Determining Trends in Water Quality Using High Resolution Land Use Data

David L. Bouck
University of Miami, dlbouck@gmail.com

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_theses

Recommended Citation
https://scholarlyrepository.miami.edu/oa_theses/693

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Theses by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.
UNIVERSITY OF MIAMI

DETERMINING TRENDS IN WATER QUALITY USING HIGH RESOLUTION LAND USE DATA

By

David L. Bouck

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

December 2017
UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

DETERMINING TRENDS IN WATER QUALITY USING HIGH RESOLUTION
LAND USE DATA

David L. Bouck

Approved:

Maria Estevanez, M.A., M.B.A
Senior Lecturer Department of
of Marine Ecosystems and Society

Gary Hitchcock, Ph.D.
Associate Professor Department
of Marine Biology and Ecology

Jill Richardson, Ph.D.
Senior Lecturer Department of
Marine Ecosystems and Society

Guillermo Prado, Ph.D.
Dean of the Graduate School

Christopher Kelble, Ph.D.
Deputy Director
Ocean Chemistry and Ecosystems Division
NOAA Atlantic Oceanographic & Meteorological Laboratory
Miami, Florida
Persistent macro-algal blooms have contributed to the decline of seagrass habitats throughout Biscayne Bay. South Florida’s canal system results in point-sources for excess nutrients that have collected from various types of anthropogenic activities in the watershed. One of the primary goals of the NOAA Habitat Blueprint Program’s Biscayne Bay Habitat Focus Area is to understand and develop mitigation strategies to combat excessive nutrient loading and increased macro-algal growth in the bay. This exploratory study utilized high resolution datasets and GIS spatial analysis techniques to analyze nutrient loading trends and relationships with adjacent land use in the Coral Gables Waterway. Nutrient concentrations throughout the canal displayed a high-low gradient from upstream to downstream sites most likely caused by physical barriers and seawater mixing. Total area grass, grass mean patch size, estimated population density, and the proximity to storm water drains within 250, 500, and 1000 meter buffer zones showed significant positive correlations with nitrogen and phosphorus concentrations. With refinement, this exploratory method could prove to be an effective means of identifying areas for further study and targeted mitigation strategies. The continued use of an intensive sampling regime in the Coral Gables Waterway is highly recommended, as it offers an invaluable dissection of the unique physical and chemical characteristics.
that govern nutrient loading into Biscayne Bay and is of the appropriate spatial resolution
to link with land-use and nutrient loading.
Acknowledgements

First, I would like to thank my committee chair Maria Estevanez for her unparalleled guidance, mentorship, support, and friendship over the past three years. My time in the Marine Ecosystems and Society Department has been a truly enlightening and rewarding experience. I would also like to thank my other committee members Dr. Chris Kelble and Dr. Gary Hitchcock for their invaluable insight and support while developing my project.

I would like to acknowledge my co-workers and friends within the National Oceanic and Atmospheric Administration who helped throughout this process. In particular, I would like to thank Dr. Joan Browder of the Southeast Fisheries Science Center for her guidance and wisdom as my internship coordinator. Also, a big thank you to Lindsey Visser of the Atlantic Oceanographic and Atmospheric Laboratory. She introduced me to a host of new fieldwork opportunities that formed the impetus for this thesis, and provided critical insight throughout its formulation. Additionally, thank you to Dr. Roland Samimy and the University of Massachusetts-Dartmouth Coastal Systems Program who developed and helped implement the Coral Gables waterway intensive water quality monitoring regime in partnership with NOAA.

Finally I would like to thank my friends and family from RSMAS and beyond for their immeasurable love and support. Thank you to Kelly Peebles for being my greatest supporter throughout everything. Without her, and obviously, all of the people above, none of this would have been possible. Last but not least, I would like to dedicate this thesis to my family. We can’t all be here to receive my gratitude, but as always, you’re eternal love and support is what drives me forward. Thank you.
# TABLE OF CONTENTS

LIST OF FIGURES ..................................................................................................... vi

LIST OF TABLES ....................................................................................................... ix

Chapter

1 INTRODUCTION ................................................................................................ 1
  1.1 Watershed Land Use and History ................................................................. 1
  1.2 Nutrient Concentrations and Distributions across Biscayne Bay ............... 2
  1.3 Ecological Impacts of Excessive Nutrient Loading in Biscayne Bay ......... 4
  1.4 Land Use Impacts on Biscayne Bay Water Quality ..................................... 5
  1.5 Land Use and Land Cover Metrics for Assessing Impacts to Water Quality .......................................................................................................................... 6
  1.6 Methodological Review ............................................................................... 8

2 PURPOSE ........................................................................................................ 10

3 METHODS ........................................................................................................ 12
  3.1 Study Area ................................................................................................... 12
  3.2 Trend Analysis of Coral Gables Intensive Sampling Regime ................... 12
  3.3 Land Use and Water Quality Analysis within the Coral Gables Waterway .......................................................................................................................... 14

4 RESULTS ........................................................................................................ 20
  4.1 Coral Gables Waterway Trend Analysis ..................................................... 20
  4.2 Land Use and Water Quality Analysis: OLS Results ............................. 21
  4.3 Comparison of GWR and OLS Models .................................................... 22

5 DISCUSSION .................................................................................................... 23
  5.1 Coral Gables Waterway Trend Analysis ..................................................... 23
  5.2 Land Use Impacts on CGW Water Quality Trends: Nitrogen Compounds ................................................................................................................. 25
  5.3 Phosphorus ................................................................................................ 27
  5.4 Limitations: Departures from Normality .................................................. 28
  5.5 Limitations: Impacts of Land Use Distribution and Classification Error... 29

6 RECOMMENDATIONS .................................................................................... 31
  6.1 Sampling and Model Improvement ............................................................ 31
  6.2 Future Study ............................................................................................... 31
  6.3 Management .............................................................................................. 32
List of Figures

Figure 1  Map of south Florida showing the location of the Biscayne Bay Watershed .................................................................................................................. 39

Figure 2a Map of land cover in south Florida. Land cover classifications are a product of the Multi-Resolution Land Characteristics Consortium National Land Cover Database. The Urban Development Boundary is delineated in black. ................................................................................. 40

Figure 2b Percentages of urban, agricultural, and wetland land cover (± SE) in the Biscayne Bay Watershed from 2001-2011. ......................................................... 41

Figure 3 Map of the Biscayne Bay Watershed showing the locations of canals and the approximate boundaries of North Bay, Central Bay, and South Bay ........................................................................................................... 42

Figure 4 Map of south Florida showing the location of the Biscayne Bay Habitat Focus Area. ........................................................................................................... 43

Figure 5 Satellite imagery showing the location of the Coral Gables Waterway. 44

Figure 6 Map of the Coral Gables Waterway showing the locations of sampling sites CGW-01 through CGW-11. The location of US-1 Highway is also included ........................................................................................... 45

Figure 7 Orthographic imagery used in land use classifications, masked to the area surrounding Coral Gables Canal ............................................................................... 46

Figure 8 Maximum Likelihood Classification of the study area. Bodies of water removed.................................................................................................................. 47

Figure 9 Example of re-sampled water quality points with 250 meter site buffers ......................................................................................................................... 48

Figure 10 Average nutrient concentrations (mg/L) at all stations. Error bars represent standard error .................................................................................................. 49

Figure 11 NO3, N+N, and DIN average concentrations (mg/L) at all stations. Error bars represent standard error .................................................................................... 50

Figure 12 NH4 and NO2 average concentration (mg/L) at all stations. Error bars represent standard error ............................................................................................ 51

Figure 13 PO4, TDP, and DOP average concentrations (mg/L) at all stations. Error bars represent standard error .................................................................................... 52
Figure 14  Average salinity level (psu) at all stations. Error bars represent standard error ................................................................. 53

Figure 15  Chlorophyll-a average concentration (ug/L) at all stations. Error bars represent standard error .................................................... 54

Figure 16  Dissolved Oxygen average concentration (ug/L) at all stations. Error bars represent standard error .................................................. 55

Figure 17  Principle Component Analysis of average nutrient values per station. Red arrows represent Eigenvectors ........................................... 56

Figure 18a  XY scatterplot of Chlorophyll-a against each explanatory variable. 250 meter scale ................................................................. 57

Figure 18b  XY scatterplot of Chlorophyll-a against each explanatory variable. 500 meter scale ........................................................................ 58

Figure 18c  XY scatterplot of Chlorophyll-a against each explanatory variable. 1000 meter scale ................................................................. 59

Figure 19a  XY scatterplot of Dissolved Inorganic Nitrogen against each explanatory variable. 250 meter scale ................................................. 60

Figure 19b  XY scatterplot of Dissolved Inorganic Nitrogen against each explanatory variable. 500 meter scale ................................................ 61

Figure 19c  XY scatterplot of Dissolved Inorganic Nitrogen against each explanatory variable. 1000 meter scale ........................................... 62

Figure 20a  XY scatterplot of Total Dissolved Phosphorus against each explanatory variable. 250 meter scale .................................................. 63

Figure 20b  XY scatterplot of Total Dissolved Phosphorus against each explanatory variable. 500 meter scale ................................................ 64

Figure 20c  XY scatterplot of Total Dissolved Phosphorus against each explanatory variable. 1000 meter scale ........................................... 65

Figure 21a  XY scatterplot of Silicon against each explanatory variable. 250 meter scale ................................................................. 66

Figure 21b  XY scatterplot of Silicon against each explanatory variable. 500 meter scale ................................................................. 67

Figure 21c  XY scatterplot of Silicon against each explanatory variable. 1000 meter scale ................................................................. 68
Figure 22a  XY scatterplot of Dissolved Oxygen against each explanatory variable.  
250 meter scale ................................................................................................. 69

Figure 22b  XY scatterplot of Dissolved Oxygen against each explanatory variable.  
500 meter scale ................................................................................................. 70

Figure 22c  XY scatterplot of Dissolved Oxygen against each explanatory variable.  
1000 meter scale ................................................................................................. 71

Figure 23  Recorded flow rate of water in cubic feet per second through the G-93 
control gate on days when water samples were collected in the CGW.  
Blue bars represent the period of sample collection ........................................... 72

Figure 24a  Chlorophyll-a OLS standard residual plots for composition metrics at 
the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.............. 73

Figure 24b  Chlorophyll-a OLS standard residual plots for configuration metrics at 
the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.............. 74

Figure 25a  Dissolved Inorganic Nitrogen OLS standard residual plots for 
composition metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 
meter analysis scales ......................................................................................... 75

Figure 25b  Dissolved Inorganic Nitrogen OLS standard residual plots for 
configuration metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 
meter analysis scales ......................................................................................... 76

Figure 26a  Total Dissolved Phosphorus OLS standard residual plots for 
composition metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 
meter analysis scales ......................................................................................... 77

Figure 26b  Total Dissolved Phosphorus OLS standard residual plots for 
configuration metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 
meter analysis scales ......................................................................................... 78
List of Tables

Table 1a  Methodological review resources 1-10. Matrix of timeframes studied, key explanatory and response variables, and analysis methods used in previous studies ................................................................. 79

Table 1b  Methodological review resources 11-20. Matrix of timeframes studied, key explanatory and response variables, and analysis methods used in previous studies ................................................................. 80

Table 2  Pearson correlation matrix of response variables. Pearson’s correlation coefficients greater than .50 (bold) indicate multicollinearity between variables ................................................................. 81

Table 3  OLS coefficients, adjusted R2 values, and GWR comparisons with AIC scores for the 250 meter analysis area. Bold values indicate a statistically significant relationship (p-value < 0.05). An asterisk next to nutrient header denotes non-normal residual distribution in that model ........................................................................................................ 82

Table 4  OLS coefficients, adjusted R2 values, and GWR comparisons with AIC scores for the 500 meter analysis area. Bold values indicate a statistically significant relationship (p-value < 0.05). An asterisk next to nutrient header denotes non-normal residual distribution in that model ........................................................................................................ 83

Table 5  OLS coefficients, adjusted R2 values, and GWR comparisons with AIC scores for the 1000 meter analysis area. Bold values indicate a statistically significant relationship (p-value < 0.05). An asterisk next to nutrient header denotes non-normal residual distribution in that model ........................................................................................................ 84

Table 6  Non-parametric Spearman correlation of water quality variables and land use metrics at the 250, 500, and 1000 meter scale. The sign of the coefficient indicates the direction of the relationship. Bold values indicate a significant relationship (p-value < 0.05) ........................................................................................................ 85

Table 7  Comparison of results from OLS and Spearman tests across all three scales. Green squares represent a matching significant relationship between both tests. A positive or negative sign indicates the direction of the relationship ........................................................................................................ 86
Chapter 1.0 Introduction

1.1 Watershed Land Use and History

Water quality issues in Biscayne Bay are inextricably linked to adjacent land use history. The Biscayne Bay watershed is located on the southeastern coast of the Florida peninsula and encompasses an area of approximately 2500 km² (Fig. 1). The majority of this watershed lies within Miami-Dade County; it includes the city of Miami, a small portion of Broward County, and borders Everglades National Park. Biscayne Bay is home to a diverse and sensitive estuarine ecosystem. Freshwater flow into the bay has facilitated the growth of extensive mangrove and seagrass habitats, both of which provide a host of important ecosystem services to the adjacent human communities. Historically, abundant summer rainfall drove sheet flows across south Florida. However, from 1907 to 1928, canals were constructed throughout the watershed to control seasonal flooding and to open up land for agriculture and metropolitan development. These canals intercept natural sheet flows and disrupt the timing and distribution of fresh water supplies to the Bay. The resultant irregularity of freshwater discharge into the bay has caused wide fluctuations in salinity (Kruczynski, 2012).

Bolstered in part by canal installation and the associated conversion of swampland to dry land, the Biscayne Bay watershed experienced rapid agricultural and urban development throughout the 1900’s. Between 1900 and 1980, Miami-Dade County’s population increased from approximately 5000 to 1.6 million individuals (US Census, 2016). The largest amount of growth occurred in the northeastern portion of the watershed, in the Miami metropolitan area. Drainage of the wetlands in the south led to agricultural development and the establishment of the American “winter vegetable...
kingdom” (Kruczynski, 2012). By the second half of the century, a distinct north-south gradient in land use existed across the watershed between urban high-density and agriculture use.

The establishment of Everglades National Park in 1947 and the Urban Development Boundary in 1975 curtailed the rate of urban expansion to the west and south. But in recent decades, sustained population growth and urban sprawl have begun to encroach on undeveloped and agricultural lands (Labiosa et al., 2009) (Fig. 2). Current US Census statistics estimate Miami-Dade’s population at approximately 2.6 million individuals and projects 3.2 million by the year 2030. This burgeoning population has imposed increasing pressure on state and county planners to expand on current development boundaries (Labiosa, 2009). The diversion of fresh water flows combined with rapid development have contributed to a decline in adjacent Biscayne Bay water quality and overall ecosystem health throughout the past century (Browder et al., 2005).

1.2 Nutrient Concentrations and Distributions across Biscayne Bay

Biscayne Bay is a shallow, semi-enclosed estuarine ecosystem. It occupies an area of approximately 700 km², with an average depth of 1.8 meters, and a maximum depth of 4 meters (Brand et al., 1991). Past research on nutrient and water flow analysis has separated the bay into three primary regions: North Biscayne Bay (NBB), Central Biscayne Bay (CBB), and South Biscayne Bay (SBB) (Alleman, 1982; Boyer, Kelble, Ortner, & Rudnick, 2009; Briceno, Boyer, Castro, & Harlem, 2013; Caccia & Boyer, 2005, 2007; SFWMD, 2013) (Fig. 3). NBB, which extends from Dumfounding Bay to Rickenbacker Causeway, is characterized by high-density urban development, industry, dense marine traffic, extensive dredging, and both natural and man-made barriers. CBB,
which extends from Rickenbacker Causeway to Featherbed Bank, is subject to both urban sprawl and agricultural land use. SBB, which extends from Featherbed Bank to Card Sound, is influenced primarily by agricultural land use and natural barriers (Boyer et al., 2009; Caccia & Boyer, 2005, 2007; SFWMD, 2013).

Overall water quality in the northern region is poor. NBB exhibits the highest levels of ammonium (NH$_4^+$; commonly associated with urban sewage pollution and wastewater treatment), as well as high turbidity, low levels of light attenuation, and high levels of primary productivity (Alleman, 1982; Caccia & Boyer, 2005; Boyer et al., 2009). Nutrient distribution in the Central Bay region is influenced by seawater exchange with the Atlantic Ocean (Alleman, 1982; Caccia & Boyer, 2005; SFWMD, 2013; Stalker, Price, & Swart, 2009). Consequently, nutrient concentrations are typically lower in this region. CBB also receives the least amount of freshwater input from adjacent canal systems, which leads to higher average salinity levels throughout the year (Alleman, 1982; Caccia & Boyer, 2005; Caccia & Boyer, 2007; Stalker et al., 2009). SBB is characterized by high concentrations of nitrate (NO$_3^-$) and nitrite (NO$_2^-$), associated with intensive fertilizer use from South-Dade agriculture which enters SBB through adjacent canals (Alleman, 1982; Caccia & Boyer, 2005; Boyer et al., 2009).

Phosphorus (P) concentrations throughout the bay are relatively low, which suggests an overall P-limited ecosystem (Brand, 1988). Greater concentrations of P are observed in the northern region of the bay and account for increased levels of primary productivity in this area (Brand, 1988; Carey et al., 2011). Canals are believed to be the primary source of P in Biscayne Bay. Phosphorus that has leached into the groundwater
from nearby landfills in both NBB and SBB also contributes to total P loading (Brand, 1988; Caccia & Boyer, 2005; Caccia & Boyer, 2007).

Nutrient concentrations in Biscayne Bay fluctuate annually due to the seasonal variation in precipitation between summer (wet) and winter (dry) (Alleman, 1982; Brand, Gottfried, Baylon, & Romer, 1991; Caccia & Boyer, 2005, 2007; Carey et al., 2011; Stalker et al., 2009). Furthermore, canal loading and seawater mixing contribute to a high–low nutrient concentration gradient that extends from nearshore to offshore regions (Brand et al., 1991; Caccia & Boyer, 2005). This gradient also fluctuates seasonally with the availability of fresh water inputs into the bay.

1.3 Ecological Impacts of Excessive Nutrient Loading in Biscayne Bay

Excessive concentrations of nitrogen (N) and P in coastal ecosystems, termed eutrophication, can cause hypoxia, anoxia, and decreases in local biodiversity (NRC, 2000). Nutrient over-enrichment can also lead to persistent phytoplankton and benthic macro-algal blooms. Prolonged phytoplankton blooms increase turbidity, which subsequently decreases light penetration to underlying seagrass beds and hampers their growth (Kelble, Ortner, Hitchcock, & Boyer, 2005). Fast-growing macroalgae often form extensive sheets that cover benthic habitats, depriving them of sunlight. A persistent macro-algal bloom has the potential to displace seagrass as the dominant form of benthic vegetation in that area (NRC, 2000).

Collado-Vides et al. (2013) documented the spatial extent of a persistent green macro-algal bloom of the genus *Anadyomene* along the inshore area of Central Biscayne Bay. A compilation of submerged aquatic vegetation (SAV) sampling from 1999 to 2012 showed bloom densities persisting from 2005 through 2012 over an area of
approximately 60 km², with up to 75% coverage recorded at some sites. This extensive bloom led to a significant loss of seagrass coverage within its area of effect (Collado-Vides et al., 2013). When seagrass die-offs such as this occur, sediment previously held by seagrass roots is re-suspended into the water column. Nutrients are released from these sediments and contribute to excess nutrient loads. Increased turbidity also compounds the effects of shading. Additionally, the diversity of resident fauna declines after the loss of food and shelter once provided by the seagrass habitat (NRC, 2000).

1.4 Land Use Impacts on Biscayne Bay Water Quality

The relationship linking anthropogenic land use to excessive nutrient loading in coastal areas has been thoroughly documented (e.g., Caccia & Boyer, 2007; Carey et al., 2011a; Carey et al., 2011b; Graves et al., 2004; Griffith, 2002; Griffith et al., 2002; Lee et al., 2009; Tong & Chen, 2002; Tu et al., 2007; Tu & Xia, 2008; Wang & Yin, 1997). Tong and Chen (2002) conducted a statistical analysis of surface water quality within the Little Miami River Basin in Cincinnati, Ohio. They found a strong relationship between excessive N and P loading and urban/agricultural land uses when compared to natural landscapes. Agriculture produced the highest concentrations of N and P, followed by urban land use. Additionally, their estimation found that impervious lands associated with urban development produced 55 times more runoff volume than pervious surfaces—a key contributing factor to urban non-point source water pollution (Schueler, 2000). Graves et al. (2004) analyzed storm water quality adjacent to predominant land cover types within the Indian River Lagoon in southeast Florida. They also found a strong link between nutrient concentrations and land use, with the highest loads of total N and P produced from agriculture and intermediate levels produced from urban/residential areas.
Caccia & Boyer (2007) conducted an evaluation of nutrient loading in Biscayne Bay between 1994-2002 from canal, groundwater, and atmospheric sources. Their study found that the composition and concentration of dissolved inorganic nitrogen (DIN) and total phosphorus (TP) loadings from canals are influenced by their proximity to varying types of development. Nitrate + nitrite comprised 95% of DIN in south bay canals, which is indicative of fertilizer use in agriculture. In the northern area, higher levels of NH$_4^+$ and TP were indicative of urban land use.

It is clear that nutrient loading and their distribution throughout Biscayne Bay is influenced by adjacent land use (e.g., Caccia & Boyer, 2005; 2007; NRC, 1993; 2000). Non-point sources of nutrients (i.e., nitrogen and phosphorus) and pollutants from urban and agricultural development across the watershed accumulate in nearby canals (Caccia & Boyer, 2005; 2007). The timed release of fresh water, via levee systems built into the canals, results in pulsed, concentrated discharges that cause rapid salinity fluctuations and high nutrient loads (Browder et al., 2005). Consequently, the southeastern Florida canal system is a primary factor influencing local water quality in Biscayne Bay.

1.5 Land Use and Land Cover Metrics for Assessing Impacts to Water Quality

Traditional point source pollution in urban settings, such as sewage or industrial outflow, can be managed through wastewater treatment methods that remove excess nutrients and contaminants. However, non-point sources of pollution arguably pose a greater threat to urban water quality because they are much more difficult to monitor and control (NRC, 2000). One possible way to address this issue is to spatially correlate land use and land cover (LULC) classifications with water quality trends in canals. The results from this type of analysis can be used by resource managers to estimate the
downstream impacts of human development and non-point source pollution (Baker, 2003; Griffith, 2002; S.-W. Lee, Hwang, Lee, Hwang, & Sung, 2009). This technique utilizes widely available geographic information systems (GIS) software to estimate LULC from remote sensing images. LULC is mapped using metrics derived from image classification, such as percent impervious surface cover, area-weighted mean shape index, or landscape development density index (LDI). A regression analysis can then be performed using these metrics and one or more ecological indicator of water quality/ecosystem health (e.g., distribution of chlorophyll-a, turbidity, temperature, salinity, dissolved oxygen, or other pollutants) (e.g., Alvarez-Romero et al., 2014; Carey et al., 2011; Griffith, 2001; Lee et al., 2009; Wang & Yin, 1997; Tu & Xia, 2008; Tong & Chen, 2002).

Previous studies in South Florida have implemented this type of analysis. Carey et al. (2011) performed a regression analysis of LULC metrics with eight canal water quality monitoring sites within the Biscayne Bay watershed. The canal sites ranged from the northern to the southern extent of the watershed, incorporating the shift in anthropogenic land use from urban to agricultural. The specific indicators used were landscape metrics, percent imperviousness, and LDI. The strongest relationship observed was between total phosphorus loads and land use within a buffer of 1000 meters surrounding sampled canals. However, these same disturbance indicators were not as effective in predicting inorganic nitrogen loading, particularly in the agricultural and less urbanized areas of the watershed. These findings suggest that specific indicators are better suited for particular land-cover classes. For example, based on this study, percent impervious surface is best suited for measuring the extent of urban developmental
impact, but fails to draw correlations between water quality and anthropogenic use in less-developed areas at the local scale. This underlines the importance of selecting a proper indicator for comparison based on the land-cover characteristics of a study area.

1.6 Methodological Review

The exact metrics, ecological indicators, and statistical methodology involved in land use and water quality correlation studies are wholly dependent on the various physical and biological factors inherent to the proposed study area (NRC, 2000; Griffith, 2002; Lee et al., 2009; Baker, 2003). In order to determine a proper approach for assessing land-use and water-quality relationships specific to the characteristics of South Florida, a methodological review of 20 publications was completed (Table 1). This review assessed various methodologies used in each study based on five primary components: (1) analysis scale, (2) timeframe, (3) explanatory variables, (4) response variables, and (5) analysis techniques. The effectiveness of individual components was assessed based on the results of each study and from specific recommendations provided by the authors. Details of the methodological review can be found in Appendix A. A brief summary of overall results follows here.

In summary, LULC-water quality studies are subject to great variability in time and space. Careful consideration of the physical and environmental characteristics of a given study area is fundamental to its proper design. A fundamental factor is the quality of base data. The accuracy of classifications and subsequent statistical analyses are dependent on the resolution of the imagery from which they are derived. Similarly, inconsistent and non-intensive water quality sampling regimes may obscure more subtle interactions between response variables and LULC metrics, leading to greater variation in
results. Decisions of which scale, timeframe, explanatory/response variables, and statistical models to employ are dependent on the inherent resolution of the available data. Studies should make every effort to maximize the overall quality of all base datasets incorporated into their analysis.

Multiple scales should be incorporated in order to elucidate any relationships that are not up-scalable (Baker, 2003). Buffer size should be relative to the resolution of base imagery, as model strength has been shown to decrease with decreasing scale and image resolution (Sliva & Williams, 2001). The greater the overall timeframe the better, and seasonality should be specified where significant variations in precipitation, vegetation growth, or other environmental factors exist (Pratt & Chang, 2012). Explanatory variables should reflect the general characteristics of the landscape and the goals of a particular study. Over-generalization of land use classifications can lead to rough assessments of LULC-water quality relationships, and land cover should be separated into constituent components if data resolution allows. Additionally, environmental variables such as slope or population density, as well as pattern metrics, all exhibit significant correlations with water quality. Ideally, a suite of metrics should be incorporated into a study’s design if possible. Response variables should be selected carefully based on the quality of data available, and the goals of the study. Finally, spatially explicit statistical methods such as Geographically Weighted Regression have shown to be more effective against traditional linear models when tested (Pratt & Chang, 2012; Jun Tu & Xia, 2008; J. Tu, Xia, Clarke, & Frei, 2007). Additional tests for normality, multicollinearity, and spatial autocorrelation, should be utilized for studies of this type.
Chapter 2.0 Purpose

The NOAA Habitat Blueprint Initiative is a multi-agency, targeted approach to marine resource management. The goal of this initiative is to focus collaborative efforts towards habitat and biodiversity conservation, fishery sustainability, ecosystem health, and socio-economic stability within important marine areas (NOAA, 2015). In 2014, NOAA leadership selected Biscayne Bay as a Habitat Focus Area (HFA) (Fig. 4). A key component of the proposed implementation plan for the Biscayne Bay HFA is to address the degradation of water quality in the bay and the subsequent increase in frequency and persistence of algal blooms (NOAA, 2015).

The ability of local resource managers to develop best management practices relies on our understanding of the complex relationships between human activities and nutrient loading. Past research has shown that land-use-water-quality interactions are highly variable in space and time (e.g., Yi et al., 2016; Pratt & Chang, 2012; Sliva & Williams, 2001). Therefore, the many unique environmental characteristics of Biscayne Bay and its adjacent watershed should be carefully considered. In keeping with the ecological and collaborative goals of the Habitat Blueprint Initiative, this study investigated different methods for assessing the impacts of anthropogenic activities on water quality indicators and the occurrence of harmful macro-algal blooms in the Bay. Its overall goal was to explore time and cost efficient procedures that utilize existing NOAA water quality monitoring programs, as well as modern GIS spatial analysis techniques, to analyze the connection between local land use, water quality, and macro-algal growth. As such, this study is divided into two primary components: (1) an analysis of water quality trends in the Coral Gables Waterway using data from an intensive
sampling regime implemented by University of Massachusetts-Dartmouth and NOAA, and (2) an exploratory analysis using GIS to assess correlations between riparian land use and water quality trends in the Coral Gables Waterway. Each section contributes to three overall goals:

1.) Identify a timely and cost efficient methodology for assessing the connection between land use and water quality at the riparian scale, which utilizes GIS and publicly available data.

2.) Conduct an exploratory analysis of correlations between different types of land cover and nutrient trends in the Coral Gables Waterway

3.) Generate recommendations for model improvement, future study, and potential mitigation strategies.
Chapter 3.0 Methods

3.1 Study Area

Coral Gables Waterway (CGW) is approximately 13 kilometers in total length. It runs southeast through Coral Gables, FL and drains into Central Biscayne Bay (Fig. 5). Between SW 40th St. and Blue Road, CGW splits into two branches—CGW east (4.5 km) and CGW west (4 km)—These two branches merge together two kilometers northwest of the canal mouth. Water flow from upstream of CGW is controlled by the South Florida Water Management District (SFWMD) G-93 control structure, which is located at the head of waterway, approximately six kilometers inland. During the dry season (November-April), CGW has little to no flow from above the G-93 gate and water movement in the canal is almost entirely governed by tidal forcing. Land use surrounding CGW can be described as medium density urban development, characterized primarily by residential properties and commercial use. The northern extent of the canal runs through residential neighborhoods and along a golf course that covers an area of approximately 1 km². The center of the waterway is intersected by US-1 and commercial areas of high imperviousness. Land use surrounding the southern extent of the canal is almost entirely residential.

3.2 Trend Analysis of Coral Gables Intensive Sampling Regime

Monthly surface-water samples were collected between October 2016 and April 2017 at 11 stations located within the Coral Gables Waterway (Fig. 6). Sampling was conducted by staff and volunteers from NOAA Atlantic Oceanographic and Meteorological Laboratory, the University of Massachusetts – Dartmouth, and Ransom Everglades School. The specific parameters that were measured include chlorophyll-a
(Chla), phosphate (PO₄), ammonium (NH₄), nitrite (NO₂), nitrate (NO₃), nitrate + nitrite (N+N), dissolved inorganic nitrogen (DIN), silicon (Si), dissolved oxygen (DO), total dissolved phosphorus (TDP), and dissolved organic phosphorus (DOP).

Chla samples were collected by filtering 150 ml of surface water through Whatman Glass microfiber filters (grade GF/F). These filters were placed in 1.5 ml cryovials and stored in liquid nitrogen for transport. Samples were stored at -80°C and were analyzed within two to three weeks of initial collection. Following the method outlined in Shoaf & Lium (1976), Chla was measured using a TD-700 Turner fluorometer (Kelble et al., 2005). Nutrient samples were collected by filtering 8 ml of surface water through 0.2 um Watman PTFE syringe filters and were stored frozen. Nutrients were determined by gas segmented continuous flow colorimetric analysis on a Seal Analytical Autoanalyzer AA3 using EPA methods: 349.0 (NH₄), 353.2 (NO₂, NO₃), 365.5 (PO₄), and 366.0 (Si).

All of the samples were collected within 1-2 hours of each other, during an ebb tide. Furthermore, in order to minimize random variation caused by nutrient spikes from rain events and/or nutrient loading from areas outside of the sampling regime, sampling efforts were conducted on days with zero precipitation. Salinity, temperature, and DO measurements were recorded on-site using a YSI Pro 30 handheld multi-parameter meter. Chla, PO₄, NH₄, NO₂, NO₃, N+N, DIN, Si, DO, TDP, and DOP values were averaged at each station and graphed to examine trends. A principle component analysis of primary nutrient averages at monitoring station was conducted using R-Studio 3.3.2 statistical software to illustrate trends in water quality and station variance.
3.3 Land Use and Water Quality Analysis within the Coral Gables Waterway

High resolution, red-green-blue, orthographic aerial imagery of the Coral Gables Waterway was obtained from the USGS Earth Explorer website. This imagery was collected at a pixel resolution of one foot per pixel (0.3048 meters) for an urban mapping project contracted by USGS in February, 2012. GeoTiff files were downloaded, projected using PCS Albers, and mosaicked using the ArcMap Image Analysis toolset. A 1-kilometer buffer was generated around the sampled extent of the Coral Gables Waterway and the mosaicked imagery file was cropped to this extent in order to generate a refined and less data-intensive study area (Fig. 7). The high resolution orthographic imagery from 2012 and 2016 imagery obtained from Google Earth were visually compared, to check for discrepancies, by separating the study area and Google imagery into quadrants.

In order to quantify land cover throughout the study area, the orthographic imagery was classified. Bodies of water, including the canal, ocean area, and ponds, were removed from the imagery using the Image Analysis toolset in preparation for classification. In order to identify primary classes of land use, the ArcMap Principle Component Analysis tool was used to analyze the raster color bands for dominant trends in reflectivity. Additionally, a process of trial and error using the ArcMap Iso-Cluster Unsupervised Classification was conducted to help identify the primary forms of land cover within the study area. Three fundamental classifications were identified: (1) Impervious Surfaces, including buildings, roads, parking lots, sidewalks, and barren lots; (2) Grass Areas, including lawns, medians, grassy lots, and golf courses; and (3) Tree Canopy, including mangroves, deciduous and evergreen trees, hedges, and bushes.
Classification training samples were created using the ArcMap Image Classification tools. Approximately 600 training samples were collected with a minimum one million pixels per class. These samples were incorporated into a Maximum Likelihood Classification using the same pixel resolution as the input imagery (Fig. 8). Ten iterations of the Majority Filter and two iterations of the Boundary Clean tools were run on the output classification raster in order to aggregate isolated pixels. Classification accuracy was assessed by generating 100 random points over the output raster and appending the class value of underlying cells to these points. Due to the extremely high resolution of the original imagery, in-situ ground-truthing of the classified values was possible. 100 random points were generated across the study area and underlying classification values were appended to each point. Each classification value was compared to the original imagery through visual assessment. Overall accuracy of the classified image was 88% with a Cohen’s Kappa coefficient of 0.79 (0.70 threshold of acceptability). The classification raster was converted to a polygon shapefile for further analysis.

Past methods for assessing riparian water quality and land use relationships have utilized drainage basins derived from slope (Mehaffey et al., 2005), or upstream buffer areas (Carey et al., 2011; Pratt & Chang, 2012) to quantify and relate LULC characteristics to water quality sampling locations. A one to one relationship of land use and water quality is necessary for regression analysis, therefore aggregated LULC metrics for drainage areas are typically compared with average water quality per each station. Due to the low sample size (n = 11) but high spatial resolution (i.e., more points per area) of the input water quality dataset, an interpolation, re-sampling, and buffer technique was employed in this study. Average water quality values for each station
were calculated and appended to a point shapefile in ArcMap. A polygon shapefile
delineation of the Coral Gables Waterway was created by tracing the canal from the study
area imagery using feature creation tools. An interpolation raster for each water quality
parameter, masked to the extent of the waterway, was generated using the Kriging
method. The output raster cell size was set to 2.5 meters in order to enhance processing
time. One hundred random points were created within the extent of the waterway
shapefile, and underlying water quality attributes were appended to each point. These re-
sampled values were used for further regression analyses.

Based on findings from the methodological review, the regional characteristics of
this study area, and available data, three composition and three configuration land cover
metrics were selected as explanatory variables. The composition variables that were
selected are as follows: (1) total area of each class, (2) sum of storm water drainage
outfalls entering the canal, and (3) average estimated population density. Average
estimated population density and sum of outfalls are measures of development intensity.
Besides runoff from areas immediately adjacent to the canal, it is likely that most water
accumulates and is directed into the canal via drainage networks. Therefore, sum of
storm water drainage outfalls entering the canal and distance to nearest drainage outfall
were also included as explanatory variables. The configuration variables that were
selected are as follows: (1) mean patch size for each class, (2) patch frequency for each
class, and (3) distance to nearest drainage outfall. Mean patch size and patch frequency
are basic indicators of habitat fragmentation within a landscape, and were selected for
this reason. Distance to outfall was included in order to assess the potential impacts of
outfall proximity on nutrient levels. Mean slope was omitted from this analysis due to
the relatively flat landscape that surrounds the study area and the poor resolution of available digital elevation models. Seasonal and climatic variables such as average rainfall, temperature, etc., were not included in this analysis because of the limited timeframe of the data set, which is confined to the dry season of 2016-2017.

Each explanatory variable was quantified within a known area surrounding the re-sampled points with the use of variable distance buffers (Fig. 9). A similar study in this area used 500, 1000, and 1500 meter buffers (Carey et al., 2011). Their analysis yielded strong correlations at the 1000-meter scale only. Taking into account the high resolution data used for this study, 1000 meters was selected as the maximum distance and 500 meters as the mid-range distance. A 250-meter minimum distance was selected in order to capture variability in runoff pathways. This is due to the inability to measure the impact of slope or delineate drainage areas. The ArcMap Buffer tool was used to create 250, 500, and 1000 meter buffers for each re-sampled point. To calculate class area (i.e., the area of impervious surfaces, grass areas, and tree canopy), these three classes were isolated from the primary classification shapefile and extracted. Each buffer distance and individual class shapefile was then overlaid and combined using the ArcMap Intersect tool. Individual class area, mean patch size, and patch frequency were calculated within each buffer area using the Patch Analyst Extension for ArcMap (Rempel, 2012). The Summary Statistic tool was used to aggregate area, mean patch size, and patch frequency values by their corresponding buffer and sample point.

Estimated average population density was calculated using 2010 census data. 2010 census block shapefiles and total population counts for Miami-Dade County were retrieved from the US Census TigerLine online database. Population counts were
spatially joined to their respective block within ArcMap by their reference GeoID. Joined shapefiles were clipped to the extent of the Biscayne Bay Watershed and projected in PCS Albers. Census blocks were converted to point format and estimated population densities were calculated with an output cell size of 20 m² using the Point Density tool in ArcMap. Estimated average population density was calculated within each buffer by intersecting the buffer layers with the population density raster. The Summary Statistics tool was used to extract the average population density within each corresponding buffer and append them to their sample point.

A line shapefile of storm water drains throughout south Florida was acquired from the Miami-Dade GIS Open Data Site. The shapefile was clipped to the extent of the study area, and outfalls were selected and exported into a separate layer. This shapefile was intersected with the buffer layers, summarized by each buffer with Summary Statistics, and appended to each re-sampling station. The Near-Distance tool was used between the re-sampled points and outfall shapefile to calculate the nearest distance to outfall for each point. This value was automatically appended to the re-sampled points shapefile.

Combined water quality and land use attribute tables were exported from ArcMap into Microsoft Excel to be reorganized and compiled. To eliminate multicollinearity, a Pearson’s Correlation matrix of explanatory variables was run for each buffer dataset. Patch frequency was found to be significantly redundant with mean patch size and was removed from the analysis entirely. Composition and configuration variables were separated into groups in order to eliminate residual multicollinearity. Water quality data was assessed for normality using qq-plots and histograms. All water quality variables
were transformed using a natural log. Re-organized datasets were re-projected into
ArcMap for regression analysis.

The Ordinary Least Squares (OLS) regression tool in ArcMap was used for linear
regression. For three scales, two groups of explanatory variables per scale, and 11 water
quality response variables, 66 OLS models were created. To test the impacts of spatial
autocorrelation on model fit, a Geographically Weighted Regression (GWR) was
completed when applicable. In total, 55 GWR models were created. Adjusted R² values
and Aikake Information Criterion (AICc) values for both OLS and GWR models were
compiled for comparison. Several OLS models exhibited non-normal residual
distribution. As a diagnostic step, a Spearman’s Rank Correlation matrix was generated
between explanatory variables and un-transformed response variables. The OLS
coefficients and p-values were compared to the Spearman’s Rho and p-values to assess
possible issues of non-normality.
Chapter 4.0 Results

4.1 Coral Gables Waterway Trend Analysis

Average nutrient levels generally display a high to low concentration gradient progressing downstream (Fig. 10). Stations CGW-11, CGW-10, CGW-09, and CGW-08 exhibit the highest nutrient concentrations overall. The highest concentration of NO$_3$ (approx. 4.0 mg/L) occurred at CGW-09, and the second highest levels of NO$_3$ (approx. 2.5 mg/L) occurred at CGW-08, along the eastern pathway (Fig. 11). NO$_3$ makes up the majority of N+N and DIN levels as NH$_4$ and NO$_2$ levels remain comparatively low (< 0.2 mg/L) throughout the canal (Fig. 12). The highest concentration of NH$_4$ (approx. 0.17 mg/L) occurred at the northernmost station, CGW-11, adjacent to the G-93 control structure. NO$_2$ concentrations were highest at CGW-11 and CGW-10 (approx. 0.045 mg/L). PO$_4$ concentrations range from 0.06 – 0.08 mg/L between CGW-11 and CGW-08 and CGW-07 (east and west pathways), then steadily decline downstream (Fig. 13). Si exhibits a more linear pattern of decline, decreasing in concentration from 2.41 mg/L at CGW-11 to 0.25 mg/L at CGW-01 (Fig. 10).

Salinity levels were highest at the mouth of the canal (32 psu), declined sharply at CGW-08 and CGW-09, and were lowest at CGW-01 (4.7 psu) (Fig. 14).

Mean Chla concentrations were highest at the upstream stations (5.0-22.0 µg/L) and lowest at the downstream stations (1.0-2.0 µg/L) (Fig. 15). Downstream station CGW-06 was the only exception. Relative to the surrounding stations, the Chla levels observed at this location were uncharacteristically high (approx. 28 µg/L).
Results of the principle component analysis between water quality variables show a clustering of variance among the downstream stations, while the upstream stations are more distinct (Fig. 17).

4.2 Land Use and Water Quality Analysis: OLS Results

Results of the OLS analysis at the 250, 500, and 1000 meter scale are provided in Tables 3, 4, and 5, respectively. These tables are supplemented by XY scatterplots that illustrate trends between Chla (Fig. 18), DIN (Fig. 19), TDP (Fig. 20), Si (Fig. 21), DO (Fig. 22), and each of the explanatory variables at the 250, 500, and 1000 meter scale.

Total area of grass was positively correlated with Chla, PO4, NH4, NO2, NO3, N+N, DIN, Si, and DO. Estimated population density was positively correlated with PO4, NO3, N+N, DIN, Si, TDP, DOP, and negatively correlated with NH4 and NO2. Total area of impervious surfaces was positively correlated with Chla and Si, and total area of tree canopy was positively correlated with Si and TDP. Total area of tree canopy also exhibited a negative correlation with DIN and a positive correlation with DO at the 500 and 1000 meter scales. Sum of outfalls exhibited positive correlations with NH4, NO2, Si, and negative with TDP and DOP at the 250 and 500 meter scales.

Grass MPS and distance to outfall both displayed consistent significant correlations at all three scales. Positive correlations for grass MPS were found with Chla, NH4, NO2, DIN, Si, and DO, while TDP and DOP were negatively correlated with Grass MPS at all three scales. Additionally, grass MPS was positively correlated with PO4 NO3, and N+N at the 250 and 500 meter scales. Distance to outfall had a significant negative correlation with Chla, PO4, NO3, N+N, DIN, Si, TDP, DOP. A positive correlation with NH4 was also found at every scale. Impervious MPS had significant
positive correlations with Chla and DOP at every scale, along with NO₃, N+N, and DIN at the 250 and 500 meter scales, and DO at the 500 and 1000 meter scales. Tree MPS displayed negative correlations with NO₃, N+N, DIN, and Si at the 500 and 1000 meter scales, and positive correlations with TDP and DOP at the 250 and 500 meter scales.

4.3 Comparison of GWR and OLS Models

Of the 65 OLS models that were generated, 55 GWR comparisons were completed (Tables 3-5). At the 1000 meter scale, configuration metrics exhibited redundancy, creating issues with multicollinearity between variables. As a result, no comparisons were made for configuration OLS models at the 1000 meter scale. Adjusted $R^2$ values for GWR models were higher than that of OLS in every case. Similarly, AICc scores for GWR were lower than OLS in every comparison.
Chapter 5.0 Discussion

5.1 Coral Gables Waterway Trend Analysis

There is a clear high-low divide in overall nutrient concentrations between the upstream stations (CGW-11 through CGW-08) and the downstream stations (CGW-07 through CGW-01). This is most likely due to higher rates of tidal flushing and mixing at the sites south of US-1. Low levels of precipitation during the dry season reduce the volume and rate of water flow within the canal, both of which are further constricted by the G-93 control structure. This gate regulates the water supply for inland water tables by preventing water from flowing into the CGW when levels drop below a specific height (SFWMD, 2017). It is therefore likely that flow rate in the CGW is primarily governed by tidal gradients and influxes from outfalls and sporadic precipitation and runoff events during the dry season. However, it is also possible that the overpass structures of US-1 on both the East and West branches of CGW create a bottle neck that significantly reduces the influence of tidal flushing north of US-1. The combined effect would likely result in higher water residency times and increased nutrient concentrations at the northern sites.

Increased water residence time has been correlated with an increase in primary production and algal blooms (S. Lee et al., 2012; S. O. Lee, Kim, Kim, Lim, & Jung, 2014). As evidenced by the relatively high Chla concentrations that were measured at stations CGW-11 and CGW-06, these locations may be within or adjacent to areas where water tends to stagnate. CGW-11 is directly downstream from the G-93 control structure and is located farthest away from the mouth of the canal. The western branch of CGW, where CGW-06 is located, also happens to be particularly narrow. Flow rate could be
measured by releasing tracers into the canal for additional correlation with Chla and nutrient concentrations.

N+N and DIN levels in CGW were predominantly composed of NO₃. NO₃ measurements were greatest at CGW-09 and CGW-08, with average concentrations almost double at CGW-09 (4.02 mg/L). But it is unclear why NO₃ levels at CGW-09 are so high. A large storm water network outfall is located approximately 100 meters upstream of station CGW-10, which could be a nitrogen point source. However, the concentration of nitrogen compounds at station CGW-10 are lower than those of both CGW-08 and CGW-09, which are located approximately 650 meters and 1300 meters downstream of CGW-10, respectively. Another possibility is that adjacent land use between CGW-10 and CGW-09 are responsible. Fertilizers and human waste are common sources of NO₃ in freshwater streams (Graham, 2009). Lawn fertilization and irrigation of the nearby golf course and adjacent residential properties could be contributing to higher concentrations. The lower levels observed at CGW-10 and CGW-08 and the precipitous drop in NO₃ between CGW-09 and CGW-06 might indicate that levels observed at CGW-09 are site specific. The addition of two sample sites approximately 500 meters north and south of CGW-09 could help to elucidate this trend.

PO₄ levels increase from 0.06 mg/L at CGW-11 and peak at stations CGW-08 (0.082 mg/L) and CGW-07 (0.080 mg/L), before declining steadily to 0.01 mg/L at CGW-01. Unlike N compounds, PO₄ concentration along the eastern branch does not decline immediately after US-1. PO₄ measurements at CGW-06 on the western branch decline after US-1, but increased primary productivity between CGW-09 and CGW-06 may explain the drop in PO₄ at this location. Conversely, less primary productivity
observed along the eastern branch might explain the higher levels of PO4 observed further downstream. Peak levels of PO4 coincide with trends in NO3 and may be indicative of runoff from lawn fertilization and irrigation, storm water, and/or human waste pollution in this section of the canal. Adding additional stations above and below these transitionary areas might help to better explain observed trends.

The highest NH4 concentration of 0.175 mg/L was observed at CGW-11. Concentrations fluctuate between 0.08 and 0.02 mg/L for the remainder of the sampling area. NO2 displays the same high to low concentration gradient observed across the study area, with its peak concentration 0.049 mg/L at CGW-10. DO levels remained fairly consistent with no observable trend throughout the canal (Fig. 16). Si levels are highest at CGW-11 at 2.41 mg/L, and steadily decline throughout the canal, with a somewhat steeper gradient along the US-1 boundary. Declining Si levels throughout the GCW could be an indication of the extent of seawater flushing or assimilation by diatoms. Si is an abundant, naturally occurring element in fresh water, and concentrations in freshwater systems typically range from 0.7 mg/L to 7.0 mg/L (Graham, 2009). Si concentrations measured in the CGW fall toward the lower end of this scale. Nominal influx of freshwater from north of the G-93 control structure, low precipitation, and gradual tidal flushing during the dry season may contribute to declining concentrations of Si throughout CGW (Fig. 23). An additional comparative sample on the northern side of the G-93 gate may help to better understand this dynamic.

5.2 Land Use Impacts on CGW Water Quality Trends: Nitrogen Compounds

The strong correlation exhibited between N compounds and total area of grass suggests that this type of land use is a significant source of excess N in the CGW. Grass
areas in an urban environment consist almost entirely of managed lawns, parks, and sports fields. As a result, these areas are subject to regular mowing, fertilization, mulching, and irrigation. Total phosphorus and nitrogen concentrations in fertilizer varies depending on the type. Florida state regulations allow for 0.25 pounds of phosphorus per 1000 square feet in a single application and a maximum of .50 pounds per year. Nitrogen in fertilizers are regulated by type and the season. Slow-release nitrogen content cannot exceed 2.0 pounds per 1000 square feet and soluble quick release nitrogen content is limited to 0.7 pounds per 1000 square feet in the spring and summer (Trenholm, 2016). Miami-Dade County restricts lawn irrigation to morning and evening hours, two days per week, per household or business (Miami-Dade County, 2017). However, lack of compliance with regulations, improper edging, and overwatering, creates runoff. The runoff from these lawns flows into local storm water drains, which empty in nearby canals. If the residence time of water in the area of CGW north of US-1 is in fact greater, it would allow normally diffuse runoff from grassy areas to accumulate and account for the relatively high concentrations of NO₃ that were measured at those sites. Targeted exploratory sampling could be used to test this possibility. Irrigation in Miami-Dade is limited to Wednesday, Thursday, Saturday, and Sunday. Comparative sampling in areas with high N concentrations could be taken on Thursday and Tuesday of the following week to test for differences between watering and non-watering days.

Grass MPS also displayed significant correlations with N concentrations. Grass MPS is a basic indicator of habitat fragmentation—meaning that as MPS increases, habitats become less fragmented. The re-sampled points in CGW with the highest grass MPS were those points adjacent to the local golf courses. These results suggest that the
greater the average size of grassy area, the greater the increase in N concentrations. This would conform to previous findings that correlate golf course land use with declines in adjacent water quality (Graves, Wan, & Fike, 2004). Impervious MPS displayed positive correlations with NO$_3$ at the 250 and 500 meter scales, but total impervious area did not display any significant relationship. It is possible that the connectivity of impervious surfaces, indicated by increasing mean patch size, may have a greater impact on N loading than total area in general.

Estimated population density displayed significant positive correlations with NO$_3$, N+N, and DIN at all three scales, and negative correlations with NH$_4$ and NO$_2$. DIN and N+N are almost entirely comprised of NO$_3$, making them somewhat redundant. Population density is a rough measure of the intensity of a certain landscape. Higher values typically indicate residential areas and higher levels of human waste. A strong correlation with NO$_3$ and population density might represent the impact of residential land use, specifically lawn care and household wastewater. However, the negative correlation displayed between NH$_4$ and population density suggest the opposite, as NH$_4$ is typically associated with decomposing organic matter in the water column (NRC, 2000). A negative correlation between NH$_4$ and NO$_2$, grass area, grass MPS, and population density, might indicate that NH$_4$ and NO$_2$ concentrations are governed more by natural processes in CGW as opposed to human wastewater. Additionally, the positive correlation between NH$_4$ and the nearest distance to a storm water outfall might represent the greater influence of natural variability on NH$_4$ concentrations.
5.3 Phosphorus

PO$_4$ serves as a key limiting nutrient for primary productivity and is particularly pertinent to the goals of the NOAA Biscayne Bay HFA. Positive correlations between grass area, grass MPS at the 250 and 500 meter scale, and population density, as well as a negative correlation with storm water outfall distance, suggest that PO$_4$ concentrations in CGW are significantly influenced by runoff from lawns and residential land use. Negative correlations for TDP and DOP were found with grass MPS and distance to outfall. Positive correlations were found between estimated population density, area tree for TDP, and impervious MPS for DOP. Tree MPS was positively correlated with both at the 250 and 500 meter scales. These relationships mirror the relationships observed for PO$_4$ with the exception of grass MPS. Results of the trend analysis showed that TDP and DOP did not exhibit the same high-low concentration gradient from upstream to downstream as PO$_4$ and other nutrients in the CGW. One possible explanation for these trends is that these aggregate measures of PO$_4$, TDP, and DOP are more influenced by natural processes in the waterway than direct inputs from neighboring land use.

5.4 Limitations: Departures from Normality

Several of the OLS models exhibited non-normal residual distribution and may be unreliable (Table 3-4). Residual maps of OLS analyses illustrate the variance in predicted versus observed values surrounding stations that exhibited higher nutrient levels, namely, CGW-09, CGW-08, CGW-07 (DIN and TDP), CGW-11, and CGW-06 (Chl-a) (Fig. 24-26). Areas of higher nutrient content likely contributed to non-normal distributions among the input datasets. Post transformation analysis of the data showed that several variables—both explanatory and response—exhibited signs of non-normality.
OLS regression is robust against violations of normality. However, in this case, the
distribution of the data may have resulted in biased models. A non-parametric Spearman
Rank Correlation matrix was created as a diagnostic using the untransformed water
quality data and land use metrics at each scale (Table 6). Results of this analysis showed
roughly 60% equivalency between OLS and Spearman results at individual scales (Table
7). Across all scales, four primary trends were maintained: (1) the positive correlation of
grass area with Chla, PO4, NO3, N+N, and DIN; (2) the positive correlation of grass MPS
with Chla, NO3, N+N, and DIN; (3) the positive correlation of estimated population
density with PO4, N+N, DIN, TDP, DOP, and a negative correlation with NH4; and (4)
the negative correlation of distance to outfall with Chla, PO4, NO3, N+N, DIN, Si, and
TDP (Table 8). This cross-validation between two models reinforces the primary
findings of the OLS/GWR analysis.

5.5 Limitations: Impacts of Land Use Distribution and Classification Error

It is possible that the significant correlations between land use and water quality
variables found in this analysis may be the result of coincidence and classification
misrepresentation. For example, distribution of grassy area across the landscape is
heavily influenced by the presence of the golf course north of US-1. As a result, area
grass and grass MPS both demonstrate a high-low gradient from upstream to
downstream, similar to observed nutrient concentrations. This trend is a direct result of
classification misrepresentation due to exceptionally large and continuous tree canopies
in the older, more residential areas south of US-1. Classification accuracy tested against
the imagery may have been relatively high. However, the imagery itself does not
accurately describe the amount of impervious and grassy areas in the southern portion of
CGW. This imbalance coincides with observed nutrient concentrations and may be misrepresenting the results. In other words, it is possible that nutrient trends may be governed primarily by the physical dynamics of CGW and downstream seawater mixing, as examined in the initial trend analysis, and not by adjacent land use trends.
Chapter 6.0 Recommendations

6.1 Sampling and Model Improvement

Additional sampling sites throughout the CGW could significantly improve model results and better explain local observed trends in nutrient loading. Increasing sampling frequency to roughly once every 500 meters would increase the overall resolution of the data set. A larger initial sample size would also improve the performance of the interpolation and re-sampling method used in this study, as well as allow the possibility to attempt different methodologies using the raw data set. For trend analysis, comparative samples from specific areas or within certain periods could be used to help better explain the dynamics of nutrient distributions in CGW. For example, a sample on the upstream side of G-93 may help to elucidate the influence of the control structure on adjacent downstream water quality.

Incorporating new classification methods and dedicated software could improve the representation of the landscape. Data that delineates road length and width can be overlaid with current classifications to reduce the impact of abundant tree canopy on total impervious area. Use of a dedicated classification program such as ERDAS Imagine software, may also prove more accurate and efficient then using the default programs available through ArcMap.

6.2 Future Study

A duplicate comparison analysis for the wet season is essential to understanding nutrient loading trends in CGW. Understanding how seasonal variables such as increased precipitation and flow rates impact nutrient concentrations will help to explain the localized trends observed during the dry season. Additionally, CGW trends from both
the wet season and dry season could be compared to long-term water quality monitoring sites surrounding the mouth of the canal, in order to make the connection between land use, nutrient loading in the canal, and conditions observed in inshore areas of Biscayne Bay.

Understanding the pathway that runoff takes to enter the canal may also help to better describe which land use features have more impact on water quality. Removing the golf course from the classification scheme may help to show the relative influence this large area has on the overall results, and the strength of relationship between grassy areas and nutrient distributions. Additionally, the impact of storm water drainage networks may not have been accurately represented by site-specific buffer areas. Water transported along these networks could travel from beyond the land area analyzed in this analysis. Therefore, a comparison analysis should be completed that measures land use metrics in a buffer area surrounding continuous drainage networks that feed into CGW, to better gauge drainage impacts.

6.3 Management

Understanding the localized interactions between land use and nutrient loading into Biscayne Bay is essential to developing targeted, efficient, and cost effective mitigation strategies that will address damage caused by excessive macro-algal growth. Intensive sampling regimes such as the one designed for CGW could be an effective means to understanding the highly variable characteristics of canals that drain into the bay. The CGW regime is relatively low-cost and logistically easy to implement, as most of the samples can be taken from land. Comparison of similar analyses from other canals could vastly improve our understanding of riparian land use composition and
configuration impacts to water quality. Additionally, costs for acquiring up-to-date high-resolution aerial imagery could be cut significantly if sampling was limited to regions immediately surrounding specific canals.
Chapter 7.0 Conclusion

Development within the Biscayne Bay watershed has led to the gradual decline in water quality and a loss of critical seagrass habitat in part due to excess nutrient loading and macro-algal blooms. Intensive sampling regimes and the use of high resolution data can significantly improve our understanding of canal-specific dynamics and local interactions with adjacent land use and water quality parameters. However, development of focused mitigation strategies will require continued sampling for multiple years and wet season comparisons. The method employed in this study of imagery classification, interpolation, and re-sampling can be an effective technique for navigating issues of low sample size, if spatial resolution of the input dataset is high. Furthermore, relevant land use data are available to the public and the costs associated with data analysis are minimal—making this method a particularly viable option for local resource managers. Careful attention should be paid to the selection of explanatory variables, and analysis variables need to be tailored to the specific characteristics of the study area. Continued support of intensive sampling regimes within canals entering the Biscayne Bay can provide valuable information to inform the design of mitigation strategies.
References Cited


Fig. 1. Map of south Florida showing the location of the Biscayne Bay Watershed.
Fig. 2a. Map of land cover in south Florida. Land cover classifications are a product of the Multi-Resolution Land Characteristics Consortium National Land Cover Database. The Urban Development Boundary is delineated in black.
Fig. 2b. Percentages of urban, agricultural, and wetland land cover (± SE) in the Biscayne Bay Watershed from 2001-2011.
Fig. 3. Map of the Biscayne Bay Watershed showing the locations of canals and the approximate boundaries of North Bay, Central Bay, and South Bay.
Fig. 4. Map of south Florida showing the location of the Biscayne Bay Habitat Focus Area.
Fig. 5. Satellite imagery showing the location of the Coral Cables Waterway.
Fig. 6. Map of the Coral Gables Waterway showing the locations of sampling sites CGW-01 through CGW-11.
Fig. 7. Orthographic imagery used in land use classifications, masked to the area surrounding Coral Gables Canal.
Fig. 8. Maximum Likelihood Classification of the study area. Bodies of water removed.
Fig. 9. Example of re-sampled water quality points with 250 meter site buffers.
**Fig. 10** Average nutrient concentrations (mg/L) at all stations. Error bars represent standard error.
Fig. 11. NO$_3$, N+N, and DIN average concentrations (mg/L) at all stations. Error bars represent standard error.
Fig. 12. NH₄⁺ and NO₂⁻ average concentration (mg/L) at all stations. Error bars represent standard error.
Fig. 13. PO₄, TDP, and DOP average concentrations (mg/L) at all stations. Error bars represent standard error.
Fig 14. Average salinity levels (psu) at all stations. Error bars represent standard error.
FIG. 15. Chlorophyll-a average concentration (μg/L) at all stations. Error bars represent standard error.
Fig. 16. Dissolved Oxygen average concentration (ug/L) at all stations. Error bars represent standard error.
Fig. 17. Principle Component Analysis of average nutrient values per station. Red arrows represent Eigenvectors.
Fig. 18a. XY scatterplot of Chlorophyll-a against each explanatory variable. 250 meter scale.
Fig. 18b. XY scatterplot of Chlorophyll-a against each explanatory variable. 500 meter scale.
Fig. 18e. XY scatterplot of Chlorophyll-a against each explanatory variable. 1000 meter scale.
Fig. 19a. XY scatterplot of Dissolved Inorganic Nitrogen against each explanatory variable, 250 meter scale.
Fig. 19b. XY scatterplot of Dissolved Inorganic Nitrogen against each explanatory variable. 500 meter
Fig. 19c. XY scatterplot of Dissolved Inorganic Nitrogen against each explanatory variable. 1000 meter scale.
Fig. 20a. XY scatterplot of Total Dissolved Phosphorus against each explanatory variable. 250 meter scale.
Fig 20b. XY scatterplot of Total Dissolved Phosphorus against each explanatory variable. 500 meter scale.
Fig. 20c. XY scatterplot of Total Dissolved Phosphorus against each explanatory variable. 1000 meter scale.
Fig. 21a. XY scatterplot of Silicon against each explanatory variable. 250 meter scale.
Fig. 21b. XY scatterplot of Silicon against each explanatory variable. 500 meter scale.
Fig. 21c. XY scatterplot of Silicon against each explanatory variable. 1000 meter scale.
Fig 22a. XY scatterplot of Dissolved Oxygen against each explanatory variable. 250 meter scale.
Fig. 22b. XY scatterplot of Dissolved Oxygen against each explanatory variable, 500 meter scale.
Fig. 22c. XY scatterplot of Dissolved Oxygen against each explanatory variable. 1000 meter
Fig. 23. Recorded flow rate of water in cubic feet per second through the G-93 control gate on days when water samples were collected in the CGW. Blue bars represent the period of sample collection.
Fig. 24a. Chlorophyll-a OLS standard residual plots for composition metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.
Fig. 24b. Chlorophyll-a OLS standard residual plots for configuration metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.
Fig. 25a. Dissolved Inorganic Nitrogen OLS standard residual plots for composition metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.
Fig. 25b. Dissolved Inorganic Nitrogen OLS standard residual plots for configuration metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.
Fig. 26a. Total Dissolved Phosphorus OLS standard residual plots for composition metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.
Fig. 26b. Total Dissolved Phosphorus OLS standard residual plots for configuration metrics at the (a) 250 meter, (b) 500 meter, and (c) 1000 meter analysis scales.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Timeframe</th>
<th>Explanatory Variables</th>
<th>Response Variables</th>
<th>Methods of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gerard McMahon and Thomas F. Coffey (2000)</td>
<td>NA</td>
<td>TIA, DCIA, Demographic Intensity</td>
<td>Nutrients, Pesticides, Iron</td>
<td>Trend analysis</td>
</tr>
<tr>
<td>Yangfan Li, Yi Li, Wei Wu (2015)</td>
<td>2000, 2004, 2008</td>
<td>Class area, Pattern</td>
<td>15 WQ variables</td>
<td>Linear regression, Non-linear regression</td>
</tr>
<tr>
<td>Xudao Wang, Zhi-Yong Yin (1997)</td>
<td>1970-1980</td>
<td>Conductivity, Class area</td>
<td>Multiple WQ variables, Conductivity</td>
<td>Correlative analysis</td>
</tr>
<tr>
<td>Marina Alberti, Derek Booth, Kristina Hill, Bekkah Coburn, Christina Auoio, Stefan Coe, Danielle Spirkedell (2007)</td>
<td>1998</td>
<td>Class area, Land use, DCIA, SHDI</td>
<td>Beothic Index of Biological Integrity</td>
<td>Stepwise regression</td>
</tr>
<tr>
<td>Susanna T. Y. Tong (1990)</td>
<td>Seasonal 1986</td>
<td>Population, Development</td>
<td>Flowrate, Multiple WQ variables, Geomorphology</td>
<td>Regression, Trend analysis</td>
</tr>
<tr>
<td>Michael D. Wise, Keith A. Greer (2005)</td>
<td>Annual Seasonal 1965-2000</td>
<td>Class area</td>
<td>Flowrate, Geomorphology, Vegetation cover</td>
<td>Multi-year trend analysis</td>
</tr>
</tbody>
</table>

**Table 1a.** Methodological review resources 1-10. Matrix of timeframes studied, key explanatory and response variables, and analysis methods used in previous studies.
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Timeframe</th>
<th>Explanatory Variables</th>
<th>Response Variables</th>
<th>Methods of Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uma Seebartusang (2012)</td>
<td>Seasonal 2006-2008</td>
<td>Land use, CI</td>
<td>Total dissolved solids, Conductance, Phosphate, Chloride concentration</td>
<td>Multiple linear regression</td>
</tr>
<tr>
<td>Sang-Woo Lee, Seong-Jia Huang, Sue-Bom Lee, Hwa-Sun Hwang, Hyes-Chun Song (2009)</td>
<td>Seasonal 2002</td>
<td>Land use, Pattern</td>
<td>Phosphorus, Nitrogen, Biological oxygen demand, Chemical oxygen demand</td>
<td>Correlative analysis</td>
</tr>
<tr>
<td>Jun Tu, Zong-Qiao Xia (2008)</td>
<td>1990-2005</td>
<td>Class area, PDLU, Population</td>
<td>Nitrogen, Phosphorus, Dissolved ions, Specific conductance</td>
<td>Linear regression</td>
</tr>
</tbody>
</table>

Table 1b. Methodological review resources 11-20. Matrix of timeframes studied, key explanatory and response variables, and analysis methods used in previous studies.
Table 2. Pearson correlation matrix of response variables. Pearson's correlation coefficients greater than .50 (bold) indicate multicollinearity between variables.

<table>
<thead>
<tr>
<th></th>
<th>Term 1</th>
<th>Term 2</th>
<th>Term 3</th>
<th>Term 4</th>
<th>Term 5</th>
<th>Term 6</th>
<th>Term 7</th>
<th>Term 8</th>
<th>Term 9</th>
<th>Term 10</th>
<th>Term 11</th>
<th>Term 12</th>
<th>Term 13</th>
<th>Term 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term 1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 2</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 3</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 4</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Term 13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Term 14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3. OLS coefficients, adjusted R² values, and GWR comparisons with AIC scores for the 250 meter analysis area.

<table>
<thead>
<tr>
<th>Distance to Outfall</th>
<th>Tree MPs</th>
<th>Grass MPs</th>
<th>Impervious MPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACC OLS</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>ACC GWR</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>GWR R</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chlor. a</th>
<th>PO4</th>
<th>NH4</th>
<th>NO3</th>
<th>ND</th>
<th>DIN</th>
<th>SI</th>
<th>DO</th>
<th>TP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

OLS Coefficients Comparison 2500 Meters 2500 Meters

<table>
<thead>
<tr>
<th>Number of Outfalls</th>
<th>Est. Population Density</th>
<th>Area Trees</th>
<th>Area Grass</th>
<th>Area Impervious</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
</tr>
<tr>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
</tr>
<tr>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
</tr>
<tr>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
</tr>
<tr>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
<td>0.0000000</td>
</tr>
</tbody>
</table>

Bold values indicate a statistically significant relationship (p-value < 0.05). An asterisk next to nutrient header denotes non-normal residual distribution in that model.
Table 4. OLS coefficients, adjusted R² values, and GWR comparisons with AIC scores for the 500 meter analysis area. Bold values indicate a statistically significant relationship (p-value < 0.05). An asterisk next to nutrient header denotes non-normal residual distribution in that model.

<table>
<thead>
<tr>
<th>Distance to Outfall</th>
<th>Tree MPs</th>
<th>Grass MPs</th>
<th>Important MPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000306</td>
<td>0.001270</td>
<td>0.000025</td>
<td>0.000170 0.000025</td>
</tr>
<tr>
<td>0.000060</td>
<td>0.000010</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
</tr>
</tbody>
</table>

OLS coefficients contribution for 500 meters

<table>
<thead>
<tr>
<th>Number of Outfalls</th>
<th>Est. Population Density</th>
<th>Area Trees</th>
<th>Area Grass</th>
<th>Area Important MPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
<td></td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
<td></td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
<td></td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
<td></td>
</tr>
<tr>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000 0.000000</td>
<td></td>
</tr>
</tbody>
</table>

OLS coefficients composition for 500 meters
Table 5. OLS coefficients, adjusted R² values, and GWR comparisons with AIC scores for the 1000 meter analysis area. Bold values indicate a statistically significant relationship (p-value < 0.05). An asterisk next to nutrient header denotes non-normal residual distribution in that model.

<table>
<thead>
<tr>
<th>Distance to Outfall</th>
<th>GWR</th>
<th>AIC GWR</th>
<th>Adjusted R²</th>
<th>OLS</th>
<th>AIC OLS</th>
<th>Adjusted R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
<tr>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
<td>0.00000000</td>
</tr>
</tbody>
</table>

OLS coefficients for intensity metrics 1000 meters.
Table 6. Non-parametric Spearman correlation of water quality variables and land use metrics at the 250, 500, and 1000 meter scale. The sign of the correlation indicates the direction of the relationship. Bold values indicate a significant relationship (p-value < 0.05).

<table>
<thead>
<tr>
<th>Distance to Oil Well</th>
<th>Population Density</th>
<th>Number of Oil Wells</th>
<th>Land USE</th>
<th>CI</th>
<th>DNI</th>
<th>SI</th>
<th>DO</th>
<th>TP</th>
<th>ODp</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500 m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000 m</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 7
Comparison of results from OLS and Spearman tests across all three scales. Green squares represent a matching significant relationship between both tests. A positive or negative sign indicates the direction of the relationship.

<table>
<thead>
<tr>
<th>Distance to Outfall</th>
<th>Number of Outfalls</th>
<th>Est. Population Density</th>
<th>Tree Mps</th>
<th>Area Tree</th>
<th>Grass Mps</th>
<th>Area Grass</th>
<th>Impervious Mps</th>
<th>Area Impervious</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

All Scales OLS/GWR - Spearman Comparison
Appendices

Appendix A: Methodological Review

Studies that attempt to demonstrate relationships between LULC characteristics and water quality indicators invariably must contend with a wide array of spatial and temporal attributes associated with the landscape. Heterogeneity between study areas makes it difficult to determine a one-size-fits-all method of analysis. Due to the unique aspects associated with each location, there is a wide range of specific variables utilized between studies. Based on an initial sampling, five features were identified as primary components within studies of this type:

- Scale – Criteria used to delineate a selected study area and the scale (i.e. catchment, riparian, both) at which the analysis was conducted.
- Timeframe – Length of measured interval and considerations of temporal/seasonal variability in the landscape.
- Explanatory Variables – Specific environmental, anthropogenic or disturbance metrics utilized to measure developmental impact on the landscape.
- Response Variables – Indicators of water quality/ecosystem degradation.
- Analysis Methods – Specific statistical models and techniques used for assessing LULC-water quality relationships.

Studies were selected for review based on their defined goals and their incorporation of the above elements. An additional three components were included in order to interpret the significance of the first five:

- Results – Trends, correlations, strength of relationships, or inconsistencies observed with each particular suite of methods use
• Possible Confounding Variables – Limitations and sources of error observed in each study.

• Recommendations – Collective recommendations for future study defined by the authors.

These eight categories served as the framework for a review matrix. Nineteen studies were included, covering a wide range of methodological mechanics and site characteristics. The following sub-sections summarize and discuss the key findings associated with the five primary categories.

Scale

Thirteen of the nineteen studies reviewed used greater area watersheds as the sole scale for their analyses. Watershed data was obtained using Hydrological Unit Code (HUC) watershed delineations pre-established by the United States Geological Survey (USGS) (U.S. Geological Survey and U.S. Department of Agriculture, 2011), or by using elevation data and catchment modeling tools such as those included in ESRI’s ArcGIS for desktop. Watersheds could also be further broken down into sub-catchments surrounding specific points (i.e. water quality monitoring stations) that varied in size depending on flow characteristics for each site. None of the studies reviewed used a smaller riparian area as the sole scale in their analysis. The five studies that included riparian buffer areas also included an equivalent watershed model for comparison. One study used local government administrative districts as the boundary area of their analysis (Seeboonruang, 2012).
There does not appear to be a clear consensus on which scale is more effective in assessing LULC-water quality relationships. Sliva & Williams directly addressed this question in their 2001 paper: “Buffer Zone Versus Whole Catchment Approaches to Studying Land Use Impact on River Water Quality”. They assessed relationships between several LULC classifications and water quality indicators in southern Ontario using a watershed scale and 100-meter stream buffers. Their watershed analysis yielded a slightly stronger relationship than the buffer scale, however, the authors hypothesize that this may have been due to lower imagery resolution at the buffer level. They recommend that high resolution imagery be utilized in future studies (Sliva & Williams, 2001). Pratt and Chang (2012) found that landscape variables significant in predicting water quality varied depending on the scale and the season. Overall, 10 out of 14 models showed higher $R^2$ values between landscape variables and water quality indicators at the watershed level (Pratt & Chang, 2012). Similarly, they suggest that low-resolution imagery used for land classification may have resulted in the weaker relationships observed at the buffer scale.

High levels of spatial variability between site characteristics of different studies may override any topical comparison of methods when considering scale. As mentioned in the introduction, Carey et al. (2011) found that LULC indicators best described TP loading trends into Miami-Dade canals within a 1000-meter riparian buffer area, as opposed to the watershed scale. This finding is especially significant when investigating occurrences of macro-algal blooms and increased rates of primary production in the Bay. Carey et al. subsequently recommends that future studies and implementation of BMPs for phosphorus focus within this area. Differences in buffer size may have played a role
in this case. Both Sliva and Williams, and Pratt and Chang, used a 100-meter riparian buffer for their analysis, while Carey et al. used a substantially larger area of 1000 meters. A larger buffer area may have helped to compensate for lower resolution landscape data. Dissimilarities in site features may also have played a role. Both the Southern Ontario and Oregon-Washington study areas boast significant topological relief and have retained natural river banks/formations. The South Florida watershed is relatively flat and has been heavily altered by the installation of the canals. All of these elements significantly alter the hydrodynamic properties of each site, which may influence correlations between land use and water quality indicators at different scales.

While there does not appear to be a definitive choice between a watershed versus buffer scale analysis, it is clear that the selection of scale depends on the goals of the study, the quality of provided data, and the physical characteristics of the local landscape. Due to the high level of physical variability surrounding a particular study area, it may be wisest to use a multi-scale analysis for investigating landscape-water quality relationships. Comparisons between watershed, sub-watershed, and several buffer delineations (100-1000 meters) may better elucidate the nuanced interactions occurring at all scales. Additionally, landscape classifications should utilize high resolution data (<1 meter) to help reduce potential error associated with low resolution at the riparian scale.

*Timeframe*

In most cases, timeframe is dependent on the availability of consistent water quality monitoring data. Fifteen of the nineteen studies reviewed used multi-year water quality data sets. Water quality data was grouped either monthly or annually depending on measurement frequency. Eight of the fifteen multi-year studies also incorporated
multiple imagery years to assess the change in landscape composition and/or pattern during these periods. The remaining seven studies used landscape imagery as close as possible to the median year for their corresponding water quality dataset. Seven studies out of the nineteen reviewed separated their water quality datasets by season. Varying precipitation and flow rates encountered at respective study sites was the driving factor for seasonal separation of data in all studies. Most studies separated seasonality into wet and dry seasons based on local precipitation patterns except for J. Tu et al. (2007) who separated their study by the four seasons.

Making a distinction between periods of high and low stream volume flow helps to account for a significant source of variation in water quality trends. Increased runoff into streams and rivers can dilute or increase nutrient concentrations and alter loading patterns (Pratt & Chang, 2012; Seeboonruang, 2012; Sliva & Williams, 2001). Additionally, wet seasons typically coincide with a period of increased vegetation growth throughout the watershed, which can also impact nutrient levels (Sliva & Williams, 2001). Based on these findings, data subject to significant seasonal variation in precipitation should be analyzed separately to reflect this factor.

Explanatory Variables

Wide varieties of explanatory variables are utilized for measuring the impact of human development on local water quality. Similar to scale, selection of suitable metrics depends heavily on the characteristics of the landscape, data available, and the specific goals of a particular study. Based on the studies reviewed, variables can be sorted into four general categories.
1) Classification metrics: define the composition of land cover and are typically derived from satellite or aerial imagery. Class metrics can be very broad, combining all facets of a particular type of land use into a generalized category (i.e. agriculture, urban, forest), or specific, representing a more fundamental component of land cover (i.e. row crops, tree crops, roads).

2) Environmental metrics: are inherent characteristics of the landscape such as, elevation, precipitation, and temperature. Human socioeconomic data such as, population density, income, or housing density can also be included.

3) Pattern metrics: measure the configuration of different land cover classes across the study area. They use statistical spatial analysis methods to evaluate the arrangement of land cover types within the landscape. Pattern metrics assign a configuration score to each class based on their respective statistical criterion for measurement (e.g. Shannon Diversity Index (SHDI), Patch Density, Edge Density).

4) Index metrics: an assigned value that describes development impacts on natural resources. Index metrics are derived from an established criterion involving class, environmental, and/or pattern metrics. Some examples include the Contamination Potential Index, or Landscape Development Index.

Classification metrics were the most common variables used, as eighteen of the nineteen studies reviewed incorporated them to some degree. Twelve studies used class metrics as their sole explanatory variables. Eight studies included environmental variables alongside class variables and only one study used an environmental metric, population change, as a sole explanatory variable (Tong, 1990). Five studies used an
index metric and were paired with class metrics. Five studies conducted a pattern analysis in conjunction with one or more of the other three categories.

Landscape classification data serves as base data for almost all methods reviewed. A statistical analysis is impossible without some quantification of land cover, and classified values are an essential component of index and pattern metrics. The question surrounding classifications is not whether they should be included, but rather, to what degree should categories of land use be generalized. Multiple studies that used grouped classifications of land use have observed that over-generalization may mask more subtle interactions with nutrient indicators (Dow & Zampella, 2000; Huang, Zhan, Yan, Wu, & Deng, 2013; S.-W. Lee et al., 2009; Pratt & Chang, 2012; Seeboonruang, 2012; Sliva & Williams, 2001). However, the resolution of available data can limit the ability to define constituent categories of land cover (Fernandez, Barquin, Alvarez-Cabria, & Penas, 2014).

The classification of agricultural land use serves as an example of potential issues with classification generalization. Almost every study reviewed included some description of cultivated land in their classification scheme. Most studies grouped multiple variations of agriculture into a singular group, such as: farmland and arable land (Li, Li, & Wu, 2016), row-crops and non-row crops (Sliva & Williams, 2001), or in-season and off-season farming (Seeboonruang, 2012). In some cases, this grouping failed to portray the inherent differences in land management practices associated with different types of agriculture. Lee et al. 2009, predicted strong correlations between agricultural land use and N/P loading due to the heavy application of fertilizers during the sampling period. Conversely, their results showed weak relationships between
agriculture, TN, and TP. It was hypothesized that differences in crop-specific water management techniques and seasonal precipitation patterns might account for the less significant impacts observed (S.-W. Lee et al., 2009). It is generally advisable that class metrics be separated into constituent components relative to the landscape and land use characteristics of the study area, and to the level of detail of available data (Baker, 2003; Dow & Zampella, 2000; Huang et al., 2013; S.-W. Lee et al., 2009; Pratt & Chang, 2012; Seeboonruang, 2012; Sliva & Williams, 2001).

Ten out of nineteen studies reviewed included environmental variables into their analysis. Environmental variables can be important predictors of water quality indicators. Where it was included, mean slope within the study area was shown to exhibit significant positive correlations with nutrient loading (Pratt & Chang, 2012; Sliva & Williams, 2001). Interactions between resident soil types and nutrients such as phosphorus have been hypothesized to play a role in loading trends, however, only one study was able to show a significant relationship (Sliva & Williams, 2001). Precipitation data is essential for calculating loading trends (Carey et al., 2011), and the occurrence of extreme weather events during a sampling period can dramatically increase nutrient influx into hydrological systems (Zhang, Kelble, Fischer, & Moore, 2009). Incorporating environmental data into an analysis can yield significant correlations. At the very least, an understanding of the natural conditions present may help to define unexplained variability that may emerge in results (S.-W. Lee et al., 2009).

Socioeconomic data has also yielded positive correlations with water quality degradation. Alberti et al. (2007) analyzed the relationship between percent impervious area and three descriptive characteristics of urban landscapes: percent transportation use,
housing density, and population density. They found that these elements help to describe a majority of the variability found within impervious areas, and hypothesize, that an analysis of their configuration within urban landscapes may better describe interrelationships with pollutants and nutrient loading trends. Additionally, a subsequent analysis showed that population density, and the number of road crossings (an indicator of transportation density) were significantly correlated with water quality degradation (Alberti et al., 2007). This study demonstrates how socioeconomic variables, when applied correctly, can serve as a fundamental metric for development impacts.

Six out of nineteen studies utilized landscape pattern metrics as key explanatory variables. All unanimously recommended the use of pattern metrics while investigating landscape-water quality connectivity. They argue that by excluding how different classes of land use are arranged in space, simple aggregate measures of land cover can only elucidate general trends (Alberti et al., 2007; Baker, 2003; S.-W. Lee et al., 2009). The same study by Alberti et al. investigated the role of pattern metrics in water quality degradation. They hypothesized that water quality conditions in urbanized landscapes are influenced by four pattern variables: land use intensity, land cover composition, landscape configuration, and connectivity of impervious areas. For their analysis, they utilized a large suite of pattern metrics such as mean patch size, aggregation index, and SHDI as their explanatory variables. They found that elements of landscape configuration (i.e. mean patch size and number of road crossings) helped to explain variance in indicators that could not be explained by class metrics alone (Alberti et al., 2007). A similar study conducted by S.-W. Lee et al. (2009), found that fragmentation of the landscape (forested land cover in particular) was correlated with degraded water
quality. This suggests that heavily fragmented vegetation and scattered land use can have a more significant impact on water quality than concentrated patches of heavy use (S.-W. Lee et al., 2009). These examples demonstrate that development impacts may need to be considered in multiple spatial dimensions.

Five of nineteen studies used index metrics in their analyses. The advantage of an index is that it can incorporate multiple interrelated landscape characteristics such as population density and TIA, into a simplified explanatory score. Additionally, an index can be easily transferred to similar studies, providing a consistent metric for comparative purposes (Carey et al., 2011; Seeboonruang, 2012). Seeboonruang (2012) utilized the Contamination Potential Index (CPI), developed by the US Department of Energy, as a measure of land use impacts. The CPI’s original purpose is to assess the risk of groundwater contamination from pesticides. It develops a disturbance score (the greater the score the greater the impact) by integrating the quantity of contaminant released, contaminant characteristics, and wastewater quality relative to standards (Seeboonruang, 2012). Their analysis found that high CPI values for different land use types can be attributed to the magnitude of certain characteristics unique to each class. For example, urban areas with high population density, or poultry farms with a large number of resident chickens, received the highest CPI scores. Subsequent regression models found positive relationships between land use types with high CPI values and incidence of increased total dissolved solids, chlorides, and conductance. In this instance, the CPI helped describe key influential characteristics associated with different land use classifications, as well as served as an accurate indicator of development impacts.
On the other hand, it is important to understand how an index is constructed before it is utilized within a particular study area. For example, a primary component of the LDI index is human population density (Carey et al., 2011). In the aforementioned study by Carey et al. (2011), the LDI index was found to have a strong positive correlation with nutrient loading trends in urbanized sub-basins. However, due to low population density in agricultural areas, LDI scores for this type of land use did not significantly correlate with water quality degradation, possibly underestimating its actual impacts. As a result, the authors recommend including LDI with a suite of other metrics relative to the landscape’s characteristics (Carey et al., 2011).

There is no global definitive answer as to which category of explanatory variable, or suite of variables, are better suited for describing LULC-water quality relationships. Similar to the case with varying scales, the selection of proper metrics depends greatly on the characteristics of the study area and the quality of data available. As stated previously, diffuse non point-source pollution is inherently difficult to quantify because it is subject to a great deal of variability within the landscape (NRC, 2000). It is common for explanatory variables to exhibit different relationships across temporal and spatial scales (Baker, 2003; Carey et al., 2011; Pratt & Chang, 2012). Therefore, careful consideration of the unique conditions present within a study area should be prioritized when selecting appropriate metrics (Baker, 2003).

*Response Variables*

There is a wide variety of nutrient, chemical, contaminant, indicator, and environmental response variables that have been used in determining developmental impacts within watersheds. Selection of proper variables depends on the data available
and the specific goals of the study. For instance, a study focusing on eutrophication issues will likely select indicators of N and P as variables because of their role in primary productivity (Atkinson, 1983; NRC, 2000). Huang et al. (2013), selected TN, TP, NH3-N, and Chemical Oxygen Demand (CODmn) in order to investigate recent bloom events in their study area, but most importantly, because there were consistent records available for these parameters. Some studies may select a response variable based on the composition of land use within an area of interest. Jun Tu and Xia (2008), used measurements of Specific Conductance (SC) to assess urban developmental impacts surrounding the Boston, MA metropolitan area, specifically because of its close relationship with urban runoff (Baker, 2003). Other studies may select a sole indicator variable such as Chlorophyll a (CHLa), or the Benthic Index of Biological Integrity as a means for evaluating water quality degradation (Alberti et al., 2007; Boyer et al., 2009).

Methods of Analysis

Overall, there is great diversity among studies reviewed, with many different individual characteristics and components utilized. However, all studies draw from similar data sources for their statistical analyses (i.e. image classifications, elevation data, water sampling data, etc.), creating a common statistical mechanistic thread. This field of research is inherently correlative, therefore, all studies that were reviewed utilized some form of regression analysis or trend comparison to assess water quality-land use relationships. Tests for normality, transformations, and non-parametric tests were common as landscape data and water sampling data were often not normally distributed (e.g. Tong & Chen, 2002). Principle component analysis and factor analysis techniques were utilized to assess the suitability for proposed explanatory metrics and reduce the
number of covariates included in subsequent regression techniques (e.g. Carey et al., 2011). Tests for multicollinearity and spatial autocorrelation were used due to high levels of interconnectivity within the landscape (e.g. Alberti et al., 2007; Jun Tu & Xia, 2008). Selection of a statistical methodology is highly dependent on all previous components reviewed: scale, timeframe, explanatory metrics and response variables.

One primary divergence between studies was whether or not they incorporated spatially explicit regression techniques for their analysis. The majority of studies reviewed used simple or multiple linear regression models, t-tests, or trend comparisons to assess the strength of relationship between explanatory and response variables. Two studies focused on comparing Ordinary Least Squares (OLS) regression methodology to the spatially explicit Geographically Weighted Regression (GWR) technique (Pratt & Chang, 2012; Jun Tu & Xia, 2008). OLS is a popular form of standard linear regression, which assumes that there is no autocorrelation and constant variance between variables (Jun Tu & Xia, 2008). GWR is based on the concept of non-stationarity throughout the landscape, meaning, features that are closer together are more interrelated than features that are further apart (Miller, 2004). It weights nearby observations over those that are further away and accounts for spatial autocorrelation by using a distance decay function (Brunsdon, Fotheringham, & Charlton, 1996). In their study, Jun Tu & Xia (2008) found that there was a significant increase in GWR model suitability over OLS results after analyzing LULC-water quality relationships in the Boston metropolitan watershed. Pratt & Chang (2012) found similar results after analyzing watersheds in the OR-WA area. Overall, their GWR models accounted for more variance in relationships over OLS.
Evidence suggests that the GWR technique is better suited for LULC-water quality studies as these relationships are not constant over spatial extents. However, GWR results may be misleading if spatial autocorrelation between variables does not previously exist. Jun Tu & Xia (2008) found that OLS model residuals that did not exhibit significant autocorrelation would perform better than their GWR comparisons. They hypothesized, that the distance decay function of GWR models may skew the results when autocorrelation is not present. Pratt & Chang (2012) similarly found that GWR $R^2$ results were not as robust when comparing upstream and downstream river sampling points. They suggested that the distance decay function might not have been able to compensate for the distance between points, despite the inherent continuity of rivers. Conversely, changing land use patterns and natural variability between the sites may also have interfered with these results. Overall, both studies recommend that careful use of OLS, GWR, and tests for autocorrelation such as the Morans I can make an effective universal combination for analyzing LULC-water quality relationships.