DELINEATING WETLANDS USING GEOGRAPHIC INFORMATION SYSTEM AND REMOTE SENSING TECHNOLOGIES

A Thesis

by

JULIE VILLENEUVE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

December 2005

Major Subject: Biological and Agricultural Engineering

DELINEATING WETLANDS USING GEOGRAPHIC INFORMATION SYSTEM
AND REMOTE SENSING TECHNOLOGIES

A Thesis

by

JULIE VILLENEUVE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Ann Kenimer
Committee Members,	Raghavan Srinivasan
	Ronald Lacey
Head of Department,	Gerald Riskowski

December 2005

Major Subject: Biological and Agricultural Engineering

ABSTRACT

Delineating Wetlands Using Geographic Information System and Remote Sensing Technologies. (December 2005)

Julie Villeneuve, Dipl., Ecole Polytechnique Feminine

Chair of Advisory Committee: Dr. Ann Kenimer

During the last century wetlands have considerably decreased. The principal cause is urbanization, especially in large urban regions such as the Houston area. In order to protect the remaining wetlands, they have to be monitored carefully. However monitoring wetland is a difficult and time-demanding task because it has to be done repetitively on large areas to be effective. This study was conducted to determine if Geographical Information System (GIS) and remote sensing technologies would allow accurate monitoring of wetland as a less time-consuming method. With this idea, a suitability model was developed to delineate wetlands in the Houston area. This model combined GIS and remote sensing technologies. The data used for this study were as high spatial resolution as possible and were generally easy to obtain. This suitability model consisted of four submodels: hydrology, soil, vegetation and multi-attribute. Each submodel generated a Wetland Suitability Index (WSI). Those WSI were summed to obtain a general WSI. The suitability model was calibrated using half of the study area. During calibration, the general model was evaluated as well as each individual index. Generally, the model showed a lack of sensitivity to changes. However, the model was slightly modified to improve the delineation of upland wetlands by increasing the weight of the soil submodel. This model was validated using the second half of the study area. The validation results improved a bit compared to the calibration results; however they remained weak. It was demonstrated that the model does not favor riverine wetlands over upland wetlands, nor large size wetlands.

The model ground truth data were evaluated and were sufficiently proven to be up to date. Those results indicated that the weakness of the model must come from inaccuracy in the input data. Therefore, the study showed that while existing computing capacity supports remote delineation, spatial accuracy is still insufficient to perform correct wetland delineation using remote sensing and GIS technologies.

To my husband Olivier, whom I love very much

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr Ann Kenimer, chair of my graduate advisory committee, for her personal and professional support. I also would like to thank Dr. Raghavan Srinivasan for his invaluable help for this project. He provided me with adequate software, data, computers ... and always took the time to answer my questions. I am grateful to Dr Ron Lacey who assisted me with his expertise in remote sensing and GIS. My appreciation also goes to Dr Sorin Popescu who always made himself available to assist me when I was struggling with my remote sensing data.

I would like to thank my family, who has always been there for me. Finally, I would like to thank Olivier whom I cherish overall. Nothing would have been possible without you.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

During the last century wetlands have considerably decreased. In Texas the loss of wetlands between the 1780's and the 1980's was estimated to be 52% [1]. The principal cause is urbanization, especially in large urban regions such as the Houston area.

Wetlands support biodiversity of both plants and animals. They shelter endangered species and play a major role in water resource management by filtering out chemicals and fertilizers [2]. They also act as buffers during storms and floods, preventing downstream damage. Consequently, many conservation programs such as the Texas Prairie Wetlands Project developed by the Texas Parks and Wildlife Department have been created to better characterize, monitor, and preserve these exceptional features.

The U.S. Army Corps of Engineers (USACE) and the Environmental Protection Agency (EPA) jointly define wetlands as: "Those areas that are inundated or saturated by surface or ground water at a frequency and duration sufficient to support, and that under normal circumstances do support, a prevalence of vegetation typically adapted for life in saturated soil conditions. Wetlands generally include swamps, marshes, bogs, and similar areas" [3]. This definition is founded on three major wetland characteristics: predominance of hydric soils, presence of hydrophytic vegetation and specific hydrologic conditions (saturation with water during part of the growing season) [4].

Those characteristics are defined as follow:

- **Hydric Soils**: Nationally, there are over 3,000 types of soil that can occur in

---

This thesis follows the style of *IEEE Transactions on Automatic Control.*

a wetland. Those soils are called hydric soils and are defined by the Natural Resources Conservation Service (NRCS) [5] as “soils that formed under conditions of saturation, flooding, or ponding long enough during the growing season to develop anaerobic conditions in the upper part of the soil”. When a soil is flooded, water stops atmospheric oxygen from penetrating the soil. Soil microbes are, therefore, required to use other sources of oxygen such as reduction of iron compounds. This reduction creates gray features that are characteristics of hydric soils [6]. Hydric soils lists are provided by the NRCS [5]. These lists were established using the criteria that were developed by the National Technical Committee for Hydric Soils (NTCHS).

- **Hydrophytic Vegetation**: Plants that grow in wetland conditions are known as hydrophytic vegetation and defined by the Environmental Laboratory [3] as “the total sum of macrophytic plant life that occurs in areas where the frequency and duration of inundation or soil saturation produce permanently or periodically saturated soils of sufficient duration to exert a controlling influence on the plant species present”. Because of adaptations that can be morphological, physiological or reproductive these plants survive when submerged by water. Hydrophytic plant lists are provided by the National Wetlands Inventory (NWI) [2] nationally or regionally.

- **Hydrology**: The wetland delineation manual [3] identifies the term “wetland hydrology” as including “all hydrologic characteristics of areas that are periodically inundated or have soils saturated to the surface at some time during the growing season.” More specifically, states of inundation and/or saturation should occur for more than 12.5 percent of the growing season. The growing season is defined as: “the portion of the year when soil temperatures at

19.7in (50cm) below surface are higher than biological zero (5^oC). However, it generally is to be determined by the number of killing frost-free days" [3].

Water saturation provokes soil reduction and plant adaptation. Hence, these three wetland characteristics interact with each other. Therefore, to accurately delineate wetlands all three characteristics have to be considered, as described in the USACE Wetland Delineation Manual, Technical Report [3]. The USACE Wetland Delineation Manual, Technical Report is the main reference used in wetland delineation. It was originally created to provide guidelines and methods to determine whether an area is a wetland for purposes of Section 404 of the Clean Water Act (formerly known as the Federal Water Pollution Control Act). This act was ratified by the US Congress in response to US waters degradation and seeks to maintain and restore them. Section 404 of this act establishes a program to regulate the discharge of dredged and fill material into waters, including wetlands. The wetland delineation manual does not establish wetland classification.

Two other publications define major classifications for wetlands: Land Use And Land Cover Classification System For Use With Remote Sensor Data [7] and Classification of Wetlands and Deepwater Habitats of the United States [4]. Anderson's classification was created to generally classify any type of land-use or land-cover in the United States. The objective was to develop a homogeneous classification for any location in the US while still keeping flexibility. This system uses a four-level code to describe land-use and cover. The first two levels of the classification are defined and the last two levels are free for any specialization. Wetlands are represented in level one by the index 6. In level two, they are divided into forested wetlands (index 61) and nonforested wetlands (index 62). Forested wetlands are defined as "wetlands dominated by woody vegetation". Nonforested wetlands are defined as “wetlands dominated by wetland herbaceous vegetation or nonvegetated” [7]. This classification is the one used by the USGS to establish the National Land Cover Data (NLCD) derived from images acquired by Landsat sensor. The NWI classification was created by wetland ecologists along with private individuals and organizations and local, state, and federal agencies. It is a classification specialized for wetlands and deep water habitats. It was adopted by the NWI in 1977. It is divided into Systems, Subsystems, Classes, Subclasses, and Dominance Types. The five systems are: Marine, Estuarine, Riverine, Lacustrine and Palustrine [4]. This classification is used by the NWI to establish wetland maps.

The NWI maps are commonly-used references for identification and location of jurisdictional wetlands. Jurisdictional wetlands are identified using field indicators and procedures determined by federal or local authorities such as the USACE [8]. In Texas, the NWI maps were developed from 1992, 1:65,000 scale photography and thus represent wetlands status in 1992.

Monitoring and maintaining wetlands requires constant attention. The NWI maps present wetlands status in 1992, but the situation has changed since. It is a difficult task to monitor changes in wetland areas over time because it has to be done repetitively on large areas to be effective. This implies processing a large amount of geographic data which requires long processing times. Handling large amounts of geographic data is the main function of Geographic Information System (GIS) and remote sensing technologies. For that reason, they appear to be appropriate tools to monitor wetlands.

The procedure applied to establish the NWI maps in 1992 used high-altitude aerial photography as a primary data source rather than satellite imagery (Landsat). The main arguments for this procedure were the lack of spatial accuracy at that time and the prohibitively high cost of computers powerful enough for data processing and analysis [9]. However, this is no longer a problem since computers become more efficient day by day. In the last few decades the number of GIS and remote sensing users has increased considerably, resulting in data cost reduction and better data accuracy.

A. Background

Delineating and classifying wetlands is a complicated task because wetlands are widely spread across the landscape and sometimes difficult to access. Aerial photographs were initially used for this task [9], [10], [11]. They are easy to obtain, the spatial and temporal resolutions can be chosen, and the prices are reasonable [9]. Their manual classification is very efficient because human beings can identify and interpret context, edges, texture, and tonal variation in color [12]. An automated classification cannot take advantage of all these characteristics. On the other hand, manual classification is extremely time consuming and requires a well-trained operator [13], [14]. In addition, employing multiple operators might introduce undesirable heterogeneity in results.

Automated classification of wetlands using satellite data may be a viable alternative to shorten processing time. When the NWI were first developed, satellite data were too expensive, difficult to obtain, required large data storage resources and unusually powerful computers [9]. However, concurrent with the development of remote sensing, computers and data storage capabilities improved and can now handle such data. In addition, the number of satellites has increased and their technology has improved, leading to cheaper prices and better data accessibility. Some of the latest sensors are very promising. Hyperspectral sensors (e.g. AVIRIS) collect tens to hundreds of narrow spectral bands nearly simultaneously. Such information could enable identification of individual wetland plant species and changes between plants of the same specie, for example one subject to stress [15], [16]. Radar sensors (e.g. ENVISAT, ERS, RADARSAT and JERS 1) are active sensors and have major advantages: they can acquire images both day and night, they function through clouds, mist, fog or smoke, and they measure the physical characteristics and geometry of the objects observed [17], [18], [19]. Sensors with very fine spatial resolution such as IKONOS (one meter for panchromatic and four meter for multispectral) may also prove useful for wetland delineation. However these latest systems are still emerging and, therefore, their data are still very expensive.

The Landsat program has been in place for several years and its data have been used with success in many studies [14], [11], [20], [21], [9], [22], [23]. Landsat data are inexpensive or free in the United States, are not computer-demanding, and have fairly respectable spatial accuracy (30m pixel). The Thematic Mapper (TM) bands relate water penetration, discriminating vegetation type and vigor, plant and soil moisture measurements [12]. Their classification has been shown adequate for wetland vegetation detection [21], [11]. The TM classification are varied and include unsupervised [11], [14], supervised [22], hybrid [21] or tasseled cap [21], [20].

Wetlands have spatial characteristics such as low slopes or drainage proximity that can be efficiently managed using GIS. Therefore, combining information derived from Landsat classification with GIS data should improve remote wetland delineation. Using soil data, Anger [14] indeed improved the accuracy of an unsupervised TM classification by eight percent. To delineate forest wetlands in Maine, Sader et al. [21] improved an unsupervised TM classification by using NWI maps, soil data, Digital Elevation Model (DEM) and drainage proximity data. However, by using NWI maps as input data, the model no longer detected forested wetlands in the study area, but it did determine which of the wetlands defined by the NWI maps were forested wetlands. This GIS rule-based model was developed with ERDAS GISMO. It consisted of four layers: one for each GIS data. Each layer had a weight assigned arbitrarily: ten for the NWI maps, eight for the soils, six for slope less than five percent, and four for locations within a 90m buffer around the water bodies. The TM classification was given a weight of 100. These weights favored the vegetation factor, even though hydric soils and hydrology are equally important factors in wetland delineation [21]. The hydrology factor was represented by a 90m buffer around the water bodies. It might have strengthened the model to consider flow accumulation in addition to drainage proximity.

In 2000, Earley [24] developed a predictive GIS model to assess the extent of wetland loss in the Richmond catchment of New South Wales. The model consisted of a multi-attribute, drainage proximity and flow accumulation submodels. The drainage proximity submodel established a series of buffers (150, 300 and 450m) around water bodies. The flow accumulation submodel used the hydrologic analysis functions available in ArcGIS Spatial Analyst to select the areas with highest flow accumulation (defined by flow accumulation values greater than one standard deviation above the mean). The combination of the drainage proximity submodel and the flow accumulation submodel presents a good hydrologic approach to wetland delineation. Earley's study [24], however, considered that the location of vegetation species was determined by specific local physical characteristics and, therefore, neglected to use a vegetation submodel. In conclusion, this predictive GIS model accentuated the hydrology factor and underestimated the vegetation factor.

B. Objectives

- The first objective of this study was to create a suitability model to delineate wetlands in the Houston area. This model combined GIS and remote sensing technologies. Joining these technologies was predicted to provide a faster and easier way to delineate wetlands.

- A second objective was to establish if this methodology could provide satisfying spatial accuracy and low cost processes. If this objective was not fulfilled, the reasons why and the improvements necessary to fulfill the objective had to be determined.

This overall goal was to delineate general wetlands. General wetlands are defined using the three major characteristics of wetlands: soil, vegetation and hydrology. However, field data are not used. General wetlands are not limited to jurisdictional wetlands. To determine jurisdictional wetlands (e.g. NWI maps), field indicators are necessary. With the absence of recent wetland location data in this study area, the suitability model was calibrated and validated using 1999 wetland locations determined by the Spatial Science Laboratory.

CHAPTER II

MODEL DEVELOPMENT

A. Synopsis

1. Study Area

The study area is located south of Houston, Texas as shown in Figure 1. It spans three counties: Harris, Galveston and Chambers and three hydrologic units: Trinity-San Jacinto (12040203), San Jacinto (12040104) and San Jacinto-Brazos (12040204). The site lies between Houston, one of most urbanized and industrialized areas in the nation, and Galveston Bay, the largest and most biologically productive estuary in Texas. The annual mean temperature is $21.7^{o}C$ and the annual total rainfall averages 102 cm. The Texas Gulf Coast has some of the most abundant and diverse wetlands in the world [25]. Many conservation programs, such as the Urban Wetland Restoration Program developed by the Texas Coastal Watershed Program have been established to protect these resources. Monitoring wetlands efficiently in this area would be of great help for conservation programs, regulatory agencies, and municipalities.

2. Suitability Model

The goal of this suitability model was to delineate wetlands in the Houston area. Wetlands exhibit three major indicators: hydrology, soil and vegetation and some implicit attributes (e.g. small slope). Therefore, combining multiple submodels was assumed appropriate for wetland delineation. Such a modeling approach allowed consideration of each of the wetland indicators individually and assignment to each a wetland suitability index (WSI) of equal weight. This study combined four sub-

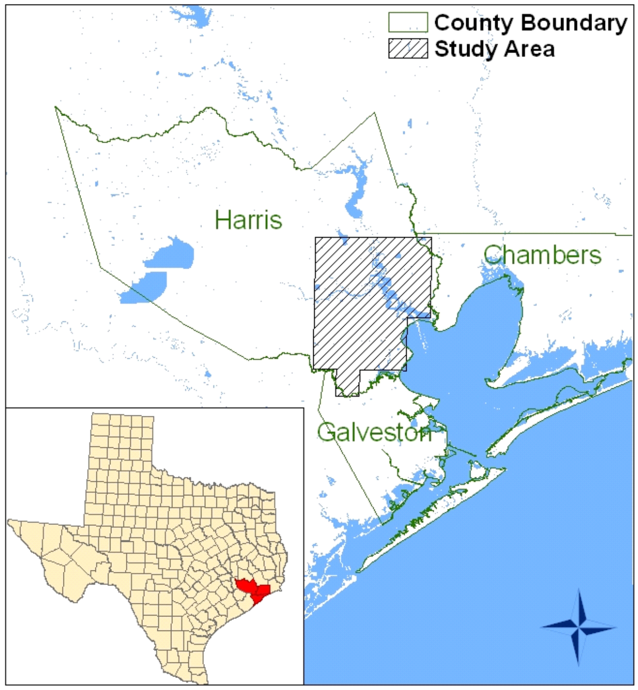

Fig. 1. Location of the study area

models. Three of them were applied to the primary wetland indicators: hydrology submodel, soil submodel and vegetation submodel. The fourth was a multi-attribute submodel, designed to select the physical landscape features not compatible with the existence of a wetland. Each submodel established a WSI. The final result of this suitability model was a general WSI obtained by summing all of the individual WSI. This model combined Remote Sensing and GIS technologies. Classifying remote sensed data such as Landsat has been shown to be adequate for wetland vegetation detection [21], [11]. Using GIS is adequate to provide accurate soil and hydrologic data for wetland delineation [24]. Therefore, the vegetation submodel used a Landsat ETM classification and the other models were GIS-based models.

a. Hydrology Submodel

The hydrology submodel was based on combination of the flow accumulation and drainage proximity submodels created by Earley [24]. The flow accumulation submodel was divided into two screening levels. The first level used the USGS National Elevation Dataset (NED) (30m of spatial resolution) and the second level sharpened the results using digital elevation model (DEM) data derived from LIght Detection and Ranging (LIDAR) data (4.57m of spatial resolution). The first level used hydrologic analysis functions available in ArcGIS Spatial Analyst. First, sinks were filled in the NED to remove imperfections in the data. A raster of flow direction from each cell to its steepest downslope neighbor was created from those modified data. This raster was processed to obtain a raster of accumulated flow to each cell [26]. Pixels with the highest flow accumulation values were selected. These selected pixels were sharpened using the DEM derived from LIDAR data. Flat areas with large accumulations of flow are favorable to wetland location [24]. To model this, slope was calculated using the DEM data, so that for each NED pixel selected in the first level, all nearby DEM pixels with a slope inferior to a predefined slope (one percent or less) were considered favorable to wetland location. Because wetlands are usually located near water bodies [21], the drainage proximity submodel created buffers around water bodies and these areas were considered favorable to wetland location. The flow accumulation and drainage proximity submodels each established a wetland suitability index and both indexes were summed to obtain a hydrology WSI.

b. Soil Submodel

Using the Soil Survey Geographic (SSURGO) database and the list of hydric soils developed by the Natural Resource Conservation Service (NRCS), all soils defined as hydric and with high frequency of flooding were assigned a favorable soil wetland suitability index.

c. Vegetation Submodel

Landsat ETM data were classified using the software ENVI. An unsupervised isodata classification was completed to detect six land cover classes: water, urban, grassland, bare ground, forest and wetland. Areas showing presence of hydrophytic vegetation were assigned a favorable vegetation indexes for wetland location.

d. Multi-Attribute Data Submodel

The multi-attribute data submodel was designed to select physical landscape features that are not compatible with the existence of wetlands. Land-cover exclusive to wetlands such as deep open water and urban area were excluded. Water bodies excluded were extracted from the NHD data. The urban areas were defined by using the results of Landsat classification developed in the vegetation submodel.

All the wetland suitability indexes (WSI) generated by these individual submodels were summed to obtain a general WSI.

3. Data Requirements

To implement the four submodels previously described, the following data were required:

- **Soil Survey Geographic (SSURGO) database**: Developed by the United States Department of Agriculture (USDA) Natural Resources Conservation Service (NRCS), SSURGO provides the highest level of detail available for soils data on a county-wide basis. The SSURGO data are available through the Soil Data

Mart USDA website [27]. The 2004 version SSURGO data for Harris, Chambers and Galveston counties were used in this study. Their format is ArcGIS shapefile, their mapping scales range from 1:12,000 to 1:63,360, their datum is NAD83, and their Projected Coordinate System (PCS) is UTM zone 15.

- **Georeferenced File Format (GeoTIFF) from the sensor system Landsat 7 Enhanced Thematic Mapper-Plus (ETM+)**: It was useful for characterizing vegetation and hydrology. Landsat data are available through the TexasView's website [28]. These data were developed in 2003. Their format is GeoTIFF, their spatial resolution is 30m, their datum is WGS 1984, and their PCS is UTM zone 15.

- **Digital Orthophoto Quadrangles (DOQ)**: A DOQ is a computer-generated image of an aerial photograph in which the image displacement caused by terrain relief and camera tilt has been removed. The DOQ combines the image characteristics of the original photograph with the georeferenced qualities of a map. DOQs are black and white (B/W), natural color, or color-infrared (CIR) images with one meter ground resolution. These data were developed in 2003. They facilitated identification of vegetation and hydrology. DOQs are available through the National Center for Earth Resources Observation and Science (EROS) website [29]. Their format is remote-sensing image, their datum is NAD83, and their PCS is UTM zone 15.

- **National Hydrography Dataset (NHD)**: NHD is "a comprehensive set of digital spatial data that contains information about surface water features such as lakes, ponds, streams, rivers, springs and wells." NHD data are available through the USGS website [30]. Their format is ArcGIS shapefile, their mapping scale is 1:100,000, their datum is NAD83, and their Coordinate Reference

System (CRS) is Decimal degrees.

- **Elevation data derived from an airborne LIDAR Topographic Mapping System (ALTMS)**: These data were created to be used as a highly accurate, inexpensive way to create digital topographic vector and raster files for implementation in GIS. These data supported computing accurate hydrology and topography. They were developed in 2002 for the Houston Galveston Area Council which shared them for this study. They have a raster format with cell size 4.57m (15ft) and elevation value accuracy better than 15cm. Their datum is NAD83, and their PCS is State Plane State Plane zone 5401 (Texas South Central). Their horizontal units are feet.

- **USGS National Elevation Dataset (NED)**: The NED dataset was developed in 1999 by merging the highest spatial resolution, best quality elevation data available across the United States into a seamless raster format. These data allowed computing accurate hydrology. The scale is 1:24,000 for the conterminous US. They have a raster format with cell size 30m and elevation expressed in centimeters. Their datum is NAD83, and their CRS is Decimal Degrees.

- **National Wetland Inventories maps (NWI)**: These maps were developed from 1992, 1:65,000 scale aerial photography. These aerial photos provide the type and acreage of wetlands present in 1992. They represent wetlands status in 1992 and will serve as the basis for comparison wetland evolution over time. Their format is ArcGIS shapefile, their datum is NAD27, and their PCS is UTM zone 15.

- **1999 wetlands location**. In 2002, the Spatial Science Laboratory (SSL) (Texas A&M University, College Station, TX) conducted a study entitled *Eval-*

uating the Loss of Wetlands in the Galveston Bay Area [31]. It used data extracted from the 1992 NWI maps and data obtained by interpretation of 1999 aerial photography (0.5m spatial resolution) to establish a 1999 wetlands location map. Its format is ArcGIS shapefile, its datum is NAD27, and its PCS is UTM zone 14.

4. Calibration and Validation of the Suitability Model

Since the SSL data [31] were the most recent available, the suitability model developed in this study was calibrated and validated by comparing its results with the results from the SSL study.

5. Expected Results

The expectations of this study were multiple. First, this study aimed to determine whether a low cost suitability model combining GIS and remote sensing technologies could be created to delineate wetlands in the Houston area. Second, the project attempted to demonstrate through model validation that this methodology can provide spatial accuracy using low cost processes. Therefore, the project aimed to demonstrate that this model was as capable of delineating general wetlands as a manual delineation using high-altitude aerial photography as the primary data source.

B. Hydrology Submodel

As previously described, the hydrology submodel was based on drainage proximity and flow accumulation. First, the drainage proximity was modeled.

1. Drainage Proximity

Three shapefiles were taken from the National Hydrography Dataset (NHD):

- Flowline: line shapefile that represents the streams, rivers, ditches and other narrow water courses.
- Areas: polygon shapefile that represents wide reaches such as estuaries.
- Waterbody: polygon shapefile that represents water bodies including lakes or seas.

As in Earley's model [24], three buffers (150, 300 and 450m) were created around water features. The three sets of buffers were unified using the ArcGIS union tool to create a unique buffer feature as shown in Figure 2. In case of buffer superposition, the buffer of smallest value was considered.

Using the editor tool, a field named index was added to the attribute table. The 150, 300 and 450m buffers were assigned, respectively, a WSI of three, two and one. Those values were arbitrarily chosen and were a main focus during the model calibration. Afterward, the flow accumulation was modeled.

2. Flow Accumulation

The hydrologic analysis functions available in Spatial Analyst in ArcGIS were used on the NED data for the full extent of the watershed.

The NED, shown in Figure 3, includes hydrologic units: 12040204 West Galveston Bay, 12040203 North Galveston Bay, 12040202 East Galveston Bay, 12040201 Sabine Lake, 12040104 Buffalo-San Jacinto, 12040103 East Fork San Jacinto, 12040102 Spring, 12040101 West Fork San Jacinto and 12030203 Lower Trinity. Sinks were filled

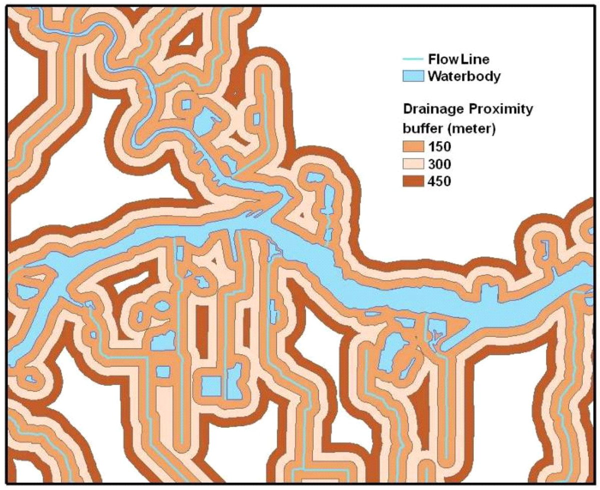

Fig. 2. Illustration of the drainage proximity submodel consisting of three buffers (150, 300 and 450m)

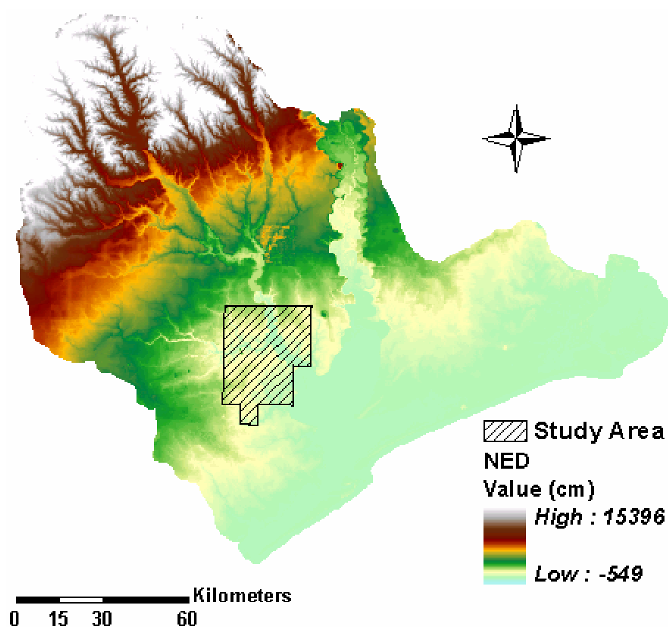

Fig. 3. NED of the watershed used for the flow accumulation submodel

in the NED to remove imperfection in the data. Then, a raster of flow direction from each cell to its steepest downslope neighbor, shown in Figure 4, was created from the filled data using the flow direction function.

Using the ArcGIS Flow Accumulation function, this raster was processed to obtain a raster of accumulated flow to each cell shown in Figure 5 [26].

Areas with a large value of accumulated flow are favorable to wetland location [24]. The flow accumulation raster has values ranging from 0 to 19984592 with a mean of 3055 and a standard deviation of 146512. The highest values of flow accumulation match up more or less with rivers. However, in urban areas in which rivers have been disturbed by human activity, the flow accumulation is important because the river path has often been changed. Therefore, during rainfall events the original path of the river may flood, and that is where wetlands are more likely located. For optimization, the raster was cut to fit the study area. Pixels with the highest flow accumulation values (more than the mean of 3055) were selected for sharpening using the DEM derived from LIDAR data. This pixel selection was transformed into a line shapefile using the ArcGIS Conversion Tools shown in Figure 6.

After flow accumulation was modeled,using Spatial Analyst on the DEM shown in Figure 7, slopes were calculated (Figure 8). Then, using the classification tool, a new raster was created containing only pixels having a slope value less than one percent. Earley's study [24] considered a slope of two percent. Because of the especially flat nature of the study area (Figure 9) this value was decreased to one percent. Using zonal statistics, approximately 47 percent of the pixels were found to have a slope value less than one percent. Since such a large percentage was selected using one percent, it was decided to use a slope of less than 0.25 percent for flow accumulation modeling (Figure 10).

In order to determine the flat areas with large flow accumulation, a 100m buffer

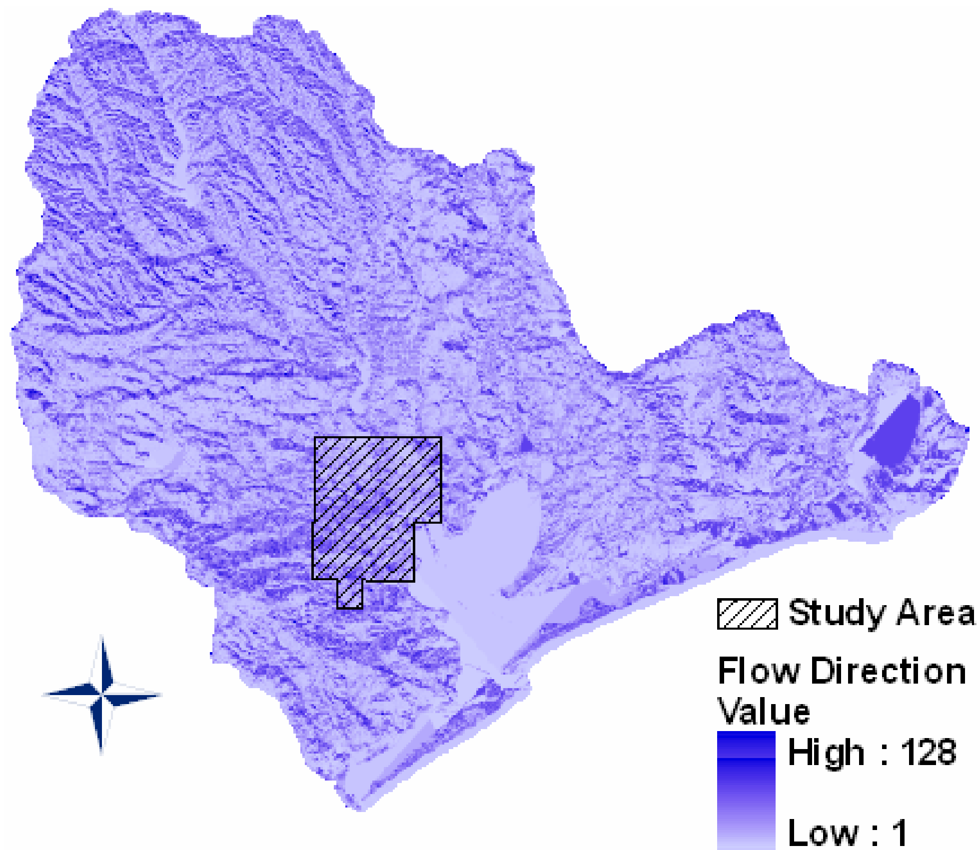

Fig. 4. Flow direction raster of the watershed obtained using the ArcGIS flow direction function on the NED

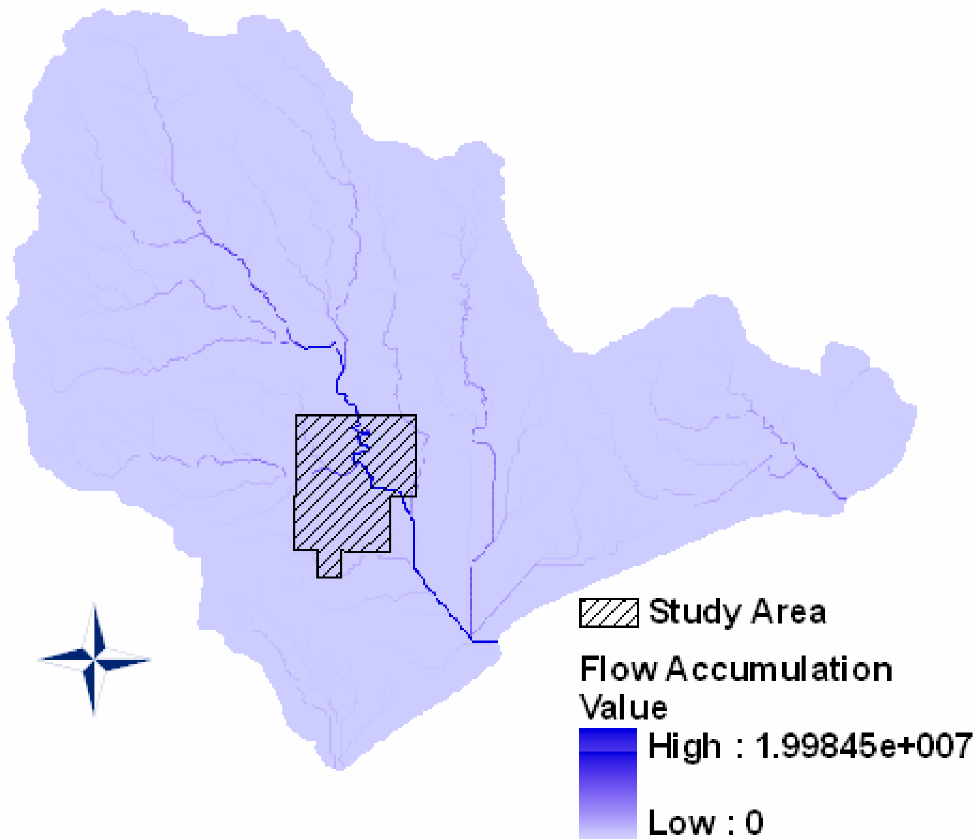

Fig. 5. Flow accumulation raster of the watershed obtained using the ArcGIS flow accumulation function on the flow direction raster

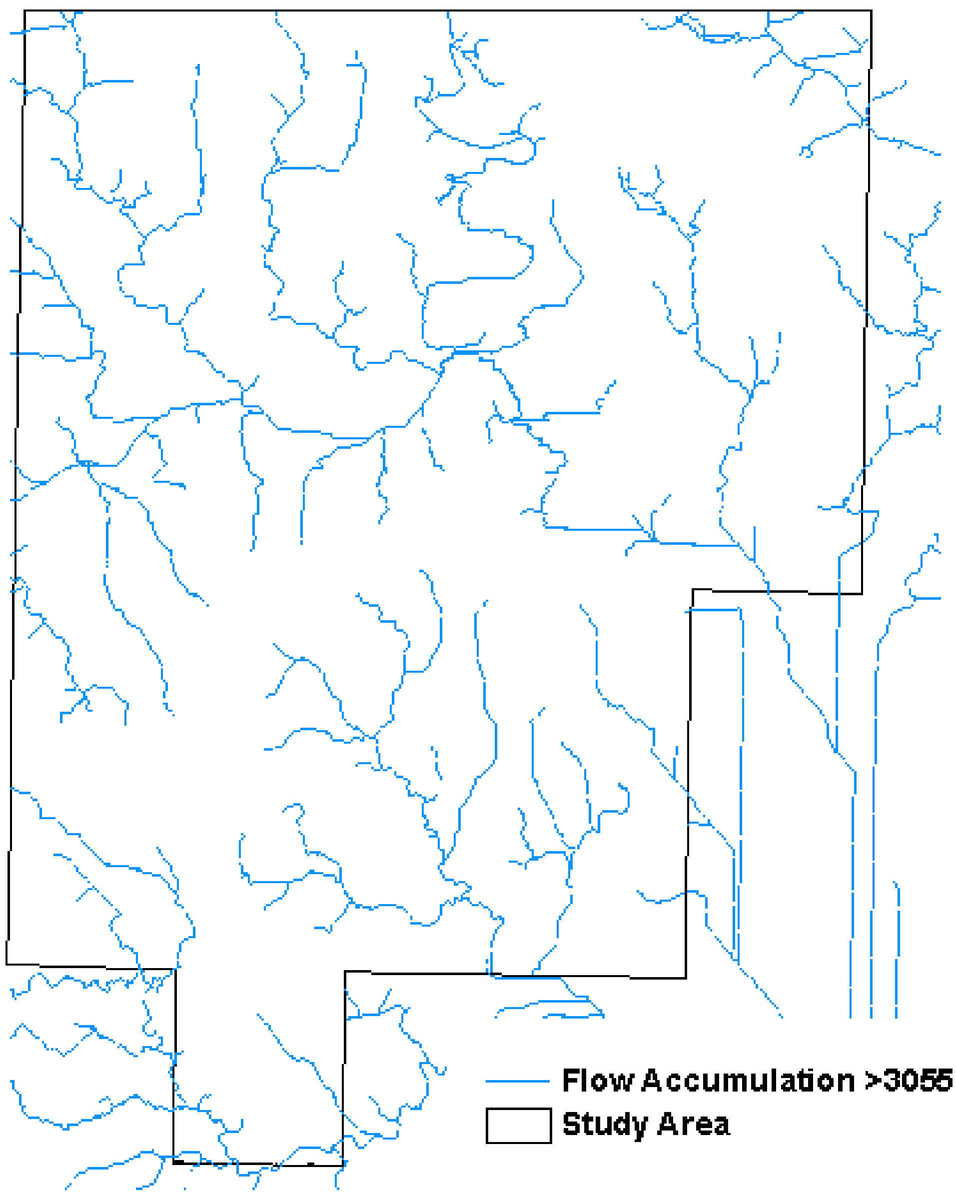

Fig. 6. Line shapefile of high flow accumulation (more than 3055) obtained from the flow accumulation raster

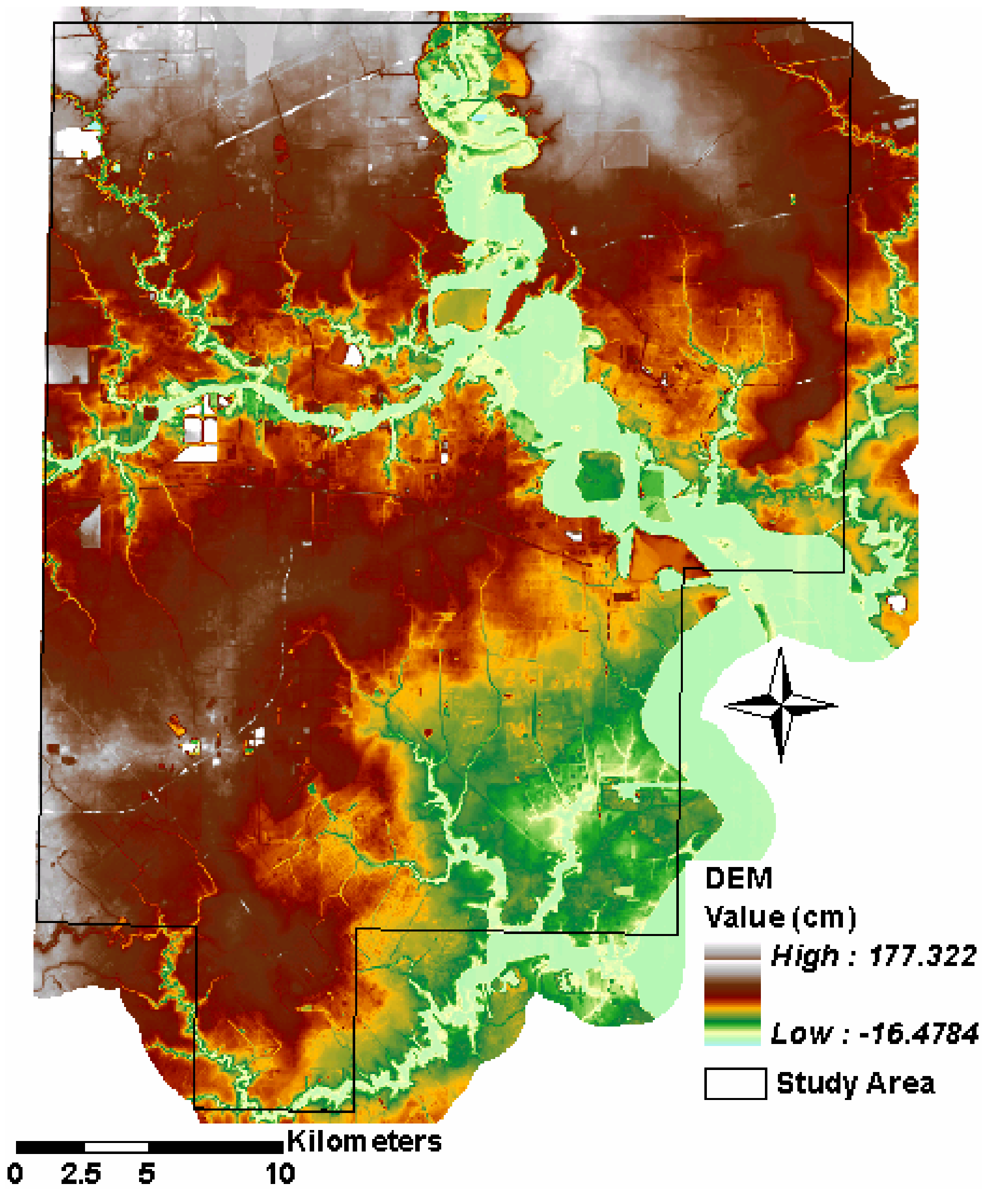

Fig. 7. DEM derived from LIDAR data (15cm vertical precision) for the study area

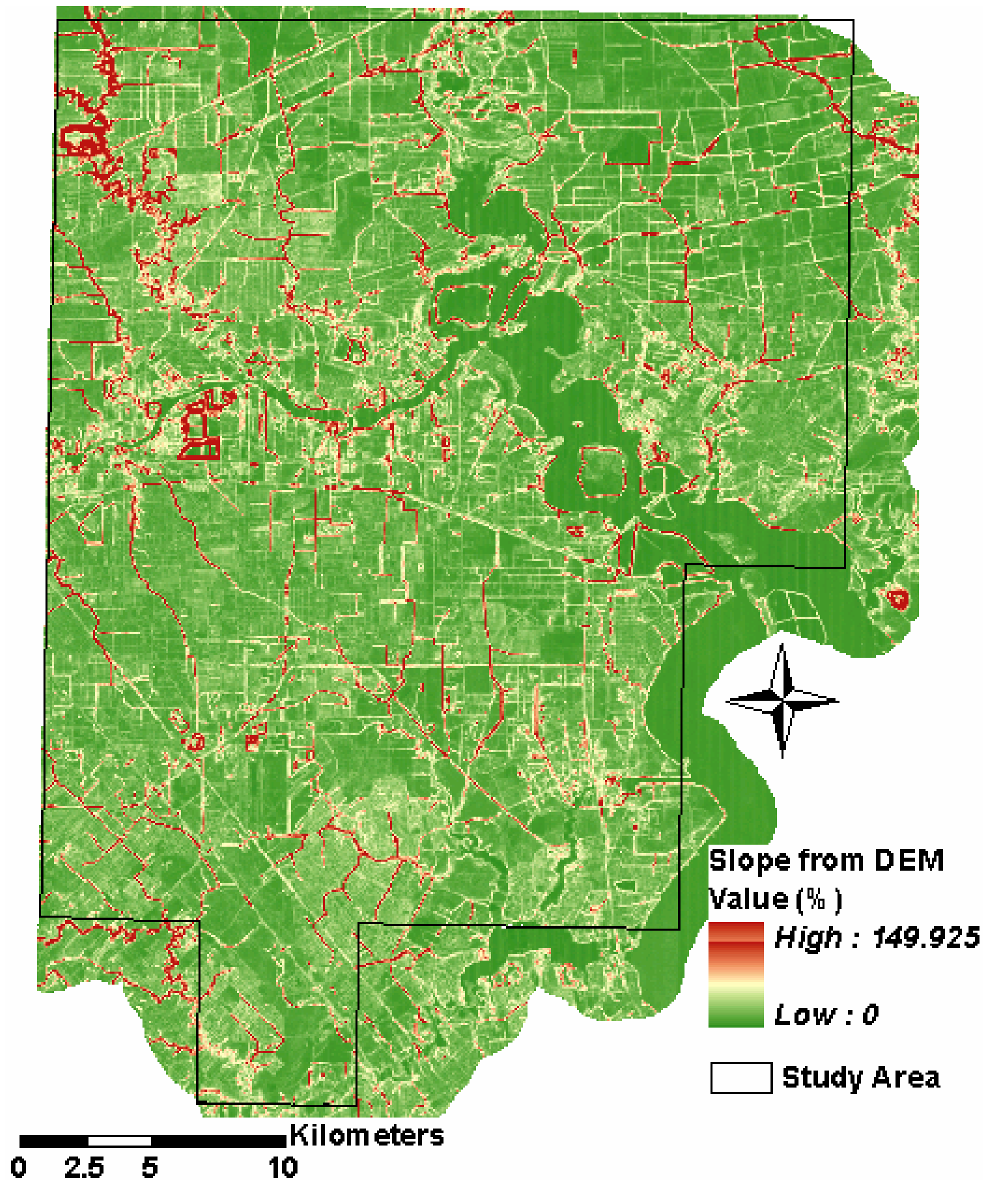

Fig. 8. Slope calculated using the ArcGIS slope function on the DEM derived from LIDAR data in the study area

was created around the flow accumulation line shapefile, and all pixel of slope less than 0.25 percent contained in this buffer were extracted. Those pixels were assigned a wetland suitability index of three. This value was arbitrary chosen and was a main focus during model calibration.

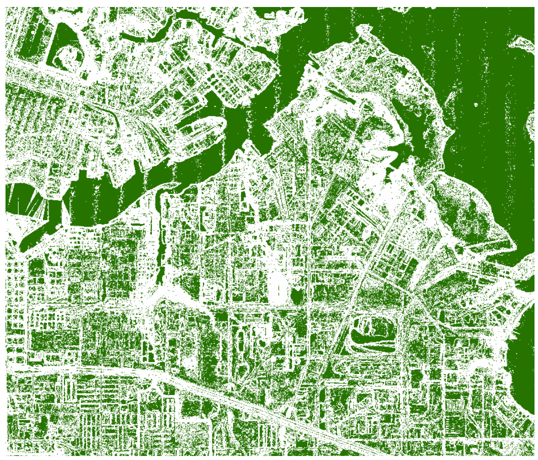

Fig. 9. Slope derived from the DEM with dark areas representing a slope less than one percent

Fig. 10. Slope derived from the DEM with dark areas representing a slope less than 0.25 percent

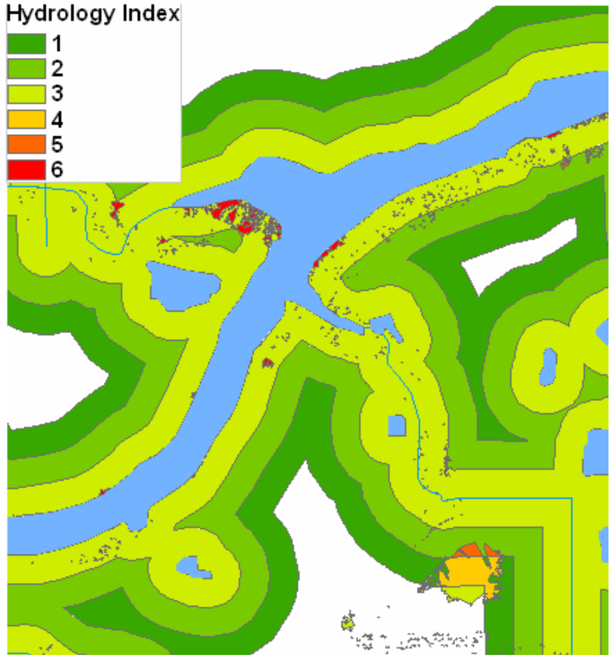

Fig. 11. Illustration of the WSI for the hydrology submodel with values ranging from zero to six

3. Wetland Suitability Index

Each of the hydrology submodels established a WSI. The two WSI were summed to establish an WSI specific to the hydrology submodel. The summed value ranged from zero up to six. Higher WSI values correspond to a greater probability that the area is a wetland. The final result for the hydrologic submodel prior to calibration is shown in Figure 11.

C. Soil Submodel

SSURGO data were downloaded from the Soil Data Mart USDA website [27] for the three counties in the study area: Chambers (TX071 version 11/08/2004), Galveston (TX167 version 10/27/2004) and Harris (TX201 version 11/05/2004). The exported files contained tabular and spatial data. Tabular data were imported in an empty MS Access SSURGO template database. Spatial data were available in the ArcView Shape File format.

1. Data Preparation

The SSURGO data have a complex structure established on three levels: map unit, soil component, and horizon. Each polygon in the spatial data represents a map unit. Each map unit can be composed of up to three components (unique soils with individual properties). Each component can include up to six horizons (different depths with individual properties) as shown in the Figure 12 [32].

The spatial data are developed at the map unit level, whereas the tabular data can be developed at the map unit, component, or horizon levels depending on the table chosen. The three main tables are map unit, component and horizon. These tables contain data about map units, soil components and soil horizons respectively. However a lot of other tables are contained in the SSURGO data. Some of these tables are specific to the map units, some to the components and some to the horizons. In this study, the two levels considered were: map unit and component. The components are characterized as hydric following the classification established by National Technical Committee for Hydric Soils [5]. If a component is characterized as hydric it is specified in the component table using a yes/no field called Hydric Rating. If rated as hydric, the specific criteria met are listed in the component hydric criteria table (cohydriccriteria). Since spatial data are only defined at the map unit level, it was decided for this study to consider the percentage of hydric soil in the map unit as the principal indicator to establish the WSI for the soil submodel. The map unit, component and cohydriccriteria tables were imported into ArcGIS. The component table was joined to the cohydriccriteria table using the cokey field, and to the map unit table using the mukey field. Using the comppct-r field (percentage of the map unit's component) from the component table and the hydriccriterion field from the Component Hydric Criteria table, a third field was created named HYDPERCT that contained the percentage of hydric soil in the map unit. This table was then joined to the soilmu-a shapefile using the mukey index to visualize the results at the map unit level.

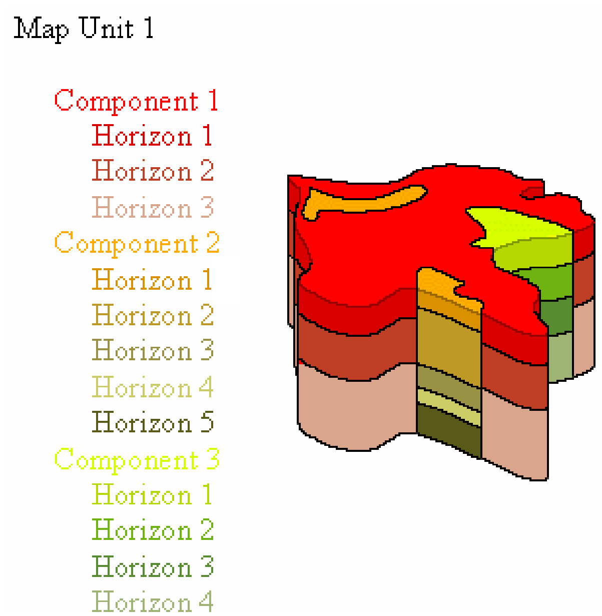

Fig. 12. Conceptual representation of SSURGO data [32]

2. Wetland Suitability Index

The WSI values for the soil submodel ranged from zero up to six as shown in Table I. The higher the WSI value, the higher the probability that the area was a wetland. Values were arbitrary chosen and were a main focus during the model calibration. The final result for the soil submodel are shown in Figure 13.

Table I. WSI depending on the percentage of hydric soil in the map unit

% of hydric soil in the map unit	WSI*
10-40 %	2
40-70 %	4
70-100 %	6

WSI*: Wetland Suitability Index

D. Vegetation Submodel

1. Image Preprocessing

Two Landsat 7 ETM+ images were downloaded from the TexasView website [28]. Those images were taken on 23 March 2003. Their locations are coded 2539 (path 25 row 39) and 2540 (path 25 row 40) as shown in Figure 14.

Other dates were available for the same locations, however, data used were chosen following the results of Hodgson et al. [33] who indicated wetlands could be better defined on imagery acquired in spring when the water table was high. March 23 was the closest available date to spring. Examination of the Texas climatologic data for March 2003, however, show that precipitation in March 2003 was below normal which might imply a low water table. The NOAA website [34] gives a precipitation depth for March 2003 between 0.97 and 1.34 inches in the Houston area (data based on weather stations at: Houston Bush International Airport, Houston Hobby Airport, Port of Houston, and Houston HWSO), whereas the average precipitation depth from 1903 to 2003 is 1.76 inches for the month of March. Precipitation variation was considered during interpretation of results.

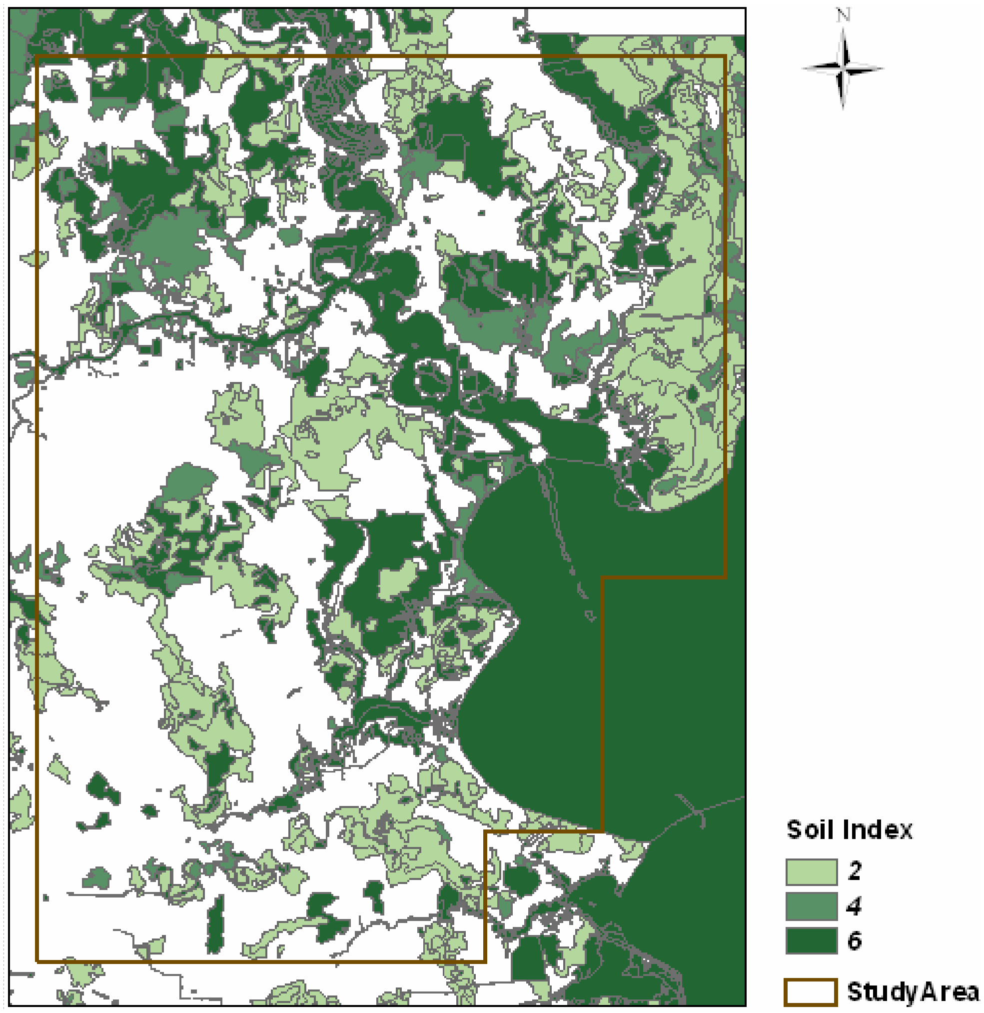

Fig. 13. WSI for the soil submodel, in the study area, with values ranging from zero to six

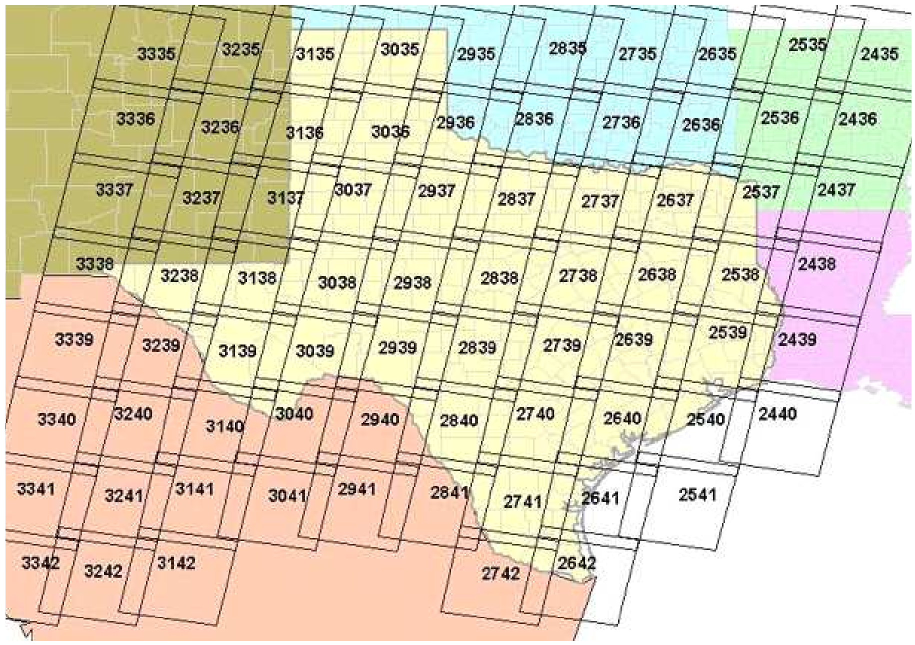

Fig. 14. Landsat 7ETM+ path and row image footprints for Texas, with study area location coded 2539 and 2540 [28]

This level of correction of these data is Terrain Correction (Level 1T). The level of correction is defined by the USGS [35] as including "radiometric, geometric, and precision correction, as well as the use of a digital elevation model (DEM) to correct parallax error due to local topographic relief. All Level 1T products are processed by the National Land Archive Production System (NLAPS)."

The two images used in this study were imported into ENVI (Environment for Visualizing Images, Research Systems, Inc.; Boulder, CO). First they were photo-calibrated using the TM Calibration Tool. This process transformed the raw data, constituted of digital numbers (DN), to exoatmospheric reflectance (reflectance above the atmosphere) [36]. This procedure was necessary because the images were not taken at the same time of the day and, therefore, the sun illumination was different. Illumination differences could have strongly affected the DN. The images were then reprojected to change their datum from WGS to North American Datum (NAD) 83, their PCS staying UTM zone 15. All the background values from band one (b1) were converted from a negative value to a zero value. Next, the images were unified as one picture using the Mosaic option in ENVI. The parameters for this function were chosen as follows: the zero background value was ignored, a feathering distance of 50 pixels was selected to smooth the transition between the two images, and no color balancing was selected. A Region Of Interest (ROI) consisting of four points located at the four corners of the study area was created. Using this ROI, the Landsat image was clipped. The resulting output image is shown in Figure 15. The preprocessing procedure is illustrated by a flow chart in Figure 16.

2. Image Enhancement

To increase the efficiency of the classification and decrease error sources, some image enhancement algorithms were used as illustrated in Figure 17. The results of these image enhancements were stacked together to form a seven band composite image which was used for classification. The image enhancements chosen were two vegetation indexes and the Principal Component Analysis (PCA). These enhancements are described in the following text.

a. Vegetation Indexes: Infrared Index and Tasseled Cap

Vegetation indexes have been used in remote sensing since the 1960s. There are more than twenty of them commonly in use [37]. Their function is to extract from remotely sensed data an index that will maximize the variance between different types of vegetation or increase the sensitivity to certain vegetation properties, such as chlorophyll content, or water content in the leaves. Many of the indexes are redundant [37]. One of the most well known vegetation indexes is the Normalized

Fig. 15. Preprocessed image obtained by the mosaic of two landsat 7 ETM+ images and then adjustment to the study area

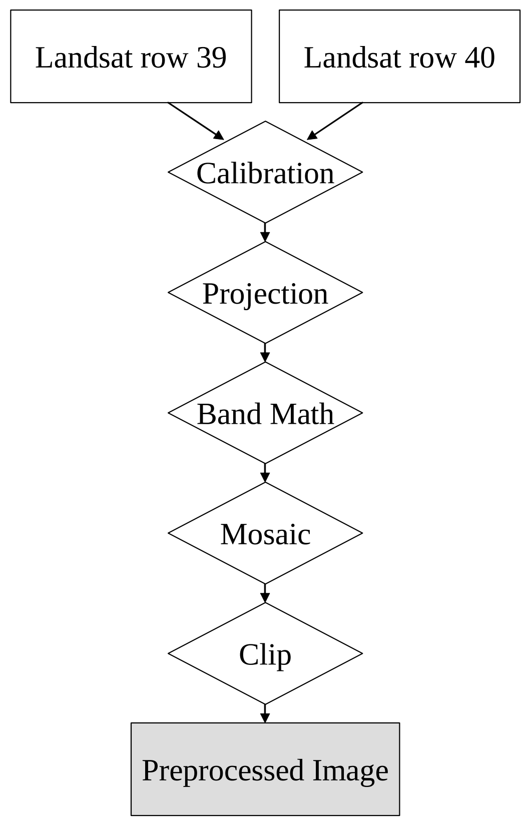

Fig. 16. Flow chart illustrating the preprocessing procedures for the vegetation submodel

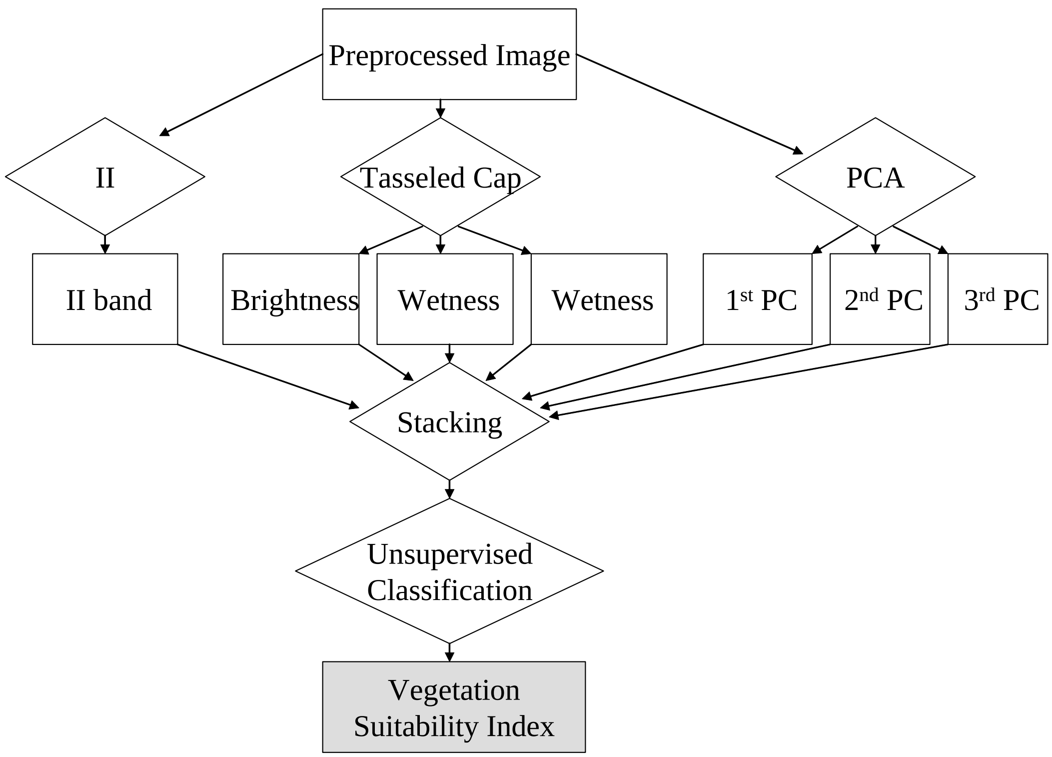

Fig. 17. Flow chart illustrating the image enhancement procedures for the vegetation submodel

Difference Vegetation Index (NDVI), developed in 1974 by Rouse et al. [38]:

$$NDVI = (NIR - Red)/(NIR + Red) \tag{2.1}$$

Where NIR is band four in ETM+ and Red is band three in ETM+. NDVI reduces noise coming from such sources as atmospheric haze or difference in solar exposure (especially for mountainous areas) [8]. Currently NDVI is used around the world to provide crop forecasts. It has also been proven useful in monitoring of wet grassland area [22]. NDVI was, therefore, considered for this study. However, Hardisky et al. [39] found that for wetlands the Infrared Index was more responsive to variations in plant biomass and water stress than was NDVI [12]. The Infrared Index (II) (shown in Figure 18) seemed a better choice for this study and was calculated using the Band Math tool.

$$II = (NIR - MidIR)/(NIR + MidIR) \tag{2.2}$$

Where NIR is band four in ETM+ and MidIR is band five in ETM+. The second vegetation index applied was the tasseled cap transformation developed by Kauth and Thomas in 1976 [40]. The value of the tasseled cap transformation in wetland detection or classification has been extensively demonstrated [20], [33], [21]. The ENVI Tassled Cap tool was used to perform this transformation. For Landsat 7 ETM+ data, this tool required calibrated reflectance. This tasseled cap transformation produces six output bands: Brightness, Greenness, Wetness, Fourth (Haze), Fifth, Sixth. In this study only the Brightness (Figure 19), Greenness (Figure 20), and Wetness (Figure 21) bands were considered in the classification.

Fig. 18. Infrared index used in the seven band composite image of the vegetation submodel

Fig. 19. Brightness band, result of the tasseled cap transformation, used in the seven band composite image of the vegetation submodel

Fig. 20. Greenness band, result of the tasseled cap transformation, used in the seven band composite image of the vegetation submodel

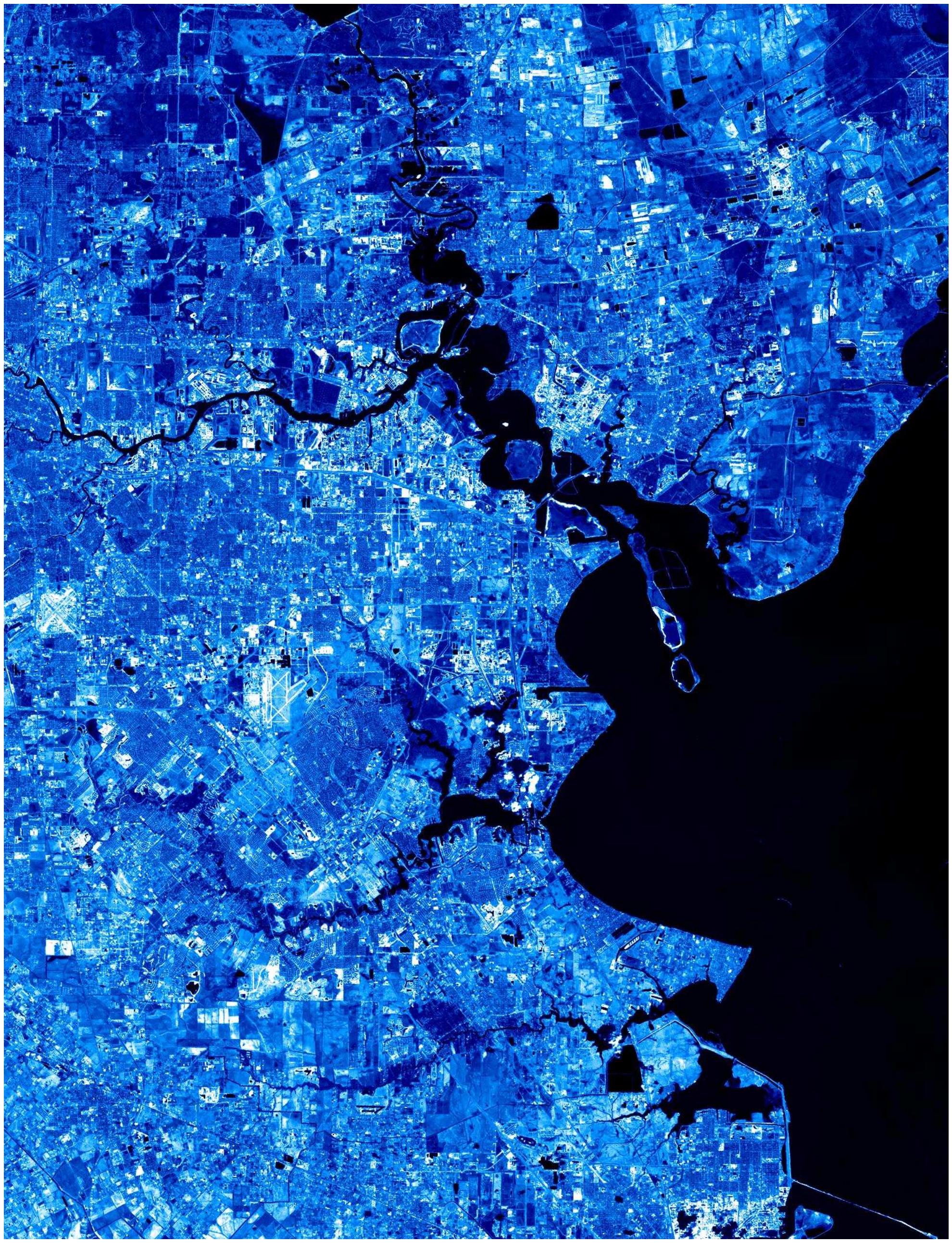

Fig. 21. Wetness band, result of the tasseled cap transformation, used in the seven band composite image of the vegetation submodel

b. Principal Components Analysis

The PCA creates bands that are linear combinations of raw sensor data. Those bands are uncorrelated and noise is extracted from them. The PCA has the ability to reduce dimensionality to two or three bands and still keep all the information content [12]. In this study, the three first components (Figures 22 to 24) were used for the classification. Those seven bands were stacked together using the ENVI Layer Stacking tool.

3. Image Classification

The three major types of classifications are hard classifications (such as supervised and unsupervised), classifications applying fuzzy logic, and hybrid classifications [12]. Hard classifications establish a membership to a unique class (usually land cover) for each pixel, whereas fuzzy logic established membership grades to multiple classes. In this study, there was no need to consider multiple memberships because the only concern was the wetland class, so fuzzy logic was not used. In supervised classification, some training areas are first delineated by the analyst; therefore, an *a priori* knowledge of the study area is required. Then, using these training areas as references, the computer established memberships to a unique land cover for all pixels. In unsupervised classification, the computer first extracts clusters of spectrally similar pixels. Then, the analyst defines the type of land cover represented by each cluster. Hybrid classifications combine supervised and unsupervised classifications [21]. Considering that the study area was not well known and that no precise wetland locations (such as GPS points) or other field data were available; it was decided to use an unsupervised classification.

The isodata classification was applied with a five percent threshold, a maximum

Fig. 22. 1st principal component, result of the PCA, used in the seven band composite image of the vegetation submodel

Fig. 23. 2nd principal component, result of the PCA, used in the seven band composite image of the vegetation submodel

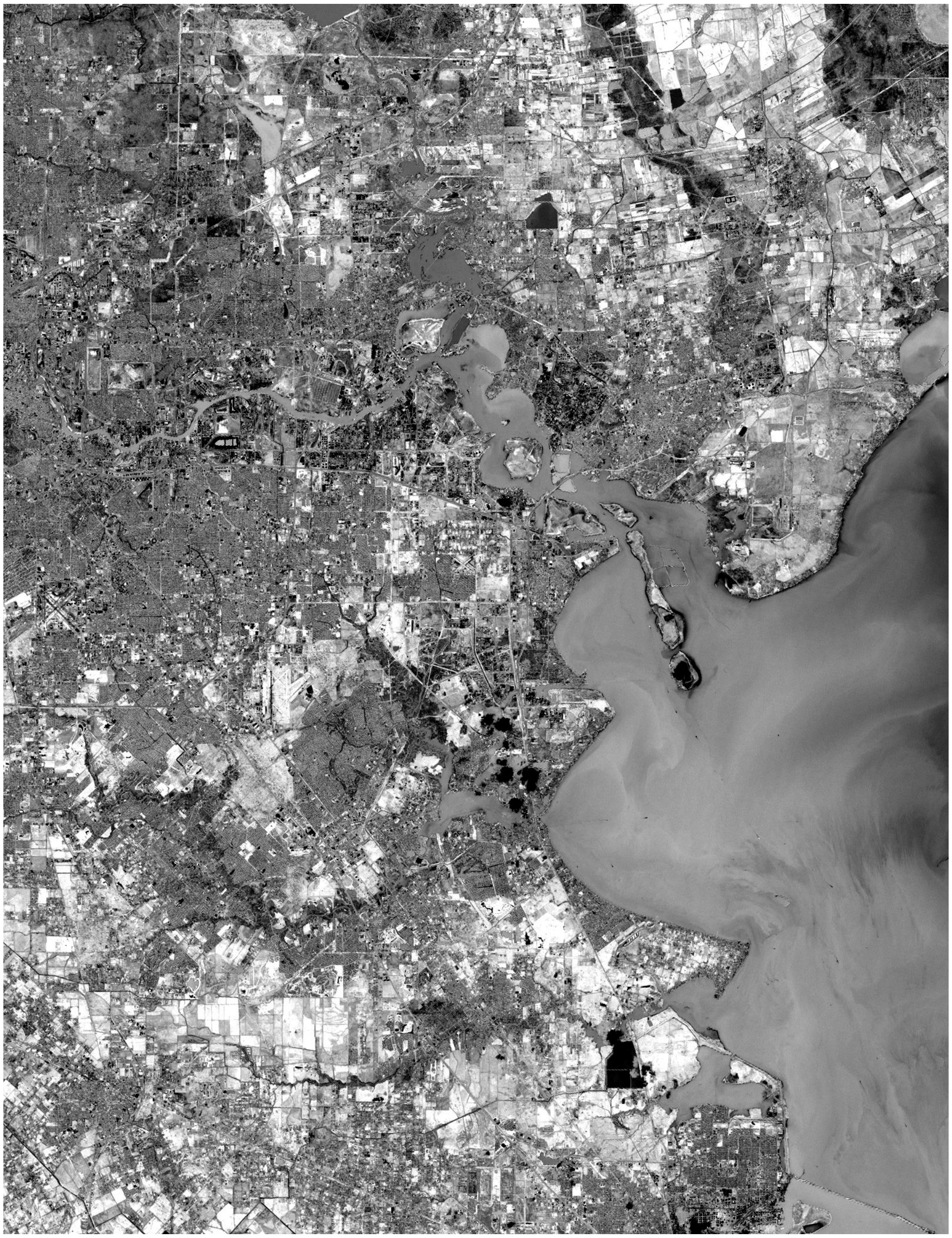

Fig. 24. 3rd principal component, result of the PCA, used in the seven band composite image of the vegetation submodel

of 40 iterations, a minimum of 40 pixels in a class, a minimum of 30 classes, and a maximum of 40 classes. The classification was produced using ENVI. Two classifications were achieved, the first using the Landsat data without enhancement and the second using the seven band composite image. This process helped in evaluating the value of the enhancements. In both cases a total of 30 classes were created. Those classes were combined in six land cover classes: water, urban, grassland, bare ground, forest and wetland. Those images were then converted in GRID format and exported in ArcGIS. All classes other than wetland were given a WSI value of zero and the wetland pixels were given a WSI value of six. The wetland location extracted from the landsat image and the seven band composite image are shown in Figure 25 and Figure 26.

E. Multi-Attribute Submodel

The Multi-Attribute Data Model was designed to select physical landscape features that were not compatible with the existence of wetlands. Land-cover exclusive to wetlands such as deep open water and urban area were excluded. Also areas with high slope were excluded.

1. Open Water

Open water areas were extracted from the NHD data. All open water areas were simply cut out from the study area. As previously described, NWI and the SSL study results [31] were applied to wetlands and open waters. In this study, the focus was only wetlands.

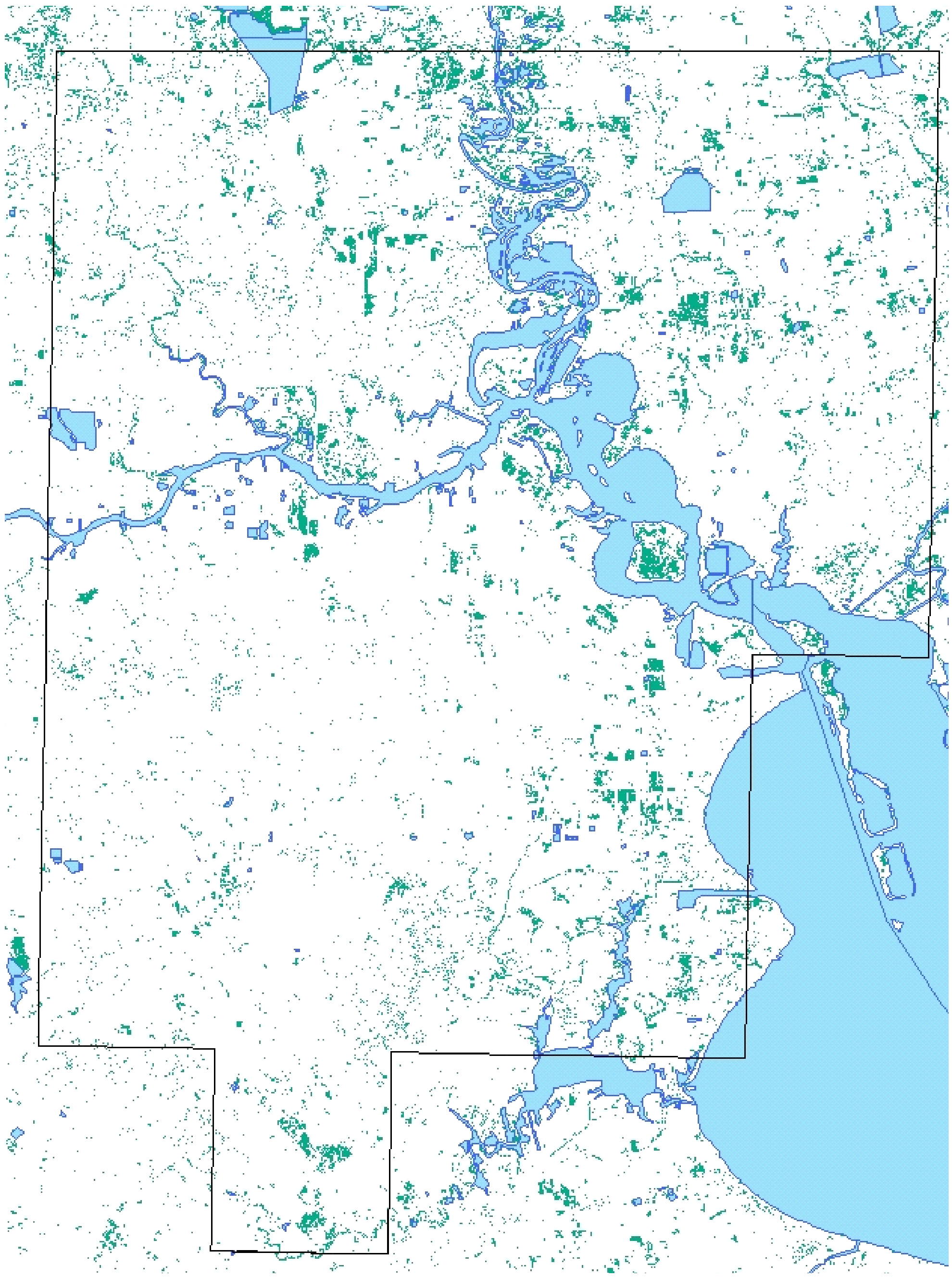

Fig. 25. Wetland vegetation location extracted from Landsat without enhancement

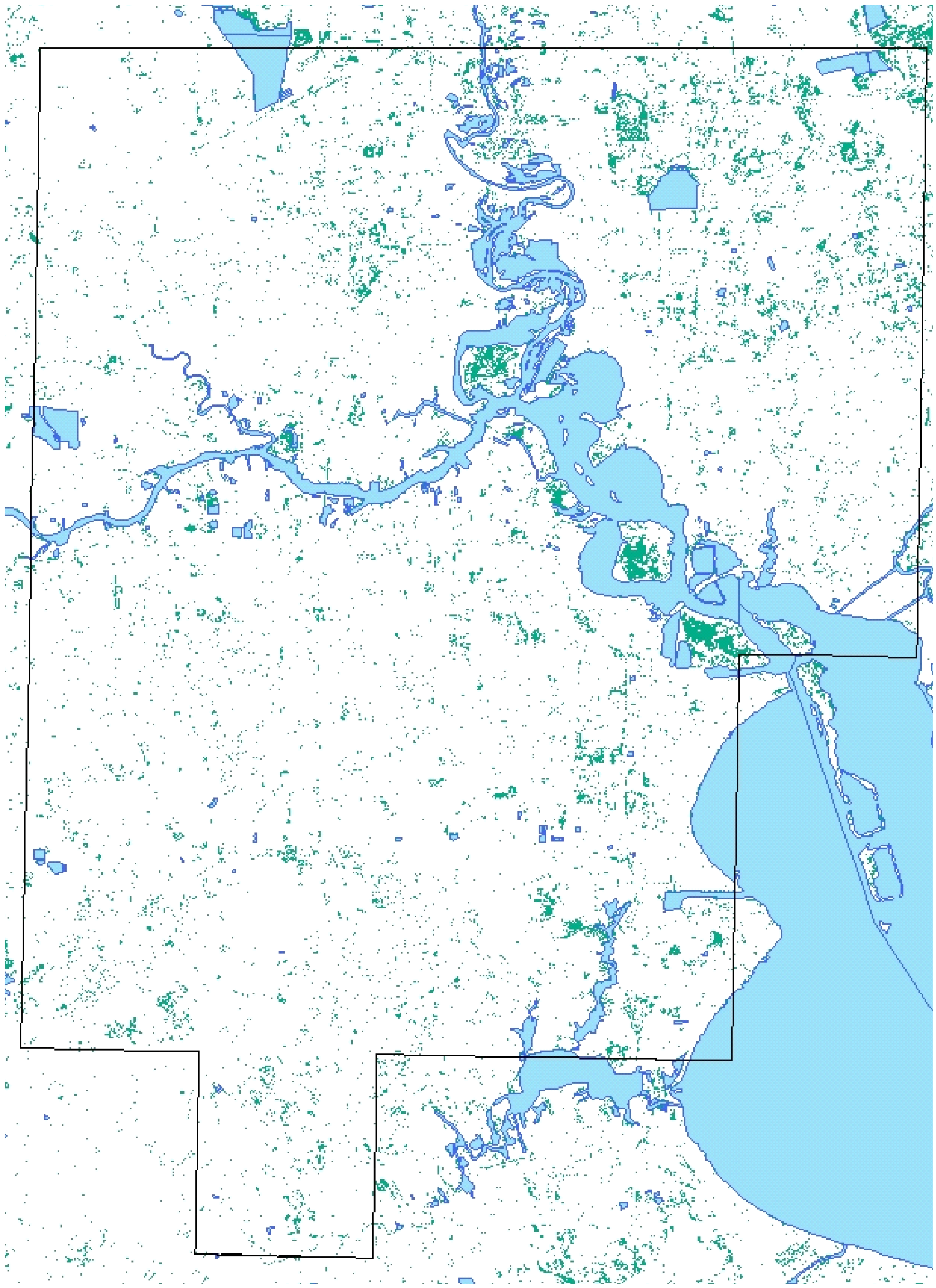

Fig. 26. Wetland vegetation location extracted from the seven band composite image

2. High Value Slope

The slope was calculated from the DEM derived from LIDAR. Wetlands are usually located in very flat areas; therefore, it was decided to exclude all areas that had a slope greater than five percent. This value was arbitrary and was intended to be examined during the model calibration. All pixels with a slope greater than five percent were given a Nodata value that would cancel all other WSI. Not knowing the real efficiency of such exclusion, it was decided to use this index only if the type II error was high.

3. Urban Area

Urban areas were defined by classification of satellite images. Using the Landsat classification defined in the vegetation submodel, pixels qualified as urban were given a negative WSI of -9 (1.5 times the weight of the other WSI. However, by doing this, wetlands located in urban area that were too small to be detected by the 30m Landsat data would be eliminated. Moreover those wetlands might be the ones that are most threaten by human activities. Therefore, it was decided to use this index only if the type II error was unacceptably high.

F. Combination of Submodels

All the submodels were converted in raster of pixel size 4.572m to match the LIDAR format. Then all the WSI were summed using the ArcGIS map algebra function.

CHAPTER III

CALIBRATION

A. Synopsis

1. Calibration Area

Half of the study area was chosen for use in model calibration and the other half was reserved for model validation. The calibration area was determined by dividing the study area into quarter quads and by choosing half of the quarter quads using the Excel random function. The calibration area is shown in Figure 27.

2. Wetland Suitability Indexes

Two main objectives of the calibration were to establish pertinent WSI values for each model and to determine a minimum value requirement for the final WSI that would allow areas to be defined as wetland. During model development, arbitrary WSI values were defined for each submodel. Initially equal weight was given to each submodel. For each submodel, the minimum WSI had been fixed to zero and the maximum WSI had been fixed to six. The WSI arbitrarily set at the beginning of calibration are shown in Table II.

3. Model Evaluation

To evaluate the results of the suitability model, resulting predictions were compared to 1999 wetland locations identified in a study made by the Spatial Science Laboratory (SSL) entitled *Evaluating the Loss of Wetlands in the Galveston Bay Area* [31] (Figure 28). These data were obtained by interpretation of 1999 aerial photography (0.5 meter spatial resolution). The result of the SSL study was an update

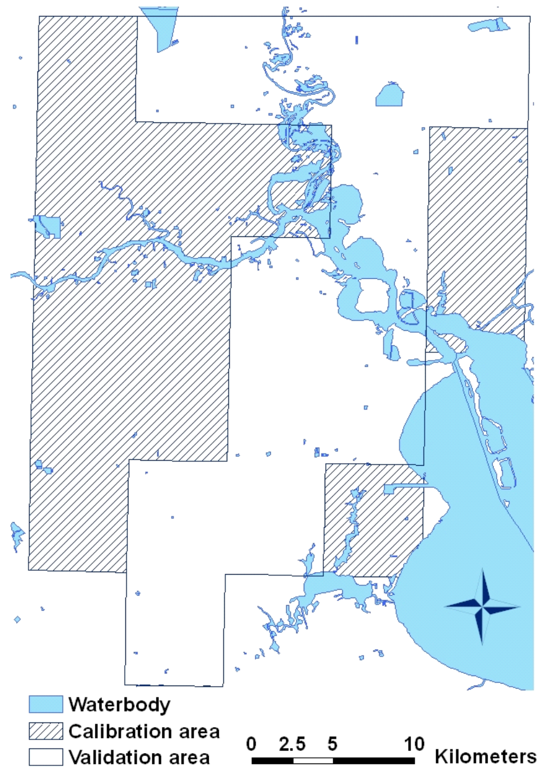

Fig. 27. Calibration area composed of half of the study area chosen randomly

Table II. WSI values used for model version one

Drainage proximity WSI* (buffer size)	Flow accumulation WSI*	Soil WSI* (% of hydric soil)	Vegetation WSI*
1 (150m)		2 (10-40%)	
2 (300m)	3	4 (40-70%)	6
3 (450m)		6 (70-100%)	

*WSI: Wetland Suitability Index

of the NWI map. Hereafter, these 1999 wetland locations are referred as 1999 NWI. Open waters were dissociated from the wetlands and, therefore, were not considered in this study's model calibration or validation. The total area of the updated 1999 NWI considered for the calibration was 30.476 km^2.

B. Model Version One

1. Qualitative Analysis

Output of model version one is shown in Figure 29. It appears that this model was particularly sensitive to riverine wetlands but did not detect upland wetlands well. Higher WSI were located around rivers and water bodies as illustrated in Figure 30. Conversely, the upland wetland areas generated low WSI as illustrated in Figure 31. The most sensitive submodel to upland wetlands seemed to be the soil submodel as illustrated in Figure 32. Consequently it was decided to give a bit more weight to the soil submodel and multiply its WSI by 1.5.

2. Quantitative Analysis

To evaluate this model, two parameters were considered. Defining the null hypothesis as: area identified as wetland in the 1999 NWI. The first parameter was determined as (1 - (type I error)) expressed in percent and named "wetland detected in %". It was equivalent to the percentage of 1999 NWI wetland area detected by this model. The second parameter, called "model accuracy", was the type II error expressed in percent and was equivalent to the 1999 NWI wetland area detected by this model over the total area detected by this model. Those results (Table III) were obtained using the zonal statistic tool.

The percentage of wetland detected was unacceptably low. Therefore, it

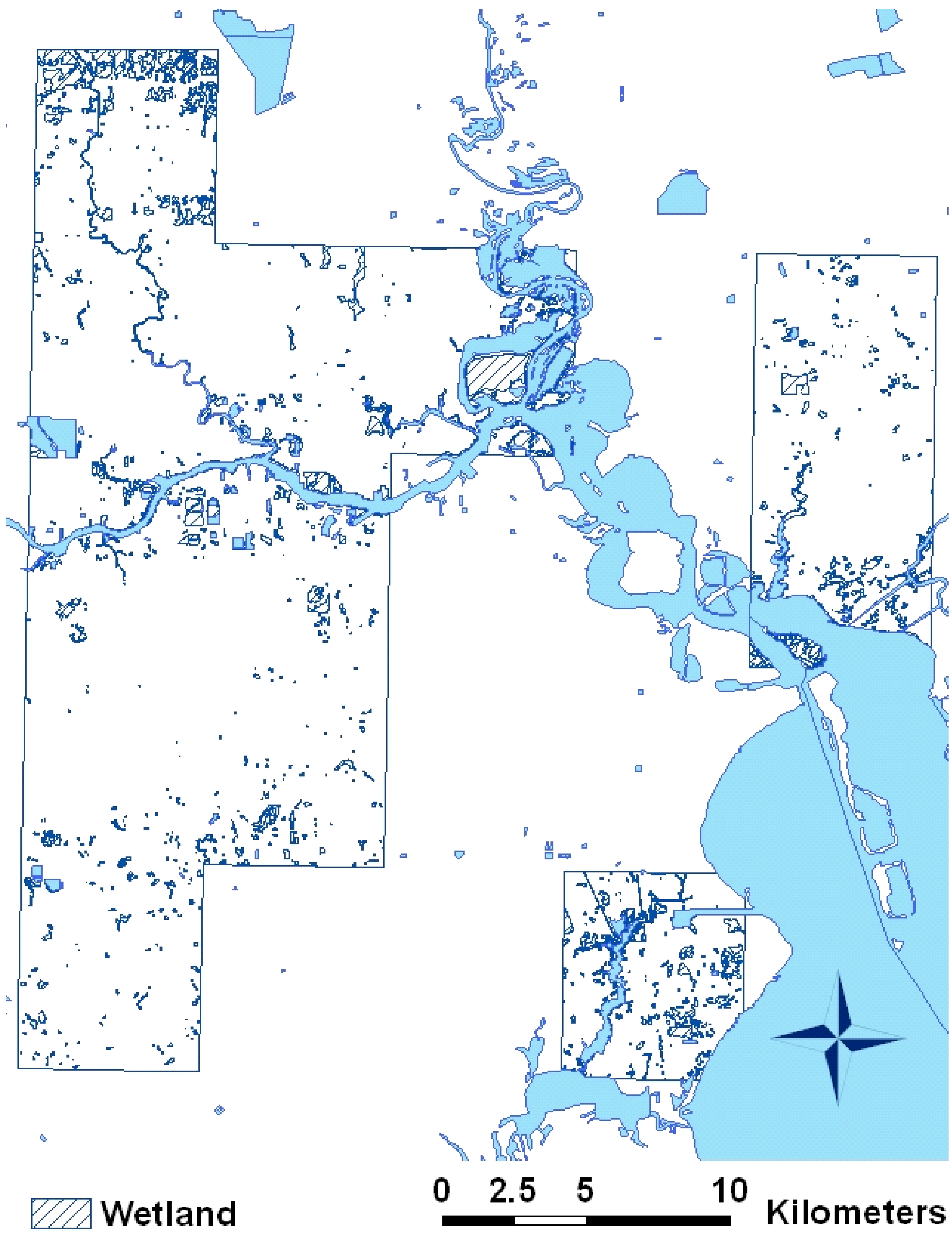

Fig. 28. Wetland locations identified in the SSL study [31] used for calibration

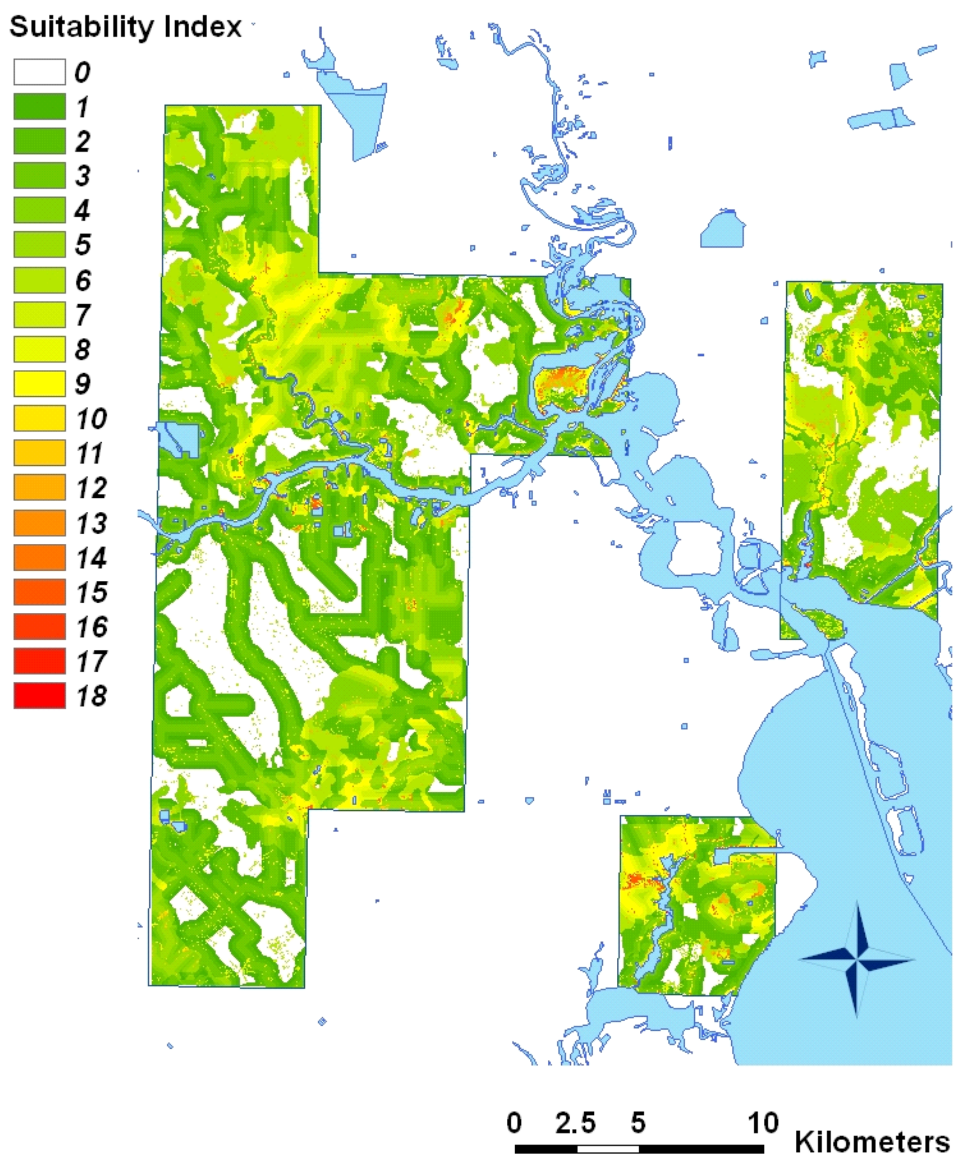

Fig. 29. Final WSI values for model version one

Fig. 30. Illustration of the presence of the highest WSI values around streams and waterbodies

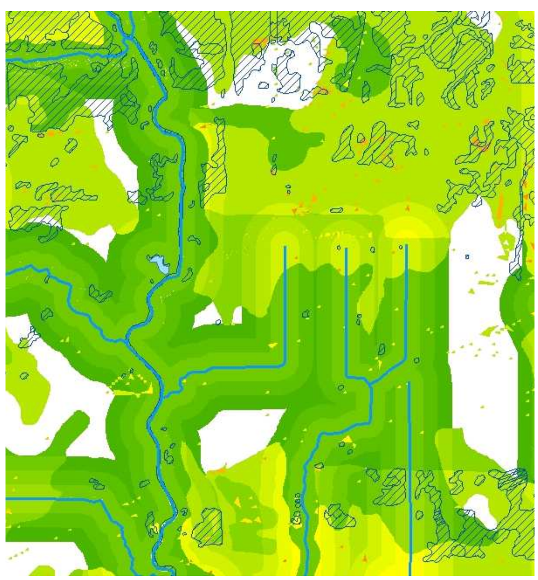

Fig. 31. Illustration of low WSI values of for upland wetlands

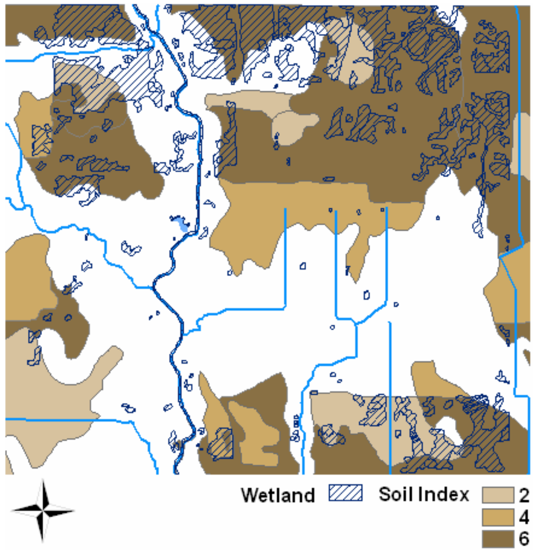

Fig. 32. Illustration of good sensitivity of the soil submodel with higher WSI values for upland wetlands

Table III. Calibration results for model version one

WSI*	Wetland detected in %	Model accuracy in %
≥ 15	4.4	59.2
≥ 12	10.0	45.7
≥ 9	27.2	27.8
≥ 6	55.0	14.3
≥ 1	94.2	7.7

*WSI: Wetland Suitability Index

was decided to introduce the slope index and the urban index to the multi-attribute submodel. The goal of this change was to reduce type II error.

C. Model Version Two

After multiplying the weight of the soil submodel by 1.5 (i.e. WSI values from three to nine) to improve upland wetland detection, and including all the components of the multi-attribute submodel (i.e. slope and urban) to improve the accuracy of the model; model version two used the WSI values shown in Table IV.

Table IV. WSI values used for model version two

Drain. prox. WSI* (buffer size)	Flow acc. WSI*	Soil WSI* (% of hydric soil)	Veg. WSI*	Slope WSI*	Urban WSI*
1 (150m)		3 (10-40 %)			
2 (300m)	3	6 (40-70 %)	6	-9	nodata
3 (450m)		9 (70-100 %)			

*WSI: Wetland Suitability Index

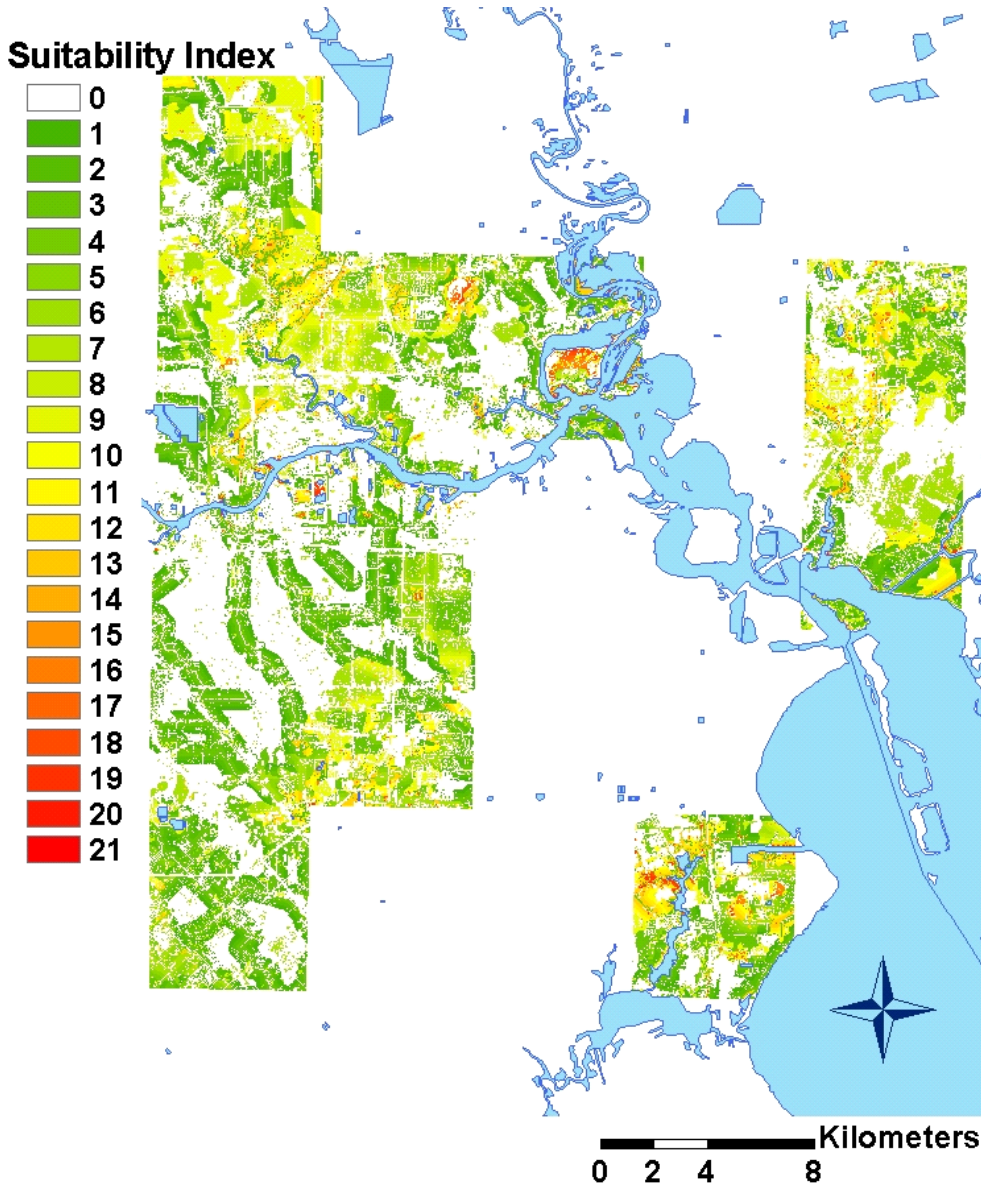

Fig. 33. Final WSI values for model version two

1. Qualitative Analysis

Output of model version two is shown in Figure 33. Riverine wetlands were still well detected. In addition, there was an improvement in the detection of upland wetlands as shown in Figure 34 . Therefore, the weight of 1.5 on the soil WSI was kept.

2. Quantitative Analysis

Model modifications slightly improved model accuracy as shown in Table V. However, the urban and slope indexes excluded 6.3% of wetlands. This percentage was considered too high. For this reason, and also considering the model's lack of sensitivity to changes, it was decided to evaluate each individual index independently of the whole model.

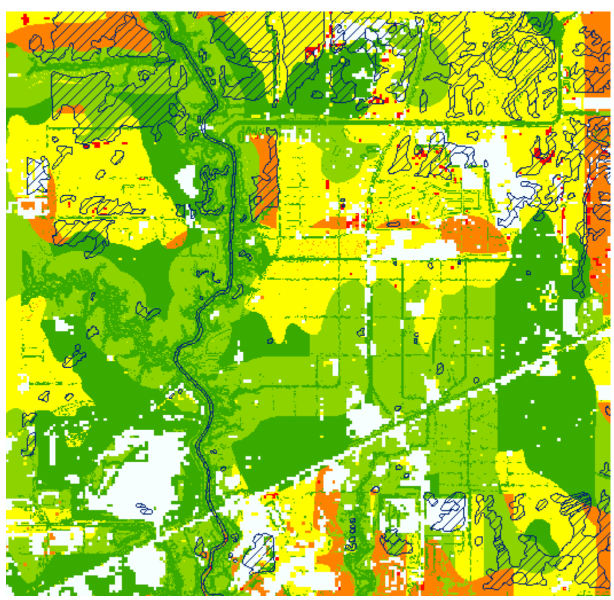

Fig. 34. Detection of the upland wetlands

Table V. Calibration results for model version two

WSI*	Wetland detected in %	Model accuracy in %
≥ 18	4.6	64.7
≥ 15	10.7	49.0
≥ 12	22.9	34.9
≥ 9	50.1	19.3
≥ 6	56.9	13.6
≥ 1	88.9	9.5

*WSI: Wetland Suitability Index

3. Indexes Evaluation

The methodology used for the model evaluation was the same as that used to evaluate the model outputs. However, in this case each index was evaluated one by one and independently of the model. For this, each index was formatted as a raster of pixel size 4.572m (matching the LIDAR data format) and the 1999 NWI was kept as a polygon shapefile. As for the model, two parameters were calculated using the zonal statistic tool. Defining the null hypothesis as: area identified as wetland in the 1999 NWI. The first parameter was determined as 1 - (type I error) expressed in percent and named “wetland detected in %”. The second, called “index accuracy”, was the type II error expressed in percent. The results are displayed in the Tables VI to IX.

As expected, percentage of wetland detected and index accuracy were inversely

Table VI. Evaluation of the soil index

Soil	Wetland detected in %	Index accuracy in %
$\geq$ 70% hydric	46.2	17.0
$\geq$ 40% hydric	48.5	11.3
$\geq$ 10% hydric	61.8	9.4

Table VII. Evaluation of the flow accumulation index

Flow accumulation	Wetland detected in %	Index accuracy in %
Slope $\leq$ 0.25%	3.5	77.4
Slope $\leq$ 0.35%	15.5	13.0
Slope $\leq$ 0.5%	28.0	10.9

Table VIII. Evaluation of the drainage proximity index

Drainage proximity	Wetland detected in %	Index accuracy in %
Buffer of 150m	40.1	10.3
Buffer of 300m	59.9	8.5
Buffer of 450m	73.5	7.7

Table IX. Evaluation of the vegetation index

Vegetation	Wetland detected in %	Index accuracy in %
Landsat	33.1	31.5
seven bands composite	30.0	32.2
Tasseled cap	22.7	46.7
Infrared index	13.6	52.4

correlated. However, such poor values for the index accuracies were not expected. In fact, except for flow accumulation, all accuracies were less than 35%.

The soil index was quite efficient, so it was decided to keep it with a higher weighting compared to other indexes. The flow accumulation had a very good reliability for a slope threshold less than 0.25%, but it dropped drastically for higher slope values. So while the percent of wetland detected was low (3.5%), it was decided to keep it at 0.25% of slope threshold.

The drainage proximity index proved insensitive to variation between buffers widths. Therefore, it was decided to dissolve the three buffers into one single buffer of 450m. In addition, it was decided not to keep both hydrologic indexes at equal weight. The flow accumulation index was given a WSI value of four and the drainage proximity index was lowered to two. This change was made to reduce type II error.

The vegetation index provided unsatisfactory wetland prediction. First, the percentage of wetland detected was low (less than 40%). Moreover, the seven bands composite image did not improve the classification. It appears individual enhancements may have canceled each other out or that some part of the information was lost during enhancement. Therefore, it was decided to classify the tasseled cap transform and the Infrared Index to determine if the Landsat classification could be improved using a single enhancement. The results are shown in Table IX . The results for single enhancement are not as good as the ones obtained by classifying the Landsat data directly; therefore, it indicates that information was lost during enhancement. Considering the poor results of the enhancement techniques, it was decided to use the classification directly from the Landsat image.

The multi-attribute submodel indexes (urban and slope) were evaluated in a slightly different way. The purpose of those indexes was to reject or lower the total WSI for areas that are exclusive to wetland. Consequently, two attribute values were calculated using the zonal statistic tool. The first value was the percentage of non-wetland areas that had been rejected. The second value was the percentage of wetland areas that had been rejected. The results are displayed in the Tables X and XI .

The urban index rejected 26.9% of non-wetland, which was considered good. However it also rejected 12.4% of wetland. This percentage was unacceptably high especially considering that those wetlands may be the most endangered by human activities and require more monitoring. Therefore, the weight of 1.5 was decreased to one for this index.

The LIDAR slope index proved to be a poor predictor of wetlands. The percentage of wetland rejected was about the same as the percentage of non-wetland rejected. It was decided that LIDAR was not efficient in this case study and this index was not used further. Not using the LIDAR data was a step back in this study. Other data were large scale data: 1:12,000 to 1:63,360 for SSURGO, 1:100,000 for NHD, 1:24,000 for NED, 1:65,000 for NWI. The small scale presented by LIDAR and the 15cm vertical accuracy added a unique level of detail to the model.

Table X. Evaluation of the urban index

Urban	% of non-wetland rejected	% of wetland rejected
Landsat	26.9	12.4

Table XI. Evaluation of the slope index

LIDAR slope	% of non-wetland rejected	% of wetland rejected
$\geq 5\%$	12.4	8.6
$\geq 10\%$	5.1	3.7
$\geq 15\%$	3.0	2.1

D. Model Version Three

For model version three, the drainage proximity buffers were dissolved as one 450m buffer of WSI two, the WSI of the flow accumulation submodel was increased to four, the soil submodel and the vegetation submodel were kept as they were in version two, the slope submodel was not used, and the urban submodel was given a WSI of -6. The indexes for model version three are presented in Table XII.

1. Qualitative Analysis

Output of model version three is shown in Figure 35. The results looked very similar to model version two results. This model appeared to be reluctant to change. However, the model cannot provide very accurate outputs if the individual indexes constituting the model are not very accurate as demonstrated previously.

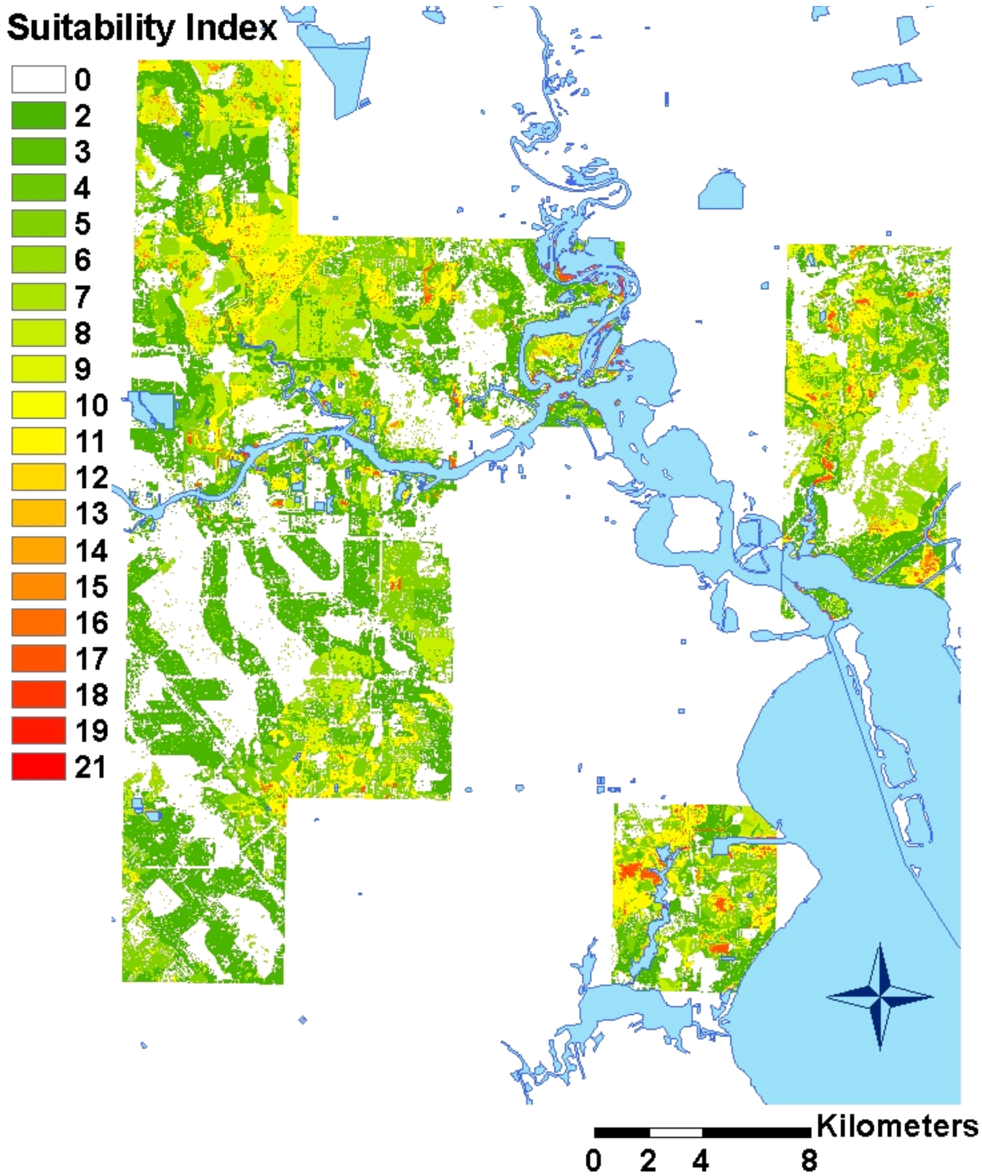

Fig. 35. Final WSI values for model version three

Table XII. WSI values used for model version three

Drain. prox. WSI*	Flow acc. WSI*	Soil WSI* (% of hydric soil)	Veg. WSI*	Urban WSI*
2	4	3 (10-40 %) 6 (40-70 %) 9 (70-100 %)	6	- 6

*WSI: Wetland Suitability Index

2. Quantitative Analysis

Model accuracy improved clearly for the highest WSI values as shown in Table XIII. However, the results were very similar to those of model version two. There was poor sensitivity to the changes. It appears that, without LIDAR data, the other data had too large a scale to delineate wetlands correctly.

Table XIII. Calibration results for model version three

WSI*	Wetland detected in %	Model accuracy in %
≥ 18	0.7	82.8
≥ 15	17.2	50.7
≥ 12	18.0	45.3
≥ 9	45.7	19.5
≥ 6	58.5	14.7
≥ 1	88.8	8.8

*WSI: Wetland Suitability Index

E. Model Version Four

In a final effort, a 2004 infrared DOQ of one meter pixel was used. This DOQ was obtained compressed as a MrSid file and, therefore, a bit of data accuracy was lost. However, considering that all the other model data were broad scale data, this DOQ still remained in comparison as a high spatial accuracy data. This DOQ was not considered for use during model development because LIDAR was expected to provide this study with sufficient accuracy for wetland delineation. Second, this DOQ offered only three bands, whereas Landsat offered seven bands particularly band five that is used for vegetation indexes such as the Infrared Index. The 2004 DOQ for Harris county was clipped to fit the study area as shown in Figure 36. Then, the resulting image was classified. As with the Landsat image, an unsupervised classification was used. The isodata classification was applied with a five percent threshold, a maximum of ten iterations, a minimum of 50 pixels in a class, a minimum of 30 classes, and a maximum of 40 classes. Thirty classes were created. Those classes were combined into six land cover classes: water, urban, grassland, bare ground, forest and wetland. Those images were then converted in GRID format and exported in ArcGIS.

The urban land cover class was used in the multi-attribute submodel with a WSI value of -6. The wetland land cover class was used in the vegetation submodel with a WSI value of six. Apart from these two modifications, this model kept the same configuration as model version three.

1. Qualitative Analysis

The two new indexes developed using the DOQ were vegetation index and urban index. With the 2004 DOQ, two considerable problems were encountered. First, the DOQ's metadata did not provide the date at which the pictures were taken during

Fig. 36. 2004 DOQ (one meter pixel) clipped to fit the study area

the year 2004. Historically in Texas, DOQ pictures are usually taken at the end of the summer to increase chances for a cloud free sky. However, the best time to detect wetlands is at the beginning of the spring when the water table is high. Therefore, the DOQ used may not have been capable of appropriately detecting wetlands. Second, the DOQ was obtained as one main file for all of Harris County. To obtain 1m pixel data, this file had to be developed from a large number of smaller pictures. Those pictures were taken at different times of the day and on different days. This implied that the sun position changed. The resulting illumination difference could have greatly affected the DN as it was seen previously for the Landsat data. Unfortunately, no photo-calibration could be undertaken because the time at which the individual pictures were taken was unknown and the pictures were already unified into one file. This problem became evident during classification. It appeared that urban pixels in the north-west of the study area had the same spectral signature as water pixels in the south-east of the study area. Those pixels could not be differentiated. Even after trying to minimize this problem, the classification yielded many pixels in the estuary that were detected as urban. Such large error was not acceptable for a satisfying land-cover classification.

2. Quantitative Analysis

The results of model version four using the DOQ were, as suspected, poor as shown in Table XIV. The wetland detected were about ten percent less than when using the Landsat data. Thus, using high scale data such as DOQ was not enough to improve the suitability model results. In fact, remote sensing data used for land-cover classification need to be corrected. Correcting them involves radiometric, geometric, precision, and exoatmospheric corrections. The DOQ was not corrected in such ways, and, therefore, was not accurate enough. Moreover, since this DOQ constituted of only three bands, it did not provide as much spectral information as the Landsat consisting of seven bands. This reduction of information was certainly important in the loss of accuracy of the indexes developed using the DOQ. Consequently, use of the DOQ was discontinued.

Table XIV. Calibration results for model version four using the 2004 DOQ

WSI*	Wetland detected in %	Model accuracy in %
≥ 18	0.6	82.5
≥ 15	16.7	53.2
≥ 12	17.4	39.3
≥ 9	44.7	16.9
≥ 6	54.5	11.9
≥ 1	90.6	8.2

*WSI: Wetland Suitability Index

F. Results and Conclusions

Model version three was chosen as the final model for validation. The accuracy of this model was improved during its development; however its performance was generally not satisfying. Theoretically, it would be acceptable to consider as wetlands those areas that present at least two of the three major wetland characteristics: predominance of hydric soils, presence of hydrophytic vegetation and specific hydrologic conditions. This correlates with a WSI value of at least nine. For model version three with a WSI value greater or equal to nine provided detection of 45.7% and accuracy of 19.5%. This percentage of wetland area detected was not enough; less than half of the 1999 NWI wetland area. In addition, the accuracy demonstrated was quite low. These results demonstrated that there was still too much error in the model. This error could have been coming from the indexes. Did error come from the scale of the data used to generate the indexes? The data scales were quite large and might have favored the detection of large wetlands. Did error come from the imprecision of the data used to generate the indexes? The broader the scale, the higher the imprecision might be. Did it come from an incorrect calibration of the indexes? This model would then be more sensitive to certain types of wetland such as riverine or upland. This error could also have come from ground truth data (1999 NWI wetland locations). The 1999 NWI map was produced to represent the state of wetlands in 1999 in the Houston area. However there might have been some changes between 1999 and the present. Could wetland change over time explain some of the difference between the 1999 NWI wetland location and the outputs of this model? The validation was performed with those questions in mind.

CHAPTER IV

VALIDATION

A. Synopsis

1. Validation Area

The second half of the study area was used to perform model validation. The validation area is shown in Figure 37.

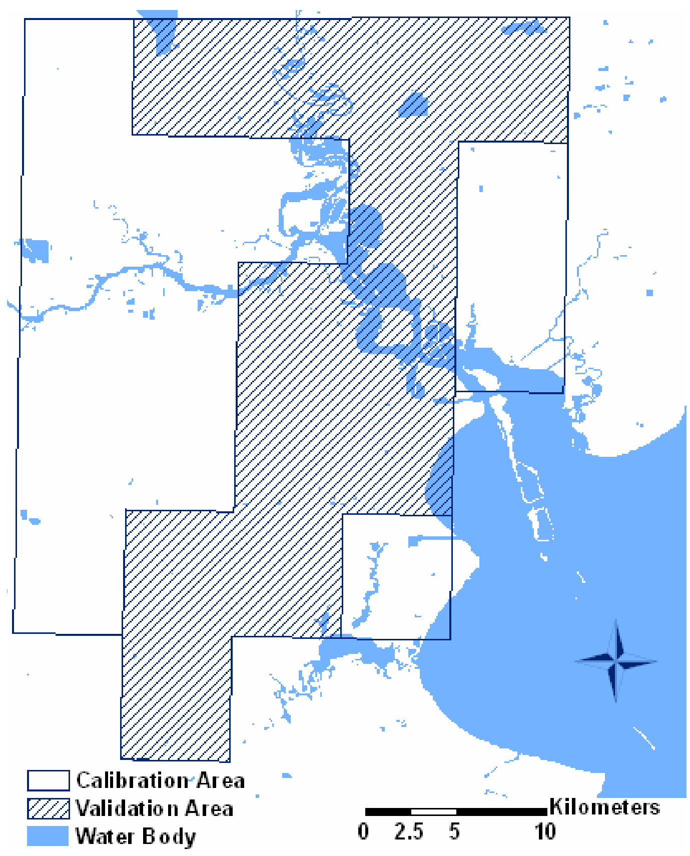

Fig. 37. Validation area composed of half of the study area chosen randomly

2. Model Evaluation

As with calibration, to evaluate the results of our suitability model, we compared model results to the 1999 NWI wetland locations presented in Figure 38. The total area of 1999 NWI considered for the validation was of 46.980 km^2.

B. Model Version Three

1. Qualitative Analysis

Visual inspection of the calibration results indicated that the partition of riverine versus upland wetlands was quite similar to that obtained during the calibration. Both appeared to be well detected, and the model did not particularly favor a certain type of wetland.

2. Quantitative Analysis

Model results, shown in Table XV, were better than calibration results, in fact the total wetland area detected (for a WSI value of one) was 5.1% greater than that observed during calibration as shown in Figure 39. However the results of the model remained weak; in fact, with a WSI value greater or equal to nine the wetland area detected was 56.2% and the accuracy was 28.2%. As with calibration, the origin of the model error remained unidentified.

To answer some of the questions relating to error, three extra analyses were performed. The first assumed an incorrect calibration of the indexes could have potentially favored riverine wetlands over upland wetlands or vice versa. The second assumed the broad scale of the data supporting the indexes might have favored detection of large size wetlands. The third assumed there might be some difference between the ground truth data (1999 wetland locations) and 2004 wetland locations.

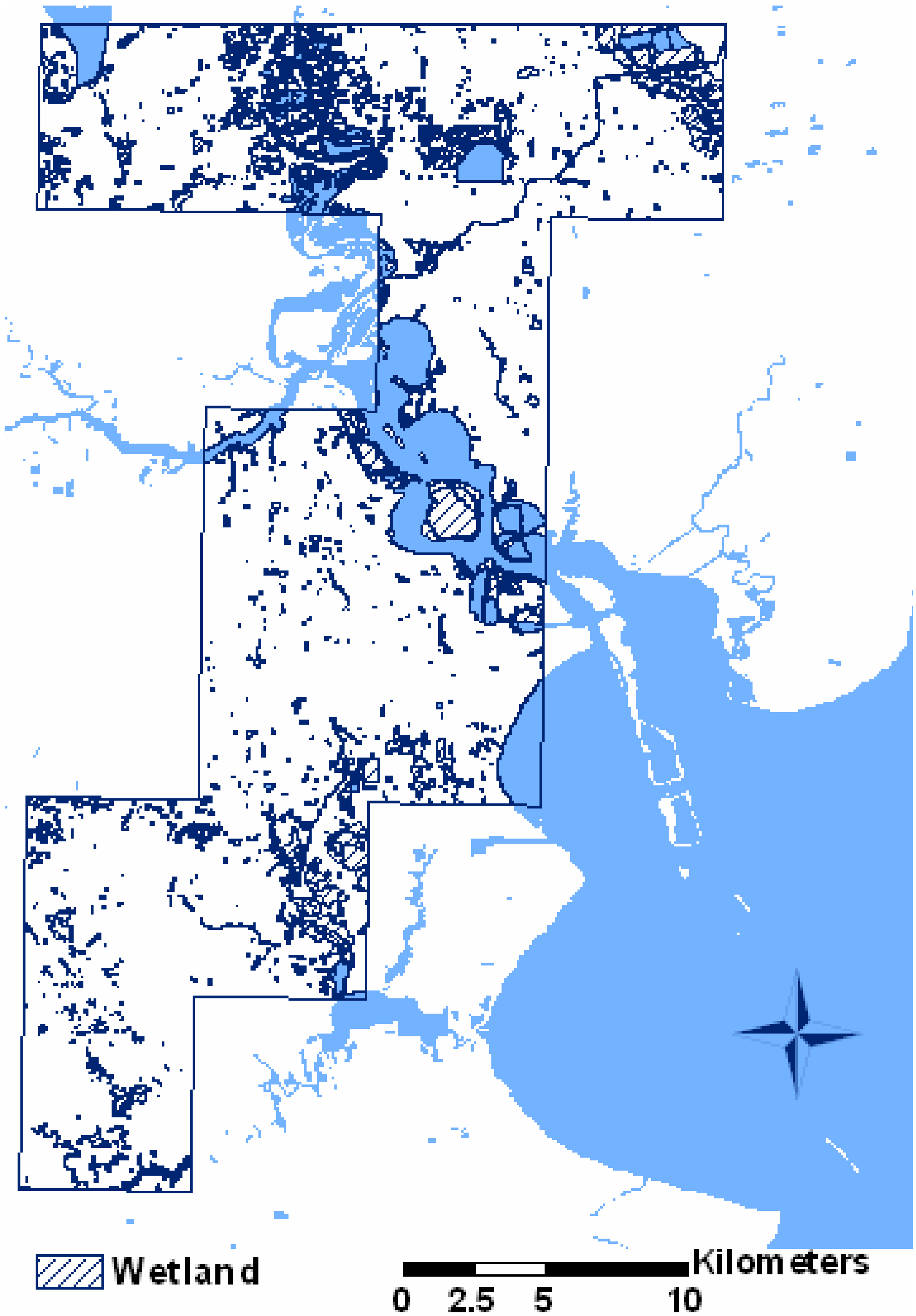

Fig. 38. Wetland locations identified in the SSL study used for the validation

Table XV. Validation results for model version three

WSI*	Wetland detected in %	Model accuracy in %
≥ 18	0.2	71.9
≥ 15	15.3	50.5
≥ 12	15.4	48.9
≥ 11	31.3	47.6
≥ 9	56.2	28.2
≥ 6	63.9	24.8
≥ 1	93.9	13.2

*WSI: Wetland Suitability Index

3. Alternative Analysis

a. Riverine Wetlands versus Upland Wetlands

To determine if the model favored riverine wetlands, all wetlands not located in a 500m buffer around any water landscape element were ignored from the ground truth data. Through this action, 20% of the wetland area was ignored. Then, as with the earlier analysis, results were obtained by zonal statistics. These results are shown in Table XVI. It appeared that riverine wetlands were not favored by the model. Riverine wetlands are better detected for WSI greater or equal to 12, however with a WSI greater or equal to nine the wetland area detected was 54.3% and the accuracy

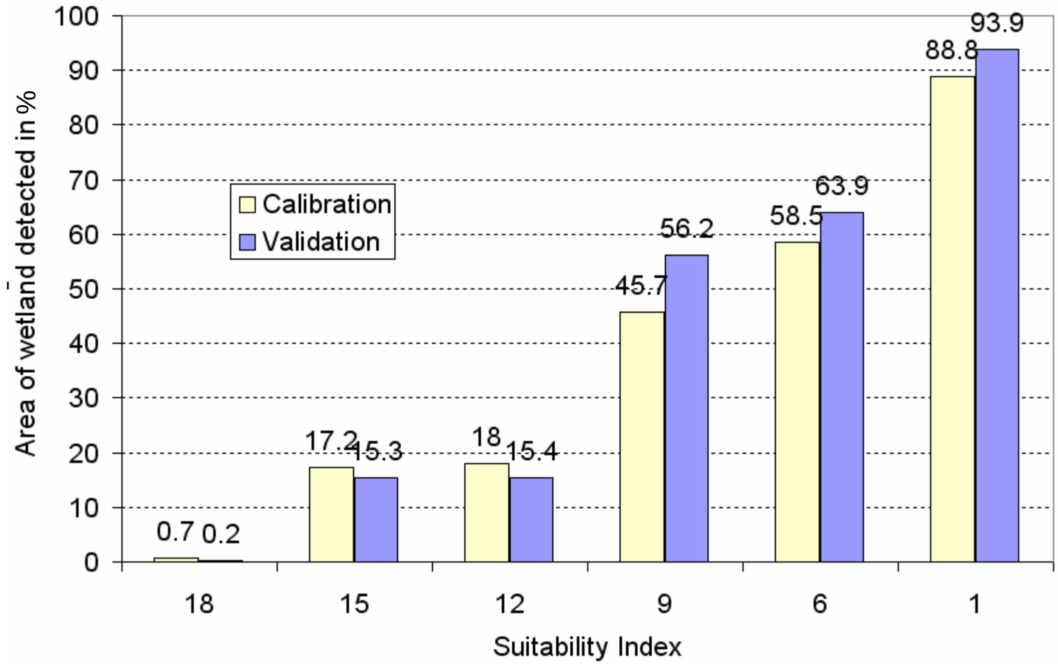

Fig. 39. Comparison of calibration results and validation results for model version three

was 23.4%. The difference between riverine wetland detection and general wetlands detection was about 1% on average. This difference was negligible.

b. Large Size Wetlands

To determine if the model favored large size wetlands, all wetlands with an area less than three hectares were ignored from the ground truth data. Through this action, 20% of the wetland area was ignored. Results were again obtained by zonal statistics (Table XVII). It appeared that large size wetlands were only very slightly favored by the model. The difference between large size wetland detection and general wetlands detection is about one percent on average. This difference is negligible.

Table XVI. Validation results for model version three for riverine wetlands

WSI*	Wetland detected in %	Model accuracy in %
≥ 18	0.6	72.0
≥ 15	17.6	42.5
≥ 12	17.5	34.5
≥ 9	57.2	23.4
≥ 6	65.3	22.2
≥ 1	96.1	10.8

*WSI: Wetland Suitability Index

Table XVII. Validation results for model version three for large size wetlands

WSI*	Wetland detected in %	Model accuracy in %
≥ 18	0.2	59.1
≥ 15	16.0	39.1
≥ 12	15.9	46.7
≥ 9	57.7	27.3
≥ 6	65.2	24.8
≥ 1	95.0	12.0

*WSI: Wetland Suitability Index

c. Ground Truth Data Accuracy

This case analysis was based on the idea that human activity expanded so quickly in the study area that 1999 NWI ground truth data were out of date. Thus, two quarter quads were chosen randomly from within the study area (Figure 40). For both quarter quads, some wetlands had been totally or partially urbanized. Therefore for both quarter quads the 1999 NWI wetland locations were updated manually using the 2004 DOQ. As before, results were obtained by zonal statistics. Tables XVIII and XIX compare those results using the 1999 NWI data as ground truth data to results using manually updated 2004 data as ground truth data. The difference between those two sets of results was always within two percent; therefore, the 1999 NWI data accuracy was considered to still be good.

Table XVIII. Evaluation of the 1999 NWI ground truth data for quarter quad 13

WSI*	Wetland detected in % of total wetland using 1999 NWI data	Wetland detected in % of total wetland using updated data	Difference in %
≥ 18	0.9	0.9	0.0
≥ 15	9.0	9.0	0.0
≥ 12	9.2	9.3	0.1
≥ 11	23.5	23.7	0.2
≥ 9	30.3	30.6	0.3
≥ 6	43.4	44.2	0.4
≥ 1	89.8	90.2	0.4

*WSI: Wetland Suitability Index

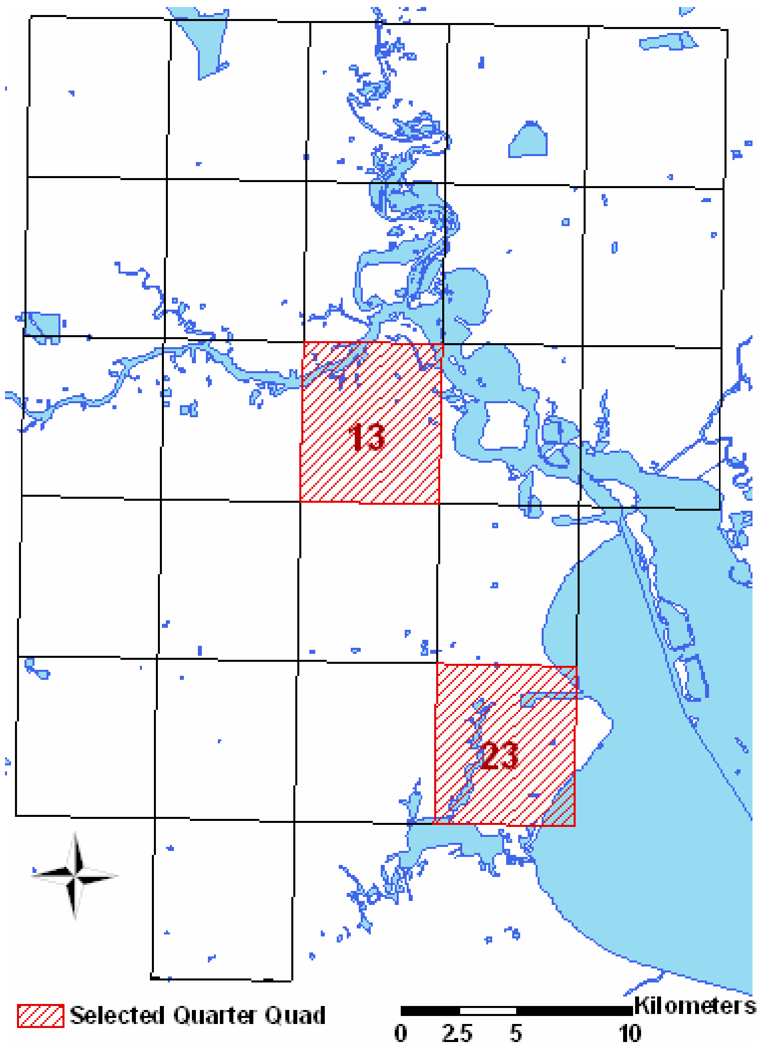

Fig. 40. Location of the quarter quads 13 and 23 randomly chosen for evaluation of the ground truth data

Table XIX. Evaluation of the 1999 NWI ground truth data for quarter quad 23

WSI*	Wetland detected in % of total wetland using 1999 NWI data	Wetland detected in % of total wetland using updated data	Difference in %
≥ 18	0.8	0.8	0.0
≥ 15	23.0	24.2	1.2
≥ 12	23.5	24.7	1.2
≥ 9	54.1	55.1	1.0
≥ 6	65.4	67.2	1.8
≥ 1	91.7	94.3	2.6

*WSI: Wetland Suitability Index

C. Results and Discussion

The validation results improved over calibration, however, the model performance remained weak. Results demonstrated that there was still too much error in the model. The alternative analyses just described indicated that the primary source of error was neither bad calibration of the model favoring riverine wetlands over upland wetlands, from the broad scale of the data constituting the indexes, nor from a lack of accuracy of the ground truth data. By elimination, there remained one more possibility: the error might have come from the inaccuracy of the broad scale data making up the indexes. This hypothesis is is supported by the scale of the data: 1:12,000 to 1:63,360 for SSURGO, 1:100,000 for NHD, 1:24,000 for NED, 1:65,000 for NWI. The problem of scale data was supported by many remote sensing oriented studies [10], [41], [11], [42]. Miyamoto et al. [10] used scale problems to justify the use of scale-independent remote sensing platforms such as kites and balloons to monitor wetlands. In these studies, only remote sensing data were used, not GIS data. Therefore, the focus was more on the scale of data itself rather than the inaccuracy coming from large scale data.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

A wetland suitability model was developed to delineate wetlands in the Houston area. This model combined GIS and remote sensing technologies. The data used for this study had to be as high scale as possible and easily accessible. This suitability model consisted of four submodels: hydrology, soil, vegetation and multi-attribute. The hydrology submodel was composed of two indexes: the drainage proximity index and the flow accumulation model. The drainage proximity index was made of a 450m buffer around all water bodies. The flow accumulation consisted of all areas of high flow accumulation with a slope less than 0.25%. The soil submodel was based on map units that presented a certain minimum percentage of hydric soil in the SSURGO database. The vegetation submodel was obtained by classification of Landsat images. The multi-attribute submodel was made of two indexes: the open water index and the urban index. The open water index cut all open water area from the study area. The urban index was determined using Landsat classification. Each one of the submodels generated a WSI. Those individual WSI values were summed to obtain a general WSI.

This model was calibrated using half of the study area. During calibration, the general model was evaluated as well as each index individually. Generally, the model showed a lack of sensitivity to changes. However, five model modifications were made during calibration. The first was to use a soil submodel with a weight 1.5 times the weight of the other submodels to improve the delineation of upland wetlands. The second was to have in the hydrology submodel a single 450m buffer for the drainage proximity index. The third was to use Landsat data for the vegetation submodel. The fourth was to use the urban index in the multi-attribute submodel. The fifth and final was to eliminate the slope index from the multi-attribute submodel. While all the models obtained weak results, model version three was selected for validation.

This model was validated using the remaining half of the study area. Validation results improved a bit compared to calibration results; however they remained weak. To determine the origin of this weakness, some alternative analyses were performed. Those analyses indicated that the model does not favor riverine wetlands over upland wetlands, nor large size wetlands. In addition, the model ground truth data were found to be in good agreement with current conditions. Results of the alternative analysis indicated that the weakness of the model must come from inaccuracy of the input data.

Results of this study indicate, 15 years after Tiner [9] and with high-power computers, that the lack of spatial accuracy was too high to perform correct wetland delineation using remote sensing and GIS technologies. However such results as 77.4% accuracy on the flow accumulation index, obtained using high accuracy LIDAR data, gave us reason to hope that data can and will improve in the near future. Such improvements could allow automatic wetland delineation. For future studies, it would be recommended to improve spatial accuracy for soil and vegetation data. Also spectral accuracy could be improved for vegetation data by using hyperspectral sensors.

REFERENCES

[1] T. E. Dahl, "Wetlands losses in the united states 1780's to 1980's," 1990. Version 16Jul97, U.S. Department of the Interior, Fish and Wildlife Service, Washington, DC Jamestown, ND: Northern Prairie Wildlife Research Center Home Page. [Online]. Available http://www.npwrc.usgs.gov/resource/othrdata/wetloss/wetloss.htm.

[2] U.S. Fish Wildlife Service (NWI), "National wetlands inventory," 2004. [Online]. Available http://wetlands.fws.gov/.

[3] Environmental Laboratory, "Corps of engineers wetland delineation manual," Tech. Rep. Y-87-1, U.S. Army Engineer Waterways Experiment Station, Vicksburg, MS, January 1987.

[4] L. M. Cowardin, V. Carter, F. C. Golet, and E. T. LaRoe, "Classification of wetlands and deepwater habitats of the united states," 1979. Version 04Dec98. [Online]. Available http://www.npwrc.usgs.gov.

[5] Natural Resources Conservation Service (NRCS), "Hydric soils," 2002. [Online]. Available http://soils.usda.gov/use/hydric/.

[6] U.S. Department of the Interior (USDI), "Riparian area management: Riparian-wetland soils," Tech. Rep. 1737-19, BLM/ST/ST-03/001+1737. 109 pp, Bureau of Land Management, Denver, CO, September 2003.

[7] J. R. Anderson, E. E. Hardy, J. T. Roach, and R. E. Witmer, "A land use and land cover classification system for use with remote sensor data." Geological Survey Professional Paper 964, 1976. [Online]. Available http://landcover.usgs.gov/pdf/anderson.pdf.

[8] J. Lyon, *Wetland Landscape Characterization; GIS, Remote Sensing and Image Analysis*. Chelsea, MI: Ann Harbor Press, 2001.

[9] R. W. Tiner, "Use of high-altitude aerial photography for inventorying forested wetlands in the united states," *For. Ecol. and Mgmt*, vol. 33-34, pp. 593–604, 1990.

[10] M. Miyamoto, K. Yoshino, T. Nagano, T. Ishida, and Y. Sato, "Use of balloon aerial photography for classification of kushiro wetland vegetation, northeastern japan," *Wetlands*, vol. 24, no. 3, pp. 701–710, 2004.

[11] K. R. Harvey and G. J. E. Hill, "Vegetation mapping of a tropical freshwater swamp in the northern territory, australia: a comparison of aerial photography, landsat tm and spot satellite imagery," *International Journal of Remote Sensing*, vol. 22, no. 15, pp. 2911–2925, 2001.

[12] J. R. Jensen, *Introductory Digital Image Processing*. Upper Saddle River, NJ: Prentice Hall, 1996.

[13] R. Kadmon and R. Harari-Kremer, "Long-term vegetation dynamics using digital processing of historical aerial photographs," *Remote Sensing of Environment*, vol. 68, no. 2, pp. 164–176, 1999.

[14] R. A. Anger, "A soil survey enhancement of landsat thematic mapper delineation of wetlands: a case study of barry county," Master's thesis, Western Michigan University, Kalamazoo, MI, 2003.

[15] Wetlands Reserve Program (WRP), "Hyperspectral imagery: A new tool for wetlands monitoring/analyses," Tech. Rep. WG-SW-2.3, May 1994.

[16] M. F. Gross and V. Klemas, “The use of airborne imaging spectrometry (ais) data to differentiate marsh vegetation,” *Remote Sensing of Environment*, vol. 19, pp. 97–103, 1986.

[17] T. Westra, K. C. Mertens, and R. R. D. Wulf, “Wavelet-based fusion of spot/vegetation and envisat/asar wide swath data for wetland mapping,” (Antwerp, Belgium), International SPOT/VEGETATION Users Conference., 2004.

[18] J. Noriega Rivera Rio and D. F. Lozano-Garcia, “Spatial filtering of radar data (radarsat) for wetlands (brackish marshes) classification,” *Remote Sensing of Environment*, vol. 73, no. 2, pp. 143–151, 2000.

[19] D. A. Roberts, M. Keller, and J. V. Soares, “Studies of land-cover, land-use, and biophysical properties of vegetation in the large scale biosphere atmosphere experiment in amaznia,” *Remote Sensing of Environment*, vol. 87, no. 4, pp. 377–388, 2003.

[20] E. P. Crist, “A tm tasseled cap equivalent transformation for reflectance factor data,” *Remote Sensing of Environment*, vol. 17, no. 3, pp. 301–306, 1985.

[21] S. A. Sader, D. Ahl, and W. Liou, “Accuracy of landsat-tm and gis rule-based methods for forest wetland classification in maine. remote sensing of environment,” *Remote Sensing of Environment*, vol. 53, no. 3, pp. 133–144, 1995.

[22] M. Bock, “Remote sensing and gis-based techniques for the classification and monitoring of biotopes, case examples for a wet grass- and moor land area in northern germany,” *Journal for Nature Conservation*, vol. 11, pp. 145–155, 2003.

[23] E. W. I. Ramsey, D. K. Chappell, D. M. Jacobs, S. K. Sapkota, and D. G.

Baldwin, "Resource management of forested wetlands: Hurricane impact and recovery mapped by combining landsat tm and noaa avhrr data," *Photogrammetric Engineering and Remote Sensing*, vol. 64, no. 7, pp. 733–738, 1998.

[24] O. Earley, "Modeling wetland loss in the richmond catchment: An example of historical vegetation modeling," 4th International Conference on Integrating GIS and Environmental Modeling: Problems, Prospects and Research Needs, Banff, Alberta, Canada, September 2000. [Online]. Available http://www.colorado.edu/Research/cires/banff/pubpapers/250/.

[25] Texas Coastal Wetlands, "Coastal wetlands: What are they and why should we care?," 2003. [Online]. Available http://www.texaswetlands.org/index.htm.

[26] Environmental Systems Research Institute Inc. (ESRI), "Gis and mapping software," 2005. [Online]. Available http://www.esri.com.

[27] Natural Resources Conservation Service (NRCS), "Soil data mart," 2005. [Online]. Available http://soildatamart.nrcs.usda.gov/.

[28] "Texasview remote sensing consortium for texas," 2005. [Online]. Available http://www.texasview.org/.

[29] U.S. Geological Survey (USGS), "National center for earth resources observation and science (eros)," 2005. [Online]. Available http://edc.usgs.gov/.

[30] U.S. Geological Survey (USGS), "National hydrography dataset," 2005. [Online]. Available http://nhd.usgs.gov/.

[31] Spatial Sciences Laboratory, "Evaluating the loss of wetlands in the galveston bay area." Unpublished project report to Texas Cooperative Extension (TCE), August 2002.

[32] Geospatial Technology Program (GTP), "Using digital soils data," 2005. [Online]. Available http://lal.cas.psu.edu/software/tutorials/soils/soils.html.

[33] M. E. Hodgson, J. R. Jensen, H. E. Mackey, and M. C. Coulter, "Remote sensing of wetland habitat: A wood stork example," *Photogrammetric Engineering and Remote Sensing*, vol. 53, pp. 1075–1080, 1987.

[34] National Oceanic Atmospheric Administration (NOAA), 2005. [Online]. Available http://www.noaa.gov.

[35] U.S. Geological Survey (USGS), "Enhanced thematic mapper plus (etm+)," 2005. [Online]. Available http://edc.usgs.gov/products/satellite/landsat7.html.

[36] ENVI, "Envi user's guide," 2004. ENVI Version 4.1 Research Systems, Inc.

[37] J. R. Jensen, *Remote Sensing of the Environment: An Earth Resource Perspective*. Upper Saddle River, NJ: Prentice Hall, 2000.

[38] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the great plains with erts," in *Third Earth Resources Technology Satellite-1 Symposium*, 1974. NASA SP-351 I: 309-317.

[39] M. A. Hardisky, V. Klemas, and M. Smart, "The influence of soil salinity, growth form, and leaf moisture on the spectralradiance of spartina alterniflora canopies," *Photogrammetric Engineering and Remote Sensing*, vol. 49, pp. 77–83, 1983.

[40] R. J. Kauth and G. S. Thomas, "The tasseled cap - a graphic description of the spectral - temporal development of agricultural crops as seen by landsat," in *Proceedings of the Symposium on Machine Processing of Remotely Sensed Data*, vol. 4, pp. 41–51, 1976.

[41] J. Barrette, P. August, and F. Golet, "Accuracy assessment of wetland boundary delineation using aerial photography and digital orthophotography," *Photogrammetric Engineering and Remote Sensing*, vol. 66, pp. 409–416, 2000.

[42] R. M. Johnston and M. M. Barson, "Remote sensing of australian wetlands: an evaluation of landsat tm data for inventory and classification," *Australian Journal of Marine and Freshwater Research*, vol. 44, pp. 223–232, 1993.

VITA

Julie Villeneuve was born in La Tronche (France). She lived all her childhood in Normandy. She received her French engineering diploma from the Ecole Polytechnique Feminine in 2000. She enrolled for a Master of Science in Biological and Agricultural Engineering at Texas A&M University in the spring of 2003. She studied under the direction of Dr Ann Kenimer. She earned her M.S in December 2005. She can be contacted at:

c/o Dr. Ann Kenimer

Dept of Biological and Agricultural Engineering

Texas A&M University

2117 TAMU

College Station, TX 77843-2117