Using SPOT satellite imagery to update the National Wetland Inventory:
A study of three Illinois counties
Janet Jarvis, Illinois Natural History Survey
Center for Wildlife Ecology
January 17, 2003
Technical report for the project funded in part by the Illinois Department of Natural Resources

Introduction
In the 1980’s, the United States Fish and Wildlife Service began working on the National Wetlands Inventory (NWI). The objective of the NWI is to map all the wetland areas throughout the United States. The Illinois portion of the NWI was completed in the middle 1980’s. Since the completion of the NWI, many errors in accuracy have been found. In addition to the problems with the accuracy of the original NWI, the landscape of Illinois has gone through many changes over the past twenty years. Many wetland areas have been destroyed due to the building of new homes and businesses, and drainage of lands for farming.

The objective of this research is to explore different remote sensing based methods for updating the National Wetland Inventory for Illinois using SPOT satellite imagery. Similar research has been conducted using Landsat MSS and TM data with a spatial resolution of 30 meters. SPOT data has a 20 meter spatial resolution, but until recently was only available with red, green and near infrared channels. The introduction of SPOT 4 brought a mid-infrared channel to 20-meter SPOT imagery. Before this time, Landsat was the leading source of imagery in wetland studies because it came with a mid-infrared band of imagery. SPOT 4 introduced a comparable image to that produced by Landsat at a higher spatial resolution.

Background
This research was completed on a Dell Dimension personal computer with an Intel Pentium 4 processor running the Windows 2000 operating system. Image processing was done in Erdas Imagine 8.5 and GIS was done in Arc View 3.1 and 3.3.

Three dates of SPOT imagery, both 20 meter multi-spectral and 10 meter panchromatic, corresponding to spring, summer and fall dates of 2001, were acquired for each study site. The areas that were chosen as study sites were Lake, Peoria and St Clair Counties, Illinois. Three dates were chosen in order to take advantage of leaves on/leaves off of trees in summer versus fall, as well as to find areas with standing water during the growing season, as is a criteria of the wetland definition provided by the U.S. Fish and Wildlife Service when the original National Wetland Inventory program was started (Cowardin, 1979).

Both multi-spectral and panchromatic images of each study site were acquired in hopes of using the 10-meter panchromatic image to resample the pixel size of the 20-meter multi-spectral image up to 10-meter resolution using the resolution merge tool built into Imagine. This would essentially produce a 10-meter multi-spectral image from the data. Unfortunately, the panchromatic and multi-spectral images were spatially shifted from one another just enough to cause the resulting image to appear blurry. Many techniques were used in an attempt to correct the shifting, but none resulted in a clear image. One advantage to using the resolution-merged image is to have a 10-meter multi-spectral image, which would obviously show more detail than a 20-meter multi-spectral or a 10-meter panchromatic image. However, due to the technique used by Imagine to accomplish the resolution merge, the pixel values are changes. The resulting image of a resolution merge process would be no more useful than a color-infrared photo in that
every pixel value would be lost and replaced, resulting in an image that can no longer be used in image processing applications (Incardona, verbal communication).

**Data Sources**

The three criteria that define a wetland are hydric soil, wetland vegetation and flooding of the land during the growing season (Cowardin, 1979). With these criteria in mind, it was decided to combine the satellite imagery with soil data in various ways to find a technique that is able to locate wetlands in the satellite images. In the various experimental mapping methods, a number of ancillary data sources were obtained and used.

**Soil**

Soil data for St Clair County was obtained from the Soil Conservation Service (SCS) of the National Resource Conservation Service (NRCS). Soil data for Lake and Peoria Counties was obtained from the Illinois State Geological Survey (ISGS). Each of the data sets was an Arc Info coverage of the soil types within the county. A list of soils considered to be hydric was obtained from the soil conservationist in each county. Each soil coverage was converted into a shapefile using Arc View and each soil entry was recoded as a hydric or non-hydric soil.

**NASS**

Landcover data was acquired from the National Agriculture Statistics Service (NASS). The NASS data is an Arc Info coverage of the agriculture land cover map for 2001. It is produced from Landsat TM data and field checked for accuracy by NASS personnel. The coverage was resampled from a pixel size of 30 meters to 20 meters to match the SPOT imagery. It was also recoded down from sixteen classes to eight, including cropland, grassland, woods, other waste, clouds, urban/residential, water and wetlands.

**DEM/SAI**

A 30-meter digital elevation model (DEM) is available for the state of Illinois. There was an attempt to resample the spatial resolution of the DEM to match the 20-meter SPOT data, and produce a slope aspect index (SAI) from the data. It was found that when the spatial resolution of the DEM was resampled, from 30 to 20-meters, it created linear artifacts in the image, and therefore, it could not be used.

**Flood Zones**

In order to investigate the hydrology of the study area, an Arc Info coverage of flood zones was used. The flood zone coverage was produced by the Federal Emergency Management Agency (FEMA) to locate areas that are prone to flooding events. It was assumed that areas that are prone to flooding have poor drainage and therefore, are likely to support wetlands.

**Methodology and Discussion**

**Spatial Modeler**

In the course of this study, four different methods were used to try to find wetlands located in the study areas. In the first method, the spatial modeler in Imagine was used to build a model that would find wet pixels in the SPOT imagery located in areas of hydric soils (figure 1). To achieve this, the mid-infrared (MIR) band from each date was used to find the wet pixels in each study area. Previous research has shown that because water absorbs mid-infrared
wavelengths of radiation, a threshold can be set where, all pixels in the MIR image below the threshold are identified as “wet” (Johnston and Barson, 1993). In this case, any pixel with a value below one standard deviation from the mean was considered a wet pixel. To be sure that all water bodies were accounted for, anything mapped as open water on the USDA’s county soil survey was also pulled out of the image. It was assumed that any pixel considered to be open water in the soil survey, or wet, according to the mid-infrared band, and also located within a hydric soil polygon in the soil survey was a wetland. Since the boundaries of wetlands can shift and be difficult to map, a 100-meter buffer was added to every wetland produced from the model.

This model had a couple of strong points, such as being quick and easy to implement, and the use of ancillary data, in the form of the county soil survey, kept human interpretation error to a minimum. However, it was not affective in locating wetlands. The model did not find all the wet areas using the mid-infrared band that were known to be in the county. Therefore, it cannot be assumed that it finds wet areas that are not known. This could be due to the fact that the marsh areas were covered with vegetation, which reflects more radiation than water. Another drawback is that the model is unable to discern wetland type. It only coded each pixel in the image as being considered a wetland or not a wetland. Ultimately, pixels that are not considered to be wetlands by the model are discarded, leaving holes in the imagery, which can be difficult to interpret if a person is not familiar with the area. Before the 100-meter buffer is applied, little more than water bodies are pulled out of the image as wetlands. After the addition of the 100-meter buffer around each wetland area, many more wetlands are pulled out of the image, but this is likely due to the fact that many wetlands occur around open water bodies. Many more non-wetland areas are pulled out after the application of the buffer as well. To alleviate some of these inherent problems, both unsupervised and supervised classifications were ran on the resulting image. Unfortunately, neither classification process produced reasonable results. In the case of the unsupervised classification, each resulting class contained pixels of multiple land cover types. This made it difficult to decide what category to put each class into. In the case of the supervised classification, with the exception of areas of open water, it was difficult to find groups of continuous pixels of the same land cover type large enough to use as training classes. Another problem with the supervised classification is that many land cover categories can appear different in different areas of the image.

**Spatial Modeler with NASS**

We tried to build on the weaknesses of the first spatial model and incorporate the NASS data to produce a model that would do a better job of classifying the wetlands. This model looks at the pixels that were classified as wetlands in the previous model (figure 2). It assigns the pixel to one of the eight classes corresponding to the recoded NASS data, according to what the NASS coverage has at the location of that pixel.

This model did not work at all. The main drawbacks were related to the accuracy of the NASS data. Because the NASS data focuses on agricultural landcover, the non-agriculture areas have a very low accuracy (Luman, verbal communication). The 2001 NASS coverage for Lake County was completed using only one date of imagery, instead of the two dates, which are usually used. Therefore, the overall accuracy of the 2001 NASS data for Lake County is low (Luman, verbal communication). Because wetlands were not the focus of the NASS mapping initiative, they were not mapped as carefully as the agricultural areas. Within the NASS data, there is a class labeled “other waste” which apparently contains non-agricultural land that the
NASS personnel did not know how to classify. Therefore, it is difficult to tell what all this class encompasses. Lastly, in areas where extensive marshes are known to exist, the NASS coverage has much of the area classified as corn. One good aspect of the NASS data is that the most reliable categories in the coverage appear to be the wooded and water classes. Therefore, a group of pixels that were identified as wetlands in the first model, and located in a wooded area according to the NASS coverage, most likely is a floodplain forest.

**20-meter GAP**

The next method is similar to that used by the Gap Analysis Program for land cover mapping. All four bands of imagery for the spring, summer and fall dates are stacked into one 12-layer image. This image was run through an unsupervised classification using Imagine’s ISODATA method with an initializing option of four standard deviations, 255 maximum classes, 0.99 convergence threshold and 500 maximum iterations. This produced a signature file containing spectral signatures for each of the 255 resulting classes. This signature file was used to run a supervised classification on the 12-layer image using all the default settings in Imagine. Each class resulting from the supervised classification was assigned to a land cover category based on reference data such as aerial photos, NASS data and the original imagery. With all 255 classes labeled according to land cover, the image was recoded in order to group each occurrence of a certain landcover category into a single class. At this point this method deviates from the GAP method. In order to eliminate the “salt and pepper” look caused when tiny groups of pixels are classified wrongly and mixed in with larger groups of a different class, the “clump” function was run on the recoded image using the 8-neighbors option, to identify the number of consecutive pixels of the same land cover type in the image. Next the “eliminate” function was used. This function finds all clumps which contain fewer than a user specified number of pixels of the same land cover type, and recodes them to match the surrounding land cover. In this case, the eliminate function was set to eliminate clumps containing fewer than 9 consecutive pixels.

This method worked much better than either spatial model. The main strong point of this method is also the main drawback. This method requires more user input and interaction than the spatial models. Each group of pixels that results from the classification process must be studied by the investigator and placed into a category depending on the land cover. Often times a class resulting from the classification process has pixels that can be placed into more than one land cover category. This forces the user to closely examine the groups of pixels and decide which category corresponds to the bulk of the pixels in the groups. Because the focus of this research is wetlands, groups of pixels with wetlands and other types of land cover are usually classified as wetlands, no matter what other types of land cover are present.

The other drawback of this method is that none of the ancillary data was used. This method relies solely on the discretion of the classification algorithms within Erdas Imagine to determine which areas are alike enough within the twelve bands of imagery to be considered the same type of land cover. Then the user goes through with a minimum amount of extra imagery to place these classes into a few discrete categories. There is no information about soil type, hydrology or vegetation besides what might come across in the imagery.

**Knowledge Engineer**

The final and most promising method used to locate wetlands in this research is the Knowledge Engineer in Imagine. Using the Knowledge Engineer tool, it is possible to build a knowledge base, or decision tree. The basic concept behind the knowledge classifier is to input
as much data as is available for the study area and set up true or false statements in a decision tree to systematically cancel possible classifications until all the criteria are met by a pixel in each data set to place it into one class. The knowledge base used the image resulting from the clump/eliminate functions in the previous method, as well as coverages of hydric soils and floodzones for the each study site (figures 3a-3d). The knowledge base was set to classify pixels that had been categorized as shallow marsh, deep marsh, or bog in the classified image, which also fell into areas of hydric soil in the county soil survey into their respective wetland categories. Any pixels that were categorized as wetlands, but did not fall within areas of hydric soils were classified as miscellaneous grasslands. Pixels that were categorized as open canopy, closed canopy or floodplain forests in the classified image, which were also mapped as flood zones in the flood zones coverage were classified as floodplain forests. Land cover categories that do not rely on hydric soil or inclusion in flood zone areas, such as urban and agriculture are classified through the knowledge engineer based solely on the results of the unsupervised/supervised classification process.

Out of the four methods used in this research, the knowledge engineer proved to be the best. In addition to the twelve layers of satellite imagery, it also used two sets of ancillary data. The spatial modeler method of wetland mapping used only soil data in addition to the satellite imagery to locate wetlands areas. There are three criteria which define wetlands, hydric soil, wetland vegetation and hydrology. This method comes the closest to using all three of the criteria needed to locate wetlands. The spatial modeler also used hydric soils, however, it paid no attention to the hydrology of the area. While there was no use of any specific vegetation data in any method throughout this research, during the classification process leading up to the knowledge engineer, the user is forced to look closely at the ground cover present in an area of the satellite image to decide if it is agricultural, grassland, forest or wetland area. In this way, it is apparent that all three wetland criteria have been used to locate wetlands using the knowledge engineer.

The wetland map produced with this method was compared to a watershed in Lake County, which was field mapped by personnel from the Illinois Department of Transportation (figure 4). There are numerous errors of omission in the satellite-mapped coverage. One reason for this is that the satellite image has a resolution of 20-meters, and is subsequently ran through a filter which rejects all land cover areas less than 9 pixels, or 3600 square meters. The field-mapped data on the other hand has no minimum required area for inclusion. However, the areas classified as wetlands by both methods are commonly similar in size and shape. This shows that the knowledge engineer method is a good starting point to locate wetlands before going into the field to map them more precisely.

Resolution Merging

In addition to the 20-meter multi-spectral SPOT imagery, the Survey also acquired 10-meter panchromatic imagery corresponding to the same dates and locations. It had been planned to use the panchromatic band for each date to perform a resolution merging process on the 20-meter data that would result in a 10-meter multi-spectral image. When this was attempted, the resulting 10-meter resolution image was somewhat blurry. This was due to the fact that the pixels in the panchromatic and multi-spectral images were shifted slightly from one another. The shifting was enough to cause the resolution merging to produce a blurry image. An attempt was made to manually line the images up, run the resolution merge and then use the knowledge engineer to locate wetlands. However, this did not work. After the images were
properly aligned, and the resolution merge process ran, the resulting 10-meter image was crisp and clear. However, after the unsupervised and supervised classifications were ran, it became apparent that there was a problem. In the classified image, there were many areas that were forested grouped into the same class as water and urban pixels. Because this was not the case when the 20-meter data of the same area was classified, this must be due to the merging process. The panchromatic image only covers visible wavelengths of light. The near and mid-infrared values in the multi spectral image may have been too highly distorted in the merging process to be useful. In the ideal case, the mid-infrared band is the best for finding water, or wet areas because the longer the wavelength of light, the less energy is reflected. Therefore, areas that reflect little mid-infrared radiation are wet. Because the panchromatic image did not cover mid-infrared wavelengths, when this image was used in the resolution merging process, it skewed the values in the MIR image too far toward the reaction expect of radiation in the visible light spectrum, and therefore, no wetlands were located.

Data Needed

DEM/SAI

The spatial resolution of the 30-meter Illinois state DEM was resampled to match the 20-meter SPOT data. This DEM was used to produce a slope aspect index (SAI). In preliminary tests, the SAI was added in as the 13th layer to the imagery. However it was discovered that when the spatial resolution of the DEM was resampled, it created linear artifacts in the image. These artifacts affected the unsupervised classification of the imagery, and therefore, it could not be used. It is possible to resample the pixel size to a lower resolution and obtain a good, smooth image. Because the software has to split pixels and make new data to fill in the gaps, resampling pixel size down to a higher resolution image proves to be more problematic. To overcome the issues associated with resampling the spatial resolution of the DEM, a 10-meter DEM would be ideal. The two most commonly used image sources are SPOT and Landsat. SPOT imagery has a 20-meter resolution and Landsat has a 30-meter resolution. Therefore, 10-meter data could easily be converted for use with 20 or 30-meter resolution images.

Soil

The National Resource Conservation Service (NRCS), Soil Conservation Service (SCS) is the agency that publishes the soil surveys for every county in the United States. They are currently trying to make all the counties available in a digital format for use with GIS software. This is obviously a very large, time consuming job. We acquired the soil survey for St. Clair County from the NRCS. The soil surveys for both Lake and Peoria Counties came from the Illinois State Geological Survey (ISGS). The Peoria County soil data consisted of three quadrangles that had been updated and digitized by ISGS personnel from the survey published in 1992 by the NRCS. Obtaining the same type of data from two different sources caused a few problems. The most obvious is the lack of soil data for the bulk of Peoria County. Because only three quadrangles were available, the study site had to be reduced to only those three quadrangles. Another problem is that there were areas within the Peoria data that were not assigned soil names or codes, and there were new soil codes that had no explanation. It was difficult to tell which soils that were updated by the ISGS were considered hydric by the NRCS.

Obviously, the NRCS is updating and digitizing the countywide soil surveys as quickly as possible. However, in the case of Lake County, the only digital soil survey available was
completed in 1970. In such a large and quickly growing county, there have been many changes to the soils, from the addition of subdivisions and digging of new ponds to filling in of old ponds and reclamation of quarries in the past 30 years. These processes would all change the soil survey. More up to date and complete digital soil surveys would be very helpful in future research.

Conclusion

At the present time, only a small portion of the Lake County study area has been checked for accuracy, however, it appears to match fairly well with the field mapped data. There are errors of omission, likely due to the requirement of the data to be 3600 square meters before it is included as a wetland. There are also errors of commission in the wetlands mapped using the knowledge engineer. Because the field-mapping project took place before the satellite mapping, it is not possible to tell the cause of these errors without further fieldwork. The objective of this research, to develop a way to update the NWI for Illinois, has been met. Provided that the proper data exists in the form of updated, digitized soil surveys, it is possible to locate wetlands in satellite imagery using the knowledge Engineer in Erdas Imagine software. This method gives a rough estimate of the size and location of wetlands, field mapping is the only way to produce a very detailed map of wetland location.

Acknowledgements:

The SPOT data was purchased on a contract with the Illinois Department of Natural Resources, Office of Reality and Environmental Planning. Image processing was done at the Illinois Natural History Survey, Center for Wildlife Ecology. Additional remote sensing/GIS assistance was provided by Brooke Bahnsen, Frank Hutto, Liane Sulloway, and Diane Szafoni from the Illinois Natural History Survey and Don Luman from the Illinois State Geological Survey. Additional soil interpretation assistance was provided by Tom Kohl from the Illinois Natural History Survey.
References


Figure 1. Spatial Model in Erdas Imagine used to locate wetlands
Figure 2. Spatial Modeler using NASS data
Land cover map of Lake County, Illinois
Land cover map of Peoria County, Illinois Using 10-meter "resolution-merged" data