Relative linkages of stream water quality and environmental health with the land use and hydrologic drivers in the coastal-urban watersheds of southeast Florida

Omar I. Abdul-Aziz1,2 and Shakil Ahmed1,2

1Department of Civil and Environmental Engineering, West Virginia University, Morgantown, West Virginia, USA,
2Department of Civil and Environmental Engineering, Florida International University, Miami, Florida, USA

Abstract A systematic data analytics was employed to determine the relative linkages of stream water quality and environmental health with the land use and hydrologic drivers in the coastal-urban watersheds of southeast Florida. Power law-based partial least squares regression models were developed to reliably estimate the linkages by appropriately resolving multicollinearity (Nash-Sutcliffe efficiency = 0.72–0.95). The analytics indicated Everglades as the external and the largest source of total nitrogen (TN) in the coastal-urban streams for both wet (June–October) and dry (November–May) seasons. The “external driver” exhibited 1.5–2 times stronger control on stream TN than that of the watershed “land use,” “hydrology,” and the “upstream reach” contributions. In contrast, Everglades appeared to be a minor source of in-stream total phosphorus (TP), which was predominantly controlled by the internal watershed processes. TP was most strongly linked with the upstream reach concentrations and watershed land uses in the wet and dry seasons, respectively. Despite the predominantly built-up fraction (74%) of the study area, agricultural land was the most substantial watershed source of in-stream nutrients. The linkages of algal biomass (Chl a) with the drivers indicated TP as the limiting nutrient. Stream dissolved oxygen was most strongly influenced by the adjacent groundwater depth and watershed land uses, respectively, in the wet and dry seasons. The estimated relative linkages and insights would be useful to identify the management targets and priorities to achieve healthy coastal-urban stream ecosystems in southeast Florida and around the world.

1. Introduction

Clean water is crucial to sustain the structure and functions of a healthy stream ecosystem. Anthropogenic activities, coupled with the natural processes, drive changes in stream water quality and environmental health [Simeonov et al., 2003; Álvarez-Cabria et al., 2016]. Linking the in-stream water quality with the watershed land uses/cover and hydrology has therefore been a focus of much research [e.g., Shrestha and Kazama, 2007; Chang, 2008; Kang et al., 2010; Nagy et al., 2012; R. Wan et al., 2014; Y. Wan et al., 2014]. However, the relative linkages of stream water quality with the land use and hydrologic drivers are yet to be understood well—particularly in the context of growing coastal-urban environments. The knowledge gaps motivate further research to understand and manage changes in urban stream water quality to ensure clean water for the future [National Research Council (NRC), 2012].

Stream water quality is generally shaped by a multitude of interacting land use/cover, hydrologic and ecological drivers. The characterization and assessment of stream water quality, therefore, often involve multivariable data sets (high mutual correlations among the drivers), necessitating the application of multivariate statistical techniques. Y. Wan et al. [2014] provided a comprehensive account of studies, involving multivariate techniques to link stream water quality indicators mainly with their land use and hydrologic drivers. Specifically, the principal component analysis (PCA), factor analysis (FA), and multiple regression modeling have widely been used in empirical investigations of stream water quality. For example, Robinson et al. [2014] employed PCA to report a notable linkage of watershed agricultural and forest lands with the stream nutrients in the Swiss Plateau. Bu et al. [2014] employed FA and multiple regression to demonstrate a strong association of agricultural lands with stream physicochemical variables (e.g., nutrients) in the Taizi River Basin, China. Using a Bayesian regression model, R. Wan et al. [2014] reported dominant controls of urban and
agricultural land uses on stream nutrients in the Tai Lake Basin, China. Kang et al. [2010] employed multiple linear regressions to demonstrate strong linkages of land uses and hydrologic variables (basin size, slope, and permeability) with the in-stream concentrations of heavy metals and bacteria counts in the Yeongsan Basin, South Korea.

Many studies (see references in Carey et al. [2013]) exclusively focused on the investigations of the sources and drivers of stream water quality in coastal-urban watersheds. Watershed storm water runoff, atmospheric deposition, sewage, and effluent of wastewater treatment plants have long been recognized as the main contributors to urban stream water pollutions [Schueler, 2003; Andersen et al., 2004; Ahn et al., 2005]. However, following implementations of improved sewage treatments, storm water runoff often appears to be the primary contributor of pollutants in urban streams [Bay et al., 2003; Ahn et al., 2005]. Tu [2009] used a geographic information system-based watershed loading model for metropolitan Boston—reporting large impacts of climate and land use changes on the watershed nitrogen loads into the receiving streams. Tran et al. [2010] demonstrated strong links of land use/cover with the stream water quality in New York. Newcomer et al. [2012] reported elevated loads of organic carbon in the metropolitan area streams of Baltimore, Maryland. Apart from the land uses, the interaction of surface water with groundwater [Menció and Mas-Pla, 2008], as well as seawater intrusion [Liu et al., 2010], is also known to affect the stream water quality in the urban-coastal watersheds.

Badruzzaman et al. [2012] provided a comprehensive review of nutrient sources in the surface waters across Florida. In a pioneering research, Caccia and Boyer [2005] reported close associations of Biscayne Bay estuary water quality indicators (e.g., nitrogen, phosphorus, dissolved oxygen, and chlorophyll a) with the watershed land uses such as the urban and agricultural lands and wetlands. In a follow-up study, Caccia and Boyer [2007] estimated high loadings of nutrients (inorganic nitrogen and total phosphorus) into the Biscayne Bay during 1994–2002 from major canals, atmospheric depositions, and groundwater inputs. Zhang et al. [2009] demonstrated strong correspondences between watershed agricultural runoff and nutrients in Biscayne Bay estuary considering prehurricane and posthurricane periods. Carey et al. [2011a] corroborated the previous studies by estimating high nutrient loadings from six canals into the Biscayne Bay during 1992–2006. Carey et al. [2011b] also reported notable links of watershed land uses and human disturbances with the nutrient loadings into five inland canals of Biscayne Bay Watershed during 1995–2004. Furthermore, Y. Wan et al. [2014] found strong linkages of stream water quality with the hydrometric variables (e.g., rainfall and flow), agricultural and wetland areas, and upstream water controls in the Indian River Lagoon Watershed, Florida.

Overall, the previous research on urban stream water quality for southeast Florida—one of the major urban population centers in the U.S.—is mostly limited to the Biscayne Bay Watershed at the south and the Indian River Lagoon Watershed at the north. Further, most studies for southeast Florida and around the world attempted to infer the linkages between the water quality indicators and their potential drivers using multivariate techniques and traditional regression modeling, which can often lead to biased results due to multicollinearity of data matrix. Although multicollinearity may be reduced by backward elimination (or forward addition) of predictors based on statistical significance, the stepwise selection can exclude mechanistically meaningful variables. Partial least squares regression (PLSR) can overcome this pitfall and resolve multicollinearity by performing the model fitting on the transformed orthogonal planes [Ishtiaq and Abdul-Aziz, 2015]. The estimated PLSR model is then transformed back to the original variable domain to compute regression weights (i.e., coefficients) of individual variables. PLSR has been successfully employed to analyze contaminants in road runoff [Mahbub et al., 2011] and watershed sediment yield and streamflows [Yan et al., 2013].

The paper aims to determine the relative linkages of land uses and hydrologic drivers with common stream water quality and environmental health indicators (TN, total nitrogen; TP, total phosphorus; Chl a, chlorophyll a; DO, dissolved oxygen) in coastal-urban watersheds. The research is conducted by integrating four complementary layers (correlation matrix, PCA, FA, and PLSR) of data analytics into a single systematic framework to reliably estimate the linkages by resolving multicollinearity. The study expands on the previous research for southeast Florida by investigating the stream water quality of the thriving urban region between the Biscayne Bay and the Indian River Lagoon Watersheds (Figure 1). The region also represents a prototype of growing coastal-urban environments around the world. The research provides an objective empirical
2. Materials and Methods

2.1. Study Area

The study watersheds represent the coastal-urban environments of southeast Florida, encompassing the Broward County and extending partly into the Palm Beach County (Figure 1). The densely populated region is located between the Florida Everglades and the Atlantic Ocean. It is drained by an intricate network of more than 266 miles of natural and dredged canals [Broward County, 2016]. The study area has six major canal basins (49–269 km²), including the Hillsboro Canal, C-14 Canal (Cypress Creek and Pompano Canal), C-13 Canal, C-12 Canal, North New River Canal, and C-11 Canal (South New River, Dania Cut-off Canal). The six canals together with the secondary and tertiary canals drain approximately 914 km² of mostly highly urbanized areas to the estuarine waters [Cooper and Lane, 1987]. Most canals are connected to the Everglades water conservation areas (WCAs) at the upstream and coastal waters at the downstream; WCAs are typically used for managing water resources and wildlife in the Everglades agricultural area. Proximity to the WCAs...
and the coastlines influences the in-stream biogeochemistry and water quality of the study watersheds [Broward County Environmental Protection Department (BCEPD), 2007].

2.2. Data Sets
Quarterly time series observations of in-stream TN, TP, Chl a, and DO during 2009–2013 were gathered for 24 monitoring stations across the study area (Figure 1) from the Broward County [Broward County Environmental Planning and Community Resilience Division, 2015]. Drainage areas (subwatersheds) for 18 monitoring sites (excluding six most upstream sites) were generated using ESRI ArcGIS 10.2 based on the waterbody identification number of the Florida Department of Environmental Protection [Florida Department of Environmental Protection, 2016], contribution area map provided by Broward County [Robert Bernhard, personal communications, 2012], and 10 m digital elevation model [United States Geological Survey (USGS), 2015a]. TN represented the summation of total Kjeldahl nitrogen, nitrate-nitrogen, and nitrite-nitrogen. TP was the aggregation of soluble reactive phosphates, condensed phosphates, and organic phosphates. The four variables are commonly used as the primary indicators of stream water quality [Chapra, 1997; Chang, 2008]. Chl a and DO further indicate the health of a stream ecosystem—representing the combined effects of different pollution sources and drivers [Boyer et al., 2009]. Further details on the water quality data, including QA/QC, are given in the supporting information.

The hydroclimatic influence on stream water quality was incorporated by conducting separate, comparative analyses for wet (June–October, representing 70% of the annual rainfall) and dry (November–May) seasons [South Florida Water Management District (SFWMD), 2015a]. The spatial data set for each season was obtained by averaging observations of the respective water quality indicators over the 5 year period (2009–2013). Subwatershed area (A), mean slope (S), and mean imperviousness (I) were used to represent the corresponding surface runoff. Percent imperviousness for each subwatershed was calculated from the 30 m data of National Land Cover Database 2011 [National Land Cover Database, 2015]. Ambient groundwater depth from the land surface (GWD, which increases with the lowering of groundwater table) was used to represent the subsurface hydrologic driver of in-stream water quality. Average GWDs for both wet and dry seasons over the 5 years were estimated for each quality station from the groundwater level data of USGS [2015b] (see Figure S1 in the supporting information for details). The canal centerline distance of individual water quality station from the coastal outlet (Dc) was used to represent the influence of both downstream coastal hydrology and upstream Everglades (except for C-12) on the in-stream water quality. The spatial data for A, S, I, Dc, and GWD were processed on the ArcGIS 10.2 platform.

The fractions of different land use areas over the subwatershed areas were estimated from the land use and land cover data of the South Florida Water Management District [SFWMD, 2015b]. Following guidelines of SFWMD [2011], different land use types were aggregated into five broad categories as follows: (1) vegetated land (VEG), including upland forests, wetland forests, and vegetated nonforested wetlands; (2) agricultural land (AGR), including lands for agriculture and grazing; (3) waterbody (WAT), including streams, waterways, lakes, and reservoirs; (4) built-up land (BUL), including residential, commercial, industrial, and transportation areas; and (5) open land (OPN), including barren and recreational lands. Further details into the land use categories are given in the supporting information.

For each water quality indicator, the final data set incorporated concentrations at the inlets and outlets (CIN and COUT, respectively) of 18 stream reaches/subwatersheds, five land use fractions (VEG, AGR, WAT, BUL, and OPN) for individual subwatersheds, the corresponding hydrologic variables (A, S, I, and GWD), and distance from coast (Dc). Observation at the immediate upstream monitoring station (CIN) was used to represent the upstream-reach contribution to the water quality of a corresponding downstream station (COUT). The sample size (N) of the spatial data set for each water quality indicator (COUT) was 18. The data summary across the study area (Table 1) and individual canal basins (Table S1 in the supporting information) indicated wide ranges and considerable gradients in water quality indicators and the land use/hydrologic drivers.

2.3. The Data Analytics Framework
The research employed a systematic data analytics framework (Figure 2), which involves a sequential application of four complementary data-mining techniques—Pearson correlation matrix, PCA, FA, and partial least squares regression (PLSR) [Ishтиaq and Abdul-Aziz, 2015]. The four-step analytics was designed to corroborate
### Table 1. Summary of the Stream Water Quality/Environmental Health Indicators and Their Land Use/Hydrologic Drivers

| Variables | Mean | Standard Deviation | Coefficient of Variation | Minimum | 25th Percentile | 50th Percentile | 75th Percentile | Maximum |
|-----------|------|--------------------|--------------------------|---------|-----------------|-----------------|----------------|---------|
| A (km²)   | 47.49| 43.08              | 0.91                     | 4.47    | 9.91            | 36.06           | 86.31          | 150.66  |
| S (%)     | 2.17 | 0.35               | 0.16                     | 1.64    | 1.88            | 2.21            | 2.43           | 2.91    |
| I (%)     | 37.15| 10.16              | 0.27                     | 19.00   | 28.63           | 36.33           | 45.55          | 52.67   |
| Dc (km)   | 9.04 | 5.18               | 0.57                     | 1.40    | 5.47            | 8.67            | 12.62          | 21.01   |
| BUL (%)   | 74.3 | 14.5               | 0.20                     | 39.3    | 64.9            | 79.8            | 85.6           | 92.2    |
| OPN (%)   | 11.2 | 8.1                | 0.73                     | 2.0     | 4.6             | 9.5             | 16.1           | 30.1    |
| AGR (%)   | 2.53 | 3.47               | 1.37                     | 0.02    | 0.25            | 1.32            | 3.61           | 13.51   |
| VEG (%)   | 3.76 | 4.10               | 1.09                     | 0.10    | 0.72            | 2.51            | 5.37           | 12.72   |
| WAT (%)   | 8.25 | 4.68               | 0.57                     | 3.48    | 5.07            | 7.72            | 9.74           | 23.42   |

**Wet Season (June–October)**

| Variables | Mean | Standard Deviation | Coefficient of Variation | Minimum | 25th Percentile | 50th Percentile | 75th Percentile | Maximum |
|-----------|------|--------------------|--------------------------|---------|-----------------|-----------------|----------------|---------|
| TN (mg/L) | 1.24 | 0.32               | 0.26                     | 0.61    | 0.95            | 1.31            | 1.45           | 1.77    |
| TP (mg/L) | 0.05 | 0.03               | 0.62                     | 0.02    | 0.02            | 0.04            | 0.05           | 0.11    |
| Chl a (μg/L) | 7.81 | 4.64             | 0.59                     | 1.84    | 5.12            | 6.49            | 8.86           | 23.13   |
| DO (mg/L) | 4.10 | 0.94               | 0.23                     | 1.61    | 3.36            | 4.43            | 4.76           | 5.67    |
| GWD (m)   | 1.84 | 0.69               | 0.37                     | 0.79    | 1.39            | 1.56            | 2.45           | 3.11    |

**Dry Season (November–May)**

| Variables | Mean | Standard Deviation | Coefficient of Variation | Minimum | 25th Percentile | 50th Percentile | 75th Percentile | Maximum |
|-----------|------|--------------------|--------------------------|---------|-----------------|-----------------|----------------|---------|
| TN (mg/L) | 1.13 | 0.37               | 0.32                     | 0.51    | 0.78            | 1.23            | 1.46           | 1.63    |
| TP (mg/L) | 0.04 | 0.03               | 0.60                     | 0.02    | 0.02            | 0.04            | 0.05           | 0.11    |
| Chl a (μg/L) | 5.93 | 5.37             | 0.91                     | 1.62    | 2.79            | 4.05            | 6.23           | 22.66   |
| DO (mg/L) | 5.97 | 1.06               | 0.18                     | 3.34    | 5.75            | 6.13            | 6.44           | 8.67    |
| GWD (m)   | 2.00 | 0.72               | 0.36                     | 0.77    | 1.56            | 1.71            | 2.52           | 3.42    |

*Notes: A, S, I, Dc, BUL, OPN, AGR, VEG, WAT, and GWD, respectively, refer to subwatershed area, slope, imperviousness, distance from coast, built-up land, open land, agricultural land, vegetated land, waterbody, and groundwater depth.*

---

**Figure 2.** The data analytics framework to determine the relative linkages of stream water quality and environmental health indicators with their drivers.
findings from different data sets and synthesize information toward the overall outcomes. The analytics methods were well-suited for our data sets, given their reputation of providing reliable results across small to large samples [Wold et al., 2001; de Winter et al., 2009]. Based on a preliminary analysis, data for all response and predictor variables were first logarithmically (log10) transformed to incorporate the nonlinear interactions. The log-transformed variables were then standardized to bring different scales and units to a comparable reference scale (dimensionless Z-scores) as follows: \( Z = (X - \bar{X})/s_X \), where \( X = \log_{10} \)-transformed variable, \( \bar{X} = \text{mean of } X \), and \( s_X = \text{standard deviation of } X \). Each layer of the data analytics was performed with the Z-scores using MATLAB.

The correlation matrix was computed to explore the multicollinear structure of the data sets and achieve first-order information (albeit impacted by multicollinearity) on the correspondences among the water quality, hydrologic, and land use variables. PCA and FA were employed to resolve multicollinearity by using projections on the orthogonal (i.e., independent) planes. Although the approaches of PCA and FA are distinctively different [Jolliffe, 2002], they, in concert, provide a relatively unbiased and confirmatory representation of the interrelationship and grouping patterns among the participatory variables. PCA derives orthogonal entities called principal components (PCs), which are weighted combinations of all original variables [Jolliffe, 2002; Peres-Neto et al., 2003]. FA inherently reduces the dimensions of a multivariate data set by decomposing the participatory variables into fewer unobservable (latent) entities called factors [Jolliffe, 2002; Johnson and Wichern, 2007]. FA is typically done using a “varimax” orthogonal rotation, which redistributes the variance of each variable to maximize the total data-system variance explained and thereby optimizes the loadings (correlation weights) of individual variables on each factor. An eigenvalue criterion (e.g., eigenvalue ≥ 1.0) is applied to represent the most variance of a data matrix with a minimum number of extracted factors.

The PLSR modeling combines features from PCA and multiple regressions [Wold et al., 2001]. However, for a conventional principal component regression, the PCs of predictors are first derived and then fitted with the response. In contrast, PLSR computes the orthogonal partial least squares (PLS) components of predictors and fits them with the response through a simultaneous decomposition of all variables [Schummann et al., 2013]. We developed PLSR model for each water quality indicator by incorporating the outlet concentrations of a stream reach (or subwatershed) as the response variable and the corresponding inlet concentrations, land use, and hydrologic drivers as the predictors. The PLSR models were reliably estimated with observed data using SIMPLS algorithm [de Jong, 1993; Hubert and Branden, 2003] and a 10-fold cross validation [Kuhn and Johnson, 2013]. The optimality (parsimony) of the models was achieved by involving a minimum number of PLS components, which was determined based on a synthesis of the minimum Akaike information criterion (AIC) [Akaike, 1974] and the maximum Nash-Sutcliffe efficiency (NSE) (see supporting information).

The final (optimal) model’s goodness-of-fit was measured by NSE, and the accuracy of predictions was evaluated by the ratio of root-mean-square error to the standard deviation of observations (RSR). Since the fitting of the optimal PLSR models was performed using the minimum number of independent PLS components, the model bias and instability arising from multicollinearity among the predictors had been optimally resolved [Kuhn and Johnson, 2013]. The estimated PLSR models were then transformed back to the original (Z-scores) variable domain to compute the relative linkages (i.e., regression weights, \( \beta \)) of each water quality variable with the land use and hydrologic predictors. The positive or negative sign of \( \beta \) indicated the direction of relationship (i.e., increasing or decreasing) between a predictor and a water quality indicator. However, the directions of relationships are mechanistically more meaningful for important predictors than that of the weak predictors. The aggregated linkages of the “land use” (\( \beta_{LU} \)), “hydrology” (\( \beta_H \)), “upstream reach” (\( \beta_{CIN} \)), and the “external driver” (i.e., upstream Everglades and/or downstream coastal; \( \beta_{OC} \)) components were calculated using the method of vector summations as follows:

\[
\beta_{LU} = \sqrt{\beta_{BUL}^2 + \beta_{OPN}^2 + \beta_{AGR}^2 + \beta_{VEG}^2 + \beta_{WAT}^2} \tag{1}
\]

\[
\beta_H = \sqrt{\beta_A^2 + \beta_S^2 + \beta_I^2 + \beta_{GWD}^2} \tag{2}
\]
Table 2. Pearson Correlation Coefficients ($r$) Between Stream Water Quality/Environmental Health Indicators and Their Drivers*

| Drivers | TN Wet | TN Dry | TP Wet | TP Dry | Chl $a$ Wet | Chl $a$ Dry | DO Wet | DO Dry |
|---------|--------|--------|--------|--------|-------------|-------------|--------|--------|
| A       | 0.27   | 0.37   | −0.12  | −0.43  | −0.22       | −0.58       | −0.15  | −0.40  |
| S       | 0.37   | 0.32   | −0.31  | −0.33  | −0.09       | −0.09       | −0.15  | 0.23   |
| $l$     | −0.68  | −0.73  | 0.15   | 0.32   | 0.11        | 0.18        | 0.60   | 0.53   |
| $D_c$   | 0.77   | 0.84   | −0.40  | −0.39  | 0.22        | 0.12        | −0.59  | −0.06  |
| BUL     | −0.56  | −0.56  | 0.12   | 0.25   | −0.09       | 0.21        | 0.65   | 0.61   |
| OPN     | 0.50   | 0.57   | −0.34  | −0.57  | −0.27       | −0.51       | −0.41  | −0.24  |
| AGR     | 0.17   | 0.11   | 0.50   | 0.51   | 0.47        | 0.36        | −0.47  | −0.67  |
| VEG     | 0.22   | 0.09   | 0.22   | 0.05   | 0.15        | −0.08       | −0.16  | −0.69  |
| WAT     | 0.57   | 0.40   | −0.32  | −0.30  | 0.00        | −0.10       | −0.24  | −0.20  |
| GWD     | −0.02  | −0.04  | −0.11  | −0.17  | 0.01        | −0.10       | 0.59   | 0.37   |
| $C_{IN}$| 0.58   | 0.65   | 0.76   | 0.64   | 0.58        | 0.64        | 0.58   | 0.21   |

*Notes: (1) Data for all variables were log$_{10}$-transformed to incorporate any nonlinear correspondences. (2) Bold indicates significant correlations at the 95% confidence level (P-values < 0.05). (3) A, S, $l$, $D_c$, BUL, OPN, AGR, VEG, WAT, GWD, and $C_{IN}$, respectively, refer to subwatershed area, slope, imperviousness, distance from coast, built-up land, open land, agricultural land, vegetated land, waterbody, groundwater depth, and inlet concentrations.

$$
\beta_{C_{IN}} = \sqrt{\beta_{C_{IN}}^2}
$$

$$
\beta_{D_c} = \sqrt{\beta_{D_c}^2}
$$

3. Results

3.1. Mutual Correspondences of the Water Quality, Hydrologic, and Land Use Variables

The nonlinear correspondences of stream water quality variables (log$_{10}$-transformed and standardized) with the land use and hydrologic drivers were first computed with the correlation coefficients (Table 2). In both wet and dry seasons, TN had a strong correspondence with the distance from coast ($r = 0.77$ to 0.84) and watershed imperviousness ($r = −0.68$ to −0.73), whereas a moderate correspondence with the built-up land ($r = −0.56$), open land ($r = 0.50$ to 0.57), and upstream reach concentrations ($C_{IN}$; $r = 0.58$ to 0.65). A moderate correspondence ($r = 0.57$) was also noted between TN and waterbody area in the wet season only. In contrast, TP had a strong correspondence with the upstream concentrations ($r = 0.64$ to 0.76) and a moderate correspondence with agricultural land ($r = 0.50$ to 0.51) in both seasons. Further, TP had a moderate correspondence with open land ($r = −0.57$) in the dry season. Similar to TP, Chl $a$ represented a moderate correlation ($r = 0.45$ to 0.58) with the upstream concentrations and agricultural land in the wet season. In dry season, a moderate correlation of Chl $a$ was noted with the inlet concentrations ($r = 0.45$), open land ($−0.51$), and subwatershed area ($−0.58$). In both seasons, DO had a moderate correlation with imperviousness ($r = 0.53$ to 0.60), built-up land ($r = 0.61$ to 0.65), and agricultural land ($r = −0.47$ to −0.67). DO further represented a moderate/strong correlation with vegetated land ($r = −0.69$) in the dry season, while a moderate correlation with the distance from coast ($r = −0.59$), groundwater depth ($r = 0.59$), and upstream concentrations ($r = 0.58$) in the wet season.

The correlation matrices (Tables S2 and S3) demonstrated high mutual correspondences among the water quality and their driving variables, indicating the presence of a strong multicollinearity in the data matrices for the two seasons. For example, the distance from coast had moderate to strong correlations with the open land ($r = 0.53$), waterbody area ($r = 0.61$), built-up land ($r = −0.63$), and imperviousness ($r = −0.73$). The negative correspondences indicate that the built-up land fraction and imperviousness increase with higher urbanization toward the coastline. In contrast, the open land and waterbody area fractions increase away from the coastline toward the Everglades. The findings were corroborated by scatter-plotting the area fractions and imperviousness with the distance from coast (not shown). Further, the substantially stronger correlation of Chl $a$ with TP than that with TN indicates that TP is the limiting nutrient in the coastal-urban streams of southeast Florida.
3.2. Relative Orientations and Groupings of the Water Quality and Driving Variables

The nonlinear loadings of different water quality indicators and the driving variables on the first two PCs are presented through biplots for both wet (Figure 3) and dry (Figure 4) seasons. The percent variances explained by all individual PCs are presented in Table S4. Among all water quality indicators and the two seasons, the first two PCs together explained approximately 56% to 62% of the total data variance.

The relative orientations of variables (Figures 3 and 4) suggested three distinct groups of drivers for all four water quality indicators across the two seasons—(1) drainage area, slope, open land, waterbody, and distance from coast; (2) agricultural and vegetated lands; and (3) built-up land and imperviousness. Based on the closeness to the outlet concentration vector ($C_{OUT}$, the red line), the driver group 1 appeared to have a strong positive linkage with TN, a moderate negative linkage with TP and Chl $a$, and a relatively weak linkage with DO. Group 2 had a moderate to strong positive linkage with TP and Chl $a$, a similar but negative linkage with DO, and a weak to moderate linkage with TN. Group 3 represented a weak linkage with TP and Chl $a$, whereas a moderate to strong negative and positive linkages with TN and DO, respectively. The upstream reach concentrations ($C_{IN}$) had a strong control on $C_{OUT}$ for all quality indicators except for DO in the dry season. The groundwater depth (GWD) appeared to be strongly and positively linked with DO in both seasons. GWD also had a strong negative linkage with group 2 and a moderate positive linkage with group 3—indicating the opposite effects of increasing agriculture/vegetation and urbanization on the water table (higher versus lower, respectively).

The PC scores on the biplots provide valuable insights into the variation of water quality across the different canal basins. For example, all sites of C-12 are located farther away from the tip of $C_{OUT}$ vector for TN than that of other canals (Figures 3 and 4) — indicating the lowest concentrations of TN in C-12 that drains an order of
magnitude smaller area (Table S1) and is not directly connected with the Everglades. However, the congregation of most sites toward the C_OUT of TN refers to much higher nitrogen concentrations in other canals that are connected with the Everglades. Further, relatively close proximity of the C-12 and Hillsboro Canal sites to C_OUT for TP and Chl a represents their higher phosphorus and algae concentrations than other canals.

3.3. Major Water Quality Drivers Based on Independent Latent Factors

FA resulted in four to five independent latent factors to optimally demonstrate the hidden patterns into water quality and the driving variables (Table 3). Variance explained by the first five factors ranged, respectively, from 38% to 47%, 15% to 20%, 10% to 13%, 9% to 11%, and 7% to 9% among the two seasons and six canal basins. Higher loadings of water quality indicators and the land use/hydrologic drivers on the same factors indicate their stronger linkages.

TN in both seasons loaded most strongly on factor 1 (0.74 to 0.79), which had very high loadings of the distance from coast (0.98 to 0.99) and moderately strong loadings of imperviousness (−0.63), built-up land (−0.53 to −0.55), and waterbody (0.52 to 0.58). In contrast, TP had the strongest loading on factor 2 (0.92) in wet season and on factor 3 (0.93) in dry season, along with the respective loadings of upstream concentrations (0.56 to 0.85) and agricultural land (0.53). Similar to TP, Chl a loaded most strongly on factor 2 (0.96) in wet season, alongside strong loadings of inlet concentrations (0.72) and a noteworthy loading of agricultural land (0.38). In dry season, the strongest loading of Chl a was on factor 4 (0.86), which had notable loadings of upstream concentrations (0.46) and open land (−0.33). DO loaded most strongly on factor 3 (0.71 to 0.96) in both seasons, alongside the high loadings of groundwater depth (0.89) in wet season, vegetated land (−0.63)

Figure 4. Biplots from principal component (PC) analysis, showing the relative orientations of the stream water quality/environmental health indicators and their drivers in dry season. Percent variance explained by each PC is shown in parenthesis.
Table 3. Major Latent Factors With Their Optimized Loadings on the Participatory Variables

| Indicators | F | A | S | I | Dc | BUL | OPN | AGR | VEG | WAT | GWD | CIN | COUT |
|------------|---|---|---|---|----|-----|-----|-----|-----|-----|-----|-----|------|
| TN Wet     | 1 | 0.005 | 0.18 | -0.63 | 0.99 | -0.55 | 0.41 | 0.06 | 0.02 | 0.58 | -0.06 | 0.24 | 0.74 |
|            | 2 | 0.91  | 0.03 | -0.66 | 0.15 | -0.60 | 0.80 | 0.02 | 0.31 | 0.13 | 0.12 | 0.45 | 0.17 |
|            | 3 | -0.01 | -0.17 | -0.28 | 0.00 | -0.53 | 0.01 | 0.80 | 0.56 | 0.27 | -0.53 | 0.02 | 0.01 |
|            | 4 | 0.18  | 0.51 | -0.24 | 0.03 | -0.18 | 0.20 | -0.09 | 0.43 | 0.45 | 0.01 | 0.83 | 0.38 |
| Dry        | 1 | 0.05  | 0.12 | -0.63 | 0.98 | -0.53 | 0.45 | 0.01 | -0.03 | 0.52 | -0.06 | 0.38 | 0.79 |
|            | 2 | 0.95  | 0.05 | -0.60 | 0.08 | -0.52 | 0.74 | 0.02 | 0.33 | 0.05 | 0.01 | 0.30 | 0.29 |
|            | 3 | 0.03  | -0.16 | -0.38 | 0.08 | -0.65 | 0.17 | 0.70 | 0.57 | 0.42 | -0.51 | 0.13 | -0.09 |
|            | 4 | 0.14  | 0.68 | -0.27 | 0.13 | -0.15 | 0.14 | -0.10 | 0.26 | 0.43 | 0.11 | 0.86 | 0.32 |
| TP Wet     | 1 | 0.66  | 0.11 | -0.98 | 0.69 | -0.89 | 0.84 | 0.18 | 0.49 | 0.58 | 0.00 | -0.32 | -0.09 |
|            | 2 | -0.01 | -0.07 | 0.08 | -0.36 | 0.16 | -0.23 | 0.22 | 0.25 | -0.30 | 0.01 | 0.85 | 0.92 |
|            | 3 | -0.10 | -0.06 | -0.12 | 0.02 | -0.35 | -0.12 | 0.95 | 0.22 | 0.14 | -0.59 | 0.00 | -0.30 |
|            | 4 | 0.08  | 0.99 | -0.11 | 0.10 | 0.01 | -0.06 | 0.11 | 0.14 | 0.18 | -0.27 | 0.16 | -0.23 |
| Chla Wet   | 1 | 0.92  | 0.13 | -0.80 | 0.28 | -0.72 | 0.85 | 0.10 | 0.43 | 0.23 | 0.03 | 0.05 | -0.13 |
|            | 2 | -0.07 | -0.09 | -0.09 | 0.07 | -0.04 | -0.23 | 0.38 | 0.18 | -0.08 | 0.14 | 0.72 | 0.96 |
|            | 3 | -0.10 | -0.13 | -0.45 | 0.92 | -0.35 | 0.28 | -0.05 | -0.19 | 0.34 | -0.09 | -0.37 | 0.19 |
|            | 4 | -0.05 | 0.20  | -0.33 | 0.26 | -0.40 | 0.18 | 0.18 | 0.46 | 0.91 | 0.11 | -0.29 | 0.05 |
|            | 5 | 0.09  | 0.12  | 0.16 | 0.00 | 0.39 | 0.07 | -0.76 | -0.35 | -0.01 | 0.77 | 0.14 | -0.15 |
| DO Wet     | 1 | 0.21  | 0.21 | -0.84 | 0.89 | -0.79 | 0.65 | 0.11 | 0.30 | 0.76 | -0.13 | -0.41 | -0.01 |
|            | 2 | 0.94  | 0.15 | -0.49 | 0.01 | -0.35 | 0.56 | 0.03 | 0.30 | -0.08 | 0.06 | -0.13 | -0.39 |
|            | 3 | 0.04  | -0.25 | -0.19 | -0.09 | -0.47 | -0.03 | 0.86 | 0.46 | 0.09 | 0.56 | 0.07 | 0.18 |
|            | 4 | -0.26 | 0.05  | 0.02 | 0.16 | 0.18 | -0.33 | 0.24 | -0.05 | -0.16 | 0.02 | 0.46 | 0.86 |

Notes: (1) Bold values indicate variables having moderate to high loadings on factors (F); F1–F5 refer to five independent factors. (2) A, S, I, Dc, BUL, OPN, AGR, VEG, WAT, GWD, and CIN and COUT, respectively, refer to subwatershed area, slope, imperviousness, distance from coast, built-up land, open land, agricultural land, vegetated land, waterbody, groundwater depth, and inlet and outlet concentrations.

in dry season, and agricultural land (−0.61 to −0.64) in both seasons. A moderate loading of wet season DO was also apparent on factor 1 (−0.54), which had moderate to high loadings of open land (0.55), built-up land (−0.69), waterbody (0.73), imperviousness (−0.74), and distance from the coast (0.89).

3.4. Estimations of the Relative Linkages of Water Quality Indicators With the Drivers

The minimum AIC criteria, in concert with an acceptable NSE, led to the inclusion of two to four PLS components to achieve optimum PLSR models for the four water quality indicators (Figure 5). Based on Moriasi et al. [2007], the fitting efficiency (NSE = 0.72–0.95) and accuracy (RSR = 0.22–0.51) of the optimal models suggested good predictions in both seasons (Table 4 and Figure S2). The power law-based PLSR models of Z-scores were significant at the 95% confidence level for all water quality indicators.

TN had the strongest linkage with the distance from coast and wet from dry seasons (β = 0.60 to 0.73; Table 4). Notable linkages of TN in both seasons were also apparent with the upstream reach concentrations (β = 0.32 to 0.34), agricultural land (β = 0.23 to 0.35), and imperviousness (β = −0.18 to −0.29). Further, the waterbody area of the draining subwatershed was moderately linked (β = −0.42) with TN in the dry season. Based on the aggregated linkages, the "external driver" component (βLN) — reflecting contributions from the upstream Everglades and/or the downstream coast — had 1.5 to 2 times stronger linkages with TN than that of the subwatershed "land use" (βLU), "upstream reach" (βCN), and "hydrology" (βH) components.
TP was most strongly linked with the upstream concentrations ($\beta = 0.47$ to 0.79; Table 4) and agricultural land ($\beta = 0.44$ to 0.48) in both seasons. Other noteworthy controls of TP included the hydrologic variables ($A$, $S$, and $I$), open land, vegetated land, and groundwater depth. However, the distance from coast was weakly linked with TP in both seasons. Contributions of the “upstream reach” component ($\beta_{CIN}$) had approximately 1.5, 2, and 12 times stronger linkages with TP in wet season than that of, respectively, the “land use” ($\beta_{LU}$), “hydrology” ($\beta_{H}$), and the “external driver” ($\beta_{DC}$) components. In contrast, the “land use” contribution component dominated TP in dry season; $\beta_{LU}$ was 1.2, 1.3, and 15.4 times stronger than $\beta_{CIN}$, $\beta_{H}$, and $\beta_{DC}$, respectively.

Similar to TP, Chl $a$ had relatively strong linkages with the upstream

![Figure 5. Plot of cross-validated (a) normalized AIC and (b) Nash-Sutcliffe efficiency (NSE) with the number of incorporated partial least squares components for both wet and dry seasons.](image)

| Predictor Variables | TN | TP | Chl $a$ | DO |
|--------------------|----|----|--------|----|
|                      | Wet | Dry | Wet | Dry | Wet | Dry | Wet | Dry | Wet | Dry |
| $A$                 | -0.07 | 0.12 | -0.12 | -0.24 | -0.29 | -0.35 | 0.09 | -0.13 |
| $S$                 | 0.20 | 0.10 | -0.25 | -0.25 | -0.01 | 0.08 | 0.01 | 0.25 |
| $I$                 | -0.18 | -0.29 | -0.21 | -0.14 | -0.16 | -0.27 | 0.13 | 0.14 |
| $D_s$               | 0.60 | 0.73 | 0.06 | -0.04 | 0.45 | 0.47 | -0.22 | 0.13 |
| BUL                 | 0.08 | 0.14 | 0.02 | 0.00 | -0.06 | 0.19 | 0.18 | 0.19 |
| OPN                 | -0.08 | -0.01 | -0.13 | -0.24 | -0.31 | -0.31 | -0.04 | 0.04 |
| AGR                 | 0.35 | 0.23 | 0.44 | 0.48 | 0.36 | 0.48 | 0.22 | -0.37 |
| VEG                 | -0.09 | -0.11 | 0.20 | 0.10 | 0.10 | 0.06 | 0.07 | -0.35 |
| WAT                 | -0.04 | -0.42 | -0.10 | 0.02 | 0.03 | -0.22 | 0.07 | 0.03 |
| GWD                 | 0.12 | 0.16 | 0.28 | 0.20 | 0.23 | 0.25 | 0.36 | 0.15 |
| $C_{IN}$            | 0.32 | 0.34 | 0.79 | 0.47 | 0.68 | 0.35 | 0.21 | -0.05 |

| Model Statistics |
|------------------|---|
| PLS components  | 3 4 4 3 2 4 2 2 |
| NSE              | 0.75 | 0.89 | 0.95 | 0.81 | 0.85 | 0.72 | 0.74 | 0.81 |
| RSR              | 0.49 | 0.33 | 0.22 | 0.43 | 0.37 | 0.51 | 0.50 | 0.43 |

| Aggregated Linkages |
|---------------------|
| $\beta_{LU}$       | 0.38 | 0.51 | 0.51 | 0.55 | 0.49 | 0.64 | 0.30 | 0.55 |
| $\beta_{H}$        | 0.30 | 0.37 | 0.45 | 0.42 | 0.41 | 0.52 | 0.39 | 0.34 |
| $\beta_{DC}$       | 0.60 | 0.73 | 0.06 | 0.04 | 0.45 | 0.47 | 0.22 | 0.13 |
| $\beta_{CIN}$      | 0.32 | 0.34 | 0.79 | 0.47 | 0.68 | 0.35 | 0.21 | 0.05 |

Notes: (1) $A$, $S$, $I$, $D_s$, BUL, OPN, AGR, VEG, WAT, GWD, and $C_{IN}$, respectively, refer to subwatershed area, slope, imperviousness, distance from coast, built-up land, open land, agricultural land, vegetated land, waterbody, groundwater depth, and inlet concentrations. (2) The aggregated relative linkages of the watershed “land use” ($\beta_{LU}$), “hydrology” ($\beta_{H}$), “upstream reach” ($\beta_{CIN}$), and the “external driver” ($\beta_{DC}$) components were calculated, respectively, by using equations (1)–(4).
reach concentrations ($\beta = 0.35$ to $0.68$; Table 4) and agricultural land ($\beta = 0.36$ to $0.48$) across the two seasons. However, unlike TP and alike TN, Chl $a$ was strongly linked with the distance from coast ($\beta = 0.45$ to $0.47$). Appreciable linkages of Chl $a$ in both seasons were also noted with the subwatershed area, imperviousness, open land, and groundwater depth. The “upstream reach” and “land use” contributions dominated Chl $a$ in, respectively, wet and dry seasons—exhibiting 1.2 to ~2 times stronger linkages with Chl $a$ than that of the other components.

DO was most strongly linked with the groundwater depth ($\beta = 0.36$; Table 4) in wet season and with the agricultural and vegetated lands ($\beta = -0.35$ to $-0.37$) in dry season. Other notable linkages of DO in both seasons include that with the subwatershed imperviousness, built-up land, distance from coast, and groundwater depth. The subwatershed “hydrology” component ($\beta_{Dc}$) had 1.3 to ~2 times stronger linkages with DO in wet season than that of “land use” ($\beta_{LU}$), “external driver” ($\beta_{Dc}$), and “upstream reach” ($\beta_{OUT}$). In contrast, the “land use” predominantly controlled DO in dry season; $\beta_{LU}$ was 1.6, 4.3, and 10.2 times stronger than $\beta_{Dc}$, $\beta_{Dc}$, and $\beta_{OUT}$, respectively.

4. Discussion

The strong positive linkages of TN with the “external driver” component (represented by the distance from coast, $D_c$; e.g., see Table 4) in both seasons underline Everglades as the primary source of nitrogen in the coastal-urban streams of southeast Florida. However, the weak linkage of TP with $D_c$ reiterates that the Everglades is a minor source of phosphorus for the southeast Florida streams [Rudnick et al., 1999]. The external control of stream TN was further manifested in C-12, which is not directly connected with the Everglades and had much lower concentrations of TN than the other canals (see Table 51). The positive linkages of TN with $D_c$ also suggest that the in-stream TN decreases away from the Everglades toward the coast. The majority of Everglades’ nitrogen comes from the water conservation area (WCA) and/or agricultural lands as organic nitrogen (particulate and dissolved), compared to the dissolved inorganic forms (NH$_3$ + NO$_3$) [BCEPD, 2007]. The dissolved organic nitrogen reduces through denitrification [Graves et al., 2004], whereas the particulate organic nitrogen tends to settle at the bottom of the shallow (depth < 3.048 m) freshwater canals. Further, increased salinity toward the coast might impact microbial activities—decreasing the rates of mineralization and transport of settled particulate organic nitrogen into the water column [Haynes, 2003; Jackson and Vallaire, 2009].

The moderate to strong positive linkages between the inlet and outlet concentrations ($C_{IN}$ and $C_{OUT}$) of individual nutrients (TN and TP) emphasize the high influence of upstream reach on the downstream water quality [Alexander et al., 2007]. The linkages of TN and TP with the watershed agricultural land can be attributed to the wash-off of fertilizers by rainfall-runoff [Li et al., 2009; Bu et al., 2014; R. Wan et al., 2014; Y. Wan et al., 2014]. Despite the predominantly built-area fraction (74%) of the study watersheds (Table 1), the agricultural land use dominated over the built-up land to contribute in-stream nutrients in both seasons (e.g., Table 4). This is consistent with the operation of regional wastewater treatment plants, which stopped discharging directly into the canals since 1989 [Broward County Development of Planning and Environmental Protection (BCDEP), 2001]. Further, the negative linkages of TN and TP with the watershed imperviousness suggested the dilution of in-stream nutrients by the corresponding surface runoff. However, the direction of predictor versus response relationship in PLSR is mechanistically more meaningful for the dominant predictors than that of the less important or weak predictors. For example, the correlation analysis (Table 2) showed a moderate positive correlation of open land area fraction (OPN) with TN. In contrast, the PLSR upon resolving multicollinearity estimated a relatively very weak (i.e., not meaningful) negative linkage between OPN and TN (Table 4).

The moderate positive correlation between OPN and TN is therefore indirect and spurious, which can be attributed to the mutual correlation (multicollinearity) between OPN and distance from the coast ($D_c$). In fact, TN, $D_c$, and OPN increase away from the coastline toward the upstream of canals near the Everglades (see more details in section 3.1). The relative linkages of Chl $a$ with the individual drivers and aggregated process components mostly reflected the corresponding linkages of TP, emphasizing the limiting nutrient in the southeast Florida canals. The finding complements the previous studies that mainly focused on the south Florida bays, estuaries, and/or Everglades. For example, Briceño et al. [2013] reported the limiting effect of TP on phytoplankton productivity in south Florida’s estuaries and bays. Noe et al. [2001] and Childers et al. [2006] reported the
phosphorus-limited nature of the large freshwater-estuarine landscape in Florida Everglades. Lapointe and Bedford (2010) reported phosphorus-limited biomass production by algal bloom on coral reefs off the south-east Florida. Our study expanded the research into the urban region between the Biscayne Bay and the Indian River Lagoon Watersheds. Our results indicate that the phosphorus limitation is also prevalent in the inland managed canals of southeast Florida.

The contributions of the “upstream reach” and “land use” components dominated both TP and Chla in wet and dry seasons, respectively. The moderate to strong linkages between the upstream and the immediate downstream concentrations (CIN and COUT) of Chla appeared to represent the corresponding linkages for the individual nutrients. The positive linkage of Chla with agricultural land can be attributed to the nutrient (TN and TP) runoff from fertilizer usage [Corkum, 1996]. Similarly, the positive linkage of Chla with D, reiterates the higher nitrogen availability in the upstream of canals that receives Everglades’ runoff, leading to higher phytoplankton biomass [Brand, 2002; NRC, 2002]. However, the increased salinity stress and turbidity (tidal mixing and sediment resuspension) may have contributed to the lower in-stream Chla near the coastline by impacting the photosynthetic rates of phytoplankton [Cloern, 1987]. Further, the negative linkages of Chla with the subwatershed area, open land, and imperviousness can be attributed to the dilution of in-stream TP and/or TN by the consequent higher runoff generation [Kang et al., 2010].

The positive linkages of the groundwater depth from the land surface (GWD) with the stream nutrients, Chla, and DO (Table 4) indicated the enhanced interactions of streamflow with the shallow groundwater table in the region (see Table 1). The groundwater control of DO in the Broward canals was also reported in a previous study [BCEPD, 2007]. The linkages can be attributed to the underlying Biscayne aquifer, which has a high transmissivity [Lietz, 1999] and is typically low in DO (median = 0.15 mg/L) and nutrients (median TN < 0.96 mg/L, median TP < 0.01 mg/L) [Bradner et al., 2005]. The positive linkage of DO with watershed imperviousness (Table 4) reiterated the dilution of in-stream nutrients and consequent enhancement of DO by higher runoff generation from the reduced permeability. Further, prohibition of direct effluent discharge from the wastewater treatment plants into the canals has largely reduced the pollution inputs from the urban land [BCDPEP, 2001]. This further emphasizes the dilution of nutrients by surface runoff from built-up land—contributing to the improved stream water quality, as evident from its intriguing positive linkage with DO (Table 4). Overall, the stronger linkage of stream DO with groundwater in the wet season eventually led to the domination of the “hydrology” component over the subwatershed “land use,” “upstream reach,” and “external driver” components. However, the dictating linkage of the “land use” component with DO in the dry season reflected the deteriorating impacts of incoming nutrients and organic matters (e.g., litters) from, respectively, the agricultural and vegetated lands during the low flow period.

Alongside the characterization and source identification of coastal water quality [e.g., Boyer et al., 1997; Boyer, 2006; Caccia and Boyer, 2007], determination of the relative dominance of one driver over the others is essential to understand the cause and effect relationships [Boyer et al., 2009]. Our study advanced the knowledge of coastal-urban water quality by estimating the relative influence of the land use and hydrologic drivers on stream water quality indicators by appropriately resolving multicollinearity. The estimated relative linkages can be utilized to identify the management targets and priorities to achieve/maintain healthy stream ecosystems in southeast Florida and similar regions, as required by the Clean Water Act (Federal Water Pollution Control Act, 2002). Since Everglades (the external driver) is the largest contributor of in-stream TN, the major reduction of nitrogen loads should be achieved in the water conservation areas (WCAs) ecologically and/or by adding engineering solutions (e.g., detention ponds). This reiterates the importance of the Comprehensive Everglades Restoration Plan to significantly decrease the nitrogen concentration in the downstream freshwater canals [Y. Wan et al., 2014]. Controlling (e.g., treatment) the watershed agricultural runoff should also receive the highest priority in land management to reduce in-stream nutrients (TN and TP). Further, removal of nutrients at the upstream locations would be effective to improve the downstream water quality. The reduction of nutrients, in turn, would reduce algal biomass (e.g., Chla) and increase DO that was typically low in the coastal-urban streams (see Table 1).

5. Conclusions

The relative linkages of stream water quality and environmental health indicators with the land use and hydrologic drivers were reliably estimated for the coastal-urban watersheds of southeast Florida by
Acknowledgments
The research was funded by a National Science Foundation (NSF) CAREER Award to Omar I. Abdul-Aziz (NSF CBET Award 1561942/1454435). We are grateful to Robert Bernhard of the Broward County, Florida, for sharing study area map and reports. All the data used in this study are adequately described in the main text, figures, tables, and supporting information. Further, the sources of all the data can be found in the reference list. Research findings and conclusions are those of the authors and do not necessarily reflect the views of NSF.

References
Ahn, J. H., S. B. Grant, C. Q. Surbeck, P. M. DiGiacomo, N. P. Nezlin, and S. Jiang (2005), Coastal water quality impact of stormwater runoff from an urban watershed in southern California, Environ. Sci. Technol., 39(16), 5940–5953.
Akaike, H. (1974), A new look at the statistical model identification, IEEE Trans. Autom. Control, 19(6), 716–723.
Alexander, R. B., E. W. Boyer, R. A. Smith, G. E. Schwarz, and R. B. Moore (2007), The role of headwater streams in downstream water quality, J. Am. Water Resour. Assoc., 43(1), 41–59.
Alvarez-Cabria, M., J. Barquin, and F. J. Perlas (2016), Modelling the spatial and seasonal variability of water quality for entire river networks: Relationships with natural and anthropogenic factors, Sci. Total Environ., 545, 152–162.
Andersen, C. B., G. P. Lewis, and K. A. Sargent (2004), Influence of wastewater-treatment effluent on concentrations and fluxes of solutes in the Bush River, South Carolina, during extreme drought conditions, Environ. Geosci., 11(1), 28–41.
Badruzzaman, M., J. Pinzon, J. Oppenheimer, and J. G. Jacangelo (2012), Sources of nutrients impacting surface waters in Florida: A review, J. Environ. Manage., 109, 80–92.
Bay, S., B. H. Jones, K. Schiff, and L. Washburn (2003), Water quality impacts of stormwater discharges to Santa Monica Bay, Mar. Environ. Res., 56(1), 205–223.
Boyer, J. N. (2006), Shifting N and P limitation along a north-south gradient of mangrove estuaries in South Florida, GeoHealth, 10.1002/2017GH000058.
Boyer, J. N. (2006), Spatial patterning of water quality in Biscayne Bay, Florida as a function of land use and water management, Mar. Pollut. Bull., 52(1), 146–1429.
Boyer, J. N., J. W. Fourqurean, and R. D. Jones (1997), Spatial characterization of water quality in Florida Bay and Whitewater Bay by multivariate analyses: Zones of similar influence, Estuaries Coasts, 20(4), 743–758.
Boyer, J. N., C. R. Kelble, B. P. Otter, and D. T. Rudnick (2009), Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA, Ecol. Indic., 9(6), 556–567.
Braddock, A. F., B. F. McNherson, R. L. Miller, G. Kish, and B. Bernard (2005), Quality of ground water in the Biscayne aquifer in Miami-Dade, Broward, and Palm Beach Counties, Florida, 1996–1998, with emphasis on contaminants, U.S. Geol. Surv. Open File Rep., 2004-1438, Reston, Va.
Briceño, H. O., J. N. Boyer, P. Harlem, and J. Castro (2013), Biogeochemical classification of South Florida’s estuarine and coastal waters, Mar. Pollut. Bull., 75(1), 187–204.
Brand, L. E. (2002), The transport of terrestrial nutrients to South Florida coastal waters, in The Everglades, Florida Bay and Coral Reefs of the Florida Keys: An Ecosystem Sourcebook, edited by J. W. Porter and K. G. Porter, pp. 361–414, CRC Press, Washington, D.C.
Broward County (2016), Environmental assessment team. [Available at http://www.broward.org/NaturalResources/Lab/AboutUs/Pages/EnvironmentalAssessmentTeam.aspx, (Accessed on April 04, 2016).]
Broward County Development of Planning and Environmental Protection (BCDPEP) (2001), Broward County, Florida historical water quality atlas: 1972–1997, Technical Report Series TR: 01-03, BCDPEP.
Broward County Environmental Planning and Community Resilience Division (2015), Broward County’s ambient water quality program. [Available at http://www.broward.org/NaturalResources/Lab/Pages/canalwaterquality.aspx, (Accessed on March 31, 2015).]
Broward County Environmental Protection Department (BCEPD) (2007), Broward County Florida water quality atlas: Freshwater Canals 1998–2003, Technical Report Series TR 07-03, BCEPD.
Bu, H., W. Meng, Y. Zhang, and J. Wan (2014), Relationships between land use patterns and water quality in the Taizi River basin, China, Ecol. Indic., 41, 187–197.
Caccia, V. G., and J. N. Boyer (2005), Spatial patterning of water quality in Biscayne Bay, Florida as a function of land use and water management, Mar. Pollut. Bull., 50(11), 1416–1429.
Carey, R. O., J. W. Fourqurean, and R. D. Jones (1997), Spatial characterization of water quality in Florida Bay and Whitewater Bay by multivariate analyses: Zones of similar influence, Estuaries Coasts, 20(4), 743–758.
Carey, R. O., K. W. Migliaccio, and M. T. Brown (2011a), Nutrient discharges to Biscayne Bay, Florida: Trends, loads, and a pollutant index, Sci. Total Environ., 409(3), 530–539.
Carey, R. O., K. W. Migliaccio, Y. Li, B. Schafer, A. K. Gregory, and M. T. Brown (2011b), Land use disturbance indicators and water quality variability in the Biscayne Bay watershed, Ecol. Indic., 11, 1093–1104.
Carey, R. O., G. J. Hochmuth, C. J. Martinez, T. H. Boyer, M. D. Dukes, G. S. Toor, and J. L. Cisar (2013), Evaluating nutrient impacts in urban watersheds: Challenges and research opportunities, Environ. Pollut., 173, 136–149.
Chang, H. (2008), Spatial analysis of water quality trends in the Han River basin, South Korea, Water Res., 42, 3285–3304.
Chapra, S. C. (1997), Water Quality Modeling, pp. 3–20, Watermark Press, Long Grove, Illinois.
Childers, D. L., J. N. Boyer, S. E. Davis, C. J. Madden, D. T. Rudnick, and F. H. Sklar (2006), Relating precipitation and water management to nutrient concentrations in the oligotrophic “upside-down” estuaries of the Florida Everglades, Limnol. Oceanogr., 51(1), 602–616.
Cloern, J. E. (1987), Turbidity as a control on phytoplankton biomass and productivity in estuaries, Cont. Shelf Res., 7(11–12), 1367–1381.
Cooper, R. M., and J. Lane (1987), An atlas of Eastern Broward County surface water management basins, DRE 231, SWFMD–Water Resources Division.

ABDUL-AZIZ AND AHMED COASTAL-URBAN STREAM HEALTH 193
Corkum, L. D. (1996), Responses of chlorophyll-a, organic matter, and macroinvertebrates to nutrient additions in rivers flowing through agricultural and forested land, Arch. Hydrobiol., 136(3), 391–411.

de Jong, S. (1993), SIMPLS: An alternative approach to partial least squares regression, Chemom. Intell. Lab. Syst., 18(3), 251–263.

de Winter, J. C., F. Dodou, and P. A. Wieringa (2009), Exploratory factor analysis with small sample sizes, Multivar. Behav. Res., 44(2), 147–181.

Federal Water Pollution Control Act (2002), Federal Water Pollution Control Act, as amended through P.L. 107–303, November 27, 2002, in U.S. Congress, Washington, D.C.

Florida Department of Environmental Protection (2016), Florida department of environmental protection geospatial open data. [Available at http://geo.data.dep.state.fl.us/datasets/d3ba2bd3c9507422a86c3e5ebf96e49_c_0, (Accessed on October 10, 2016.).]

Graves, G. A., Y. Wan, and D. L. Fike (2004), Water quality characteristics of storm water from major land uses in south Florida, J. Am. Water Resour. Assoc., 40(6), 1405–1419.

Hubert, M., and K. V. Branden (2003), Robust methods for partial least squares regression, J. Chemom., 17(10), 537–549.

Ishita, K. S., and O. I. Abdul-Aziz (2015), Relative linkages of canopy-level CO2 fluxes with the climatic and environmental variables for deciduous forests, Environ. Manage., 55(4), 943–960.

Jackson, C. R., and S. C. Vallaire (2009), Effects of salinity and nutrients on microbial assemblages in Louisiana wetland sediments, Wetlands, 29(1), 277–287.

Johnson, R. A., and D. W. Wichern (2007), Applied Multivariate Statistical Analysis, 6th ed., pp. 481–538, Pearson Prentice Hall, Upper Saddle River, N. J.

Jolliffe, I. T. (2002), Principal Component Analysis, 2nd ed., pp. 150–165, Springer, New York.

Kang, J. H., S. W. Lee, K. H. Cho, S. J. K, M. S. Cha, and J. H. Kim (2010), Linking land-use type and stream water quality using spatial data of fecal indicator bacteria and heavy metals in the Yeonggan river basin, Water Res., 44(14), 4143–4157.

Kuhn, M., and K. Johnson (2013), Applied Predictive Modeling, pp. 101–137, Springer, New York.

Lapointe, B. E., and B. J. Bedford (2010), Ecology and nutrition of invasive Caulerpa brachypus f. parvifolia blooms on coral reefs off southeast Florida, U.S. Harmful Algae, 9(1), 1–12.

Li, S., S. Gu, X. Tan, and Q. Zhang (2009), Water quality in the upper Han River basin, China: The impacts of land use/land cover in riparian buffer zone, J. Hazard. Mater., 165, 317–324.

Lücke, A. C. (1999), Methodology for estimating nutrient loads discharged from the east coast canals to Biscayne Bay, Miami-Dade County, Florida, Water-Resources Investigations Report 99-4094, U.S. Geological Survey, Tallahassee, Fl.

Liu, D., X. Chen, and Z. Lou (2010), A model for the optimal allocation of water resources in a saltwater intrusion area: A case study in Pearl River delta in China, Water Resour. Manage., 24(1), 63–81.

Mahbub, P., A. Goonetilleke, and G. A. Ayoko (2011), Prediction model of the buildup of volatile organic compounds on urban roads, Environ. Sci. Technol., 45(10), 4433–4439.

Menció, A., and J. Mas-Pla (2008), Assessment by multivariate analysis of groundwater-surface water interactions in urbanized Mediterranean streams, J. Hydrol., 329(3), 355–366.

Moriasi, D. N., G. J. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith (2007), Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, Trans. Am. Soc. Agric. Biol. Eng., 50(3), 885–900.

Nagy, R. C., B. G. Lockaby, L. Kalin, and C. Anderson (2012), Effects of urbanization on stream hydrology and water quality: The Florida Gulf Coast, HydroL Processes, 26(13), 2019–2030.

National Land Cover Database (2015), Available at http://www.mrlc.gov/nlcd11_data.php, (Accessed on October 10, 2016.).

Newcomer, T. A., S. S. Kaushal, P. M. Mayer, A. R. Shields, E. A. Canuel, P. M. Groffman, and A. J. Gold (2012), Influence of natural and novel organic carbon sources on denitrification in forest, degraded urban, and restored streams, Ecol. Monogr., 82(4), 449–466.

Noe, G. B., D. L. Childers, and R. D. Jones (2001), Phosphorus biogeochemistry and the impact of phosphorus enrichment: Why is the Everglades so unique?, Ecosystems, 4(7), 603–624.

Peres-Neto, P. R., D. A. Jackson, and K. M. Somers (2003), Giving meaningful interpretation to ordination axes: Assessing loading significance in principal component analysis, Ecology, 84(9), 2347–2363.

Rietz, D. N., and R. J. Haynes (2003), Effects of irrigation-induced salinity and sodicity on soil microbial activity, Soil Biol. Biochem., 35(6), 845–854.

Robinson, C. T., N. Schwirth, S. Baumgartner, and C. Stamm (2014), Spatial relationships between land-use, habitat, water quality and lotic macroinvertebrates in two Swiss catchments, Aquat. Sci., 76, 375–392.

Rudnick, D. T., Z. Chen, D. Childers, J. N. Boyer, and T. Fontaine (1999), Phosphorus and nitrogen inputs to Florida Bay: The importance of the Everglades watershed, Ecosystems, 99, 398–416.

Schueler, T. (2003), Impacts of Impervious Cover on Aquatic Systems, pp. 55–91, Center for Watershed Protection, Ellicot City, Md.

Schumann, S., L. P. Nolte, and G. Zheng (2013), Comparison of partial least squares regression and principal component regression for pelvic shape prediction, J. Biomech., 46(1), 197–199.

Shrestha, S., and F. Kazama (2007), Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji River Basin, Japan, Environ. Modell. Softw., 22, 464–475.

Simeonov, V., J. A. Stratis, C. Samara, G. Zachariadis, and D. Voutsas (2003), Assessment of the surface water quality in northern Greece, Water Res., 37(17), 4119–4124.

South Florida Water Management District (SFWMD) (2011), 2009 South Florida Water Management District photoindex interpretation key (version 1.5).

South Florida Water Management District (SFWMD) (2015a), Available online at http://www.sfwmd.gov/porta/portal/page/portal/levelthree/ weather%20%20water (Accessed on March 01, 2015).

South Florida Water Management District (SFWMD) (2015b), Available online at http://sfwmmp maps.arcgis.com/home/item.html?id=989eb66b13b3bf3c9d3d3042eaf0904 (Accessed on March 02, 2015).

Tran, C. P., R. W. Bode, A. J. Smith, and G. S. Kleppel (2010), Land-use proximity as a basis for assessing stream water quality in New York State (USA), Ecol. Indic., 10, 727–733.

Tu, J. (2009), Combined impact of climate and land use changes on streamflow and water quality in eastern Massachusetts, USA, J. Hydrol., 379, 268–283.
United States Geological Survey (USGS) (2015a), Available online at http://viewer.nationalmap.gov/viewer/ (Accessed on March 01, 2015).
United States Geological Survey (USGS) (2015b), Available online at http://maps.waterdata.usgs.gov/mapper/index.html (Accessed on April 22, 2015).
Wan, R., S. Cai, H. Li, G. Yang, Z. Li, and X. Nie (2014), Inferring land use and land cover impact on stream water quality using a Bayesian hierarchical modeling approach in the Xitiaoxi River watershed, China, J. Environ. Manage., 133, 1–11.
Wan, Y., Y. Qian, K. W. Migliaccio, Y. Li, and C. Conrad (2014), Linking spatial variations in water quality with water and land management using multivariate techniques, J. Environ. Qual., 43, 599–610.
Wold, S., M. Sjöström, and L. Eriksson (2001), PLS-regression: A basic tool of chemometrics, Chemom. Intell. Lab. Syst., 58(2), 109–130.
Yan, B., N. F. Fang, P. C. Zhang, and Z. H. Shi (2013), Impacts of land use change on watershed streamflow and sediment yield: An assessment using hydrologic modelling and partial least squares regression, J. Hydrol., 484, 26–37.
Zhang, J. Z., C. R. Kelble, C. J. Fischer, and L. Moore (2009), Hurricane Katrina induced nutrient runoff from an agricultural area to coastal waters in Biscayne Bay, Florida, Estuarine Coastal Shelf Sci., 84(2), 209–218.