An Evaluation of Nutrient and Geospatial Data Relationships in Selected Watersheds of the Houston-Galveston Region

October 24, 2013

Prepared for:

Houston-Galveston Area Council

Prepared by:

Michael Baker Corp.

This report was prepared in cooperation with the
Houston-Galveston Area Council, the Texas Commission on Environmental Quality, and
the Environmental Protection Agency and
Authorized under the Texas Clean Rivers Act and applicable Federal Law

Table of Contents

Tables

Figures

1 Introduction and Background

The Houston-Galveston Area Council (H-GAC) is the Clean Rivers Program (CRP) lead agency for the San Jacinto River Basin and three associated coastal basins – the Trinity-San Jacinto, the San Jacinto-Brazos and the Brazos-Colorado. H-GAC is a Council of Governments (COG), the regional authority for the Gulf Coast State Planning Region, and has been actively involved in regional water quality planning and public outreach activities since the 1970's.

The 2010 State of Texas Integrated Report (which includes a List of Impaired Water Bodies and is required under Section 305(b) of the Clean Water Act) identifies 44 of the 51 watersheds (classified segments) located within H-GAC's four Clean Rivers Program basins as impaired or with water quality concerns. These pollutants come from point sources such as domestic and industrial wastewater discharges, and non-point sources like on-site sewage facilities (OSSFs) and runoff from the landscape. The impact of runoff on water quality varies with the type of land cover in the watershed, and the uses to which the land is put (agriculture, forestry, residential use, parks). This project explores the impact of land use and land cover types on nutrient levels in associated waterways.

This project was conducted to identify potential correlations between land cover and/or inferred land use and ambient nutrient concentrations in selected streams in the region. The analysis includes evaluation of spatial and temporal variation. The information provided by this analysis is intended to increase the understanding of water quality concerns due to nutrient loads in runoff from watersheds and the sources of these loads. In addition, any identified correlations could help identify sources of nutrient loads to help prioritize implementation of structural and non-structural best management practices (BMPs) at locations where these would be most effective to improve receiving water quality. This approach was suggested by the U.S. Environmental Protection Agency (EPA) in a March 16, 2011 memo from Acting Assistant Administrator, Nancy Stoner, titled *Working in Partnership with States to Address Phosphorus and Nitrogen Pollution through Use of a Framework for State Nutrient Reductions*. For these purposes, H-GAC acquired available water quality data, geospatial data and modeling land cover information from already existing sources in order to perform the proper analyses.

This project included advanced statistical analyses of water quality and geospatial data by evaluating ambient nutrient data using GIS technology and modeling of land information to help develop correlations based on watershed characteristics. Analysis included an evaluation of the association of land cover/ land use changes over time and nutrient concentration trends. Nutrient trends were assessed using a variety of parametric and non-parametric methods.

2 Data

H-GAC currently has seven local partners collecting ambient water quality data, including nutrient data, through the Clean Rivers Program. All sampling and laboratory analysis methods are specified in H-GAC's Texas Clean Rivers Program FY 2012-2013 Regional Monitoring Activities Quality Assurance Project Plan (QAPP) and the Texas Commission on Environmental Quality (TCEQ) Nutrient Monitoring QAPP FY2012 and FY2013, as well as all past QAPP versions.

Routine ambient water quality data collected by CRP partners as well as the TCEQ Field Operations Division are stored in TCEQ's Surface Water Quality Monitoring Information System (SWQMIS) database and have undergone rigorous validation and verification processes outlined in the applicable QAPPs. Additionally, since 2007, only water quality data produced by a National Environmental Laboratory Accreditation Conference (NELAC) accredited laboratory may be added to SWQMIS.

H-GAC also maintains a centralized geospatial warehouse of both tabular (non-geographic) and spatial (geographic) datasets. Geographical Information System (GIS) staff in the Community & Environmental Planning Department (C&E) capture, manipulate, develop, analyze, store and display spatially referenced data to support a wide variety of applications ranging from sites assessments, environmental planning, urban planning, and spatial analysis.

2.1 Data Quality

Water quality data were acquired from one primary source, the TCEQ SWQMIS. Only 'non-qualified', routine, ambient, fixed station water quality data from SWQMIS collected after December 31, 1995, for the H-GAC region were used for statistical modeling of nutrient/land cover relationships. The data included all nutrient data and associated field parameters collected with a calibrated data sonde. All acquired water quality data were collected under TCEQ approved QAPPs. Some existing data collected prior to 1996 was used for the trend analyses in order to temporally "bracket" the land cover datasets to which the trends were compared.

The GIS data sets used in this project were acquired from reliable sources such as U.S. Geological Survey (USGS), U.S. Department of Agriculture (USDA), Texas Natural Resources Information System (TNRIS), TCEQ, US Census Bureau, COGs, and other local, regional, state and federal organizations or governments. A complete list of files and sources is provided in Appendix 5 of H-GAC's C&E Data Management Plan.

2.2 GIS Data

The GIS data used in the project includes geospatial software and special databases currently developed, stored, and/or maintained by H-GAC's C&E. The data were used as was available. The following data were considered:

- CRP stream network and station datasets
- 1996, 2001, 2006, and 2011 land cover data sets (these are only developed every five years)
- Soils
- Elevation
- Texas Road network
- Imperviousness
- Wastewater outfall dataset
- USGS HUC 8 and HUC 12 layers

2.3 Hydrology / Hydraulics

Hydrological and hydraulic data were obtained from several sources:

- Flow data from USGS
- Daily precipitation from National Oceanic and Atmospheric Agency (NOAA) Climactic Data Center and from Harris County Flood Control District
- Monthly average wastewater discharge data from Discharge Monitoring Reports provided by the TCEQ

2.4 Water Quality Data

The primary source for water quality data used in this analysis was TCEQ's SWQMIS database. The acquired data includes all nutrient data and associated field parameters collected with a calibrated data sonde. All acquired water quality data were collected in compliance with TCEQ's approved QAPPs.

For the monitoring stations selected in this project, the following parameters were obtained from SWQMIS:

- Total phosphorus
- Orthophosphate phosphorus
- Nitrogen as nitrate+nitrite
- Nitrogen as nitrate
- Nitrogen as ammonia
- Nitrogen as TKN
- *E. coli*
- Enterococci
- Temperature
- Specific conductance
- Dissolved oxygen

- Secchi transparency
- Total suspended solids
- pH
- Chlorophyll *a*

Table 1 shows the data that were used, the monitoring sites, number of data points, and drainage area characteristics. Figure 1 and Figure 2 present boxplot graphs of total phosphorus and total nitrogen by monitoring station and by land use. For values that were below the limit of detection or reporting limits, a value of half of the reporting limit was used.

Monthly average wastewater discharge data were obtained from Discharge Monitoring Reports provided by the TCEQ. Wastewater permit data were obtained from an H-GAC database that consists of data provided by the TCEQ.

Table 1: Summary of Data Ranges, Parameter Counts, and Categorical Variables

Monitoring Station	Watershed	Earliest Data	Most Recent Data	Number of Total Phosphorus Results	Number of Nitrate Results	Number of Total Nitrogen Results	Dominant Land use	Watershed Type	Wastewater Influence	Wastewater Discharge Level
11120	Cedar Bayou Above Tidal	01/30/1996	12/06/2011	9	9	8	Agricultural	Nonurban	Not Effluent Dominated	M
11125	Greens Bayou Above Tidal	01/31/1996	12/14/2011	19	19	4	Urban	Urban	Effluent Dominated	M
11135	Houston Ship Channel/Buffalo Bayou Tidal	01/29/1996	11/29/2011	9	9	2	Urban	Urban	Effluent Dominated	M
11139	Houston Ship Channel/Buffalo Bayou Tidal	01/29/1996	11/21/2011	23	23	6	Urban	Urban	Effluent Dominated	H
11312	Spring Creek	02/06/1996	12/13/2011	18	16	6	Forest	Nonurban	Effluent Dominated	H
11332	Cypress Creek	02/06/1996	12/13/2011	14	13	5	Agricultural	Nonurban	Effluent Dominated	M
11334	Caney Creek	01/03/1996	11/30/2011	14	24	2	Forest	Nonurban	Not Effluent Dominated	L
11367	Lake Creek	03/31/2011	05/17/2011	1	0	0	Agricultural	Nonurban	Not Effluent Dominated	L
11369	Greens Bayou Above Tidal	01/03/2001	12/08/2011	20	20	7	Urban	Urban	Effluent Dominated	H
11387	Whiteoak Bayou Above Tidal	01/29/1996	11/17/2011	10	12	8	Urban	Urban	Effluent Dominated	H
11467	Dickinson Bayou Above Tidal	03/13/1996	12/07/2011	11	17	6	Agricultural	Nonurban	Not Effluent Dominated	L
11484	Chocolate Bayou Above Tidal	03/13/1996	09/28/2011	5	7	6	Agricultural	Nonurban	Not Effluent Dominated	M

Monitoring Station	Watershed	Earliest Data	Most Recent Data	Number of Total Phosphorus Results	Number of Nitrate Results	Number of Total Nitrogen Results	Dominant Land use	Watershed Type	Wastewater Influence	Wastewater Discharge Level
12147	San Bernard River Above Tidal	03/13/1996	10/19/2011	9	10	8	Agricultural	Nonurban	Not Effluent Dominated	L
17746	Peach Creek	01/19/2006	01/12/2011	3	2	2	Forest	Nonurban	Not Effluent Dominated	L

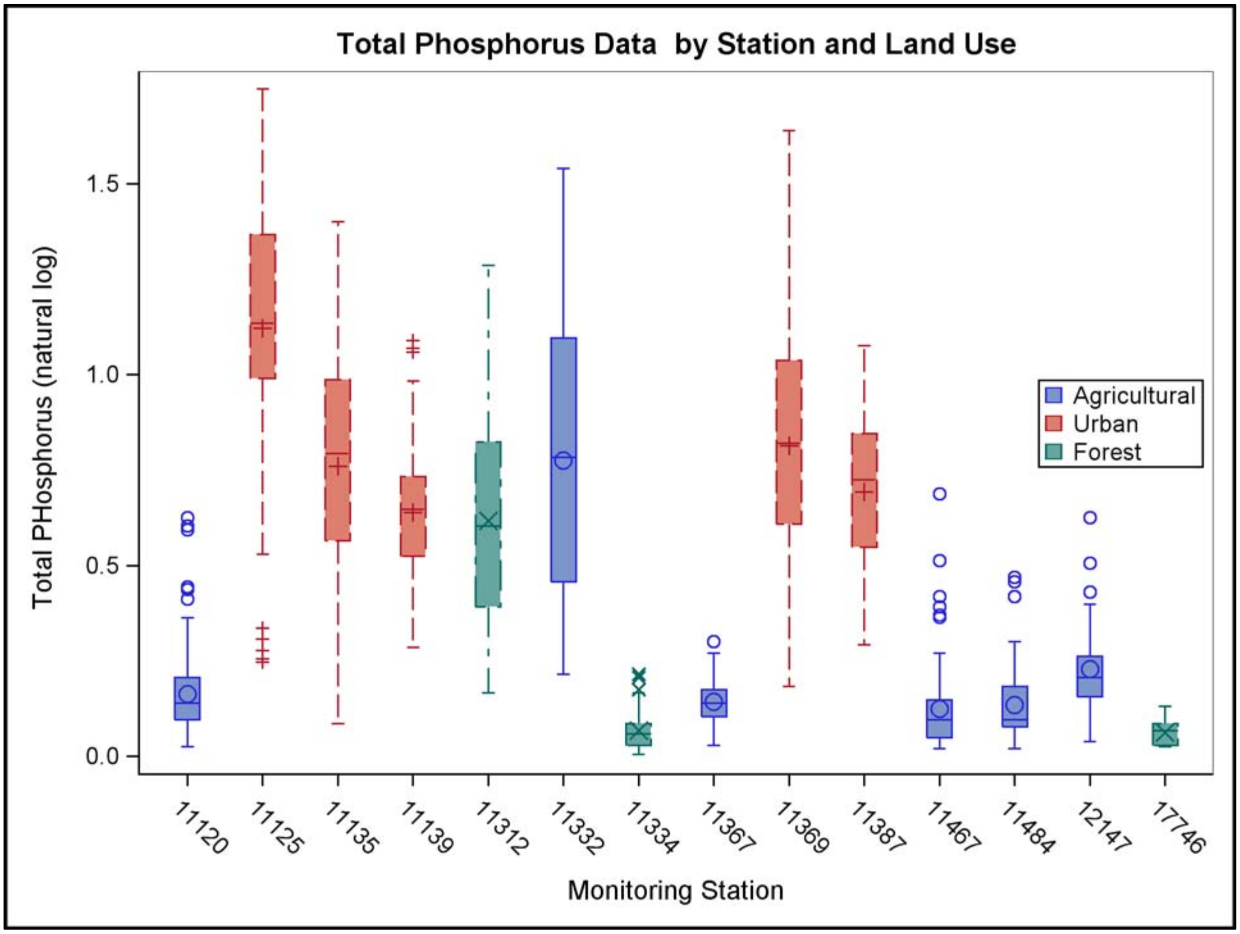

Figure 1: Boxplots of Total Phosphorus Concentration, by Station and Dominant Land Use

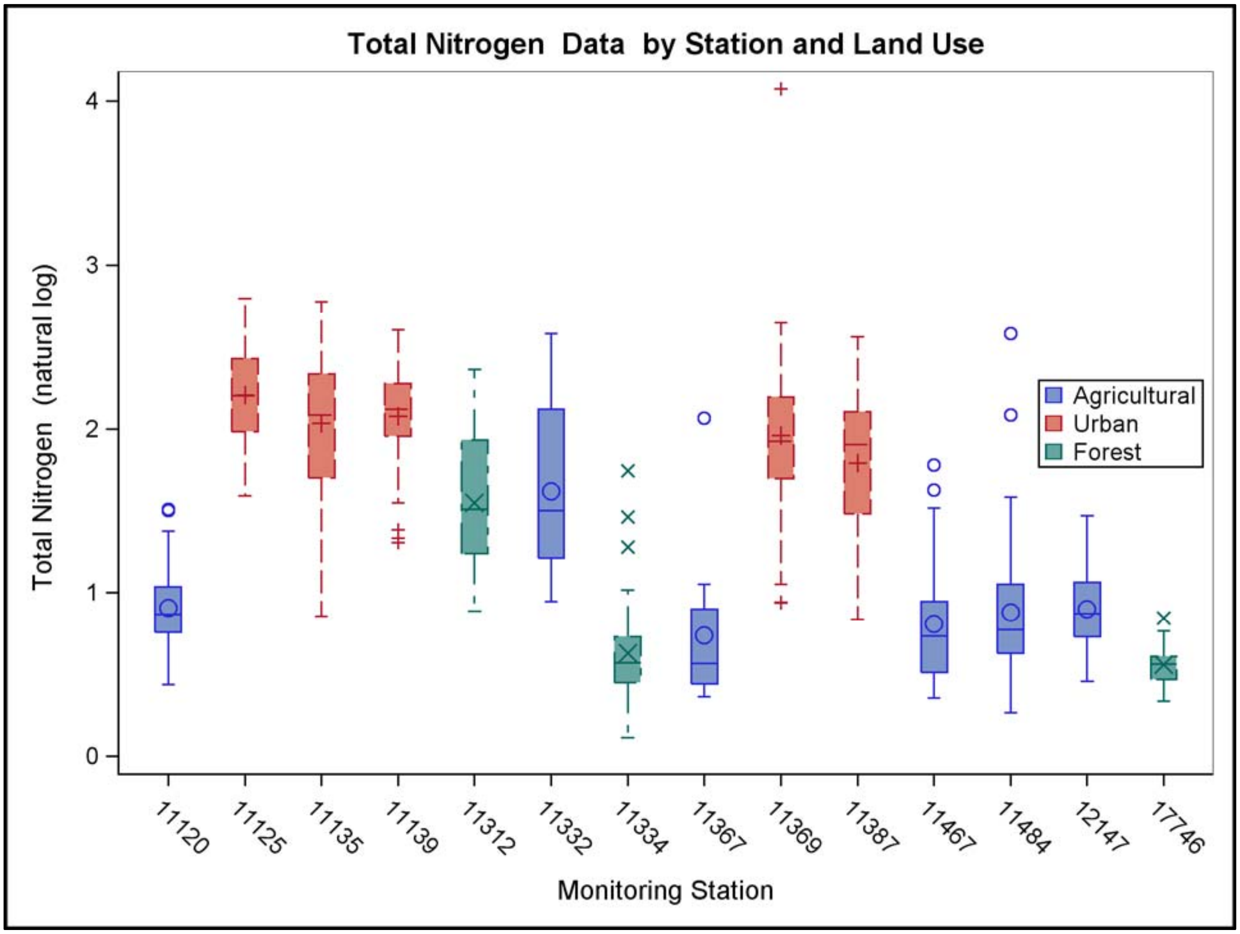

Figure 2: Boxplot of Total Nitrogen Concentration, By Station and Land Use

3 Methodology

This project includes several steps to define and understand the relationship between and cover and receiving water quality with a focus on nutrients. This project includes the evaluation of geospatial data and ambient nutrient data using GIS technology and advanced statistical analyses. For this report, H-GAC efforts are grouped into six primary steps:

1. Identification of monitoring stations with sufficient data,
2. Subwatershed identification and delineation,
3. Evaluation and definition of the characteristics of subwatersheds,
4. Land cover change detection analysis,
5. Statistical analyses to understand relationships between land cover data and receiving water quality, and
6. Development of load estimates for constituents of concern.

Each of these steps includes different types of actions. H-GAC used GIS and statistical software in several of these steps to process data, analyze data, and to assist in the analysis of results. The statistical analyses included an evaluation of the association of land cover changes over time and nutrient concentration trends. Regression and other statistical models were developed to relate land cover / land use data to nutrient concentrations in selected streams. Nutrient characteristics were assessed using parametric and non-parametric methods.

3.1 Selection of Monitoring Stations

The first step for this analysis was the identification of ambient monitoring stations in the region served by H-GAC where nutrient data were collected. H-GAC reviewed monitoring data collected at ambient monitoring stations to identify locations where sufficient data are available to produce statistically significant results. The initial list included monitoring stations with more than 20 data points for flow. From that list, 14 stations were selected for this project. These final selections were based on the presence of WWTPs upstream of the monitoring station, the spatial distribution of the stations, the major land types of the watersheds draining to the monitoring stations and land cover change over time. Table 2 includes a list of the monitoring stations that were selected for additional analysis in addition to the location of the stations (waterbody, longitude, and latitude) and station IDs.

Table 2: Selected Monitoring Stations

Station ID	Location	Longitude	Latitude
11367	LAKE CREEK	-95.578629	30.253798
11334	CANEY CREEK	-95.192123	30.148779
17746	PEACH CREEK	-95.169838	30.137611
11312	SPRING CREEK	-95.405762	30.092131

Station ID	Location	Longitude	Latitude
11120	CEDAR BAYOU	-94.985440	29.972281
11125	GARNERS BAYOU	-95.234062	29.933887
11332	CYPRESS CREEK	-95.598610	29.973663
11369	GREENS BAYOU	-95.228333	29.849722
11387	WHITEOAK BAYOU	-95.396942	29.775000
11139	BRAYS BAYOU	-95.412033	29.697258
11135	SIMS BAYOU	-95.445953	29.618767
11467	DICKINSON BAYOU	-95.170050	29.435925
11484	CHOCOLATE BAYOU	-95.323160	29.371076
12147	SAN BERNARD RIVER	-95.893330	29.313055

3.2 Subwatershed Identification and Delineation

After the preliminary selection of monitoring stations, the next step was to identify the watersheds that drain to the stations and the characteristics of these watersheds. H-GAC used GIS software and geospatial data to identify the drainage areas to the selected monitoring stations and to evaluate the characteristics of these drainage areas. The main geospatial data used for this step included:

- HUC-8 watershed boundaries (from USGS),
- H-GAC CRP stream network data layer for the 15 county region,
- H-GAC CRP monitoring stations,
- Digital Elevation Model (DEM) with 10 meter resolution (USGS), and
- 1-foot contour line data (H-GAC).

Where H-GAC determined that the USGS subwatersheds should be revised, H-GAC delineated watersheds using the ArcSWAT tool package (ArcSWAT, 2012; Spatial Science Lab, Texas A&M University). These were delineated with respect to each selected CRP station and based on the elevation values from DEM files and other terrain information (including contour line data), and stream network data. The methodology used within ArcSWAT was:

1. Load the DEM and stream network files through ArcGIS environment;
2. Extract DEM data by raster based masks for total drainage area for each selected stations (HUC8 watershed boundaries were used as the reference in deciding the extent of total drainage area);
3. Burn the stream network into the elevation data files to force the water flow towards existing streams (otherwise in flat land areas, flow accumulation may not be accurately represented);
4. Fill DEM to convert waterbodies and streams as sinks;

5. Create the flow direction grids to force the flow toward streams;
6. Create the flow accumulation grids up to minimum size of 75 hectares (accumulation is estimated based on the flow directions);
7. Define watershed outlets at the location of the selected CRP monitoring stations (the ArcSWAT tool can snap the defined outlet to the nearest stream, if the GIS data has the CRP monitoring station located away from the stream); and
8. Define each new watershed by merging the flow accumulated grids together.

The Better Assessment Science Integrating point and Non-point Sources (BASINS) automated/manual watershed delineation tool was also used to verify the delineated watersheds through a comparison of outputs from the two methods. Both the shapefile outputs from ArcSWAT and BASINS were exported into ArcMap and compared to determine boundary difference.

The delineated boundaries were used to identify the characteristics of the contributing area to each monitoring station. The sub-watersheds generated in the delineation process from individually-selected CRP station contributions were also merged to create larger watersheds where necessary. After completing the delineation process, the watersheds and stations were spatially joined to incorporate the station information into the attribute table for each watershed.

3.3 Watershed Characterization

After the subwatersheds draining to each monitoring station where identified, ArcGIS/Info 10.1 was used to identify the characteristics of the areas draining to each monitoring station. The delineated subwatersheds were assigned attributes, including[1]:

- Land cover types,
- Number of upstream wastewater treatment plants (WWTPs) with average permitted discharge, where available,
- Soil types (including hydrological soil groups),
- Drainage density and drainage area,
- Basin Relief (including minimum, maximum, and average elevation),
- Road density, and
- Imperviousness.

The data used and the methodologies for each of this are described in 3.3.1 below.

3.3.1 Land Cover Data and Change Analysis

Much research has been done to study the relationship between vegetative cover and hydrologic response. Vegetation stabilizes soils and prevents landslides and excessive soil

[1] Mohamoud, Y. 2004. Comparison of Hydrologic Responses at Different Watershed Scales. Ecosystem research Division, US EPA, Athens, GA. for characterization description for each parameter

erosion. In addition, cover also influences hydrologic response in a number of other ways. For example, forest cover directly affects such hydrologic processes as interception, rainfall infiltration, evaporation from plant canopy, and evapotranspiration.

The land use/ land cover datasets used were the NOAA Coastal Change Analysis Program (C-CAP) data sets for 1996, 2001, 2006, and 2011. Dataset of 2011 is not publicly available from NOAA C-CAP data center. This dataset was created by NOAA for the purposes of H-GAC's Clean River Program (Figure 3). The land cover data is in 30m resolution.

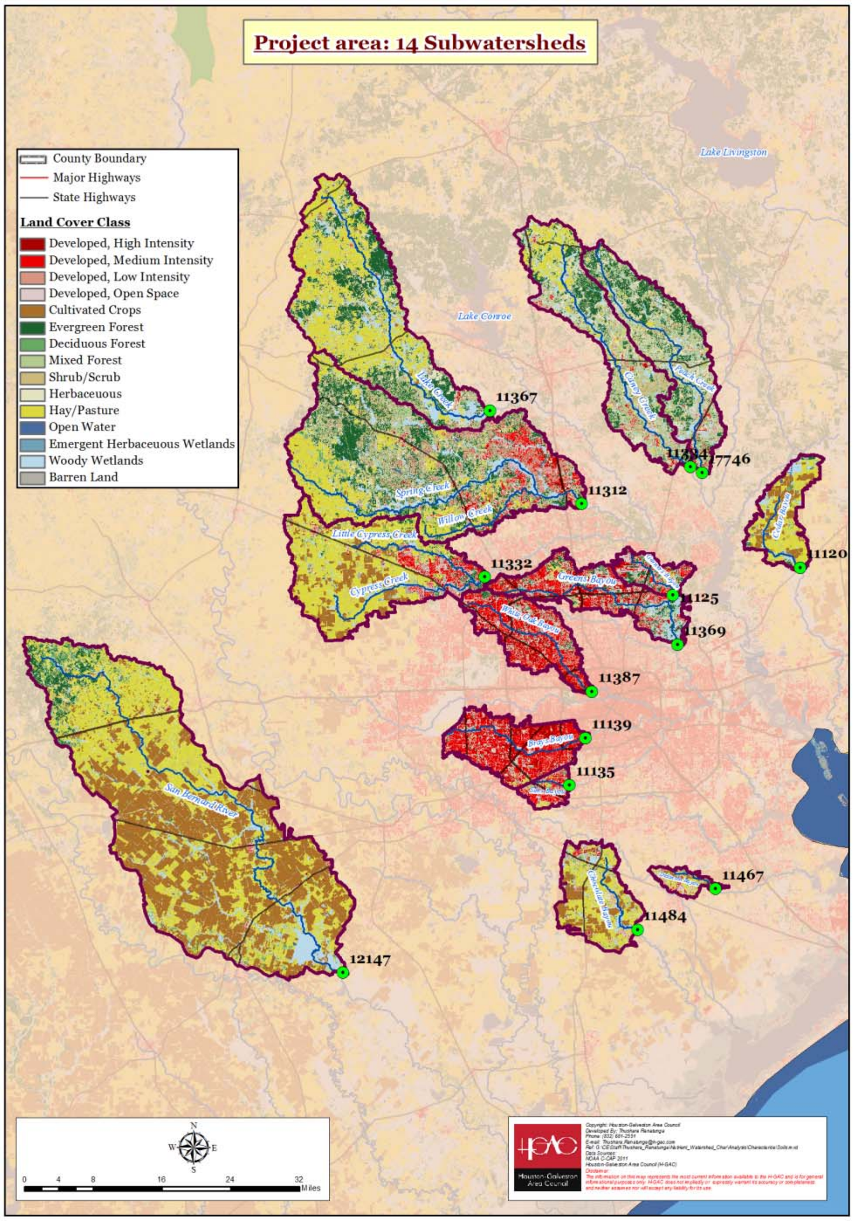

Figure 3: NOAA C-CAP Land Cover 2011

H-GAC reclassified the existing land use/ land cover types into categories that fit the needs of the project. The data was reclassified into nine classes based on Anderson land use and land cover classification (Anderson et al. 1976). The land cover classes of the NOAA C-CAP dataset are shown in Table 3. These are consistent with the National Land Cover Database (NLCD) datasets and include detailed land cover classes. The land use / land cover categories of the NOAA C-CAP datasets and the new classification categories are shown in Table 3.

Table 3: Land Use / Land Cover Categories Used in this Project

NOAA C-CAP 22 Class Classification	HGAC 9 Class Classification
Open Water (21)	Water (1)
Palustrine Aquatic Bed (22)	
Estuarine Aquatic Bed (23)	
Open Space Developed (5)	Developed Open Space (2)
Low Intensity Developed (4)	Developed (3)
Medium Intensity Developed (3)	
High Intensity Developed (2)	
Unconsolidated Shore (19)	Bare Land (4)
Barren Land (20)	
Deciduous Forest (9)	Forest (5)
Evergreen Forest (10)	
Mixed Forest (11)	
Scrub Shrub (12)	Scrub/Shrub (6)
Grassland (8)	Grasslands (7)
Pasture/Hay (7)	Cultivated (8)
Cultivated Land (6)	
Palustrine Forested Wetlands (13)	Wetlands (9)
Palustrine Scrub Shrub Wetlands (14)	
Estuarine Forested Wetlands (15)	
Estuarine Scrub Shrub Wetlands (16)	
Palustrine Emergent Wetlands (17)	
Estuarine Emergent Wetlands (18)	

After the categories were defined, the land cover data for each year were used to estimate the rates of change for each land cover type over time. To prepare the land cover files, the raster datasets were converted into vector format and the selected watershed boundaries were overlaid with the 9 categories. The polygons were simplified to combine the land cover class and watershed IDs. The land cover areas were estimated in acres for each watershed and area percentages were calculated. During the raster to vector conversion process data resolution was not changed and it retained the original resolution.

Nutrient water quality trends were estimated using statistical procedures and compared to land cover change derived from GIS analysis. The ArcGIS program was used to extract land cover data into tabular formats. The results of the analysis are shown in Section 4.1.1 and a map of land cover/ land use change is included to show the areas where land cover/ land use designations have changed. The trend analyses were performed based on the percentage of change for each land cover. Three types of trends were identified:

- Increasing: > 5% change
- Stable: + 5% to - 5%
- Decreasing: < -5%

The five percent threshold criteria was chosen by H-GAC staff based on the distribution of land cover percentage change. The tables of land cover change and nutrient trends were merged into the shapefile attributes. GIS maps were produced for each combination of land cover type and nutrient type. These maps were color coded based on the unique combination of increasing, stable, and decreasing land cover and nutrient trends.

Besides the above analysis, land cover conversions of one class to another from 1996 to 2011 were analyzed in ArcGIS spatial analysis. This procedure was used to identify the amounts of change of each land cover class and the spatial locations of these changes. The following procedure was used in the processing of land cover class change detection.

1. ArcGIS spatial analysis "combine" operation was applied for the two datasets (NOAA 1996 and NOAA 2011)
2. Run a logical class assignment (changed code) based on the changes in the cell values in python environment.
   EX:
   ```
   If lc_1996 = "forest" AND lc_2011 = "developed"
           return 1
   If lc_1996 = "forest" AND lc_2011 = "cultivated"
           return 2
   ```
3. ArcGIS "intersect" analysis was performed for the land cover change output and watershed layers. This provides a shapefile with watershed names and land cover change types.
4. Then the area was calculated for each change in acres.

Based on the land cover classification, watersheds were also defined as urban or non-urban for statistical analysis. If the dominant land use types in the watershed are "Developed" and "Developed Open Space", the watershed was designated to be "Urban." In addition, "Cultivated" land cover was defined as "Agricultural" land cover in the statistical analysis results.

3.3.2 Influence of WWTP Effluent

H-GAC took steps to limit the influence of WWTPs on ambient concentrations at the locations analyzed in this project. H-GAC identified the WWTPs in each subwatershed draining to one of the selected monitoring stations. Monitoring stations with the least amount of upstream WWTP flow were given preference when selecting the final list of monitoring stations for this project.

The ArcGIS spatial analysis point density estimation tool was used to estimate the number of WWTP outfalls in each watershed (output into a raster file as the number of points per square mile). Information about the permitted discharge amounts in the past is not readily available, and the current permitted discharge was assumed for all years at each station. This variable was not found to be significant in most analyses. Where it was found to be a significant predictor, however, the uncertainty in past values introduces additional uncertainty into the statistical models. In these situations parameter estimates, measures of model fit, and other statistical measures were interpreted with caution. Since effluent levels in 1996 vary from those in 2011 and current levels. The number of permitted outfalls in each of the watersheds is shown in Figure 4.

H-GAC also defined stations as either effluent dominated or not-effluent dominated. The stations were defined based on the ratio of the average monthly discharge reported in Discharge Monitoring Reports (DMR) obtained from TCEQ to the monthly average of measured flow at a nearby USGS gaging station. If the ratio was seen to exceed 0.5 at more than one point, the stream was considered to be effluent-dominated and the appropriate value was assigned to the variables. DMR data are not available before 2006, so values are less reliable for older data.

3.3.3 Soil Data Analysis

Soils play an important role in defining hydrologic responses of the watershed. They impact rainfall infiltration, percolation, and moisture storage. Their characteristics have a direct impact on the level of nutrient transport to nearby streams. Soil properties such as taxonomic groups, topographic locations, geomorphic features, and hydrologic groups can provide valuable information on the hydrologic characteristics of the area.

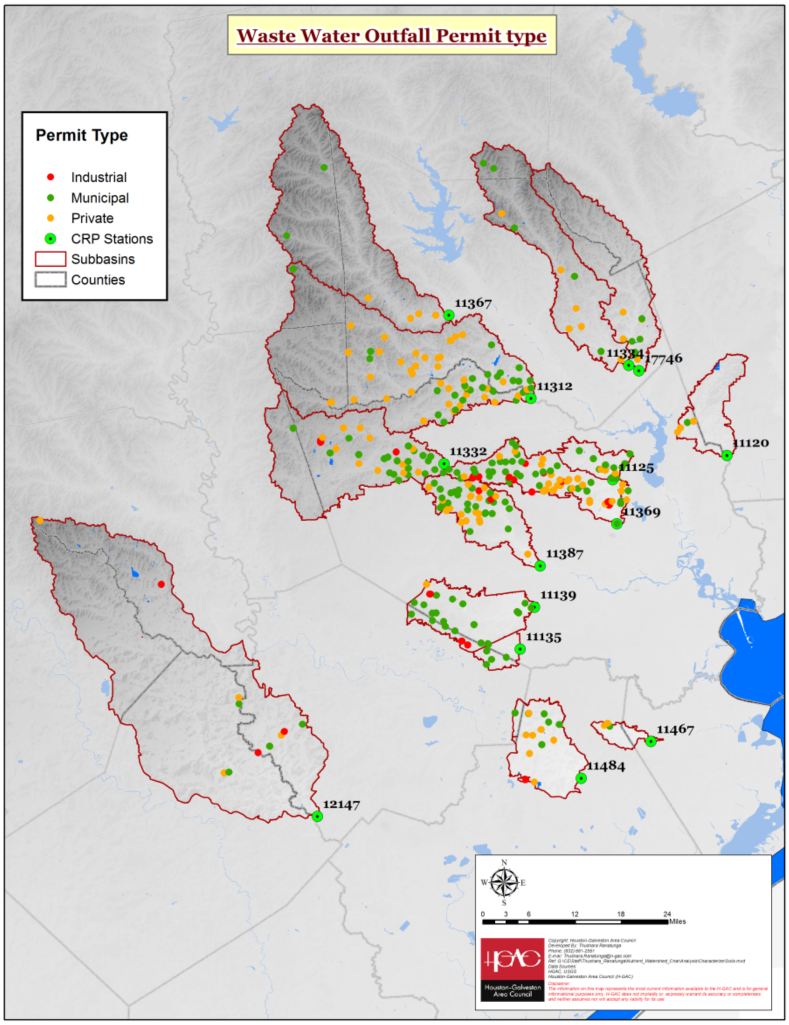

Figure 4: Waste Water Outfall Permit Type

Soil Survey Geographic (SSURGO) 2012 tabular and spatial soil data were acquired from the USGS National Resource Conservation Survey (NRCS) Web Soil Service (WSS)[2] for the H-GAC 15-county region. H-GAC performed an analysis of the soil data to define hydrologic characteristics of each watershed. The steps in the analysis used by H-GAC are provided below:

1. Spatial soil data was merged into one shapefile for the 15-county region and simplified with the map unit key (mukey) and map unit symbol (musym).
2. Using SAS, the soil taxonomic classifications and hydrologic groups were extracted from the downloaded tabular data in relation to the map unit keys.
3. Taxonomic and hydrologic soil group information were joined with spatial data (map unit key was used as the unique identifier).

[2] The WSS provides online soil data and data produced by the National Cooperative Soil Survey at http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm.

4. The output was overlaid with the selected watersheds, and soil polygons were simplified for each watershed.
5. Percentage area was estimated to identify the major soil types of each watershed.
6. Soil polygons with areas greater than 10% were selected and considered as the major soil types.
7. The classification descriptions were defined based on the soil taxonomy, drainage capacity, and hydrologic soil groups.

3.3.4 Drainage Density

The drainage density of a basin is the total line length of all the streams in a watershed divided by the watershed area. The drainage density is largely dependent on slope. A high density may indicate one or more of the following:

- A "mature," well developed channel system,
- Surface runoff that moves rapidly from hill slopes (overland flow) to channels,
- Thin/deforested vegetation cover, or
- Basin rocks/soils/surface with a generally low infiltration rate (highly impervious geology or abundant impervious manmade surfaces).

In this analysis, drainage density was estimated in units of miles of stream lengths per square mile. The drainage density for all of the 14 watersheds was estimated using the following process:

1. The CRP streams (both major and minor) were clipped into the 14 watersheds and linear units were set to US feet.
2. Drainage density was calculated using the ArcGIS line density tools (within the Spatial Analysis extension) per square mile. The spatial resolution was selected as 10 miles in the output raster.
3. Rasters were split into each watershed using split tools in ArcGIS.
4. Drainage density maps were generated and the average, minimum, and maximum drainage density were also estimated for each watershed. The output file displays the miles of streams per square mile area.

3.3.5 Road Density

Road data for the 15-county area were downloaded from the Texas Strategic Mapping Program (StratMap) of the Texas Natural Resources Information System (TNRIS), a division of the Texas Water Development Board (TWDB). The version of the dataset is "Phase 2 data" for the entire state of Texas published in 2012. The feature layer compasses all public roadways in the state of Texas including city streets, county roads, state and federal highways, and interstates. H-GAC used ArcGIS to clip the data for the watersheds of interest. These were then overlaid on the project watersheds simplified with watershed names. H-GAC estimated the total length of roadways for each watershed from the new shapefile created.

3.3.6 Imperviousness

Impervious areas increase the amount of runoff in watersheds and often have a negative impact on water quality in receiving streams. Imperviousness is defined as the sum of roads, parking lots, sidewalks, rooftops, and other impermeable surfaces of the urban landscape. This variable is used in measuring all scales of developments, as the percentage of area that is not "green." Impervious surfaces collect and accumulate pollutants deposited from the atmosphere, leaked from vehicles, or derived from other sources. During storms, these accumulated pollutants were quickly washed off to nearby water bodies. Most of monitoring and modeling studies have suggested that pollutant loads are directly related to watershed imperviousness (Schueler, 1994).

H-GAC used ArcGIS to estimate the amount of impervious area in each watershed. The information was obtained from the NLCD 2006 Imperviousness layer that includes percent imperviousness. Imperviousness raster data was converted into vector types based on the impervious percentages. The information was overlaid with the watershed boundaries and added to the project shapefile. H-GAC estimated the total acreage of impervious area for each watershed from the updated shapefiles.

3.3.7 Basin Relief

H-GAC also estimated the relief for the watersheds draining to monitoring stations. Basin relief is an indicator of the potential energy of the water being drained from the system. It is also highly correlated to drainage area and is an indicator of the overall watershed gradient. High relief may also indicate the presence of high elevation summits, thus high precipitation inputs and large recharge and discharge areas within a watershed. These parameters may not have a direct influence on hydrologic response, but can be useful for estimating other geomorphologic parameters.

Basin or watershed relief is measured as the difference between the maximum and minimum watershed elevations. These values were based on the USGS 10-meter DEM files. The maximum elevation of a basin is the highest watershed elevation. The minimum elevation is the lowest elevation point of a watershed.

The average elevation is the arithmetic mean of all the digital elevation model (DEM) data points within a watershed. The average elevation has important hydrologic and climatic influence because elevation influences soil, geology, vegetation, and microclimate of a watershed that, in turn, influence the hydrologic response. Average elevation is a reasonable measure of the overall watershed elevation, but it can be indirectly influenced by the presence of very low or very high elevation points. The elevation standard deviation is a measure of the variability in watershed elevation.

H-GAC used ArcGIS to produce individual raster DEM files for each watershed (using the "split" tool) and to estimate watershed relief statistics based on elevation records. H-GAC estimated the maximum, minimum, and average watershed relief for each watershed.

3.4 Statistical Analysis

The results of the GIS analyses described above were used to perform statistical analyses to evaluate relationships between nutrient and geospatial data. Numerous statistical tests were performed to evaluate the data. H-GAC used SAS (v9.3 / SAS/STAT 12.1) software to perform the statistical analysis.

Several tests were run (Shapiro-Wilk, Komolgorov-Smirnov, Anderson-Darling, and Cramer-von Mises) to examine the distribution of the data analyzed. The tests indicate that none of the variables with the exception of the natural log of total nitrogen in non-effluent dominated watersheds were normally distributed. A SAS macro and PROC TRANSREG were run to examine multiple potential transformations; however, none of the tests identified satisfactory transformations. For most variables, an approximately bimodal distribution of values was observed, reflecting the differences in concentrations between effluent-dominated (with primarily urbanized land use types) and non-effluent dominated watersheds.

Many of the statistical tests that were run are non-parametric (i.e. Kendall correlation, Kruskall-Wallis ANOVA) and robust to departures from normality (ANOVA by GLM), or require conditional rather than absolute normality (evaluated by the normality of the residuals from a complete regression or GLM model). Canonical correlation and discriminant analysis are very sensitive to departures from normality and the validity of the analysis requires variables that exhibit both univariate and multivariate normality. The normality of residuals from multiple regression and GLM (ANCOVA) models is discussed in the sections below. Formal tests of normality are very sensitive to slight departures from normality and may identify distributions as non-normal where graphical analysis (e.g. histograms) appears to show normal conditions, particularly for larger samples.

For parametric tests, data transformations were employed as needed. In cases where parametric tests were not appropriate due to the nature of the data, semi-parametric or non-parametric tests were applied. No data were disqualified on a statistical basis alone. Outliers were not removed since all data that might be considered to be outliers were confirmed as correct prior to inclusion in SWQMIS.

Due to the nature of this study and the limitations of available data, it was not feasible to obtain random samples from the existing dataset. Therefore, it was not possible to calculate and select a sample size required to obtain a specific statistical power. Post-hoc statistical power was calculated where appropriate. Details are found in the discussion of individual analyses.

In general, the null hypothesis used in most of the statistical tests is that there is no relationship between nutrient concentrations and the independent variables (these can be measures of land cover, effluent dominance, watershed type, wastewater discharge level, etc.). The specific null hypotheses are described in the sections describing the individual tests in the appropriate sections below. Details concerning candidate independent/ predictor variables are also found in the discussions of the specific analyses.

H-GAC applied a significance level of 0.05 in all statistical tests to control the Type I error rate. Data was also evaluated to ensure that the assumptions of specific statistical tests were met. The null hypothesis was rejected if the p-value of the applicable test was less than 0.05. It should be noted that statistical significance does not indicate practical importance. Where applicable, the results of statistical analysis are summarized in tables. The p-value in these tables is equivalent to the achieved Type I error rate. The p-value is the probability that the observed results would be obtained if the null hypothesis was "true." In other words, if the p-value is less than 0.05, the probability that the null hypothesis has been falsely rejected is less than five percent.

H-GAC evaluated the 24-hour dissolved oxygen data associated with the stations selected for this project and concluded that there was insufficient available data to allow valid statistical inferences regarding the relationship between dissolved oxygen, nutrients in regional waterways, and land cover influences.

3.4.1 Trend Analysis

H-GAC performed trend analysis to evaluate temporal variation in nutrient concentrations over the period considered. Trends were estimated for nutrients at each monitoring station. Trend analyses were performed on individual nutrient sample concentrations and also for annual median concentrations over time. H-GAC considered several types of trends using several statistical methods. The trend analyses performed were:

- Temporal trends of nutrient concentrations for individual samples over time (Kendall correlation, and robust regression)
- Flow adjusted trends in individual nutrient sample concentrations over time (described below)
- Nutrient concentration trends for individual samples over time, controlling for flow (robust regression)
- Seasonally adjusted trends in nutrient concentrations (Kendall / Sen Slope estimation)
- Annual median nutrient concentration by year (ordinary least squares method, robust regression, and Kendall correlation)
- Survival analysis (described below)

To estimate the flow adjusted trends in individual nutrient sample concentrations over time (item #2 in the list above), H-GAC implemented the following procedure. First, H-GAC transformed the data (the natural logarithms of nutrient concentration and the base-10 logarithm of flow). A LOESS regression (in SAS, PROC LOESS was used) fitting concentration to flow was fit to the data to obtain residuals. A Kendall correlation analysis between residuals and time was performed to assess the trend of the flow adjusted nutrient concentrations over time.

Some of the datasets used had numerous values that were entered in SWQMIS as less than the limit of quantitation (also referred to as censored data). If more than fifteen percent of the data for a nutrient at a monitoring station were censored, survival analysis (SAS PROC LIFEREG) was used to estimate the trends.

3.4.2 Regression Models

Regression analysis is a statistical technique used to determine the direction and magnitude of the relationship between one continuous dependent variable and one or more continuous independent variables (predictors or regressors). Regression analysis is frequently used for exploratory data analysis and the development of predictive models. In the case of simple linear regression, two quantities (a slope and intercept) that relate the dependent variable and a single independent variable are estimated using the method of least squares. A line derived from these parameters and the value of the independent variable, if plotted on a graph of the dependent vs. independent variable, would be a straight line that is closest to all of the data points.

The model can be written as:

$$Y = \beta_0 + \beta_1 X + \mathcal{E}$$

where

Y = the dependent (response) variable

X = the independent (predictor) variable

β_0 and β_1 are unknown parameters

$\mathcal{E}$ = the error term

Multiple regression analysis is applied to explore or predict the influence of two or more independent variables on a single dependent variable. The method of least squares cannot be used for this case, and matrix algebra is used to estimate parameters (sometimes called partial regression coefficients) that relate each independent variable to the dependent variable. The multiple regression model can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_k X_k + \mathcal{E}$$

where

Y = the dependent (response) variable

$X_1, X_2 \ldots X_k$ = the independent (predictor) variables

β_0 ,β_1, $\beta_2 \ldots \beta_k$ are unknown parameters

$\mathcal{E}$ = the error term

K= the number of independent variables

The predicted value (Y) is technically a prediction of the mean of a distribution of Y for given values of the independent variables, and is a normally distributed random variable. The values of the beta terms are estimated from the data. The variance in the dependent variable that is not captured by the beta coefficients is retained by the error term. Regression models may include polynomial terms (squares, cubes, and so forth) if their inclusion improves the model fit and reduces the error. The coefficient of determination (R^2) represents the proportion of the variation in the dependent variable attributed to the independent variable, and is the primary indicator of how closely the model fits the data. The coefficient of determination is the square of the Pearson correlation coefficient (R) in the case of regression of one variable on another.

As the number of independent variables in the model increases, R^2 increases even if the parameters are not statistically significant predictors of the dependent variable. An alternative measure of model fit, adjusted R^2, accounts for this problem and is a better estimate of model fit where there are several predictors.

The validity of a model developed by regression analysis is highest when several assumptions about the data are met (Ott and Longnecker 2010):

- The mean value of the dependent variable is linearly related to the value(s) of the independent variable(s).
- The observations are randomly sampled from the population of interest.
- The conditional values of the dependent variables are normally distributed. This was evaluated through formal tests of normality and examination of histograms of the residuals. Tests of normality are sensitive to small departures from normality as the sample size increases (beyond 30 for some tests), and may suggest non-normal residuals for models that fit the data very closely. In these cases, residual histograms may be more informative.
- The residuals have equal variance (are homoscedastic) at all levels of the dependent variable. This can be evaluated through examination of several residual plots.
- The observations are independent of one another (i.e., knowledge of the value of one does not provide information about the value of another). Independence was evaluated by calculating the Durbin-Watson statistic, which provides an estimate of serial correlation (autocorrelation) in the dataset.

The null hypothesis (falsified or not by F-tests, t-tests, and others) is that the values of the beta parameters are all zero. The values of the beta parameters can be considered slopes of a regression line, so the null hypothesis can also be expressed as “all the slopes are zero;” if this is not rejected, the best model (baseline model) is one that predicts that the value of the dependent variable will be the mean of the dependent variable at all values of the independent variables. If the F-test suggests the null hypothesis should be rejected, one can conclude that the beta/slope of at least one parameter is not zero. The Student’s t-test is used to determine the probability that the partial regression coefficient (slope) of each independent variable is zero. If the p-value of the t-test for the slope of a parameter is below an assigned threshold, the parameter is retained in the model. It is important to note that the results of the t-test are only valid if all other variables are in the model, but remain “fixed” in value. Removal of one variable will affect the standard error of the estimate of all other predictors and alter the outcome of the t-test. Variables that are found to be non-significant are removed one at a time, generally beginning with the variable that has the highest p-value. Frequently, automatic model selection methods will include variables that are collinear (correlated with one another). The partial regression coefficients (slope) calculated for collinear variables may not be physically interpretable. For example, one of them may have a different sign than the other, but collinearity does not affect the predictive power of the model. Important predictors may be left out of a model because collinearity increases the standard error of the partial regression coefficients, making statistical significance harder to detect.

Regression analysis was performed using SAS PROC REG and PROC ROBUSTREG. H-GAC staff developed regression models through comparison of the results of two “automatic” variable selection methods, followed by iterative elimination of variables after examination of collinearity diagnostics and other measures. The process may be summarized as follows:

- Stepwise regression of the full model was performed, which evaluates all combination of model parameters and retains those with a specified p-value.
- The full model was run with a variable selection option that ranks a selected number of models by the value of Mallows C(p), adjusted R^2, Schwartz’ Bayesian Criterion (SBC), and Akaike’s Information Criterion (AIC).
- One or more candidate models was selected, and the regression was run with options that produce collinearity statistics, tests of model specification error, formal tests of the normality of residuals, residual histograms, and autocorrelation diagnostics. The output was reviewed and variables were removed iteratively (a variable was removed, the regression was re-run and diagnostics evaluated, and so forth).

Regression models were developed to identify the temporal and spatial relationships between land cover and/or inferred land use data and nutrient concentrations in selected streams. Fourteen watersheds were identified that have sufficient data to conduct comparisons against the land cover and/or inferred land use data. Nutrient trends were assessed using a variety of parametric and non-parametric methods, as discussed on the pages that follow. Total phosphorus and total nitrogen were chosen as the dependent variables. The following were candidate predictors/ independent variables in the initial (full) model:

- Flow – Base-10 logarithm, CFS
- Watershed Flow – the sum of permitted discharges in the watershed, MGD
- Rainfall on the previous day (inches, untransformed and natural log) , inches
- Temperature (degrees F)
- Effluent dominance (indicator variable: “1” = dominated, “0” = non-dominated)
- Raw and log-transformed values of the following
  - Developed area (percent)
  - Agricultural area (percent)
  - Forest (percent)
  - Grasslands (percent)
  - Wetlands (percent)

H-GAC hoped to more fully utilize the discharge monitoring data obtained from TCEQ. However, H-GAC did not have discharge data predating 2002, and TCEQ staff reported that flow data entered prior to the last few years was unreliable. Because this data could not be confidently applied to more than two land-cover datasets, it was not used.

3.4.3 Correlation Analysis

Correlation analysis quantifies the magnitude and direction of the association between two continuous variables. Parametric (for example, the Pearson product moment correlation) and nonparametric (such as Spearman and Kendall correlation) methods are available. As always, parametric tests are more powerful but are not reliable if the data are not normally distributed. Because the data used in this project are not normally distributed, nonparametric tests were selected. These tests rank all observations, and compare the ranks of each member of a pair. If all ranks are the same, the correlation would be perfect (-1.0 or 1.0). Spearman's Rho method calculates the Pearson correlation coefficient from the ranks, while the Kendall Tau B correlation is calculated from the numbers of concordant (same rank) and discordant (different rank) pairs:

$$\tau = \frac{[\text{(number of concordant pairs)} - \text{(number of discordant pairs)}]}{0.5\, n\, (n-1)}$$

Where n = number of pairs

H-GAC selected Kendall Tau B for evaluation of the correlation of nutrient data to land cover. Analysis was conducted using SAS PROC CORR. Several correlation analyses were performed:

- Nutrient species with the land cover types obtained through GIS analysis by watershed and monitoring station,
- Nutrient species and a smaller set of land cover categories obtained by combining land cover types into groups with common characteristics,
- Nutrient species with land cover categories for all watersheds and stations,
- Nutrient species with land cover categories defined by the dominant land use, and
- Nutrient species with land cover categories separated by the presence of significant effluent from WWTP(s).

The nutrients analyzed at each site were:

- Total Phosphorus (tphos)
- Nitrate + Nitrite (nit)
- Total Nitrogen (TN)
- Ammonia-N (amm)

The value of the correlation coefficient ranges from -1.0 to 1.0; the strength of the relationship is indicated by the absolute value of the coefficient, while the sign indicates the direction of correlation (direct or inverse). If the two variables are independent, the coefficient will be near zero.

3.4.4 Analysis of Variance

Analysis of variance (ANOVA) techniques compare the mean (or median) levels of the dependent variables across categories or between groups. In general, the null hypothesis for all ANOVA technique is that there is no difference in the mean or median of the dependent variable in different groups. Both parametric and nonparametric versions of ANOVA have been developed. The general form of the ANOVA model is:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$$

where

Y_{ij} = the jth sample measurement in population i

μ = the mean of all populations

τ_i = the effect of population i

ε_{ij} = error term (sample measurement deviation from the population mean)

As noted previously, parametric procedures have greater power (that is, the ability to correctly reject the null hypothesis) than nonparametric procedures (Helsel and Hirsch, 2002). Parametric ANOVA was performed using the GLM (general linear model) procedure in SAS. Least square means (LSMEANS) rather than simple arithmetic means were used for comparison of differences. Least square means are superior to simple means in a complex model because they are adjusted for the influence of other variables in the model and can be applied to unbalanced data. If certain assumptions are met, ANOVA using the general linear model is more powerful than nonparametric ANOVA (see Ott and Longnecker, 2010, section 8.4 for the assumptions of ANOVA). Multi-way analyses are possible, the procedure is robust to moderate violations of normality, and the analysis is weighted to account for unbalanced data. In addition, the homogeneity of variance (HOV) test can be performed; if variance differs between groups, Welch's statistic can be calculated to identify statistically significant differences between the groups. HOV (Brown-Forsythe) and Welch tests cannot be performed if a least square means adjustment is made, so two analyses (one with the LSMEANS option, which adjusts estimates for other parameters in the model, and one with the MEANS option using raw means), were performed for total nitrogen and total phosphorus concentration.

As described above, ANOVA techniques are used to compare the mean or median of a dependent variable at different levels of a categorical variable or groups. Nonparametric procedures incorporate tests that are robust to departures from normality and are insensitive to extreme values (outliers). Nonparametric ANOVA (in SAS, PROC NPAR1WAY) produces Wilcoxon scores for the dependent variable across groups, among other measures. The Wilcoxon statistic is derived from a rank sum test, and the null hypothesis is that the sum of the ranks is the same for all groups. This method was used by H-GAC to evaluate the difference between medians values of the concentrations of total nitrogen and total phosphorus and the values of categorical variables (watershed type, effluent dominance, and level of effluent discharge). Only one-way analysis is possible, so each of the categorical variables was modeled separately. Validity of the results depends on compliance with several assumptions, the most important of which is homogeneity of variance (HOV) within each group. This method is not sensitive to moderate departures from homogeneity if the sample size is large.

3.4.5 Analysis of Covariance (General Linear Model)

Analysis of covariance (ANCOVA) using the general linear model may include continuous (numeric) and categorical predictors in the same model. Modeling categorical predictors in standard regression analysis requires calculation of indicator (dummy) variables with one of two numeric values, while ANOVA is restricted to categorical predictors. The analysis of covariance is a technique that combines regression and ANOVA in a single model. H-GAC conducted the ANCOVA analysis using SAS PROC GLM. The method allows comparison of group means or medians while taking the values of continuous variables (covariates) into account. In randomized controlled experimental designs, a categorical "treatment" variable is the predictor of the dependent variable, and covariates are added to assess the possibility that the effect of the treatment varies with the level of the covariate. The current study relies on observational data only, as it is not possible to randomly assign subjects to treatment, and the analogue for treatment in this case is "exposure," the preexisting membership of subjects (nutrient concentrations at specific stations) in groups (effluent-dominated or not, etc.). For a discussion of application of ANCOVA to observational data, see Riggs 2008.

Variables representing the interaction between categorical grouping variables and continuous covariates can (and, for the initial model development, must) be included in the model. The general form of an ANCOVA model for one group and one covariate is:

$$Y_{ij} = \mu + \alpha_j + \beta_w (X_{ij} - \mu_X) + e_{ij}$$

Where

Y_{ij} = value of the dependent variable, ith observation in the jth group

X_{ij}= value of the covariate, ith observation in the jth group

μ = the mean of the dependent variable across all groups

α_j = jth group effect (μj - μ) with the covariate X held constant

β_w = slope of the covariate with group membership constant

e_{ij} = residual or unexplained variance for observation i in group j

ANCOVA was used in this study to evaluate the relationship between individual nutrient species and predictors that include both categorical and continuous variables and interaction terms between categorical variables and continuous covariates. Similar results would be obtained from multiple regression with indicator (dummy) variables to represent categorical variables. H-GAC used ANCOVA (in SAS, PROC GLM was used) to provide additional output similar to that produced by standard ANOVA. As in multiple regression, collinearity between predictor variables can produce counterintuitive parameter estimates (for example, differing signs for partial regression coefficients of the untransformed variable and the natural log of a variable).

Interaction effects exist where the levels of one variable affect the relationship of two other variables. For example, a drug intended to treat high blood pressure might be evaluated by comparing dosage and total reduction in BP. The gender of the subject might affect drug response, so one would include gender as well as an interaction term (variable) in the model. This term is the multiple of, in this example, gender and total reduction in BP. If the term is statistically significant, there is evidence that men and women respond differently and the slope of the regression line differs by gender; typically in a regression plot with one line for each gender, the lines would cross if the interaction is significant. If only the intercepts are different, gender alone can be used as a variable. Interactions can be modeled in all models derived from generalized linear models, but in this project were only evaluated in ANCOVA models.

H-GAC staff developed models of the relationship between total phosphorus and total nitrogen, with land cover, rainfall, flow, effluent dominance, wastewater discharge level, watershed type, and interactions between rainfall and land cover types. No significant interactions were suggested by the data. Wastewater-related and watershed-type variables were significantly associated with the concentration of nutrients in the models that included data from all stations but were not important when effluent-dominated streams were excluded.

All variables were entered in an initial model and were then removed iteratively to produce a final model that included only statistically significant predictors. The sample size was taken into account during this process. In general, no more than one predictor for ten observations was retained. The following candidate predictors/ independent variables were selected for the initial (full) model:

- Watershed type (categorical, two levels)
- Wastewater discharge (categorical, three levels)
- Effluent dominance (categorical, two levels)
- Flow – Base 10 logarithm, CFS
- Watershed Flow – the sum of permitted discharges in the watershed, MGD
- Rainfall on the previous day, inches (untransformed and natural log)
- Temperature, degrees F
- Raw and log-transformed values of the following
    - Developed area (percent)
    - Agricultural area (percent)
    - Forest (percent)
    - Grasslands (percent)
    - Wetlands (percent)
- Interaction between rainfall and all land cover variables

In addition, models using the land cover acreage rather than the percent of the total area of the watershed were evaluated. None of these models explained more than 25 percent of the variance in nutrient concentrations.

3.4.6 Canonical Correlation

Canonical correlation was used to identify and measure the associations between sets of variables. Derived variables are created by weighting the values of individual variables to create a linear combination that maximizes the correlation within each set of variables. The general form of the model is

$$w_1 = a_1x_1 + a_2x_2 + \ldots \quad a_kx_k$$

$$v_1 = b_1y_1 + b_2y_2 + \ldots \quad b_py_p$$

r_{w1v1} = first canonical correlation

Where

w_1 = Derived variate for first group

v_1 = Derived variate for second group

k = variables in first derived variate

p = variables in second derived variate

Typically, several canonical variables are derived from each set. The variables selected for each set can refer to different (real or presumed) dimensions. Correlation analysis is performed between each derived variable. The method produces estimates of canonical correlation coefficients that are analogous to multiple correlation coefficients, and the square of the canonical correlation can be interpreted as a measure of the variance in the dependent set of variables that is explained by the other set of canonical variables. Canonical correlation can be viewed as a generalization of multivariate regression, multivariate analysis of variance, and multivariate analysis of covariance. With additional coding, canonical correlation can produce the same results as bivariate correlation and simple linear regression. In this sense it is one of the "most general" models.

For example, a researcher might be interested in analyzing survey data to explore the relationship between personality and intellectual aptitude. Clusters of questions related to each of these general concepts would be identified, and several derived variables (canonical variates) that maximize the correlation between the members of each cluster would be calculated. The correlation between each derived variable would then be evaluated. One might also correlate two different survey instruments intended to measure the same general feature (e.g., comparing two psychological tests for gambling addiction).

Canonical correlation was applied in this study to evaluate the relationship between nutrients and land cover in the aggregate, rather than between individual nutrient species and individual land cover types. H-GAC evaluated correlations between sets of derived variables (variates) that represent linear combinations of nutrient concentration and land cover variables using SAS PROC CANCORR. Multivariate normality in at least one of the pairs of canonical variates is essential if valid inferences of statistical significance are to be made. In addition, the stability of parameter estimates depends on the sample size. Most commonly, 20-40 observations per variable is recommended (various authors, cited in Nash and Chaloud, 2002). In order to maximize the sample size, nitrate-N was used in place total nitrogen in the calculation of nutrient canonical variates. Multivariate normality was achieved by using the square root of the natural logarithms of total phosphorus and nitrate-N. These transformed variables were used to calculate the first canonical variate, representing nutrients. The natural logarithm of five land cover category percentages (developed, agriculture, forest, wetlands, and grassland) was used to create the second derived variable (land cover variates). Two canonical variates were calculated for each group. Two analyses were performed: all stations combined, and separately by effluent domination status.

3.4.7 Discriminant Analysis

Discriminant analysis is an application of canonical correlation analysis to the cases where one is interested in the correlation between group (categorical) variables and continuous variables. It can also be viewed as the reverse of multivariate analysis of variance (MANOVA); discriminant analysis predicts the value of categorical variables from continuous variables, while MANOVA predicts the means of continuous variables from the value of categorical variables. Discriminant analysis can also be viewed as a means of establishing boundaries between groups. Discriminant analysis quantifies how well observations fit into pre-existing groups. This function is a linear combination of the values of each variable that places each observation in the group to which it is "closest," and is a means of maximizing the distance between the groups on the basis of observations of specific variables. Discriminant functions can be calculated for individual variables or for derived variables that are weighted combinations of variables. H-GAC applied the SAS procedure PROC DISCRIM to calculate the Fisher linear discriminant function for a group of categorical or continuous variables. Each observation is scored on the basis of distance from a group centroid. The classification error rate can reveal problems with the assignment of the values of categorical variables to specific watersheds, assuming there is a relationship between land cover and/or inferred land use and nutrient concentration.

3.5 Load Duration Curves

Load Duration Curves (LDC) present the corresponding relationship between contaminant loadings and streamflow conditions at monitoring sites and show the percent of time that flow rates are exceeded.

The use of duration curves provides a technical framework for identifying daily loads in TMDL development and account for the variable nature of water quality associated with different streamflow rates. Specifically, a maximum daily concentration limit can be used with basic hydrology and a duration curve to identify a TMDL that covers the full range of flow conditions. With this approach, the maximum daily load can be identified for any given day based on the streamflow. Identification of a loading capacity using the duration curve framework is driven by the flow duration curve (FDC) and a water quality criterion or target value. The target may be constant across all flow conditions or the target may vary with flow.

Flow data was obtained from the CRP sampling database. The flow values that were collected along with water quality measurements were used. All of the monitoring stations used in this analysis had at least 20 data points for flow. The process used to develop the LDCs was:

1. The flow data were ranked from minimum to maximum flow values.
2. The percent exceedance was estimated using:

   $$P_i = \frac{i}{n+1}$$

   where *pi* is the exceedance probability or plotting position, and *i* is the rank number for a given number of observations 1, 2, 3,…, *n*[3].
3. The FDC graph was plotted on a log normal probability grid with the observed flow values (y-axis) and the percent exceedance of each flow count (x-axis).
4. Nutrient concentrations for each sample time that had flow data measurements were obtained using the SAS statistical tool.
5. The total loading for each event was estimated by multiplying the nutrient concentrations by the flow values.
6. Total daily loads were estimated using the following equation:

   $$Total\ Loading\ \left(\frac{mg}{Day}\right) = C(\frac{mg}{L}) * Q(\frac{cf}{s}) * 28.3168\left(\frac{L}{cf}\right) * 3600\left(\frac{s}{hr}\right) * 24(\frac{hrs}{day})$$

   Where, C is the measured concentration in mg/L, and Q is the measured flow rate in CFS.
7. Values were plotted against the percent exceedance in a semi-log normal plot with the FDC.
8. A TMDL (screening level) graph was developed based on the numeric targets provided by the TCEQ:
   - Total Phosphorus: 0.69 mg/L
   - Total Nitrogen: 6.8 mg/L
   - Nitrate: 1.95 mg/L
9. Screening levels were multiplied by the measured flow rated using the appropriate unit conversions to generate total maximum daily loads (TMDLs).
10. Flow regimes were identified and marked in each plot based on USEPA LDC guidelines and as shown in Table 4.

[3] For additional information, refer to the *Development of Duration-Curve Based Methods for Quantifying Variability and Change in Watershed Hydrology and Water Quality* (USEPA, 2008)

Table 4: Flow Duration Intervals used in the Flow Regime

Flow Duration Interval	Hydrologic Condition Class*
0-10%	High flows
10-40%	Moist Conditions
40-60%	Mid-Range Conditions
60-90%	Dry Conditions
90-100%	Low Flows

Source: Cleland 2003

An interpretation of the flow regimes used in the FDC is shown in Table 5.

Table 5: Interpretation of the Flow Regimes Designated in the FDC

Contributing Source Area	Duration Curve Zone				
	High Flow	Moist	Mid-Range	Dry	Low Flow
Point Source				M	H
On-site wastewater systems			H	M	
Riparian Areas		H	H	H	
Storm water: Impervious Areas		H	H	H	
Combined sewer overflows	H	H	H		
Storm water: Upland	H	H	M		
Bank erosion	H	M			

Note: Potential relative importance of source area to contribute loads under given hydrologic condition (H: High; M: Medium)
Source: USEPA 2007

LDCs were calculated and plotted for each selected station for the following constituents of concern:

- Nitrate
- Total nitrogen (TN)
- Total Phosphorus (TP)

4 Results and Discussion

4.1 Watershed Characterization

Watershed characterization results provide the information on the physical and natural characteristics of watersheds that could support understanding the influence on nutrient loading of each watershed. For example, land cover changes can help to understand the potential growth of the area and provide the logical basis for identifying the potential sources of nutrient level changes of the watersheds. H-GAC used ArcGIS/Info 10.1 to evaluate hydrogeological and land use characteristics of the subwatersheds. Many of these characteristics are considered to have potential impacts on nutrient levels. H-GAC also used some of the results to compare versus nutrient data. The results of the GIS analyses are described in the sections below.

4.1.1 Land Cover Change Analysis

The output results of land cover change analysis based on NOAA Coastal Change Analysis (C-CAP) datasets of year 1996 and 2011 are presented in this section. The estimated change in the percent of each land cover type is presented in Table 6. The table also includes the area of each land cover type (in acres) for each year. The percent area of land cover types for each watershed and each year is also included. The table column “change in %” represents the change of land cover on a percentage basis for each watershed. The trends are also noted to be increasing, decreasing or stable for each land cover. According to the criteria given in the section 3.3.1, these trend assignments were allocated. This information was exported into SAS and used in the statistical analysis. The majority of the watersheds are show an increasing trend in developed land cover types (for ex; watershed ID 11125, 11135, 11139, 11312, 11332, 11369, 11387). Watersheds such as Caney creek (11334) and Peach Creek (17746) show an increasing trend of natural and unmanaged grasslands. Cedar Bayou watershed (11120), Dickinson Bayou (11467), Chocolate Bayou (11484) and San Bernard River (12167) show no trends in any of the land cover types, which reflects that these watersheds have not shown significant changes over the years from 1996 to 2011 and are fairly stable. Almost all of the watersheds have a decreasing trend of forest land cover percentages.

Table 6: Results of Land Cover Change Analysis

Watershed	Station ID	Land Cover	Area 1996 (Acres)	1996 %	Area 2011 (Acres)	2011 %	Change in %
Cedar Bayou	11120	Bare Land	59.29	0.14	40.06	0.10	-0.05
		Cultivated	29750.48	72.36	29769.30	72.40	0.05
		Developed	344.46	0.84	396.87	0.97	0.13
		Developed Open Space	311.71	0.76	333.24	0.81	0.05
		Forest	2677.33	6.51	2487.26	6.05	-0.46
		Grasslands	1343.39	3.27	1378.01	3.35	0.08
		Scrub/Shrub	1280.34	3.11	1416.80	3.45	0.33
		Water	909.86	2.21	708.06	1.72	-0.49
		Wetlands	4439.28	10.80	4586.54	11.16	0.36
Garners Bayou	11125	Bare Land	35.16	0.18	269.08	1.41	1.23
		Cultivated	794.45	4.16	486.52	2.55	-1.61
		Developed	4778.91	25.04	7594.51	39.79	14.75
		Developed Open Space	2461.55	12.90	3695.84	19.36	6.47
		Forest	8450.94	44.27	4892.56	25.63	-18.64
		Grasslands	620.56	3.25	407.39	2.13	-1.12
		Scrub/Shrub	161.49	0.85	357.84	1.87	1.03
		Water	35.48	0.19	71.64	0.38	0.19
		Wetlands	1749.76	9.17	1312.92	6.88	-2.29
Sims Bayou	11135	Bare Land	88.02	0.74	33.03	0.28	-0.46
		Cultivated	1546.79	12.93	1188.37	9.94	-3.00
		Developed	6173.51	51.62	7124.06	59.57	7.95
		Developed Open Space	2088.37	17.46	2141.71	17.91	0.45
		Forest	1189.17	9.94	658.50	5.51	-4.44
		Grasslands	289.14	2.42	269.07	2.25	-0.17
		Scrub/Shrub	233.48	1.95	218.66	1.83	-0.12
		Water	68.97	0.58	112.87	0.94	0.37
		Wetlands	281.85	2.36	213.04	1.78	-0.58

Watershed	Station ID	Land Cover	Area 1996 (Acres)	1996 %	Area 2011 (Acres)	2011 %	Change in %
Brays Bayou	11139	Bare Land	144.45	0.24	40.76	0.07	-0.17
		Cultivated	2640.45	4.42	637.08	1.07	-3.35
		Developed	45327.36	75.86	49317.55	82.54	6.68
		Developed Open Space	7701.57	12.89	7591.49	12.71	-0.18
		Forest	2665.66	4.46	1107.33	1.85	-2.61
		Grasslands	563.18	0.94	338.41	0.57	-0.38
		Scrub/Shrub	260.30	0.44	233.57	0.39	-0.04
		Water	67.32	0.11	348.08	0.58	0.47
		Wetlands	378.98	0.63	135.01	0.23	-0.41
Spring Creek	11312	Bare Land	786.68	0.29	2067.50	0.77	0.47
		Cultivated	55046.52	20.41	53616.91	19.88	-0.53
		Developed	29489.98	10.93	43919.02	16.28	5.35
		Developed Open Space	4177.61	1.55	8219.20	3.05	1.50
		Forest	117407.24	43.53	88961.76	32.98	-10.55
		Grasslands	10298.02	3.82	16303.19	6.04	2.23
		Scrub/Shrub	24284.51	9.00	30012.84	11.13	2.12
		Water	1338.08	0.50	1997.86	0.74	0.24
		Wetlands	26882.82	9.97	24613.18	9.13	-0.84
Cypress Creek	11332	Bare Land	676.15	0.49	1424.45	1.03	0.54
		Cultivated	94327.89	68.47	88598.94	64.31	-4.16
		Developed	6593.06	4.79	14485.65	10.52	5.73
		Developed Open Space	3587.56	2.60	6132.89	4.45	1.85
		Forest	11545.64	8.38	7303.51	5.30	-3.08
		Grasslands	5582.00	4.05	5419.91	3.93	-0.12
		Scrub/Shrub	6362.51	4.62	5509.90	4.00	-0.62
		Water	691.04	0.50	1131.27	0.82	0.32
		Wetlands	8395.26	6.09	7754.58	5.63	-0.47
Caney Creek	11334	Bare Land	245.97	0.21	396.06	0.34	0.13
		Cultivated	16549.10	14.31	16421.86	14.20	-0.11
		Developed	6284.57	5.43	8215.81	7.10	1.67
		Developed Open Space	991.52	0.86	1437.46	1.24	0.39
		Forest	51399.44	44.44	41451.75	35.84	-8.60
		Grasslands	11042.20	9.55	17148.34	14.83	5.28
		Scrub/Shrub	12022.27	10.40	14680.46	12.69	2.30
		Water	607.32	0.53	661.20	0.57	0.05
		Wetlands	16510.18	14.28	15239.62	13.18	-1.10

Watershed	Station ID	Land Cover	Area 1996 (Acres)	1996 %	Area 2011 (Acres)	2011 %	Change in %
Lake Creek	11367	Bare Land	311.76	0.17	624.65	0.34	0.17
		Cultivated	80205.75	43.08	80134.17	43.04	-0.04
		Developed	4193.69	2.25	5405.96	2.90	0.65
		Developed Open Space	167.25	0.09	434.38	0.23	0.14
		Forest	56655.16	30.43	49000.59	26.32	-4.11
		Grasslands	7750.10	4.16	11034.16	5.93	1.76
		Scrub/Shrub	10088.01	5.42	12880.83	6.92	1.50
		Water	1117.88	0.60	1225.17	0.66	0.06
		Wetlands	25703.07	13.80	25452.78	13.67	-0.13
Greens Bayou	11369	Bare Land	172.32	0.27	414.81	0.66	0.39
		Cultivated	3373.10	5.37	1444.30	2.30	-3.07
		Developed	24035.22	38.25	31641.40	50.35	12.10
		Developed Open Space	7784.76	12.39	9184.33	14.62	2.23
		Forest	17368.17	27.64	10689.62	17.01	-10.63
		Grasslands	2275.92	3.62	2326.34	3.70	0.08
		Scrub/Shrub	964.52	1.53	1328.08	2.11	0.58
		Water	142.90	0.23	235.53	0.37	0.15
		Wetlands	6721.37	10.70	5573.88	8.87	-1.83
Whiteoak Bayou	11387	Bare Land	227.17	0.41	98.88	0.18	-0.23
		Cultivated	1658.65	3.00	167.14	0.30	-2.70
		Developed	36876.74	66.80	42473.63	76.94	10.14
		Developed Open Space	7811.20	14.15	7928.66	14.36	0.21
		Forest	7175.05	13.00	3510.23	6.36	-6.64
		Grasslands	667.11	1.21	385.62	0.70	-0.51
		Scrub/Shrub	120.15	0.22	233.29	0.42	0.20
		Water	108.48	0.20	135.62	0.25	0.05
		Wetlands	557.42	1.01	268.90	0.49	-0.52
Dickinson Bayou	11467	Bare Land	6.70	0.07	6.01	0.06	-0.01
		Cultivated	5165.64	50.08	5024.63	48.71	-1.37
		Developed	1073.66	10.41	1285.15	12.46	2.05
		Developed Open Space	1313.32	12.73	1373.96	13.32	0.59
		Forest	868.73	8.42	824.22	7.99	-0.43
		Grasslands	785.91	7.62	781.60	7.58	-0.04
		Scrub/Shrub	871.71	8.45	781.96	7.58	-0.87
		Water	18.08	0.18	57.73	0.56	0.38
		Wetlands	210.87	2.04	179.38	1.74	-0.31

Watershed	Station ID	Land Cover	Area 1996 (Acres)	1996 %	Area 2011 (Acres)	2011 %	Change in %
Chocolate Bayou	11484	Bare Land	31.53	0.06	39.14	0.07	0.01
		Cultivated	37647.55	66.21	36487.08	64.17	-2.04
		Developed	2200.04	3.87	3519.36	6.19	2.32
		Developed Open Space	2071.74	3.64	2371.56	4.17	0.53
		Forest	1787.87	3.14	1650.91	2.90	-0.24
		Grasslands	2849.85	5.01	2822.93	4.96	-0.05
		Scrub/Shrub	3630.42	6.38	3388.41	5.96	-0.43
		Water	261.76	0.46	453.56	0.80	0.34
		Wetlands	6382.90	11.22	6130.72	10.78	-0.44
San Bernard River	12147	Bare Land	940.36	0.21	966.54	0.21	0.01
		Cultivated	337804.19	75.10	339628.11	75.51	0.41
		Developed	3439.49	0.76	3846.33	0.86	0.09
		Developed Open Space	2410.85	0.54	2654.69	0.59	0.05
		Forest	31327.84	6.96	29698.72	6.60	-0.36
		Grasslands	15585.19	3.46	16808.70	3.74	0.27
		Scrub/Shrub	25655.59	5.70	23222.50	5.16	-0.54
		Water	1263.43	0.28	1485.50	0.33	0.05
		Wetlands	31378.91	6.98	31494.75	7.00	0.03
Peach Creek	17746	Bare Land	119.62	0.12	247.31	0.25	0.13
		Cultivated	2160.78	2.15	2260.09	2.25	0.10
		Developed	3095.33	3.08	3910.75	3.90	0.81
		Developed Open Space	656.16	0.65	863.74	0.86	0.21
		Forest	61891.79	61.66	52211.18	52.02	-9.64
		Grasslands	6530.66	6.51	15289.60	15.23	8.73
		Scrub/Shrub	13055.09	13.01	13085.11	13.04	0.03
		Water	200.02	0.20	246.38	0.25	0.05
		Wetlands	12664.82	12.62	12260.10	12.21	-0.40

4.1.1.1 Land Cover Change Detection

Besides the above land cover percentage change estimates from year 1996 to 2011, H-GAC ran a land cover conversion analysis. The results of the analysis provide information about the area coverage of each land cover class conversion from one class to another and the spatial distribution of each conversion type. Figure 5 shows the locations of land cover changes for the 14 watersheds in the study area. It represents the areas that have been changed to any type of land cover class. Watersheds such as 11125, 11369, 11312, 11367, 11334, 1776, and 11332 show considerable amounts of land cover changes over the 15 year study period.

The estimated changes in acres for each watershed are presented in Table 7 through Table 20. As shown in these tables, the majority of the watersheds have forest or cultivated lands that have been converted into developed lands. The watersheds with the largest amounts of development include:

- The Brays Bayou watershed (11139) had over 4700 acres that were converted to developed or developed open space;
- The Cypress Creek watershed (11332) had over 8800 acres converted to developed or developed open space;
- The Whiteoak Bayou (11387) watershed had almost 6300 acres converted to developed or developed open space;
- The Greens Bayou watershed (11369) had over 9400 acres converted to developed or developed open space;
- The Spring Creek watershed (11369) had over 18,000 acres converted to developed or developed open space.

The spatial pattern maps of each land cover change for the 14-watershed area is attached in Appendix B.

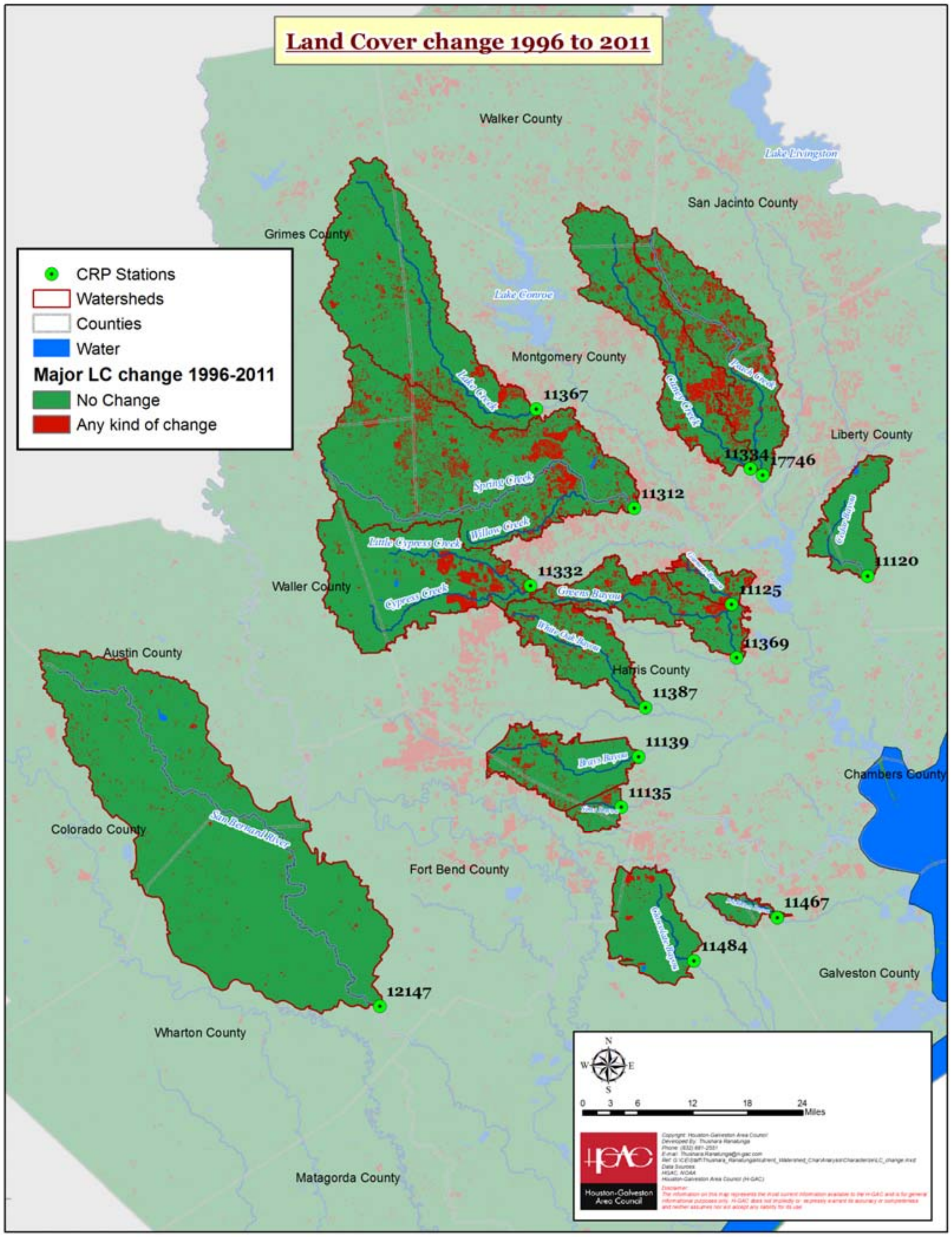

Figure 5: Land Cover Change 1996 to 2011

Table 7: Land Use / Land Cover Change by Area for Brays Bayou (11139)

FROM	TO	Area (Acres)
Bare Land	Developed	111.14
Cultivated	Developed Open Space	408.39
	Developed	1594.47
Developed Open Space	Developed	897.29
Forest	Developed	914.11
	Developed Open Space	394.95
Grassland	Developed	207.52
Wetlands	Developed	172.91
	Developed Open Space	62.66

Table 8: Land Use / Land Cover Change by Area for Cedar Bayou (11120)

FROM	TO	Area (Acres)
Cultivated	Scrub/Shrub	113.21
Forest	Grassland	88.78
	Cultivated	77.28
Open Water	Wetlands	137.78

Table 9: Land Use / Land Cover Change by Area Caney Creek (11334)

FROM	TO	Area (Acres)
Forest	Scrub/Shrub	6810.90
	Grassland	5347.93
	Developed	1140.81
	Developed Open Space	167.08
Grassland	Scrub/Shrub	924.18
Scrub/Shrub	Forest	3745.83
	Grassland	1095.52
	Developed	496.48
	Developed Open Space	155.75

FROM	TO	Area (Acres)
Wetlands	Bare Land	104.47
	Grassland	640.94
	Scrub/Shrub	374.16
	Developed	103.84

Table 10: Land Use / Land Cover Change by Area for Chocolate Bayou (11484)

FROM	TO	Area (Acres)
Cultivated	Developed Open Space	231.87
	Developed	969.25
	Open Water	178.13
Forest	Developed	62.42
	Cultivated	71.95
Scrub/Shrub	Forest	61.31
	Developed	64.04
	Cultivated	147.82
Wetlands	Developed	172.77

Table 11: Land Use / Land Cover Change by Area for Cypress Creek (11332)

FROM	TO	Area (Acres)
Cultivated	Bare Land	620.15
	Developed	4122.09
	Open Water	334.89
Developed Open Space	Developed	115.32
Forest	Scrub/Shrub	251.07
	Grassland	423.12
	Developed	2590.07
	Bare Land	117.45
	Developed Open Space	621.83
	Cultivated	205.87
Grassland	Scrub/Shrub	145.43
	Developed Open Space	210.54
	Developed	377.29

FROM	TO	Area (Acres)
Scrub/Shrub	Developed	284.93
	Cultivated	621.48
	Developed Open Space	158.49
Wetlands	Cultivated	124.92
	Developed	263.99
	Developed Open Space	122.81

Table 12: Land Use / Land Cover Change by Area for Dickinson Bayou (11467)

FROM	TO	Area (Acres)
Cultivated	Developed	110.29

Table 13: Land Use / Land Cover Change by Area for Garners Bayou (11125)

FROM	TO	Area (Acres)
Cultivated	Developed Open Space	98.06
	Developed	315.75
Forest	Scrub/Shrub	227.41
	Grassland	140.74
	Developed	1915.51
	Bare Land	114.84
	Developed Open Space	1038.07
Grassland	Developed	192.25
Wetlands	Developed	308.35
	Developed Open Space	121.83

Table 14: Land Use / Land Cover Change by Area for Peach Creek (17746)

FROM	TO	Area (Acres)
Forest	Scrub/Shrub	6047.72
	Grassland	7923.06
	Developed	615.28

FROM	TO	Area (Acres)
Grassland	Scrub/Shrub	486.01
Scrub/Shrub	Forest	5155.14
	Grassland	1204.71
	Developed	119.88
	Cultivated	99.57
Wetlands	Grassland	109.94

Table 15: Land Use / Land Cover Change by Area for Greens Bayou (11369)

FROM	TO	Area (Acres)
Cultivated	Developed Open Space	300.01
	Developed	1656.73
Developed Open Space	Developed	630.48
Forest	Scrub/Shrub	590.45
	Grassland	629.65
	Developed	3819.61
	Bare Land	182.27
	Developed Open Space	1220.08
	Cultivated	117.33
Grassland	Developed Open Space	172.29
	Developed	344.40
Scrub/Shrub	Developed	225.54
Wetlands	Developed	759.32
	Developed Open Space	341.87

Table 16: Land Use / Land Cover Change by Area for Lake Creek (11367)

FROM	TO	Area (Acres)
Cultivated	Scrub/Shrub	170.23
	Grassland	250.01
Forest	Scrub/Shrub	3587.79
	Grassland	4070.99
	Developed	984.34
	Bare Land	284.59

FROM	TO	Area (Acres)
	Developed Open Space	189.97
Grassland	Scrub/Shrub	1071.20
	Forest	349.92
Scrub/Shrub	Forest	1178.15
	Grassland	353.84
	Developed	122.57
	Cultivated	435.65

Table 17: Land Use / Land Cover Change by Area for Whiteoak Bayou (11387)

FROM	TO	Area (Acres)
Bare Land	Developed	182.99
Cultivated	Developed Open Space	221.19
	Developed	1259.30
Developed Open Space	Developed	865.16
Forest	Scrub/Shrub	184.78
	Developed	2587.00
	Developed Open Space	679.08
Grassland	Developed	269.52
Wetlands	Developed	221.13

Table 18: Land Use / Land Cover Change by Area for San Bernard River (12147)

FROM	TO	Area (Acres)
Cultivated	Scrub/Shrub	187.33
	Wetlands	168.95
	Developed Open Space	170.38
	Bare Land	132.96
	Developed	238.45
	Open Water	164.61
Forest	Scrub/Shrub	246.42
	Grassland	1188.85
	Cultivated	438.62
Grassland	Scrub/Shrub	132.59
Scrub/Shrub	Forest	259.12
	Grassland	303.68
	Cultivated	2193.69
Wetlands	Cultivated	120.67

Table 19: Land Use / Land Cover Change by Area for Sims Bayou (11135)

FROM	TO	Area (Acres)
Cultivated	Developed	270.69
Developed Open Space	Developed	143.67
Forest	Developed	370.38

Table 20: Land Use / Land Cover Change by Area for Spring Creek (11312)

FROM	TO	Area (Acres)
Bare Land	Scrub/Shrub	127.40
	Open Water	170.26
Cultivated	Forest	129.32
	Developed Open Space	595.90
	Bare Land	261.44
	Developed	954.49
Developed Open Space	Developed	193.69

FROM	TO	Area (Acres)
Forest	Scrub/Shrub	9573.14
	Grassland	6182.74
	Developed	10384.34
	Bare Land	954.28
	Wetlands	140.66
	Developed Open Space	2500.79
	Cultivated	387.54
	Open Water	242.94
Grassland	Scrub/Shrub	928.41
	Forest	189.41
	Developed Open Space	160.91
	Developed	207.99
Scrub/Shrub	Forest	2101.98
	Grassland	1142.91
	Developed	1207.70
	Cultivated	200.42
	Developed Open Space	385.89
Wetlands	Bare Land	292.82
	Grassland	103.87
	Open Water	220.02
	Developed	1266.29
	Developed Open Space	558.27

4.1.2 Influence of WWTP Effluent

H-GAC performed GIS analyses to identify the spatial distribution of wastewater outfalls in the study region. Table 21 presents the number and types of outfalls in each of the subwatersheds draining to monitoring stations in the project area. Figure 6 shows the density of outfalls (number of outfalls per square mile) in each of the project subwatersheds.

Many of the project watersheds have few outfalls. However, Brays Bayou and Cypress Creek each have over 25 outfalls and Spring Creek, Greens Bayou, and Whiteoak Bayou have over 50 outfalls. In addition, as shown in Figure 6, there are some areas southeast of project subwatersheds with higher densities of outfalls. Other than this characterization, H-GAC assigned three categories of influence from WWTPs: low, medium, and high. The categories were assigned on the basis of the sum of permitted flow at domestic wastewater treatment facilities, as recorded in an in-house database, for each watershed/station.

Table 21: Permitted Outfalls

Watershed Name	Watershed ID	Outfall Counts			
		Municipal	Industrial	Private	Uncategorized
Cedar Bayou	11120	1	0	3	0
Garners Bayou	11125	11	0	4	0
Sims Bayou	11135	4	0	0	0
Brays Bayou	11139	20	6	1	1
Spring Creek	11312	28	0	38	1
Cypress Creek	11332	21	3	11	0
Caney Creek	11334	5	0	6	0
Lake Creek	11367	2	0	4	1
Greens Bayou	11369	35	17	28	3
Whiteoak Bayou	11387	24	2	22	2
Dickinson Bayou	11467	1	0	2	0
Chocolate Bayou	11484	6	1	7	1
San Bernard River	12147	4	4	4	0
Peach Creek	17746	6	0	4	0

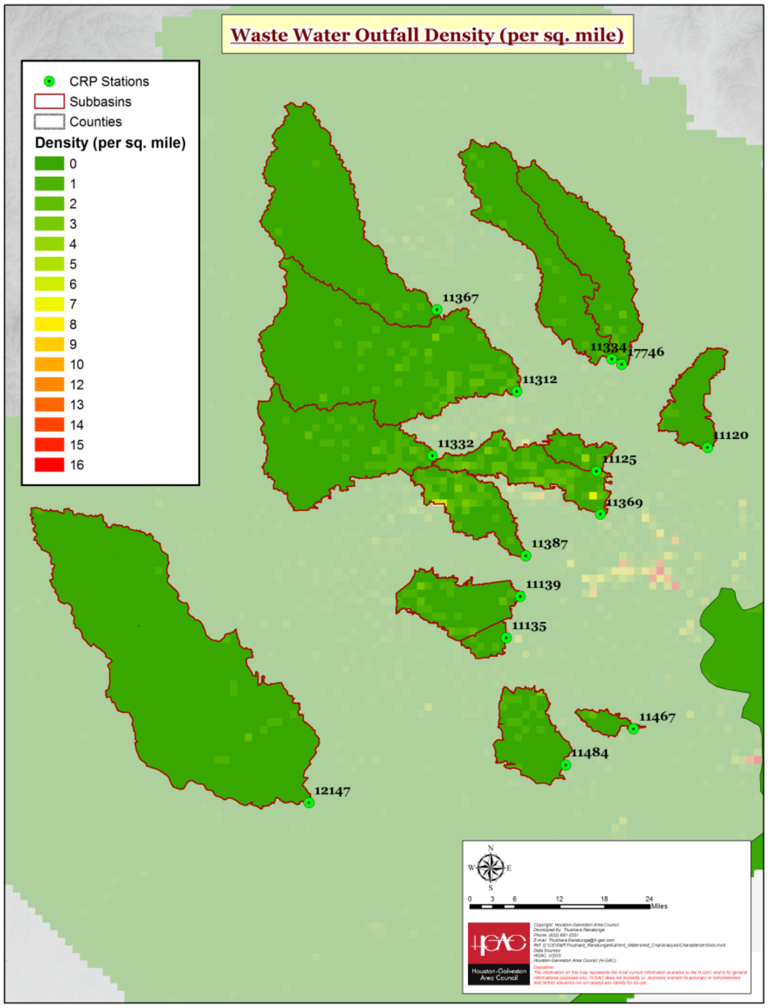

Figure 6: Waste Water Outfall Density

4.1.3 Soil Data Analysis

Figure 7 shows a map of the soil types in the project region. In addition, Table 22 includes the drainage characteristics (drainage class), geomorphic descriptions, the predominant hydrologic groups, and the area (in acres) for each of the taxonomic soil orders present in each watershed. Geographic features in the area include meander scrolls and gilgai on flats on coastal plains. Mollisols, Vertisols, Alfisols, and Utisols are the primary soils in the area and the watersheds in the region contain combinations of these soils that range from hydrologic group B to D soils. The majority of soils in the project watersheds are poorly draining hydrologic group D soils which have less infiltration capacity and high runoff potential. The majority of the Cedar Bayou watershed soils, for example, are classified as Mollisols and Vertisols with somewhat poorly drained to moderately well drained soil types. In addition, Garners Bayou is primarily Mollisols and Alfisols with poorly drained B and D hydrologic type soils. Water movement through the soil is restricted or very restricted. Group D soils typically have greater than 40% clay, less than 50% sand, and have clayey texture (USDA, 2007 – Hydrologic Soil groups).

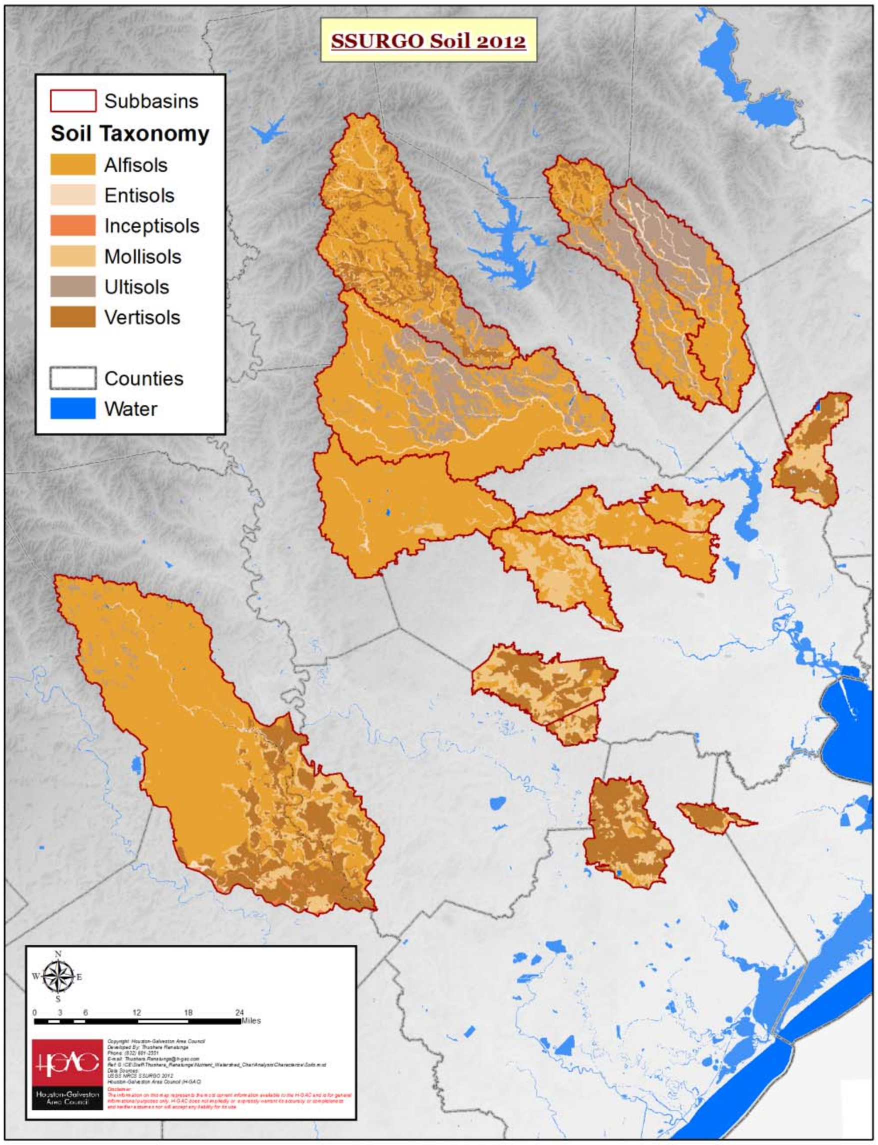

Figure 7: Soil Characteristics of Project Watersheds

Table 22: Major Soil types of each watershed

Watershed	Station ID	Taxonomic Order	Drainage Class	Geomorphic Description	Hydrologic Group	Area
CEDAR BAYOU	11120	Mollisols	Moderately well drained	meander scrolls on coastal plains	C/D	4985.69
CEDAR BAYOU	11120	Vertisols	Somewhat poorly drained	gilgai on flats on coastal plains	D	5940.87
CEDAR BAYOU	11120	Mollisols	Somewhat poorly drained	meander scrolls on coastal plains	D	9088.7
CEDAR BAYOU	11120	Vertisols	Poorly drained	gilgai on depressions on flats on coastal plains	D	12367.7
GARNERS BAYOU	11125	Mollisols	Poorly drained	flats on coastal plains	B/D	3429.7
GARNERS BAYOU	11125	Alfisols	Poorly drained	depressions on coastal plains	B/D	4402.0
GARNERS BAYOU	11125	Alfisols	Poorly drained	flats on coastal plains	B/D	5117.2
SIMS BAYOU	11135	Vertisols	Moderately well drained	gilgai on flats on coastal plains	D	4268.9
SIMS BAYOU	11135	Mollisols	Somewhat poorly drained	meander scrolls on coastal plains	D	6893.1
BRAYS BAYOU	11139	Mollisols	Somewhat poorly drained	meander scrolls on coastal plains	D	25448.5
BRAYS BAYOU	11139	Vertisols	Moderately well drained	gilgai on flats on coastal plains	D	26302.4
SPRING CREEK	11312	Alfisols	Somewhat poorly drained	low hills on coastal plains	C/D	32843.6
SPRING CREEK	11312	Ultisols	Moderately well drained	interfluves on coastal plains	B	35014.6
CYPRESS CREEK	11332	Alfisols	Poorly drained	depressions on coastal plains	B/D	24975.4

Watershed	Station ID	Taxonomic Order	Drainage Class	Geomorphic Description	Hydrologic Group	Area
CYPRESS CREEK	11332	Alfisols	Somewhat poorly drained	low hills on coastal plains	C/D	58278.5
CANEY CREEK	11334	Alfisols	Poorly drained	flats on coastal plains	C/D	14649.1
CANEY CREEK	11334	Ultisols	Moderatel y well drained	interfluves on coastal plains	B	20730.6
LAKE CREEK	11367	Alfisols	Moderatel y well drained	interfluves on coastal plains	D	18536.3
LAKE CREEK	11367	Alfisols	Somewhat poorly drained	interfluves on coastal plains	A	21122.8
GREENS BAYOU	11369	Mollisols	Poorly drained	flats on coastal plains	B/D	9624.8
GREENS BAYOU	11369	Alfisols	Poorly drained	depressions on coastal plains	B/D	11419.9 6
GREENS BAYOU	11369	Alfisols	Poorly drained	flats on coastal plains	B/D	26374.4
WHITEOAK BAYOU	11387	Alfisols	Poorly drained	flats on coastal plains	C	6845.5
WHITEOAK BAYOU	11387	Alfisols	Poorly drained	flats on coastal plains	B/D	10451.4
WHITEOAK BAYOU	11387	Alfisols	Poorly drained	depressions on coastal plains	B/D	13241.8
WHITEOAK BAYOU	11387	Mollisols	Poorly drained	flats on coastal plains	B/D	21120.2
DICKINSON BAYOU	11467	Mollisols	Somewhat poorly drained	meander scrolls on coastal plains	D	2540.9
DICKINSON BAYOU	11467	Vertisols	Moderatel y well drained	gilgai on flats on coastal plains	D	6903.3
CHOCOLAT E BAYOU	11484	Mollisols	Somewhat poorly drained	meander scrolls on coastal plains	D	14312.9
CHOCOLAT E BAYOU	11484	Vertisols	Moderatel y well drained	gilgai on flats on coastal plains	D	38151.9

Watershed	Station ID	Taxonomic Order	Drainage Class	Geomorphic Description	Hydrologic Group	Area
SAN BERNARD RIVER	12147	Alfisols	Somewhat poorly drained	flats on coastal plains	D	44099.4
SAN BERNARD RIVER	12147	Vertisols	Moderately well drained	gilgai on flats on coastal plains	D	52355.1
SAN BERNARD RIVER	12147	Alfisols	Moderately well drained	meander scrolls on coastal plains	D	57124.8
PEACH CREEK	17746	Alfisols	Somewhat poorly drained	hills on coastal plains	C/D	13038.8
PEACH CREEK	17746	Alfisols	Poorly drained	flats on coastal plains	C/D	14292.9
PEACH CREEK	17746	Ultisols	Moderately well drained	interfluves on coastal plains	B	28931.3

Note: the different shading used in the table represents the different study watersheds.

4.1.4 Drainage Density

Table 23 shows the drainage density in miles per square mile for each watershed, including the minimum, maximum, and average drainage densities and the standard deviation. As described in Section 3.3.4, the drainage density of a basin is the total line length of all the streams in a watershed divided by the watershed area. As shown in the table, Cedar Bayou, Spring Creek, Lake Creek, Dickinson Bayou, and Chocolate Bayou have the highest drainage densities of the project subwatersheds. Garners Bayou, Brays Bayou, Greens Bayou, Whiteoak Bayou, and Peach Creek have the lowest average drainage densities. Figure 8 shows a map of the drainage densities for the study area.

Table 23: Drainage Density Results

Station ID	Watershed Name	Drainage Density (Miles per sq. mile)			
		Min	Max	Average	Standard Deviation
11120	Cedar Bayou	1.32	3.35	2.67	0.35
11125	Garners Bayou	0.94	2.18	1.65	0.27
11135	Sims Bayou	1.45	2.25	1.9	0.13
11139	Brays Bayou	0.42	2	1.55	0.3
11312	Spring Creek	1.15	3.03	2.08	0.42
11332	Cypress Creek	1.18	2.54	1.86	0.34
11334	Caney Creek	0.88	2.76	1.81	0.46
11367	Lake Creek	2.14	3.95	3.2	0.28
11369	Greens Bayou	1.09	2.2	1.56	0.23
11387	Whiteoak Bayou	0.78	1.94	1.52	0.18
11467	Dickinson Bayou	1.89	2.41	2.21	0.08
11484	Chocolate Bayou	1.95	3.33	2.8	0.22
12147	San Bernard River	0.88	3.03	1.79	0.42
17746	Peach Creek	0.82	2.69	1.65	0.35

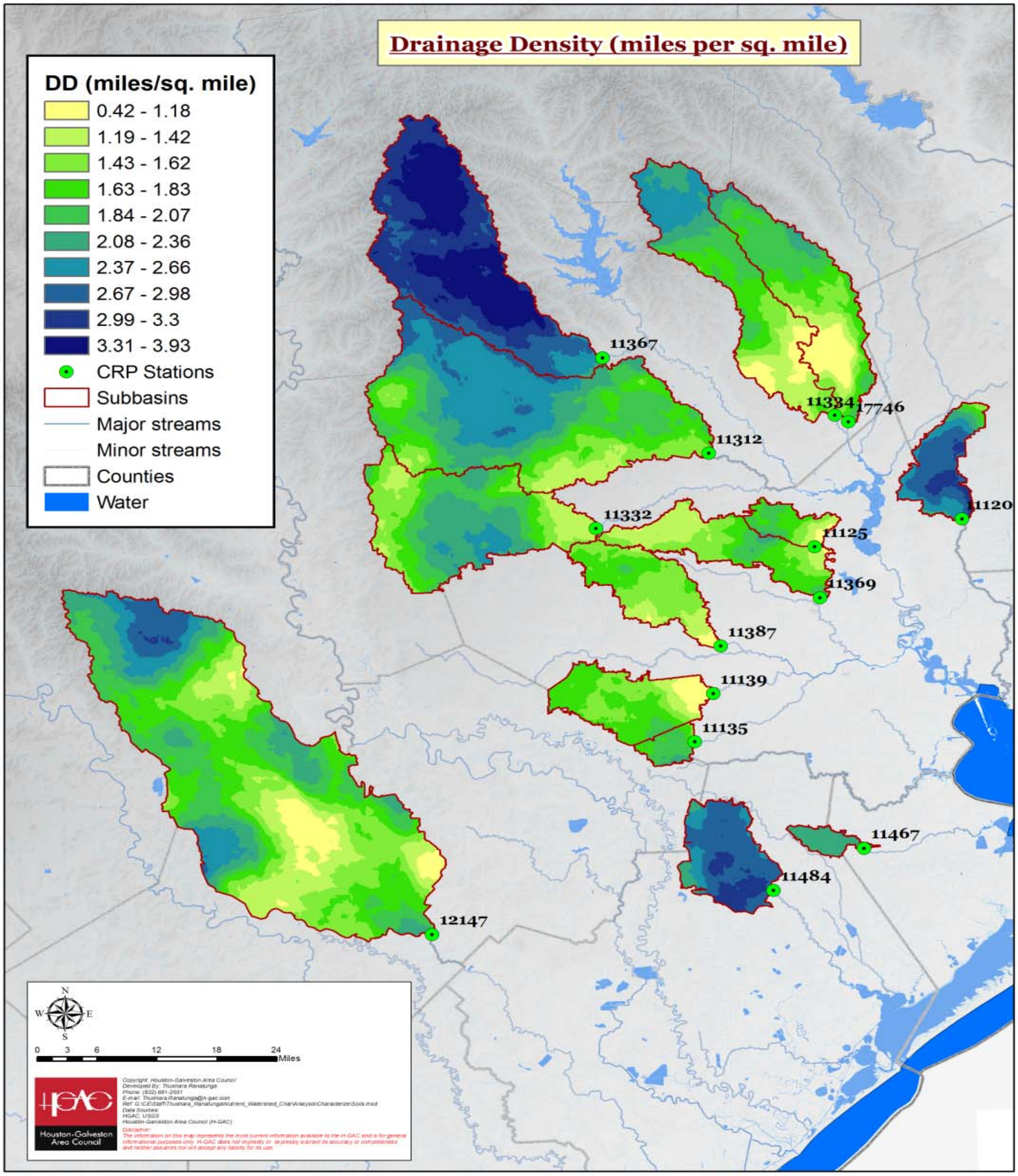

Figure 8: Drainage Density

4.1.5 Road Density

Table 24 includes a list of the length of roadway within each of the project watersheds. Brays Bayou, Spring Creek, Greens Bayou, Whiteoak Bayou, and the San Bernard River subwatersheds have the most miles of road within the study area. Dickinson Bayou and Cedar Bayou have the fewest miles of road within the study area. Figure 9 includes a map of that shows the roads in each of the subwatersheds.

Table 24: Total Road Miles by Watershed

Watershed Name	Station ID	Road Lengths (miles)
Cedar Bayou	11120	102.5
Garners Bayou	11125	248.5
Sims Bayou	11135	300.4
Brays Bayou	11139	2154.0
Spring Creek	11312	2156.1
Cypress Creek	11332	838.1
Caney Creek	11334	565.5
Lake Creek	11367	628.6
Greens Bayou	11369	1270.0
Whiteoak Bayou	11387	1552.2
Dickinson Bayou	11467	64.9
Chocolate Bayou	11484	294.0
San Bernard River	12147	1572.8
Peach Creek	17746	533.0

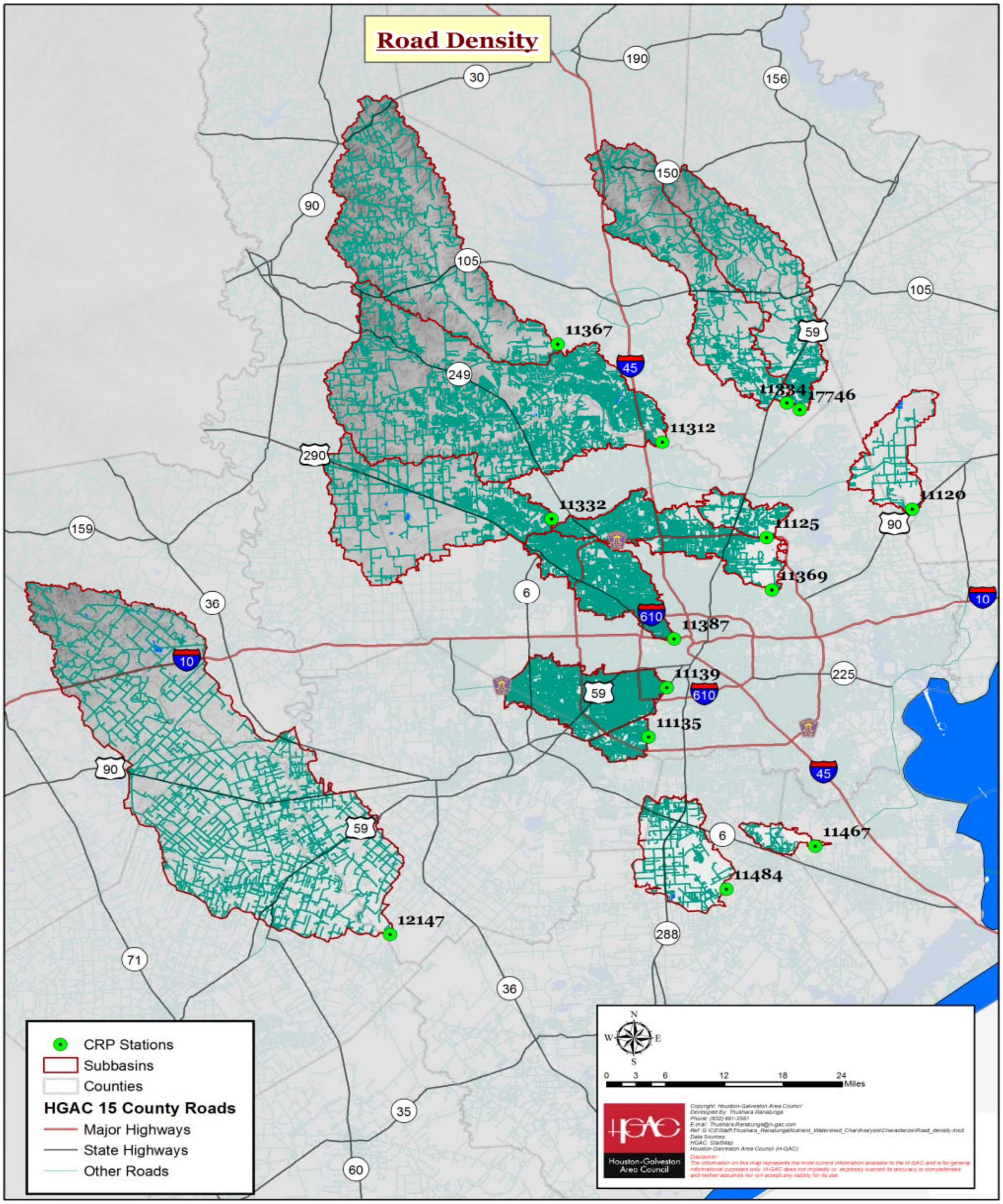

Figure 9: Road Density

4.1.6 Imperviousness

Table 25 lists the total impervious acreage for each watershed of the project area. Spring Creek has the highest amount of impervious area, almost 75,000 acres. Brays Bayou, Whiteoak Bayou, and Greens Bayou also had comparatively large amounts of impervious area with greater than 40,000 acres of impervious area. Cedar Bayou, Sims Bayou, Dickinson Bayou, and Chocolate Bayou had the least amount of impervious area. All had under 10,000 acres. The percentages are also presented in Figure 10.

Table 25: Impervious Acres by Watershed

Watershed Name	Station ID	Total Impervious Area (Acres)
Cedar Bayou	11120	1,844.6
Garners Bayou	11125	12,355.4
Sims Bayou	11135	9,455.3
Brays Bayou	11139	57,966.9
Spring Creek	11312	74,744.9
Cypress Creek	11332	24,040.5
Caney Creek	11334	18,396.3
Lake Creek	11367	10,989.6
Greens Bayou	11369	44,439.0
Whiteoak Bayou	11387	53,140.0
Dickinson Bayou	11467	3,101.6
Chocolate Bayou	11484	7,471.4
San Bernard River	12147	22,560.6
Peach Creek	17746	11,825.6

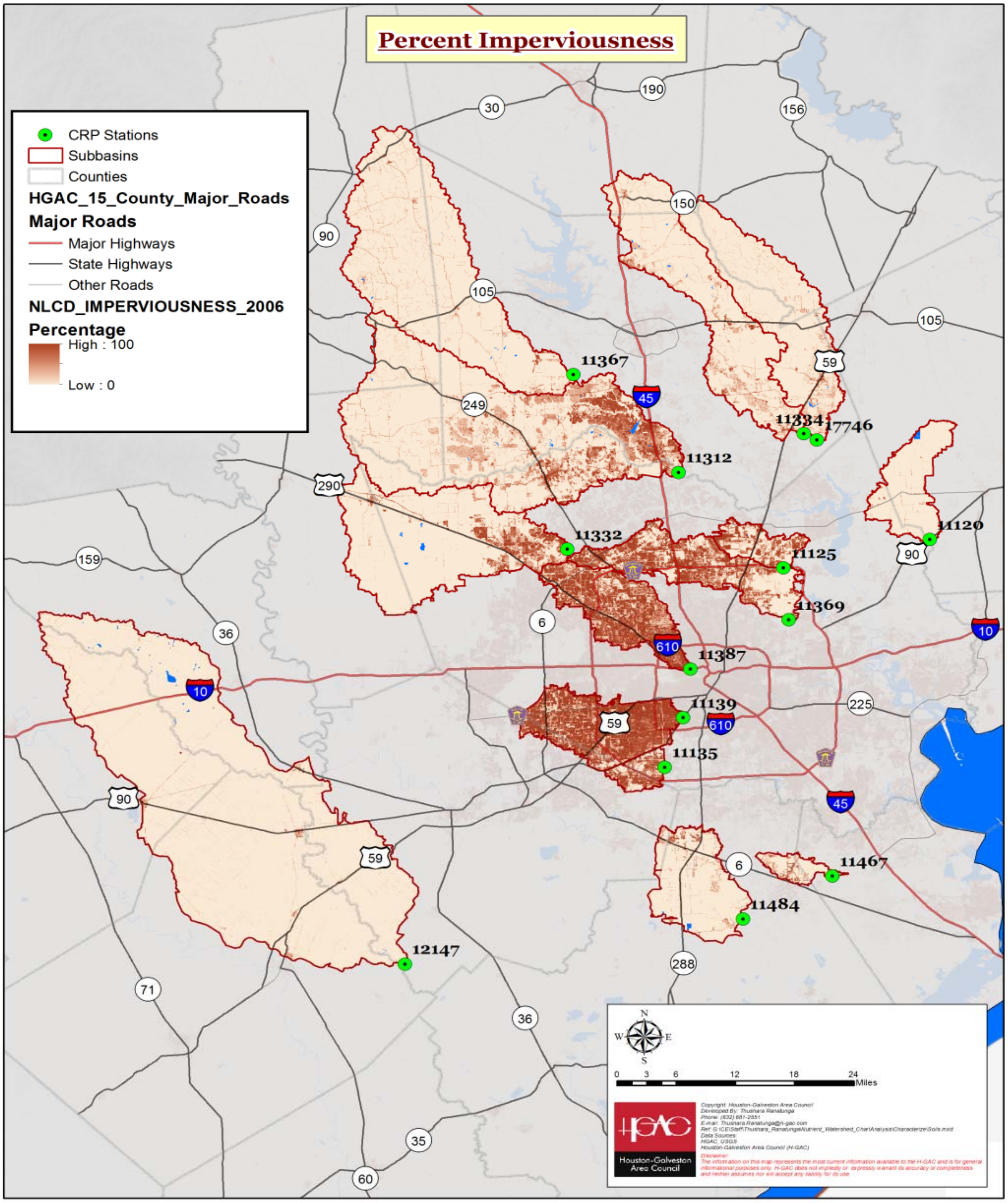

Figure 10: Percent Impervious

4.1.7 Watershed Relief

The elevation parameter values are listed in Table 26, including the minimum, maximum, average values, the standard deviation, and the relief for each watershed. Cypress Creek, Caney Creek, Lake Creek, and Peach Creek have the highest watershed relief elevations. All are above 55 meters. Higher relief can cause greater erosion and eventually create additional stream channels. Therefore, those watersheds show high drainage density compared to other watersheds. The lowest values for reliefs are found at Dickinson Bayou, Chocolate Bayou, and Sims Bayou that are all below 20 meters. These are also presented in a map in Figure 11.

Table 26: Watershed Relief Results

Station ID	Watershed Name	Elevation (meters)				
		Min	Max	Average	Standard Deviation	Relief
11120	Cedar Bayou	11.92	32.07	22.94	3.12	20.15
11125	Garners Bayou	8.36	58.15	23.22	3.86	49.79
11135	Sims Bayou	9.37	31.62	19.83	2.14	22.25
11139	Brays Bayou	4.84	50.32	21.54	3.9	45.48
11312	Spring Creek	19.87	135.59	68.7	20.66	115.72
11332	Cypress Creek	30.7	96.12	59.69	12.03	65.42
11334	Caney Creek	17.36	137.25	69.61	27.1	119.89
11367	Lake Creek	39.15	145.11	89.99	18.59	105.96
11369	Greens Bayou	1.52	49.29	25.28	7.51	47.77
11387	Whiteoak Bayou	2.42	49.25	29.04	6.39	46.83
11467	Dickinson Bayou	-3.72	20.85	11.1	1.89	24.57
11484	Chocolate Bayou	2.04	25.95	14.78	2.69	23.91
12147	San Bernard River	8.5	136.21	49.71	22	127.71
17746	Peach Creek	15.5	134.4	60.81	25.12	118.9

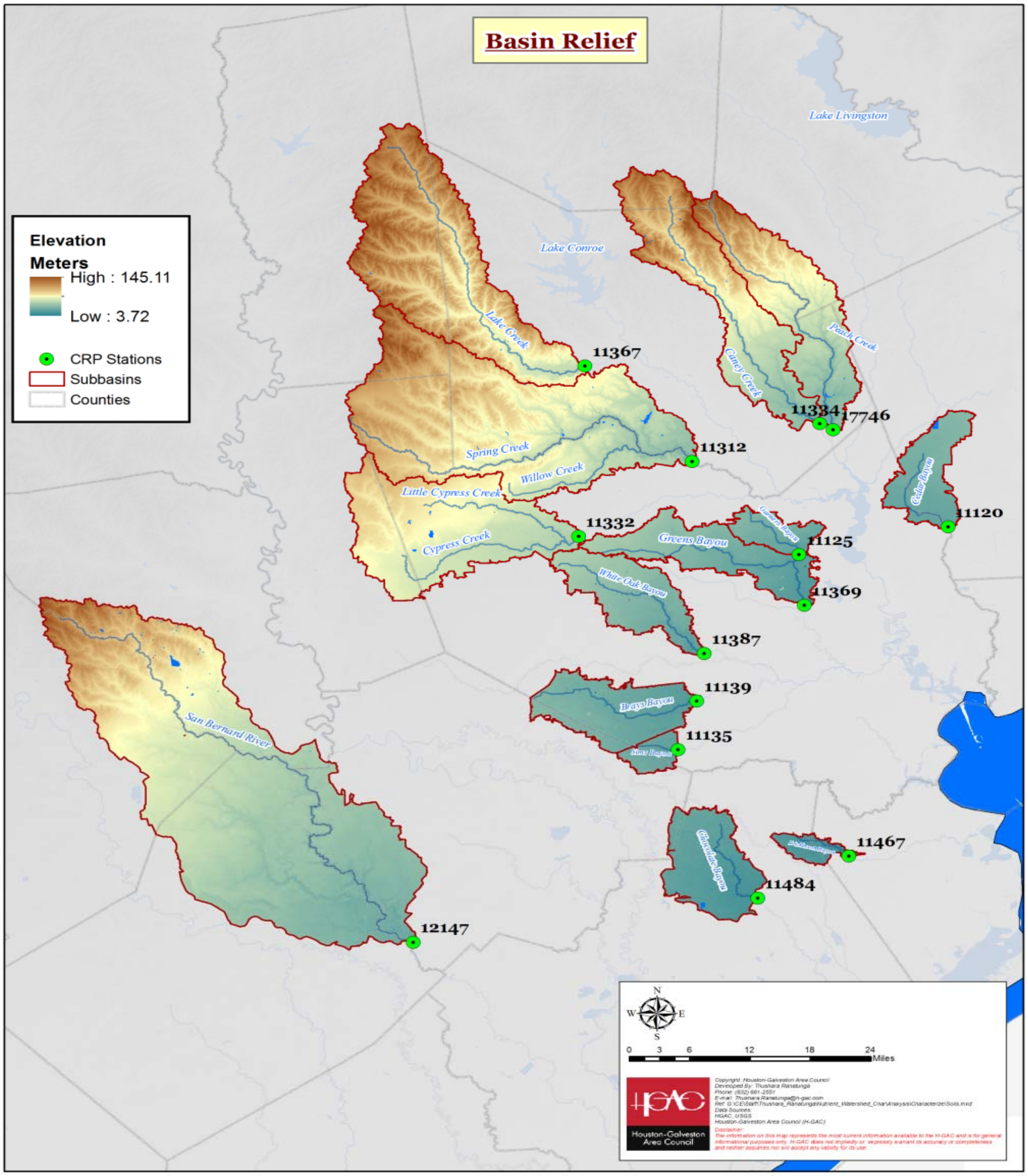

Figure 11: Basin Relief

4.2 Trend Analysis

H-GAC analyzed nutrient data for each site to identify and characterize temporal trends. The results are summarized in Table 27 below. Multiple approaches were evaluated, and H-GAC considers three of these as the most reliable:

1. Non-parametric correlation analysis of the residuals from a LOESS fit to flow (flow adjusted concentration)
2. Robust regression using both date and flow as predictors
3. Ordinary least squares (linear) regression of the annual median of nutrient concentration on the year.

If the trends identified by these three methods were not supported by several of the auxiliary methods the discrepancies were reviewed and appropriate action taken. Method 1 is the most widely recommended trend analysis technique (Hirsch 1991; Helsel *et al* 1991; Helsel and Hirsch 2002; McFarland and Millican 2011), and was used in the land cover change analysis. When the requisite assumptions are met, parametric methods (robust and linear regression, for example) are more powerful than non-parametric methods. Trends identified using these methods (2 and 3) but not by method 1 were graphically evaluated and included in the table for comparison if the existence of a trend could not be discounted. Finally, if more than fifteen percent of the data are censored, survival analysis was performed. Results of survival analysis appear in the comments field of the table below. Method 3 (regression of annual medians on the year) was employed to eliminate trends introduced by seasonality in the data. Seasonal Kendall (Helsel and Hirsch 2002), Sen Slope estimation (Helsel and Hirsch 2002), Theil regression (Hess *et al* 2002), and time series methods such as ARIMA or the unobserved component model (Ragavan and Fernandez 2006) are better ways to control for seasonality, but an evenly spaced time series is required to properly identify trends. This requirement was not met by most data series, and this created significant differences between the results of these analyses and non-seasonal tests, even in cases where the seasonality test did not reveal seasonal effects.

Analysis of flow-adjusted concentrations identified twelve statistically significant trends. Each of these trends was also identified by one or more of the other two methods. If the flow-adjusted concentration trend was not significant but one of the other two methods suggested that a statistically significant trend was present, a time series of the natural log was plotted and examined.

1. The following trends were identified by robust regression with control for flow, but were not included in the land cover change/ nutrient trend comparison maps:

i. Decreasing total nitrogen concentrations at station 11120. A plot of the time series did not support this, but the LOESS fit showed that total nitrogen concentration began to increase in 2010, near the beginning of the current drought. The trend is not included in the summary below.
ii. Increasing nitrate concentration at station 11125, but the trend was weak. A similar pattern in the total phosphorus data was observed. These trends are not included in the summary below.
iii. Increasing total phosphorus at station 11139. Graphical examination did not disqualify the trend, and it is included in the summary below.
iv. Increasing nitrate concentration at station 11367. Graphical examination revealed a cluster of censored and/or low values between 2007 and 2009. A return to higher values after 2010 may have created the appearance of a trend. The trend is not included in the summary. Additionally, increasing total phosphorus at this station was suggested by robust regression. Examination of a LOESS fit showed that the concentration generally increased until 2009-2010, and then began a decline. The nitrate trend suggested by flow-controlled robust regression trend is not included in the summary. However, given two potential nutrient trends, and a poor temporal representation of total phosphorus, the existence trends toward increasing concentrations cannot be ruled out. It should be noted that the data set at this station consists of 16 total phosphorus and nitrate results, and 8 total nitrogen results. The latter is insufficient for reliable inference.
v. A weak trend toward increasing total nitrogen at station 11369. A plot of the time series showed the trend is an artifact of an outlier in 2009.The trend is not included in the summary.

2. Regression of the annual median of nitrate concentration suggested decreasing concentrations. More than fifteen percent of the data was censored, so survival analysis was performed, and the same trend was identified. It is possible that outliers in 1999 were responsible for some of the trend, but there are more censored data in the latter half of the period that in the beginning. The trend appears in the summary below.

3. Increasing total phosphorus at station 11334 was suggested by flow-adjusted concentration analysis and robust regression. Because more than 15 percent of the data are censored, survival analysis (PROC LIFEREG) was applied, and the presence of a trend was not supported. Graphical analysis show most of the censored data are found between 2004 and 2010, and the highest values are also seen during this period. The trend was not removed from the summary, but the possibility it is the result of laboratory practices rather than environmental conditions must be considered.

4. Survival analysis suggested declining nitrate at station 11484; no other test identified a trend. A review of the time series does not support the existence of a trend, and it is not listed in the table below.

Table 27: Trend Analysis - Summary

		LOESS Residual / Kendall Correlation		Robust Regression with Flow		Annual median OLS		Comments
Station	**Parameter**	**Trend**	**N**	**Trend**	**N**	**Trend**	**N**	
11120	Nitrate-N		103		103		16	>15% Censored ; no trend
11120	Total Nitrogen		71		51		16	
11120	Total Phosphorus	Increasing	117	Increasing	117		16	
11125	Nitrate-N		28		28		10	
11125	Total Nitrogen		8		8		2	
11125	Total Phosphorus		28		28		10	
11135	Nitrate-N	Decreasing	107	Decreasing	107	Decreasing	13	
11135	Total Nitrogen		23		20		6	
11135	Total Phosphorus		113		113		14	
11139	Nitrate-N		82		159		16	
11139	Total Nitrogen		48		36		16	
11139	Total Phosphorus		75	Increasing	154		16	
11312	Nitrate-N		19		19		13	
11312	Total		16		10	Increasing	5	

	Nitrogen							
11312	Total Phosphorus		20		20		13	
11332	Nitrate-N	Increasing	130	Increasing	130		15	
11332	Total Nitrogen	Increasing	48	Increasing	29		15	
11332	Total Phosphorus	Increasing	130	Increasing	130		15	
11334	Nitrate-N		33		33		16	
11334	Total Nitrogen		33		12		14	
11334	Total Phosphorus	Increasing	25	Increasing	25		14	>15% Censored; no trend
11367	Nitrate-N		16		16		8	
11367	Total Nitrogen		12		5		3	
11367	Total Phosphorus		21		21		8	
11369	Nitrate-N		21		21		13	
11369	Total Nitrogen		19		16		13	
11369	Total Phosphorus		21		21		13	
11387	Nitrate-N		81		81		16	
11387	Total Nitrogen		68		21		16	
11387	Total Phosphorus		74		74		16	
11467	Nitrate-N		30		30	Decreasing	16	>15% Censored; Decreasing
11467	Total Nitrogen		22		8		12	

11467	Total Phosphorus		37		37		15	
11484	Nitrate-N		67		67		16	>15% Censored; See item 4 above
11484	Total Nitrogen		22		30		16	
11484	Total Phosphorus	Increasing	63	Increasing	63		16	
12147	Nitrate-N	Decreasing	81	Decreasing	81		16	
12147	Total Nitrogen		69		34		16	
12147	Total Phosphorus	Increasing	61		81		16	
17746	Nitrate-N	Increasing	20	Increasing	20	Increasing	9	
17746	Total Nitrogen		19		10		9	
17746	Total Phosphorus		14	Decreasing	14		9	>15% Censored; Decreasing

Note: Stations located on effluent-dominated streams are identified by a gray background.

4.3 Land Cover Change vs. Water Quality Trends

Using the trend analysis results from the section 4.2, H-GAC did a spatial mapping of land cover changing trends and nutrient trends as explains in the section 3.3.1. For the land cover changing trends, three categories were identified - Increasing, Stable, and Decreasing. In selecting land cover classes, only Developed, Forest, Cultivated, Grassland, and Wetlands were selected as the other classes did not show considerable net changes over the time period from 1996 to 2011. For the nutrient trends, LOESS Residual Kendall Correlation, with the flow adjusted nutrient concentration results from Table 27 was used. In the trend table, the empty values of each nutrient trend were considered as “No Trend” in the maps. Table 28 below summarizes the trends of land cover change over time and by nutrient for the 14 watersheds. The table provides the number of watershed counts used in each combination of land cover and nutrient trends. This mapping effort may not necessarily reflect the practical correlation of land cover change and nutrient trends. The actual relationships and correlations can be further augmented by the results from the statistical analysis.

Table 28: Land Cover and Water Quality Trends in Study Area

Land Cover Type	Land Cover Trend	Nutrient Type	Watershed Count		
			Decrease	No Trend	Increase
Cultivated	Stable	Nitrate	2	10	2
Cultivated	Stable	Total Nitrogen	0	13	1
Cultivated	Stable	Total Phosphorus	0	9	5
Developed	Increase	Nitrate	1	5	1
Developed	Increase	Total Nitrogen	0	6	1
Developed	Increase	Total Phosphorus	0	6	1
Developed	Stable	Nitrate	1	5	1
Developed	Stable	Total Nitrogen	0	7	0
Developed	Stable	Total Phosphorus	0	3	4
Forest	Decrease	Nitrate	0	5	1
Forest	Decrease	Total Nitrogen	0	6	0
Forest	Decrease	Total Phosphorus	0	5	1
Forest	Stable	Nitrate	2	5	1
Forest	Stable	Total Nitrogen	0	7	1
Forest	Stable	Total Phosphorus	0	4	4
Grasslands	Increase	Nitrate	0	1	1

Land Cover Type	Land Cover Trend	Nutrient Type	Watershed Count		
			Decrease	No Trend	Increase
Grasslands	Increase	Total Nitrogen	0	2	0
Grasslands	Increase	Total Phosphorus	0	1	1
Grasslands	Stable	Nitrate	2	9	1
Grasslands	Stable	Total Nitrogen	0	11	1
Grasslands	Stable	Total Phosphorus	0	8	4
Wetlands	Stable	Nitrate	2	10	2
Wetlands	Stable	Total Nitrogen	0	13	1
Wetlands	Stable	Total Phosphorus	0	9	5

4.3.1 Developed Land Cover Change

As shown in Figure 12, half of the study watersheds have an increasing trend in the percentage of developed land cover. Majority of these watersheds show no trends in nutrients with the increase of developed land. However, one watershed, draining to station 11332, is an exception having increases for all three nutrient types (as shown in Figure 12 through Figure 14).

The remaining watersheds show a somewhat constant percentage of developed land cover (as presented in Table 28) over the study period. With no change in developed land cover, total phosphorus in four watersheds (11120, 11334, 11484, and 12147) has increased (Figure 13). All the other watersheds show no trends in nutrients.

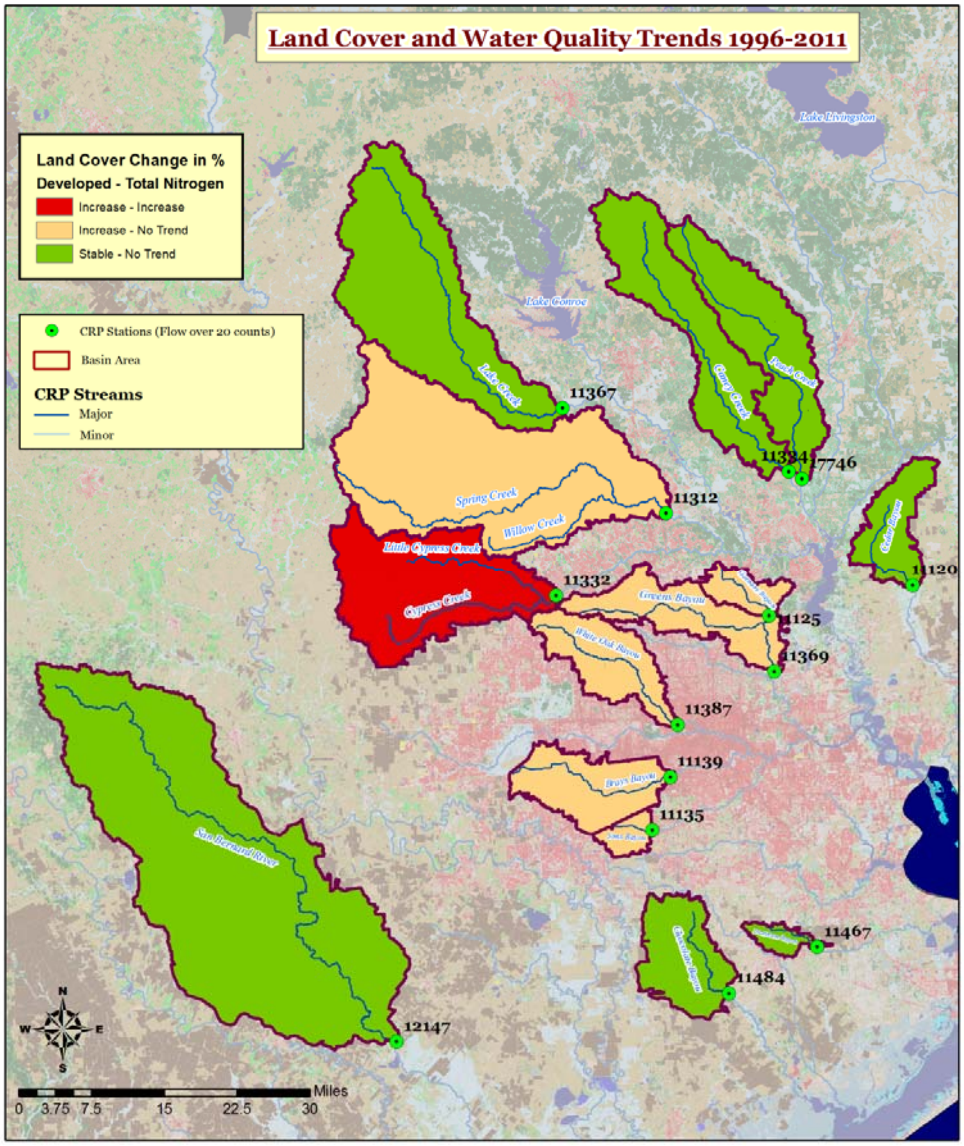

Figure 12: Land Cover and Water Quality Trends (Developed Land and Total Nitrogen)

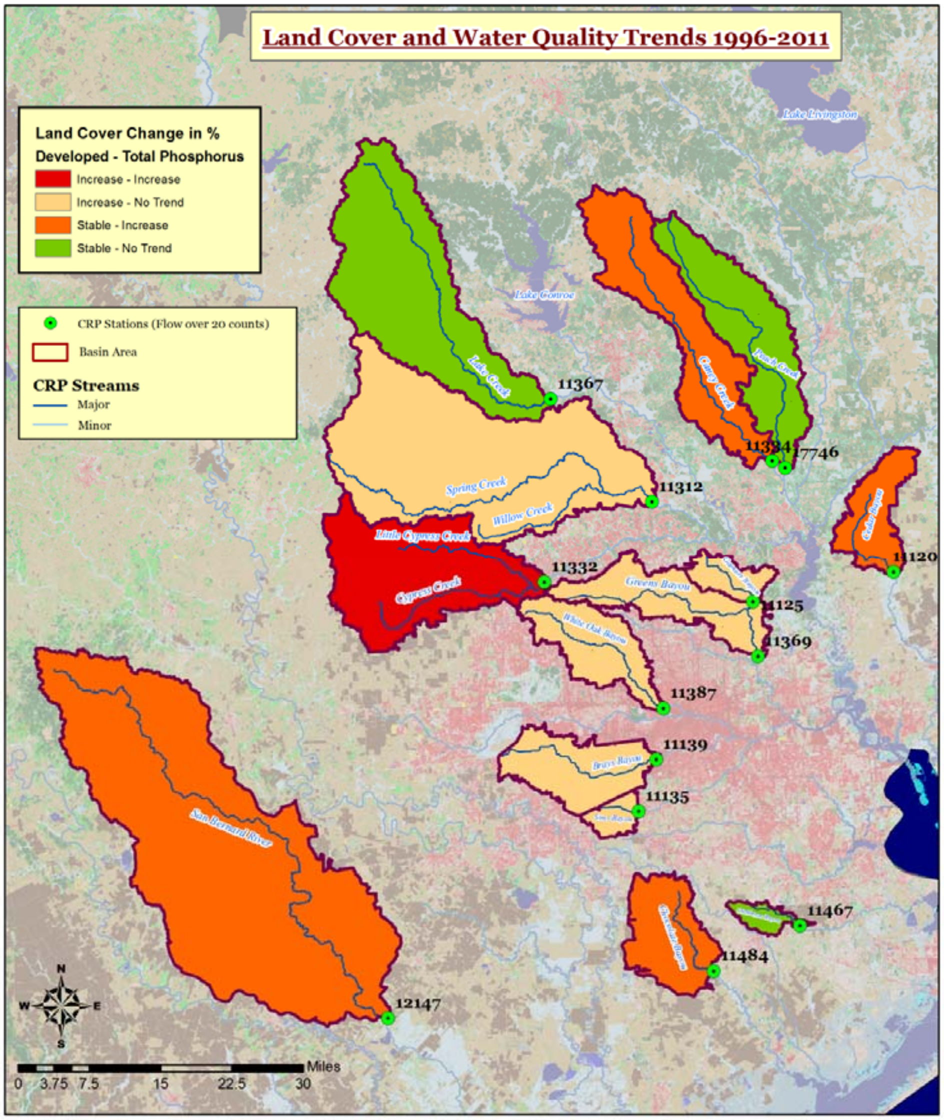

Figure 13: Land Cover and Water Quality Trends (Developed Land and Total Phosphorus)

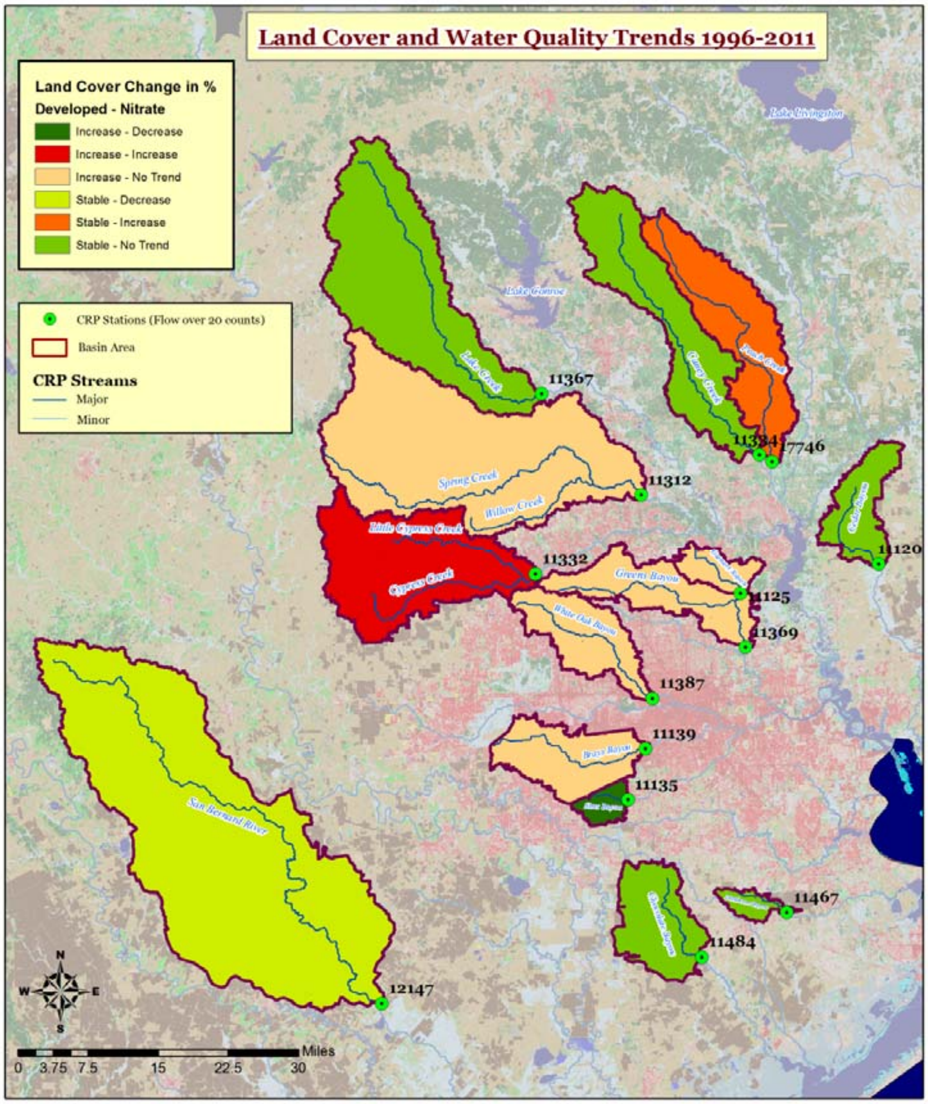

Figure 14: Land Cover and Water Quality Trends (Developed Land and Nitrate)

4.3.2 Forest Land Cover Change

As shown in Figure 15 through Figure 17, five of the watersheds had a decreasing trend in the percentage of forest land over the study period. All of these showed no trend in total nitrogen concentrations. The Caney Creek watershed (11334) showed an increase in total phosphorus concentrations and Peach Creek (17746) had an increase in nitrate, while the others had no change in these nutrients over the study period. The other watersheds showed no change in the amount of forest land.

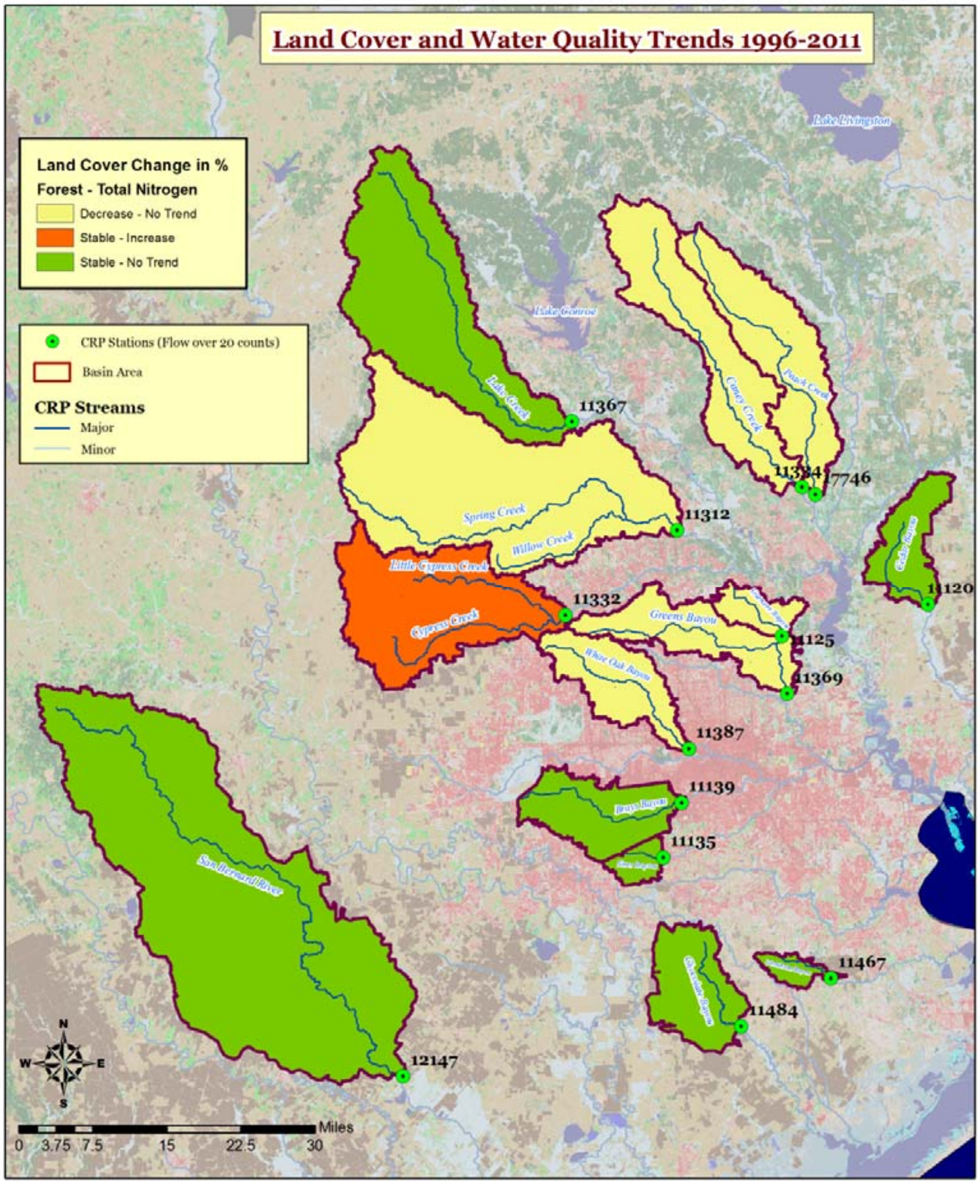

Figure 15: Land Cover and Water Quality Trends (Forest Land and Total Nitrogen)

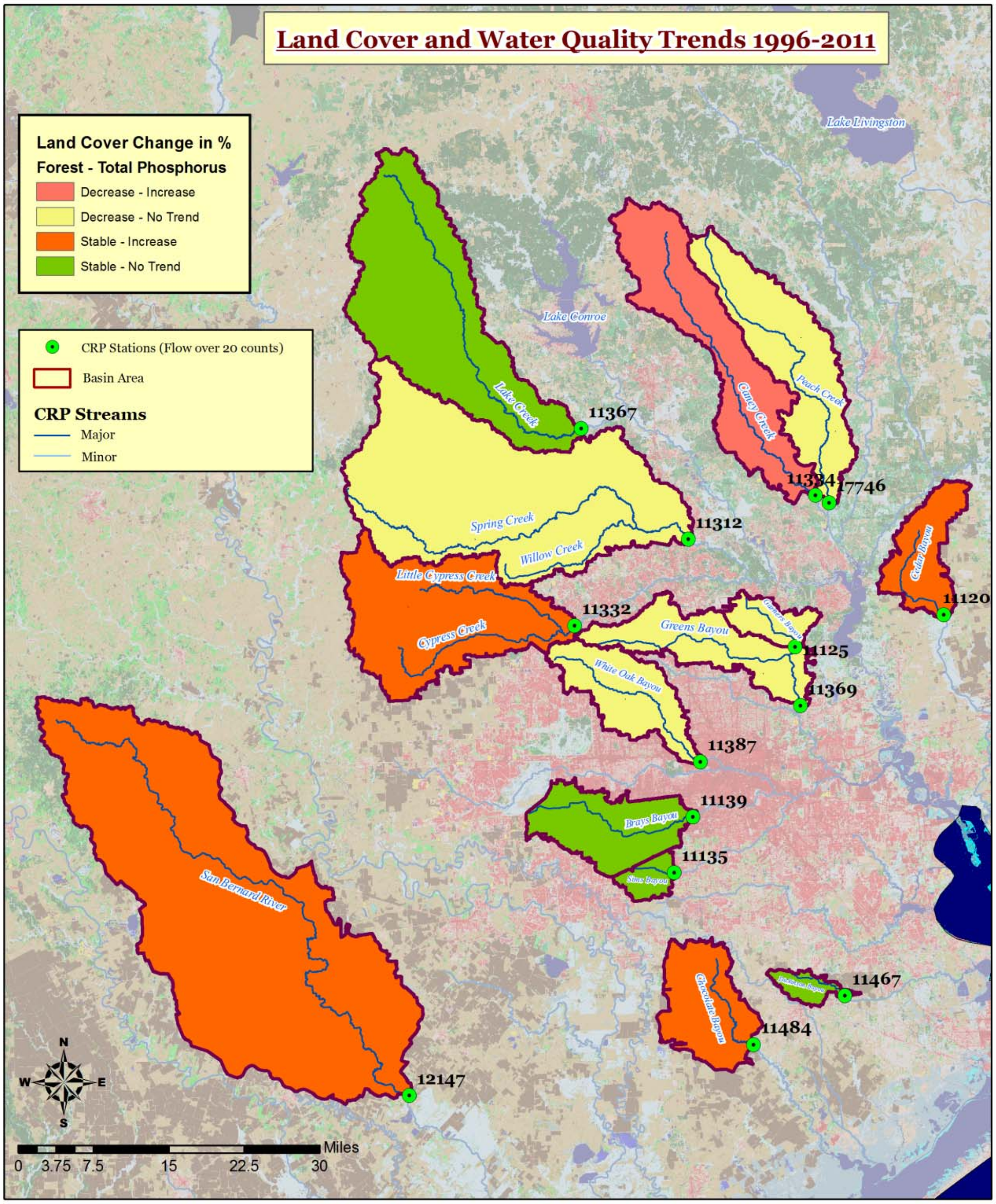

Figure 16: Land Cover and Water Quality Trends (Forest Land and Total Phosphorus)

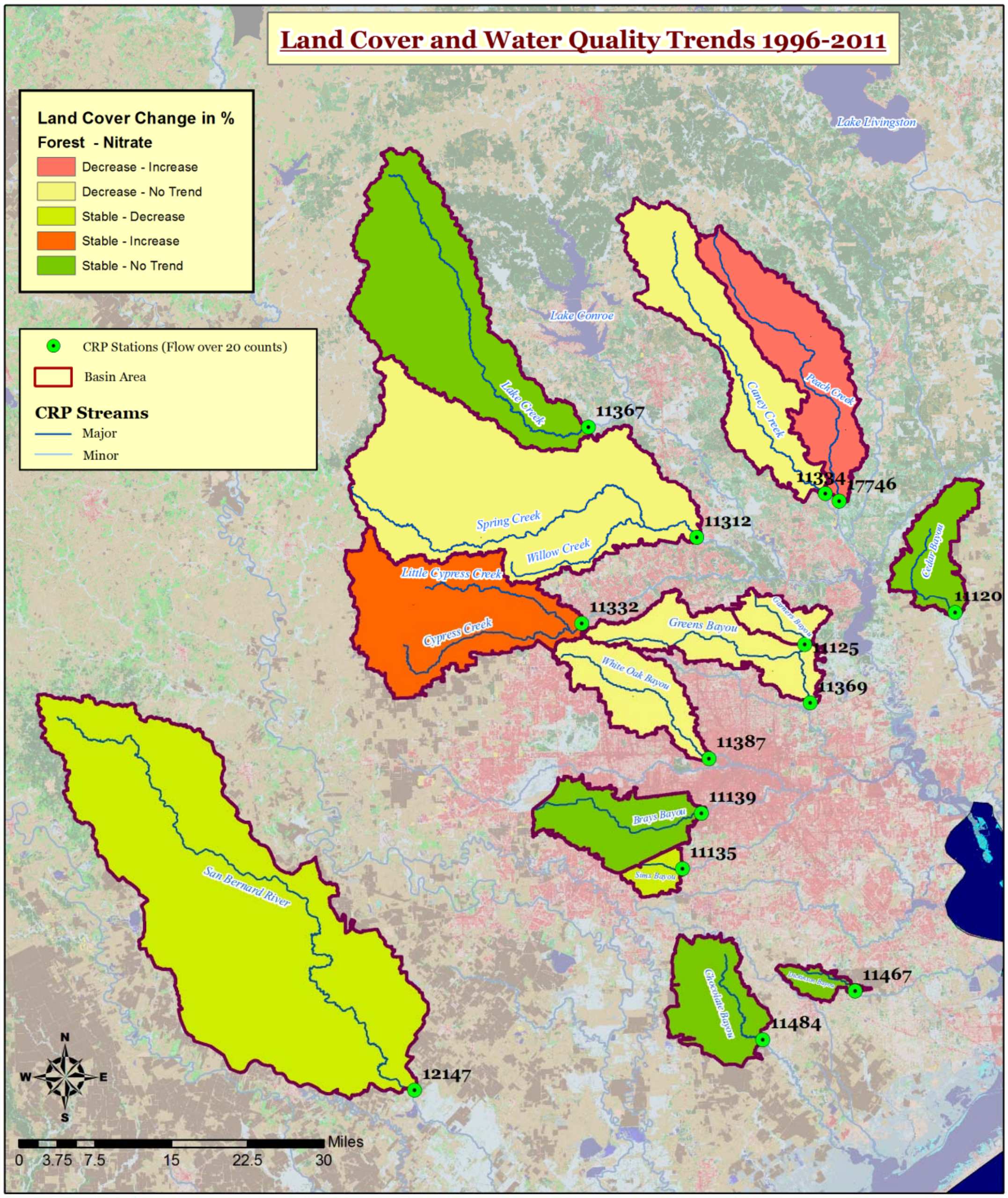

Figure 17: Land Cover and Water Quality Trends (Forest Land and Nitrate)

4.3.3 Cultivated Land Cover Change

As shown in Figure 18 through Figure 20, all of the watersheds in the study area had no trend in the percentage of cultivated land over the study period. One of these showed an increase in total nitrogen concentrations, two had an increase in nitrate concentrations, two had a decrease in nitrate concentrations, and five had an increase in total phosphorus concentrations, over the study period.

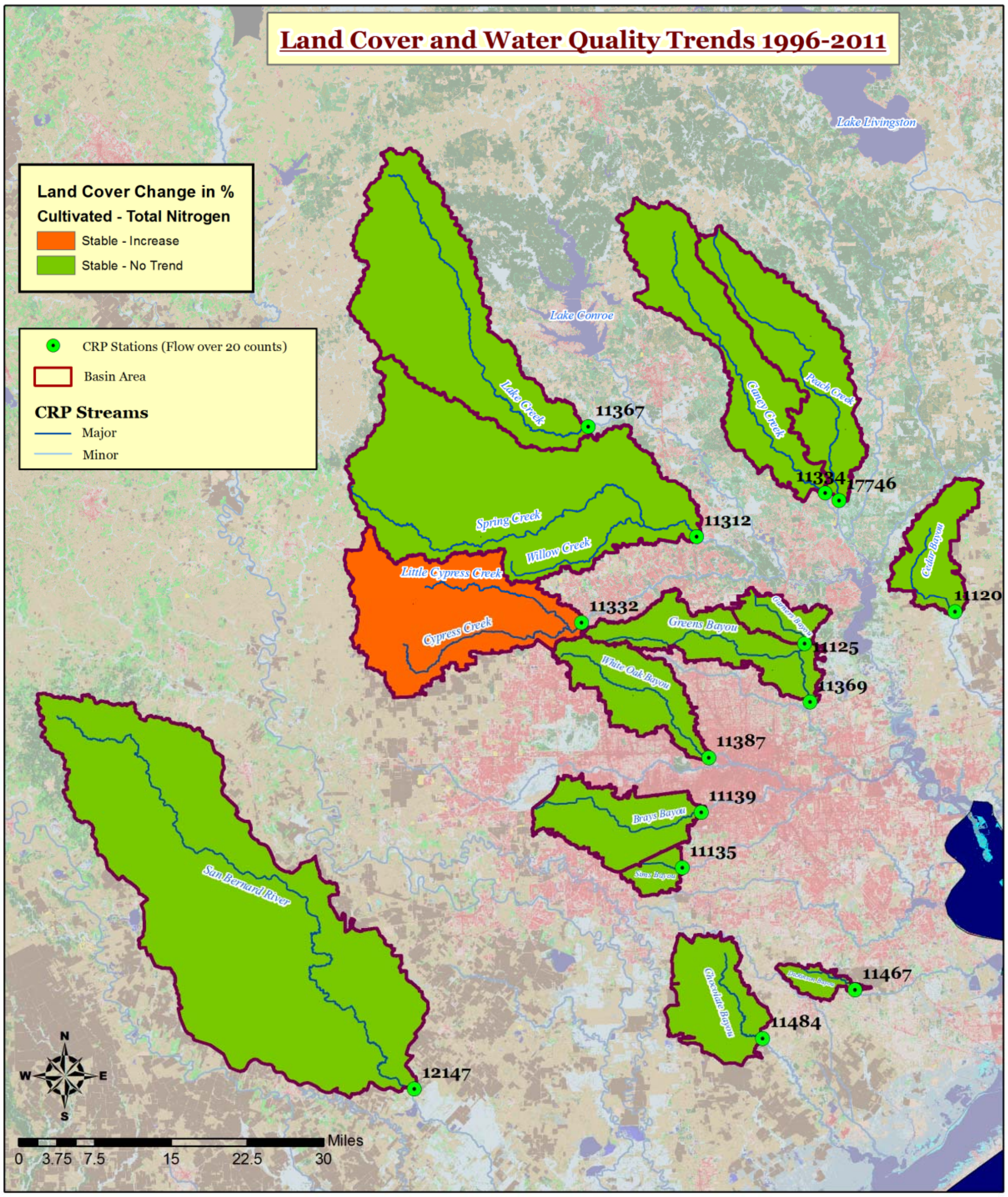

Figure 18: Land Cover and Water Quality Trends (Cultivated and Total Nitrogen)

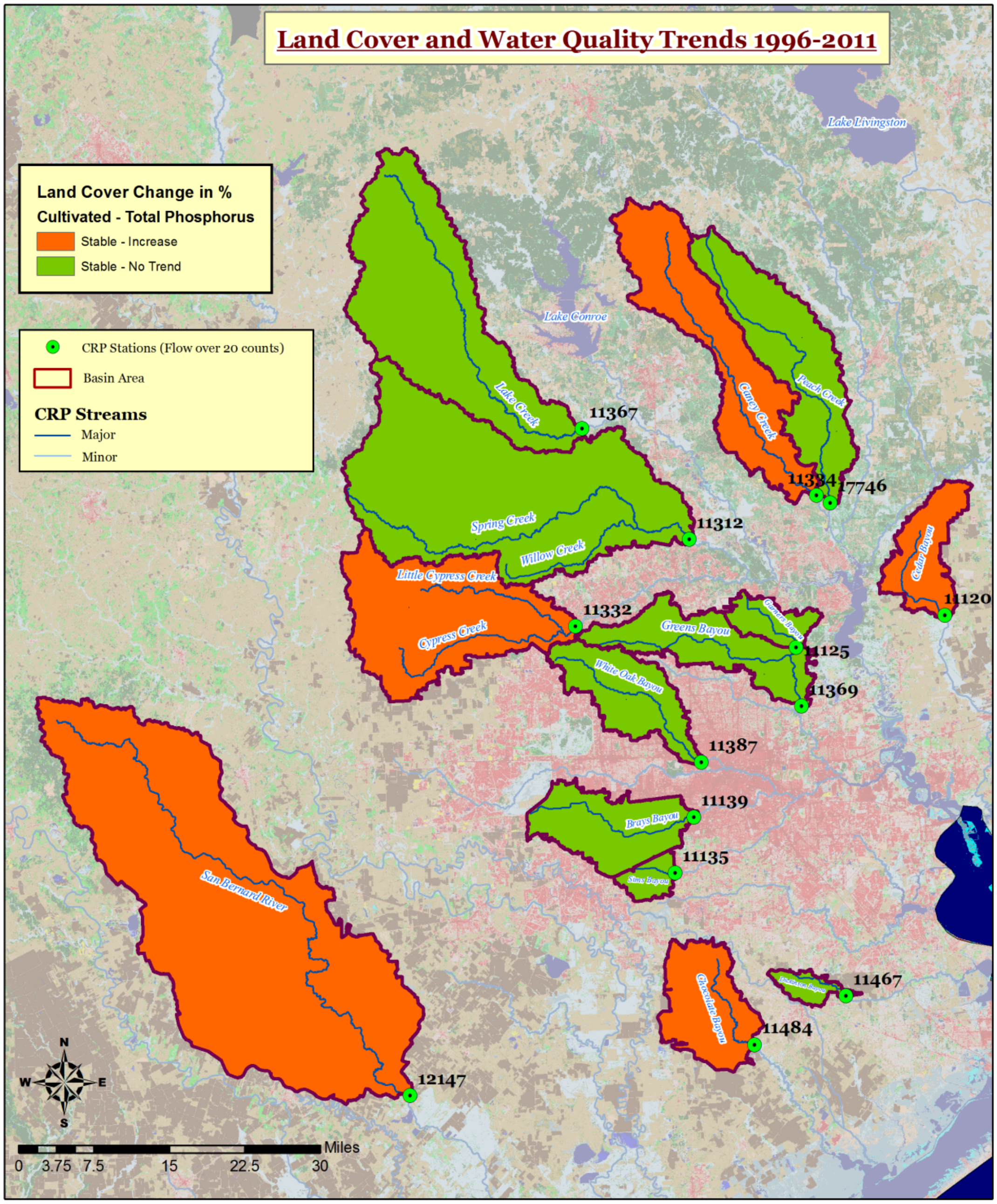

Figure 19: Land Cover and Water Quality Trends (Cultivated and Total Phosphorus)

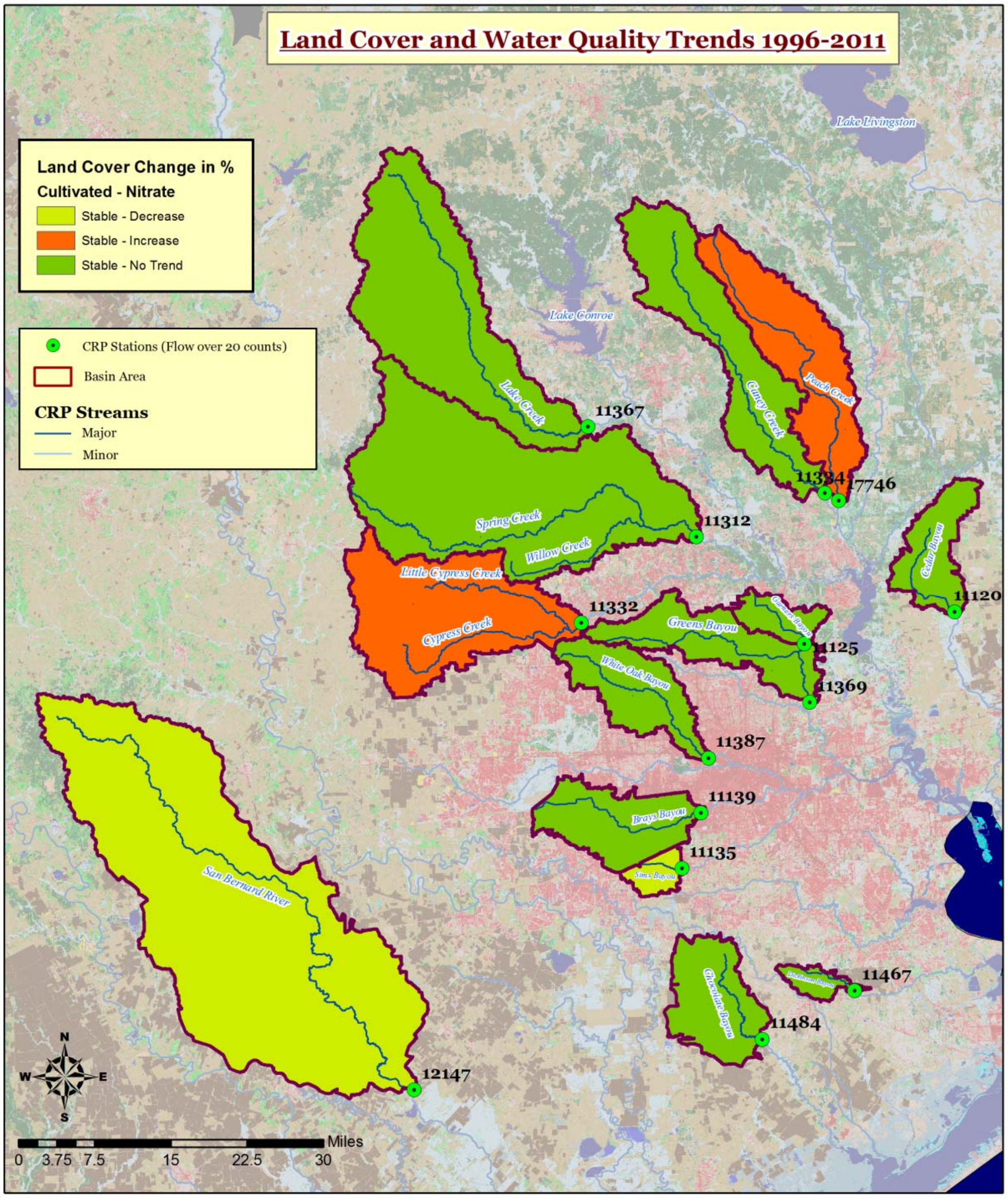

Figure 20: Land Cover and Water Quality Trends (Cultivated and Nitrate)

4.3.4 Grassland Change

As shown in Figure 21 through Figure 23, two of the watersheds had an increasing trend in the percentage of grassland over the study period. Neither of these watersheds showed an increasing or decreasing trend in total nitrogen concentrations. Peach Creek (17746) had an increasing trend for nitrate, while Caney Creek (11334) had no change over the study period. The Caney Creek watershed had an increase in total phosphorus concentrations. The other watersheds showed stable levels in the amount of grassland.

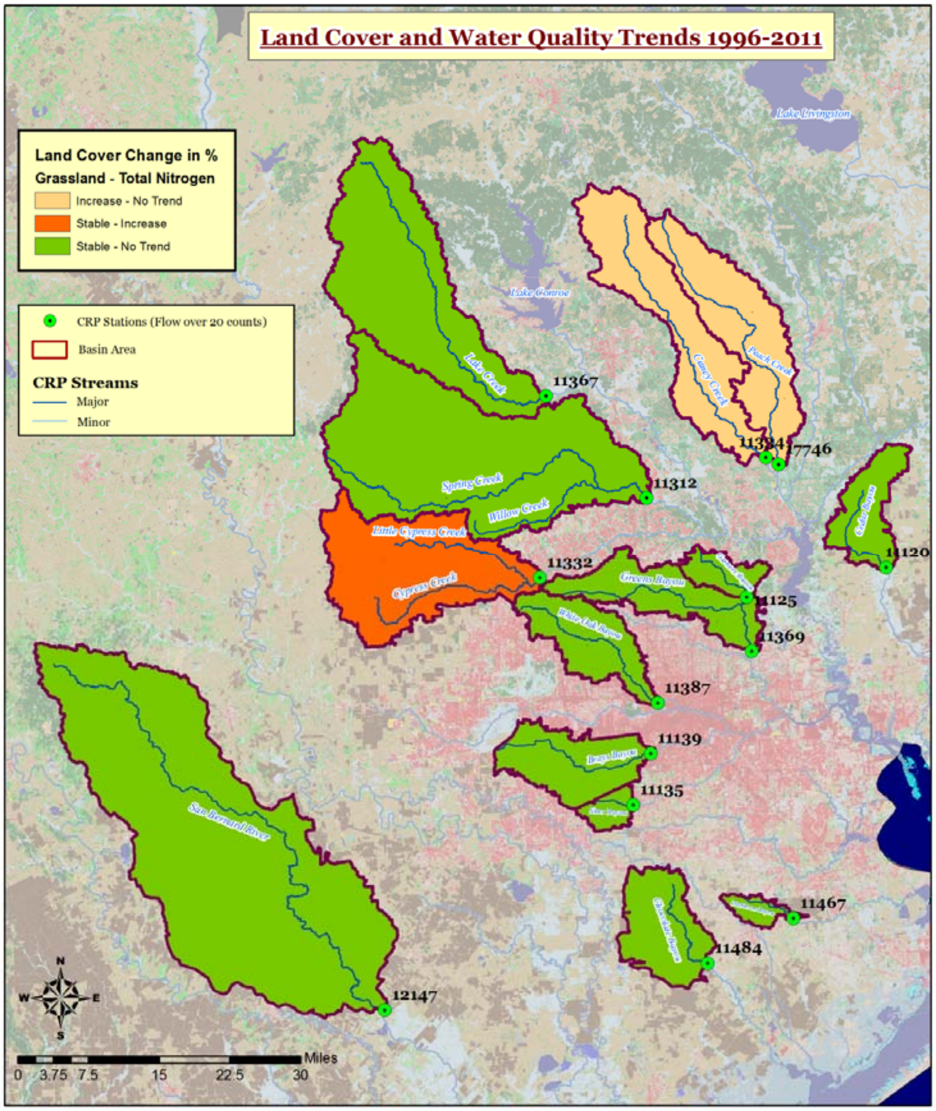

Figure 21: Land Cover and Water Quality Trends (Grassland and Total Nitrogen)

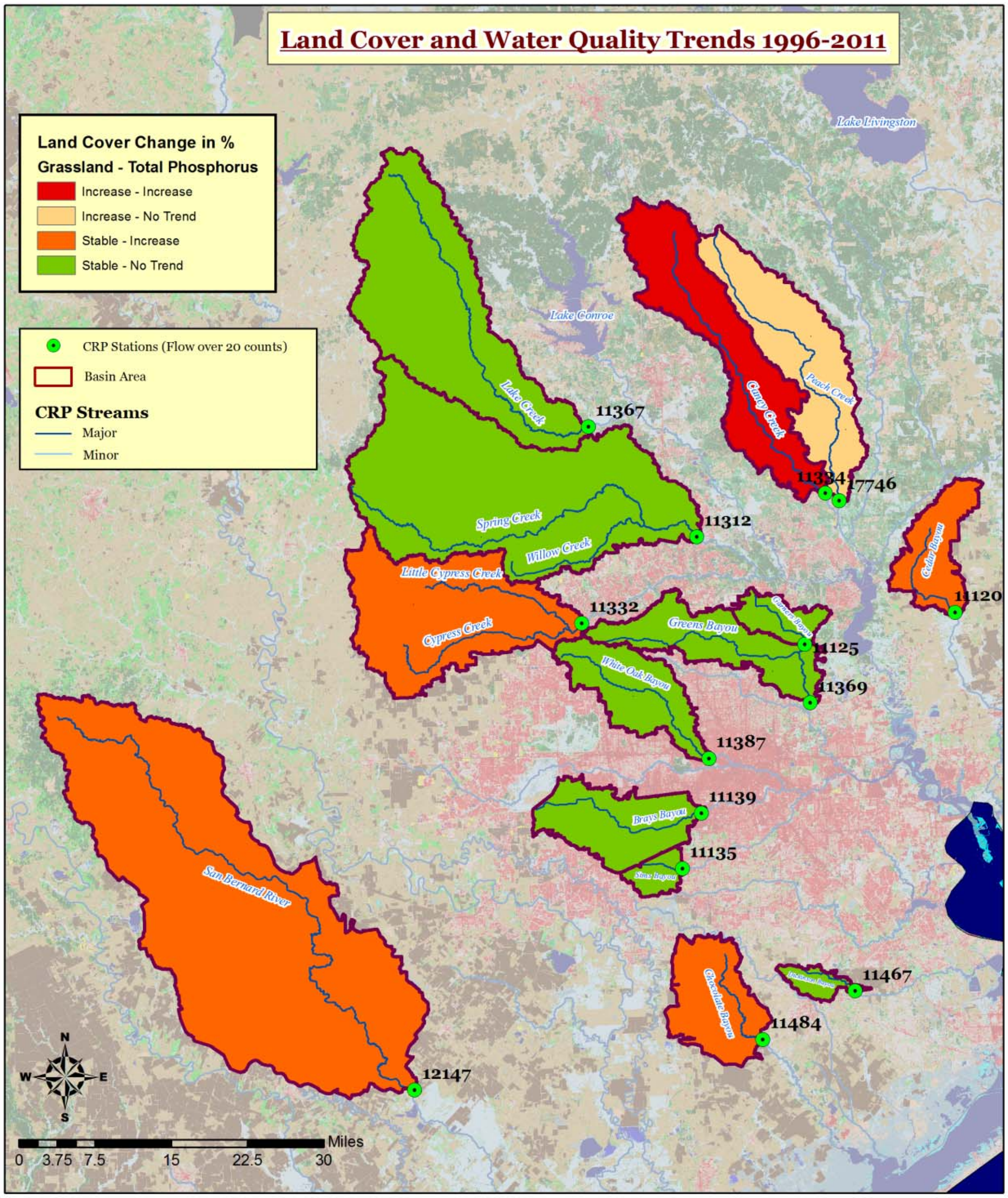

Figure 22: Land Cover and Water Quality Trends (Grassland and Total Phosphorus)

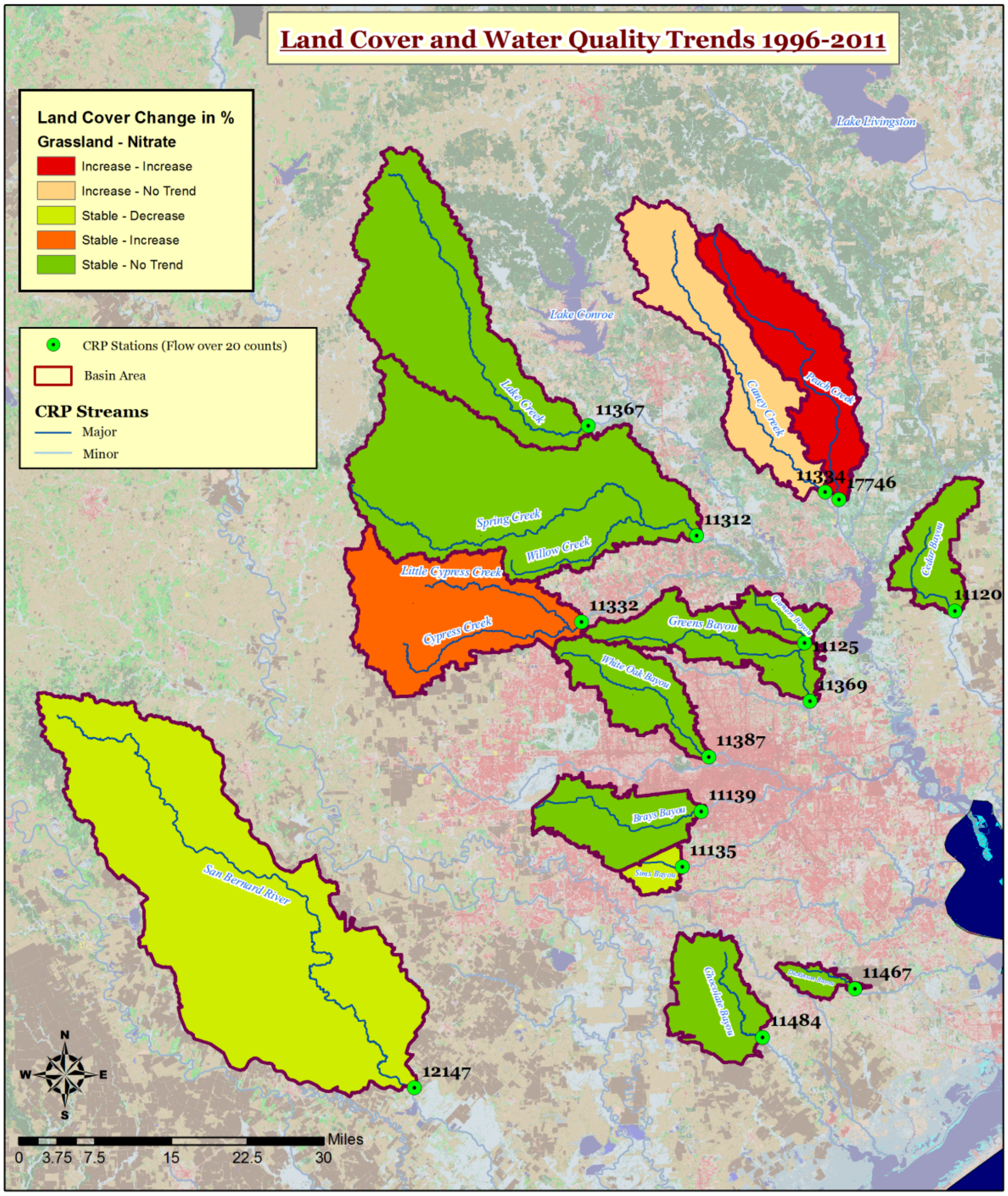

Figure 23: Land Cover and Water Quality Trends (Grassland and Nitrate)

4.3.5 Wetland Land Cover Trends

As shown in Figure 24 through Figure 26, all of the study watersheds had no trend in the percentage of wetlands over the study period. One of the watersheds had an increase in total nitrogen concentrations, two had an increase in nitrate concentrations, two had a decrease in nitrate concentrations, and five had an increase in total phosphorus concentrations.

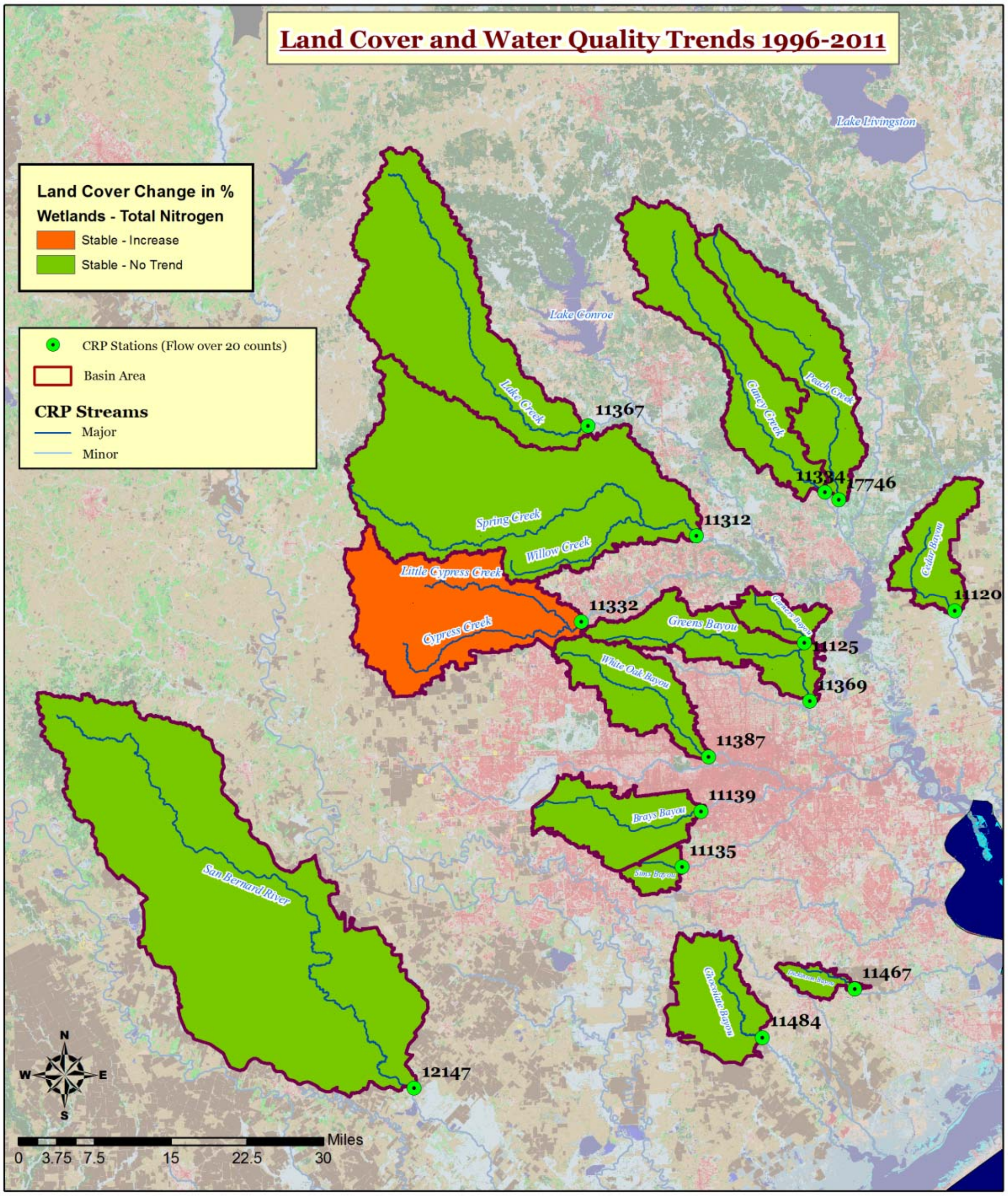

Figure 24: Land Cover and Water Quality Trends (Wetlands and Total Nitrogen)

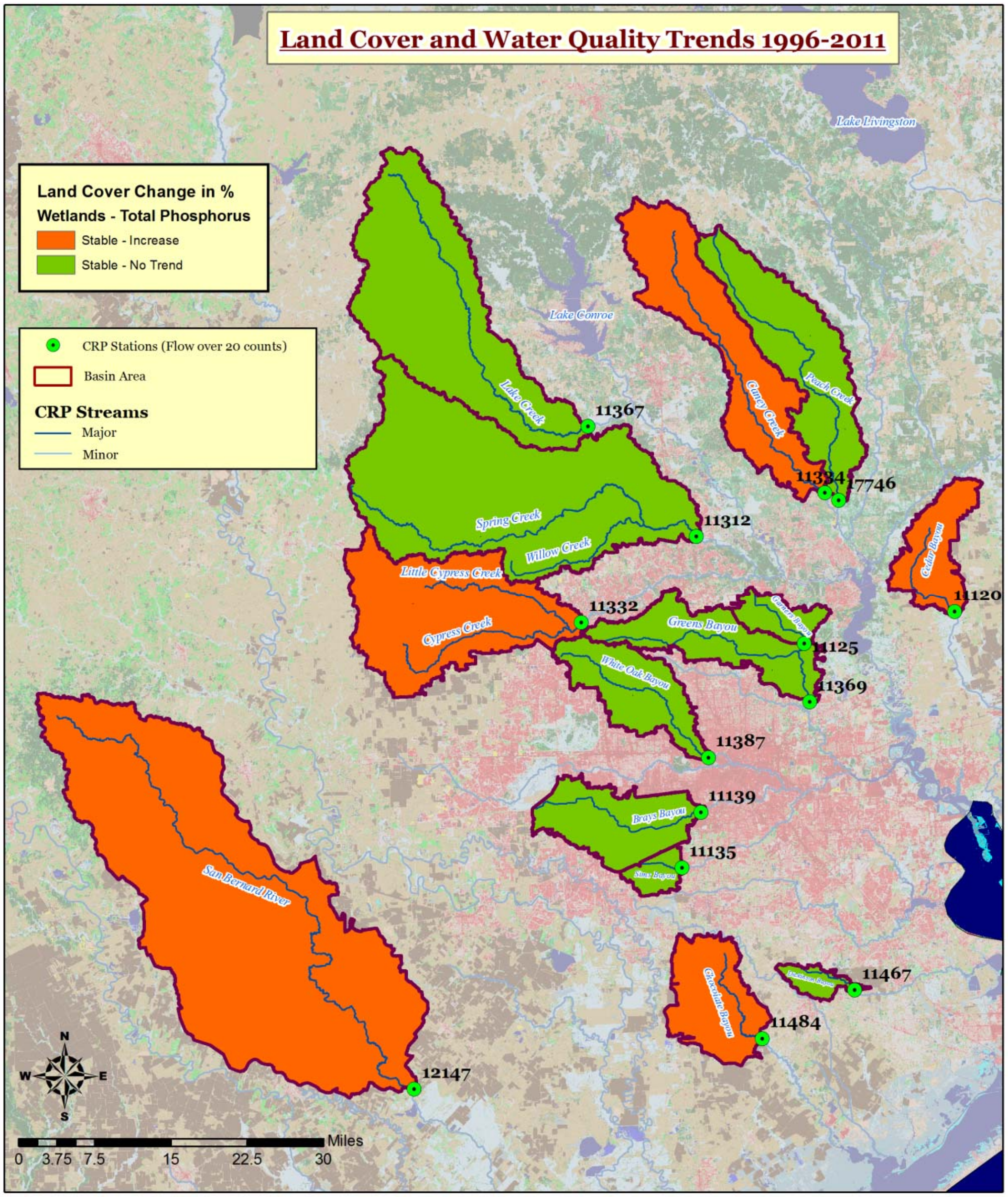

Figure 25: Land Cover and Water Quality Trends (Wetlands and Total Phosphorus)

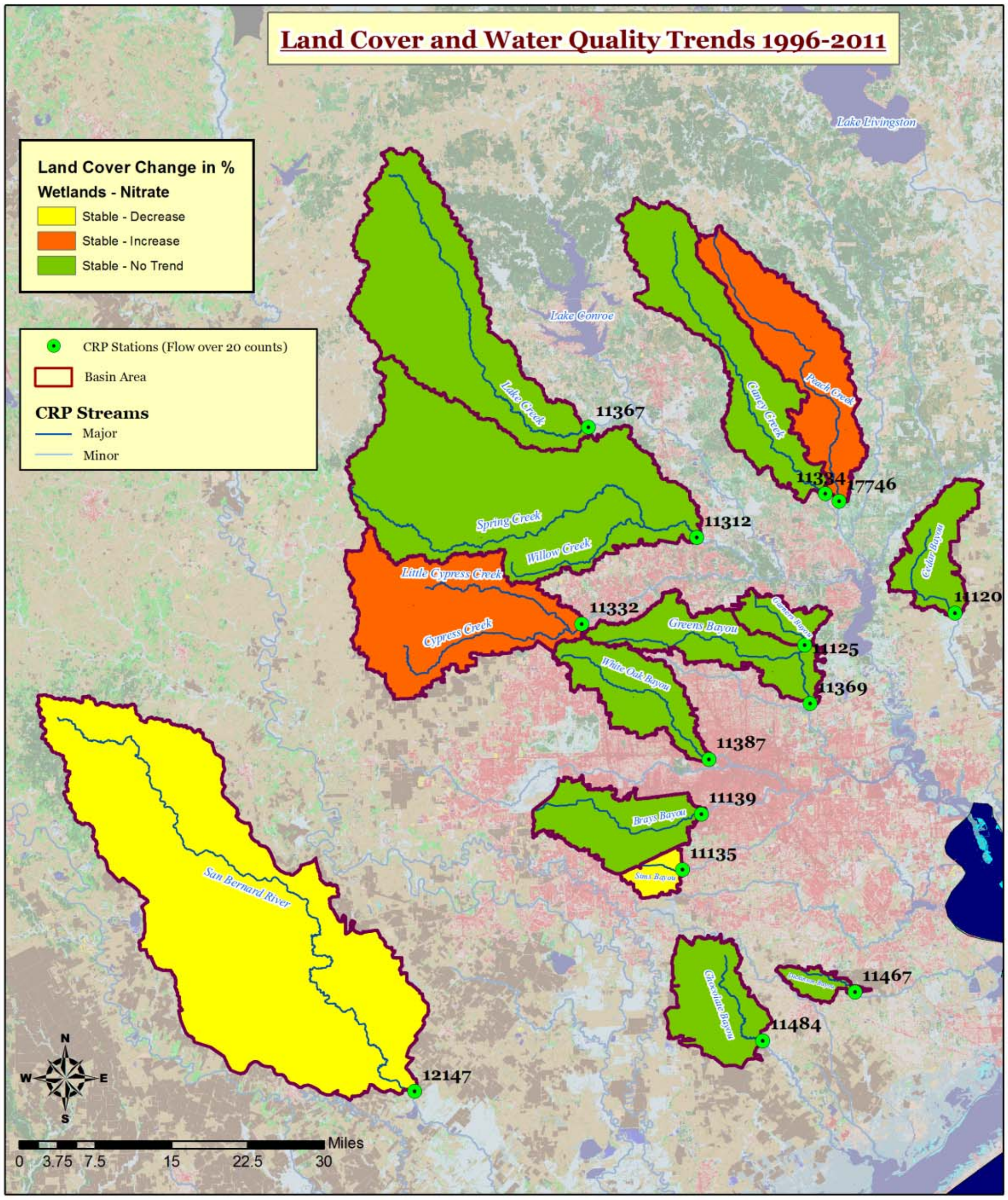

Figure 26: Land Cover and Water Quality Trends (Wetlands and Nitrate)

4.4 Regression Analysis

In this section and those that follow, associations and relationships between water quality and land cover variables are described in the form of statistical models. Do not assume that correlations between variables or the predictive accuracy of a cluster of variables imply causation. It must be borne in mind that identification of a statistically significant relationship between two variables does NOT establish a causal relationship. The ability to mathematically predict the value of one variable from another does not have any inherent bearing on causal relationships; predictive power is important for many reasons, but does not imply a causal relationship. Please note that it is common to find claims (in professional literature as well as this report) that variance of one set of variables “explains” a specific portion of variance in another variable, but this is not “explanation” in the sense of describing the reasons an event occurs. In this context, the proportion of variance “explained” is an indicator of the fit of a mathematical model to observed data and how well the dependent variable can be “predicted” from the independent variables.

H-GAC developed regression models relating total phosphorus and total nitrogen to a variety of land cover and other variables. Log-transformed concentrations and annual median concentrations were each regressed on the predictors. As discussed in Section 4.5.2, ANOVA results suggest that wastewater treatment plant effluent is a significant source of nutrients in area waterways; therefore, models were also developed from a dataset that excluded effluent-dominated streams. The results are summarized in Table 29 through Table 32.

Table 29: Regression Analysis – Total Phosphorus (Natural Logarithm of Concentration)

	All Stations		Effluent-Dominated Stations Excluded	
Model R^2	0.7156		0.2906	
Model Adjusted R^2	0.7035		0.2616	
Number of Observations	99		52	
	Estimate	**P-Value**	**Estimate**	**P-Value**
Intercept	0.3300	-	0.3757	-
Flow (base 10 log)	-0.1848	0.0004	-	-
Rainfall (previous day, natural log)	-0.0799	0.0086	-	-
Effluent Dominance	0.7625	<0.0001	-	-
Wetlands (natural log of percent)	0.0618	0.0491	-	-
Wetlands	-	-	-0.0186	<0.0001
Developed (percent)	-	-	-0.0088	0.0003

Table 30: Total Nitrogen- Regression of Concentration (log) on Land Cover

	All Stations		Effluent-Dominated Stations Excluded	
Model R^2	0.7878		0.4806	
Model Adjusted R^2	0.7782		0.4037	
Number of Observations	70		32	
	Estimate	**P-Value**	**Estimate**	**P-Value**
Intercept	1.3252	-	-0.6304	-
Rainfall (previous day)	-0.0913	0.0426	0.2610	0.0088
Effluent Dominance	0.9520	<0.0001	-	-
Grassland (natural log of percent)	-0.2990	0.0010	-	-
Grassland (percent)	-	-	0.0869	0.0290
Agricultural (percent)	-	-	0.0160	0.0128
Developed (percent)	-	-	-0.0117	0.0068

Table 31: Total Phosphorus – Regression of Annual Median on Land Cover and Flow

	All Stations		Effluent-Dominated Stations Excluded	
Model R^2	0.7160		0.6206	
Model Adjusted R^2	0.7007		0.5784	
Number of Observations	40		21	
	Estimate	**P-Value**	**Estimate**	**P-Value**
Intercept	-0.2445	-	0.3234	-
Effluent Dominance	1.46281	<0.0001		
Flow (annual median, base 10 log)	-0.0053	0.0073	-	-
Developed (natural log of percent)			-0.0621	<0.0001
Wetlands (percent)			-0.0106	0.0040

Table 32: Total Nitrogen – Regression of Annual Median on Land Cover and Median Flow

	All Stations		Effluent-Dominated Streams Excluded	
Model R^2	0.8071		0.7050	
Model Adjusted R^2	0.7969		0.6703	
Number of Observations	41		20	
	Estimate	**P-Value**	**Estimate**	**P-Value**
Intercept	0.7406	-	1.4329	-
Effluent Dominance	4.9401	<0.0001	-	-
Developed (percent)	0.0231	0.0301	-	-
Developed (natural log of percent)	-	-	-0.2530	<0.0001
Forest (percent)	-	-	-0.0057	0.0292

In addition, H-GAC used robust regression (in SAS, PROC ROBUSTREG was used) to analyze the data. Statistical models were developed using only variables identified as significant correlates of total phosphorus, total nitrogen, and nitrate. Robust regression is insensitive to departures from normality and extreme values (outliers) in the dataset. Because most analyses suggest that effluent dominance is the most significant predictor of nutrient concentrations overall, models were tested only for watersheds not classified as effluent-dominated to assess the role of land cover alone. The results are summarized in Table 33 below. The results of the regression suggest that none of the predictors influences the variance of the different concentrations.

Table 33: Robust Regression Results

Variable	Total Phosphorus		Total Nitrogen		Nitrate + nitrite	
Observations	52		30		32	
Model R^2	0.1994		0.1990		0.1163	
	Estimate	p-value	Estimate	p-value	Estimate	p-value
Agricultural	0.0047	0.0766	0.0063	0.0718	-0.0013	0.4407
Forest	0.0037	0.1235				
Wetlands	-0.0054	0.0807				
Grasslands	-0.0059	0.5335				
Flow			0.0007	0.0926	0.1671	0.0167

4.4.1 Discussion

Wastewater influence explains most of the variation in nutrient concentration when data from all watersheds are regressed on the significant predictors. In the case of total phosphorus concentration, 63 percent of the variance is explained by effluent dominance (an indicator variable with values of 0 or 1); only one land cover predictor was found to be significant, accounting for one percent of the variance. The other two predictors – flow and rainfall – are likely related to the impact of effluent discharge; the negative partial regression coefficients suggest that total phosphorus concentration is reduced at higher flow levels after significant rainfall. H-GAC staff has observed such a pattern in previous analyses. An example from station 11332 is presented in Figure 27.

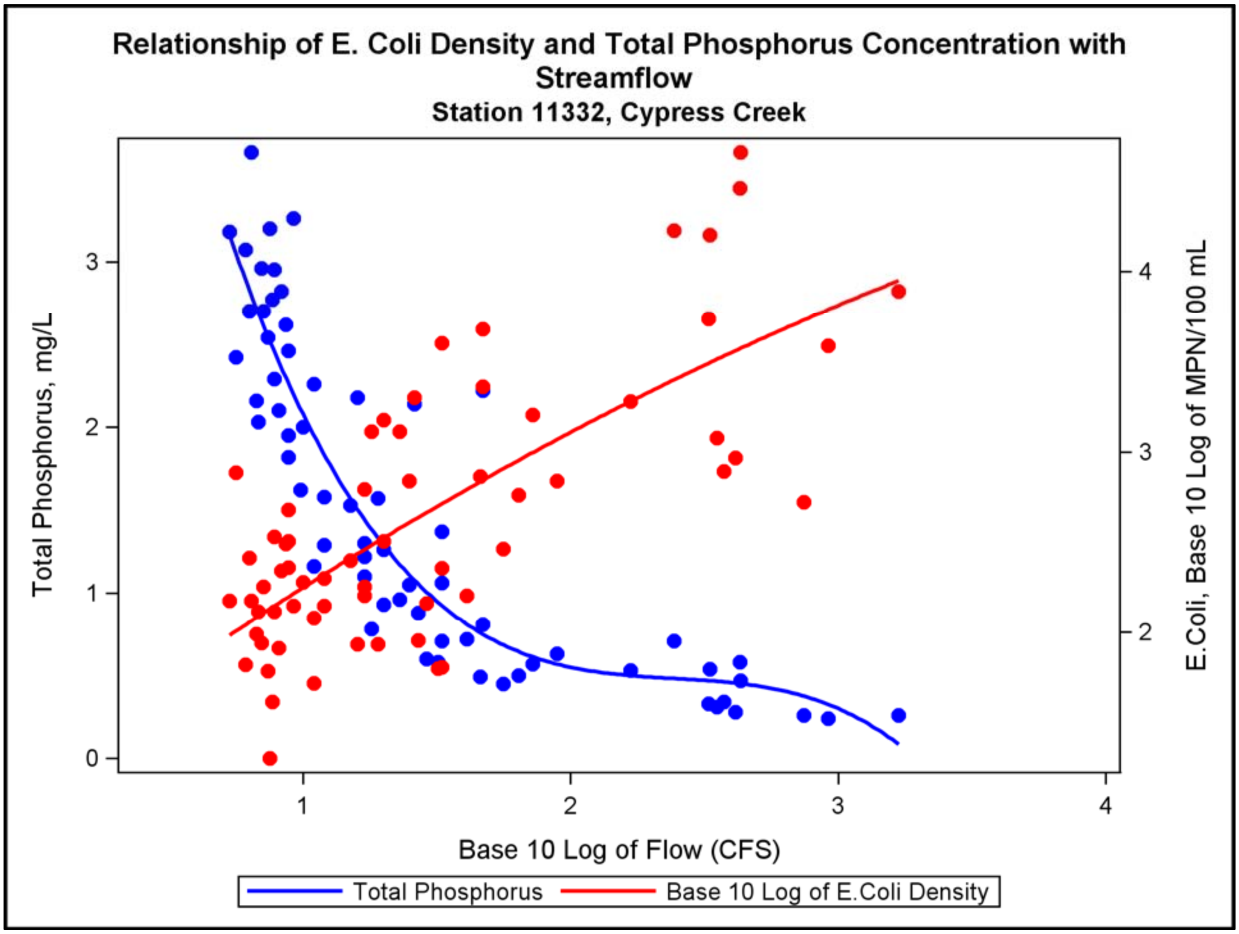

Figure 27: Relationship of *E. coli* Density and Total Phosphorus Concentration with Streamflow at Station 11332 (Source: H-GAC 2013 Basin Highlights Report)

A different relationship between flow and nutrient concentration is seen in streams that do not have a significant influx of domestic wastewater. This pattern was observed at station 11120 on Cedar Bayou as shown in Figure 28.

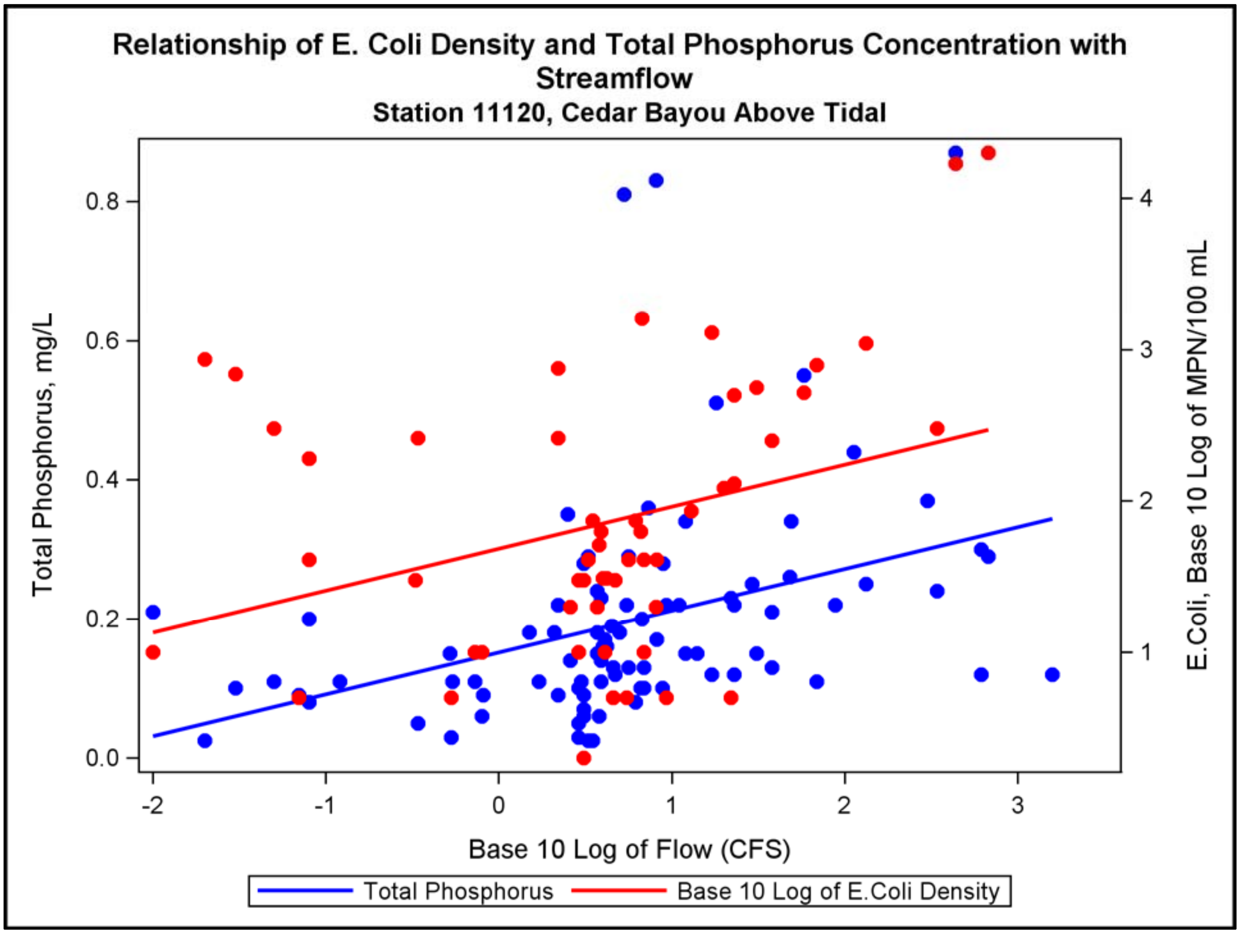

Figure 28: Relationship of E. coli Density and Total Phosphorus Concentration with Streamflow at Station 11120 (Source: H-GAC 2013 Basin Highlights Report)

The only significant land cover variable in the total nitrogen model for all watersheds is grassland coverage. Rainfall was also found to be important, reducing the concentration. The models for both total phosphorus and total nitrogen (natural logs) have a fairly high R^2, but examination of the plot of observed versus predicted concentration shows the bimodal distribution in the data noted previously; high values will inflate the R^2. As expected, when effluent dominated streams are removed from the models, the sample size is reduced, resulting in lower power to detect effects. Developed space is a significant predictor for both nutrient species, explaining 22 percent of total phosphorus and 12 percent of total nitrogen variance. In each case, concentration varies inversely with developed area. Agricultural coverage explains 15 percent of total nitrogen variance. The models as a whole explain 26 and 40 percent of the total variance respectively. Low power may account the low proportion of variance explained, particularly for total nitrogen.

Analysis of annual medians has two advantages over individual concentrations: seasonal variation is taken out of the picture, and the effect of extreme values is reduced. The main disadvantage is a reduction in statistical power. In each case, the variance explained by the models when effluent dominated watersheds are excluded is much higher than when individual concentration is considered. Unsurprisingly, effluent dominance explains most of the variance in total phosphorus (65 percent) and total nitrogen (78 percent) medians; as in the case of concentrations, streamflow is also significant (7 percent of the variance). Developed area is a significant predictor when effluent dominated streams are excluded, and also for total nitrogen in all watersheds. It explains little of the variance for all watersheds, but accounts for 39 percent of total phosphorus and 61 percent of total nitrogen variance in streams that are not effluent-dominated. However, the direction of influence differs: developed area is positively correlated with total nitrogen when all watersheds are considered. Because the partial r-square is less than 0.03, it might be a statistical artifact introduced by the selection process, and may merely serve to "adjust" the predicted value in a manner that minimizes model error.

As noted earlier, the best regression models will satisfy a number of assumptions and requirements. The total phosphorus and total nitrogen (concentration) models for all watersheds show unacceptable residual variance, but all other diagnostics are satisfactory. The total phosphorus model that excludes effluent-dominated watersheds has several problems. Residual variance is acceptable, but the residuals are not normal and there is evidence of positive autocorrelation. In a larger sample size that would not be too troublesome, but in this case it suggests model misspecification. The results should be interpreted with caution, and parameter estimates may be unreliable. No problems were found in the total nitrogen model. Both total phosphorus median models and the total nitrogen model for all watersheds failed the residual normality criterion. Plus, the small sample size compounds the problem. The total nitrogen model that excludes effluent-dominated watersheds did not fail the key diagnostic criteria and is the most reliable model of the four.

4.5 Correlation Analysis

H-GAC evaluated the correlation between the nutrient species (total phosphorus, total nitrogen, nitrogen as nitrate, and nitrogen as ammonia) and watershed characteristics. The correlation analysis was part of an initial exploratory analysis. The statistically significant correlations are included in Appendix A.

Table 34 and Table 35 show the correlation matrices for the key variables analyzed. The tables show correlations and statistical results between nutrients (total phosphorus concentration, total nitrogen concentration, nitrate and nitrite concentration) and watershed characteristics (land cover types and categories, flow, and previous day's rainfall for all stations). Table 35 shows the results for watersheds where there are relatively low levels of flow from WWTP dominant land use and effluent domination status.

4.5.1 Kendall Rank Correlation Coefficient

As discussed in Section 3.4.3, the Kendall Tau B is a rank correlation coefficient based on agreement between the ranked (ordered) values of observational pairs. The absolute value of the coefficient is a measure of the strength of the association, and the sign indicates the direction (direct or inverse correlation).

The results of nonparameteric correlation analysis (Kendall's Tau A) are presented below in Table 34 through Table 38. It must be noted that the p-value is the probability that the observed results would be obtained from a sample of a population in which the variables are uncorrelated. As the sample size increases, the p-value will fall simply because the confidence level is function of sample size, and the odds that the correlation in the sample reflects the correlation in the actual population increase. A correlation that is statistically significant but with a small coefficient (for example, less than 0.2) has little practical or explanatory significance.

Table 34: Correlation Results for all Stations

Variable	Total Phosphorus		Total Nitrogen		Nitrate + nitrite	
	(N = 165)		(N=70)		(N=181)	
	Correlation	p-value	Correlation	p-value	Correlation	p-value
Effluent Domination	0.64945	<.0001	0.6956	<.0001	0.6782	<.0001
Developed	0.3111	<.0001	0.4898	<.0001	0.4034	<.0001
Agricultural	-0.2529	<.0001	-0.4404	<.0001	-0.4051	<.0001
Forest	-0.0502	0.3475	-0.0757	0.3584	-0.1118	0.0277
Wetlands	-0.2243	<.0001	-0.3902	<.0001	-0.2468	<.0001
Grasslands	-0.2951	<.0001	-0.4162	<.0001	-0.3906	<.0001
Flow	-0.0237 (n=99)	0.7303	0.2142 (n=60)	0.0159	0.1385 (n=99)	0.0434
Rainfall (previous day)	-0.1882	0.0021	-0.0846	0.3647	-0.1480 (n=174)	0.0132

Table 35: Correlation Results for Stations without Significant Amount of WWTP Effluent

Variable	Total Phosphorus (N=52)		Total Nitrogen (N=32)		Nitrate + nitrite (N=32)	
	Correlation	p-value	Correlation	p-value	Correlation	p-value
Developed	-0.1523	0.1295	-0.2418	0.0551	-0.2891	0.0468
Agricultural	0.3752	0.0002	0.2459	0.0511	-0.0632	0.4595
Forest	-0.3095	0.0021	-0.1188	0.3458	0.1361	0.1109
Wetlands	-0.3287	0.0011	-0.0450	0.7206	0.2536	0.7150
Grasslands	-0.3399	0.0007	-0.1844	0.1435	0.0222	0.6200
Flow	0.2293	0.0609	0.2240	0.0833	0.2168 (n=30)	0.0833
Rainfall (previous day)	0.1422	0.1999	0.2686	0.0544	0.26857	0.0544

Table 36: Correlations by Dominant Land Use - Agriculture

Kendall Tau b Correlation Coefficients
Prob > |tau| under H0: Tau=0
Number of Observations

	Total Phosphorus	**Nitrate + Nitrite**	**Total Nitrogen**
Effluent Dominance	0.73131 <.0001 35	0.68667 <.0001 42	0.73030 0.0105 10
Percent Developed	0.68904 <.0001 35	0.63341 <.0001 42	0.78113 0.0029 10
Percent Agricultural	0.44921 0.0005 35	0.37297 0.0011 42	0.47875 0.0679 10
Percent Forest	-0.39211 0.0025 35	-0.10747 0.3473 42	-0.52915 0.0436 10
Wetlands	-0.70427 <.0001 35	-0.66628 <.0001 42	-0.78113 0.0029 10
Percent Grassland	-0.59387 <.0001 35	-0.55755 <.0001 42	-0.32757 0.2117 10
Flow, Base 10 Log	0.08989 0.7194 10	0.11111 0.6767 9	0.11111 0.6767 9
Rainfall, Previous Day	-0.10605 0.4468 35	-0.09269 0.4553 42	-0.10847 0.6977 10

Table 37: Correlations by Dominant Land Use - Forest

Kendall Tau b Correlation Coefficients
Prob > |tau| under H0: Tau=0
Number of Observations

	Total Phosphorus	Nitrate + Nitrite	Total Nitrogen
Effluent Dominance	0.73131 <.0001 35	0.68667 <.0001 42	0.73030 0.0105 10
Percent Developed	0.68904 <.0001 35	0.63341 <.0001 42	0.78113 0.0029 10
Percent Agricultural	0.44921 0.0005 35	0.37297 0.0011 42	0.47875 0.0679 10
Percent Forest	-0.39211 0.0025 35	-0.10747 0.3473 42	-0.52915 0.0436 10
Wetlands	-0.70427 <.0001 35	-0.66628 <.0001 42	-0.78113 0.0029 10
Percent Grassland	-0.59387 <.0001 35	-0.55755 <.0001 42	-0.32757 0.2117 10
Flow, Base 10 Log	0.08989 0.7194 10	0.11111 0.6767 9	0.11111 0.6767 9
Rainfall, Previous Day	-0.10605 0.4468 35	-0.09269 0.4553 42	-0.10847 0.6977 10

Table 38: Correlations by Dominant Land Use - Urban

Kendall Tau b Correlation Coefficients
Prob > |tau| under H0: Tau=0
Number of Observations

	Total Phosphorus	Nitrate + Nitrite	Total Nitrogen
Effluent Dominance	. . 81	. . 83	. . 27
Percent Developed	-0.37194 <.0001 81	-0.03719 0.6299 83	0.17576 0.2085 27
Percent Agricultural	0.13791 0.0787 81	-0.12011 0.1196 83	-0.12303 0.3787 27
Percent Forest	0.40602 <.0001 81	0.09328 0.2268 83	-0.13475 0.3349 27
Wetlands	0.29993 0.0001 81	0.01097 0.8869 83	-0.25193 0.0714 27
Percent Grassland	0.25043 0.0014 81	-0.03048 0.6928 83	-0.25779 0.0651 27
Flow, Base 10 Log	-0.47782 <.0001 46	-0.05352 0.5936 48	-0.05251 0.7396 21
Rainfall, Previous Day	-0.20533 0.0209 81	-0.14624 0.0953 83	-0.30675 0.0500 27

4.5.2 Discussion

It is important to note that individual correlations may suggest candidate predictors for more sophisticated models, but are of limited value when considered alone. Regression models can assess the significance of a variable when other variables are included, but correlation cannot reveal the inter-relationships and interactions that exist in the real world, and can seldom be taken at face value as an explanation. For example, there is a high correlation between effluent dominance and developed area. Both are highly correlated with total phosphorus. Both could be incorrectly claimed to "cause" high nutrient levels. Both could be considered "confounding variables" that mask or distort the influence of other independent variables.

Examination of the correlation tables above shows that effluent dominance is significantly correlated with total phosphorus in forest and agricultural areas. This correlation results from the existence of one effluent-dominated stream in each watershed. The nutrient concentration in those streams is far higher than the other streams (see the discussion of watershed classification, Section 4.1).

Examination of the correlations for all stations reveals a positive correlation between all of the nutrient types and the percentage of developed land in the watershed. This probably results from the fact that all but two effluent-dominated streams are in highly developed urban areas, and there is significant developed space in those two. There are negative correlation between nutrient concentrations and the percentage of agricultural, forest, grassland, and wetland land cover types. When effluent-dominated streams are excluded, the pattern changes significantly. Nutrients are positively correlated with development and agriculture, while nutrient concentrations are inversely related to the percentage of forest, wetlands, and grasslands. Again, the negative correlation is probably due to the absence of effluent domination, which is highly correlated with developed area. The other analyses performed by H-GAC suggest that these associations are excellent examples of the effect of confounding variables.

Table 35 shows results for stations that do not have significant upstream sources of WWTP effluent. The table shows several differences from Table 34. In contrast to Table 34, these stations show a negative correlation between development in watersheds and nutrient concentrations downstream. In addition, the table shows positive correlations between agriculture and nutrients. As in the analysis that included all of the monitoring sites, the percentage of forest, wetlands, and grasslands land use area is negatively correlated with nutrients. Again, the correlation between developed area and effluent dominance is masked, distorting the meaning and significance of the correlations. The relationship between land cover and nutrient concentration cannot be inferred from correlation analysis alone.

4.6 Analysis of Variance (ANOVA)

As discussed in the methodology section above, all types of ANOVA involve comparisons of the mean (or median) levels of a continuous outcome variable across categories or groups. H-GAC performed parametric ANOVA (SAS PROC GLM) and nonparametric (SAS PROC NPAR1WAY) to test the null hypothesis that there was no difference in the means or median total phosphorus and total nitrogen concentration across the following groups:

- WWTP Effluent domination (Yes/No)
- Wastewater discharge level (Low/Medium/High)
- Watershed type (urban/non-urban).

In addition, the Brown-Forsythe test of homogeneity of variance was performed and the Welch statistic calculated. If the variance between groups is not homogenous, the Welch statistic is the more reliable indicator of the statistical significance of mean differences. Nonparametric ANOVA is based on a rank-sum test, and a Wilcoxon or Kruskall-Wallis score is calculated. As in the case of parametric ANOVA, the means or medians of total phosphorus and total nitrogen concentrations were compared across levels of the same three categorical variables.

Statistically significant differences between total phosphorus and total nitrogen means were found for different levels of effluent domination and watershed type by all methods. Mean total phosphorus was found to be significantly different across levels of wastewater discharge by all methods. Total nitrogen was also found to differ between levels of wastewater discharge by the Welch test. Results are summarized in Table 39. Box plots from nonparametric ANOVA are presented in Figure 29 through Figure 35.

Heterogeneous variances were detected in the watershed discharge level and effluent dominance groups for both nutrients. The Welch test, which adjusts for heterogeneity, showed that all differences between means were statistically significant, but there is no Welch equivalent for nonparametric analysis of variance. However, the results of the Welch test support the findings of nonparameteric ANOVA. The similar variance in the two “watershed type” groups, not seen in the groups within the other two categories, could be due to the exclusion of Cypress Creek and Spring Creek from the “urban” category. The addition of the high nutrient concentrations observed in samples from these streams to the “nonurban” group increases the variance in the group, reducing the difference in variance between these two groups.

Table 39: Analysis of Variance/ General Linear Model Results

Parameter	Comparison	N	Kruskal-Wallis p-value	GLM/ ANOVA p-value (model)	Model R^2	Statistically Significant Differences (GLM/ Welch)	Brown-Forsythe HOV Test	Welch p-value
Total Phosphorus	Effluent Dominated (Y/N)	Y: 113 N: 52	<0.0001	<0.0001	0.6807	<0.0001	<0.0001	<0.0001
	Wastewater Discharge in Watershed (H/M/L)	H:71 M: 56 L: 38	<0.0001			<0.0001 (H-L)	<0.0001	<0.0001
	Watershed Type (urban/nonurban)	U:81 NU: 84	<0.0001				0.1097	<0.0001
Total Nitrogen	Effluent Dominated (Y/N)	Y: 38 N: 32	<0.0001	<0.0001	0.7770	0.0001	0.0031	<0.0001
	Wastewater Discharge in Watershed (H/M/L)	H: 27 M: 25 L: 18	<0.0001				0.0179	<0.0001
	Watershed Type (urban/nonurban	U: 27 NU: 43	<0.0001			0.0004	0.4058	<0.0001

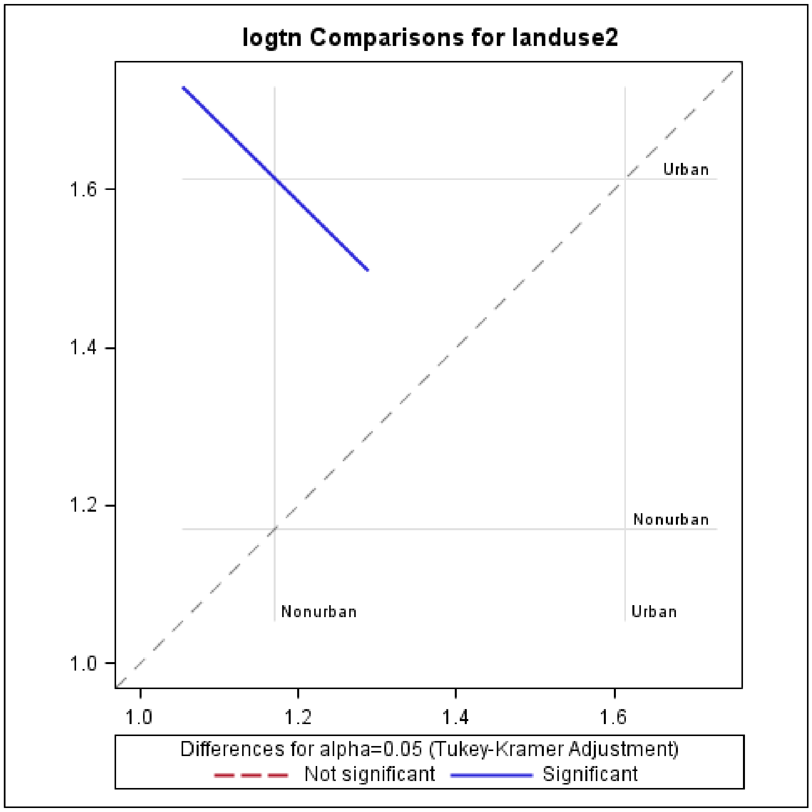

Figure 29: Comparison of Total Nitrogen to Land Cover

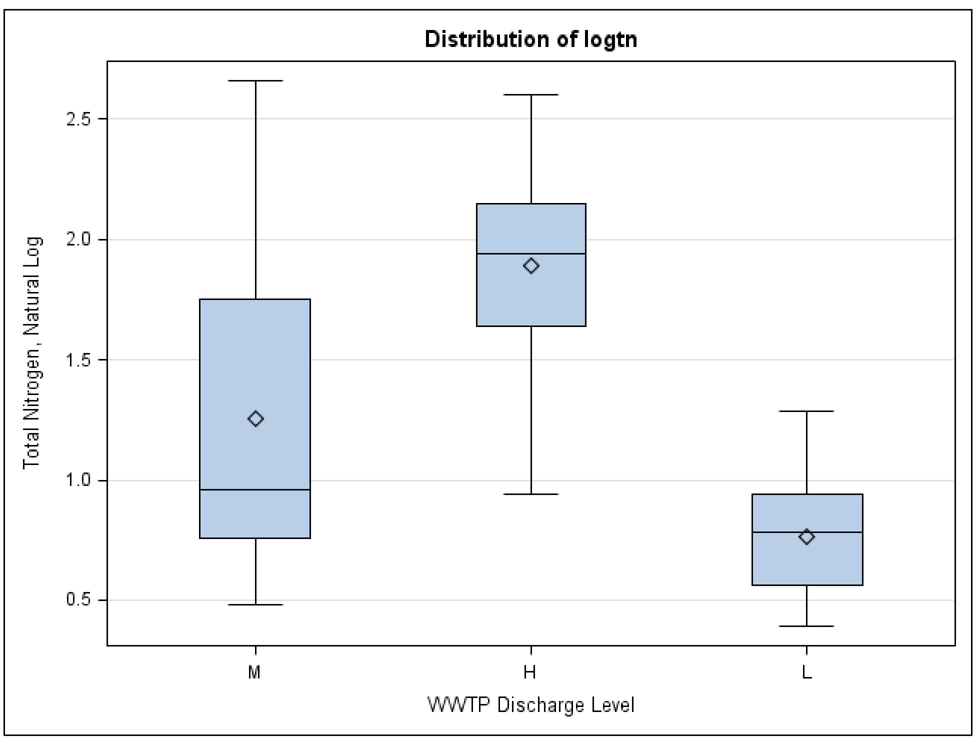

Figure 30: Distribution of Total Nitrogen for WWTP Discharge Levels

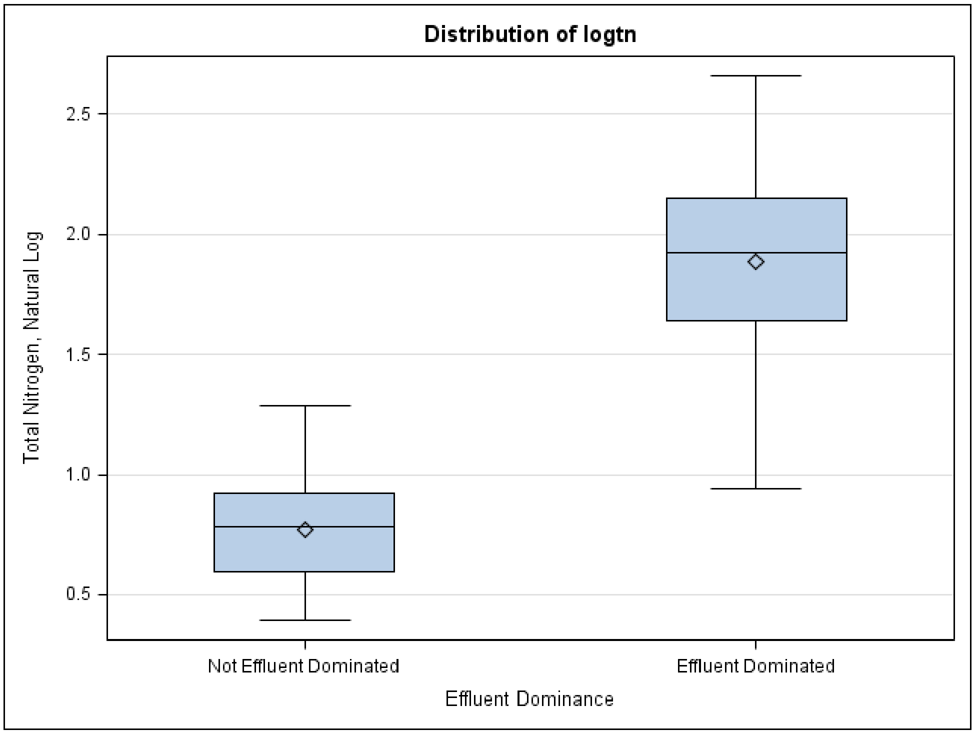

Figure 31: Distribution of Total Nitrogen for Ratio of Effluent Levels

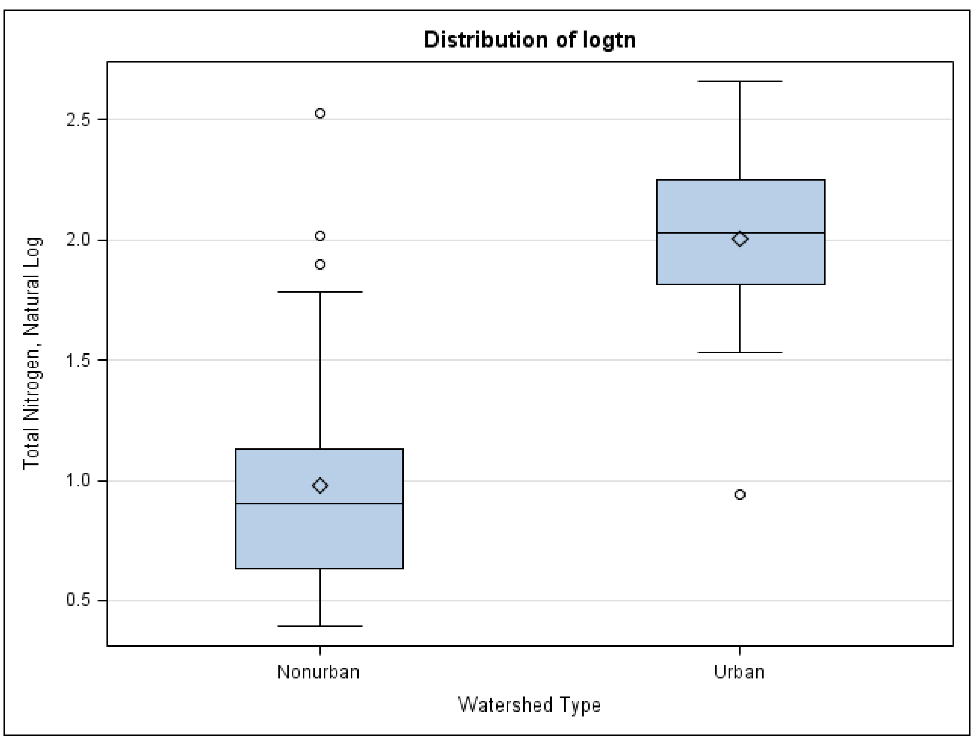

Figure 32: Distribution of Total Nitrogen by Watershed Type

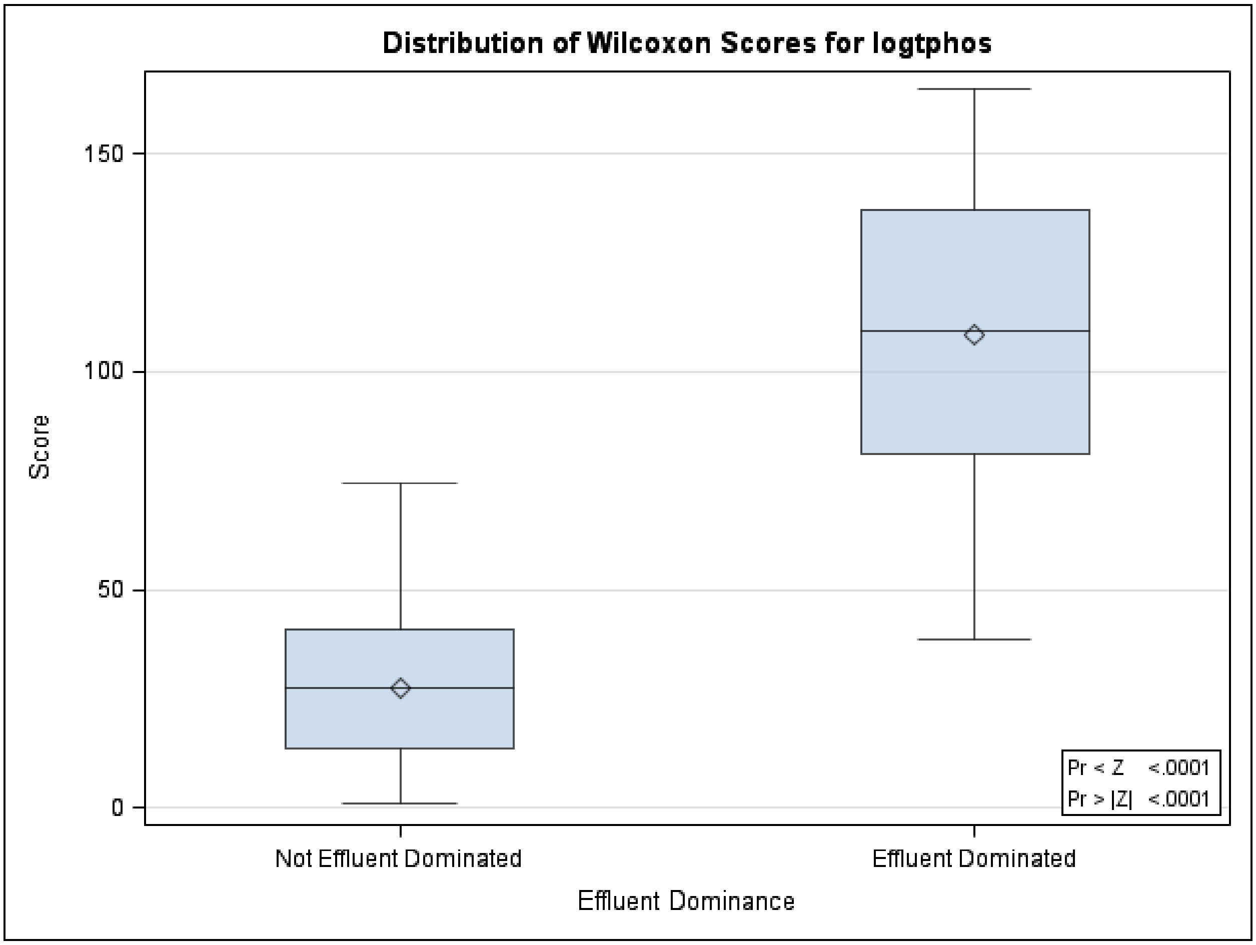

Figure 33: Distribution of Wilcoxon Scores for Total Phosphorus by Effluent Dominance

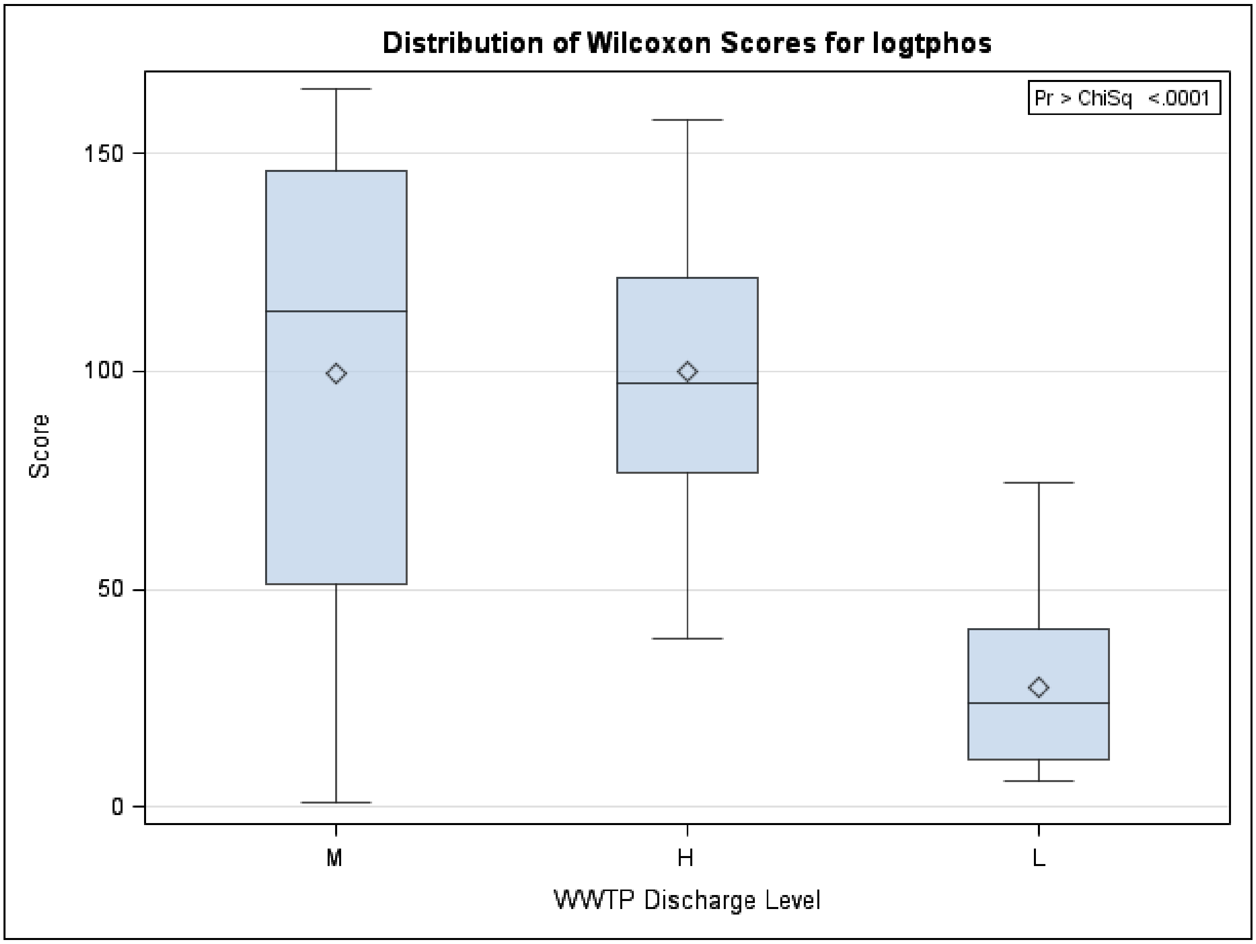

Figure 34: Distribution of Wilcoxon Scores for Total Phosphorus by WWTP Discharge Level

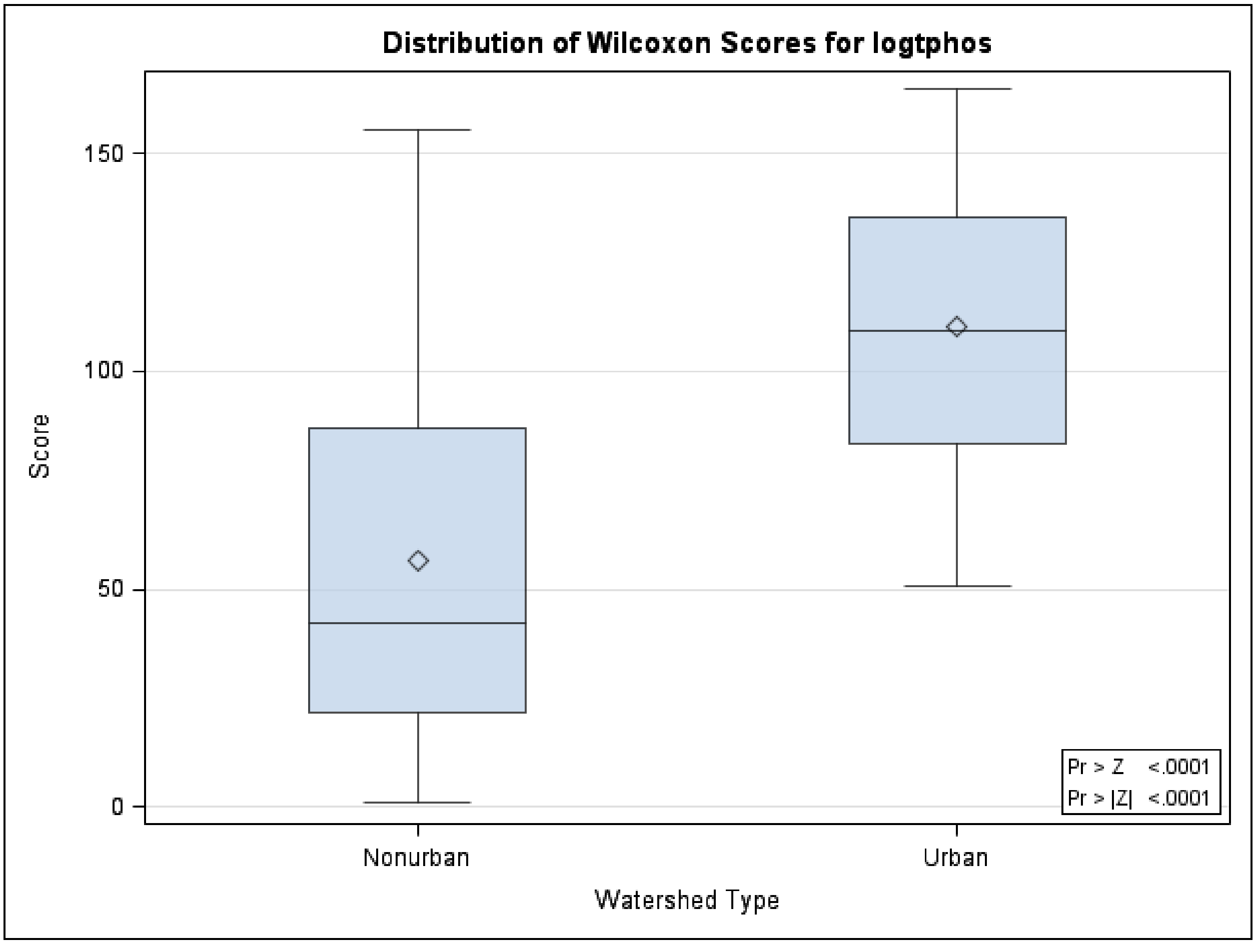

Figure 35: Distribution of Wilcoxon Scores for Total Phosphorus by Watershed Type

4.7 Analysis of Covariance (ANCOVA)

As described in Section 3.4.5, ANCOVA is a statistical analysis which has features of both regression and ANOVA. ANCOVA models were developed by H-GAC using SAS PROC GLM. These are similar to ANOVA insofar as the categorical (group) predictors may be used to explain nutrient concentration. They are similar to regression analysis in that continuous variables ("covariates") such as percent land cover, rainfall, and flow may be included in the model to account for covariance between the mean of the dependent variable in a given group and the value of the continuous variable. In other words, ANCOVA is useful when the influence of the group variable depends upon the level of a covariate. H-GAC hoped to develop and evaluate predictive models using ANCOVA that would allow prediction of nutrient concentrations from land cover variables. This was not feasible given the small sample size from non-effluent dominated watersheds.

The ANCOVA model for total phosphorus (natural log) concentration at all stations has an R^2 of 0.7657. Residuals are normally distributed, the variance of the residuals is acceptable, and there is no pattern suggesting serial correlation. Three of the predictors also appear in the regression model (effluent dominance, wetland coverage, and flow). Effluent dominance explains more variance in the total phosphorus concentration than flow and the land cover variables combined, while wastewater discharge level explains more than any single land cover variable.

When effluent-dominated watersheds are excluded, only wastewater discharge level, forest coverage, and wetlands are significant predictors and explain about 36 percent of the total variance (R^2 = 0.3632). The residuals are not normally distributed, suggesting poor model fit. Only wetlands and wastewater discharge level are significant in both models. Forest area coverage explains more variance than wastewater discharge level or wetlands. As noted earlier, one watershed in which forest is the dominant land cover type (Spring Creek) is also effluent-dominated and nutrient concentrations are fairly high. This would obscure the influence of forest land cover in a model that included all watersheds.

The models developed for total nitrogen showed better fit to the data (normally distributed residuals, no problem with residual variance or serial correlation) and included some of the same predictors as the total phosphorus models. When all watersheds are included, the model has an R^2 of 0.7019. Forest, grasslands, developed area, and wastewater discharge level were found to be significant predictors of total nitrogen concentration. More variance is explained by developed area than the other predictors. As noted earlier, developed area is highly correlated with effluent dominance and a model that included that variable rather than developed area would explain almost as much variance. It should be noted that because aquatic organisms (particularly cyanobacteria) can fix nitrogen from the atmosphere, nitrogen levels can vary independently of point or nonpoint source contributions, and different relationships between wastewater and runoff contributions are possible for nitrogen and phosphorus.

The model excluding effluent-dominated watersheds has an R^2 of 0.4469 and shares several significant predictors with the total phosphorus model. Forest coverage, wetlands, and wastewater discharge level appear in both models and the total nitrogen model includes the previous day's rainfall. As before, forest coverage explains more variance than the other predictors although wetland coverage explains almost as much. Rainfall explains the least but is fairly important when the effect of other variables is removed. A summary of results is provided in Table 40 and Table 41. Diagnostic plots are found in Figure 36 through Figure 43. See attached Appendix A for complete set of results and figures created when running the ANCOVA analysis.

Table 40: General Linear Model – Total Phosphorus

	All Stations		Effluent-Dominated Stations Excluded	
Model R^2	0.765736		0.3632	
Model Adjusted R^2	-		-	
Number of Observations	99		52	
	Estimate	**P-Value**	**Estimate**	**P-Value**
Intercept	1.14419	-	0.50102	-
Effluent Dominance	-	<0.0001		
Not Effluent Dominated	-0.77482	<0.0001		
Effluent Dominated	0.0000	-		
Wastewater Discharge Level		<0.0001		0.0002
High	-0.52190	0.0019		
Medium	-0.29856	<0.0001	-0.21308	0.0002
Low	0.0000	-	0.0000	-
Flow (base 10 log)	-0.18124	0.0009		
Developed (natural log of percent)	0.06519	0.0488		
Grasslands (natural log of percent)	-0.35971	<0.0001		
Wetlands (natural log of percent)	0.21149	0.0007	0.08429	0.0088
Forest (natural log of percent)			-0.1662	<0.0001

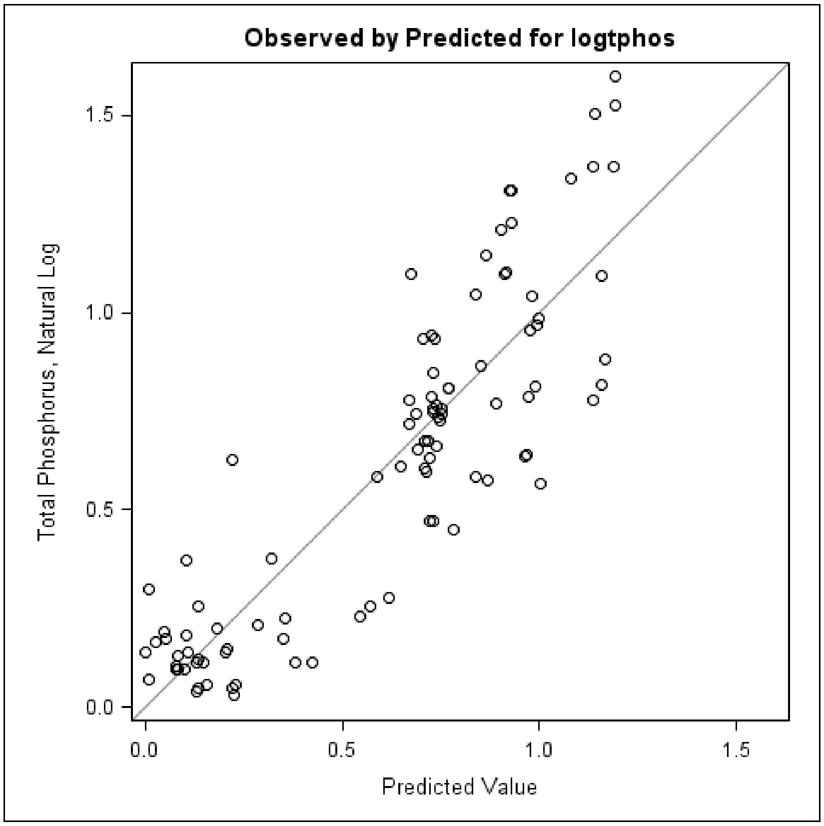

Figure 36: Observed and Predicted Values (Total Phosphorus at all Stations)

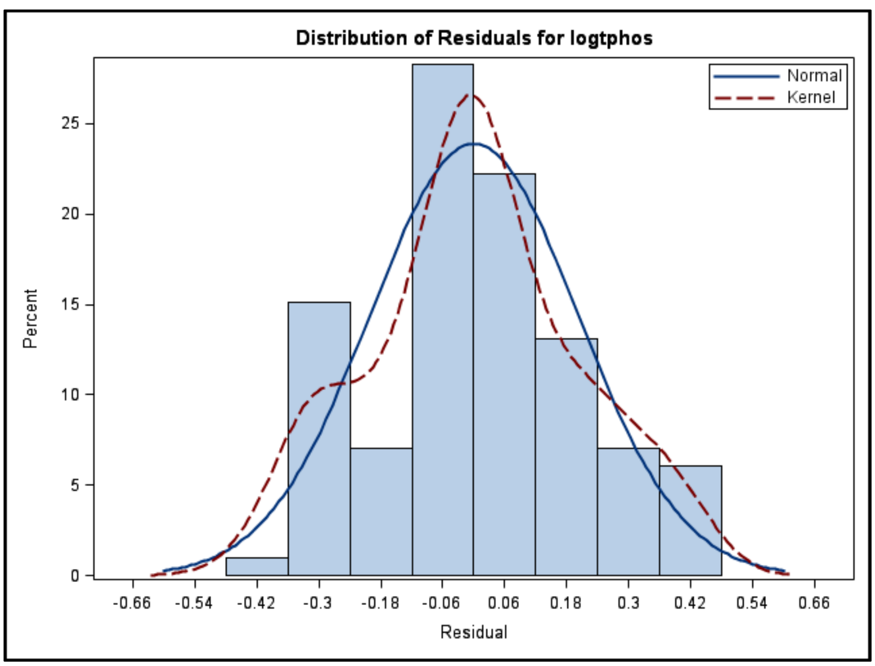

Figure 37: Distribution of Residuals (Total Phosphorus at all Stations)

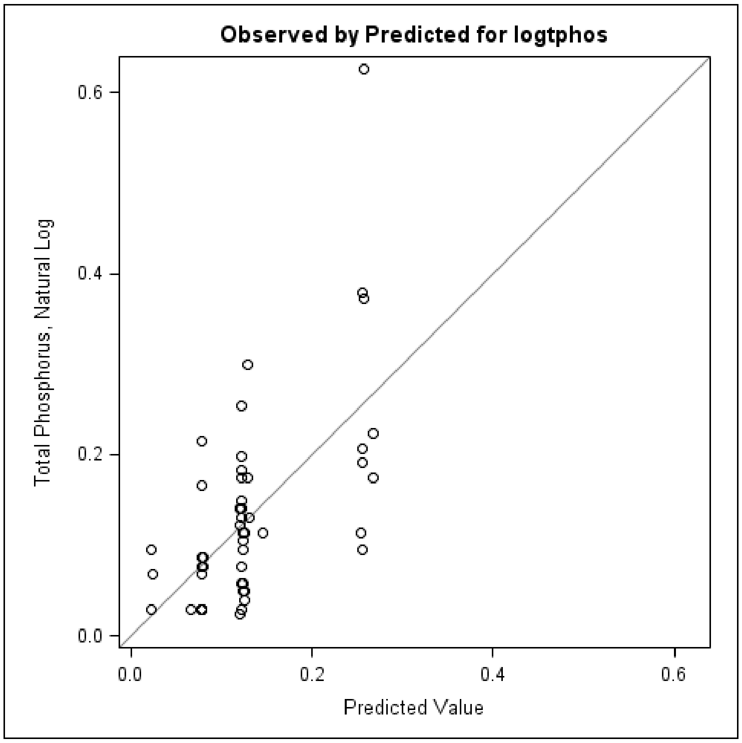

Figure 38: Observed and Predicted Values (Total Phosphorus with Effluent Dominated Streams Excluded)

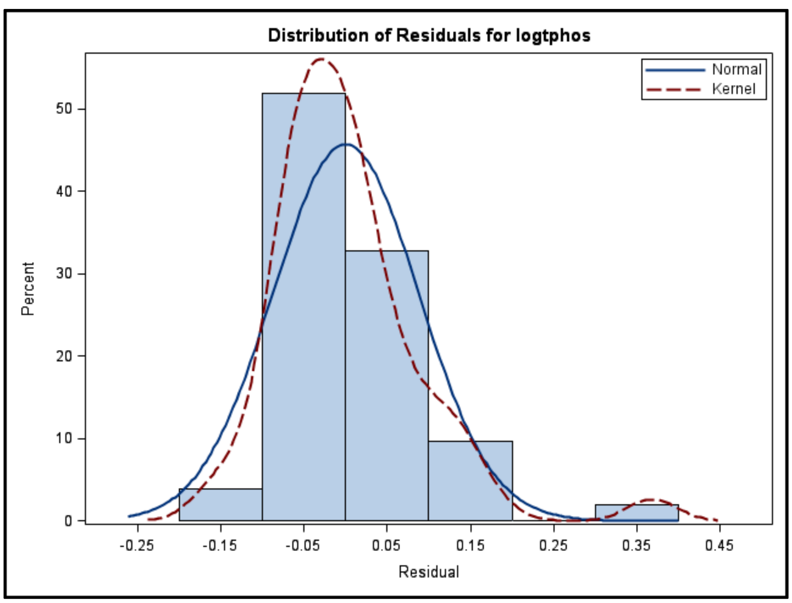

Figure 39: Distribution of Residuals (Total Phosphorus with Effluent Dominated Streams Excluded)

Table 41: General Linear Model Results – Total Nitrogen

	All Stations		Effluent-Dominated Stations Excluded	
Model R^2	0.7019		0.4469	
Model Adjusted R^2	-		-	
Number of Observations	70		32	
	Estimate	**P-Value**	**Estimate**	**P-Value**
Intercept	1.29458		2.4004	-
Rainfall (previous day)			0.27075	0.0072
Wastewater Discharge (level)	-	0.0505	-	0.0039
High	0.19608	0.2788		
Medium	0.31546	0.0156	-0.31996	0.0039
Low	0.00000	-	0.00000	-
Developed (percent coverage)	0.01037	<0.0001		
Forest (percent coverage)	0.01384	0.0005	-0.00968	0.0010
Grasslands (natural log of percent)	-0.48689	<0.0001		
Wetlands (natural log of percent)			0.30750	0.0012

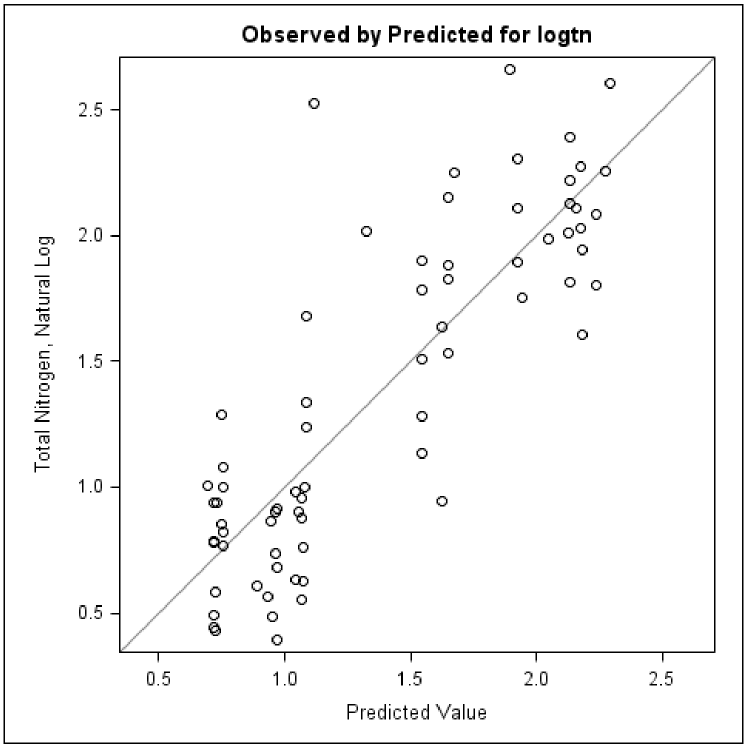

Figure 40: Observed and Predicted Values (Total Nitrogen at all Stations)

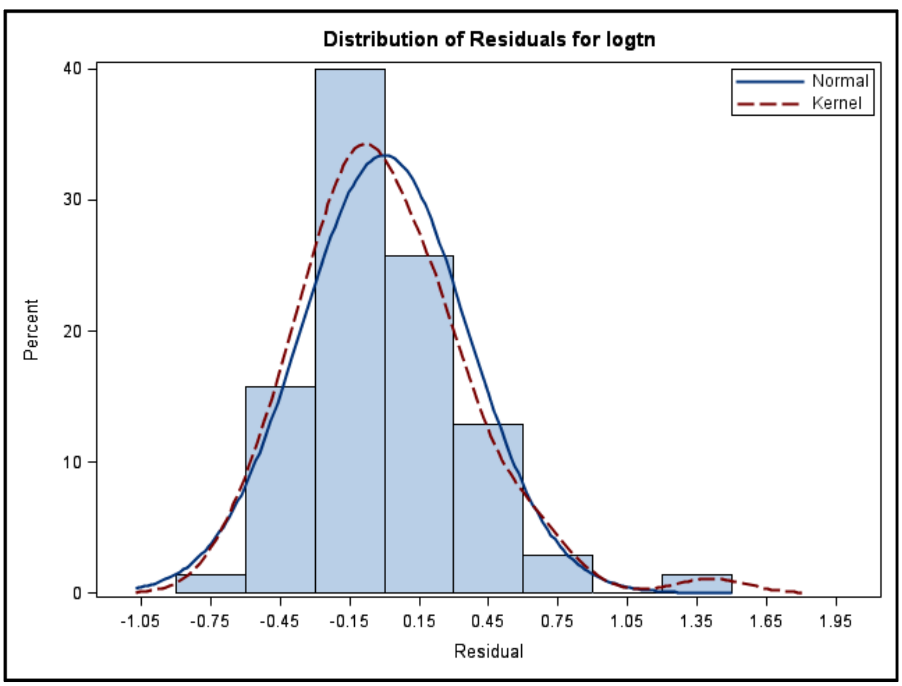

Figure 41: Distribution of Residuals (Total Nitrogen at all Stations)

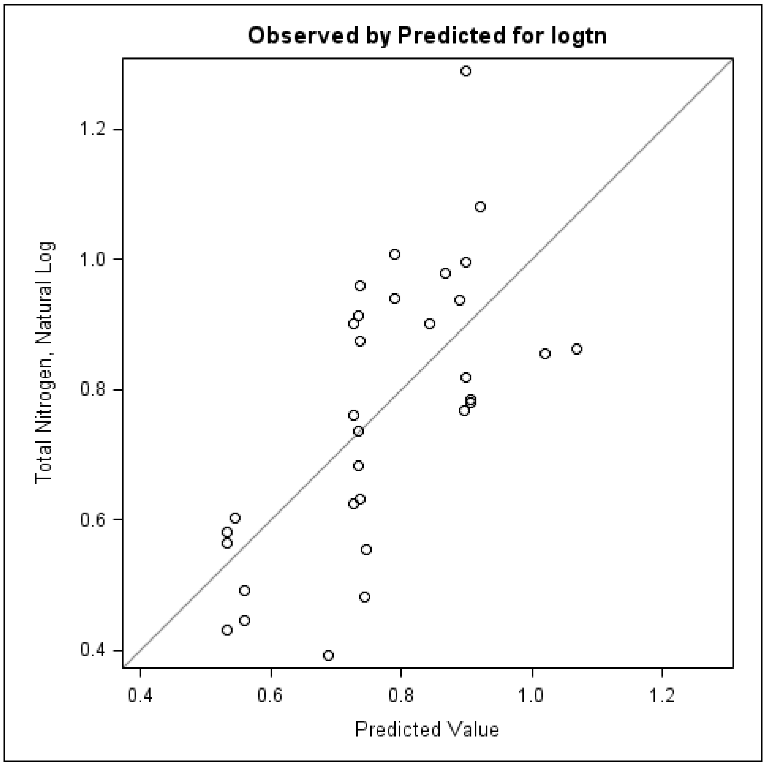

Figure 42: Observed and Predicted Values (Total Nitrogen with Effluent Dominated Streams Excluded)

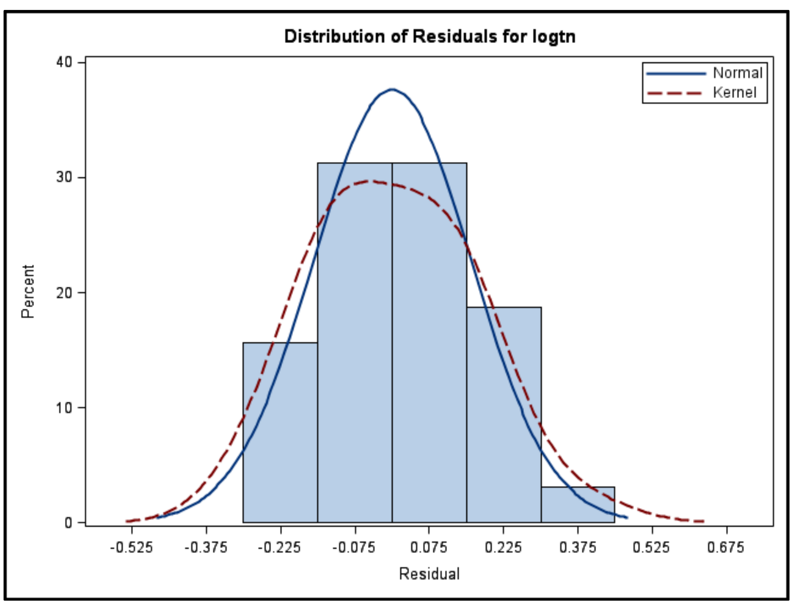

Figure 43: Distribution of Residuals (Total Nitrogen with Effluent Dominated Streams Excluded)

4.8 Canonical Correlation

H-GAC conducted canonical correlation analysis to examine the relationship between nutrients and land cover in the aggregate. Multivariate normality was evaluated using a SAS macro, which showed that the square root of the natural logarithms of total phosphorus and nitrate-N has a multivariate normal distribution. The ratio of observations to variables is far below the minimum of 20 frequently recommended which limits the interpretability of the correlation coefficients; the ratio is 14.1 for effluent dominated watersheds, and 5.3 for non-effluent dominated watersheds. The goal of this analysis is to compare the results for watersheds with and without significant wastewater influence and perhaps to estimate the importance of land cover alone. Given that our intent is general description, the sample size/variable ratio may be less important than in other contexts.

Separate analyses were performed for effluent dominated and non-dominated streams. The results of the analysis are summarized in the Table 42. Both pairs of variates are significant in both analyses. The most significant finding is that the correlation between nutrient levels and land cover types is higher in the watersheds where streamflow at the monitoring station is not dominated by wastewater plant effluent.

- The canonical correlation coefficient is higher (0.6118 compared to 0.4728)
- Land cover variates explain twice the variance in the data (34% compared to 16 percent)
- Canonical R^2 is higher (this is a measure of the predictive ability of one variate from the other, and does not combine the effect of both variates; e.g., predicting nutrient variate 1 from land cover variate 1).
- The cumulative squared multiple correlations is higher in non-dominated watersheds
  - Prediction of total phosphorus from land cover: 0.3532 vs. 0.1877
  - Prediction of nitrate-N: 0.3243 vs. 0.1298

These results suggest that the influence of land cover variability may be higher than the regression and GLM results indicate in all watersheds.

Table 42: Summary of Canonical Correlation Analysis

	Statistic	Effluent-Dominated			Not Effluent-Dominated		
Variate Pair 1	Canonical Correlation	0.4728			0.6118		
	Significance [i]	<.0001			0.0023		
	Proportion [ii]	0.6683			0.5590		
Variate Pair 2	Canonical Correlation	0.3536			0.5662		
	Significance [iii]	0.0086			0.0149		
	Proportion [4]	0.3317			0.4410		
Model Significance	Wilks’ Lambda p-value	<.0001			0.0023		
Redundancy Analysis	Standardized Variance in Nutrient Variates Explained by Land Cover Covariates	Nutrient 1/ LC1: 0.0765 Canonical R^2: 0.2236 Nutrient 2 / LC2: 0.0823 Canonical R^2: 0.1251 Cumulative Variance Explained: 0.1588			Nutrient 1/ LC1: 0.1264 Canonical R^2: 0.3743 Nutrient 2 / LC2: 0.2123 Canonical R^2: 0.3206 Cumulative Variance Explained: 0.3387		
	Squared Multiple Correlations between Nutrients and First M Canonical Variates of Land Cover	M	**1**	**2**	M	1	2
		Total Phosphorus	0.1422	0.1877	Total Phosphorus	0.2271	0.3532
		Nitrate-N	0.0107	0.1298	Nitrate-N	0.0257	0.3243

1, 3 Significance: test of the null hypothesis that the correlation is zero
[1, 4] Proportion: Proportion of variability in the data accounted for by the derived variate pair
[5 Wilks’] Lambda: Ratio of generalized error variance to generalized total variance

4.9 Discriminant Analysis

The dominant land use in each watershed was identified on the basis of the distribution of land cover types. Generally, the dominant land use thus identified accounted for more than fifty percent of the area of the watershed. A categorical variable was created having possible values of “Developed”, “Agricultural”, or “Forest.” H-GAC staff used SAS PROC DISCRIM to create a Fisher linear discriminant function from the natural logs of total phosphorus, total nitrogen, and nitrate concentration. The proportion of watersheds in each category was used as the prior probability of group membership. Each watershed was classified (by the software procedure) into one of these categories. The results are found in Table 43 below.

Table 43: Number of Observations and Percent Classified into Land Cover Categories

From land use	Agricultural	Forest	Urban	Total
Agricultural	28 93.33	0 0.00	2 6.67	30 100.00
Forest	6 60.00	1 10.00	3 30.00	10 100.00
Urban	2 7.41	0 0.00	25 92.59	27 100.00
Total	36 53.73	1 1.49	30 44.78	67 100.00
Priors	0.45	0.15	0.4	

Table 44 identifies the proportion of samples that were incorrectly classified (6.67 percent for agricultural, 90 percent for forest, and 7.41 percent for urban land) for a total of 19.46 percent of the 67 samples or 13 samples.

Table 44: Error Count Estimates for Land Cover Types

	Agricultural	Forest	Urban	Total
Rate	0.0667	0.9000	0.0741	0.1946
Priors	0.4500	0.1500	0.4000	

Table 45: Misclassified Observations

Station ID	From Land Cover	Classified into Land Cover	Agricultural	Forest	Urban
11312	Forest	Urban	0.1013	0.0973	0.8014
11312	Forest	Urban	0.1782	0.1255	0.6962
11312	Forest	Agricultural	0.8561	0.1236	0.0203
11312	Forest	Agricultural	0.5117	0.1811	0.3072
11312	Forest	Urban	0.0376	0.0763	0.8861
11332	Agricultural	Urban	0.2294	0.1624	0.6082
11332	Agricultural	Urban	0.0036	0.0589	0.9375
11334	Forest	Agricultural	0.9000	0.1000	0.0000
11334	Forest	Agricultural	0.8497	0.1415	0.0088
11369	Urban	Agricultural	0.9117	0.0847	0.0036
11387	Urban	Agricultural	0.4436	0.1500	0.4064
17746	Forest	Agricultural	0.9186	0.0812	0.0001
17746	Forest	Agricultural	0.9282	0.0717	0.0001

Table 45 lists the probabilities of samples being from the given land cover type based on total phosphorus and total nitrogen. For example, the first sample shown in the table collected from Station 11312 (Spring Creek watershed), had a probability of 80.14 percent of being from an urban land cover type although the dominant land cover type is forest (forest plus scrub/shrub). This is undoubtedly a reflection of the effluent domination of Spring Creek. It should also be noted that there has been considerable suburban development in this watershed in the past 15 years and, while “Forest” is a significant land cover type for the watershed as whole, it may not reflect the land cover and land uses closest to the site of sample collection. Samples collected from Station 11332 were classified as from urban land. Stations 11312 and 11332 are each located in suburban areas that were classified on the basis of land cover delineation as forest and agricultural respectively. H-GAC also evaluated the accuracy of classification of watershed into effluent-dominated or non-effluent dominated groups from characteristics of the derived nutrient variable. The results are presented in Table 46. The error count estimates for the estimates are presented in Table 47.

Table 46: Classification of Watersheds by Ratio of Effluent Levels

	N	Y	Total
N	28	1	29
Y	3	35	38
Total	31	36	67
Priors	0.5	0.5	

Note:
Y – Effluent dominated
N – Not effluent dominated

Table 47: Error Count Estimates for Ratio of Effluent Levels

	N	Y	Total
Rate	0.0345	0.0789	0.0567
Priors	0.5000	0.5000	

The results indicate that samples at four of the stations were misclassified and those stations are included in Table 48. The results show that there is a 3 percent chance of a sample from a non-effluent group being misclassified and a 7 percent chance of a sample from an effluent-dominated group being misclassified. The following four samples were misclassified:

- Spring Creek (11312) – an effluent-dominated stream has one sample indicative of a non-effluent dominated stream,
- Cypress Creek (11332) – an effluent-dominated stream has one sample indicative of a non-effluent dominated stream,
- Greens Bayou (11369) – an effluent-dominated stream has one sample indicative of a non-effluent dominated stream,
- San Bernard River (12147) – a non-effluent dominated stream has one sample indicative of an effluent-dominated stream.

Table 48: Posterior Probability of Membership in an Effluent Dominated Watershed

Station_ID	From diseff	Classified into diseff	N	Y
11312	Y	N	0.8970	0.1030
11332	Y	N	0.9406	0.0594
11369	Y	N	0.9779	0.0221
12147	N	Y	0.0041	0.9959

4.10 General Comments on Statistical Analyses

4.10.1 Statistical Power and Sample Size

Type II Error is the probability that the null hypothesis is not rejected when it should be. Power is the ability of a test to “correctly” reject the null hypothesis and is calculated as (1-Type II Error Rate). The Type I error rate can be controlled by the researcher by simply selecting it but the Type II rate depends upon the sample size, the actual value of the population parameter, the true variance of the parameter in the population, and the magnitude of the effect (effect size) if the alternative hypothesis is true.

Various “rules of thumb” have been proposed for the appropriate sample size for developing predictive regression or ANCOVA models with no general agreement. Many propose a minimum of 10 observations per independent variable; others suggest between 10 and 30, and some authors believe 30 are too few. If the intent of the model is prediction, the sample size required is a function of the effect size. In multiple regression, this is usually estimated by the squared multiple correlation coefficient (Knofczynski and Mundfrom 2008). If one is designing a controlled experiment, the sample size required to obtain the desired statistical power (typically 0.8) is calculated before the experiment begins. In observational studies, one might calculate the number of existing observations to be randomly sampled as “subjects” of the study. Power calculations prior to the analyses can be considered *a priori* power calculation. Given the number of watersheds for which land cover data was created and the number of observations made in the four years for which land cover datasets existed, H-GAC had to use all the observations. However, we now have some sense of the effect sizes attributable to land cover and can estimate the sample size necessary to detect other effects of that magnitude, as well as determine the probability that other effects were not detected. Exploratory and *post hoc* power analysis was performed using G*Power 3, a freeware application developed at the Institute for Experimental Psychology in Düsseldorf, Germany (Faul *et al* 2007).

G*Power 3 estimates post-hoc statistical power using Cohen’s effect sizes (Cohen 1988). Table 49 shows the sample size needed to detect effects of the size detected in the study with no more than a 20 percent type II error rate (odds of missing a significant predictor).

Table 49: Sample Size Needed to Detect Effects with Less Than a 20 Percent Type II Error Rate

Dependent Variable	Sample Size	Number of predictors	Model R^2	Statistically significant land cover variables	Partial R^2	Cohen's f^2	Sample Size Needed to Detect with power=0.8
Total Phosphorus (natural log)	**52**	2	0.2906	Wetlands	0.0664	0.0711	**90**
				Developed	0.2242	0.2890	**23**
Total Nitrogen (natural log)	**32**	4	0.4806	Agricultural	0.1446	0.1689	**39**
				Developed	0.1247	0.1425	**45**
				Grasslands	0.1023	0.1140	**56**

With the exception of the strong effect of developed area on total phosphorus concentration, the analyses had a high probability of failing to detect significant predictors. Some research has shown that the interaction of rainfall with land cover types is an important variable for an understanding of relationships. The small sample size available in the study left a low probability of detecting the effect of interactions between these variables although all interactions were tested in the initial ANCOVA models.

4.10.2 The Influence of Wastewater Discharge

H-GAC recognized, on the basis or prior analysis, that wastewater was a significant source of nutrients in area waterways and its influence on the variance of concentrations must be controlled if the effects of land cover and land use are to be revealed. We hoped to use TCEQ DMR data to gauge the contribution of wastewater effluent to measured flow and use effluent discharge as a variable in the models. However, reliable data were only available for one of the study years (2011). As a substitute, streams were categorized as either effluent-dominated or not and a wastewater discharge level variable was created from the sum of permitted discharges for all permittees upstream of the monitoring site. These variables proved to be significant predictors of nutrient concentration but the following caveats must be considered:

- Most of the streams in the study receive some domestic wastewater, even if they are not effluent dominated
- The effluent domination variable is a dichotomous dummy variable and is inferior to continuous estimates of effluent discharge
- The wastewater discharge level variable was based on data in an internal database that is updated infrequently.
    - It is likely that some of the permitted flow limits in the database are wrong
    - Some permittees that discharge effluent have no specific flow limit

  - o We assumed that the relative discharge in the past is proportional to the discharge levels at present, and this is unlikely but unavoidable if the variable is to be used

- Numerous studies have found that that impervious cover is positively correlated with nutrient concentrations in storm water runoff (Carey et al 2013). Impervious cover was calculated only for one year and was not available when statistical analysis began. Impervious cover is higher in “developed” land cover and the positive correlations between developed area, effluent dominance, and nutrient concentrations were noted earlier. Separating the contributions of runoff and wastewater effluent in urban watersheds is no small task; isotopic tracer methods can be useful but the cost prohibits their inclusion in most routine ambient monitoring programs.
- All levels of development intensity were combined in the “developed” land cover type making it difficult to assess the contribution from one nutrient-intensive land cover type: suburban lawns. This is undoubtedly a significant source of nutrients in storm water runoff, but this land cover type could not be isolated from other land cover types. There is substantial sub- and exurban development in the Cypress and Spring Creek watersheds but GIS analysis found agriculture and forest (respectively) to be the dominant land cover types.

Finally, as noted earlier there is a “bimodal” distribution in much of the data. The high concentrations associated with effluent dominated streams significantly influence the model R^2 when both types of watershed are modeled. When they are modeled separately, the model R^2 drops significantly for both groups from those where they are combined. Segregating these data in predictive models would result in more realistic estimates of the model’s explanatory power.

4.10.3 Watershed and Station Selection

Four land cover datasets were available from which to categorize and delineate each watershed. In retrospect, our selections limited the amount of data from non-effluent dominated watersheds. Given the overwhelming influence of wastewater discharges, we are more likely to discover relationships between land cover and nutrient concentrations (if they exist) in these watersheds. Additionally, discriminant analysis shows that nutrient data from the watersheds where “forest” is the dominant land use have little in common; discriminant analysis showing a 90 percent classification error rate. In some cases, the land cover closest to the monitoring station is drastically different than in the watershed as a whole. This is particularly true for station 11332 on Cypress Creek, located in a suburban area but in a watershed that (as defined) is dominated by agricultural land cover. Data from this station and from the Spring Creek “forest” station, also located in a rapidly developing sub-or-ex urban area, frequently showed up as outliers and influential observation in regression diagnostics.

Some researchers have found land cover/land use in a buffer zone along a stream is more closely associated with water quality than land use/land cover in the catchment as a whole (Tran *et al* 2010) while other work has shown the opposite (Sliva and Williams 2001). In future work, it might be useful to create and characterize a riparian buffer. It should also be noted that agricultural land use did not correlate with nutrient concentration in most models. The cultivated area (the "agriculture" variable in the models) included both pasture/grazing land and land devoted to row-crop agriculture, rice fields, sod farming, etc. Different types of agricultural land uses exhibit differing nutrient runoff profiles (Haggard *et al* 2003). The impact of one could be masked by another. However, given the small sample sizes at hand, limiting the number of independent predictors was vital. In the future, consideration should be given to variable reduction methods such as factor and principal components analysis.

4.11 Load Duration Curves

H-GAC developed Load Duration Curves (LDCs) and Flow Duration Curves (FDCs) to characterize the watershed responses to precipitation and other inputs and understand the load capacity of receiving waters for pollutants of concern.

FDCs characterize a watershed's response to precipitation and other inputs, integrating multiple factors that affect streamflow at a point (topography, soil distribution, climate, land use, flow controls such as dams, etc). A flat FDC implies a greater level of storage in the basin and a steeper FDC implies a flashy watershed, where streamflow increases quickly following precipitation.

An LDC with a decreasing trend indicates that the largest concentrations occur at high flow rates. For constituents with decreasing trends, the supply in the watershed is available for transport by runoff from a terrestrial source, and/or may be mobilized via in-stream sediment transport processes associated with increased stream velocities and higher flows from precipitation. An increasing trend implies that constituent supply is limiting, and/or dilution occurs during precipitation events, and indicates that the largest concentrations occur at lower flow rates. Larger concentrations may occur at lower flow rates, for example, because base flow is derived from stored water having long contact times within the aquifer or because of continuous discharges that dominate at low flow (e.g., point sources). A curve with a static trend (flat curve) indicates that constituents have no relationship with flow rates and constant loads are released to water. During high flow events, the concentrations are diluted and loads are maintained at constant levels.

Plotting TMDL screening levels with an LDC is useful in providing information whether the watershed is impaired or not. If impaired, the LDCs may provide information about flow conditions where the stream has nutrient levels above the screening limit. Examples of LDCs are provided below. The complete set of LDCs for each of the watersheds is included in Appendix C of this report. These can be compared against Table 5 to identify the likely contributing source areas for flow regimes where there are concerns. The LDCs estimate that:

- Station 11120, 11334, 11367, 11467, 11484, 12147, and 17746 comply with or are close to complying with screening levels for total nitrogen and total phosphorus during all flow conditions. This suggests that loads from different sources to these watersheds are below or near targets. Several examples of the LDCs are provided below in Figure 44 through Figure 47.

- Station 11125 and 11369 (for total phosphorus) exceed the screening levels during all flow conditions shown in the LDC. This suggests that there are likely many sources of concern contributing high loads of nutrients including point sources and storm water. An example is shown in Figure 48.
- Station 11135, 11139, 11312, 11332, 11369 (TP and Nitrate), and 11387 (for total phosphorus) exceed screening levels for most flow conditions, but comply during some higher flows. This suggests that the sources present during low to mid-range flows are the largest sources of concern, such as point sources, on-site wastewater systems, riparian areas, and storm water. An example of an LDC that characterizes these LDCs is shown in Figures 49 and 50. These watersheds are characterized with high densities of WWTP outfalls and major land cover types are developed categories. These suggest that the major sources for the nutrient levels are from wastewater effluents.
- LDCs for station 11334, 11367, 11467, 11484, and 17746 are all below the screening levels and do not provide any indicator of particular source. An example is shown in Figure 51. Considering the dominant land cover types and wastewater outfall density, these watersheds can be classified as rural (with less developed land cover) and lower outfall densities. Therefore, the data and results indicate nonpoint sources contribution which may be from on-site wastewater systems or nutrients contribution from the agricultural lands.

LDC analysis helps not only in recognizing the potential nutrient sources but also in identifying the watersheds where nutrient reductions are needed. Based on the above LDC analysis there are several watersheds where nutrient reduction strategies need to be applied. Watersheds such as Garners Bayou (11125) and Brays Bayou (11369) both have nutrient levels above the screening criteria in all flow conditions. Any control measures established would be beneficial to water quality. Watersheds such as Sims Bayou (11135), Brays Bayou (11139), Spring Creek (11312), Cypress Creek (11332), Greens Bayou (11369), and Whiteoak Bayou (11387) have nutrients levels above the screening criteria in medium to low flow conditions. Therefore H-GAC recommends conducting further investigations in these watersheds to identify problem point sources and establish essential control measures to reduce nutrient releases to water ways in those flow conditions.

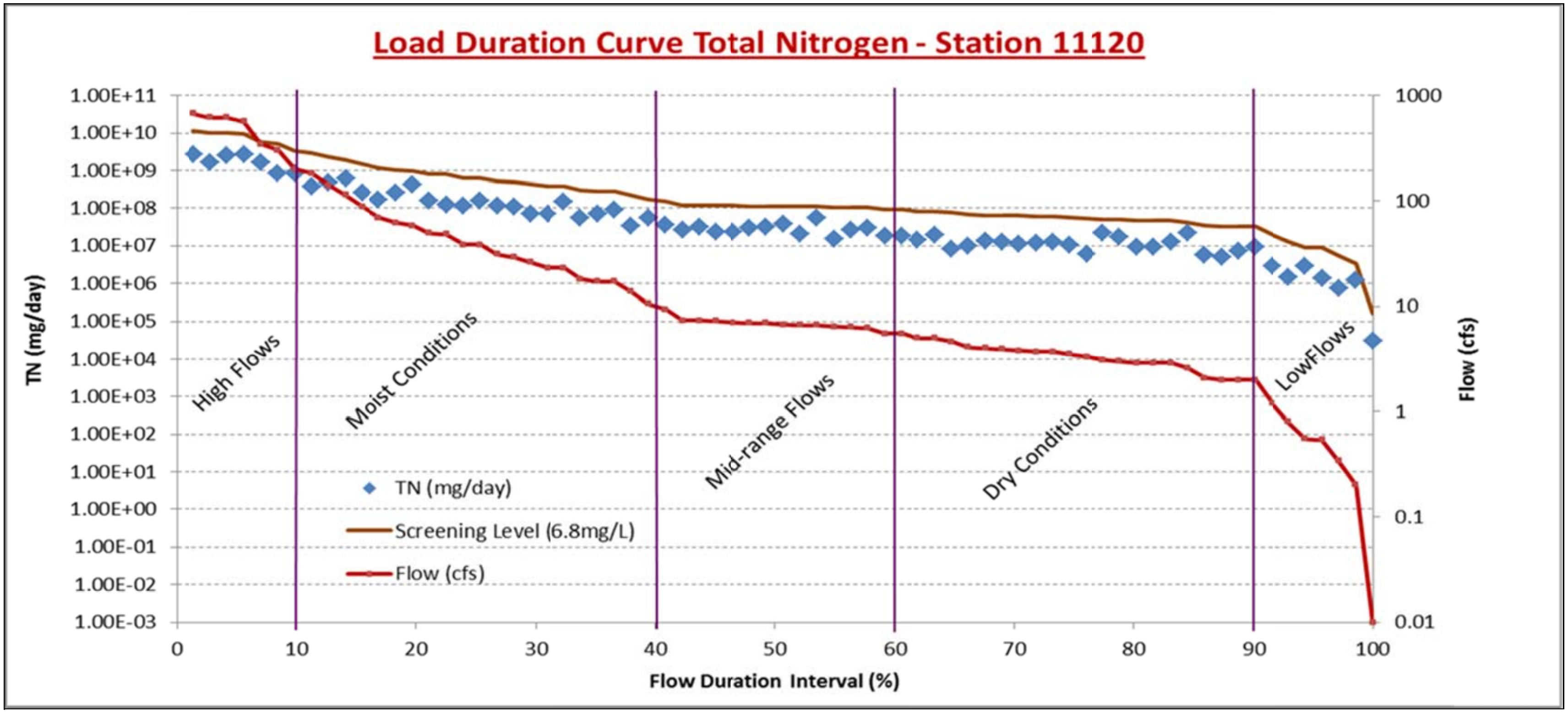

Figure 44: Load Duration Curve for Total Nitrogen (11120)

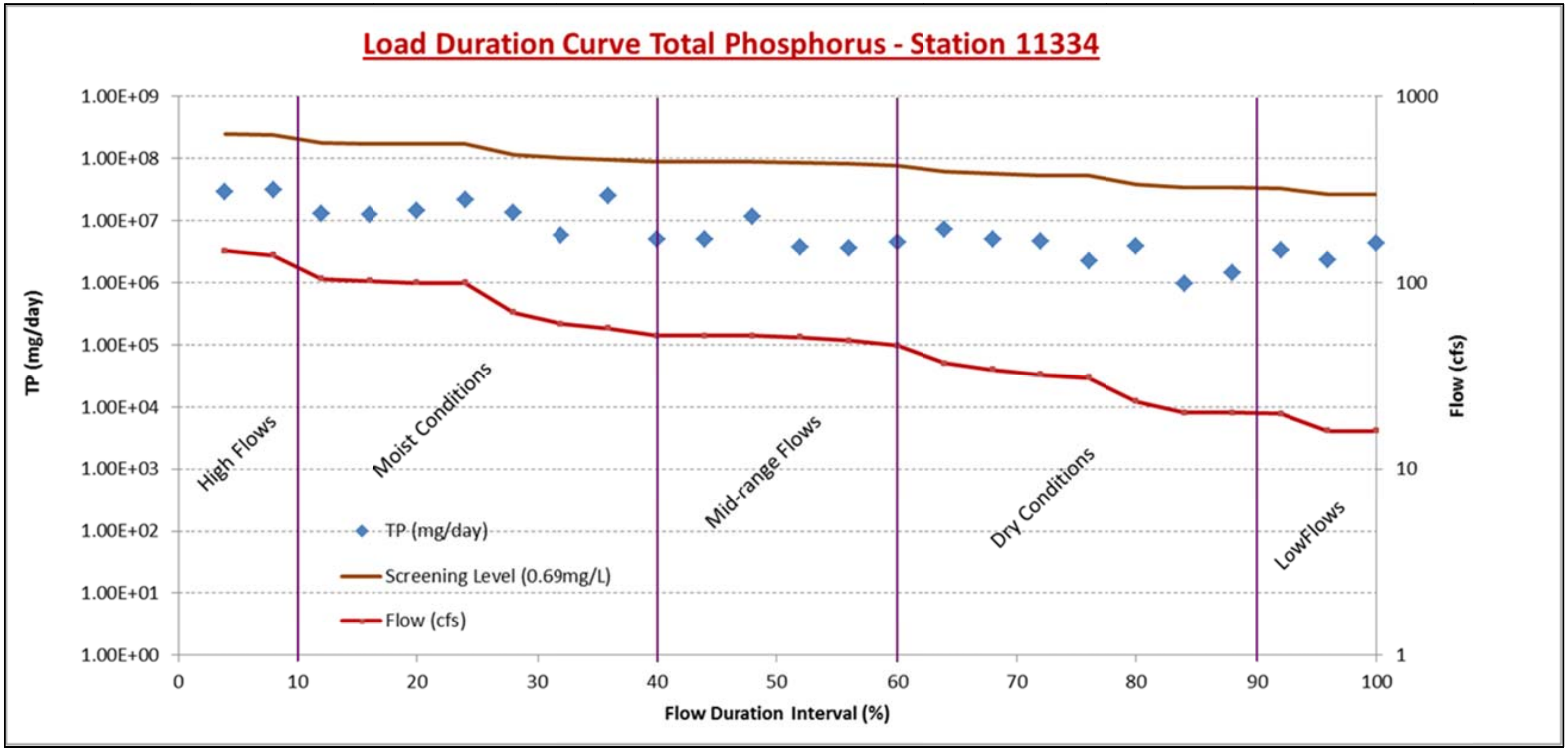

Figure 45: Load Duration Curve for Total Phosphorus (11334)

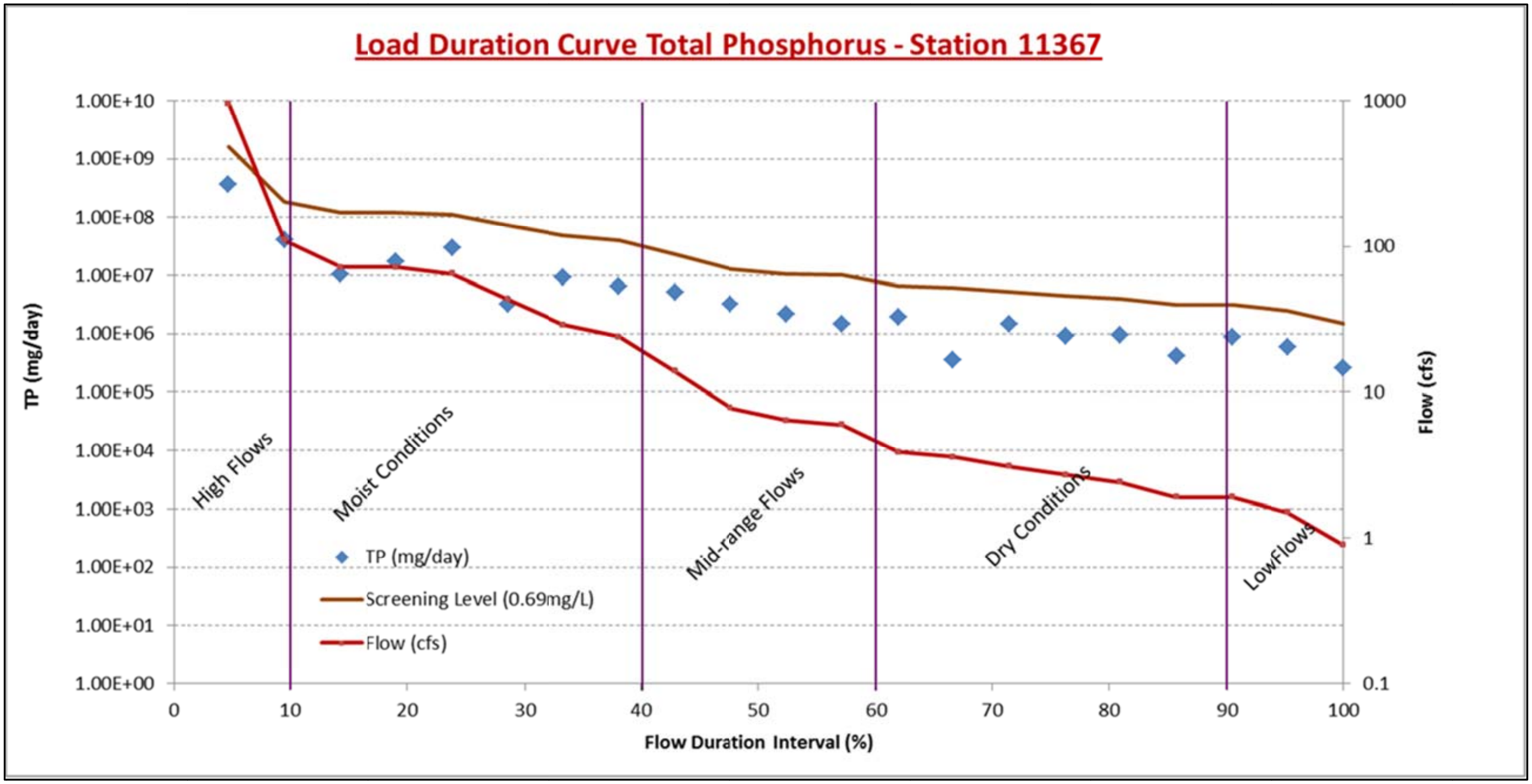

Figure 46: Load Duration Curve for Total Phosphorus (11367)

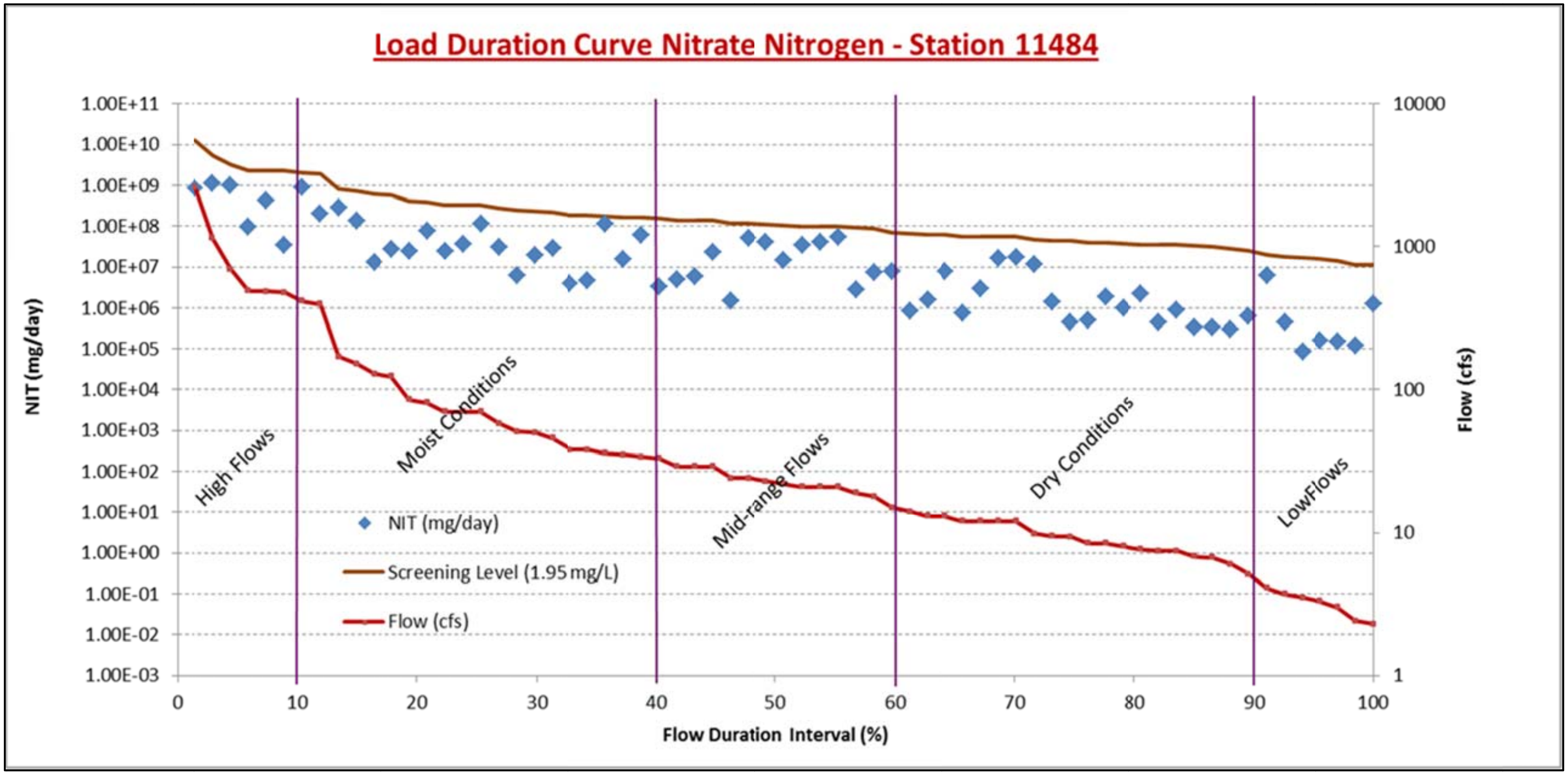

Figure 47: Load Duration Curve for Nitrate Nitrogen (11484)

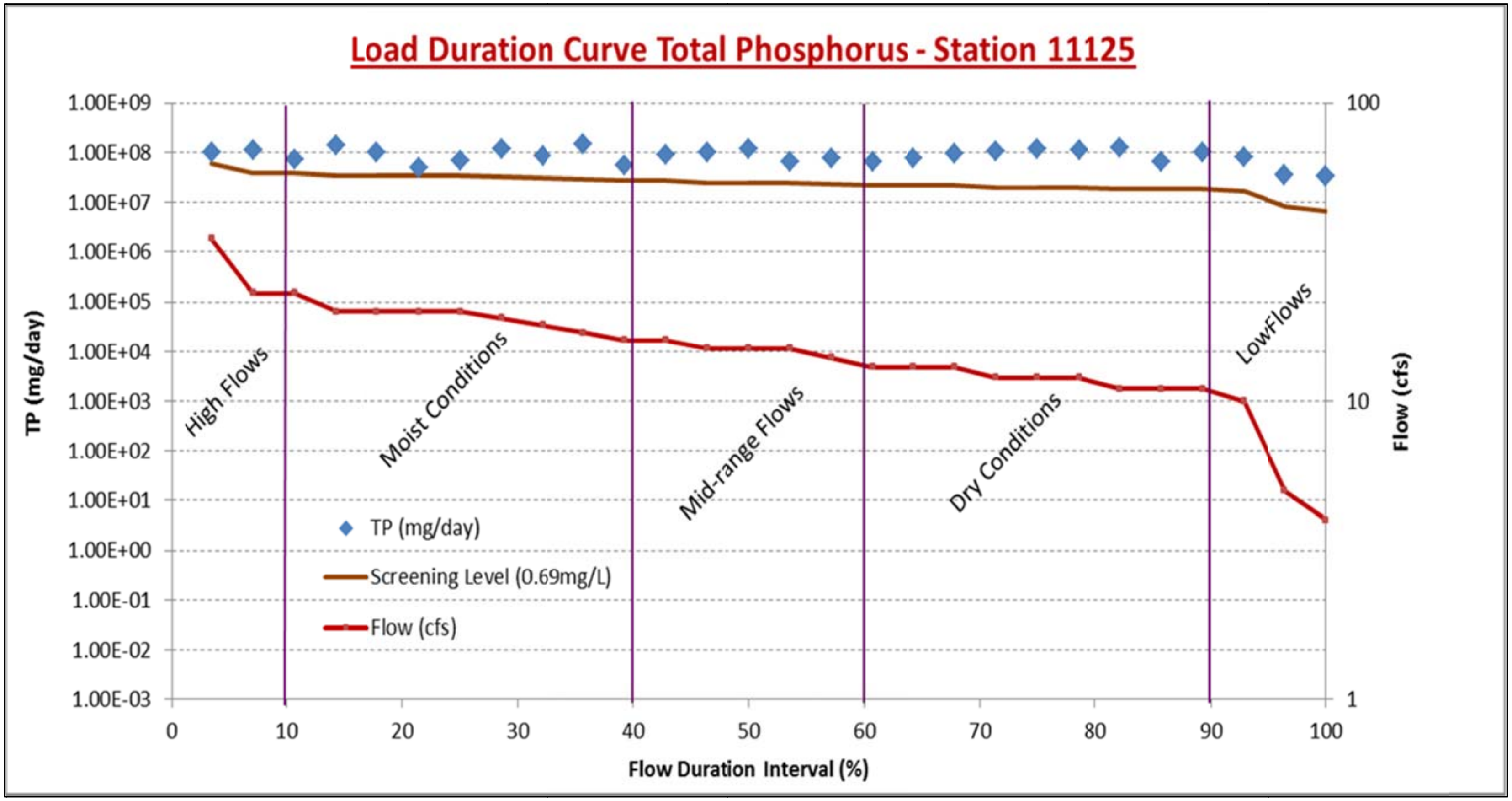

Figure 48: Load Duration Curve for Total Phosphorus (11125)

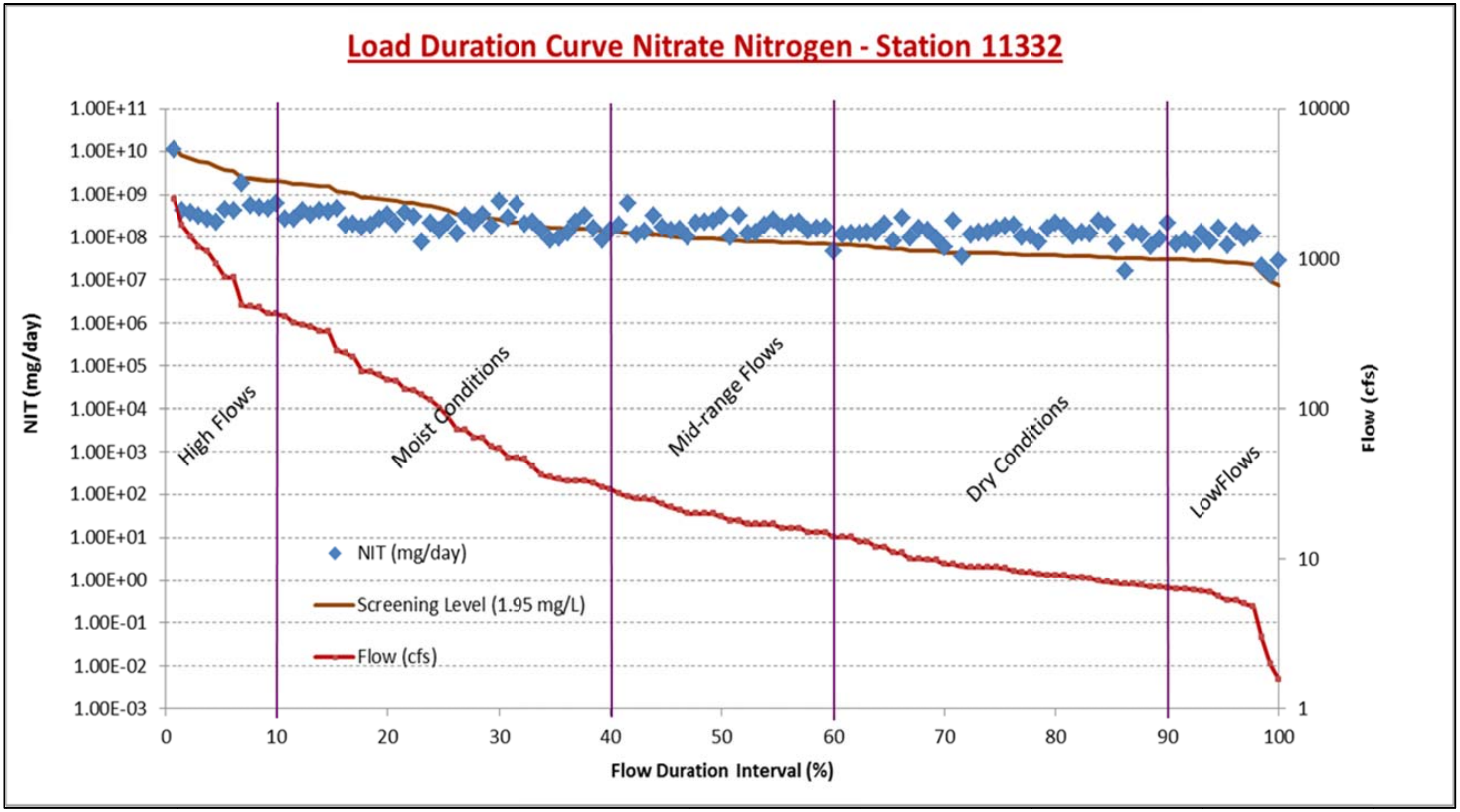

Figure 49: Load Duration Curve for Nitrate Nitrogen (11332)

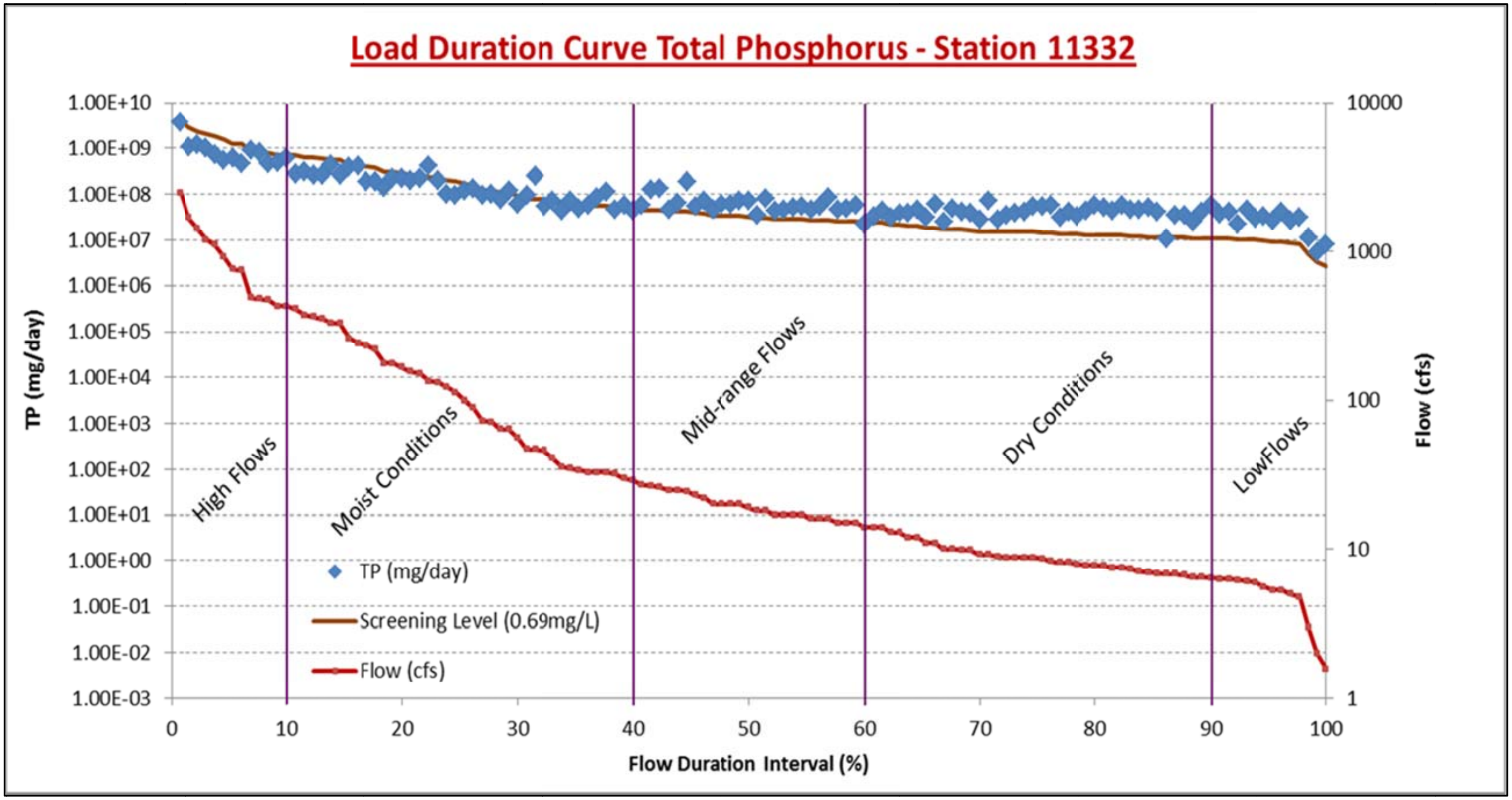

Figure 50: Load Duration Curve for Total Phosphorus (11332)

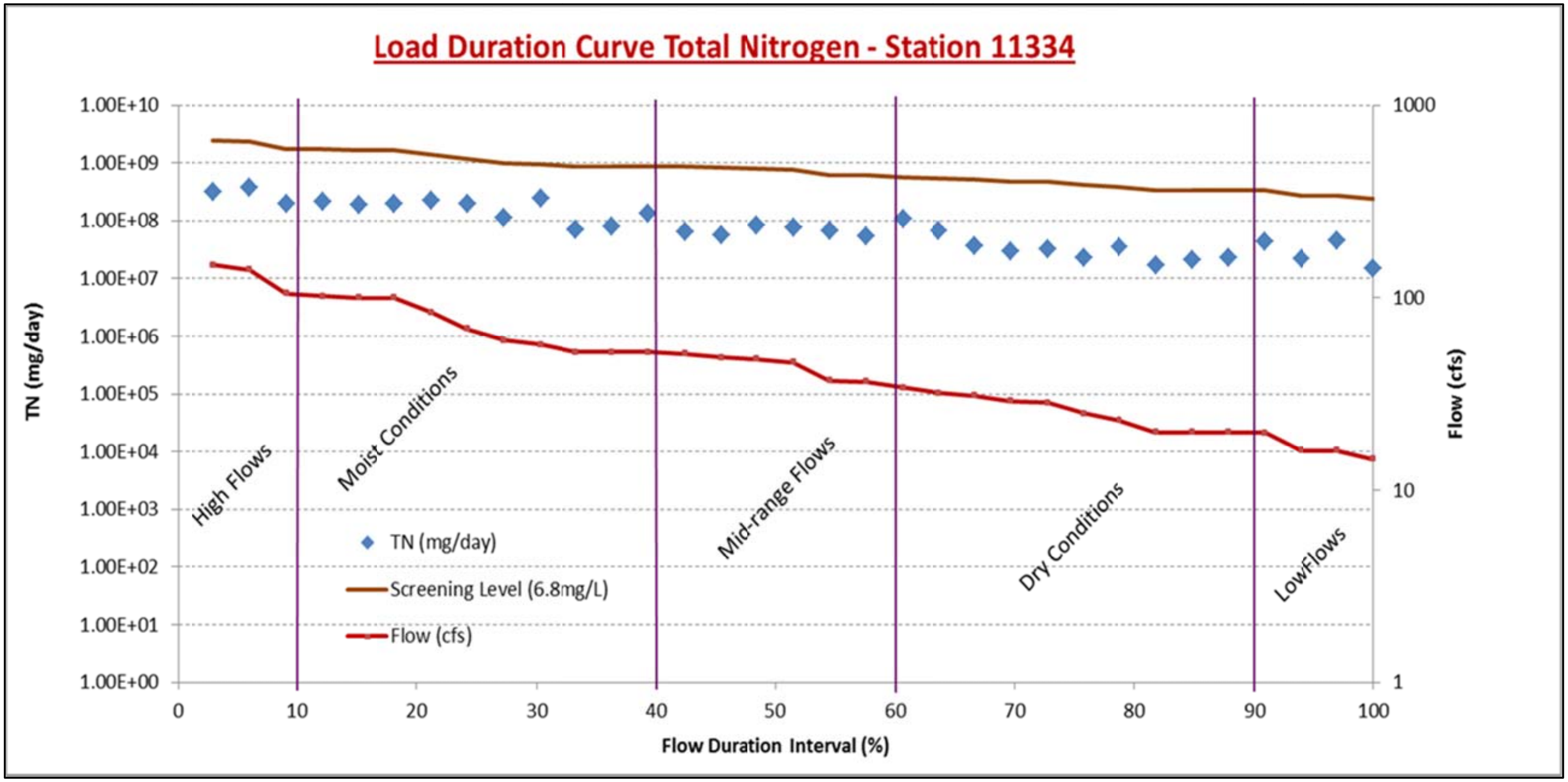

Figure 51: Load Duration Curve for Total Nitrogen (11334)

5 Summary and Conclusions

H-GAC conducted this project to identify potential correlations between land cover and/or inferred land use and ambient nutrient concentrations in selected streams in the region. The information provided by this analysis is intended to increase the understanding of water quality concerns due to nutrient loads in runoff from watersheds and the sources of these loads. In addition, any identified correlations could help identify sources of nutrient loads to help prioritize implementation of structural and non-structural BMPs at locations where these would be most effective to improve receiving water quality.

The analyses performed under this project were based on existing water quality data. H-GAC acquired available water quality data, GIS technology, and modeling land use/land cover information. Water quality data were acquired from one primary source, the TCEQ SWQMIS. Only 'non-qualified', routine, ambient, fixed station water quality data from SWQMIS post-December 31, 1995 for the H-GAC region were used. All acquired water quality data were collected under TCEQ approved QAPPs. Other physical and natural characteristics, geospatial data, and land cover data were acquired from H-GAC's GIS spatial database and from other reliable federal and state government databases.

This project included advanced statistical and spatial analyses of water quality and geospatial data by evaluating ambient nutrient data using GIS technology and modeling of land cover and/or inferred land use information to help develop correlations based on watershed characteristics. Analysis included an evaluation of the association of land cover changes over time and nutrient concentration trends.

Data from fourteen monitoring locations in the Houston region were analyzed. Drainage areas (sub-watersheds) for each monitoring station were delineated and hydrologic, land cover, and water quality characteristics of each were defined. In addition, H-GAC investigated the potential impact of domestic wastewater discharges on water quality measured at the monitoring stations.

Land cover change analysis indicates that majority of the watersheds have an increasing trend of developed land cover types between years of 1996 to 2011. Only two watersheds show an increasing trend in natural and unmanaged grasslands. Three watersheds show no significant change in any land cover classes. Watersheds that experienced either increasing or decreasing trends in any type of land class have significant decreasing of forest lands. This trend indicates that forest is the primary land cover type consumed in creating other land types. Other than forest lands, cultivated land shows a considerable amount of loss over the period of study.

Statistical analysis suggests that wastewater treatment plant effluent is a major source of nutrients in urban watersheds and the level of wastewater discharge is positively correlated with nutrient concentration overall. Several land cover variables were found to be statistically significant predictors of nutrient concentration and some land cover predictors appeared in several models. Wetland and forested land cover were found to be significant in several analyses. Wetland and forested land cover are negatively correlated with nutrient concentration in watersheds not dominated by wastewater effluent, although the effective size is modest. Unfortunately, due to the small sample size reliable predictions cannot be made with these models. Regression, ANCOVA, and canonical correlation analyses all suggest that land cover explains 30-40 percent of the variance of nutrient concentration when effluent-dominated streams are excluded. This is not generally considered “high explanatory power”, but the probability of Type II errors was found to be high and important predictors may not have been detected.

For several reasons, the results of the regression and ANCOVA models should be interpreted with caution:

- Boxplots of the distribution of nutrients in each watershed (Examples figures 29-35) show that nutrient concentrations are higher in urban watersheds. ANOVA analysis suggests these differences are statistically significant. Models that include all data (urban watersheds with significant wastewater influx and nonurban watersheds with less impact) show an approximate bimodal distribution, with urban data clustered at high concentrations, while data from nonurban watersheds generally have lower concentrations. Due to the nature of regression analysis, observations with higher values of dependent and independent contribute more to the overall sum of squares, inflating the R^2 of the model. When the two clusters are modeled separately:
    - Different predictors are significant
    - R^2 falls dramatically for each model
    - The model fit is better for data from rural areas without much wastewater influx than for urban areas

- The models for urban watersheds do not adequately reflect surface runoff. Wastewater influence and developed area are highly correlated (collinear) and would not usually be included in the same model. Nutrients entering waterways from impervious-surface runoff in urban areas could not be separated from wastewater contributions with the data available - a problem which has been noted elsewhere (Carey *et al* 2013). Impervious cover and quantitative effluent discharge data could improve this study.
- Given that it is difficult to properly partition nutrient concentration variance between wastewater and runoff in urban areas, more can be learned about the relationship between land cover and nutrient concentrations when wastewater influence in minimized. The sample of data from waterways receiving minimal discharges form WWTPs was too small to achieve sufficient statistical power. Consequently, some important relationships might not be detected and the statistical models will be limited to a small number of predictors.
- Previous research has found important interactions between variables, for instance between rainfall and land cover type. The sample size was too small to properly model such interactions.
- Partial residual plots suggested that in some cases, polynomial models could provide a better fit. A larger sample would be required to develop and assess polynomial models.

H-GAC also developed LDCs for nitrate, total nitrogen and total phosphorus for each of the watersheds analyzed in this project. These can be evaluated to identify the likely contributing source areas for flow regimes where there are water quality concerns. The LDCs estimated that several of the stations do not exceed nutrient screening levels for total nitrogen and total phosphorus during all or most flow conditions, suggesting that most of the sources are contributing nutrients to these watersheds at very low levels. Two of the stations were identified as exceeding screening criteria for one or more nutrient types during all flow conditions shown in the LDC. This suggests that there are likely many sources of concern contributing high loads of nutrients including point sources and storm water. Five of the stations exceed screening levels for one or more nutrient types during most flow conditions, but comply during higher-medium flows. This suggests that the sources present during low to mid-range flows are the largest sources of concern, such as point sources, on-site wastewater facilities, and riparian areas. Watersheds with less developed land cover types are showing nutrient levels below the screening levels for most of the constituents and no increasing or decreasing trends associated with flow levels. For these watersheds, the likely nutrient sources are non-point sources.

In conclusion, this study found real, but relatively weak, associations between land cover/land use, and in-stream nutrient concentrations. More than half of the data came from effluent-dominated streams where the influence of land cover variation would be masked by that of waste water discharge. When effluent-dominated streams are removed from an analysis to highlight the impact of land cover variation, the sample size is too small to reliably detect variables that may have a real, but relatively modest, effect on nutrient concentrations. This study was, first and foremost, exploratory. Reliable waste water discharge data, rainfall data collected closer to the monitoring stations, delineation of smaller study areas upstream of monitoring stations, and more data from stations on streams that are not dominated by wastewater effluent would improve a study of this type. Multiple models suggest that nutrient concentrations are positively correlated with developed area, and inversely related to forested land and wetlands. An unrelated analysis has shown that wetlands and forest are rapidly being replaced by developed area in the H-GAC region and the protection of these areas is essential. H-GAC believes that this study has identified relationships between nutrient concentrations and other variables that provide a good basis for future analysis.

6 References

Carey, Richard 0. , George J. Hochmuth, Christopher J. Martinez,Travor H. Boyer, Michael D. Duke, Gurpal S. Toor, and John L. Cisar. 2013. Evaluating nutrient impacts in urban watersheds: Challenges and research opportunities. *Environmental Pollution* 173:138-149.

Cleland, B.R., 2003. *TMDL Development from the "Bottom Up" –Part III. Duration Curves and Wet-Weather Assessments. National TMDL Science and Policy 2003 – WEF Specialty Conference*. Chicago, IL.

Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences,* Second Edition. Hillsdale NJ: Lawrence Erlbaum

Faul, Franz *et al.* 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* 39(2): 175-191.

H-GAC, 2013. H-GAC's Regional Compilation and Evaluations of Nutrient and Geospatial Data Acquired from Existing Sources Quality Assurance Project Plan (QAPP) – Revision #0. USEPA QTRAK# 13-289, Under EPA Grant No. 19-8665306 FY12-13, CWA 106. TCEQ Grant No.: 998813. Submitted to TCEQ: 4/18/2013

Haggard, B.E., P.A. Moore Jr., I. Chaubey, E.H. Stanley. 2003. Nitrogen and phosphorus concentrations and export from an Ozark plateau catchment in the United States. *Biosystems Engineering* 86(1):75-85.

Helsel, Dennis R. 1991. Technical Memorandum 91.1 – "Statistics – Trend Analysis: Review of Techniques. Briefing paper prepared for staff of USGS Branch of Systems Analysis. http://water.usgs.gov/admin/memo/BSA/BSA91.01.pdf ; retrieved 10 September 2013.

Helsel, D.R. and R.M. Hirsch. 2002. *Statistical Methods in Water Resources;* Chapter A3 of United States Geological Survey *Techniques of Water-Resources Investigations of the United States Geological Survey; Book 4, Hydrologic Analysis and Interpretation.* http://pubs.usgs.gov/twri/twri4a3/pdf/twri4a3-new.pdf ; retrieved 10 September 2013.

Hess, Ann, Paul Patterson and Harry Iyer. 2002. A SAS Macro for Theil Regression. http://vista.cira.colostate.edu/improve/publications/graylit/026_theilreg/theilregressionl.pdf , retrieved 10 September 2013.

Hirsch, Robert M. , Richard B. Alexander and Richard A. Smith. 1991. Selection of methods for the detection and estimation of trends in water quality. *Water Resources Research* 27 (May):803-813

Kendall, M.G., 1939. The geographical distribution of crop productivity in England. Journal of the Roal Statistical Society, 102: 21-62.

Knofczynski, Gregory T. and Daniel Mundfrom. 2008. Sample sizes when using multiple linear regression for prediction. *Educational and Pychological Measurement* 68:431-442

Manly, Bryan F.J. 2001. *Statistics for Enviromental Science and Management.* Boca Raton: Chapman and Hall/CRC

McFarland, Anne and Jimmy Millican. 2011. Assessment of Water Quality Trends for the North Bosque River through 2010. Final Project Report to Surface Water Quality Monitoring Team, Texas Commission on Environmental Quality. http://www.tceq.texas.gov/assets/public/compliance/monops/nps/projects/10079NorthBosque2011Report.pdf, retrieved 26 March 2013.

Mohamoud, Y. 2004. Comparison of Hydrologic Responses at Different Watershed Scales. Ecosystem research Division, US EPA, Athens, GA. for characterization description for each parameter

Nash, Maliha S. and Deborah J. Chaloud. 2002. Multivariate Analyses (Canonical Correlation and Partial Least Square (PLS)) to Model and Assess the Association of Landscape Metrics to Surface Water Chemical and Biological Properties Using Savannah River Basin Data. United States EPA Publication EPA/600/R-02/091.

Ott, R.Lyman and Michael Longnecker. 2010. *An Introduction to Statistical Methods and Data Analysis,* Fifth Edition*.* Belmont CA: Brooks/Cole

Ragavan, Anpalaki J. and George C. Fernandez. 2006. Modeling Water Quality Trend in Long Term Time Series. SAS Users Group International (SUGI) publication 205-31.

Rheem, Sungsue and Golde I. Holtzman. 1991. A SAS®Program for Seasonal Kendall Trend Analysis. SAS Users Group paper, SuGI91. http://www.sascommunity.org/sugi/SUGI91/Sugi-91-209%20Rheem%20Holtzman.pdf , retrieved 15 September 2013.

Riggs, Michael R, Kurt J. Haroldson, and Mark A. Hanson. 2008. Analysis of Covariance Models for Data from Observational Field Studies. *Journal of Wildlife Management* 72(1):34-43.

SAS Institute, Inc. 2012. *SAS/STAT 12.1 User's Guide.*

- 2011. *SAS 9.3 Base Procedures.* Cary NC: SAS Institute Inc.

- 2011. *SAS ODS Procedures Guide, Second Edition.* Cary NC: SAS Institute Inc.

-2008. *Statistics 2: ANOVA and Regression Course Notes* Cary NC: SAS Institute Inc.

- 2008. *Multivariate Statistical Methods: Practical Research Applications Course Notes* Cary NC: SAS Institute Inc.

Schueler, T. 1994. The Importance of Imperviousness. Watershed Protection Techniques. 1(3): 100-111.

Sliva, Ludie and D.Dudley Williams. 2001. Buffer zone versus whole catchment approaches to studying land use impact on river water quality. *Water Research* 35(14):3462-3472.

Stevens, James, 2002. *Applied Multivariate Statistics for the Social Sciences Fourth Edition.* Matwah NJ : Lawrence Erlbaum

Tian, Jie and George C.J. Fernandez. 1998. Seasonal Trend Analysis of Monthly Water Quality Data. University of Alberta School of Business Research Paper No. 2013-1203.

Tran, Christopher P., Robert W. Bode, Alexander J. Smith, and Gary S. Kleppel. 2010. Land-use proximity as a basis for assessing stream water quality in New York State (USA). Ecological Indicators 10:727-733

USDA NRCS, 2013. Web Soil Survey. http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm (last updated on 2/15/2013)

USDA, 2007. Hydrologic Soil Groups

USEPA, 2011. *Working in Partnership with States to Address Phosphorus and Nitrogen Pollution through Use of a Framework for State Nutrient Reductions.* Technical Memorandum – March 16, 2011.

USEPA, 2008*. Development of Duration-Curve Based Methods for Quantifying Variability and Change in Watershed Hydrology and Water Quality.*

USEPA, 2007. *An Approach for Using Load Duration Curves in the Development of TMDLs*. EPA 841-B-07-006. http://www.epa.gov/owow/tmdl/techsupp.html

Winkler, Steve. 2004. A User-Written SAS Program for Estimating Temporal Trends and their Magnitude. St. Johns River Water Management District Technical Publication SJ2004-4. http://www.sjrwmd.com/technicalreports/pdfs/TP/SJ2004-4.pdf, retrieved 18 September 2013.