**Geographic Information Systems (GIS) Materials and Methods.**

GIS methods section of this report describes the process by which source data, i.e. aerial photography, its interpretations, and field observations becomes a final spatial data layer; viewable on computer screens, printable as a map, and capable of various types of summary reports, and analysis.

One of the first steps in the planning process was decide among myriad techniques, and multiple paths to accomplish the end goal. The various options included

1) digitizing vegetation delineations directly from aerial photos,

2) transferring vegetation delineations to DOQQ, then digitizing from these,

3) registration of digital aerial photos and heads up digitizing,

4) digital classification,

5) scanning the vegetation delineation, followed by raster to vector line following conversion (Arc/Scan).

Various options were ruled out due to relative ranking in terms of what was believed to be the most efficient, accurate, and utilitarian approach. Digitizing from aerial photos is relatively quick, but sometimes insufficiently corrects for the inherent distortion within an aerial photo. Transferring line-work to DOQQ's allows an effective registration since the DOQQ's are planimetric, but the transfer process relies on multiple stepwise adjustments between an overlay, and the DOQQ, since creating a DOQQ at exactly the same scale as an aerial photo would be impossible. Given a 6 year interval between the date of the DOQQ and the vegetation study, spatial control may be difficult to identify. Digitial classification sounded interesting and fast, but hue and brightness variation between flights and within a single frame could have posed edge matching issues. At large scale displays, the stairstep edge artifact of raster conversion can be detracting. Scanning the linework had been tested before. Editing and edge matching issues are considered costs that counter the speed of digital conversion. With some reluctance, option 3) was chosen. The benefit of image registration of aerial photos is a data product that can be shared with various agencies, and reused in the future, or for different purposes. This process was considered to be efficient because the digitizing would be done in a single, seamless coverage, which avoids the cost of stitching together photo-based coverages, and edge-matching the line-work, and attributes.

**Georeferencing:**

1) Photos were scanned at 300 dots / inch on a HP 6300 scanner, saved in compressed jpeg format, using "excellent" quality. Jpeg compression can cause degradation if this parameter is set to maximum compression. Output file size per frame is ~ 5 Megabytes. Fiducial marks were not included in the scan, or used in the process of registration. Note:

to speed processing every other photo was skipped. Adjacent photos had a 60% overlap, which is perfect for stereo interpretation, and orthoregistration. Skipping every other photo resulted in photos with about a 20% overlap.

2) Image to map registration. ERDAS Imagine was used to transform the scanned aerial photography to map projection. Source control points were selected from 1993 USGS Digital Ortho Quarter Quads at 1 meter resolution, and a real world positional accuracy of a 1:12,000 scale map. The cell sampling rate (or resolution) on the registered aerial photos is 1 meter on the ground.

For this project, the 2nd order polynomial transformation was used. A second order polynomial fits the typical scale changes in an aerial photograph of flat terrain very well.

Photo-scale changes due to terrain effects were not important because the project area is in very flat terrain, so orthorectification was not performed. The residuals (how far each measured point deviates from its mathmatically predicted location), which are reported as Root Mean Square Error (RMSE), may be interpreted as how well the image matches the map projection, but the mean spatial error in the image may be higher than the residuals imply. Solutions for the polynomial Root Mean Square Error (RMSE) values were targeted for less than 1 meter, or equivalently, one pixel. For some frames, the RMSE averaged slightly greater than one. The anicdotal test for goodness of fit was to overlay the registered photos with the Orthophoto quad, and to compare the fit between adjacent quads.

**Digitizing:**

Digitizing is the process by which lines on a map are captured in an electronic format. Lines are represented by a series of x, y coordinate pairs representing the locations of line start and end points, and the positions of line direction changes. This process can be achieve with special electronically sensitive tracing boards or by capturing on-screen mouse movements, to name a few.

Aerial Photos were used as a backdrop on the screen in an Arc/Info environment. ArcEdit was used to trace polygons using the patterns seen on the displayed aerial photo, and using the delineated photos as reference.

An Arc Macro Language (aml) menu was written to handle the basic editing functions: Add an image to the backdrop, set editing scale, set the feature type, file save, etc. The scale set during digitizing was typically 1:4800, but often a larger scale was set to digitize finer detail.

Editing of polygons was undertaken to utilize built in routines in Arc/Edit to build polygon topolog, and to automatically add label points.

At the outset, on-screen digitizing of delineated vegetation proved to be troublesome. Comparing the patterns delineated on the photo and replicating those patterns while

digitizing required a lot of visual referencing of two separate sources, which was a very time consuming process. The process was originally visualized to only use the patterns in the digital version, without requiring a match to delineations drawn on photos.

To improve the process, a test was performed to see if modification of our technique could increase efficiency. A sample photo was scanned with the delineations, then registered. This combined photo was then used as a backdrop. Personnel performing the digitizing reported that they could capture the delineations many times faster, and were more assured that they were following the delineation more precisely. As a result of this test, all of the photos were re-scanned with the mylar overlay showing the delineations.

Display response tended to slow down as polygons were digitized. Another deviation to the original plan was to digitize in nine separate coverages, then merge these when the process was completed. To facilitate the merging process, the edge polygons of the completed coverage were copied into the new coverage. Digitizing would continue, building onto the row of copied polygons. At the completion of digitizing, the polygons which were copied from the adjacent coverage were deleted, so that they would not be redundant entries when the separate coverages were joined back together.

At the completion of a digitizing session, the topology of polygons was rebuilt, adding label points to newly digitized polygons, etc. Another check performed was to list label errors. This would list any illegally formed polygons, such as those not containing a label point, or containing two label points with different id numbers.

At the completion of digitizing, but before joining adjacent coverages, unique identification numbers were assigned to each polygon. The first coverage had on the order of five to six thousand polygons. This was assumed to be typical of each of the following coverages as well. In order to maintain a numbering system that was unique, and since a system assigned numbering system began with 1 in each separate coverage, an offset factor of 10,000 was applied to the first coverage. This unique numbering, referred to as Unique-id, began with 10,001, and extending to 10,000 plus n, where n is the number of polygons in the coverage. Arc/Info automatically creates a number for each polygon, but these numbers change when editing, thereby necessitating a user assigned identifier. Subsequent coverages were assigned unique identifiers using an offset of coverage number * 10,000. Coverage 2 identifiers range from 20,001 through 20,000 + number of polygons in coverage 2. When coverages are merged the system assigned identifiers are unpredictable, but the Unique-id remains unchanged for the polygon.

These unique numbers served as a key field to link to the Access data base. Attributes for polygons were keyed into the database using the Unique-id numbers, and all the vegetation attributes.

Subsequent editing of the polygons was required because gaps / overlaps were detected. To accomplish these edits, the coverage was converted to an ArcView shapefile to facilitate editing without Arc/Info. The cause for these errors has not been determined.

**Random Selection for Accuracy Assessment**:

An accuracy assessment of the photo interpretation by field visit was desired. All of the polygons could not be field checked due to time, budget constraints. A random selection was desired, so that the results of the sample selected could be an indicator for map accuracy. The sample selection was constrained to public properties, and selected private properties for which access was granted. Due to limited time to perform field studies, only certain classes of vegetation were assessed. The number of polygons was selected for each class based on estimated variance of proportion correct, and a bounding variable.

The selection process proceeded as follows:

1) select all polygons in the sample frame of properties accessible.

2) remove as candidates for selection any polygon that had been visited in the field.

3) for each class to be assessed, use a random number generation to select n polygons. A standard ArcView script is included to do this, it was modified to select a certain number, rather than percent. The random selection process is based on records, giving equal probability to both small and large polygons.

4) centroids for polygons were downloaded into a GPS unit, and maps of selected polygon boundaries, and centroids were plotted over aerial photos to provide field crews a means to reconnoiter to the polygon which was checked.

The formula for sample size is based on Cochran (1963), <u>Sampling Techniques, 2<sup>nd</sup></u> <u>Edition</u> (p. 75):

$$n = (t^2 pq)/d^2$$

> n = number of samples

> t = abscissa of a normal curve that cuts off an area of a (alpha)

> p = estimated variance, proportion correct

> q = 1- p

> d = discrepancy.

For this sampling exercise, the following parameter were set for all classes: alpha = .05, t = 1.96, d = .2, p is estimated for each class in the table below, under the column Estimated Proportion correct.

For the first class, the number of samples, n, is calculated by:

$$n = (1.96^2 * .95 * .05) / .2^2$$

$$n = ( 3.8416 * 0.0475) / .04$$

n = 4.5, or rounded up, 5 samples

Following is a table of vegetation classes assessed for accuracy, the estimated variance (proportion correct), and number of samples needed.

| Final Vegetation Code | Classification Name | Estimated Percent Correct | Number of Samples |
|---|---|---|---|
| 103 | Phragmites australis | 95 | 5 |
| 116 | S. califoricus/S. Acutus | 80 | 16 |
| 123 | Typha species (generic) | 80 | 16 |
| 137 | Scirpus maritimus | 75 | 19 |
| 141 | Distichlis spicata | 90 | 9 |
| 142 | Distichlis / annual Grasses | 90 | 9 |
| 157 | Scirpus (ca or acutus)-Typha sp. | 80 | 16 |
| 159 | echinocloa-polygonum-xanthiu | 90 | 9 |
| 160 | Distichlis-Juncus-Triglochin-Gla | 90 | 9 |
| 162 | Sc. ca-Sc. ac/ Rosa | 90 | 9 |
| 227 | Annual Grasses/Weeds | 90 | 9 |
| 231 | Annual Grasses (generic) | 95 | 5 |
| 311 | Atriplex trangularis | 75 | 19 |
| 316* | Atriplex/Sesuvium | 75 | 19 |
| 324 | Lepidium (generic) | 95 | 5 |
| 342 | Cotula coronopifolia | 95 | 5 |
| 344 | Lotus corniculatus | 95 | 5 |
| 346 | Salicornia virginica | 95 | 5 |
| 347 | Salicornia / Annual Grasses | 95 | 5 |
| 348 | Salicornia / Atriplex | 80 | 16 |
| 356 | Salicornia / Sesuvium | 95 | 5 |
| 357 | Sesuvium verrucosum | 90 | 9 |
| 402 | Conium maculatum | 95 | 5 |

| 413 | Centaurea (generic) | 90 | 9 |
|---|---|---|---|
| 514 | A. lentiformis (generic) | 95 | 5 |
| 604 | Rosa californica | 90 | 9 |

**Hardware / Software Configuration:**

A variety of personal computers, and laptop computers were used to accomplish the GIS processing, and attribution. Registration of aerial photos, and mosaicking of photos was accomplished with Erdas Imagine version 8.3 on computers with a processor speed of 333 Mhz, and 128 MB ram. Most of the digitizing was accomplished with Arc/Info, version 7.2. Additional edits have been performed in the ArcView environment with ArcView version 3.x. The attribute database was developed in Access 2000 on a laptop computer. Links between the polygons and Access are performed dynamically by connecting through a Open Data Base Connection (ODBC), then joining the ArcView virtual table to the shapefile by the Unique-id key field.

Data backups have been written to Jazz diskettes, and written to CD's.

Field sample locations were documented using a Trimble GeoExplorer II Global Positioning System (GPS) receiver. Stationary positions were read for approximately 3 minutes to collect 180 readings, which were differentially corrected in Trimble Pathfinder Office from base station files collected from the US Forest Service Community Base Station in Sacramento, CA.

**Data Sets:**

The following data sets were created during the 1999 Suisun Marsh vegetation mapping process:

1) Vegetation coverage interpreted from aerial photography.

2) Access database of attributes for each polygon.

3) Vegetation Classification tables, and crosswalks to other classification schemes.

4) Registered natural color aerial photographs at one meter resolution. Photomosaics of aerial photos by 7 ½' minute quad sheet areas, and within the Suisun Marsh Study Area.

5) Satellite image classification for sample stratification.

6) Stratified random samples for field data collection.

7) Field sample GPS locations.

8) Field data on species, and relative composition. Data stored in Calif. Vegetation Information System (CVIS).

9) Additional field verification locations.

10) Accuracy assessment locations.

11) Suisun Marsh Study area, defined by a combination of boundaries and limited by the 10 foot contour line.

12) Property boundaries for selected owners, but without owner information.

The following data sets were acquired for the project:

1) SPOT satellite image, both 20 meter multispectral, and 10 meter panchromatic, June 1999. This data set is licensed by SPOT Image Corp., which limits redistribution rights.

2) Digital Orthophoto Quarter Quads at one meter resolution, produced by USGS.

The following existing data sets were used, and maybe redistributed:

1) USGS 7 ½ minute topo quads, in digital format.

2) 1:100,000 scale roads, levees.

3) Dept. of Fish and Game Lands, and Rush Ranch Open Space Preserve, Solano County.

4) Hydrology at 1:24,000.