5-2015

Aquaculture and Coastal Habitats Mapping of Vietnam, Thailand and Cambodia: Summer Internship with Clark Labs

Xiao Xiao
Clark University

5-1-2015

Aquaculture and Coastal Habitats Mapping of Vietnam, Thailand and Cambodia: Summer Internship with Clark Labs

Xiao Xiao
Clark University

Aquaculture and Coastal Habitats Mapping of Vietnam, Thailand and Cambodia:

Summer Internship with Clark Labs

Xiao Xiao

May 2015

Submitted to the faculty of Clark University, Worcester,

Massachusetts, in partial fulfillment of the requirements for

the degree of Master of Science in the department of International Development,

Community, and Environment

And accepted on the recommendation of

Yelena Ogneva-Himmelberger, Academic Advisor

ABSTRACT

Aquaculture and Coastal Habitats Mapping of Vietnam, Thailand and Cambodia:

Summer Internship with Clark Labs of Clark University

XIAO XIAO

The report provides a detailed description of my internship experience with Clark Labs during the summer of 2014. As a research assistant, I was involved in the Aquaculture and Coastal Habitats Mapping project. This project was funded by the Gordon and Betty Moore Foundation. My responsibility was to create a baseline map of coastal habitats and the distribution of aquaculture (mostly shrimp) in Thailand, Vietnam, and Cambodia. The internship was valuable, and gave an opportunity to learn and grasp more skills. It will influence my career goal in the future. I will definitely recommend this internship to people who have interest in GIS and Remote Sensing, who enjoy facing challenges and want to use their power to develop the world

____

Yelena Ogneva-Himmelberger ,Ph.D.
Chief Instructor

ACADEMIC HISTORY

Name: Xiao Xiao **Date**: May 2015

Place of Birth: Chongqing, China **Date**: August 31th 1990

Baccalaureate School: Wuhan University

Baccalaureate Subject: B.S., Remote Sensing Science and Technology

DEDICATION

I would like to dedicate this to my family, friends and all those who helped me during my journey at Worcester.

ACKNOWLEDGEMENTS

I thank Professor J. Ronald Eastman for providing the great opportunity to work at Clark Labs. Your support and guidance helped me a lot to fulfill the goal of the internship. I also thank Professor Yelena Ogneva-Himmelberger, for her instructions of my internship paper. Finally, I thank to my colleagues at Clark Labs, too.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION

Remote Sensing (RS) and Geographic Information Science (GIS) captured my heart since the first day I learned about remotely sensed imagery. I chose Remote Sensing science as my major in my undergraduate school without hesitation. After 4 years of study, I not only grasped the concepts, principles and applications of RS/GIS, but also was inspired to go abroad to pursue further study. GISDE program at Clark University is specially designed for people who are interested in GIS/RS, and it gave me the opportunity to learn more theoretical and practical knowledge and technologies of GIS and RS. The skills I learned at Clark made me fully prepared to work at Clark Labs in the summer 2014.

As a research assistant in Clark Labs, I was involved in the Aquaculture and Coastal Habitats Mapping project, which was funded by the Gordon and Betty Moore Foundation. I was responsible for creating a baseline map of coastal habitats and the distribution of aquaculture (mostly shrimp) in Thailand, Vietnam, and Cambodia, and providing the transition potential for coastal habitats to pond aquaculture in the next step. The internship was meaningful and beneficial, and gave me a great opportunity to learn and grasp more practical skills of GIS and RS. For example, I learned different classification methods and their corresponding advantages and disadvantages. Also, after studying the aquaculture habitat in these three countries, I learned the importance of aquaculture habitats and their influences on local economies and environments. Furthermore, this internship helped to improve my communication skill and bolster my qualifications for future career.

This paper consists of 5 chapters each with a specific focus. Chapter 2 describes the Clark Labs. Chapter 3 discusses my job responsibilities and my accomplishments. Chapter 4 provides general assessment of the internship.

CHAPTER 2: DESCRIPTION OF ORGANIZATION

2.1 History

Clark Labs is a research organization founded in 1987 at Clark University which is world-renown for its Geography Department. It is engaged in the development of geospatial software named “IDRISI”. A variety of modules within IDRISI support GIS and Remote Sensing related analysis. While in 2015 probably, it will release the new version with a totally different name “Terrset”, which represents its strong geospatial functionality in GIS and RS.

Clark Labs has many partners all over the world, including well-known organizations and companies, such as The Gordon and Betty Moore Foundation, Google.org, USDA, the United Nations and Conservation International. Clark Labs provides strong academic and practical technologies to solve challenging geospatial problems.

2.2 Products and Achievements

The flagship product of Clark Labs is the IDRISI GIS and Image Processing software, which has been widely used in over 180 countries by a wide range of professionals since 1987. IDRISI is famous for its spatial analysis tool packages. It includes nearly 300 modules to perform geospatial analysis. It provides an easy, convenient and trustable way for researchers to perform image processing and analysis. GIS/RS analysts, environmental scientists and urban planners can get much help and benefit from it. Moreover, the modules in IDRISI are best known for applications to decision making, classification technologies, time series analysis, land change analysis and forecasting, earth trend analysis and dynamic modeling.

2.3 Past and Recent Research

Clark Labs engages in a variety of research covering a wide range of environmental, economic development, resource, human issues and others. It is dedicated to solving the most challenging problems in the real-world. IDRISI is designed to support the analysis. The research that Clark Labs has done includes Climate and Ecosystem Dynamics, Land Change Analysis, Reducing Emissions from Deforestation and Forest Degradation (REDD), Machine Learning and Neural Networks, Multi-Criteria / MultiObjective Decision Making, Dynamic Modeling and others (Clark Labs 2014).

Clark Labs has played a pioneering role in uncertainty management, which is still a challenge in the world. The uncertainty management includes the soft reclassification procedure that allows one to map the probability of a location being above or below a threshold, and the generation of normal and rectilinear distributions for uncertainty analysis such as Monte Carlo methods (Clark Labs, 2014). Monte Carlo methods are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results; typically one runs simulations many times over in order to obtain the distribution of an unknown probabilistic entity (Hazewinkel, Michiel, ed. 2001). IDRISI is the only system that has implemented dynamic modeling using a graphical interface (Clark Labs, 2014).

At present, Clark Labs is working with Esri, implementing its Land Change Modeler Extension to the Esri Cloud Platform. Moreover, Clark Labs is working with the Gordon and Betty Moore Foundation to research coastal habitats at risk for shrimp farm conversion in Southeast Asia (Clark Labs, 2014). All the projects contribute to the solution of real world problems.

2.4 Funders and Organization Structure

Dr. J. Ronald Eastman is a professor of Geography in Graduate School of Geography who founded the Clark Labs in 1987 at Clark University, He pursued his Ph.D. at Boston University and joined Clark in 1981 (Clark Website, 2014). Clark Labs has a staff of approximately 30 employees, who come from a variety of countries and backgrounds. The diverse culture leads to a harmonious, complementary atmosphere, in which it can stimulate staff's creativity, imagination, cooperation and efficiency.

There are 3 core groups working in Clark Labs. One is the programming group, which is responsible for algorithm designing and developing new modules in Delphi computer language to support geospatial analysis or other research. One is the testing group, which is in charge of algorithm testing of the new modules, as well as technical support. The last is the analyst group, which works for various analysis projects using IDRISI software. Clark Labs not only provides full-time jobs, but also gives opportunities for both undergraduate and graduate students to gain work experience and learn more practical knowledge.

CHAPTER 3 INTERNSHIP RESPONSIBILITY

As an intern at Clark Labs, I was involved in the Aquaculture and Coastal Habitats Mapping Project. This project was funded by the Gordon and Betty Moore Foundation. The goal of this project was to create a baseline map of coastal habitats and the distribution of aquaculture (mostly shrimp) in Thailand, Vietnam, and Cambodia, and to provide the transition potential for coastal habitats to pond aquaculture. This summer I worked on creating the baseline map. 34 Landsat-8 scenes were used for classification of coastal marine and terrestrial land cover. "Landsat-8 provides us with a nice combination of cost effectiveness, synoptic coverage, spectral resolution and the ability to pan sharpen the critical bands needed for water penetration," said Dr. Ron Eastman. Figure 1 shows the study area in Thailand, Vietnam and Cambodia.

I worked directly under the guidance of Dr. Ron Eastman. In summer, there were three graduate students working for the project including me. In the first week of my internship I started to get familiar with the land cover classes I would be mapping using 15 m pan-sharpened Landsat scenes and high resolution imagery in Google Earth™. Figure 2 displays a list of the land cover classes mapped for this project. There are 7 categories in total: Aquaculture, Marine Habitats, Wetland Habitats, Dryland Habitats, Barren Habitats, Anthropogenic Habitats, and Missing Data. For each category, we have several sub-classes, which lead to a more detailed level and a total of 30 classes.

After studying all the classes I learned the whole process of classification, including data downloading, data pre-processing, and classification methods. Data pre-processing included atmospheric correction, water depth correction and upscaling all the bands to 15 m resolution. Data pre-processing provides us with a solid foundation to perform classification. For classification we used the supervised classification methods Mahalanobis Distance and K-Nearest Neighbors (KNN). Accuracy assessment will be performed in the winter of 2014with the aid of Google Earth™. Ground truthing will be conducted in the spring of 2015 to improve the final accuracy.

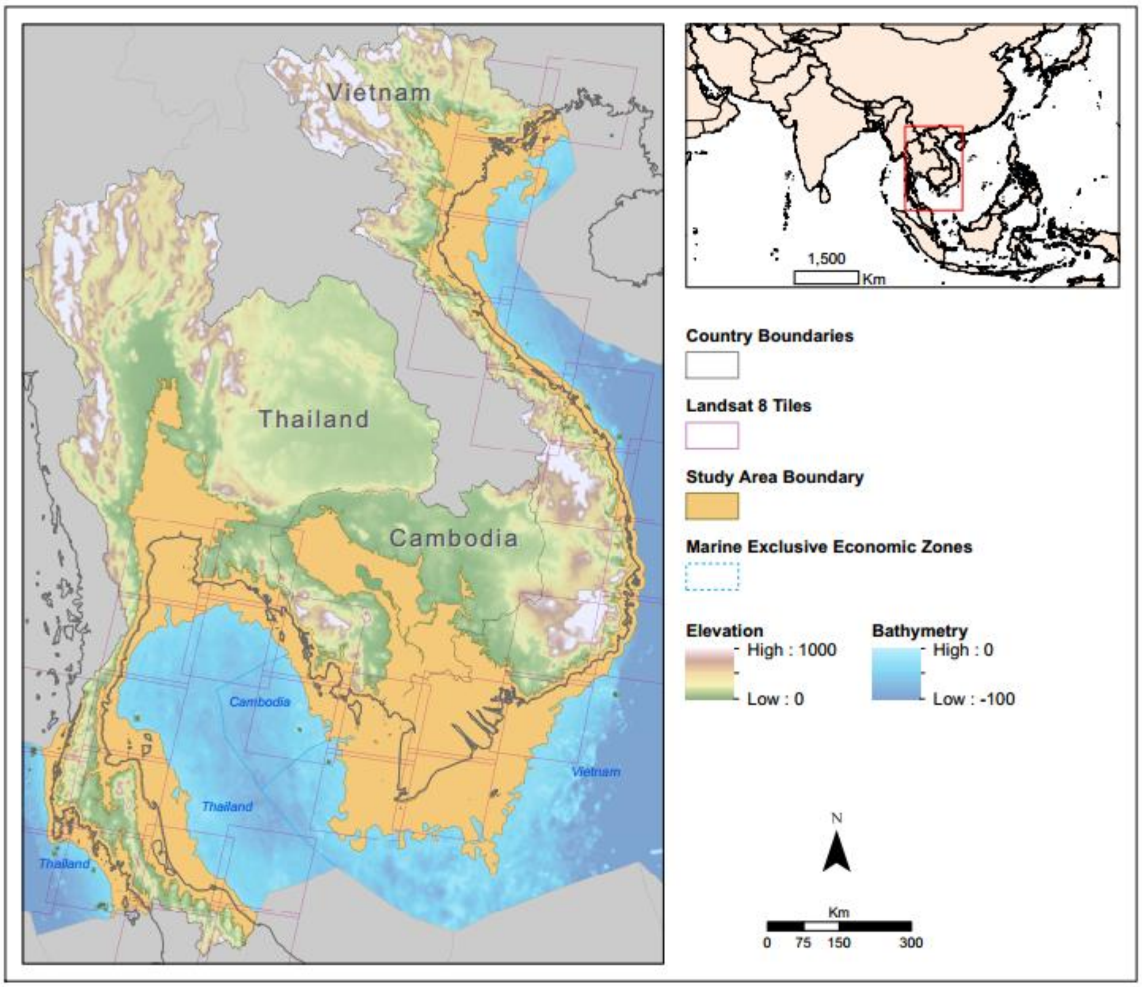

Figure 1. The study area in Thailand, Vietnam and Cambodia for coastal habitats project.

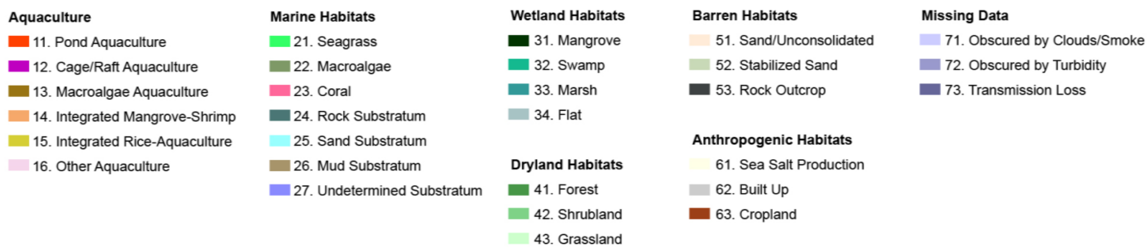

Figure 2. Legend of classes in the aquaculture and coastal habitats mapping project.

3.1 Remote Sensing Data: Landsat-8

On February 11th 2013, the National Aeronautics and Space Administration (NASA) launched the Landsat-8 satellite, which is the eighth satellite in the Landsat program (USGS-Landsat-8). Landsat-8 carries two instruments: 1) the Operational Land Imager (OLI) sensor includes upgraded legacy bands (eight 30 m multispectral bands and one 15 m panchromatic band); along with three new bands (compared to Landsat-7), a deep blue band for coastal/aerosol studies, a shortwave infrared band for cirrus detection, and a Quality Assessment band that displays pixels affected by instrument noise or cloud-contamination; 2) the 100 m Thermal Infrared Sensor (TIRS) sensor provides two thermal bands (USGS-Landsat-8). Figure 3 shows the comparison of Landsat-8's OLI and TIRS bands and Landsat-7's ETM+ and TIRS bands (NASA-Landsat-8). Landsat- 8 images were downloaded from the United States Geological Survey (USGS) Earth Explorer website.

The advantages of Landsat data are that they are free, and captured using high quality sensors with moderate spatial and radiometric resolution. Radiometric resolution determines how finely a system can represent or distinguish differences of intensity, which is essential for the marine classification. The visible light bands (1-4) can be pan-sharpened and up-scaled from 30 m to 15 m using the panchromatic band 8. For this project pan-sharpening improved our ability to visually interpret the images and place training sites to be used in classification.

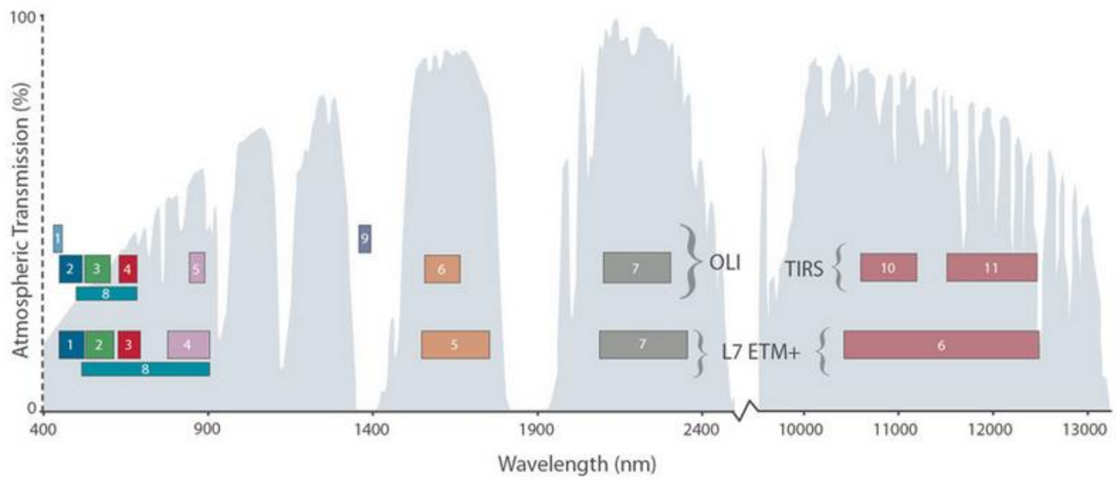

Figure 3. Comparison of Landsat 8’s OLI bands and Landsat 7’s ETM+ bands

3.2 Image Pre-processing

Image pre-processing consists of data operations which typically correct distorted or degraded images before proceeding with analysis of the image. In this project image pre-processing involved atmospheric correction, pan-sharpening the visible bands, up-scaling the infrared bands and water depth correction.

3.2.1 Atmospheric Correction

The atmosphere can affect the nature of remote sensing images in a number of ways. At the molecular level, atmospheric gases cause Rayleigh scattering which progressively reduces shorter wavelengths arriving to the ground. Atmospheric components such as oxygen, carbon dioxide, ozone and water vapor cause absorption of energy at selected wavelengths. Aerosol particulates will lead to Mie scattering, caused when the size of an aerosol particle is similar to some incident wavelength creating atmospheric haze. Atmospheric correction aims to correct remotely sensed images for these distortions. To perform atmospheric correction, Landsat-8 bands should be converted to top of atmosphere (TOA) reflectance from quantized and calibrated DN according to the following equation:

$$\mathrm{TOA}' = \mathrm{M\rho} * \mathrm{DN} + \mathrm{A\rho} \qquad \text{Eq.1}$$

M_ρ is the band-specific multiplicative rescaling factor from the metadata. A_ρ is the band-specific additive rescaling factor from the metadata. Sun angle will affect the TOA reflectance, therefore it should be corrected according to the following equation:

$$\mathrm{TOA} = \frac{\mathrm{TOA}'}{sin(\theta\mathrm{SE})} \qquad \text{Eq.2}$$

θ_{SE} is the local sun elevation angle. It is provided in the metadata in degrees.

After calculating TOA reflectance for each band, the Dark Object Subtraction (DOS) method was applied to remove atmospheric scattering. DOS looks for the darkest pixel value in each band and assumes it has no reflectance, thereby any value greater than zero must be due to atmospheric scattering. The value above zero, known as the digital number (DN) haze value is subtracted from all other pixels, thus removing the atmospheric scattering. DOS is a simple yet effective method of atmospheric correction.

3.2.2 Pan-sharpen

The spatial resolution of a Landsat-8 true color composite (bands 2, 3, 4 as blue, green, and red) image is 30 m, which can make it difficult to visually distinguish between some coastal habitats relevant to this project, such as seagrass and coral. The panchromatic band 8 was used to upscale the visible bands (bands 1 to 4) through pan-sharpening. Pan-sharpening is a process of merging high-resolution panchromatic and lower resolution multispectral data to create higher-resolution color images (Brower, 2000). Pan-sharpening essentially increases the resolution of the visible light information contained within a scene.

A modified IHS (Intensity-Hue-Saturation) technique was performed to undertake pan-sharpening. All multispectral bands were up-scaled to 15 m first. Multiple regression was used to establish the relationship between the band 8 and the reflectance of the visible bands. A predicted Intensity (I') image is be obtained from the multi-regression model and used create pan-sharpened versions R', G', B' following equations:

$$R' = R + Pan - I' \quad \text{Eq.3}$$

$$G' = G + Pan - I' \quad \text{Eq.4}$$

$$B' = B + Pan - I' \quad \text{Eq.5}$$

Figure 4 displays true color composite images before pan-sharpening and after pan-sharpening. There is an improvement in the visual interpretability after pan-sharpening.

The pan-sharpen method only works for bands 1 to 4, upscaling the infrared bands (bands 5 to 7) to 15 m is required for consistency of classification and presentation. Bilinear resampling method was taken to accomplish this process, which minimizes the degree to which values are interpolated.

Figure 4.True color composite image created using visible bands 2-4 before pan-sharpening (left) and after pan-sharpening (right)

3.2.3 Water Depth Correction

For water features, variance of water depth will lead to inconsistent reflected electromagnetic waves as water causes energy to attenuate with greater depth. In the project, band ratios were created to correct depth variance of reflectance. Band ratios are depth invariant because the light attenuation coefficients of the two bands involved are identical. After some experiments, we have therefore found that ratios of Landsat 8 bands 2/1, 3/2 and 4/3 work fairly well. The 2/1 ratio provides the most information while 4/3 is the least. However, the use of three ratios increases spectral resolution substantially and allows for the production of color composites. However, the relationship between the bands is non-linear; therefore the natural logarithm was taken for the results of each ratio. These log ratio bands were used exclusively for classification of marine habitats.

3.2.4 Water Mask Creation

This project is mapping terrestrial and marine habitats, as such land and ocean surfaces need to be separated. In order to separate land and ocean I developed an effective method of creating a water mask. Principle components analysis (PCA) was performed yielding the first principle component, which contains the most variance in the reflectance from bands 1 to 7. Furthermore, I explored the difference of pixel values in the PCA image between land and water carefully using the Profile module in Idrisi to determine one specific pixel value as a threshold to separate them. However, pond aquaculture was included into the water mask due to its similar pixel values to water. To separate the pond aquaculture and water, the Group module was applied in Idrisi. Group determines contiguous groupings of identically valued integer cells in an image, and cells belonging to the same contiguous grouping are given a unique identifier, numbered consecutively in the order found (Idrisi Selva Help Manual, 2014). With the Group module, we can obtain a series of numbers which represent different groups and the water group could be easily separated from pond aquaculture and land groups by performing reclassification using the Reclass module. However, there is still a portion of pond aquaculture that connects directly to the ocean. These remaining ponds were manually reclassified using the Digitize and Rastervector modules. The polygons that only contain ponds were carefully drawn with the Digitize module. These polygons were converted from vector format to raster format using Rastervector module. This new raster is used to subtract ponds from the water mask using the Overlay module.

3.3 Coastal Zone Definition

The focus of this project is to map the coastal regions of Thailand, Vietnam, and Cambodia. Therefore a definition and mask of this coastal region was required. The coastal zone in this project is interpreted as the environment formed by the coexistence of two margins: coastal land defined as the terrestrial edge of continents, and coastal waters defined as the littoral section of the adjacent ocean. The terrestrial portion of the coastal zone is defined by the area extending 10 km landward from the coastline or the area with less than 5 m elevation (based on SRTM data) within 60 km from the coast. The 5 m elevation threshold is based on comparison of salt intrusion maps and tidal charts with elevation data around Bangkok, Thailand. The maximum elevation of salt intrusion surrounding Bangkok was 4.03 m (Arlai, Phatcharasak. 2007). As Bangkok has the max tide for all locations within the study area 5 m elevation is enough to include most coastal habitats that could be used for shrimp aquaculture. The maximum salt intrusion during the dry season for estuaries in Northern Vietnam was 40km (Cat, Vu Minh and Bui Du Duong. 2006), and the maximum salt intrusion during the dry season months for all estuaries measured in the Mekong were below 60km (Nguyen, Anh Duc. 2008). Therefore 60 km from the coast was used as previous studies of salt intrusion measured distance from the coast along the river whereas our 60 km is straight from the coastline The marine part of coastal zone is defined as a zone extending 10 km offshore or the area with less than 30 meters depth water (based on GEBCO bathymetry data). Only islands that are less than 200 km from the mainland coast were mapped. Figure 5 shows the coastal zone which is part of our study area.

Figure 5. The coastal zone of part of our study area

3.4 Classification

After pre-processing classification of the Landsat scenes can proceed. In this project, we used supervised classification methods, which is the development of a classification by manually selecting representative training areas using the aid of reference sources, and algorithms to characterize the statistical patterns of the representative areas and thereby classify the image (Dietterich, Thomas G, 1998).

The most important part of supervised classification is creating training sites which are used to derive a representative sample of the spectral signatures of each class. The quality of training data can significantly influence the performance of the classification accuracy (Chen.D and Douglas Stow, 2002). The training sites should be polygons with similar pixels. When selecting training sites, a range of factors should be considered, such as the number of pixels used in training and making the training polygons equally distributed in the images for each class. Sometimes, one class contains different spectral signatures. For example, cropland may have a block of green pixels and brown pixels. In this case, cropland should be divided into two classes and two different sets of training sites are needed.

After digitizing an adequate number of training sites, the Makesig module in Idrisi was used to extract the spectral signatures from each training site. These spectral signatures can be used for classification of remotely-sensed imagery using any number of data classifiers. Literature review along with case studies we performed lead us to use the K-nearest Neighbor (KNN) and Mahalanobis Distance classifiers in classification of the Landsat8 images in Thailand, Vietnam and Cambodia.

In KNN classification, an object is classified by a majority vote of its training data neighbors, with the object being assigned to the class most common among its k nearest neighbors within feature space (Weinberger, 2005). KNN classification is a fairly simple method but can support both hard and soft classification. For a hard classification, a pixel is assigned to the class which dominates the k-nearest neighbors. When soft classification is performed, each category's proportion among the k-nearest neighbors is assigned to the pixel as a degree of membership to that category. In our project, KNN is used primarily as a soft classifier when the class contains a non-linear signature, such as built up, cropland and pond aquaculture.

The Mahalanobis Distance classifier is a soft classifier, which indicates the degree of membership of that pixel to training data. The output of Mahalanobis classifier is a set of images that expresses typicality of pixel's reflectance relative to those described by the training sites (Idrisi Selva Help Manual, 2014). More specifically, typicality expresses the probability of finding a pixel with a Mahalanobis distance greater than or equal to that of the pixel being evaluated (Idrisi Selva Help Manual, 2014). In our project, Mahalanobis Distance classifier is used when the class has a uniform signature, such as mangrove, sand, and others.

After running the KNN and Mahalanobis distance soft classifiers for various classes, a set of images (one per class) were obtained. Each pixel in the images contains the probability relative to the class described by the training sites with the range 0 to 1. Each image is examined and a threshold was empirically set to remove the pixels that were unlikely to belong to the class using the Reclass or Harden modules. The Harden module can be used to reclassify multiple images at different thresholds in a single step. The individual classes were combined into a single classified land and marine cover map. Figure 6 shows the completed classification images I made over the course of my summer internship. These images have not yet been masked to the coastal zone.

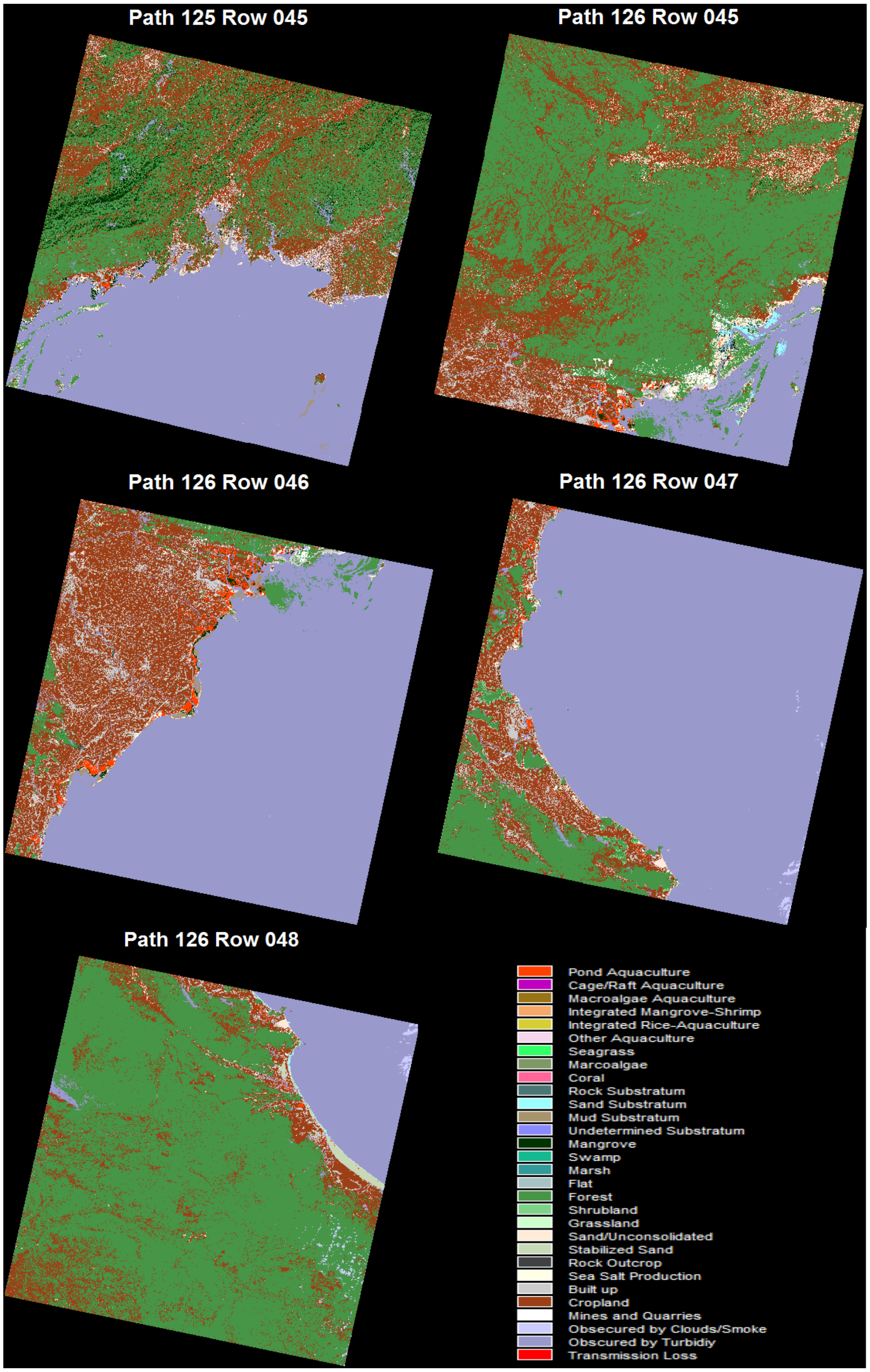

Figure 6. Completed classification images with their paths and rows

CHAPTER 4: INTERNSHIP ASSESSMENT

4.1 Clark University

Clark University is well-known for its Geography programs. The GISDE program in Geography is specially designed for people who are interested in GIS. It attracts plenty of students every year. The knowledge and skills learned from Clark University in GISDE Program made me fully prepared to work at Clark Labs. The courses at Clark include both lectures and laboratories, and provided me with important GIS concepts and practical exercises. The skills learned in Intro to Remote Sensing and Advanced Raster GIS courses in particular helped me understand my work quickly and develop new ideas.

4.2 Skills Learned at Clark Labs

The internship at Clark Labs solidified the skills that I obtained at Clark University and provided me a chance to learn new skills. My responsibilities at Clark Labs included performing land cover classification and related analysis in Cambodia, Thailand and Vietnam. Therefore, I learned a variety of classification methods and familiarized myself with their advantages and disadvantages. Also, through studying the aquaculture habitat in these three countries, I learned the importance of aquaculture habitats and their influences on local economies and environments. Furthermore, good communication is really essential in the work place. This internship gave me an opportunity to improve my communication skill and bolster my qualifications for future career.

4.3 Recommendation

My experiences, as an intern in Clark Labs, helped me to consolidate what I learned at school and to develop my future career goals. I was excited to see and get familiar with a more comprehensive GIS and Remote Sensing (RS) world. I would like to work as a GIS/RS analyst, using my skills learned at Clark University and Clark Labs, to solve related problems in the future. I will recommend this great internship to those who have interest in GIS and Remote Sensing, who enjoy facing challenges and want to use their power to develop the world.

CHAPTER 5: CONCLUSION

Clark University, which is famous for its geographical study, equipped me with solid theoretical and practical foundation of GIS and Remote Sensing. During the summer of 2014, I found an internship at the Clark Labs, as a research assistant, which was a meaningful experience. The internship provided me an opportunity to combine the academic world and real world. The project was amazing and valuable. Not only I could grasp more GIS and RS skills, but also learned more about costal habitat in Thailand, Cambodia, and Vietnam. The people in Clark Labs were professional and friendly. The internship prepared me well to meet future challenges, and it will definitely influence my career. I will recommend this internship to those who have strong interest in GIS and RS, and who want to challenge themselves.

BIBLIOGRAPHY

Brower, B. V. and C. A. Laben (2000). Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening, Google Patents.

Chen, D. and D. Stow (2002). "The effect of training strategies on supervised classification at different spatial resolutions." Photogrammetric Engineering and Remote Sensing 68(11): 1155-1162.

Dietterich, T. G. (1998). "Approximate statistical tests for comparing supervised classification learning algorithms." Neural computation 10(7): 1895-1923.

Hazewinkel, M. (2001). Encyclopaedia of mathematics, supplement III, Springer.

Labs, C. (2014). "http://www.clarklabs.org/."

NASA (2014). "http://www.nasa.gov/."

USGS (2014). "http://www.usgs.gov/."

Weinberger, K. Q., et al. (2005). Distance metric learning for large margin nearest neighbor classification. Advances in neural information processing systems.