classification Sanborn overlaid low water imagery on top of high water imagery when season sets merged together. This reduced the amount of area exposed to high water conditions because there was substantial overlap between east and west scenes. In addition, it was important to note that as long as the land cover of interest is distinguished on at least one scene, Classification and Regression Tree (CART) will likely be able to accurately classify the land cover of interest. For each Landsat path and row we had 3 images that cover a range of phenological conditions and also met low water level conditions.

NAIP

Technical Specifications

NAIP acquires digital ortho imagery during the agricultural growing seasons in the continental United States. A primary goal of the NAIP program is to enable availability of ortho-imagery within one year of acquisition. The tiling format of NAIP imagery is based on a 3.75' x 3.75' quarter quadrangle with a 300 meter buffer on all four sides. The NAIP quarter quads are formatted to the UTM coordinate system using the North American Datum of 1983 (http://www.apfo.usda.gov/FSA). Sanborn acquired NAIP imagery that was collected in 2009 over Oregon and Washington. Imagery was collected by two different contractors (Aerial Services Inc. for Oregon and Northwest for Washington) each capturing 4 band 1.0 meter data.

NAIP River Level Analysis

River level is a combination of many factors, the two most substantial are tidal influence and discharge from the Bonneville Dam. Given the priority for mapping tidal freshwater habitats imagery collected at low river levels are preferable to those collected at higher water levels. Many river and estuarine areas in the study area were covered by both Oregon and Washington NAIP collection efforts. For the majority of these areas the lower river stage imagery was supplied by NAIP to Sanborn. If the river was covered by multiple NAIP scenes the lower water imagery was selected by Sanborn for classification. A good example of the benefit of multiple NAIP coverage with regards to water level can be seen by comparing water levels for the collection of imagery over Multnomah County and Cowlitz County (Table 3). Flight lines from both of these counties cover the same portion of the Columbia River, but the Cowlitz County imagery was collected at a much lower water level compared to the Multnomah County imagery. Thus, for the portion of the study area where these two flight lines overlapped, the Cowlitz County imagery was chosen for classification.

Table 3. Table shows two sets of overlapping NAIP flights that covered a portion of the river. The Cowlitz County imagery was selected over the Multnomah County imagery.

County	Date	Acquisition Start Time (UT)	Acquisition End Time (UT)	Feet to MLLW
Multnomah County, OR	06/23/2009	10:21	10:29	5.5
Cowlitz County, WA	**08/01/2009**	**7:18**	**7:37**	**0.2**
Multnomah County, OR	06/23/2009	15:22	15:30	4.5
Cowlitz County, WA	**08/01/2009**	**7:43**	**8:00**	**0**

The major exception to this availability to choose among multiple sets of NAIP imagery occurs at the mouth of the Columbia River, where the width of the river prevented coverage by both Oregon and Washington NAIP flights. The Clatsop County imagery is at a much lower tide and was chosen where it overlapped with the Pacific County imagery. However, given the width of the river at its terminus, the Clatsop County did not extend completely across the river so the Pacific County imagery was the only available NAIP imagery for the most northern western portion of the river. The Pacific County imagery was flown at a high river stage (Table 4).

Table 4. Table shows feet to MLLW for Pacific County imagery. Areas covered by these images were not duplicated by the Oregon NAIP flights.

County	Date	Acquisition Start Time (UT)	Acquisition End Time (UT)	Feet to MLLW
Pacific County, WA	**09/11/2009**	**4:15**	**4:34**	**7.5**
Pacific County, WA	**09/11/2009**	**4:15**	**4:34**	**7.5**
Pacific County, WA	**09/11/2009**	**3:56**	**4:12**	**7.3**
Pacific County, WA	**09/11/2009**	**3:32**	**3:52**	**6.8**
Pacific County, WA	**09/11/2009**	**3:12**	**3:28**	**6.2**

Although the imagery that covers the north western portion of the river (Pacific County) was flown at a high river stage, many of the marshes and some of the sand bars and mud flats are still spectrally distinct from the surrounding water. The marshes and flats with distinct spectral signatures were identified and grouped by the segmentation software. Classification of these areas were not affected by the high river stage in the Pacific County NAIP imagery because spatial extent of these features were delineated in spite of river conditions and spectral information from low river stage Landsat images enabled CART modeling to correctly identify the features. Areas that were not adequately delineated by segmentation due to high river levels in this area include Knappton Cove and some portions of a large marsh and mud flat system extending west from the middle portion of the Astoria-Megler Bridge. These areas needed additional localized conditional modeling during the post classification editing stage of the project.

Atmospheric Corrections

Landsat Image Processing

The Estuary Partnership's request for proposals (RFP) for 2009-2010 land cover mapping along the lower Columbia River indicated that image processing should include radiometric and atmospheric corrections to the selected imagery. Although spectral information from Landsat 5 TM imagery was used in Sanborn's classification, the classification relied on multiple data sources and ultimately resulted in a high spatial resolution mapping product. Given the differences in resolution and methodology between the initial Landsat based classification and a high resolution product, it is important to clarify the methodology for radiometric and atmospheric corrections applied to the Landsat imagery included in the project.

Sanborn removed unwanted radiometric and atmospheric influences in Landsat imagery by normalizing these conditions between the images contained in each seasonal dataset used in the classification. Image normalization proceeded using techniques pioneered by personnel of the U.S. Bureau of Land Management (Eckhardt et al. 1990) and used in image processing performed by the Multi-Resolution Land Characteristics (MRLC) Consortium. Image normalization is achieved by developing regression equations between the brightness values of "normalization targets" present in the base scene and the scene to be normalized. Image normalization reduces pixel brightness value (BV) variation caused by non-surface factors so variations in pixel BVs between dates can be related to actual changes in surface conditions. The non-surface factors accounted for by this method include the majority of variables included in the Estuary Partnership's initial RFP. These variables include differences in solar irradiance (solar distance and angle), atmospherically emitted path radiance and path transmittance, ground-level downward irradiance, optical thickness, and upward radiance. Surface temperature is the one variable listed in the Estuary Partnership's RFP not fully accounted for by this processing method. However, given the large spatial resolution of the Landsat TM 5 thermal band (120 m) compared to other Landsat bands and image segment size, it was decided that a surface temperature variable would not aid substantially in discerning land cover types for this project.

Image processing resulted in 3 sets (spring, summer, and fall) of normalized Landsat TM 5 imagery that covered the entirety of the study area. This enabled the use of image analysis logic developed for a base scene to be applied to the other scenes. It also developed a base dataset to which other images could be normalized for change detection in the future. Extension of image analysis logic throughout the study area helped ensure a robust and accurate classification and helped meet the goal of producing a highly accurate land cover map of the lower Columbia River.

Image Processing

Description of steps: LCREP Land Cover Classification

Step 1: Image Quality Review

Imagery underwent a quality control review before processing. The extents and projections were checked to match project requirements. The imagery was assessed for tonal balance, pixel drop out, shadow length, cloud and shadow cover, and haze.

Step 2: Ancillary Data Used

A LiDAR derived DEM and a Vegetation Height layer were generated from LiDAR data were provided by the Army Corps of Engineers. These elevation datasets were very useful for distinguishing between grass, shrub, and forest and for the identification of wetland communities.

The National Wetlands Inventory was used as a base for identifying wetlands; data collected in the field supplemented this dataset. NOAA provided a shapefile of the coastline that separated land from several benthic classes. In many cases this was used to help distinguish reef from shoreline.

Step 3: Creation of Land Cover Product: Initial Segmentation

Photo interpretation and manual delineation of image objects has historically been the means of generating polygons from imagery. The elements of photo interpretation are the guiding factors people use to visually group land cover objects; color, shape, texture, context, brightness, etc. Automated methods for developing polygons from imagery have been rapidly advancing, and are widely utilized in remote sensing of land cover products. Image segmentation is the process of separating an image into spectrally homogeneous polygons that depict distinct regions on the ground. Sanborn used eCognition software to generate image segments. Image segments are derived using several parameters, such as image texture, shape, and size. Multi-scale segmentations will be calculated with varying "scale" parameters. The scale parameter affects the relative size of output polygons, although there is not a direct relation between the input scale and the number of pixels per polygon. Image segmentation was completed using the multi-spectral (1 m) imagery in order to group like spectral and textural objects within the imagery. Several iterations of review were conducted before a final segmentation parameter optimization was reached. Image segments served as the basic unit of land cover classification, as opposed to traditional "per-pixel" imagery classifications.

Step 4: Initial Classification

Automated classification of the image segments was performed in a Classification and Regression Tree (CART) analysis that included a wealth of spectral and ancillary information for each segment. CART is predictive software that uses nonparametric statistics to generate rule sets for classifying objects. Rule sets are defined based on how the values of the dependent variable compare to the values of the known data (in this case the training data). CART has been known to produce accurate automated land cover classifications from training data. The statistics associated with remotely sensed land cover data can be highly complex and may not subscribe to a regular or parametric distribution therefore the classification tree approach is appropriate. Also, the software is capable of handling large databases with minimal processing time requirements.

This initial automation was done using See5, the CART application for predicting thematic data. The original See5 program is designed to run on raster data. Sanborn developed a program that enabled streamlined See5 processing for vector data so that vector image segments could be classified instead of image pixels.

The resulting classification was further refined through logical rule-sets built within ArcMap software. These rule sets were applied both globally and locally.

Automated Classification Edits

As with any automated or semi-automated land cover classification there are often inconsistencies in the land cover map. The final step before map finalization was to remove inaccuracies through manual segment labeling as interpreted by an analyst. Editing was done in the ArcMap software environment where individual segments or groups of segments can be reshaped and/or recoded. Manual edits focused on edges of objects such as expanding a grassy field, which may be incorrectly classified due to within class spectral or textural inconsistencies. Further refinement is also often done on areas within the image affected by cloud and cloud and terrain shadow.

Wetland Selection Methodology

The methodology for wetland selection was of particular importance to this project. Several ancillary data layers were of particular importance during this phase: NWI, a tidal/diked spatial boundary produced by LCREP, and the USACE LiDAR data. Merging areas classified as tidal/diked with NWI, followed by selecting image segments whose centroids fall within this merge. This initial selection was further refined by creating a slope and elevation mask derived from the LiDAR data. The segments that were selected then served as our wetland/upland mask, and land cover classes were revised as a result.

Map Finalization

Sanborn received comments from LCREP and NOAA regarding the quality of the draft map. These comments were incorporated into the final map, redelivered and several iterations of QC and editing were completed. Each comment was tracked and responded to. The final map was submitted to NOAA as the final delivery, including metadata.

Accuracy Assessment

Purpose of Accuracy Assessment

The purpose of an accuracy assessment is to provide a quantitative measure of reliability for the vegetation map.

The accuracy of the draft vegetation map was assessed quantitatively by using an error matrix. The error matrix is a square array of numbers set out in rows and columns which express the number of pixels assigned to a particular category in one classification relative to the number of pixels assigned to a particular category in another classification. The columns usually represent this reference data while the rows indicate the classification generated from the remotely sensed data (Congalton and Green 1999).

Accuracy Assessment Point Collection

The primary goals of the AA sampling strategy were: 1) to have a broad distribution of points throughout the study area, 2) to ensure AA points were non-coincidental with training points, and 3) to avoid clustering of points for one vegetation class.

The AA was conducted at the scale of image segments. Although the initial field sampling campaign produced a large number of reference samples, a high percentage of samples within each class were contiguous and thus highly spatially auto correlated.

In order to assess the number of non-spatially auto correlated field samples and develop target amounts of AA points per class, contiguous segments labeled as the same vegetation class were dissolved (Figure 7). To further reduce spatial autocorrelation, the dissolved sample segments were buffered 15 meters and dissolved into any intersecting segment of the same class.

The buffering and second dissolve process was necessary to eliminate artifacts produced by very small segments that separated large contiguous areas labeled as the same class. For example, adjacent agricultural fields separated by a narrow hedge row that were identified as two sampling polygons after the initial dissolve were combined after the buffer and second dissolve process.

Target numbers for AA segments per class were developed after the first field sampling campaign and dissolving processes were completed. Target AA numbers were based on the number of dissolved samples within each vegetation class. A minimum target of 50 segments was set for each vegetation class (Congalton per comm.). Additional AA segments were selected for vegetation classes with greater than 100 dissolved polygon samples based on the proportion of field samples available, a priori knowledge of the relative distribution of the class across the study area (excluding water), and the spatial distribution of sample polygons. Ultimately, the number of AA segments was limited by the number and spatial distribution of field samples.

Only one AA segment was selected from each buffered and dissolved polygon. If an AA segment was selected from a dissolved polygon, all segments within that polygon were removed from the training data. This process of selecting AA segments served to help increase the distribution of AA across the study area and reduce the spatial autocorrelation between AA and training sites.

Figure 7. Dissolving and Buffering Process

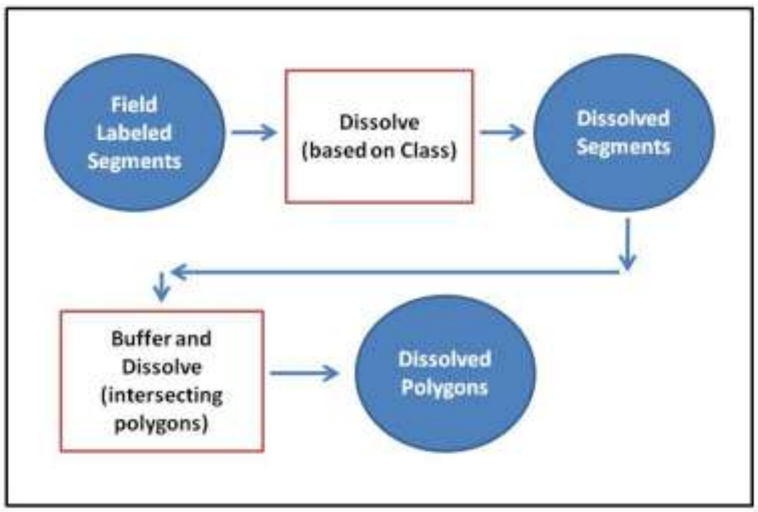


Table 5. Buffered and dissolved polygons and AA segments collected for each vegetation class.

Name	Total buffered and dissolved Polygons	Total AA Segments selected
Coniferous Upland Forest	105	54
Deciduous Upland Forest	235	54
Coniferous Wetland Forest	103	35
Deciduous Wetland Forest	204	50

Upland Shrub/Scrub	170	50
Wetland Shrub/Scrub	99	52
Upland Herbaceous	231	50
Wetland Herbaceous	352	55
Agriculture	129	55
Tree Farms	72	43
Bare	87	50
Mud	170	50
Sand	135	50
Urban - Impervious	113	50
Water	119	51

Target amounts for AA segments were met for all classes except Coniferous Wetland Forest, Tree Farms, and Sand. Tree Farm AA points were limited by the number of sample polygons and the close spatial distribution of many of the tree farm polygons. Of the 72 distinct polygons that were available to choose Tree Farm AA segments from, many polygons were located within very close proximity to each other. Coniferous Wetland Forest samples were also limited by a small number of distinct sample polygons (the lowest off all naturally occurring vegetation classes). A low number of sample polygons for the Wetland Coniferous Forest Class is not unexpected given the proportionally small and clustered distribution of this vegetation system within the study area.

Results

Results of the Accuracy Assessment

Overview

- Overall Accuracy for all assessed classes is 86%
- Kappa statistic is 85%
- Average Accuracy for all natural vegetated systems is 81%

Overall accuracy for the ecological systems is 86%. The Kappa statistic was 85%, representing good agreement between the reference data and the map (Congalton and Green, 1999). The Kappa statistic adjusts the estimate of overall accuracy for the accuracy expected from a purely random assignment of map labels and is useful for comparing different matrices.

Most ecological systems fell above 75% per-class accuracies (user's and producer's) and overall accuracy for all natural vegetated systems is 81%. Overall accuracy for all managed and non-vegetated classes is 93% (user's and producer's). In summary, this assessment shows that there is a high degree of agreement between reference data and the land cover map.

Confusion Among Ecological Systems

The majority of confusion for any one forested system is found within among the other forest classes. This pattern is strongest for the two coniferous forest classes. Shrub-Scrub and Herbaceous Classes confusion is dispersed over a larger number of ecological systems compared to the forest classes. The majority of misclassified mud and sand reference points are constrained between these two classes.

The Upland Shrub-Scrub, Wetland Shrub-Scrub, and Upland Herbaceous have the lowest producer's accuracies. The Upland and Wetland Deciduous Forest and Upland Herbaceous classes have the lowest user's accuracies. These classes have user's or producers accuracies below 80%, but most have accuracies above 70%.

Wetland Herbaceous and Agriculture classes have high user's accuracies and compared to their producer's accuracy. This large difference indicates that these systems may be over represented on the map. Additionally, the Agriculture class is most often confused with Upland Herbaceous and Upland Shrub-Scrub. This result is not unexpected give that land use the most important variable for distinguishing these classes and this variable can be difficult to derive from imagery and ancillary data across the study area.

The majority of all naturally vegetated AA segments that were misclassified were confused with the upland or wetland equivalent within each class, or were assigned the correct upland/wetland call in another class. The overall accuracy of the upland and wetland division was assessed by creating an error matrix with the naturally vegetated classes dissolved together on the upland/wetland attribute. The overall accuracy of the upland/wetland division was found to be 94%.

Table 6. Error Matrix

Reference Data

Map Data	Upland Coniferous Forest	Upland Deciduous Forest	Wetland Coniferous Forest	Wetland Deciduous Forest	Upland Shrub-Scrub	Wetland Shrub-Scrub	Upland Herbaceous	Wetland Herbaceous	Agriculture	Tree Plantations	Bare	Mud	Sand	Urban - Impervious	Water	**Total**	**User's Accuracy**
Upland Coniferous Forest	**45**	4	1													50	90%
Upland Deciduous Forest	8	**45**		7	2											62	73%
Wetland Coniferous Forest		1	**30**													31	97%
Wetland Deciduous Forest	1	4	4	**42**		3				2						56	75%
Upland Shrub-Scrub					**33**	5	1		1							40	83%
Wetland Shrub-Scrub				1	4	**38**		3								46	83%
Upland Herbaceous					5		**37**				3					45	82%
Wetland Herbaceous					4	5	3	**52**	2		1	2	1			70	74%
Agriculture					2	1	8		**52**	2	1					66	79%
Tree Plantations										**39**						39	100%
Bare											**45**					45	100%
Mud												**40**	2			42	95%
Sand							1					6	**47**			54	87%
Urban - Impervious														**50**		50	100%
Water												2			**51**	53	96%
Total	**54**	**54**	**35**	**50**	**50**	**52**	**50**	**55**	**55**	**43**	**50**	**50**	**50**	**50**	**51**	**749**	
Producer's Accuracy	83%	83%	86%	84%	66%	73%	74%	95%	95%	91%	90%	80%	94%	100%	100%		

Overall Accuracy	86%
KHAT	0.85

Appendix A: Dichotomous Land Cover Key

The minimum mapping unit is 0.25 acres.

1 Is the segment > 75% water?

If yes, then **1.1**
If no, then **2**

1.1 Is the water covered by > 80% of aquatic vegetation? (examples include: algal mats, detached floating mats, and rooted vascular plant assemblages)
Then ………………………………………………………… **Aquatic Beds**

1.2 Else.……………….…………………….……………………….……….. **Water**

2 Is the segment ≥ 50% urban development and Impervious surfaces such as roads, railroads, compacted bare earth (dirt roads), piers, docs, bridges (permanent fixtures) ?

If yes, then **2.1**
If no, then **3**

2.1 Is the land cover composed of urban development and Impervious surfaces such as roads, railroads, compacted bare earth (dirt roads), piers, docs, bridges (permanent fixtures)?.. **Urban - Impervious**

2.2 Is the land cover composed of relatively permanent in-water and over-water structures such as bridges, pilings and abandoned log rafts? ………… **Urban – Other**
(Urban – Other will largely be added via GIS modeling and editing)

3 Does vegetation cover < 10% of the segment

If yes, then **3.1**
If no, then **4**

3.1 Is ≥ 50% of the barren area Mud?
(Mud is defined here as unconsolidated material characteristic of particle size smaller than sand, frequently with a high moisture content. These materials are subject to inundation and redistribution due to the action of water. Characterized by substrates lacking vegetation except for pioneering plants that become established during brief periods when growing conditions are favorable. Erosion and deposition by waves and/or currents produce a number of landforms representing this class. Examples include Estuarine/Palustrine Emergent Mud Flats or Bars.)
If yes, then……………………………………………………………..**Mud**

3.2 Is ≥ 50% of the barren area Sand?
If yes, then……………………………………………………………..**Sand**

3.3 Else ……………………………………………………………………….**Bare**
(Bare is further split into Rock and Cobble using GIS modeling)

4 Is > 50% of the vegetation adapted to intensive human manipulations?

If yes, go to **4.1**
If no go to **5**

4.1 Is > 50% of the vegetation anthropogenically manipulated in structure and composition, including: parks, lawns, athletic fields, golf courses, and natural grasses occurring around airports and industrial sites.
If yes, then………………..……………………………**Open Space Developed**

4.2 Is > 50% of the vegetation active agricultural crops, frequently mowed pastures, or tree plantations?
Is > 75% of the vegetation tree plantation?
I If yes………………………………………………………… **Tree Plantation**
If no, then …………………………………………………….. **Agriculture**

5 Is ≥ 20% of the vegetation cover tree canopy?
(Trees are defined here as woody vegetation > 5m in height)

If no, go to **6**.
If yes, then,

Is the segment > 50% upland?
(Uplands generally lack hydric soils. Examples of wetland indicator plants would include, Trees: *Fraxinus latifolia*, *Thuja plicata*, and *Alnus rubra,* Shrubs: *Cornus stolonifera, Malus fusca, Physocarpus capitatus, Spiraea douglasii, Spiraea douglasii,* and *Salix spp.,* Herbs: *Mentha spp. Polygonum spp., Typha spp., Lysichiton americanum,* and *Lythrum salicaria,* Grasses: *Phalaris arundinacea, Alopecurus spp.,* and *Agrostis spp.,* Other: *Equisetuim spp., Carex spp.,* and *Juncus spp.*)

If yes, then go to **5.1**
If no, go to **5.3**

5.1 Is the tree canopy within the segment ≥ 50% coniferous tree species?
If yes, then……………………………………….**Coniferous Upland Forest**

5.2 Else …………………………………………………. **Deciduous Upland Forest**

5.3 Is the tree canopy within the segment ≥ 50% coniferous tree species?
If yes, then……………………………………..**Coniferous Wetland Forest**
If no, go to **5.4**

Modifier – These classes will be added via GIS modeling

5.3.1 *If yes, is ≥ 50% of the segment diked?*
*If yes, then……………………….****Coniferous Wetland Forest – Diked***

5.3.2 *Is ≥ 50% of the segment under tidal influence?*
If yes, then …………………….. ***Coniferous Wetland Forest – Tidal***

5.3.3 *Else …………………………..* ***Coniferous Wetland Forest – Non-tidal***

5.4 Is the tree canopy within the segment > 50% deciduous tree species?
If yes, then………………………………………. **Deciduous Wetland Forest**

Modifier – These classes will be added via GIS modeling

5.4.1 *If yes, is ≥ 50% of the segment diked?*
If yes, then………………………. ***Deciduous Wetland Forest – Diked***

5.4.2 *Is ≥ 50% of the segment under tidal influence?*
If yes, then ………………………. ***Deciduous Wetland Forest – Tidal***

5.4.3 *Else …………………………….* ***Deciduous Wetland Forest – Non-tidal***

6 Is ≥ 20% of the vegetation covered in shrub canopy?
(Shrubs are defined here as woody vegetation ≤ 5m in height)

If no, go to **7**
If yes, then,

6.1 Is the segment > 50% upland?
(See *Uplands* definition under **5**)
If yes, then ………………………………………. **Upland Shrub/Scrub**

6.2 Else ………………………………………………. **Wetland Shrub/Scrub**

Modifier – These classes will be added via GIS modeling

6.2.1 *Is ≥ 50% of the segment diked?*
If yes, then……………………………. ***Wetland Shrub/Scrub – Diked***

6.2.2 *Is ≥ 50% of the segment under tidal influence?*
If yes, then……………………………. ***Wetland Shrub/Scrub – Tidal***

6.2.3 *Else …………………………..* ***Wetland Shrub/Scrub – Non-Tidal***

7 **Else** *(Is ≥ 20% of the vegetation grammanoid or herbaceous?)*

7.1 Is the segment > 50% upland?
(See *Uplands* definition under **5**)

If yes, then ………………………………………….. **Upland Herbaceous**

7.2 Else ………………………..……………………..……. **Wetland Herbaceous**

Modifier – These classes will be added via GIS modeling

7.2.1 *Is ≥ 50% of the segment diked?*
If yes, then………………………………..…… ***Wetland Herbaceous – Diked***

7.2.2 *Is ≥ 50% of the segment under tidal influence?*
If yes, then……………………………………. ***Wetland Herbaceous – Tidal***

7.2.3 *Else ……………………………* ***Wetland Herbaceous – Non-Tidal***

Appendix B. Numbers of segment-based samples by sample source.

Class	Sample Sources								Grand Total
	Field Session 1	Field Session 2	Photo Interpreted	NRI	CASI 1	CASI 2	CASI 3	CASI 4	
Coniferous Upland Forest	247	199	39		23	39		3	550
Deciduous Upland Forest	640	106		1794		45	7	146	2738
Coniferous Wetland Forest	1350		3		155	5			1513
Deciduous Wetland Forest	931	55	1	199	4	34	468	138	1830
Upland Shrub/Scrub	91	3	26	226				22	368
Wetland Shrub/Scrub	650	29	15	44	1	6	7	2	754
Upland Herbaceous	106	27	2	2025	19		121	69	2369
Wetland Herbaceous	977	23		668	146	1405	91	554	3864
Agriculture	443	211	46	11	643	233		544	2131
Tree Farms	82	229	105						416
Bare	13	70	91		2				176
Mud	453	14				926			1393
Sand	147	58	22	66	7	207	19	23	549
Urban - Impervious				5442	25	1		264	5732
Water				164					164
Grand Total	6130	1024	350	10639	1025	2901	713	1765	24547

Appendix C. Number of segment-based samples by sample zone.

	Sample Zones									
Class	1	2	3	4	5	6	7	8	9	Grand Total
Coniferous Upland Forest	131	67	65		5	41	105	87	49	550
Deciduous Upland Forest	32	123	34		3	179	419	1780	168	2738
Coniferous Wetland Forest	237	611	502	158		5				1513
Deciduous Wetland Forest	114	21	41	84	558	304	314	365	29	1830
Upland Shrub/Scrub	10	3				89	12	207	47	368
Wetland Shrub/Scrub	16	11	134	453	13	15	1	61	50	754
Upland Herbaceous	19	6	3		121	66	51	2057	46	2369
Wetland Herbaceous	249	203	221	1514	79	570	96	856	76	3864
Agriculture	650	260	79			682	351	93	16	2131
Tree Farms		41	352		1		21	1		416
Bare	6	1				92	30	41	6	176
Mud	208	8	65	1092			1		19	1393
Sand	7	10	17	215	47	51	102	90	10	549
Urban - Impervious	25	1				3	259	5211	233	5732
Water								145	19	164

Grand Total	**1704**	**1366**	**1513**	**3516**	**1762**	**827**	**2097**	**10994**	**768**	**24547**

Appendix D. Number of accuracy assessment segments by class.

Name	Total AA Segments selected
Coniferous Upland Forest	54
Deciduous Upland Forest	54
Coniferous Wetland Forest	35
Deciduous Wetland Forest	50
Upland Shrub/Scrub	50
Wetland Shrub/Scrub	52
Upland Herbaceous	50
Wetland Herbaceous	55
Agriculture	55
Tree Farms	43
Bare	50
Mud	50
Sand	50
Urban - Impervious	50
Water	51

Literature Cited:

Congalton, R. and K. Green. 2009. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*. 2nd Edition. CRC/Taylor & Francis, Boca Raton, FL 183p.

Eckhardt, D.W., Verdin, J.P., and Lyford, G.R. 1990. *Automated update of an irrigated lands GIS using SPOT HRV imagery*. Photogramm. Eng. Remote Sens. 56:1515-1522.