2016

Remote Sensing of Coastal Wetlands: Long term vegetation stress assessment and data enhancement technique

Subrina Tahsin
University of Central Florida

REMOTE SENSING OF COASTAL WETLANDS: LONG TERM VEGETATION STRESS ASSESSMENT AND DATA ENHANCEMENT TECHNIQUE

by

SUBRINA TAHSIN
BURP, Bangladesh University of Engineering and Technology, 2009
MSc, Florida International University, 2014

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2016

Major Professor: Stephen C. Medeiros

ABSTRACT

Apalachicola Bay in the Florida panhandle is home to a rich variety of salt water and freshwater wetlands but unfortunately is also subject to a wide range of hydrologic extreme events. Extreme hydrologic events such as hurricanes and droughts continuously threaten the area. The impact of hurricane and drought on both fresh and salt water wetlands was investigated over the time period from 2000 to 2015 in Apalachicola Bay using spatio-temporal changes in the Landsat based NDVI. Results indicate that salt water wetlands were more resilient than fresh water wetlands. Results also suggest that in response to hurricanes, the coastal wetlands took almost a year to recover while recovery following a drought period was observed after only a month. This analysis was successful and provided excellent insights into coastal wetland health. Such long term study is heavily dependent on optical sensor that is subject to data loss due to cloud coverage. Therefore, a novel method is proposed and demonstrated to recover the information contaminated by cloud.

Cloud contamination is a hindrance to long-term environmental assessment using information derived from satellite imagery that retrieve data from visible and infrared spectral ranges. Normalized Difference Vegetation Index (NDVI) is a widely used index to monitor vegetation and land use change. NDVI can be retrieved from publicly available data repositories of optical sensors such as Landsat, Moderate Resolution Imaging Spectro-radiometer (MODIS) and several commercial satellites. Landsat has an ongoing high resolution NDVI record starting from 1984. Unfortunately, the time series NDVI data suffers from the cloud contamination issue. Though simple to complex computational methods for data interpolation have been applied to recover

cloudy data, all the techniques are subject to many limitations. In this paper, a novel Optical Cloud Pixel Recovery (OCPR) method is proposed to repair cloudy pixels from the time-space-spectrum continuum with the aid of a machine learning tool, namely random forest (RF) trained and tested utilizing multi-parameter hydrologic data. The RF based OCPR model was compared with a simple linear regression (LR) based OCPR model to understand the potential of the model. A case study in Apalachicola Bay is presented to evaluate the performance of OCPR to repair cloudy NDVI reflectance for two specific dates. The RF based OCPR method achieves a root mean squared error of 0.0475 sr^{-1} between predicted and observed NDVI reflectance values. The LR based OCPR method achieves a root mean squared error of 0.1257 sr^{-1}. Findings suggested that the RF based OCPR method is effective to repair cloudy values and provide continuous and quantitatively reliable imagery for further analysis in environmental applications.

ACKNOWLEDGMENTS

I would like to express my outstanding gratitude to my advisors, Dr. Stephen C. Medeiros and Dr. Arvind Singh. I would like to thank Dr. Mayo for her involvement in my thesis work. Without their patience, guidance, and support this research would not have been possible, nor would I have been capable of completing it. I would like to thank my lab mate, Milad Hooshyar for his support in my research. A special thank you to my research colleagues who have helped me achieve all that I have thus far with their help and support: Karim Alizad, Yin Tang, Han Xiao, Daljit Sidhu, Marwan Kheimi. Finally, thanks to my parents, siblings, husband and in-laws for all of the love and encouragement they have given me over the years in furthering my education and providing a mountain of support that has allowed me to accomplish this goal.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

1.1 Application of Vegetation Index in Wetland Stress Analysis

Advances in remote sensing applications and data analysis systems are bringing cutting edge research techniques to real world practice and enabling cost efficient, quantitative biophysical analysis more accessible. For example, wetland extent mapping, leaf area index, canopy density and closure, etc., are making the assessment of biophysical parameters doable at regional scales. These great resources also present new challenges. Staffs responsible for environmental monitoring as well as ecosystem modelers are handling large uncertainties in data as a result of weather, environment and vegetation. Having comprehensive and up to date information is crucial to optimize wetland and forest management throughout the season especially before and after extreme natural hazards. In particular, vegetation index maps consist of detailed imagery that abstracts a measure of the green vegetation present in their study area. Time series analyses of the trend of greenness in vegetation can play a crucial role in identifying vegetation/wetland stress and relate the impact of hydrologic events. Long term impacts of extreme events on the ecosystem can range from small to massive, depending on the severity and duration of the event. A crucial component to time series analyses is establishing baseline characteristics of the study area so that changes can be identified.

The Apalachicola region in the Florida Panhandle has a very large estuarine ecosystem comprised of both salt and freshwater wetlands. The research presented focuses on the resilience

of both kinds of wetlands in this area, and uses the region as a tested for machine learning based data enhancement technique.

1.2 Normalized Difference Vegetation Index (NDVI)

Normalized Difference Vegetation Index (NDVI) derived from Landsat has excellent spatial resolution compared to other publicly available satellite imagery. Landsat NDVI has many environmental applications including the ability to analyze changes in land use, transformation of urban heat islands, and impacts of extreme events. Landsat NDVI carries valuable information regarding land surface properties for modeling terrestrial ecosystems on the global, continental, and regional scales, since 1984. Such a long time record is unique in the satellite remote sensing community. Theoretically, NDVI, calculated from a normalized transform of the near-infrared (NIR) and red reflectance ratio, is an index used to characterize the reflective and absorptive features of vegetation in the red and NIR portions of the electromagnetic spectrum. However, there are almost always disturbances in these time series, caused by cloud contamination, atmospheric variability, and bi-directional effects. These disturbances greatly affect the monitoring of land cover and terrestrial ecosystems and show up as undesirable noise. Although the most often-used NDVI data sets are the post-processed 16-day Maximum Value Composite (MVC) products, they still include such noise. For this reason, a number of methods for reducing noise and constructing high-quality NDVI time series data sets for further analysis have been formulated, applied, and evaluated.

1.3 Cloud Concerns in Optical Sensor Data

Predicting missing data is a challenge in any analysis of time series data derived from satellite imagery. Landsat NDVI is not without the same problem. Missing data is inevitable due to the presence of clouds, especially in warm coastal regions where water evaporation and frequent storms combine to produce cloud coverage. Cloud coverage hinders scientific research that depends on optical remote sensing imagery. Moreover, observations are often incomplete because of sensor failure or outliers causing anomalous data. Therefore, it is very important to carry out research on the filtering and gap filling of time series satellite images.

1.4 Scope of the Study

The current study performed a long-term wetland stress analysis using a 30 year NDVI time series. It analyzed the impact of extreme events on the ecosystem that can range from massive to small. Before analyzing these event based impacts, significant pre-processing and reclassification was done in order to make the data more manageable. Once the data were prepared, two research questions were addressed: Which wetlands among SWW, FFW, FEW are more resilient to hurricane and drought? How long it take to recover the wetlands after an extreme event? Can Landsat pixels obscured by clouds be recovered? In an attempt to compare the resiliency of each wetland types, probability density function (PDF) for each wetland was developed. Also the recovery time after an event was computed by the time gap to return from an anomalous NDVI range to s regular NDVI range. All computation was derived with regard to seasonality removed time series. Seasonality was defined in the current study as monthly mean data over the whole time series.

After observing the limitations imposed on the analysis by cloudy pixels, the study further proposed a novel approach using machine learning techniques based on multi-parameter time series data to repair missing NDVI reflectance values. The unique and novel method was named Optical Cloud Pixel Recovery (OCPR). High spatio-temporal resolution raster based temperature, precipitation, and spatial locations along with water levels from a nearby tide gage and corresponding month were selected as the feature vector (predictor) components associated with NDVI (label). To reconstruct cloud contaminated pixel values from the time-space-spectrum continuum, the random forest (RF) machine learning tool was utilized. Approximately 30 years of time series data were collected for the training and testing of the OCPR model. All of these variables contained periods of missing data that were filtered out of the training and test data. RF is used to model the data distribution which is adapted to handle missing values. The RF, along with mean only and linear regression models, was assessed using the root mean square error (RMSE) between the simulated and the observed NDVI values in the test data set. The result is a deep, functioning model that can be used on Landsat as well as other satellite images worldwide, subject to further refinements and testing.

CHAPTER 2: RESILIENCE OF COASTAL WETLANDS TO EXTREME HYDROLOGIC FLUCTUATION

2.1 Background and Introduction

Hurricanes and droughts are climatically-instigated pulse events that cause enormous ecosystem perturbation. Such pulses occur frequently and the corresponding change in the ecosystem can persist for varying lengths of time [*Yang et al.*, 2008]. Ecosystem resilience can be understood by investigating the effects of these pulse-induced changes, including their persistence through time [*Switzer et al.*, 2006]. The impacts of hurricanes or droughts on coastal wetlands can vary depending on the wetland type, i.e.,- freshwater forested wetland (FFW), freshwater emergent wetland (FEW) and saltwater wetland (SWW) ecosystems [*Mo et al.*, 2015].

Hurricanes cause physical damage to wetlands by high velocity winds and flows as well as salt water flood submergence [*Stanturf et al.*, 2007]. The effect of even a short duration of saltwater storm surge inundation can have devastating effect on FWW [*Conner and Ozalpl*, 2002; *Stanturf et al.*, 2007], while SWW are more tolerant of elevated salinity. Depending on other hydrological factors such as rainfall [*Huang et al.*, 2015a] and groundwater recharge, the salinity levels in the surface water can remain elevated for months following a hurricane [*Steyer et al.*, 2007], with freshening not occurring for as long as one year [*Chabreck and Palmisano*, 1973]. Although hurricanes bring storm surge and large amounts of rainfall in a relatively short time period, droughts are another natural hazard that can last from months to years [*Florida Climate Center*, 2014].

Remote sensing can be used to detect and track the wetland dynamics at the regional scale. Multiple satellite sensors such as Landsat [*Han et al.*, 2015; *Tian et al.*, 2015], Formosat

[*Tian et al.*, 2015], MODIS [*Landmann et al.*, 2013] and AVHRR [*Ramsey III et al.*, 1997] can provide data for this application; these data are often processed into vegetation indices. These vegetation indices can be obtained from sensor reflectance data in spectral bands that are responsive to vegetation characteristics. The most well known is the Normalized Difference Vegetation Index (NDVI) and it is frequently used to identify and characterize vegetated areas. It has been shown to be highly correlated with parameters associated with plant health and productivity such as vegetation density and cover [*Wiegand et al.*, 1974; *Ormsby et al.*, 1987] vegetation dynamics over time [*Wellens*, 1997]; vegetation classification [*Evans and Geerken*, 2006] and many other related aspects [*Wang and Tenhunen*, 2004; *Pettorelli et al.*, 2011]. For example, *Ramsey III et al.* [1997] analyzed forest damage caused by Hurricane Andrew in 1992, using NDVI derived from Advanced Very High Resolution Radiometer (AVHRR) multi-temporal images. Their main finding was the utility of regionally averaged NDVI change as an indicator of damage severity. *Wang* [2012] also identified severe mangrove forest damage after Hurricanes Katrina and Wilma that took two to three years to recover. Numerous other post hurricane studies also focused on damage to coastal mangrove forests owing to hurricane winds [*Middleton*, 2009, 2016] and storm surge [*Conner and Ozalpl*, 2002; *Stanturf et al.*, 2007].

Apalachicola in the Florida Panhandle is located in a high risk hurricane zone [*Passeri et al.*, 2015]. While hurricanes bring storm surge and large amounts of rainfall in a relatively short time period, droughts are another natural hazard that can last from months to years [*Florida Climate Center*, 2014]. The lower marshes of the Apalachicola River are primarily composed of natural wetlands area with few anthropogenic disturbances. The minimal human influence on

wetlands for this region make it an ideal candidate for assessing the impacts of extreme events on wetland recovery [*Matlock*, 2009].

While the effects of wind, surge duration and salinity on wetland ecosystem have been investigated previously [*Ramsey III et al.*, 1997; *Wang*, 2012][*Ramsey III et al.*, 1997; *Conner and Ozalpl*, 2002; *Wang et al.*, 2010; *Wang*, 2012], the differences between effects on FWW and SWW remains largely uninvestigated. A very limited number of studies have attempted to untangle the relative impacts of drought on SWW and FWW ecosystems [*Ji and Peters*, 2003; *Lloret et al.*, 2007]. To the above aim, using Appalachicola Bay as the study area, the purpose of this paper is to investigate the stress on SWW, FFW and FEW due to hurricanes and droughts using an empirical vegetation index

2.2 Study Area and Extreme Events

2.2.1 Study Area: Apalachicola Bay

Apalachicola Bay in the Florida Panhandle is located in a high risk hurricane zone and has recived significant attention in the past few decades. Apalachicola Bay is home to rich natural resources including oyster beds and a vast span of marshes. Apalachicola oysters is accounted for 90% of Florida's oyster production that contributes to the economics of Apalachicola Bay [*Huang and Jones*, 2001] . The lower marshes of the Apalachicola River are primarily composed of natural wetlands area with few anthropogenic disturbances. The minimal human influence on wetlands for this region make it an ideal candidate for assessing the impacts of extreme events on wetland recovery.

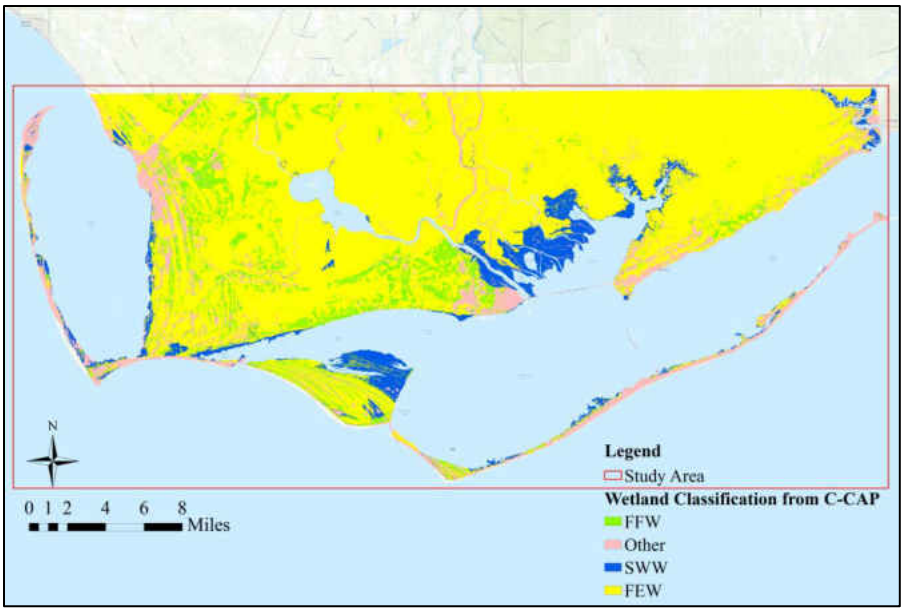


Figure 2.1. Apalachicola Bay and lower river marshes, derived from C-CAP wetland classification (2006), used as the study area. Freshwater Forested Wetland (FFW) represents 8.36% of the total study area, Freshwater Emergent Wetland (FEW) represents 73.77%, whereas Saltwater Wetland (SWW) represents 6.83%. The majority of "other" is agriculture and cropland

The National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (C-CAP) classified wetlands along the eastern seaboard and Gulf coasts of the United States. C-CAP is considered a reliable, integrated digital database that enables researchers to track development in coastal regions [*Klemas et al.*, 1993]. This study used the C-CAP classification as basis and resampled it to three wetland types - SWW, FFW and FEW. The resampling of SWW included all estuarine forested, emergent and scrub wetlands; FFW represented freshwater forested wetlands; and FEW included freshwater emergent and scrub wetlands [*Klemas et al.*, 1993]. An elevated salinity gradient (> 0.5%) charecterized SWW, while low salinity gradient (< 0.5%) charecterized FWW. The red box in Figure 2.1 is the study domain for current study.

2.2.2 Hurricane and Drought Years

A number of significant hurricanes impacted the Applachicola Bay from 2000 to 2015. The Apalachicola is a micro-tidal estuary with a tide range ~ 1m [*Passeri et al.*, 2015]. Heavy rainfall in 2003 caused flooding in several counties adjacent to the study area that lead to declaring local state of emergency. Due to the localized intense rainfall, local rivers swelled to reach severe flood stages [*The Florida State Emergency Response Team*, 2003]. Hurricane Frances made its second landfall near St. Marks, FL in August 2004, after crossing the Florida peninsula and weakening to a tropical storm; however, storm surge impacts were still significant along the Florida Panhandle [*National hurricane center*, 2004]. Tropical Storm Bonnie and Hurricane Ivan also made landfalls in 2004 to the west of Apalachicola Bay with Ivan causing up to 3.65 meters of surge along the coast in Apalachicola Bay [*Edmiston et al.*, 2008]. In July 2005, Hurricane Dennis caused a 2.74 m surge on the barrier islands protecting Apalachicola Bay [*Beven*, 2005]. Tropical Storm Claudette hit the Florida Panhandle in 2009; although the intensity of this storm was comparatively less than those in preceding years, Apalachicola Bay received significant surge as it was positioned in the northeast quadrant of the storm. The eastern half of a hurricane, and the northeast quadrant in particular, contains the most intense winds and therefore storm surge due to the wind speed and the hurricane's forward velocity acting in the same direction and compounding each other. Hurricane Isaac in 2012 caused 1.0 m of surge in Apalacicola Bay. In terms of the other hazard considered in this study, a significant drought occurred in the Apalachicola Bay River watershed from May 2011 to June 2012. Lastly and most recently, Hurricane Andrea made landfall in Florida's big bend region in 2013. Storm surge

inundation level in Apalachicola Bay during that time was as much as 1.7 m above mean sea level (MSL) [*Beven*, 2005].

2.3 Data Collection

The Landsat-derived composite multiband vegetation index imagery (procesed to NDVI) were obtained from USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA). These images have a ground resolution of 30 m at the mean solar zenith angle of each 16-day period [*Zhu and Woodcock*, 2012] and were collected from Landsat 5, 7 and 8 over the time period from 2000 to 2015. Like other multi-spectral satellites, Landsat data are contaminated by clouds and cloud shadows. ESPA provides standalone "cfmask" layers that account for atmospheric gases, aerosols, and clouds (including thin cirrus clouds). Landsat 7 imagery required additional processing for stripe removal due to the documented failure of the scan line corrector (SLC) in 2003 [*She et al.*, 2015]. After removing images that were unusable due to cloud cover, 134 months of data were available for use out of the entire 192 months over the study time period. From the Landsat imagery, the NIR and RED spectral bands were used to compute NDVI according to equation 1.

$$\boldsymbol{NDVI} = \frac{\boldsymbol{NIR-RED}}{\boldsymbol{NIR+RED}} \quad (1)$$

NDVI values range from 0.0 (i.e. no vegetation) to 1.0 (vigorous, dense green biomass) [*Lane et al.*, 2014].

2.3.1 Data Pre-processing

NDVI time-series data have been used in the past to detect long term land-use / land-cover (LULC) changes [*Pirotti et al.*, 2014]. However, the utility of NDVI to detect vegetation stress in wetlands is often limited by poor quality data resulting from atmospheric and other effects. Studies typically assume that the NDVI time-series follows annual cycles of growth and decline of vegetation, and that clouds or poor atmospheric conditions usually depress observed NDVI values [*Chen et al.*, 2004]. Previous research has applied methods to construct NDVI time-series by filling gaps and smoothing out noise in the time-series data. Overcoming missing or poor quality NDVI data has been primarily accomplished through spatial [*Myneni et al.*, 1998; *Lim and Kafatos*, 2002; *Potter et al.*, 2003] or temporal [*Justice et al.*, 1985; *DeFries et al.*, 1995; *Loveland et al.*, 2000] averaging / filtering. A particular method that has proven to be acceptable for constructing a high-quality NDVI time-series is based on the Savitzky–Golay (S-G) Filter [*Savitzky and Golay*, 1964; *Chen et al.*, 2004; *Luo et al.*, 2005]. This study used an empirical S-G filter for both interpolating missing data and discounting negative and anomalously low NDVI values.

2.3.2 Data Analysis

Seasonality describes the phenomenological dynamics of terrestrial ecosystems that reflect the response of the atmosphere to inter and intra-annual dynamics of the hydrologic regimes and Earth's climate [*Myneni et al.*, 1997; *White et al.*, 1997; *Schwartz*, 1999]. In other words, seasonality of vegetation refers to the regular periodic change that a terrestrial ecosystem

experiences. For example, coastal Louisiana wetlands showed distinct seasonality and all marshes peaked within one month from late July to mid-August [*Mo et al.*, 2015]. A similar phonological cycle was observed in other ecosystems [*Chidumayo*, 2001; *Zhang et al.*, 2003; *Mo et al.*, 2015]. The seasonality of Apalachicola Bay wetlands was filtered out and time-series were plotted for the three wetland types. Seasonally adjusted NDVI is represented by $\widetilde{NDVI}_{IJ}$ and computed using equation 2.

$$NDVI_{IJ} - NDVI_I' = \widetilde{NDVI}_{IJ} \quad (2)$$

Here, $NDVI_{IJ}$ is NDVI in month *I* of year *J*; $NDVI_I'$ is the mean NDVI over the time-series for month *I*. The seasonally adjusted time-series were used to indicate abnormal NDVI peaks or drops.

2.4 Results

2.4.1 NDVI Variability under Extreme Natural Events

Flooding associated with storm surge was observed as a result of the major hurricanes Dennis, Frances, Claudette and Andrea that made landfalls in the last decade in Apalachicola Bay. Figure 2.2 shows NDVI variability during a regular year (a), after Hurricane Frances (b), after Hurricane Dennis (c) and during drought (d). While 2002 was a regular year, 2004 and 2005 had significant storm surge from Hurricane Frances and Dennis and 2012 was classified as a drought year [*Hatter*, 2015]. The mean annual NDVI values in the study area were found to be 0.52, 0.49, 0.34 and 0.41 in 2002, 2004, 2005 and 2012, respectively. The aftermath of each hurricanes mentioned above was observed for a year from the day it made landfall. Recall that

low NDVI values represent wetland with less vigor, a high NDVI represents wetlands with more vigor. 2004 and especially 2005 showed the most stress (loss of vigor) for wetlands due to repeated hurricane strikes. Drought also impacted the average NDVI range in 2012-2013. Any sharp deviation from the average range can indicate the wetland stress. To quantify the wetland stress, box plots of yearly NDVI were constructed over the study period.

Figure 2.3 shows the box plots of yearly NDVI for each wetland type. As can be seen from Figure 2.4 . SWW has lower range of NDVI (median value 0.37) than FFW (median value 0.56) and FEW (median value 0.51) during the study period. The largest reduction occurred in 2005 following the 2004-2005 back to back hurricane landfalls. NDVI was found to be the lowest during that time and the values were 0.15 for SWW and 0.20 for both FFW and FEW. Other significant reductions were observed in 2009 after Hurricane Claudette and in 2013 during the drought period. The significant NDVI reduction of FFW can be attributed towards partial to total uproot of the wetland to major foliage damage of SWW.

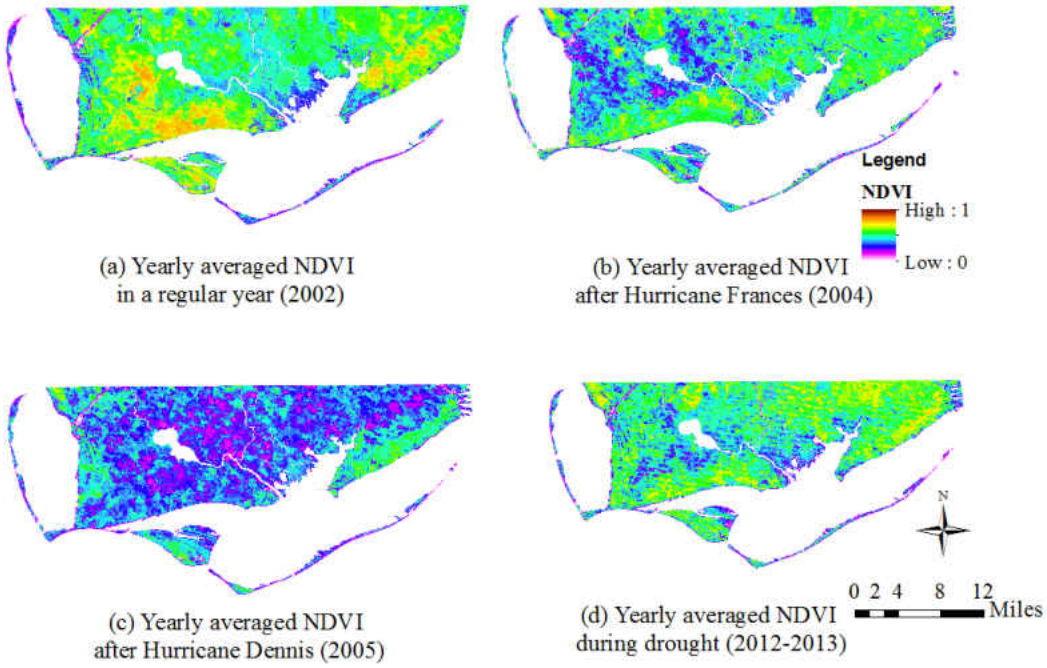


Figure 2.2. Yearly averaged NDVI values at the Apalachicola Bay for a regular year (a), and for different extreme hydrologic events (b-d).

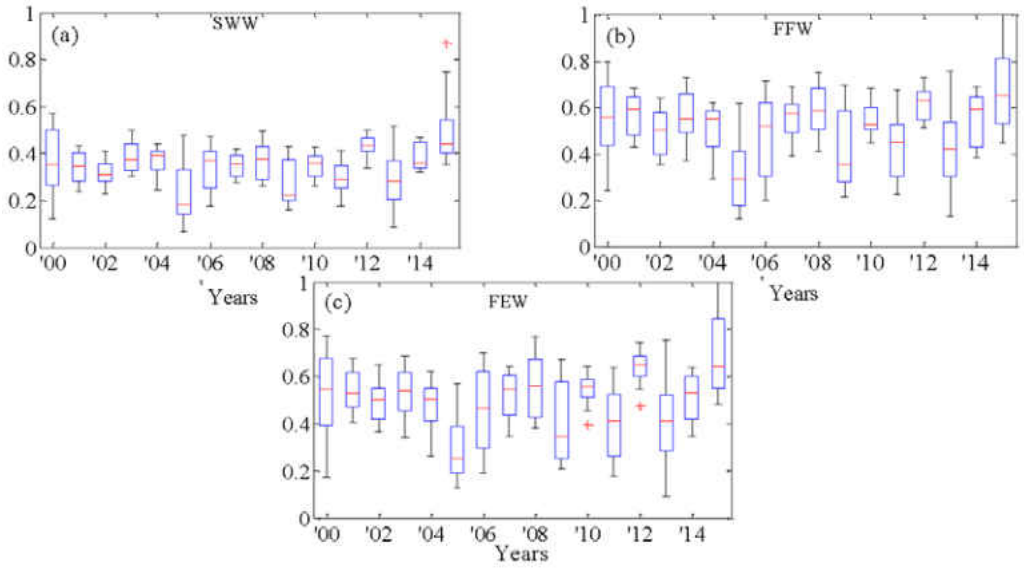


Figure 2.3. Boxplots of the computed NDVI for the Freshwater Forested Wetland (FFW), Freshwater Emergent Wetland (FEW), and Saltwater Wetland (SWW) of the Apalachicola Bay.

Vertical line (black line) in each box plots indicates median demarcating 50% data either above or below the median. Upper and lower quartiles of the box refer to 25% and 75% of the data from the median, whereas upper and lower whiskers indicate maximum and minimum values, respectively, excluding outliers. + Symbols indicate outliers in the data.

2.4.2 SWW vs. FWW Resilience to Extreme Natural Events

Figure 2.4 shows the comparison between the PDFs of three types of wetlands NDVI for SWW, FFW and FEW after the extreme natural events over the study period. As discussed in the previous section, SWW had the lowest average NDVI value over the study period. The range of the NDVI PDF for SWW was 0.5 compared to 0.70 for FFW and 0.78 for FEW. The narrow PDF for SWW indicated more stability over the study period which we interpret as an indicator of resilience. Therefore, SWW demonstrated more resilience to hydrologic hazards over the study period.

Figure 2.5 represents the time-series plot of the three types of wetlands in Apalachicola Bay Florida after filtering out the seasonality. Any sharp drop or peak in the time-series (Figure 2.5) can be attributed towards an extreme event. With this hypothesis, Figure 2.5 reveals major NDVI reductions associated with major hurricane years 2004 and 2005. Hurricane Claudette is associated with the second highest NDVI reduction in 2009 since the second highest drop in Figure 2.5 is right after 2009, another significant reduction was observed during the 2012-2013 period when a drought occurred concurrently with Hurricane Andrea.

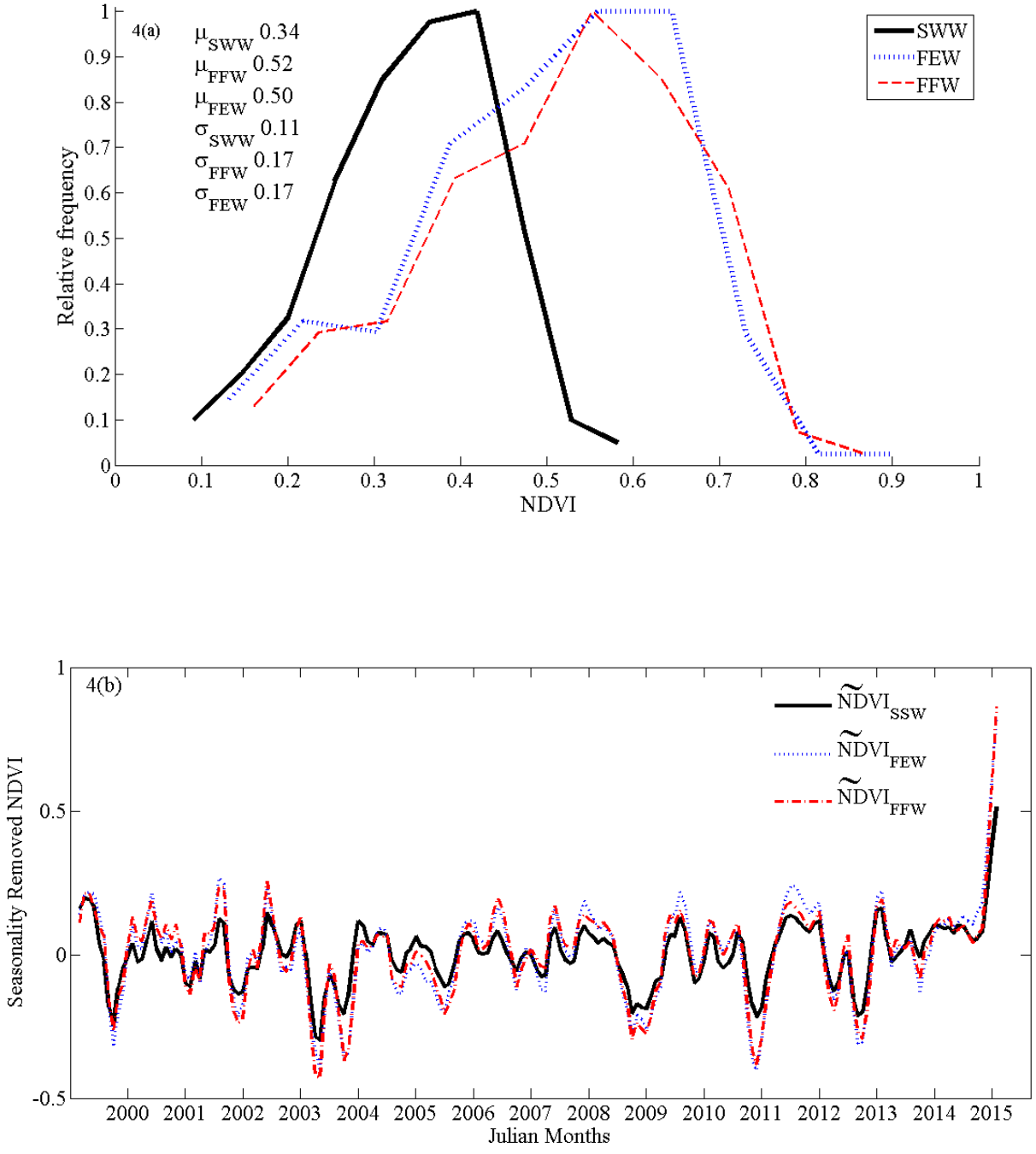


Figure 2.4. Probability density function (PDF) of the NDVI (A), and seasonality removed time-series of NDVI (B) For the three different wetland types used in this study. The box inset in (A) indicates the mean and the standard deviation of the NDVI.

2.4.3 Recovery after Hurricanes and Droughts

The ability of a wetland to recover after a hydrologic event depends on both the type of event and the type of wetland. SWW showed distinctly different responses compared to FWW for both hurricane and drought events. Figure 2.5 shows the differences in seasonally adjusted NDVI between FFW and SWW (Figure 2.5-a) and FEW and SWW (Figure 2.5-b) over the 16 year study period. The period from 2004 to 2006 shows a negative difference indicating that the SWW NDVI was greater than that of FWW. In regular years (i.e. with no extreme hydrologic event), the FWW have consistently higher NDVI difference than SWW; therefore a positive difference is considered normal here. Along that line, a negative difference indicates an anomaly (or anomalies) and wetland damage by submergence, flattening or uproot/extraction of the wetland vegetation by the hurricanes and their associated storm surge. The situation was exacerbated as a result of the repeated hurricanes in 2004 and 2005. Figure 2.5 also indicates that after hurricane years, the anomalous negative differences revert back to regular positive difference after approximately one year. However, in 2012 the difference remains negative during most of the drought (lasting approximately six months, indicated on Figure 2.5 as a yellow shaded region) but reverts to positive towards the end.

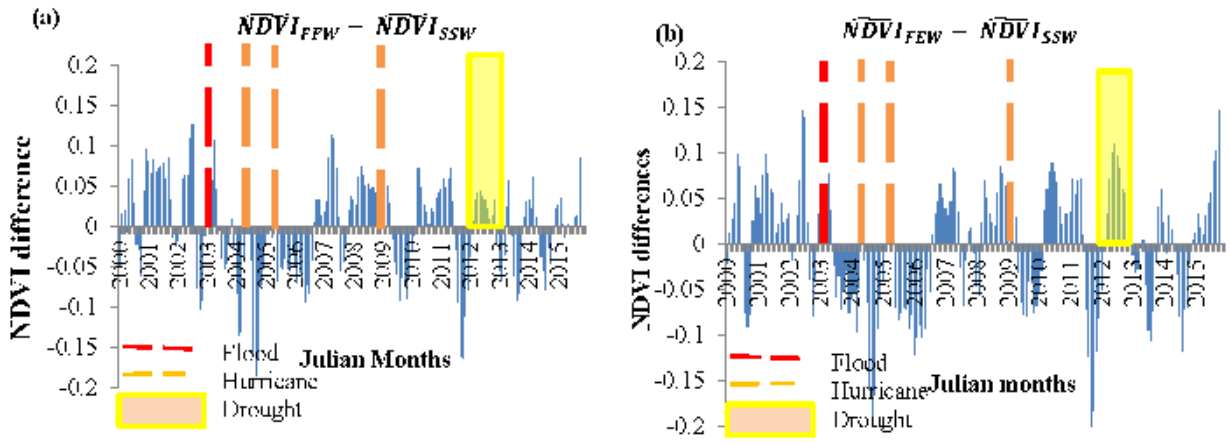


Figure 2.5. Time-series of seasonality-removed NDVI (shown in Figure 4b) differences between FFW and SWW (a), and FEW and SWW (b).

2.5 Concluding remarks

Hydrologic disturbances like hurricanes and droughts cause variable levels of damage indifferent wetland ecosystems. Exploratory analysis of the NDVI showed that Hurricanes Frances, Dennis, Claudette, and Isaac combined with a drought and Tropical Storm Andrea in 2004, 2005, 2009, 2012 and 2013 caused stresses in Florida's Big Bend Region wetland. Using NDVI derived from Landsat 5, 7 and 8 as a proxy for wetland health, we showed that both response and recovery are influenced by the event (flood, hurricane, drought) and wetland (fresh or saltwater) types. Hurricanes and their associated saltwater storm surge caused NDVI reductions (i.e. stress) lasting a year or more before recovery was indicated in the NDVI trend for all wetland types. Recovery after droughts was much shorter, often beginning at the tail end of the drought and requiring only a month to recover to baseline levels. Freshwater wetlands were observed to be less resilient than saltwater wetlands to these hazards demonstrated by larger reductions in NDVI post-event. These results can be used to guide resource management

practices such additional freshwater releases from upstream controls after a surge event to help flush and freshen freshwater wetlands. The Apalachicola River is controlled by the Jim Woodruff Dam at Lake Seminole near the Florida-Georgia border. Additionally, future research to investigate the spatial distribution or zonation of wetlands as a function of the hydrologic attributes of hurricanes (storm surge followed by hydrologic flood) would be worthwhile.

2.6 References

Beven, J. (2005), Tropical Cyclone Report Hurricane Dennis 4-13 July 2005, Natl. Weather Serv. Natl. Hurric. Center. Trop. Predict. Cent.

Chabreck, R., and A. Palmisano (1973), The effects of Hurricane Camille on the marshes of the Mississippi River delta, Ecology, 54(5), 1118–1123, doi:10.2307/1935578.

Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh (2004), A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter, Remote Sens. Environ., 91(3-4), 332–344, doi:10.1016/j.rse.2004.03.014.

Chidumayo, E. N. (2001), Climate and phenology of savanna vegetation in southern Africa, J. Veg. Sci., 12(3), 347–354, doi:10.2307/3236848.

Conner, W. H., and M. Ozalpl (2002), Baldcypress Restoration in a Saltwater Damaged Area of South Carolina, Ecology, 365–369.

DeFries, R., M. Hansen, and J. Townshend (1995), Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data, Remote Sens. Environ., 54(3), 209–222.

Edmiston, H. L., S. a. Fahrny, M. S. Lamb, L. K. Levi, J. M. Wanat, J. S. Avant, K. Wren, and N. C. Selly (2008), Tropical Storm and Hurricane Impacts on a Gulf Coast Estuary: Apalachicola Bay, Florida, J. Coast. Res., 10055(10055), 38–49, doi:10.2112/SI55-009.1.are.

Evans, J. P., and R. Geerken (2006), Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure, Remote Sens. Environ., 105(1), 1–8, doi:10.1016/j.rse.2006.05.017.

Florida Climate Center, F. S. U. (2014), Drought, Florida state Univ. Available from: http://climatecenter.fsu.edu/topics/drought (Accessed 7 February 2014)

Gresham, C. A. (1993), Changes in baldcypress-swamp tupelo wetland soil chemistry caused by Hurricane Hugo induced saltwater inundation, in Proceedings of the Seventh Biennial Southern Silvicultural Research Conference, pp. 171–185, U.S. Dept. Agric. Forest Service Southern Forest Experiment Station, New Orleans, LA.

Grodsky, S. A., N. Reul, G. Lagerloef, G. Reverdin, J. A. Carton, B. Chapron, Y. Quilfen, V. N. Kudryavtsev, and H. Y. Kao (2012), Haline hurricane wake in the Amazon/Orinoco plume: AQUARIUS/SACD and SMOS observations, Geophys. Res. Lett., 39(20).

Han, X., X. Chen, and L. Feng (2015), Four decades of winter wetland changes in Poyang Lake based on Landsat observations between 1973 and 2013, Remote Sens. Environ., 156, 426–437, doi:10.1016/j.rse.2014.10.003.

Hatter, L. (2015), Apalachicola Bay Part 2: Climate Change And Collapse, WFSU. Available from: http://news.wfsu.org/post/apalachicola-bay-part-2-climate-change-and-collapse (Accessed 23 December 2015)

Herndon, R. (2012), Storm Data and Unusual Weather Phenomena: August 2012.

Huang, W., and W. K. Jones (2001), Characteristics of long-term freshwater transport in Apalachicola Bay, J. Am. Water Resour. Assoc. , 37(3), 605–616.

Huang, W., S. Hagen, P. Bacopoulos, and D. Wang (2015), Hydrodynamic modeling and analysis of sea-level rise impacts on salinity for oyster growth in Apalachicola Bay, Florida, Estuar. Coast. Shelf Sci., 156, 7–18.

Ji, L., and A. J. Peters (2003), Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices, Remote Sens. Environ., 87(1), 85–98, doi:10.1016/S0034-4257(03)00174-3.

Justice, C. O., J. R. G. Townshend, B. N. Holben, and C. J. Tucker (1985), Analysis of the phenology of global vegetation using meteorological satellite data, Int. J. Remote Sens., 6(8), 1271–1318.

Klemas, V. V, J. E. Dobson, R. L. Ferguson, and K. D. Haddad (1993), A coastal land cover classification system for the NOAA Coastwatch Change Analysis Project, J. Coast. Res., 9(3), 862–872.

Landmann, T., M. Schramm, C. Huettich, and S. Dech (2013), MODIS-based change vector analysis for assessing wetland dynamics in Southern Africa, Remote Sens. Lett., 4(2), 104–113, doi:10.1080/2150704X.2012.699201.

Lane, C., H. Liu, B. Autrey, O. Anenkhonov, V. Chepinoga, and Q. Wu (2014), Improved Wetland Classification Using Eight-Band High Resolution Satellite Imagery and a Hybrid Approach, Remote Sens., 6(12), 12187–12216, doi:10.3390/rs61212187.

Lim, C., and M. Kafatos (2002), Frequency analysis of natural vegetation distribution using NDVI/AVHRR data from 1981 to 2000 for North America: Correlations with SOI, Int. J. Remote Sens., 23(17), 3347–3383, doi:10.1080/01431160110110956.

Lloret, F., a. Lobo, H. Estevan, P. Maisongrande, J. Vayreda, and J. Terradas (2007), Woody plant richness and NDVI response to drought events in Catalonian (northeastern Spain) forests, Ecology, 88(9), 2270–2279, doi:10.1890/06-1195.1.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant (2000), Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, Int. J. Remote Sens., 21(6-7), 1303–1330, doi:10.1080/014311600210191.

Luo, J., K. Ying, and J. Bai (2005), Savitzky-Golay smoothing and differentiation filter for even number data, Signal Processing, 85(7), 1429–1434, doi:10.1016/j.sigpro.2005.02.002.

Matlock, M. (2009), Apalachicola National Estuarine Research Reserve, Florida, Encycl. Earth.

Middleton, B. A. (2009), Effects of Hurricane Katrina on the forest structure of baldcypress swamps of the Gulf Coast, Wetlands, 29(1), 80–87.

Middleton, B. A. (2016), Differences in impacts of Hurricane Sandy on freshwater swamps on the Delmarva Peninsula, Mid-Atlantic Coast, USA, Ecol. Eng., 87, 62–70, doi:10.1016/j.ecoleng.2015.11.035.

Mo, Y., B. Momen, and M. S. Kearney (2015), Quantifying moderate resolution remote sensing phenology of Louisiana coastal marshes, Ecol. Modell., 312, 191–199, doi:10.1016/j.ecolmodel.2015.05.022.

Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani (1997), Increased plant growth in the northern high latitudes from 1981 to 1991, Nature, 386(6626), 698–702, doi:10.1038/386698a0.

Myneni, R. B., C. J. Tucker, G. Asrar, and C. D. Keeling (1998), Interannual variations in satellite-sensed vegetation index data from 1981 to 1991, J. Geophys. Res. Atmos., 103(D6), 6145–6160, doi:10.1029/97JD03603.

National hurricane center (NHC) (2004), Hurricane Frances Advisory Archive, Natl. Hurric. Cent. Available from: http://www.nhc.noaa.gov/archive/2004/FRANCES.shtml

National Weather Service (NWS) (2016), Tropical Cyclone History for Southeast South Carolina and Northern Portions of Southeast Georgia.

Ormsby, J. P., B. J. Choudhury, and M. Owe (1987), Vegetation spatial variability and its effect on vegetation indices, Int. J. Remote Sens., 8(9), 1301–1306, doi:10.1080/01431168708954775.

Passeri, D. L., S. C. Hagen, S. C. Medeiros, M. V Bilskie, K. Alizad, and D. Wang (2015), The dynamic effects of sea level rise on low-gradient coastal landscapes : A review, Earth's Futur., 3, 1–23.

Pettorelli, N., S. Ryan, T. Mueller, N. Bunnefeld, B. Jedrzejewska, M. Lima, and K. Kausrud (2011), The Normalized Difference Vegetation Index (NDVI): Unforeseen successes in animal ecology, Clim. Res., 46(1), 15–27, doi:10.3354/cr00936.

Pirotti, F., M. A. Parraga, E. Stuaro, M. Dubbini, A. Masiero, and M. Ramanzin (2014), NDVI from Landsat 8 Vegetation indices to study movement dynamics of Capra Ibex in

mountain areas , Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci., XL(7), 147–153, doi:10.5194/isprsarchives-XL-7-147-2014.

Potter, C., P.-N. Tan, M. Steinbach, S. Klooster, V. Kumar, R. Myneni, and V. Genovese (2003), Major disturbance events in terrestrial ecosystems detected using global satellite data sets, Glob. Chang. Biol., 9(7), 1005–1021, doi:10.1046/j.1365-2486.2003.00648.x.

Ramsey III, E. W., D. K. Chappell, and D. G. Baldwin (1997), AVHRR lmagery Used to Identify Hurricane Damage in a Forested Wetland of Louisiana, Photogramm. Eng. Remote Sens., 63(3), 293–297.

Savitzky, A., and M. J. E. Golay (1964), Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Anal. Chem., 36(8), 1627–1639, doi:10.1021/ac60214a047.

Schwartz, M. D. (1999), Advancing to full bloom: planning phenological research for the 21st century, Int. J. Biometeorol., 42, 113–118, doi:10.1007/s004840050093.

She, X., L. Zhang, Y. Cen, T. Wu, C. Huang, and M. H. Al Baig (2015), Comparison of the Continuity of Vegetation Indices Derived from Landsat 8 OLI and Landsat 7 ETM+ Data among Different Vegetation Types, Remote Sens., 7(10), 13485–13506, doi:10.3390/rs71013485.

Stanturf, J. a., S. L. Goodrick, and K. W. Outcalt (2007), Disturbance and coastal forests: A strategic approach to forest management in hurricane impact zones, For. Ecol. Manage., 250(1-2), 119–135, doi:10.1016/j.foreco.2007.03.015.

Steyer, G. D., B. C. Perez, S. Piazza, and G. Suir (2007), Potential Consequences of Saltwater Intrusion Associated with Hurricanes Katrina and Rita, Sci. Storms-the USGS response to hurricanes 2005 US Geol. Surv. Circ. 1306, 137–146.

Switzer, T. S., B. L. Winner, N. M. Dunham, J. a Whittington, and M. Thomas (2006), Influence of sequential hurricanes on nekton communities in a southeast Florida estuary: short-term effects in the context of historical variations in freshwater inflow, Estuaries and coasts, 29(6A), 1011–1018.

The Florida State Emergency Response Team (2003), Spring Floods of 2003.

Tian, B., Y.-X. Zhou, R. M. Thom, H. L. Diefenderfer, and Q. Yuan (2015), Detecting wetland changes in Shanghai, China using FORMOSAT and Landsat TM imagery, J. Hydrol., 529(1), 1–10, doi:10.1016/j.jhydrol.2015.07.007.

Wang, Q., and J. D. Tenhunen (2004), Vegetation mapping with multitemporal NDVI in North Eastern China Transect (NECT), Int. J. Appl. Earth Obs. Geoinf., 6(1), 17–31, doi:10.1016/j.jag.2004.07.002.

Wang, W., J. J. Qu, X. Hao, Y. Liu, and J. a. Stanturf (2010), Post-hurricane forest damage assessment using satellite remote sensing, Agric. For. Meteorol., 150, 122–132, doi:10.1016/j.agrformet.2009.09.009.

Wang, Y. (2012), Detecting Vegetation Recovery Patterns After Hurricanes in South Florida Using NDVI Time Series, University of Miami.

Wellens, J. (1997), Rangeland vegetation dynamics and moisture availability in Tunisia: an investigation using satellite and meteorological data, J. Biogeogr., 24(6), 845–855, doi:10.1046/j.1365-2699.1997.00159.x.

White, M. a., P. E. Thornton, and S. W. Running (1997), A continental phenology model for monitoring vegetation responses to interannual climatic variability, Global Biogeochem. Cycles, 11(2), 217, doi:10.1029/97GB00330.

Wiegand, C. L., H. W. Gausman, J. A. Cueller, A. H. Gerbermann, and A. J. Richardson (1974), Vegetation density as deduced from ERTS-1 MSS response.

Yang, L. H., J. L. Bastow, K. O. Spence, and A. N. Wright (2008), What can we learn from resource pulses, Ecology, 89(3), 621–634, doi:10.1890/07-0175.1.

Zhang, X., M. a. Friedl, C. B. Schaaf, A. H. Strahler, J. C. F. Hodges, F. Gao, B. C. Reed, and A. Huete (2003), Monitoring vegetation phenology using MODIS, Remote Sens. Environ., 84(3), 471–475, doi:10.1016/S0034-4257(02)00135-9.

Zhu, Z., and C. E. Woodcock (2012), Object-based cloud and cloud shadow detection in Landsat imagery, Remote Sens. Environ., 118, 83–94, doi:10.1016/j.rse.2011.10.028.

CHAPTER 3: OPTICAL CLOUD PIXEL RECOVERY VIA MACHINE LEARNING

3.1 Introduction

Normalized Difference Vegetation Index (NDVI) conveys valuable information relating to vegetation properties on the land surface [*Justice et al.*, 1985; *Myneni et al.*, 1997]. NDVI is a vegetation index derived from optical remote sensors and represents the reflective and absorptive characteristics of vegetation in the red and near infrared (NIR) bands of the electromagnetic spectrum. For this reason, a chronological analysis of NDVI can indicate changes in vegetation conditions proportional to the absorption of photo-synthetically active radiation [*Sellers*, 1985]. Such time series analyses of NDVI can detect the impact of natural events or anthropogenic disturbances on vegetation and can play an important role in natural resource management [*Jovanović and Milanović*, 2015]. The role of NDVI change detection can provide multi-dimensional information such as differences in urban land use land cover changes [*Chen et al.*, 2004], vegetation dynamics, surface elevation and floodplain dynamics [*Marchetti et al.*, 2016]. NDVI can be downloaded at no cost from several publicly available optical remote sensors such as the Advanced Very High Resolution Radiometer (AVHRR), Moderate Resolution Imaging Spectro-radiometer (MODIS)-TERRA, MODIS-AQUA or Landsat satellites. Commercial satellites such as Satellite Pour l'Observation de la Terre (SPOT) also provides NDVI information. Among the publicly available sensors, Landsat has longest data record and is applicable for modeling terrestrial ecosystems on the global, continental, and regional scales.

The main drawback of Landsat, as with any optical satellite sensor, is that the imagery can obscured, shadowed, or saturated. The effects of clouds and cloud shadows, atmospheric

variability, and bi-directional effects include the outright omission or skewing of readings in the image. These issues hamper the monitoring of terrestrial ecosystems and introduce undesirable noise [*Gutman*, 1991; *Cihlar et al.*, 1997]. This is especially significant in change detection analyses at climatic time scales.

In addition to the optical issues mentioned above, sensors can also experience instrument failure. For example, the scan line corrector (SLC) of the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) sensor failed permanently in 2003. Therefore, about 22% of the pixels in an SLC-off image are not scanned. Several algorithms are available to accurately reconstruct the missing values using multi-temporal images of the corresponding pixels as referable information in a regression model [*Zeng et al.*, 2013].

All of these errors, especially those introduced by the presence of clouds, introduce large uncertainties in satellite images during information retrieval, signal processing, data compression and procedures causing anomalous results that are often difficult to correct [*Melgani*, 2006; *Eckardt et al.*, 2013]. To mitigate these effects, the commonly used Landsat NDVI data sets are multiple-day Maximum Value Composite (MVC) products [*Holben*, 1986]; however, some noise remains in the final imagery products that must be dealt with.

Data from Landsat, or optical satellites in general, are often discontinuous and faulty in warm coastal areas such as Florida's gulf and west coasts due to heavy near shore evapotranspiration [*Gutman et al.*, 2004]. Therefore, it is crucial to mitigate or eliminate faulty information and recover missing information in a defensible and repeatable manner to enhance NDVI as a viable tool for long term change detection analyses.

A number of methods for reducing noise and building high-quality NDVI time series data sets have been proposed, applied, and examined in the last two decades in accordance with data availability and research application. However, research into the recovery of missing NDVI pixels due to cloud cover is limited. The most common and widely used NDVI cloud pixel recovery methods are threshold based methods such as the best index slope extraction algorithm (BISE) [*Viovy et al.*, 1992]; Fourier-based fitting methods [*Cihlar*, 1996; *Roerink et al.*, 2000]; and asymmetric function fitting methods such as the asymmetric Gaussian function fitting approach [*Jonsson Lars Eklundh*, 2002] and the weighted least-squares linear regression approach [*Swets et al.*, 1999]. Some simpler approaches were also practiced before the year 2000 such as substitution and interpolation. Substitution approaches were used to fill in cloudy pixels by using information from adjacent cloud-free pixels in the same time period. Otherwise information was retrieved for the corresponding pixels from previous time periods with spatial relationships [*Long et al.*, 1999; *Lin et al.*, 2014]. In addition to methods applied directly to NDVI, data recovery techniques for other types of data are also informative. Several studies have used similar techniques to estimate missing rainfall data including inverse distance weighting [*Simanton and Osborn*, 1980], Expectation Maximization Algorithm [*Makhuvha et al.*, 1997]; and regression [*Lynch*, 2003].

Machine learning has also been applied to this problem. Artificial neural networks (ANNs) have been proven to produce reasonable relationships from small datasets while remaining relatively robust in the presence of noisy or missing input [*Ilunga and Stephenson*, 2005]. In that study, they utilized ANNs to fill in missing hydrological data in South Africa. Recently, artificial intelligence and machine learning research has accelerated and new

applications are constantly being brought forth. This is mainly the result of their capability to capture complex, nonlinear and dynamic relationships in function generalization and regression as well as classification of data. Specifically, evolutionary algorithms, including genetic programming, feed forward back propagation neural networks, support vector machines, and deep learning algorithms are effective at recognizing subtle patterns and thus have be employed to characterize the complex relationships between the cloudy and cloud-free pixels in the historical time series over spatial and spectral domains [*Jerez et al.*, 2010]. However, frequent clustered missing data such as seasonal storm clouds over multiyear time scales make the compilation of viable training and test data difficult. To combat this, ancillary predictor variables can aid in developing models to capture the complex distribution of the data. Developing these ancillary variables for use in change detection analyses requires that they have compatible spatial and temporal scales. Therefore, most conventional methods have suffered from lower learning capacity which has hampered their applicability and broader impacts.

Decision tree based methods such as Random Forests (RF) [*Breiman*, 2001] are recognized for their ability to recover missing values as well as accommodate high dimensional data and complex relations among variables. For example, Hapfelmeier and others in 2014 used RF to improve missing value prediction. RF utilizes an efficient training algorithm that generates an ensemble of decision trees. Each decision tree in a RF is a set of hierarchically organized restrictions or conditions, that are successively applied from a root (parent) node to a leaf (or terminal) node to make repeated predictions of the phenomenon represented by training data [*Breiman*, 2001]. In the regression case, the final result is the mean prediction of all of the decision trees in the RF. This ensemble prediction tends to provide good generalization

performance with efficient learning speed accuracy compared to conventional neural computing algorithms such as back propagation, and will be used here to recover NDVI from cloud pixels using hydrologic predictor data derived from the time-space-spectrum continuum.

While each of the approaches mentioned above has unique advantages and applicability to the subject problem, some have already been successfully applied to NDVI time series preprocessing. These methods are all based on modeling complex spatial, temporal or spatio-temporal data for recovering the value of missing pixels. However, it is still difficult to systematically reconstruct information under cloudy pixels at regional or larger scales with sufficient accuracy in warm coastal areas with frequent storms and broad cloud cover. Due to the stochastic nature of clouds, it is difficult to build a consistent relationship to recover the value of pixels beneath them. As a start, relationships in the time-space-spectrum domain exist between cloud and cloud-free pixels, which are useful as a historic memory of prevailing conditions. The addition of ancillary hydrologic data such as rainfall or water level, known to influence the vegetation characteristics, can enhance the predictive performance of models capable of synthesizing disparate data sources. The combination of time-space-spectrum memory coupled with additional relevant variables provides us with the tools to develop a novel approach for recovering Landsat NDVI values from beneath cloud cover. To summarize the work done to date and illustrate the novelty of the proposed method, Table 3.1 shows previous missing data prediction methods along with their advantages and disadvantages.

Table 3.1. Description of existing methods to recover missing values from geospatial

Data type	Method	Advantages	Disadvantages
NDVI	The Best Index Slope Extraction (BISE) [*Zhu et al.*, 2012][*Viovy et al.*, 1992]	Effective noise removal	Dependence on threshold value and predefined time period; resulting profiles insensitive to timing of NDVI change
	Fourier based fitting [*Roerink et al.*, 2000]	Retain amplitude of local maxima and minima in time series	Only determines overall curve shape, rather than identifying particular cycles; needs to rerun over the entire time series every time new data are added
	Savitzy-golay (S-G) [*Chen et al.*, 2004]	Preserves shape, timing and amplitude of time series for a broad range of phonologies	Running mean and median filters alter the timing of local maxima and minima, even when weighted.
	Asymmetric function fitting [*Zhu et al.*, 2012]	Preserves aesthetic value and geometric accuracy	Successive relaxation of parameters depending on fit requires trial and error
Missing Data (climate and remote sensing)	Nonlinear filter, ANN [*Tu*, 1996]	Computationally efficient; detects complex nonlinear relationships; multiple training algorithms are supported	"black box" nature; tendency to over fit

Data type	Method	Advantages	Disadvantages
	Multi-temporal regression	Efficient; applicable to small data sets	Sensitive to outliers; Therefore it gives doubtful estimates for prediction.
	Extreme learning machine [*Sovilj et al.*, 2015]	Less training time compared to BP and SVM/SVR; outperforms BP in many applications	Can over fit and get trapped in local minima
	Random Forest	No expectation of linear features; handles a wide range of training set sizes	Tendency to over fit for regression using limited, noisy data.

A careful investigation of the literature regarding the links between NDVI and other climatic parameters guided the selection of predictor variables. Barbosa and Lakshmi Kumar (2016) showed the links between NDVI and rainfall in north eastern Brazil. They explored vegetative drought in the region and found rainfall as the dominant causative factor to the event. Fu and Burgher (2015) found that the maximum temperature primarily splits NDVI values, followed by previous rainfall and then inter-flood dry period and resulting groundwater levels. Warmer months required more rain compared to cooler months to attain similar mean NDVI values in areas of high NDVI such as riparian zones, likely due to higher local evaporation. Inter-flood dry periods were found to be important for maintenance of NDVI levels, especially when rainfall is limited. Another contributing factor in NDVI dynamics is the groundwater level. Shallower groundwater levels tend to enhance NDVI and thereby vegetation greenness primarily due to the wetter environment. Wang and other in 2001 examined spatial responses of NDVI to

precipitation and temperature during a 9-year period (1989-1997). Among the considered climatic factors, precipitation and temperature strongly influence both temporal and spatial patterns of NDVI. Hao et al. (2012) explored the linkage of NDVI to temperature and precipitation in northern china. The NDVI response for grassland and forest to three climatic indices (i.e., yearly precipitation and highest and lowest temperature) was analyzed showing that the yearly precipitation and highest temperature were correlated with NDVI.

However, powerful tools are needed here to explore these complex relationships accurately and efficiently. In recent years, artificial intelligence and machine learning techniques have been rigorously proven effective for characterizing the complex relationships in classification and function generalization applications. Therefore, a novel hybrid information recovery method, named Optical Cloud Pixel Recovery, is proposed here by reconstructing the value of cloudy pixels through the established multi-parameter time-space-spectrum relationships with cloud-free pixels. In current study, OCPR predicted NDVI using a RF model trained and tested using a large high resolution spatio-temporal multi-parameter (temperature, precipitation, water level and months) data set. Therefore, the objective of this study was to develop and optimize the OCPR method and assess its performance in terms of information recovery from cloud coverage in remotely sensed Landsat MVC images. A comparison of the performance of RF with the linear regression and mean only methods to predict NDVI is also included to justify the complexity of the RF model.

3.2 Methods and Material

To restate, the objective of this study is to recover NDVI values from beneath cloudy and faulty pixels within Landsat MVC imagery. The method is based on three assumptions: (1) NDVI data are a proxy for vegetation vigor, therefore a monthly NDVI time series will follow the annual cycle of growth and decline; (2) The "cfmask" product provided with Landsat MVC imagery accurately identifies clouds and cloud shadows [*Zhu and Woodcock*, 2012]; (3) coastal NDVI dynamics are related to local hydrologic variables such as rainfall, temperature, and water level [*Simanton and Osborn*, 1980; *Makhuvha et al.*, 1997]. In line with these three assumptions, the OCPR method was developed. In the following sections, a brief description of the study area is provided, followed by the OCPR method development and assessment. Figure 3.1 presents a flowchart that summarizes the methodology.

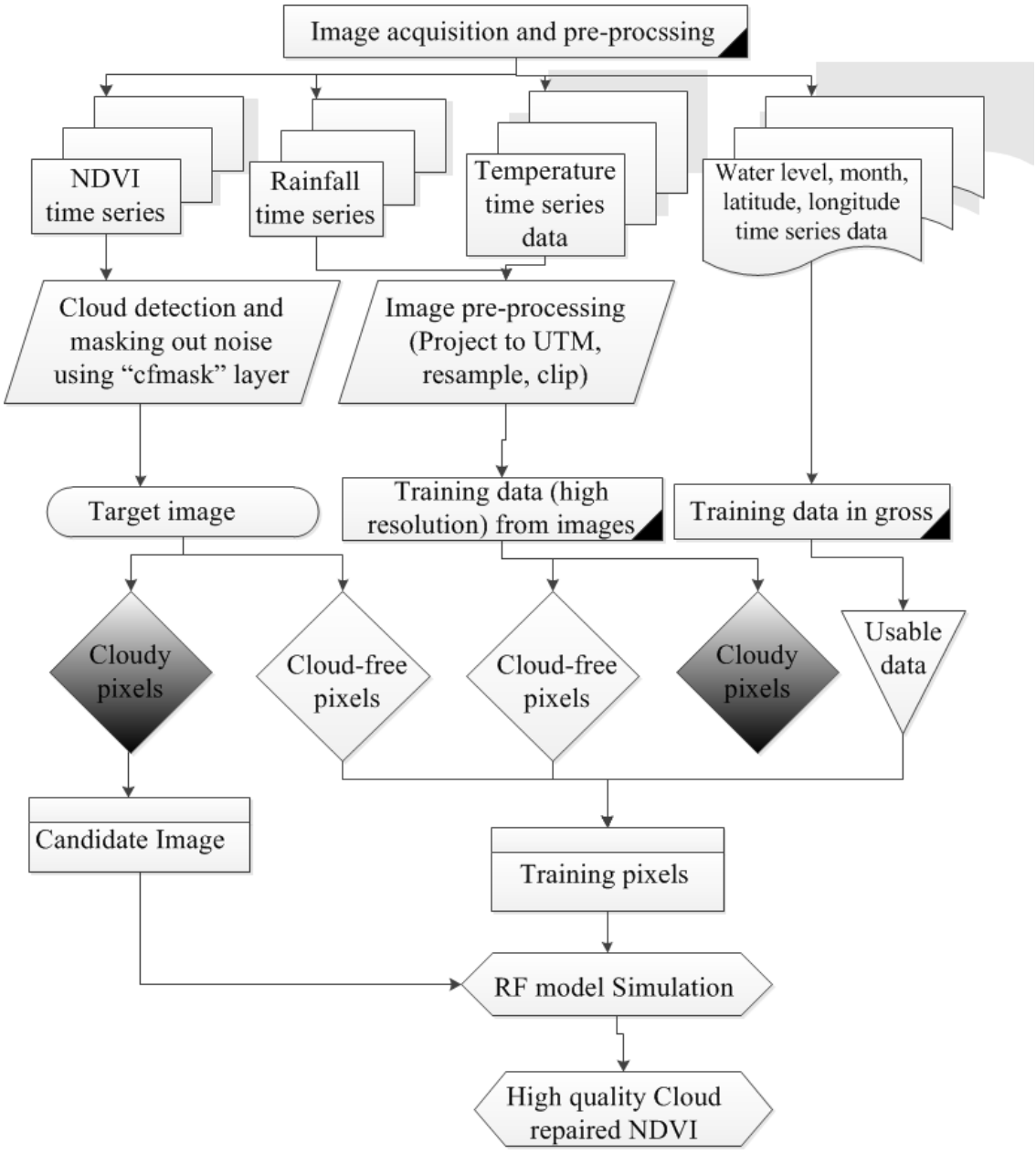


Figure 3.1. Schematic flowchart of the proposed OCPR method

3.2.1 Study Area

The study area is comprised of a part of Landsat TM scene L4-5 TM, Path 19/Row 39, located in Apalachicola Bay, Florida (See **Figure 3.2**). Apalachicola Bay is renowned for the largest oyster fishery in Florida [*Huang et al.*, 2015b]. It is a home to a rich variety of wetland plant, animal and microbial species. The Apalachicola lower river region as a whole is a nearly uninterrupted series of natural habitats including marshes, swamps, upland vegetation, and flood plains. Much of the basin vegetation has the appearance of a mature forest because of rapid regrowth. Although some municipalities (Apalachicola and Eastpoint) are situated near or within the riverine and tidal flood plains, they are not major urban centers. Therefore, there is very little industrialization in the basin. The study area includes both Gulf and Franklin counties. Wetland areas, including both forested and non-forested wetlands, comprise about 42% of the study area excluding open water and urban areas. The non-wetland and non-forested areas are mainly covered by agriculture, buildings and invasive vegetation.

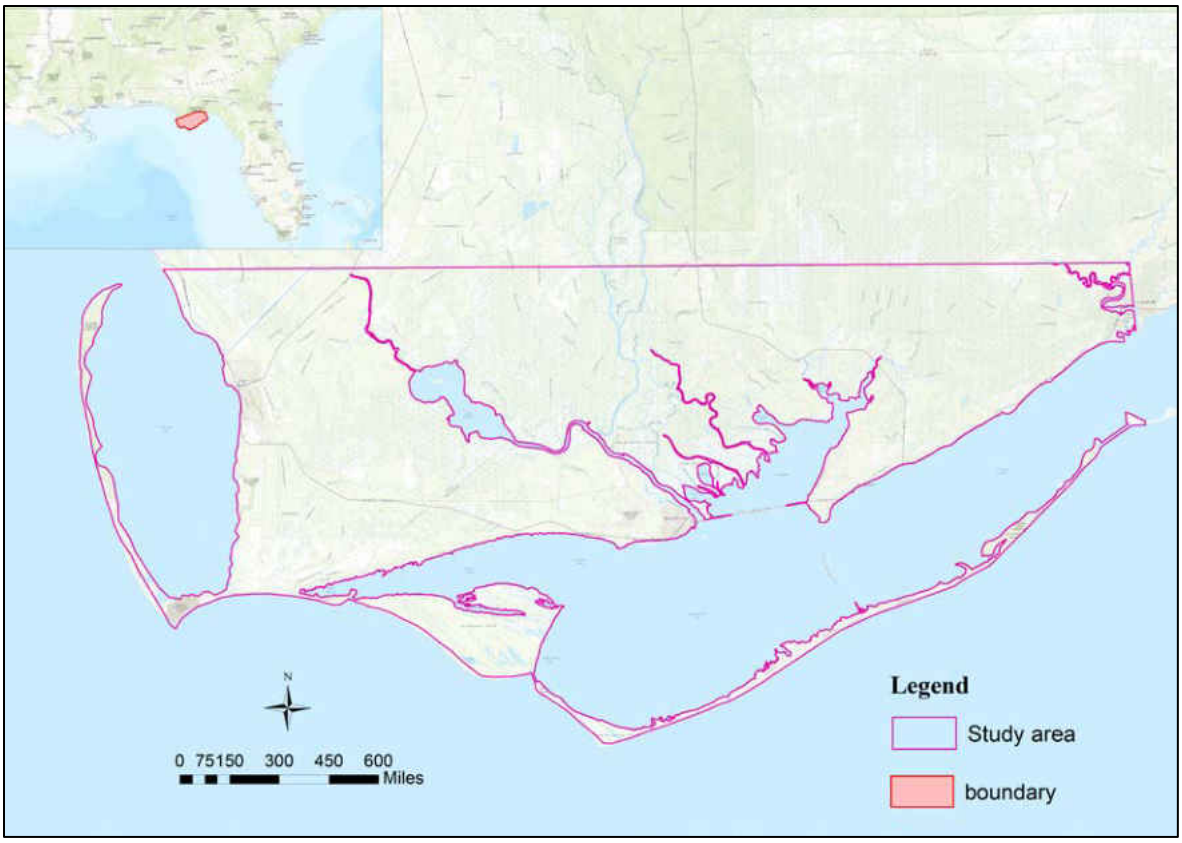


Figure 3.2. Study Area in Apalachicola Bay, Florida.

3.2.2 Optical Cloud Pixel Recovery

The proposed method to recover cloud from optical cloud is unique and computes an NDVI value for cloudy or faulty pixels is computed through an empirically trained random forest model. The base dataset is composed of the relevant spatially distributed hydrologic time series data associated with the available body of Landsat MVC NDVI images. The predictor variables include mean monthly temperature, cumulative monthly precipitation, mean monthly water level, calendar month encoded as a sequential number from 1 to 12, northing and easting (coordinates). The objective of the method is to train and test the model using historic data and validate its

performance in terms of its ability to accurately recover hypothetical (i.e. synthetic) cloudy pixels manually inserted into the validation images.

3.2.2.1 Random Forest

Random Forest (RF) is a decision tree based method for classification and regression. It is an ensemble of multiple decision trees, or a set of hierarchically ordered conditions that each produce individual predictions (i.e. class or regression value) that are aggregated into a single prediction by majority vote (classification) or averaging (regression). The conditions are sequentially applied to a randomly selected subset of the data from a root (parent) node to a terminal (or child) node to make repeated predictions of the phenomenon represented by training data [*Breiman*, 2001]. The child nodes can be thought of, metaphorically, as the leaf of a tree. The trees used in developing the RF algorithm are referred to as Classification and Regression Trees (CART). The prediction of missing data values is generally achieved through a model developed by the algorithm and a set of training data. The model comprises a number of CART trees as set by the model developer. Training and testing (and sometimes validation) datasets are extracted from the total data corpus to train the model and then test (and validate) the model's prediction capabilities. The predictions for a RF regression model are trained by finding the mean of all the predictions of each CART tree that best minimize the error function. Each decision tree in a RF utilizes a randomly chosen training subset and then replaced for a number of times equal to the number of trees in the ensemble [*Breiman*, 1996]. The prediction output of the RF is based on the average of the prediction of all the regression trees, or the classification receiving the highest number of "votes" from the trees. Recursive splitting and multiple

classifications or regressions are carried out to run the analysis of the decision trees [*Rodriguez-Galiano et al.*, 2014]. In other words, the RF algorithm is initiated by dividing the target variable or parent node into binary parts, where the child nodes are purer than the parent node. Throughout this procedure, the decision tree progresses through all candidate splits to determine the optimal split that maximizes the purity of the resulting tree. Residual sum of squares (RSS) shown in equation (3) is used as the splitting criteria.

$$RSS = \sum_{left}(y_i - y_L)^2 + \sum_{right}(y_i - y_L)^2 \quad (3)$$

where, $\sum_{\text{left}}$ mean y value for left node and $\sum_{\text{right}}$ mean y value for right node.

While classic regression trees are typically pruned (reducing the number of leaves or child nodes) according to a specific condition, decision trees in RF grow to maximum purity, constrained in most computer implementations by a maximum depth parameter. Each tree may share similar or different conditions as set by the model developer. Each tree sees part of the training data sets and captures part of the information it contains .The RF algorithm uses the Gini impurity index [*Breiman et al.*, 1984] to calculate the information purity of child nodes compared to that of their parent node. From the parent node, the data splitting process in each internal node of a condition of the tree is repeated until a pre-specified stop condition is reached. Each of the child nodes has a simple regression model attached to it, which applies to that node only.

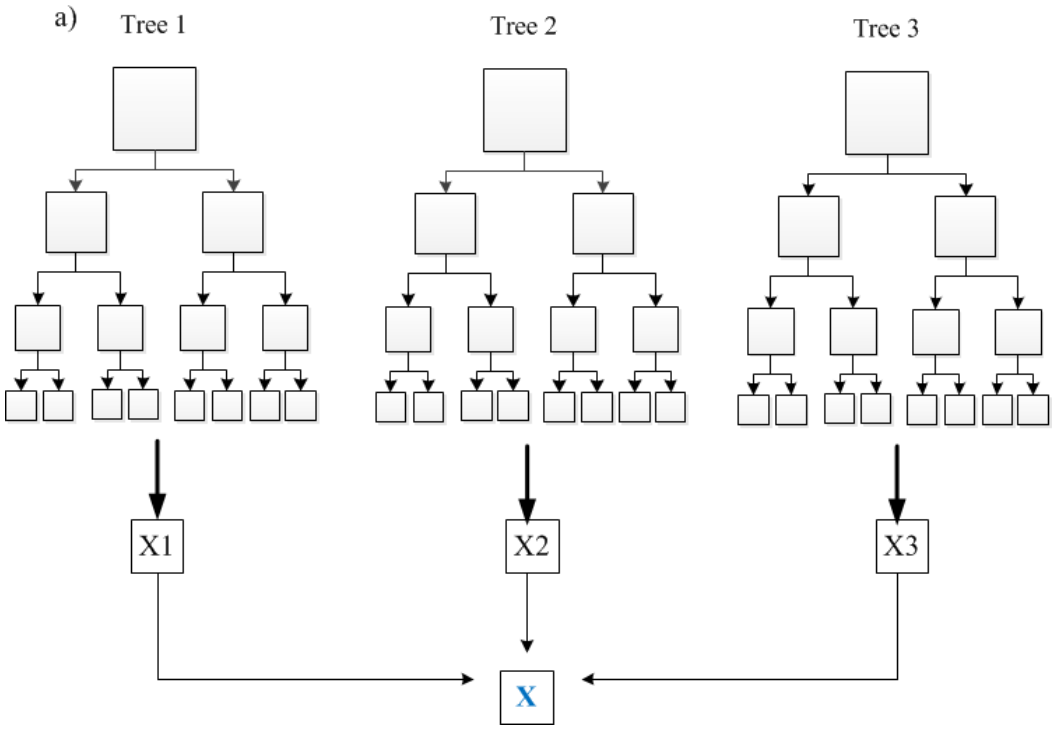
It has been observed in several studies that the RF algorithm offers characteristics that make it appealing for different areas of application. These include built-in feature selection capabilities, relatively high levels of accuracy in predictions, and a means for evaluating the

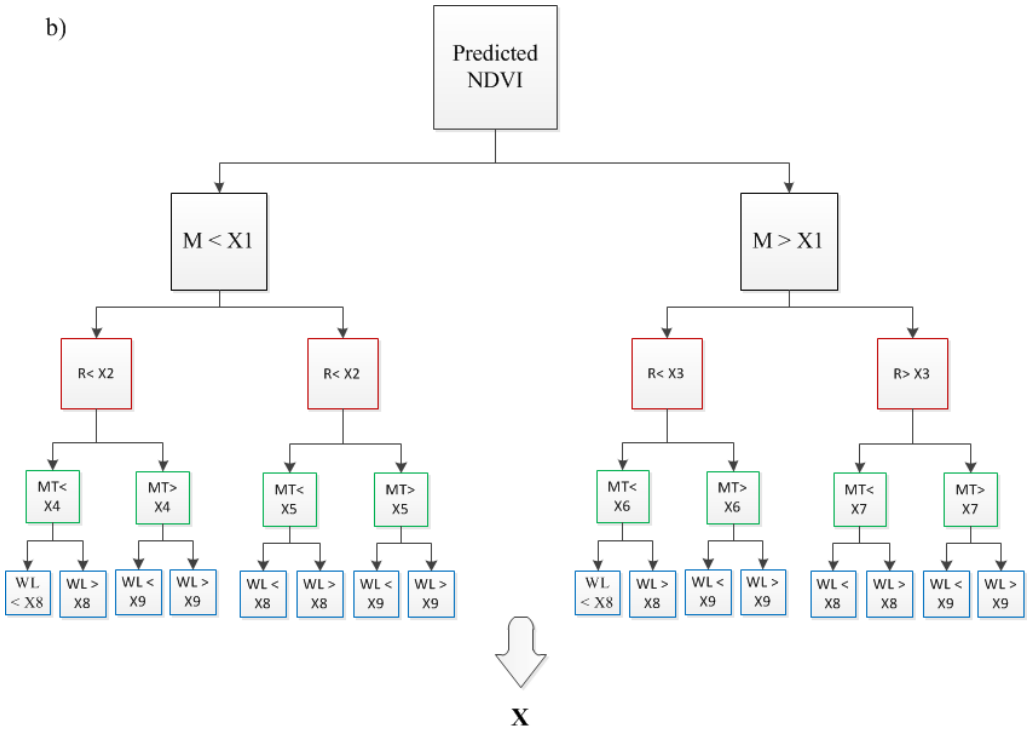
influence of each feature to the algorithm [*Palmer et al.*, 2007]. The theoretical background of RF regression was discussed in detail in [*Breiman*, 2001; *Svetnik et al.*, 2004; *Biau et al.*, 2008]. The essential attribute of RF is that it trains each tree individually, using a random sample of the data. This randomness helps to make the model more robust than a single decision tree and less likely to over fit to the training data. The ensemble of decision trees performs predictions of continuous variables by averaging the predictions of all trees. Finally a regression or classification is obtained either by using weighted or un-weighted majority voting mechanism. Random vectors are often generated in order to grow each tree in this ensemble of decision trees [*Breiman*, 2001]. A popular example of this is bootstrap aggregation, commonly referred to as bagging, where a random selection is made from the dataset in the training set without replacement [*Breiman*, 1996; *Biau et al.*, 2008].

The RF algorithm also provides an additional level of randomness to the bagging process. While nodes of standard trees are split by making use of the best possible split from the full list of predictor variables, RF uses a randomly selected subset of these variables; this drastically speeds up the tree growing process. However, the procedure in a RF is such that every node utilizes the best possible split from the randomly selected subset of predictors at the node to perform the node splitting procedure. The best splitter might either be the best overall, or just a fairly good splitter, or may not be of any help at all. If the splitter is not very helpful, the outcome from the split is two nodes which are essentially the same. One of the major benefits of the RF algorithm is that it is very easy to implement because there are only two important control variables: predictor sub-setting control for splitting at the nodes and the number of trees in the forest. Once sufficient values for these two parameters are determined, the algorithm is not

particularly sensitive to them [*Liaw and Wiener*, 2002]. Figure 3.3 shows a synopsis of an ensemble with four trees.

Critical parameters used to constrain a RF model are number of tree in a forest and maximum depth of the trees. Numerous opinions have been put forth for selecting the optimal number of the parameters. Previous researches indicated that sometimes, a larger number of trees in a forest only increase its computational cost without any significant gain in performance. It is also possible that there is a threshold beyond which there is no significant gain, unless a huge computational environment is available. As the number of trees grows, it does not always mean the performance of the forest is significantly better than previous forests (fewer trees), and doubling the number of trees is worthless [*Oshiro et al.*, 2012].





*** M= month; R= rainfall, MT= maximum temperature; WL= water level

Figure 3.3. a) Typical scheme of a Random forest regression tree structure; b) Detail of "Tree 1"

3.2.3 Application of OCPR in Apalachicola Bay

The schematic flowchart of the OCPR is represented in Figure 3.1. Recall that the OCPR includes four crucial steps: 1) image and data acquisition and input (target: NDVI and predictors: temperature, precipitation, water level, month) preparation for RF training and testing; 2) cloud and faulty value detection from NDVI; 3) training and testing the model as well as determining optimum model parameters; and 4) validation.

3.2.3.1 Image and Data Acquisition and Input Preparation for Machine Learning

3.2.3.1.1 Target Variable: NDVI

Surface Reflectance NDVI data were acquired between 1984 and 2014 from USGS Earth Resources Observation and Science (EROS) Center Science Processing Architecture (ESPA) archive. Since OCPR works based on the historical time series of NDVI, a sufficient body of data is required to characterize the relationship between NDVI and its predictor variables.

A threshold of 70% or less cloud cover was used for the acquisition of Landsat MVC imagery over the above-referenced time period, resulting in 252 usable scenes out of 384 (Figure 3.4). Since NDVI is released as a 16 day composite, two images per month are often available. Considering that the month is a feature in the predictor variable vector, when two images were available for a given month the one with less cloud coverage was selected. Of these, 93% were from Landsat-5 as it was the only data source from 1984 to 1998 and continued to acquire data until 2013. Landsat-7 data was avoided as it was contaminated by stripes as a result of the scan line corrector (SLC) in the Landsat Enhanced Thematic Mapper Plus (ETM+) sensor that failed permanently in 2003 [*Goward et al.*, 2006]. The remainder of the data came from Landsat-8.

In performing the NDVI OCPR, data availability was considered at the pixel level, as per-pixel cloud cover and the swath side lap between two adjacent paths were evaluated. Ancillary cloud mask, cloud shadow mask, adjacent cloud mask, snow mask, and water mask were available from the USGS earth explorer for the study area and were used for data quality assessment (QA). The QA layer, namely “cfmask”, identified water, cloud, cloud shadow, and snow [*Zhu and Woodcock*, 2012] and was included in the ESPA NDVI product used in this study.

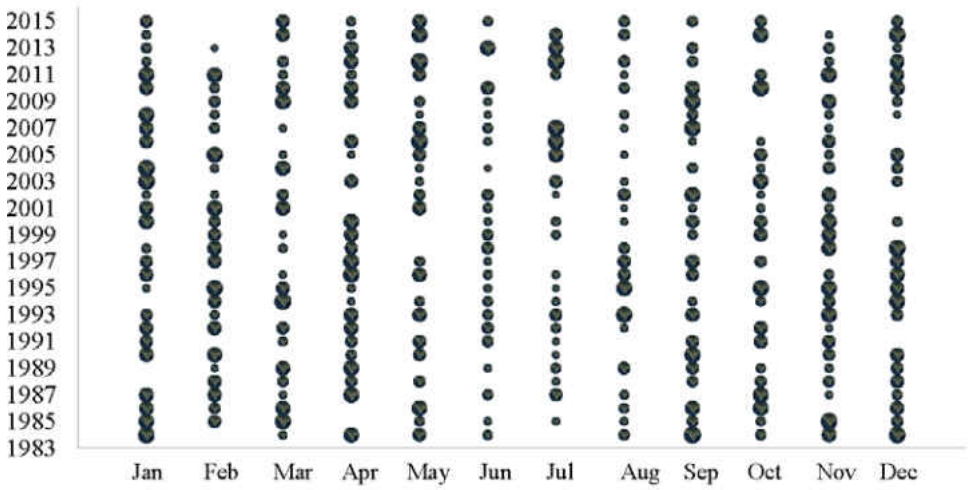


Figure 3.4. Availability and usability of 16-day composite NDVI images over the time series (1984-2015), the size of bubbles indicate % of available data in corresponding NDVI images

After NDVI image acquisition, all the images were registered and clipped to the spatial extent of the project. Spatial registration, projection and resampling using WGS1984 UTM Zone 16N was implemented to ensure that each 30 meter pixel location is consistent throughout the time series. These two processes (i.e., spatial registration and spatial clipping) were implemented using ArcGIS. NDVI was calculated as the ratio of red and NIR bands of a sensor system as shown in equation (4):

$$NDVI = \frac{NIR-R}{NIR+R} \quad (4)$$

3.2.3.1.2 Predictor Variables: Rainfall, Temperature, Water Level and Month

The data for the predictor variables rainfall and temperature were collected for the same spatio-temporal domain. The PRISM Climate Group collects climate observations from a large number of monitoring networks and builds spatial climate datasets to analyze short and long-

term climate patterns. Time series data for precipitation and temperature were available at a spatial resolution of 4 km and the temporal coverage starts from 1981 to the present. This dataset was available online free of charge. These data were modeled using climatologically-aided interpolation (CAI), which uses the long-term average patterns as first-guess of the spatial pattern of climatic conditions for a given month or day. CAI is robust to wide variations in station data density, which is necessary when modeling long time series. The data used herein were available based on either monthly or daily interpolation. Monthly average values were abstracted from daily values by averaging. These data use all station networks and data sources collected by the PRISM Climate group. After the precipitation and temperature data collection for the time domain was complete, they were registered, projected and clipped to the specifications of the NDVI data. Spatial resolution for each precipitation (inches) and temperature (Kelvin) pixel was resampled from 4 km to 30 m which did not add any new information but simply reallocated the information to correspond to the NDVI data. The processed NDVI, temperature (K) and precipitation (inches) images are shown in Figure 3.5.

Next, water level data over the same temporal domain were collected. Water level data are measured by NOAA at coastal gauge stations across the U.S. (see Figure 3.6). One gauge station is located in the study area (8728690 Apalachicola, FL). Monthly water level data were collected in meters measured from mean sea level (MSL) relative to the NAVD88 orthometric datum. The tide gauge data were added to the predictor variable vector and used as a proxy for prevailing water levels. If a storm surge event were experienced, the value would be higher and according to our current hypothesis and previous work, would lower the NDVI of impacted areas, especially freshwater wetlands.

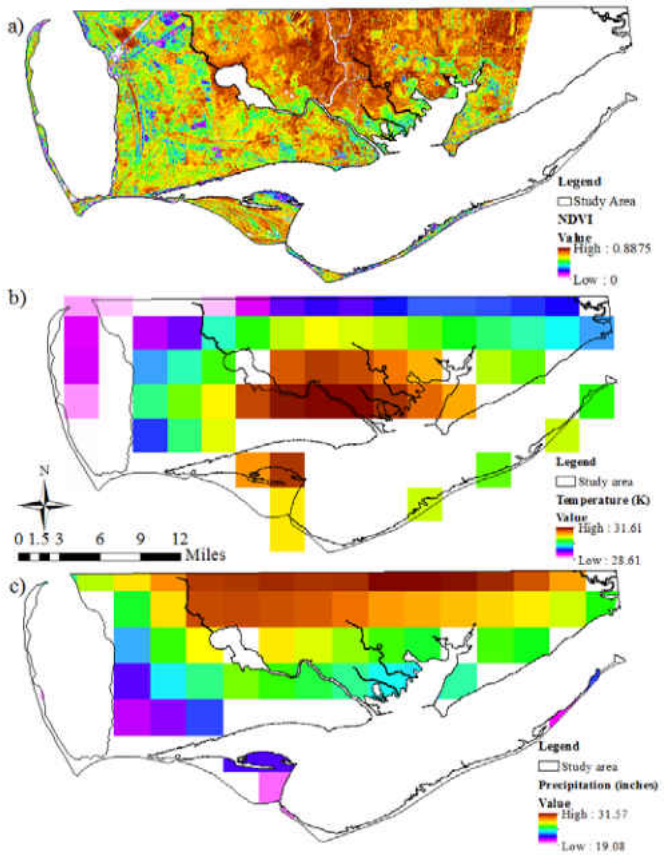


Figure 3.5. Sample NDVI (a), Temperature (b), and Precipitation (c) raster data

Figure 3.6. NOAA tide gauge station 8728690 – Apalachicola, FL (red box)

Table 3.2. Sample input data for training OCPR model

NDVI	Northing (m)	Easting (m)	Month	Precipitation (mm)	Temperature (˚ C)	Water Level (m)
0.38	720735	3307665	1	252.692	18.425	0.17
0.46	720765	3307665	1	298.636	21.775	0.17
0.18	720795	3307665	3	252.692	18.425	0.18
0.23	720825	3307665	4	321.608	23.45	0.21
0.45	720855	3307665	4	275.664	20.1	0.21
0.13	720885	3307665	6	344.58	25.125	0.26
0.09	720915	3307665	7	275.664	20.1	0.24
0.45	720945	3307665	8	367.552	26.8	0.2
0.58	720975	3307665	4	390.524	28.475	0.21
0.43	721005	3307665	2	298.636	21.775	0.13
0.26	721035	3307665	3	275.664	20.1	0.18

3.2.3.2 Cloud and Faulty Value Detection from NDVI

In order to ensure that the labeled data were clean for training/testing of the RF and linear regression, pixels were classified as cloudy or cloud-free. After this preprocessing step, climatologic maps were generated for each month of each year to calculate the averaged percentage of clouds (POCs) over the area of interest. Percentage of cloud cover was calculated by averaging the corresponding cloud-free pixel values in each image, pixel-by-pixel. Since the data were collected from different sources, each dataset had their own label for anomalous data. For instance, NDVI time series gives "*NaN*" to a void value to a pixel in the climatologic map.

Again, “cfmask” layers were processed to make a binary map that reclassifies all quality flag pixels as “0.0” and rest of the valid pixels as “1.0”. A raster multiplication was done using the binary reclassified time series “cfmask” and the NDVI time series. Appendix A shows an example of the procedure and the resultant NDVI after the raster multiplication by the binary “cfmask layer”. This function removes the faulty values from the NDVI time series and leaves a void pixel in those faulty pixels. Again, temperature and rainfall time series gives “0.0” to a void value to a pixel in the climatologic map. For consistency in the input and target data, “0.0” were given to all void pixels to all time series. POC of the climatologic map at month, *j* can be calculated as

$$\mathrm{POC}^{j} = 100 \times \frac{N_c j}{N_j} \tag{5}$$

Where N_c^j is the total number of cloudy pixels and *Nj* is the total number of pixels (i.e., both cloudy and cloud-free pixels) in the climatologic map.

3.2.3.3 Selection of Input for Training OCPR Model

It is very important to select reliable inputs for the construction of the OCPR model. The final prediction performance is highly dependent on the trained model. Candidate/training pixels are those cloud free pixels or information from target and training dataset. Among the dataset, NDVI is the target dataset and temperature, rainfall, water level, month, northing, and easting are the predictor datasets. As shown in Figure 3.7, target pixels (i.e., cloudy pixels) are shown in red color in candidate image and training inputs are all corresponding pixels in all parameters except

the target pixel. While NDVI, temperature, rainfall, northing, and easting are pixel based information, month and water level had a unique data for each time period (months of each year).

3.2.3.4 Building the Prediction Model

The random Forest package that was used in the current study was implemented in Python. The scikit-learn (sklearn) [*Pedregosa and Varoquaux*, 2011]module was used to train and run the RF model and the GDAL[*GDAL*, n.d.] module was used to get the spatial information from the geo referenced raster images of all the target and predictor variables. 70% of the data corpus was randomly selected, without replacement, as the training data with the remaining 30% held out for testing. The parameters that need to be set are the number of trees (k) to be generated and the number of predictors randomly sampled at each split. Figure 3.3 (a) and Figure 3.3 (b) explained in detail about the number of trees and splitting conditions. For the maximum purity of the RF model, the cases of missing parameter for most of the predictor variables were removed. Release 0.17.of sklearn has a class "Imputer" that handles simple per-feature missing value imputation. Arrays containing "NaN"s can be processed by the algorithm to have those replaced by the mean, median, mode or selected condition set by developer of the corresponding feature. In current study (i) a predictor variable with missing values issued as the pseudo-target variable and is fitted with all the other predictor variables without missing values and (ii) missing values in the pseudo-target variable are predicted or imputed using the RF-fit. After imputation of missing values in a predictor layer, RF training and prediction of the real target variable follow. It was found that the RF can be generalized and implemented with much

faster training speed than other training algorithms by using a cloud service (Databricks) which is at least 50 times faster than a laptop or desktop workstation.

3.2.3.5 Validation and Performance Metrics

The OCPR models were evaluated against a linear regression (LR) model by comparing prediction accuracy. For quantitative validation of the model, hypothetical clouds were created where the underlying image has little to no cloud actual cover. This provides labeled data for validation purposes. The images containing the hypothetical clouds were deliberately excluded from the training / testing corpus. A performance matrix was developed for the hypothetical cloud pixels using RF based OCPR model and LR based model. Later, a demonstration of the method on images with little, moderate and heavy natural cloud cover was presented. These cases visually demonstrated the application potential of the new algorithm. The following prediction metrics were also used in the hypothetical cloud validation in order to compare the original data to model predictions.

3.2.3.5.1 Root-mean-square error (RMSE)

The root-mean-square error is defined as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left[\widehat{T}\ (S_t, t_t) - T\ (S_t, t_t)\right]^2} \quad (6)$$

where $\widehat{T}\ (S_t, t_t) - T\ (S_t, t_t)$ represents the difference between the predicted and observed NDVI at space–time points (S_t, t_t) and m is the length of the time series of observations for each location.

3.2.3.5.2 Correlation coefficient (COR)

Perhaps the simplest overall measure of performance, the correlation coefficient is defined as:

$$\text{Coefficient of determination} = \frac{Cov\left(\hat{T}\left(S_i,t_i\right),T\left(S_i,t_i\right)\right)}{S_{\hat{j}}S_j} \quad (7)$$

Where S_j and S_T indicate the standard deviations of predicted and observed NDVI values, respectively. The correlation coefficient measures the linear association between prediction and observation. However, it only performs well when data are normally distributed and it is sensitive to large values and outliers.

3.3 Results

3.3.1 Suitability and sensitivity analysis of RF Model

The best and most stable result was found using six hydrological predictors in the RF model. Table 3.3 shows that success-rate of RF model according to number of trees and maximum depth of trees. Twelve trees with a maximum depth of 30 was the optimum combination in this case as shown by the minimum RMSE and coefficient of determination (R^2) has been found highest. If we increased the number of trees and maximum depth beyond those values, we did not gain much improvement and the computation time significantly increased. Therefore, all the hydrologic predictor variables were kept in the RF based OCPR model with the selected number and maximum depth of trees.

Table 3.3. Sensitivity analysis of RF model using tree number and depth of tree in forest

Tree number	Tree depth	RMSE	R^2
10	12	0.0802	0.6987
12	12	0.0802	0.6989
22	12	0.0802	0.6985
35	12	0.0800	0.7001
60	12	0.0798	0.7023
120	12	0.0798	0.7017
10	30	0.0461	0.8949
12	30	0.0475	0.8944
22	30	0.0468	0.7017
35	30	0.0470	0.8951
60	30	0.0473	0.8946
120	30	0.0473	0.7018
10	60	0.0473	0.8949
12	60	0.0472	0.8955
22	60	0.0473	0.7025
35	60	0.0474	0.8967
60	60	0.0474	0.8973
120	60	0.0474	0.7040

3.3.2 Prediction of Missing Value Using RF/LR based OCPR Model

Results suggested that the RF based OCPR model using hydrologic parameters outperforms the traditional LR based OCPR model especially in terms of prediction accuracy. 30% of the total data corpus was selected, without replacement, for testing the models. Figure 3.9 shows scatter plots of the predicted versus true NDVI for the pixels in testing dataset. Figure 3.9 (a) shows the scatter plot of the RF-based OCPR model; Figure 3.9 (b) shows scatter plots of the LR model. The RF based model has an R^2 value of 0.8944 and an positively sloped linear trend while the LR model has a significantly weaker R^2 value of 0.2594 and a much flatter linear trend. The RMSE values shown in Table 3.4 also suggest that the RF based OCPR was able to

reconstruct the cloudy pixels quite closely. Based on this evidence, the RF OCPR outperforms LR in terms of prediction accuracy.

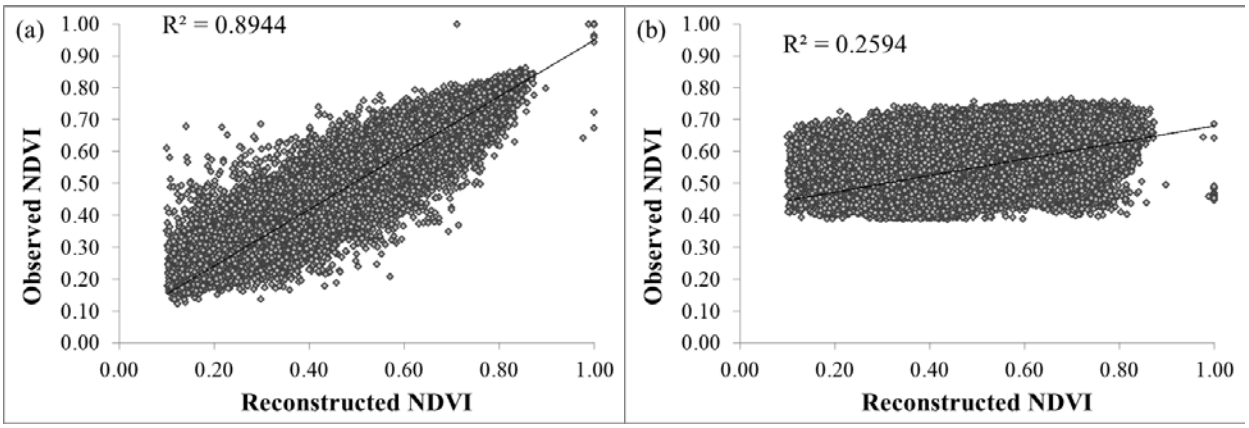


Figure 3.7. Scatter plots between the observed and reconstructed pixel values using a testing dataset with (a) RF-based OCPR; (b) LR based OCPR

Table 3.4. Comparison among different algorithms

Method	RMSE	R2
RF	0.0475	0.8944
LR	0.1257	0.2594

The performance of OCPR was also validated by comparing the predicted and observed NDVI reflectance data for synthetic or hypothetical clouds manually applied to selected images that were held out of the training / test data corpus. An area of naturally cloud-free pixels with heterogeneous nature was selected for demonstration as shown in the red solid color of highlighted area in Figure 3.10. These pixels were manually labeled as cloudy pixels (i.e., given a value of "NaN"). Then, the OCPR method was utilized to reconstruct the values of these pixels.

Image dates were selected so that the first date was before the study time period, the second date was near the middle and the final date was after. The three months were selected in three different times of year to capture seasonal variations. The first scenario shown in Figure 3.10 (a) was conducted for the images collected in April, 1984, the second scenario shown in Figure 3.10 (b) was conducted for the image collected in August, 2002 and the third scenario shown in Figure 3.10 (c) was conducted for the image collected in February, 2015. All the figures showed that the OCPR model is capable of reconstructing missing NDVI values caused by cloud contamination with visually plausible results. The predictions did not produce any extremely low or high values in the hypothetically clouded pixels.

These results suggest that RF has many advantages in fast and accurate learning capability when characterizing complex time-space-spectrum relationships in real world studies. The proposed RF based OCPR method is capable of recovering missing information with high efficacy for operational use. Possible improvements of the prediction accuracy for those extreme conditions (i.e., peak and valley) could be made by adding more high resolution dynamic input parameters in the model training process that can capture changes at the pixel scale.

In terms of model training time, LR is significantly faster than RF. Overall, RF shows much better prediction accuracy than LR in this real world application based on the patches or spatial patterns of NDVI reflectance values. Figure 3.11 further demonstrated the application of the RF based OCPR on three selected dates that had little, moderate and heavy cloud cover. The demonstration also further showed that the OCPR model is capable of reconstructing all missing values caused by cloud contamination with visually plausible results.

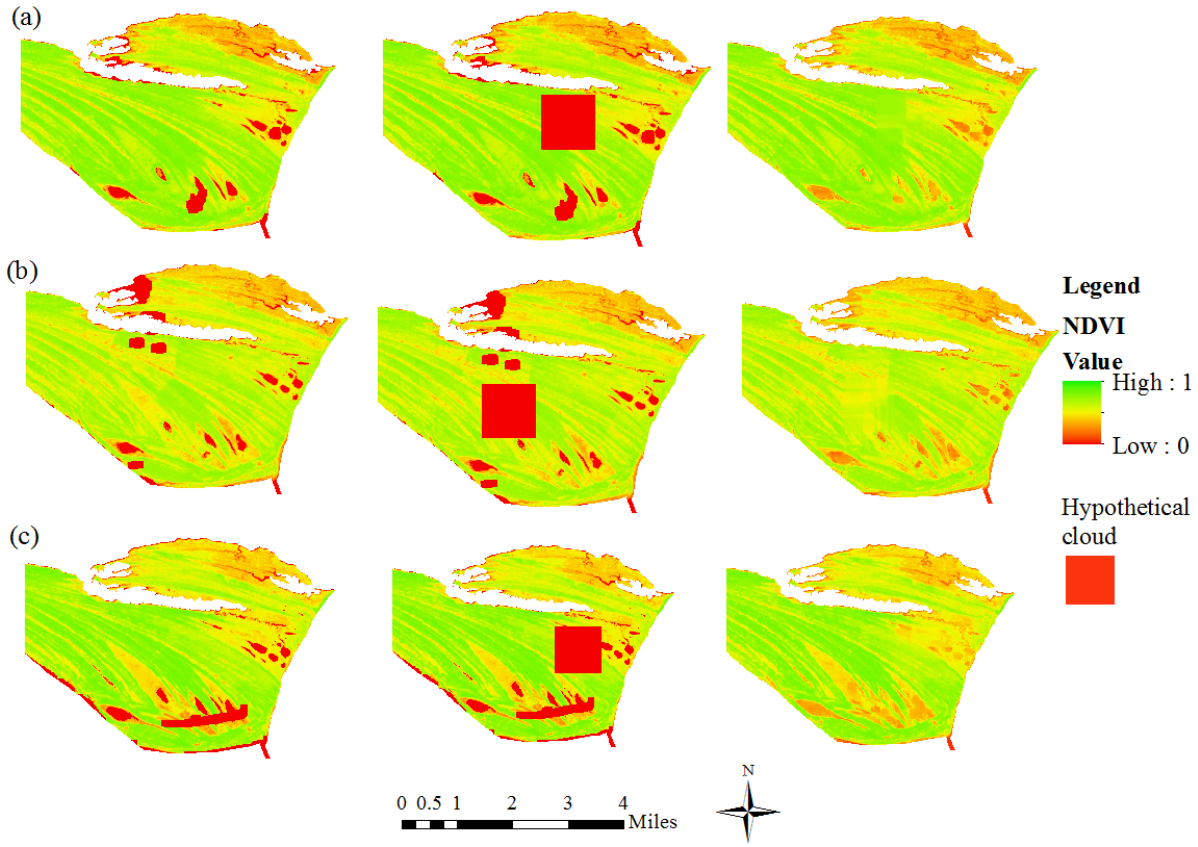


Figure 3.8. Application of OCPR model to reconstruct NDVI over hypothetical clouds on selected dates (a) April, 1984 (b) August, 2002; (c) February, 2015.

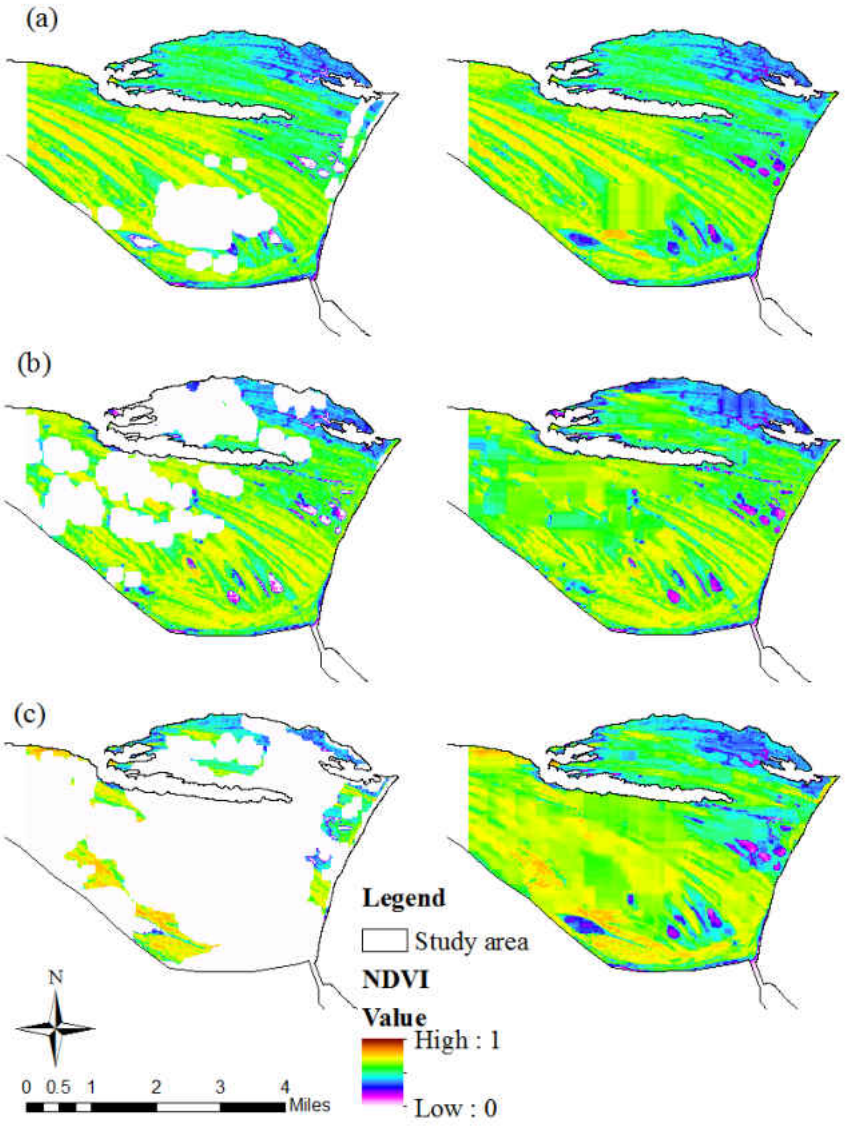


Figure 3.9. Comparisons of NDVI reflectance images under severe cloud cover before and after cloud removal with different training algorithms utilized by OCPR. (a) Cloudy images on February, 2010. (b) Reconstructed image from RF-based OCPR.

3.4 Discussion and Conclusion

OCPR via machine intelligence through RF is proposed in this paper to address the very important cloud repair issue in visible and infrared remote sensing images. This novel method takes advantage of the RF machine algorithm to characterize the complicated relationships among the NDVI, hydrologic parameters and spatial locations over the time-space spectrum. Inclusion of location (encoded as the northing and easting coordinates of the pixels) into the feature vector restricts model to reconstruct a value close to that of the neighboring pixels as well as a plausible value for that pixel in historical terms . The main limitations of this research involve the availability of high resolution feature data. The predictor variables found in the feature vector included temperature, precipitation and water level. These variables are currently not available with as high resolution as NDVI. This is especially true of water level data recorded in very limited locations such as National Oceanic and Atmospheric Administration (NOAA) tide gages. However, despite these limitations, the OCPR model works well and should enhance future analyses. The method was quite feasible to perform well even in very heterogeneous landscape where other approaches might fail. The experimental results suggest that OCPR is capable of reconstructing all missing information over the cloud-contaminated region with not only visually plausible but quantitative promising results, even under severe cloud cover situations. The basic RF is employed as a machine-learning tool in the OCPR method currently, and the final performance of the algorithm, to some extent, is still dependent on its training accuracy. Therefore, improvements can be achieved by further optimizing the training algorithms and architectures of RF with the newer ideas of treating missing values in predictor variables. Focusing on screening and selecting suitable inputs for the OCPR models is

critical to the prediction accuracy. It should be noted that the OCPR method is limited by the availability of the historical time series to characterize the complex time–spatial–spectral relationships between the cloudy and cloud-free pixels over the multiple parameters in a specific regoin. Therefore, the final prediction accuracy might be constrained by having fewer inputs for building the prediction model at some pixel locations. The idea of spatial information recovery via machine learning provides a promising and efficient approach to mitigate and eliminate cloud contaminations from the remote sensing images, which face highly heterogeneous land surfaces over which traditional methods have not worked well.

3.5 References

Barbosa, H. A., and T. V. Lakshmi Kumar (2016), Influence of rainfall variability on the vegetation dynamics over Northeastern Brazil, *J. Arid Environ.*, *124*, 377–387.

Beven, J. (2005), Tropical Cyclone Report Hurricane Dennis 4-13 July 2005, *Natl. Weather Serv. Natl. Hurric. Center. Trop. Predict. Cent.*

Biau, G., L. Devroye, and G. Lugosi (2008), Consistency of random forests and other averaging classifiers, *J. Mach. Learn. Res.*, *9*(2008), 2015–2033.

Breiman, L. (1996), Bagging Predictors, *Mach. Learn.*, *24*, 123–140.

Breiman, L. (2001), Random forests, *Mach. Learn.*, *45*(1), 5–32.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (1984), *Classification and Regression Trees*.

Chabreck, R., and A. Palmisano (1973), The effects of Hurricane Camille on the marshes of the

Mississippi River delta, *Ecology*, *54*(5), 1118–1123, doi:10.2307/1935578.

Chen, J., P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh (2004), A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter, *Remote Sens. Environ.*, *91*(3-4), 332–344, doi:10.1016/j.rse.2004.03.014.

Chidumayo, E. N. (2001), Climate and phenology of savanna vegetation in southern Africa, *J. Veg. Sci.*, *12*(3), 347–354, doi:10.2307/3236848.

Cihlar, J. (1996), Identification of contaminated pixels in AVHRR composite images for studies of land biosphere, *Remote Sens. Environ.*, *56*(3), 149–163.

Cihlar, J., H. Ly, Z. Li, J. Chen, H. Pokrant, and F. Huang (1997), Multitemporal, multichannel AVHRR data sets for land biosphere studies - Artifacts and corrections, *Remote Sens. Environ.*, *60*(1), 35–57.

Conner, W. H., and M. Ozalpl (2002), Baldcypress Restoration in a Saltwater Damaged Area of South Carolina, *Ecology*, 365–369.

DeFries, R., M. Hansen, and J. Townshend (1995), Global discrimination of land cover types from metrics derived from AVHRR Pathfinder data, *Remote Sens. Environ.*, *54*(3), 209–222.

Eckardt, R., C. Berger, C. Thiel, and C. Schmullius (2013), Removal of optically thick clouds from multi-spectral satellite images using multi-frequency SAR data, *Remote Sens.*, *5*, 2973–3006, doi:10.3390/rs5062973.

Edmiston, H. L., S. a. Fahrny, M. S. Lamb, L. K. Levi, J. M. Wanat, J. S. Avant, K. Wren, and N. C. Selly (2008), Tropical Storm and Hurricane Impacts on a Gulf Coast Estuary: Apalachicola Bay, Florida, *J. Coast. Res.*, *10055*(10055), 38–49, doi:10.2112/SI55-

009.1.are.

Evans, J. P., and R. Geerken (2006), Classifying rangeland vegetation type and coverage using a Fourier component based similarity measure, *Remote Sens. Environ.*, *105*(1), 1–8, doi:10.1016/j.rse.2006.05.017.

Florida Climate Center, F. S. U. (2014), Drought, *Florida state Univ.* Available from: http://climatecenter.fsu.edu/topics/drought (Accessed 7 February 2014)

Fu, B., and I. Burgher (2015), Riparian vegetation NDVI dynamics and its relationship with climate, surface water and groundwater, *J. Arid Environ.*, *113*, 59–68.

GDAL, 201x. (n.d.), GDAL - Geospatial Data Abstraction Library: Version x.x.x, *Open Source Geospatial Found.*

Goward, S., T. Arvldson, D. Williams, J. Faundeen, J. Irons, and S. Franks (2006), Historical record of landsat global coverage : Mission operations, NSLRSDA, and international cooperator stations, *Photogramm. Eng. Remote Sensing*, *72*(10), 1155–1169.

Gutman, G., A. C. Janetos, C. O. Justice, E. F. Moran, J. F. Mustard, R. R. Rindfuss, D. Skole, B. L. Turner II, and M. a Cochrane (2004), Land Change Science: Observing, Monitoring, and Understanding Trajectories of Change on the Earth's Surface, *Remote Sens. Digit. Image Process.*, *6*, 482.

Gutman, G. G. (1991), Vegetation indices from AVHRR: An update and future prospects, *Remote Sens. Environ.*, *35*(2-3), 121–136.

Han, X., X. Chen, and L. Feng (2015), Four decades of winter wetland changes in Poyang Lake based on Landsat observations between 1973 and 2013, *Remote Sens. Environ.*, *156*, 426–437, doi:10.1016/j.rse.2014.10.003.

Hao, F., X. Zhang, W. Ouyang, A. K. Skidmore, and A. G. Toxopeus (2012), Vegetation NDVI linked to temperature and precipitation in the upper catchments of Yellow River, *Environ. Model. Assess.*, *17*(4), 389–398, doi:10.1007/s10666-011-9297-8.

Hapfelmeier, A., T. Hothorn, C. Riediger, and K. Ulm (2014), Estimation of a Predictor's Importance by Random Forests When There Is Missing Data: RISK Prediction in Liver Surgery using Laboratory Data, *Int. J. Biostat.*, *10*(2), 165–183, doi:10.1515/ijb-2013-0038.

Hatter, L. (2015), Apalachicola Bay Part 2: Climate Change And Collapse, *WFSU*. Available from: http://news.wfsu.org/post/apalachicola-bay-part-2-climate-change-and-collapse (Accessed 23 December 2015)

Holben, B. N. (1986), Characteristics of maximum-value composite images from temporal AVHRR data, *Int. J. Remote Sens.*, *7*(11), 1417–1434.

Huang, W., and W. K. Jones (2001), Characteristics of long-term freshwater transport in Apalachicola Bay, *J. Am. Water Resour. Assoc.* , *37*(3), 605–616.

Huang, W., S. Hagen, P. Bacopoulos, and D. Wang (2015a), Hydrodynamic modeling and analysis of sea-level rise impacts on salinity for oyster growth in Apalachicola Bay, Florida, *Estuar. Coast. Shelf Sci.*, *156*, 7–18.

Huang, W., S. Hagen, P. Bacopoulos, and D. Wang (2015b), Hydrodynamic modeling and analysis of sea-level rise impacts on salinity for oyster growth in Apalachicola Bay, Florida, *Estuar. Coast. Shelf Sci.*, *156*, 7–18, doi:10.1016/j.ecss.2014.11.008.

Ilunga, M., and D. Stephenson (2005), Infilling streamflow data using feed-forward back-propagation (BP) artificial neural networks: Application of standard BP and pseudo Mac Laurin power series BP techniques, *Water SA*, *31*(2), 171–176.

Jerez, J. M., I. Molina, P. J. Garc??a-Laencina, E. Alba, N. Ribelles, M. Mart??n, and L. Franco (2010), Missing data imputation using statistical and machine learning methods in a real breast cancer problem, *Artif. Intell. Med.*, *50*(2), 105–115.

Ji, L., and A. J. Peters (2003), Assessing vegetation response to drought in the northern Great Plains using vegetation and drought indices, *Remote Sens. Environ.*, *87*(1), 85–98, doi:10.1016/S0034-4257(03)00174-3.

Jonsson Lars Eklundh, P. (2002), Seasonality extraction by function-fitting to time-series of satellite sensor data, *IEEE Trans. Geosci. Remote Sens.*, *40*, 1824–1832, doi:10.1109/TGRS.2002.802519.

Jovanović, M. M., and M. M. Milanović (2015), Normalized Difference Vegetation Index (NDVI as the basis for local forest management. Example of the municipality of Topola, Serbia, *Polish J. Environ. Stud.*, *24*(2), 529–535.

Justice, C. O., J. R. G. Townshend, B. N. Holben, and C. J. Tucker (1985), Analysis of the phenology of global vegetation using meteorological satellite data, *Int. J. Remote Sens.*, *6*(8), 1271–1318.

Klemas, V. V, J. E. Dobson, R. L. Ferguson, and K. D. Haddad (1993), A coastal land cover classification system for the NOAA Coastwatch Change Analysis Project, *J. Coast. Res.*, *9*(3), 862–872.

Landmann, T., M. Schramm, C. Huettich, and S. Dech (2013), MODIS-based change vector analysis for assessing wetland dynamics in Southern Africa, *Remote Sens. Lett.*, *4*(2), 104–113, doi:10.1080/2150704X.2012.699201.

Lane, C., H. Liu, B. Autrey, O. Anenkhonov, V. Chepinoga, and Q. Wu (2014), Improved

Wetland Classification Using Eight-Band High Resolution Satellite Imagery and a Hybrid Approach, *Remote Sens.*, *6*(12), 12187–12216, doi:10.3390/rs61212187.

Liaw, a, and M. Wiener (2002), Classification and Regression by randomForest, *R news*, *2*(December), 18–22.

Lim, C., and M. Kafatos (2002), Frequency analysis of natural vegetation distribution using NDVI/AVHRR data from 1981 to 2000 for North America: Correlations with SOI, *Int. J. Remote Sens.*, *23*(17), 3347–3383, doi:10.1080/01431160110110956.

Lin, C. H., K. H. Lai, Z. Bin Chen, and J. Y. Chen (2014), Patch-based information reconstruction of cloud-contaminated multitemporal images, *IEEE Trans. Geosci. Remote Sens.*, *52*(1), 163–174.

Lloret, F., a. Lobo, H. Estevan, P. Maisongrande, J. Vayreda, and J. Terradas (2007), Woody plant richness and NDVI response to drought events in Catalonian (northeastern Spain) forests, *Ecology*, *88*(9), 2270–2279, doi:10.1890/06-1195.1.

Long, D. G., Q. P. Remund, and D. L. Daum (1999), A cloud-removal algorithm for SSM/I data, *IEEE Trans. Geosci. Remote Sens.*, *37*(1), 54–62.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant (2000), Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *Int. J. Remote Sens.*, *21*(6-7), 1303–1330, doi:10.1080/014311600210191.

Luo, J., K. Ying, and J. Bai (2005), Savitzky-Golay smoothing and differentiation filter for even number data, *Signal Processing*, *85*(7), 1429–1434, doi:10.1016/j.sigpro.2005.02.002.

Lynch, S. D. (2003), *Development of a RASTER Database of Annual, Monthly and Daily*

Rainfall for Southern Africa.

Makhuvha, T., G. Pegram, R. Sparks, and W. Zucchini (1997), Patching rainfall data using regression methods. 1 Best subset selection, EM and pseudo-EM methods: Theory, *J. Hydrol.*, *198*(1-4), 289–307.

Marchetti, Z. Y., P. G. Minotti, C. G. Ramonell, F. Schivo, and P. Kandus (2016), NDVI patterns as indicator of morphodynamic activity in the middle Paran?? River floodplain, *Geomorphology*, *253*, 146–158.

Matlock, M. (2009), Apalachicola National Estuarine Research Reserve, Florida, *Encycl. Earth*.

Melgani, F. (2006), Contextual reconstruction of cloud-contaminated multitemporal multispectral images, *IEEE Trans. Geosci. Remote Sens.*, *44*, 442–455, doi:10.1109/TGRS.2005.861929.

Middleton, B. A. (2009), Effects of Hurricane Katrina on the forest structure of baldcypress swamps of the Gulf Coast, *Wetlands*, *29*(1), 80–87.

Middleton, B. A. (2016), Differences in impacts of Hurricane Sandy on freshwater swamps on the Delmarva Peninsula, Mid-Atlantic Coast, USA, *Ecol. Eng.*, *87*, 62–70, doi:10.1016/j.ecoleng.2015.11.035.

Mo, Y., B. Momen, and M. S. Kearney (2015), Quantifying moderate resolution remote sensing phenology of Louisiana coastal marshes, *Ecol. Modell.*, *312*, 191–199, doi:10.1016/j.ecolmodel.2015.05.022.

Myneni, R. B., C. D. Keeling, C. J. Tucker, G. Asrar, and R. R. Nemani (1997), Increased plant growth in the northern high latitudes from 1981 to 1991, *Nature*, *386*(6626), 698–702, doi:10.1038/386698a0.

Myneni, R. B., C. J. Tucker, G. Asrar, and C. D. Keeling (1998), Interannual variations in satellite-sensed vegetation index data from 1981 to 1991, *J. Geophys. Res. Atmos.*, *103*(D6), 6145–6160, doi:10.1029/97JD03603.

National hurricane center (NHC) (2004), Hurricane Frances Advisory Archive, *Natl. Hurric. Cent.* Available from: http://www.nhc.noaa.gov/archive/2004/FRANCES.shtml

Ormsby, J. P., B. J. Choudhury, and M. Owe (1987), Vegetation spatial variability and its effect on vegetation indices, *Int. J. Remote Sens.*, *8*(9), 1301–1306, doi:10.1080/01431168708954775.

Oshiro, T. M., P. S. Perez, and J. A. Baranauskas (2012), How many trees in a random forest?, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7376 LNAI, pp. 154–168.

Palmer, D. S., N. M. O'Boyle, R. C. Glen, and J. B. O. Mitchell (2007), Random forest models to predict aqueous solubility., *J. Chem. Inf. Model.*, *47*(1), 150–8, doi:10.1021/ci060164k.

Passeri, D. L., S. C. Hagen, S. C. Medeiros, M. V Bilskie, K. Alizad, and D. Wang (2015), The dynamic effects of sea level rise on low-gradient coastal landscapes : A review, *Earth's Futur.*, *3*, 1–23.

Pedregosa, F., and G. Varoquaux (2011), Scikit-learn: Machine learning in Python, … *Mach. Learn. …*, *12*, 2825–2830.

Pettorelli, N., S. Ryan, T. Mueller, N. Bunnefeld, B. Jedrzejewska, M. Lima, and K. Kausrud (2011), The Normalized Difference Vegetation Index (NDVI): Unforeseen successes in animal ecology, *Clim. Res.*, *46*(1), 15–27, doi:10.3354/cr00936.

Pirotti, F., M. A. Parraga, E. Stuaro, M. Dubbini, A. Masiero, and M. Ramanzin (2014), NDVI

from Landsat 8 Vegetation indices to study movement dynamics of Capra Ibex in mountain areas , *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, *XL*(7), 147–153, doi:10.5194/isprsarchives-XL-7-147-2014.

Potter, C., P.-N. Tan, M. Steinbach, S. Klooster, V. Kumar, R. Myneni, and V. Genovese (2003), Major disturbance events in terrestrial ecosystems detected using global satellite data sets, *Glob. Chang. Biol.*, *9*(7), 1005–1021, doi:10.1046/j.1365-2486.2003.00648.x.

Ramsey III, E. W., D. K. Chappell, and D. G. Baldwin (1997), AVHRR lmagery Used to Identify Hurricane Damage in a Forested Wetland of Louisiana, *Photogramm. Eng. Remote Sens.*, *63*(3), 293–297.

Rodriguez-Galiano, V. F., M. Chica-Olmo, and M. Chica-Rivas (2014), Predictive modelling of gold potential with the integration of multisource information based on random forest: a case study on the Rodalquilar area, Southern Spain, *Int. J. Geogr. Inf. Sci.*, *28*(7), 1336–1354, doi:10.1080/13658816.2014.885527.

Roerink, G. J., M. Menenti, and W. Verhoef (2000), Reconstructing cloudfree NDVI composites using Fourier analysis of time series, *Int. J. Remote Sens.*, *21*(9), 1911–1917.

Savitzky, A., and M. J. E. Golay (1964), Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.*, *36*(8), 1627–1639, doi:10.1021/ac60214a047.

Schwartz, M. D. (1999), Advancing to full bloom: planning phenological research for the 21st century, *Int. J. Biometeorol.*, *42*, 113–118, doi:10.1007/s004840050093.

Sellers, P. J. (1985), Canopy reflectance, photosynthesis and transpiration, *Int. J. Remote Sens.*, *6*(8), 1335–1372.

She, X., L. Zhang, Y. Cen, T. Wu, C. Huang, and M. H. Al Baig (2015), Comparison of the

Continuity of Vegetation Indices Derived from Landsat 8 OLI and Landsat 7 ETM+ Data among Different Vegetation Types, *Remote Sens.*, *7*(10), 13485–13506, doi:10.3390/rs71013485.

Simanton, J. R., and H. B. Osborn (1980), Reciprocal-Distance Estimate of Point Rainfall, *J. Hydraul. Div.*, *106*(7), 1242–1246.

Sovilj, D., E. Eirola, Y. Miche, K.-M. Björk, R. Nian, A. Akusok, and A. Lendasse (2015), Extreme learning machine for missing data using multiple imputations, *Neurocomputing*, *174*, 220–231, doi:10.1016/j.neucom.2015.03.108.

Stanturf, J. a., S. L. Goodrick, and K. W. Outcalt (2007), Disturbance and coastal forests: A strategic approach to forest management in hurricane impact zones, *For. Ecol. Manage.*, *250*(1-2), 119–135, doi:10.1016/j.foreco.2007.03.015.

Steyer, G. D., B. C. Perez, S. Piazza, and G. Suir (2007), Potential Consequences of Saltwater Intrusion Associated with Hurricanes Katrina and Rita, *Sci. Storms-the USGS response to hurricanes 2005 US Geol. Surv. Circ. 1306*, 137–146.

Svetnik, V., A. Liaw, C. Tong, and T. Wang (2004), Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules, *Mult. Classif. Syst.*, 334–343.

Swets, D. L., B. C. Reed, J. D. Rowland, and S. E. Marko (1999), A weighted least-squares approach to temporal smoothing of NDVI, in *ASPRS Annual Conference, From Image to Information*, edited by American Society for Photogrammetry and Remote and Sensing., Portland, Oregon.

Switzer, T. S., B. L. Winner, N. M. Dunham, J. a Whittington, and M. Thomas (2006), Influence

of sequential hurricanes on nekton communities in a southeast Florida estuary: short-term effects in the context of historical variations in freshwater inflow, *Estuaries and coasts*, *29*(6A), 1011–1018.

The Florida State Emergency Response Team (2003), *Spring Floods of 2003*.

Tian, B., Y.-X. Zhou, R. M. Thom, H. L. Diefenderfer, and Q. Yuan (2015), Detecting wetland changes in Shanghai, China using FORMOSAT and Landsat TM imagery, *J. Hydrol.*, *529*(1), 1–10, doi:10.1016/j.jhydrol.2015.07.007.

Tu, J. V. (1996), Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes, *J. Clin. Epidemiol.*, *49*(11), 1225–1231.

Viovy, N., O. Arino, and a. S. Belward (1992), The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series, *Int. J. Remote Sens.*, *13*(8), 1585–1590.

Wang, J., K. P. Price, and P. M. Rich (2001), Spatial patterns of NDVI in response to precipitation and temperature in the central Great Plains, *Int. J. Remote Sens.*, *22*(18), 3827–3844, doi:10.1080/01431160010007033.

Wang, Q., and J. D. Tenhunen (2004), Vegetation mapping with multitemporal NDVI in North Eastern China Transect (NECT), *Int. J. Appl. Earth Obs. Geoinf.*, *6*(1), 17–31, doi:10.1016/j.jag.2004.07.002.

Wang, W., J. J. Qu, X. Hao, Y. Liu, and J. a. Stanturf (2010), Post-hurricane forest damage assessment using satellite remote sensing, *Agric. For. Meteorol.*, *150*, 122–132, doi:10.1016/j.agrformet.2009.09.009.

Wang, Y. (2012), Detecting Vegetation Recovery Patterns After Hurricanes in South Florida Using NDVI Time Series, University of Miami.

Wellens, J. (1997), Rangeland vegetation dynamics and moisture availability in Tunisia: an investigation using satellite and meteorological data, *J. Biogeogr.*, *24*(6), 845–855, doi:10.1046/j.1365-2699.1997.00159.x.

White, M. a., P. E. Thornton, and S. W. Running (1997), A continental phenology model for monitoring vegetation responses to interannual climatic variability, *Global Biogeochem. Cycles*, *11*(2), 217, doi:10.1029/97GB00330.

Wiegand, C. L., H. W. Gausman, J. A. Cueller, A. H. Gerbermann, and A. J. Richardson (1974), *Vegetation density as deduced from ERTS-1 MSS response*.

Yang, L. H., J. L. Bastow, K. O. Spence, and A. N. Wright (2008), What can we learn from resource pulses, *Ecology*, *89*(3), 621–634, doi:10.1890/07-0175.1.

Zeng, C., H. Shen, and L. Zhang (2013), Recovering missing pixels for Landsat ETM+ SLC-off imagery using multi-temporal regression analysis and a regularization method, *Remote Sens. Environ.*, *131*, 182–194.

Zhang, X., M. a. Friedl, C. B. Schaaf, A. H. Strahler, J. C. F. Hodges, F. Gao, B. C. Reed, and A. Huete (2003), Monitoring vegetation phenology using MODIS, *Remote Sens. Environ.*, *84*(3), 471–475, doi:10.1016/S0034-4257(02)00135-9.

Zhu, W., Y. Pan, H. He, L. Wang, M. Mou, and J. Liu (2012), A changing-weight filter method for reconstructing a high-quality NDVI time series to preserve the integrity of vegetation phenology, *IEEE Trans. Geosci. Remote Sens.*, *50*(4), 1085–1094.

Zhu, Z., and C. E. Woodcock (2012), Object-based cloud and cloud shadow detection in Landsat imagery, *Remote Sens. Environ.*, *118*, 83–94, doi:10.1016/j.rse.2011.10.028.

CHAPTER 4: DISCUSSION AND CONCLUSION

4.1 Introduction

A long time series NDVI was analyzed to specify the impact of hydrologic event on wetland stresses. Such a long term analysis is much credible comparing to single event based before-after analysis that bring potential doubt about non-uniformity for all similar events. NDVI is highly responsive towards chlorophyll and measure the greenness of plants. The NDVI can identify flood stresses in plants since flood stressed plants reflect more blue and less infrared radiation. Landsat derived NDVI (30 m spatial resolution) which is a composite of 16 day data. An empirical data smoothing and prediction filter was applied over the time series to predict the missing monthly mean NDVI reflectance data. Thus an uninterrupted time series NDVI was obtained. NDVI is a widely used index to measure density of live green vegetation at global and regional scale. Usually the impact of hurricane on ecosystem can range from very massive to small. It depends on the hurricane trajectory or the nature of the forest. To deeply understand the stresses of different species of wetland, the study area was further classified in fresh and salt water wetlands. During 2004 and 2005, two hurricanes hit Apalachicola Bay, hurricane Dennis on 2004 and hurricane Katrina on 2005. Salt water wetlands showed less dynamic behavior before and after extreme events over the time series than freshwater wetlands. The evidence suggested that salt water wetland has high resiliency to natural hazard than freshwater wetlands. This research also showed that it took a year for wetlands to recover after a hurricane event, while it took very quick, a month to recover after a drought event for all wetland types. Though empirical S-G filter was used for prediction of mean NDVI, it takes into account only temporal observations within a selected window to predict the missing values. It is to be noted that we

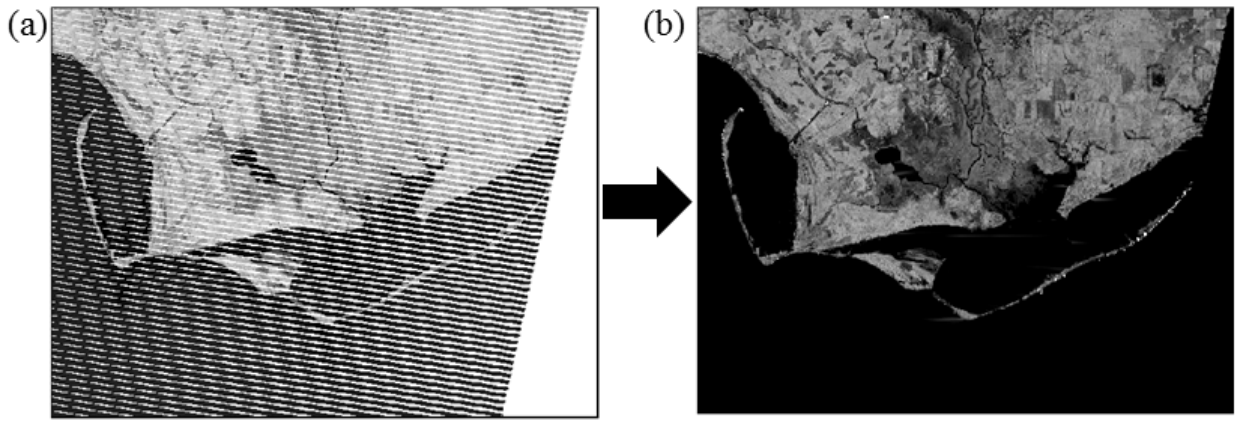
used the filter for predict mean NDVI instead of pixel by pixel NDVI prediction. A much un-investigated and under applied method to predict missing data was incorporation of big and multi variable data and use of machine learning technique. Against this backdrop, the current research went ahead to fill the research gap.

Missing data due to cloud contamination is a big hindrance in earth system analysis when processing remote sensing images retrieved from visible or infrared spectral ranges. Numerous computational methods for interpolation have been implemented to repair missing data caused by cloud and other reasons. All these algorithms are subject to many shortcomings. In order to provide reliable estimates for the missing value approximation, a novel and unique Optical Cloud Pixel Recovery (OCPR) method was proposed and applied in Apalachicola Bay. Multi parameter 30 year time series data were utilized to repair missing data in NDVI reflectance in Landsat data based on machine learning approach, random forest (RF). The study area is a coastal area and similar to other coastal areas, covered by heavy cloud most of the year especially wet seasons. OCPR enables to devise the cloud repair in a step by step strategy towards final estimation. The proposed methodology has longer running times compared to simple methods, but the overall increase in accuracy justifies this trade-off. For the purpose of demonstration, the performance of OCPR is evaluated by reconstructing the missing NDVI reflectance of Landsat over the study area for two specific dates. For comparison, the traditional artificial neural network (ANN) and linear regression (LR) were also implemented to reconstruct the same missing values. Experimental results show that the RF outperforms the ANN and LR algorithms by an enhanced machine learning capacity with simulated memory effect embedded in Landsat due to linking the complex time-space-spectrum continuum between cloud-free and cloudy pixels over a good

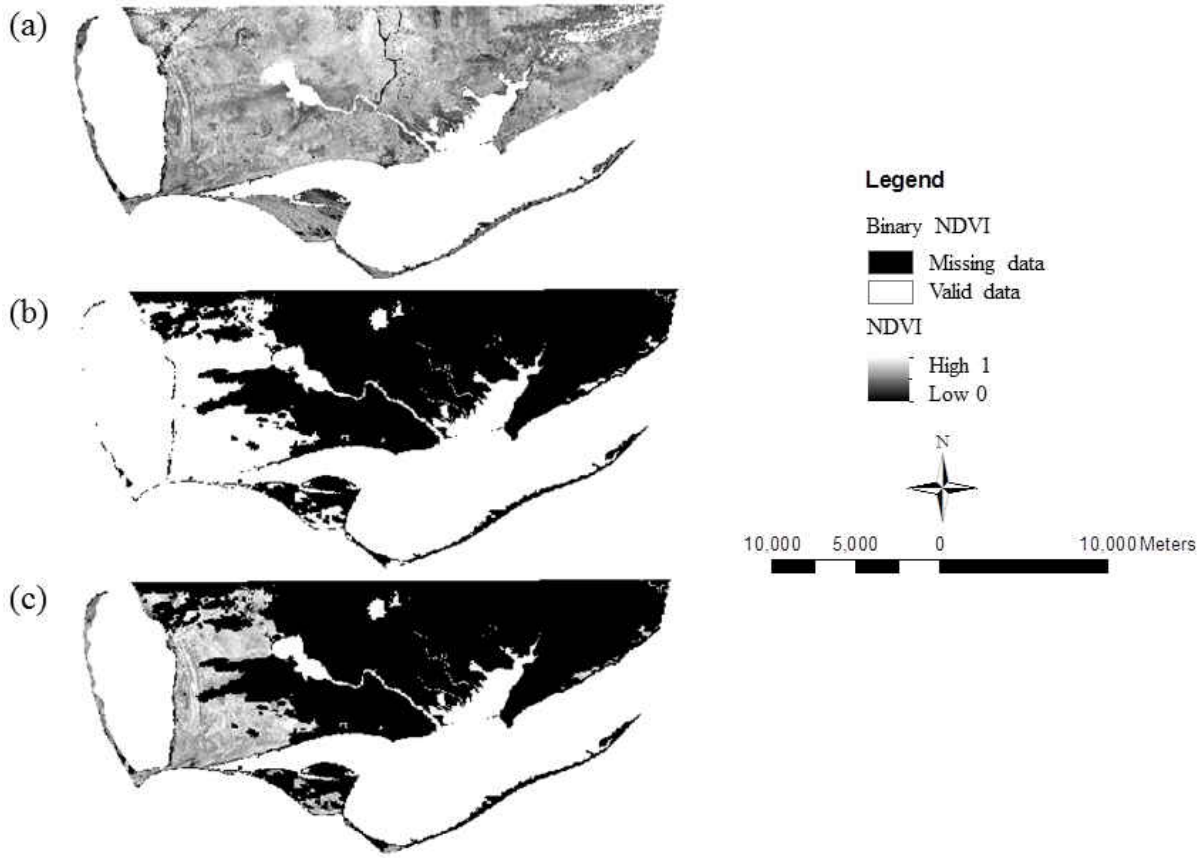
number of predictor variables. Temperature, precipitation, water level, month, spatial locations were selected as predictor variable to define the NDVI. The RF-based OCPR practice presents a correlation coefficient of 0.88 with root mean squared error of 0.09 between predicted and observed reflectance values. These findings suggest that the OCPR method is effective at recovering missing NDVI information and providing visually logical and quantitatively assured images for further scientific research about earth surface and landscape changes.

Machine learning methods, which have been used in predictive modeling of missing NDVI, such as RF, usually require a large training dataset and are debatable for data with missing values. However, as shown in this study, the RF algorithm, which is also a machine learning method, can be used in data-driven predictive modeling while large dataset is available. Nevertheless, this proposition may gain much popularity by further verification by testing RF modeling in other areas of homogeneous or more heterogeneous landscape. An advantage of RF over artificial neural networks and support vector machines is the linked imputation technique for representation of missing values in evidential data. This is an important advantage because evidential data with missing values is a common feature of areas with few available dataset.

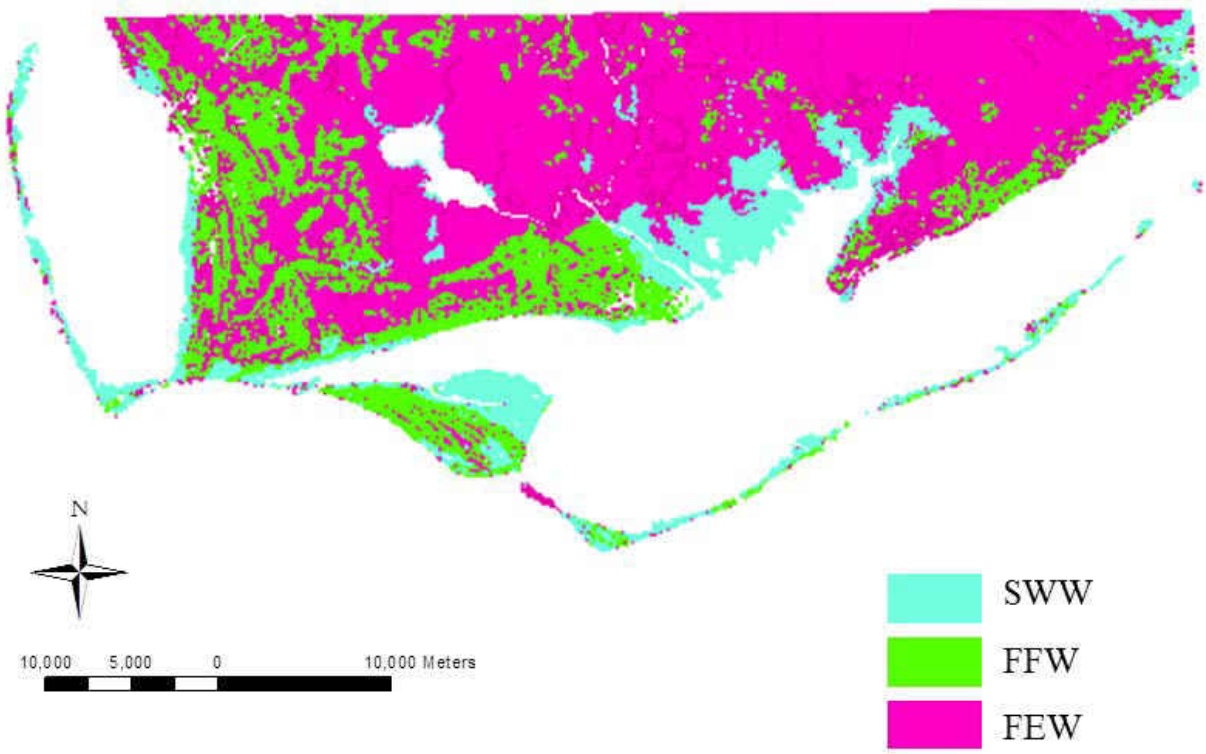
APPENDIX A: LIST OF FIGURES



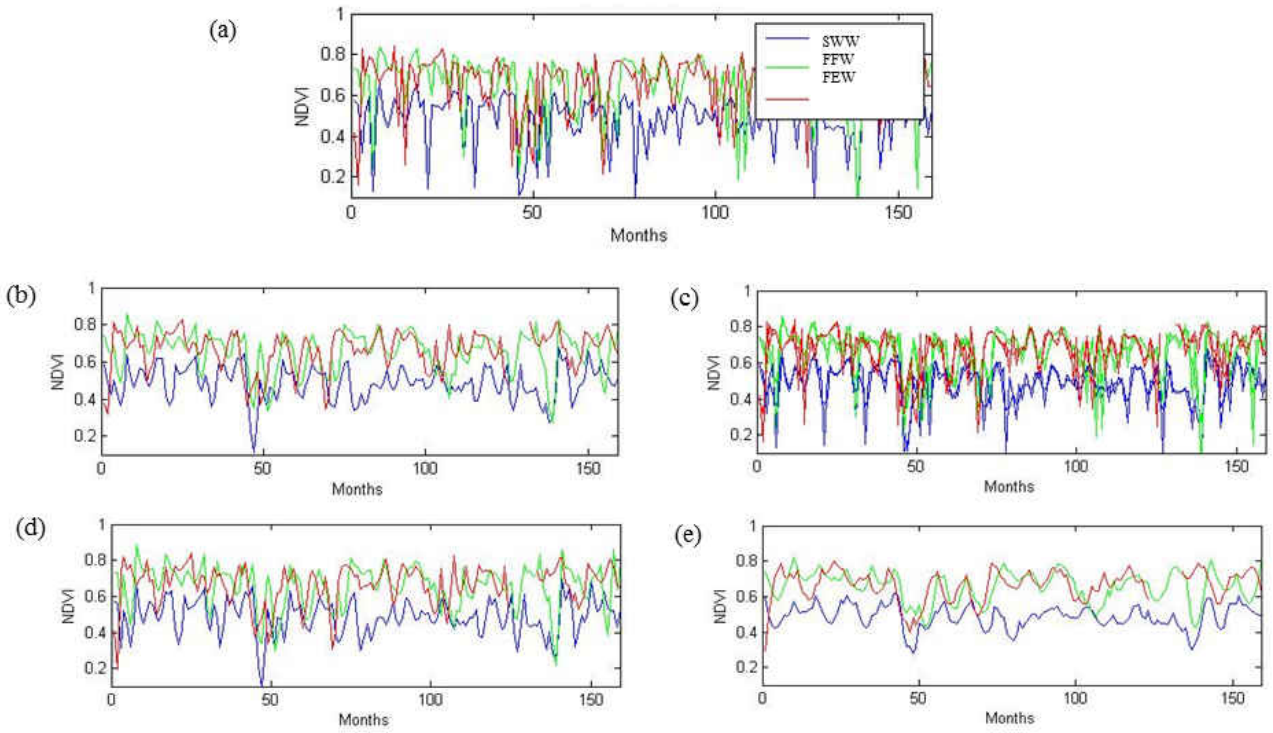
An example of Landsat NDVI de-stripping (a) stripped Landsat NDVI reflectance data; (b) De-stripped Landsat NDVI reflectance data using "cfmask" layer provided by



Landsat NDVI Cloud masking (a) raw NDVI reflectance data (b) binary cloud mask layer (c) final NDVI reflectance after adjusting cloudy and noisy data using "cfmask"



Study area wetland classification



Comparison of filtered data using S-G filters using different degree and window size. a) Observed NDVI from time series (2000-2015); b) Filtered NDVI with Degree 3, moving window size 5; c) Filtered NDVI with Degree 2, moving window size 3; c) Filtered NDVI with Degree 5, moving window size 7; d) Filtered NDVI with Degree 3, moving window size 9.

APPENDIX B: LIST OF TABLES

Water level and precipitation data

Year and Month	Water Level (m)	Precipitation (inches)
19831	0.015	4.300
19841	-0.039	4.730
19842	-0.030	3.930
19843	0.028	6.080
19844	0.067	9.180
19845	0.058	0.320
19846	0.073	3.370
19847	0.131	18.070
19848	0.104	4.720
19849	0.122	1.250
198410	0.147	1.780
198411	0.037	2.160
198412	0.015	0.910
19851	-0.085	5.580
19852	-0.070	1.780
19853	-0.049	2.550
19854	-0.064	0.860
19855	-0.003	2.720
19856	-0.006	3.910
19857	0.037	7.660
19858	0.177	16.180
19859	0.147	5.380
198510	0.226	11.230
198511	0.134	6.480
198512	-0.036	4.240
19861	-0.106	3.820
19862	-0.009	5.410
19863	-0.085	2.230
19864	-0.006	0.260
19865	0.104	4.360
19866	0.079	2.010
19867	0.019	3.340
19868	0.055	12.040
19869	0.174	9.290
198610	0.150	9.190
198611	0.113	5.180

Year and Month	Water Level (m)	Precipitation (inches)
198612	0.031	9.680
19871	-0.024	6.010
19872	0.009	4.180
19873	0.150	10.530
19874	-0.064	0.130
19875	0.028	1.960
19876	0.092	4.420
19877	0.113	3.090
19878	0.046	5.790
19879	0.095	5.220
198710	-0.067	0.150
198711	0.015	5.590
198712	-0.021	0.900
19881	-0.113	2.940
19882	-0.134	8.450
19883	-0.085	5.130
19884	-0.033	3.760
19885	-0.033	0.890
19886	0.022	3.600
19887	0.000	7.950
19888	0.037	13.300
19889	0.183	10.440
198810	0.070	1.770
198811	0.049	2.950
198812	-0.082	1.170
19891	-0.091	1.240
19892	-0.125	1.950
19893	-0.042	6.000
19894	-0.052	0.860
19895	-0.052	4.240
19896	0.092	8.880
19897	0.104	6.990
19898	0.076	4.170
19899	0.122	10.430
198910	0.000	2.630
198911	0.012	3.900
198912	-0.116	7.150

Year and Month	Water Level (m)	Precipitation (inches)
19901	-0.070	2.430
19902	-0.006	3.940
19903	0.147	4.160
19904	0.055	2.230
19905	0.095	0.510
19906	0.043	2.820
19907	0.064	9.340
19908	0.110	2.320
19909	0.107	5.200
199010	0.070	1.960
199011	0.083	1.580
199012	-0.039	1.590
19911	-0.015	20.800
19912	-0.073	0.700
19913	0.083	11.390
19914	0.095	8.320
19915	0.208	12.140
19916	0.073	3.110
19917	0.101	17.400
19918	0.101	9.400
19919	0.156	1.810
199110	0.116	0.980
199111	-0.009	0.680
199112	-0.088	1.510
19921	-0.061	6.300
19922	0.034	8.940
19923	-0.018	
19924	-0.018	1.020
19925	-0.033	
19926	0.104	
19927	0.028	2.270
19928	0.104	13.710
19929	0.140	5.840
199210	0.148	7.620
199211	0.084	5.970
199212	0.043	1.560
19931	0.065	6.790

Year and Month	Water Level (m)	Precipitation (inches)
19932	0.024	3.920
19933	-0.007	5.560
19934	0.054	0.920
19935	0.073	0.330
19936	0.032	3.950
19937	0.007	2.880
19938	0.050	9.880
19939	0.127	3.350
199310	0.149	6.170
199311	-0.021	4.840
199312	-0.041	2.750
19941	-0.128	6.530
19942	-0.049	2.430
19943	0.003	5.260
19944	0.057	1.390
19945	0.014	1.410
19946	0.049	12.810
19947	0.281	12.530
19948	0.220	15.140
19949	0.205	9.080
199410	0.226	8.660
199411	0.045	0.520
199412	0.019	1.420
19951	-0.055	3.390
19952	-0.105	2.080
19953	0.061	3.810
19954	0.060	3.530
19955	0.088	4.520
19956	0.047	9.350
19957	0.075	4.540
19958	0.214	7.970
19959	0.208	0.650
199510	0.202	3.190
199511	0.013	7.520
199512	-0.050	2.530
19961	-0.079	2.780
19962	-0.052	3.900

Year and Month	Water Level (m)	Precipitation (inches)
19963	-0.035	10.620
19964	0.006	4.250
19965	0.005	0.560
19966	-0.008	1.990
19967	0.039	10.490
19968	0.101	11.240
19969	0.140	7.140
199610	0.176	20.520
199611	0.080	1.110
199612	-0.019	5.470
19971	-0.047	5.780
19972	-0.022	4.830
19973	0.045	0.700
19974	-0.003	5.000
19975	0.016	1.430
19976	0.062	3.240
19977	0.072	9.770
19978	0.084	6.430
19979	0.134	2.290
199710	0.110	8.290
199711	0.000	4.680
199712	-0.072	7.090
19981	0.037	6.500
19982	0.129	9.880
19983	0.246	8.160
19984	0.253	1.460
19985	0.100	0.480
19986	0.004	0.200
19987	-0.001	1.560
19988	0.075	5.990
19989	0.322	18.010
199810	0.152	0.200
199811	0.038	1.200
199812	-0.026	1.220
19991	0.001	4.460
19992	0.034	1.690
19993	0.023	3.560

Year and Month	Water Level (m)	Precipitation (inches)
19994	0.030	2.160
19995	0.080	3.770
19996	0.064	5.930
19997	0.071	2.330
19998	0.121	5.250
19999	0.197	4.300
199910	0.115	3.670
199911	0.026	2.830
199912	-0.038	2.670
20001	-0.101	3.010
20002	-0.080	0.980
20003	0.019	4.500
20004	0.028	0.510
20005	0.044	0.010
20006	0.068	3.400
20007	0.098	3.370
20008	0.099	3.040
20009	0.155	17.980
200010	0.086	0.500
200011	0.095	6.780
200012	-0.098	1.500
20011	-0.176	1.880
20012	-0.152	0.550
20013	0.004	11.830
20014	0.013	0.140
20015	0.019	0.000
20016	0.016	7.840
20017	0.075	8.650
20018	0.128	9.910
20019	0.120	5.200
200110	0.087	2.530
200111	0.088	1.500
200112	0.007	2.730
20021	-0.121	8.010
20022	-0.093	0.820
20023	-0.042	4.970
20024	0.034	0.700

Year and Month	Water Level (m)	Precipitation (inches)
20025	0.043	2.410
20026	0.093	10.390
20027	0.057	8.050
20028	0.124	6.610
20029	0.265	7.350
200210	0.185	5.020
200211	0.007	2.230
200212	-0.062	3.360
20031	-0.118	0.160
20032	-0.046	5.750
20033	0.108	6.180
20034	0.070	1.390
20035	0.101	2.170
20036	0.138	13.160
20037	0.165	7.800
20038	0.190	9.050
20039	0.223	5.540
200310	0.181	6.670
200311	0.126	3.520
200312	-0.012	2.170
20041	-0.043	1.450
20042	-0.036	6.080
20043	-0.021	0.060
20044	-0.030	1.280
20045	0.059	0.400
20046	0.086	6.080
20047	0.075	5.510
20048	0.086	4.940
20049	0.148	5.180
200410	0.183	3.920
200411	0.152	5.060
200412	-0.060	3.150
20051	-0.012	1.930
20052	0.069	5.580
20053	0.009	8.500
20054	0.228	9.330
20055	0.120	4.760

Year and Month	Water Level (m)	Precipitation (inches)
20056	0.202	5.280
20057	0.249	3.820
20058	0.212	6.460
20059	0.241	1.310
200510	0.110	1.110
200511	0.039	4.370
200512	-0.049	1.960
20061	-0.046	1.820
20062	-0.056	4.900
20063	0.023	0.400
20064	0.033	1.750
20065	0.060	3.950
20066	0.074	4.930
20067	0.035	2.580
20068	0.075	3.890
20069	0.159	6.620
200610	0.123	2.180
200611	-0.007	2.300
200612	-0.065	8.250
20071	-0.080	4.580
20072	-0.086	4.110
20073	-0.053	1.100
20074	-0.034	1.710
20075	0.097	0.180
20076	0.080	0.410
20077	0.067	3.700
20078	0.125	2.960
20079	0.145	9.140
200710	0.165	5.490
200711	0.016	2.830
200712	-0.030	1.010
20081	-0.087	3.750
20082	-0.044	2.880
20083	-0.002	3.460
20084	0.039	1.620
20085	0.086	2.920
20086	0.061	2.420

Year and Month	Water Level (m)	Precipitation (inches)
20087	0.111	4.490
20088		11.27
20089		0.2
200810	0.154	3.92
200811	0.055	2.58
200812	0.009	2.01
20091	-0.106	1.5
20092	-0.107	2.98
20093	0.054	4.03
20094	0.219	7.4
20095	0.09	1.86
20096	0.1	3.62
20097	0.113	5
20098	0.175	12.08
20099	0.296	13.15
200910	0.218	1.5
200911	0.18	5.03
200912	0.173	7.66
20101	0.043	6.05
20102	0.067	3.3
20103	0.028	4.83
20104	0.033	0.7
20105	0.127	3
20106	0.158	5.04
20107	0.189	9.03
20108	0.218	5.55
20109	0.284	2.95
201010	0.123	1.05
201011	0.091	3.95
201012	-0.144	1.6
20111	-0.129	4.96
20112	-0.106	3.34
20113	0.014	4.91
20114	0.033	0.74
20115	0.046	0.31
20116	0.102	1.4
20117	0.167	10.08

Year and Month	Water Level (m)	Precipitation (inches)
20118	0.142	3.63
20119	0.189	7.27
201110	0.099	7.63
201111	0.097	1.41
201112	0.019	1.75
20121	-0.048	2.72
20122	-0.014	2.99
20123	0.06	2.11
20124	0.091	1.31
20125	0.15	0.79
20126	0.287	21.6
20127	0.197	8.03
20128	0.262	9.99
20129	0.228	5.17
201210	0.151	0.65
201211	0.139	0.27
201212	0.063	1.83
20131	0.002	0.8
20132	0.043	9.73
20133	0.055	3.91
20134	0.092	3.46
20135	0.101	1.34
20136	0.148	3.88
20137	0.241	12.31
20138	0.219	7.87
20139	0.257	6.06
201310	0.235	2.62
201311	0.127	5.14
201312	0.056	6.6
20141	-0.045	5.48
20142	0.016	6.22
20143	0.03	7.32
20144	0.139	6.88
20145	0.151	1.28
20146	0.13	3.46
20147	0.081	4.61
20148	0.15	3.95

Year and Month	Water Level (m)	Precipitation (inches)
20149	0.229	5.53
201410	0.19	4.39
201411	0.064	5.16
201412	0.132	3.64
20151	0.009	4.94
20152	0.006	4.52
20153	-0.002	1.41
20154	0.095	3.7
20155	0.135	0
20156	0.112	6.89
20157	0.123	5.33
20158	0.209	4.68
20159	0.225	3.69
201510	0.314	1.88
201511	0.247	14.3