https://doi.org/10.25923/2wcd-7q26

Correlation and Prediction of Land-Based Pollution Impacts on Sediment Contaminant Concentrations within National Estuarine Research Reserves of the Southeastern US and Selected Chesapeake Bay Watersheds

NOAA National Centers for Coastal Ocean Science
Stressor Detection and Impacts Division
William L. Balthis

NOAA Office for Coastal Management
David L. Eslinger

July 2018
NOAA Technical Memorandum NOS NCCOS 250
NOAA NCCOS Stressor Detection and Impacts Division

Acknowledgments

The authors would like to thank A.K. Leight, Jacqueline Tay, and Bob Wood for assembling the Chesapeake Bay data used in this study.

Citation:

Balthis, W. L. and D. L. Eslinger. 2018. Correlation and Prediction of Land-Based Pollution Impacts on Sediment Contaminant Concentrations within National Estuarine Research Reserves of the Southeastern US and Selected Chesapeake Bay Watersheds. NOAA Technical Memorandum NOS NCCOS 250, NOAA National Ocean Service, Charleston, SC 29412-9110. 31 pp. doi:10.25923/2wcd-7q26

Cover photo credit: NOAA

Correlation and Prediction of Land-Based Pollution Impacts on Sediment Contaminant Concentrations within National Estuarine Research Reserves of the Southeastern US and Selected Chesapeake Bay Watersheds

July, 2018

William L. Balthis[1] and David L. Eslinger[2]

[1]NOAA National Ocean Service, National Centers for Coastal Ocean Science, Charleston, SC

[2]NOAA National Ocean Service, Office for Coastal Management, Charleston, SC

United States Department	National Oceanic and Atmospheric Administration	National Ocean Service
Wilbur L. Ross, Jr. Secretary of Commerce	RDML Timothy Gallaudet, Ph.D., USN Ret. Assistant Secretary of Commerce and Acting Under Secretary	Russell Callender, Ph.D. Assistant Administrator

Table of Contents

List of Figures

List of Tables

1 Introduction

Between 1970 and 2010, the number of people living in coastal watershed counties of the U.S. increased from 112.9 million to 163.8 million people, an increase of 45 % (NOAA 2013). Rising populations lead to increased impervious surface area in coastal watersheds and higher volumes of urban nonpoint-source (NPS) pollution resulting from stormwater runoff from diffuse sources, such as streets, parking lots, and construction sites (EPA 2015). Because of the diffuse nature of nonpoint sources, it is difficult to assess the impacts of this type of pollution at any particular location. Modeling and simulation approaches provide a means of estimating runoff and expected pollutant loads at the local and watershed level, while field-derived measurements of estuarine sediment pollutant concentrations can be used in calibration and parameter estimation in such models. In this study, we use an open-source version of the Nonpoint Source Pollution and Erosion Comparison Tool (OpenNSPECT, Eslinger et al. 2012) and measured pollutant concentrations to evaluate relationships between land-use modeling of NPS pollution runoff and observed patterns of ecosystem health measured in NOAA's National Estuarine Research Reserve System (NERRS) and in selected non-NERRS estuaries in the southeastern U.S. and Chesapeake Bay.

OpenNSPECT is a raster-based GIS model that estimates runoff, pollution and erosion in a GIS setting. It was initially released as the Nonpoint-Source Pollution and Erosion Comparison Tool (N-SPECT), an Esri® extension, for ArcView 9.0, in 2004. In 2011, an open-source version of the tool, OpenNSPECT, was released as a plugin running in the MapWindow open-source GIS environment. OpenNSPECT is designed to allow users to estimate the relative impact of changes in land use and/or environmental conditions. Generally, it is used as a screening tool, comparing two different land use scenarios to examine the differences between nearly identical model runs. This allows users to understand the potential impact of making land use/land cover changes, even if the absolute concentrations estimated by OpenNSPECT are not accurate. Hence the classification as a "screening tool:" a tool to help users identify areas of concern that will need additional study.

However, many users have employed OpenNSPECT (and N-SPECT before it) to estimate the actual concentrations of pollutants. This may be appropriate in some environments, in which the assumptions made by the model are not violated severely. In this analysis, we will test OpenNSPECT to see how well it can be used to estimate sediment pollutants in a variety of coastal environments, an application which has not been rigorously tested previously.

To have a meaningful test of the accuracy of the model in estimating pollutant values, it is necessary to derive local pollutant coefficients for pollutants of interest. As part of this research, we will be attempting to develop a standard method that others can use to do this. As a future project, OCM will be creating more guidance and "how-to" documentation for users on the methods developed in this study.

NOAA's National Centers for Coastal Ocean Science (NCCOS) has conducted a series of regional ecological assessments throughout many estuarine and coastal shelf areas of the southeastern U.S.

aimed at evaluating condition of living resources and ecosystem stressors in these coastal areas, including NERRS. Information from these studies is intended to provide a means of assessing the current status of ecological condition and stressor impacts throughout these areas, and to serve as a baseline for evaluating future changes due to natural or human-induced disturbances (Balthis et al. 2012). Additionally, NOAA's Cooperative Oxford Laboratory (COL), along with a team of federal, state, and academic partners, has conducted regional assessments of land use, water and sediment quality, and aquatic animal health in watersheds of Chesapeake Bay (Leight et al. 2014). At the state level, the South Carolina Estuarine and Coastal Assessment Program (SCECAP) was established in 1999 to begin evaluating the overall health of the state's estuarine habitats on a periodic basis using a combination of water quality, sediment quality, and biotic condition measures (Van Dolah et al. 2013). SCECAP is a collaborative program led by two state agencies, the South Carolina Department of Natural Resources (SCDNR) and the South Carolina Department of Health and Environmental Control (SCDHEC), as well as NOAA's NCCOS Charleston lab and the Hollings Marine Laboratory (HML).

This project aims to evaluate relationships between modeled levels of water runoff, suspended or dissolved pollutants, and observed pollutant levels in receiving water sediments. The basic assumptions required in this modeling approach are that 1) pollutants are related to land cover type; 2) the concentration of pollutants from a given land cover can be represented by a single value, the event mean coefficient, which can be thought of as the average concentration over the course of all rainfall events that produce runoff; 3) pollutants are carried conservatively downstream, where they are precipitated out and found in the sediments; 4) there are no significant sources of sediment pollutants other than those coming down from upstream lands.

2 Methods

2.1 Study Areas

Initially, six potential study areas were identified: three reserves within the National Estuarine Research Reserve System (NERRS) and three smaller watersheds in the Maryland portion of the Chesapeake Bay (Figure 1). As the study progressed, however, it became clear that the large NERRS sites were particularly poor candidates for this work due to the large expanses of salt marsh between the estuarine sampling sites and potential land-based pollutant sources. Therefore, we eventually dropped the Sapelo Island and North Inlet/Winyah Bay NERRS, and instead added Charleston, SC estuarine monitoring sites, where sampling sites were located closer to land. Ultimately, we analysed watershed runoff, potential pollutant loading and sedimentary pollutants in five watersheds: the ACE Basin NERR and Charleston Harbor (Figure 2) and the Magothy River, Corsica River, and Rhode River estuaries in the Chesapeake Bay, MD (Figure 3).

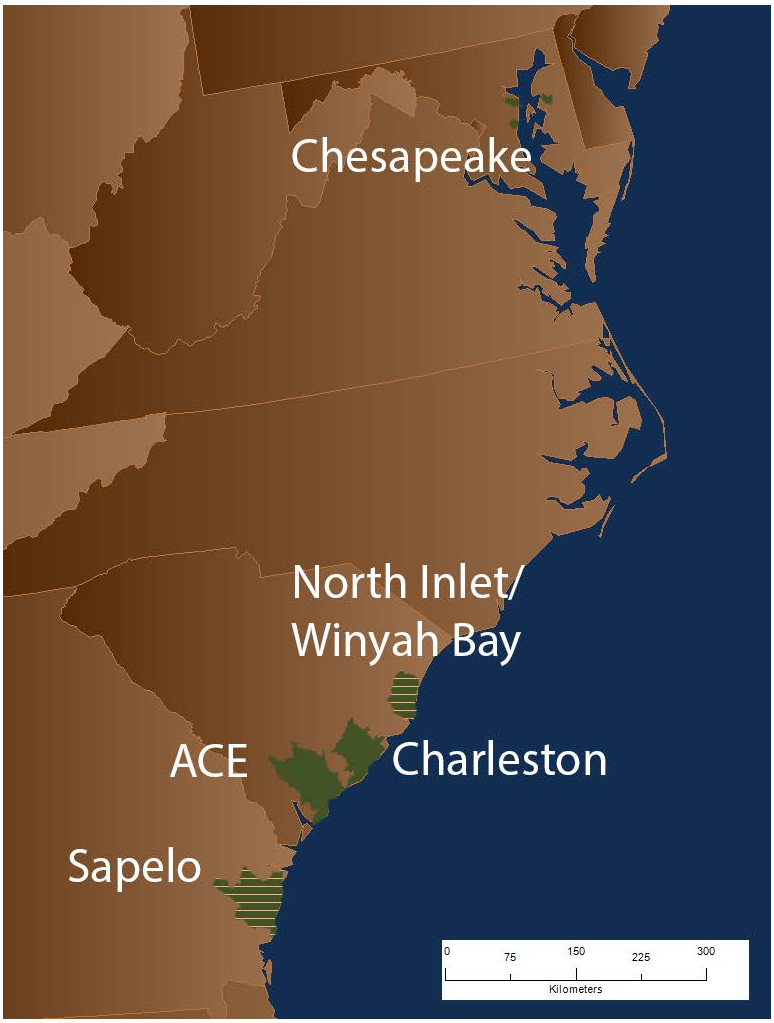

Figure 1. Southeastern watersheds considered for inclusion in this study.

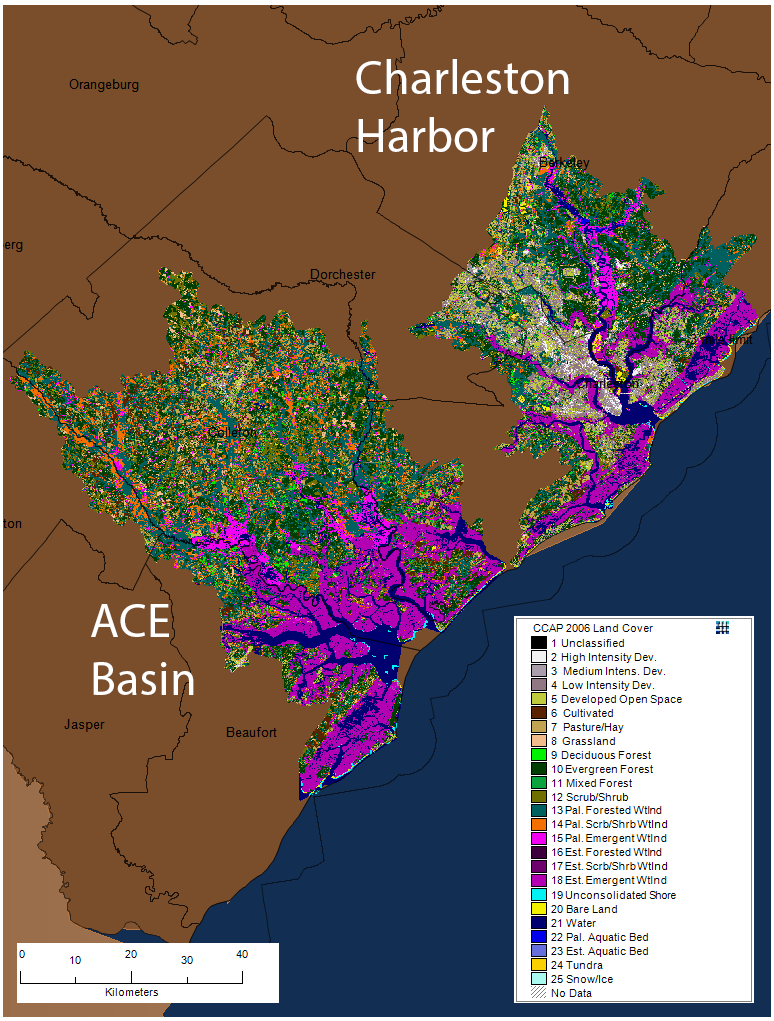

Figure 2. Watersheds in the Charleston area included in this study.

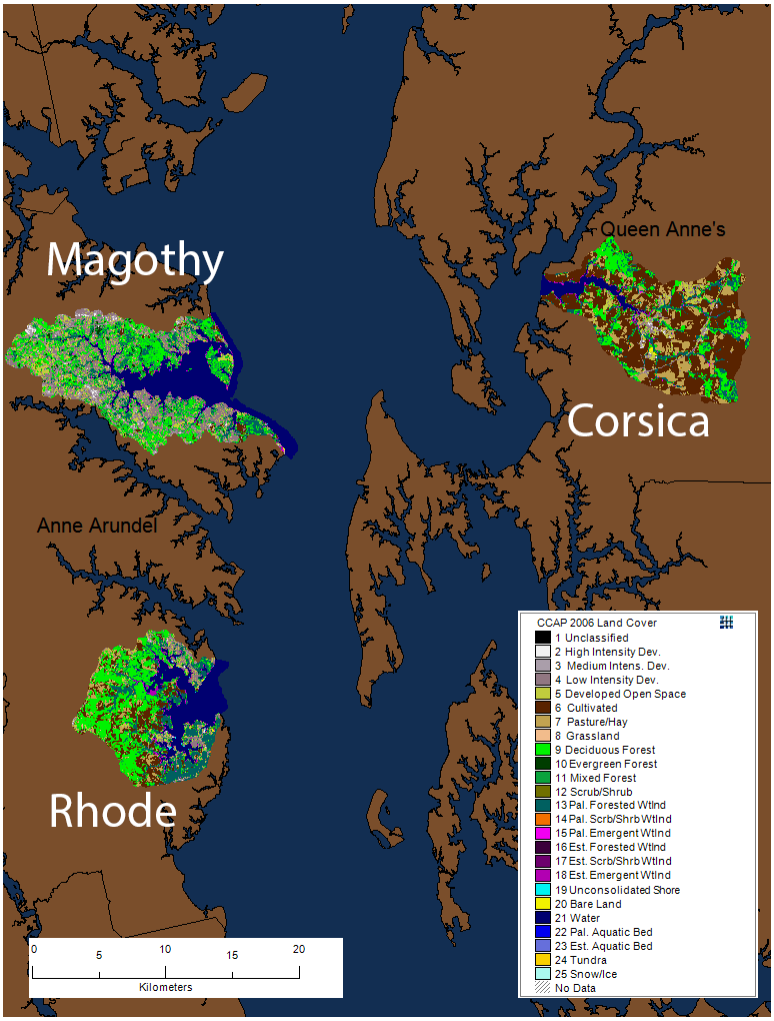

Figure 3. Watersheds in Chesapeake Bay included in this study.

2.2 Data Sources

OpenNSPECT requires up to five data inputs: classified land cover, digital elevation, precipitation over the time period to be modeled, rainfall erosivity (if examining erosion), and soil characteristics pertaining to permeability and erodibility. Fortunately, all of these data sets are readily available from Federal data archives. The resolution used here (30 m) matches that of the Thematic Mapper-derived classified land cover available from the Office for Coastal Management's Coastal-Change Analysis Project (C-CAP). The C-CAP data are very similar to, and are used as the base data for the coastal component of, the National Land Cover Database (NLCD) data produced by the federal Medium Resolution Land Characteristics Consortium (Homer et al. 2015). However, the C-CAP data have wetlands divided into additional freshwater and estuarine categories, which makes it a more suitable data set for this

investigation. C-CAP data were downloaded by state from the NOAA OCM Digital Coast (Digital Coast Data Registry).

Elevation data are available from the National Map web site (National Map) and were downloaded in 1x1 degree IMG-formatted files for areas of interest. Elevation data are available at a variety of spatial resolutions in unprojected, geographic coordinates. We used the one arc-second data, which are approximately 30-meter resolution for much of the US. Each 1x1 degree DEM is approximately 50MB in size. Grids were mosaicked together as needed before being projected and clipped to the respective areas of interest.

Precipitation values for 30-year climatological normals were taken from the PRISM Climate Group at Oregon State University (PRISM). These data are available in 800-meter and 4-km horizontal resolution. We used the 1981-2010 30-year normal data at 800-meter resolution.

Rainfall erosivity data are available from the NOAA Office for Coastal Management in a raster format. However, since erosion was not considered in the present analysis, erosivity data were not used.

Soils data are available by county in a shapefile and related database format from the USDA Natural Resources Conservation Service Geospatial Data Gateway (Geospatial Data Gateway). The data used here were from the Soil Survey Geographic Database (SSURGO data) and consisted of shapefiles identifying different soil types and associated data tables with a wealth of details about each soil polygon. OpenNSPECT requires only Hydrologic Soil Group and K-factor information.

The various data sets all have different projections, resolutions, and spatial extents. The necessary DEM and soils data generally required multiple data sets. These data sets were combined in the original projection and resolution to cover the needed areas. The rasters (land cover, elevation, and precipitation) were reprojected, binned to the same cell size, and clipped to the area of interest polygon using an OpenNSPECT data preparation module. The soils data, in shapefile format, were linked with the appropriate data tables and rasters were produced for the needed K-factor and hydrologic soils group.

Measurements of in-situ sediment pollutant concentrations were derived from three separate but related sampling programs. Data for southeastern NERRs and neighboring non-NERRs estuaries were generated from a series of ecological assessments conducted by NCCOS. Additional data for South Carolina estuaries, including SC NERRS sites (ACE Basin), were provided by SCECAP. Data for Chesapeake Bay were obtained from the NOAA/NCCOS Cooperative Oxford Laboratory (COL), which conducted a multi-year assessment of land use, water and sediment quality, and aquatic animal health in several small watersheds of the Chesapeake Bay. Each of these efforts used similar sampling and analytical protocols described in the respective reports (Balthis et al. 2012, Leight et al. 2014, Van Dolah et al. 2013). The data used in this study can be accessed through the NCCOS Regional Ecological Assessments website (Regional Ecological Assessments) or through NOAA's Office for Coastal Management (OCM) Digital Coast data registry (Digital Coast Data Registry).

2.3 Pollutant Coefficient Development

Pollutant coefficients were developed using multiple linear regression, by regressing field-measured sediment concentrations of total polycyclic aromatic hydrocarbon (PAH) against the accumulated runoff values for the various land cover types located upstream of sampling sites. A similar approach has been applied previously to develop regional pollution export coefficients for total nitrogen (Schenk et al. 2008). In order to do this, local topographic data were used to derive watershed areas. Stream centerlines were burned into the topographic data as necessary, generally in the open marsh and bay areas. Once the watersheds and flow direction had been calculated based on the topography, it was possible to estimate the areas that contributed water to a particular sampling point.

Because of the artificial way in which water flows in the model, upstream contributing area calculations only create relatively realistic drainages when given pour points that are on the stream centerline. Since the actual field sampling locations did not always correspond to the topographically-derived stream centerlines, some sampling points needed to be moved to the nearest stream centerline location (Figure 4). This movement was done based on the relationship of the sampling point and the presumed local flow. However, because the sampling points were in tidal marshes, some with interconnecting channels, we recognize that there could be multiple “reasonable” places to move a sampling point. The guiding principle used in this analysis was to move points to the closest streamline without crossing a terrestrial CCAP land cover class. Sampling points were dropped from the analysis if they were more than about 500 m away from the DEM-derived stream flow line.

Overall, assuming some horizontal homogeneity, and given the expected tidal mixing in these estuarine waters, the relocation and alignment of sampling points to stream centerlines in order to retain as many sites as possible for analysis provided a reasonable starting point for these simulations. It is not a perfect solution, but a reasonable one, given the artificial nature of our flow regime.

Once sampling points were relocated to stream locations, they were used as pour points for defining the upstream watersheds that contributed runoff to each sampling point. Because sampling points were frequently upstream and downstream from others, some of these watersheds formed a nested set (i.e., a larger watershed with a pour point far downstream may contain smaller watersheds and sampling points within it).

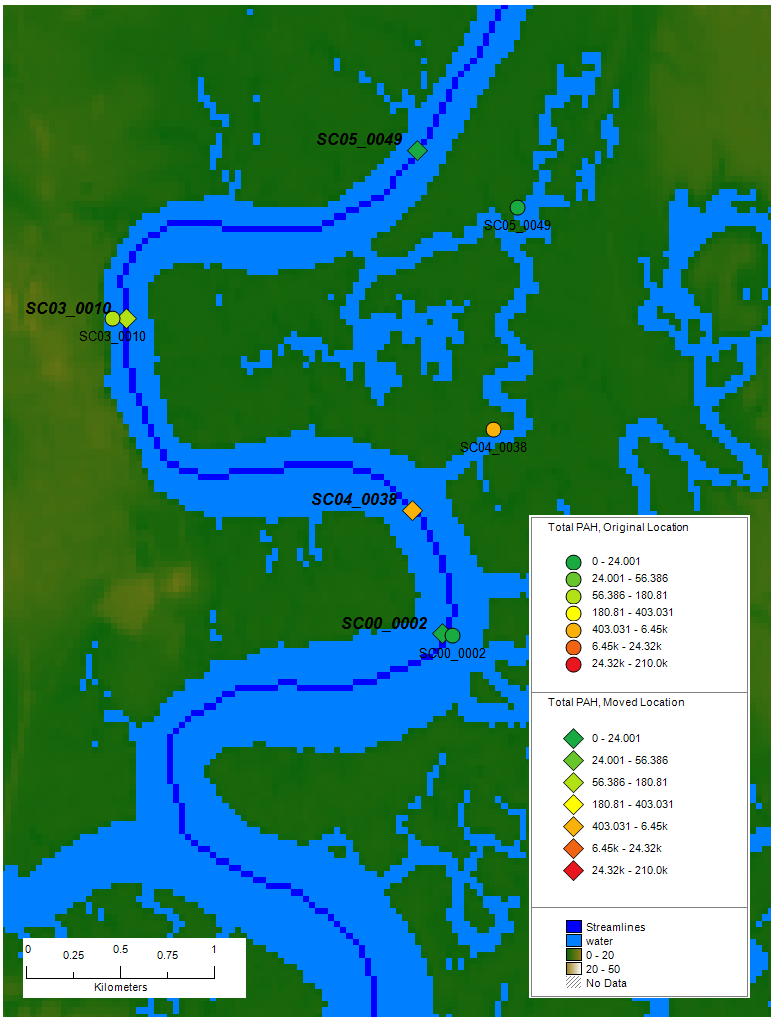

Figure 4. Example illustrating relocation of sampling points to topographically-derived stream centerlines. Original sampling locations are indicated by filled circles; relocated points along stream centerlines are indicated by filled diamond symbols.

The watershed data and the 30-year annual precipitation normals were used as inputs to an initial OpenNSPECT run which produced only local runoff and accumulated runoff output estimates. Using watershed polygons, the runoff produced at each grid cell, and the C-CAP land cover within each polygon, the total contributing runoff by land cover class was calculated at each sampling point. The calculated runoff volume, the counts of each land cover class, and the pollutant sampling data were combined into one data table. The resulting table contained the total amount of different pollutants at the bottom of each watershed, the number of cells for each land cover type, the total runoff from each land cover type, and the total runoff and total area of the watershed. As mentioned previously, there were many nested sets of sampling points/watersheds. However, since they contained different amounts of different land covers, these nested watersheds were considered independent in the rest of the analysis.

Since the present study is intended to understand how well a simple surface runoff model such as that used in OpenNSPECT can be used to predict benthic sediment concentrations, we performed multiple regressions of contaminant concentrations on accumulative land cover cell counts and on accumulated runoff. The estimated regression coefficients represent the pollutant loadings from one grid cell of each respective land cover class, and hence were used as pollutant coefficients in OpenNSPECT.

2.4 Land Cover Proximity Analysis

As the study progressed, it became clear that relating surface runoff to accumulated sediment concentrations was not particularly well-modeled with an upstream-flow model that ignored deposition or any dynamics outside of the cell of origin. In considering other methods to look at land use, we tested a simple model which used the number and relative distribution of different land cover types within a circular proximity of the sampling points. Given the relatively large tidal signal present in all of our study areas, this seemed to be a better approach than the surface flow-driven OpenNSPECT model.

To perform the proximity analysis, land cover was analyzed iteratively for sets of concentric buffered circles around each sampling point. Linear regressions of log-transformed sediment pollutant concentrations versus land cover counts were performed for each buffer distance. Plots of correlation versus distance for each study location were used to help assess the potential influence of buffer distance on the spatial relationship of land cover and pollutants.

3 Results

The initial regression results did not suggest a significant relationship between total PAH versus land cover-specific runoff in the ACE Basin NERR, and the value of R^2 indicated a poor fit of the model. Since the C-CAP data are classified into a number of related land cover types, we tried different grouping methods to see if the poor results might simply be an artifact of subdividing related classes. We therefore looked at four different levels of classification: Ungrouped (19 classes), Group A (5 classes), Group B (6 classes), and Group C (7 classes). Land cover types included in the different groupings are shown in Table 1. Unfortunately, none of the alternate groupings worked well for the ACE Basin (Table 2).

In working with the ACE Basin data, we observed that most of the sampling points were quite distant from land (Figure 5). Given the approximately two-meter tidal range in this area, we felt that the distance between the assumed pollutant sources on land, in combination with the complicated tidally-dominated flow, might be invalidating the main model assumption of a direct causal link between observed pollutant concentration and upstream land cover. The concept of “upstream” breaks down in these tidally-dominated waters.

Table 1. Land cover classifications used in the present study.

Ungrouped (19 classes)	Group A (Five classes)	Group B (Six classes)	Group C (Seven classes)
LC_2: High Intensity Developed	Urban	Urban	Urban
LC_3: Medium Intensity Developed	Urban	Urban	Urban
LC_4: Low Intensity Developed	Urban	Urban	Urban
LC_5: Developed Open Space	Urban	Urban	Urban
LC_6: Cultivated Land	Cropland	Farmed	Cropland
LC_7: Pasture/Hay	Pasture	Farmed	Pasture
LC_9: Deciduous Forest	Forest	Uplands	Uplands
LC_10: Evergreen Forest	Forest	Uplands	Uplands
LC_11: Mixed Forest	Forest	Uplands	Uplands
LC_8: Grassland	Other	Uplands	Uplands
LC_12: Scrub/Shrub	Other	Uplands	Uplands
LC_13: Palustrine Forested Wetland	Other	Lowlands	Lowlands
LC_14: Palustrine Scrub/Shrub Wetland	Other	Lowlands	Lowlands
LC_16: Estuarine Forested Wetland	Other	Lowlands	Lowlands
LC_17: Estuarine Scrub/Shrub Wetland	Other	Lowlands	Lowlands
LC_15: Palustrine Emergent Wetland	Other	Marsh	Marsh
LC_18: Estuarine Emergent Wetland	Other	Marsh	Marsh
LC_19: Unconsolidated Shore	Other	Mudflats	Mudflats
LC_20: Bare Land	Other	Mudflats	Mudflats

Note: Water classes (LC_21: Water and LC_22: Palustrine Aquatic Bed) were excluded from all analyses

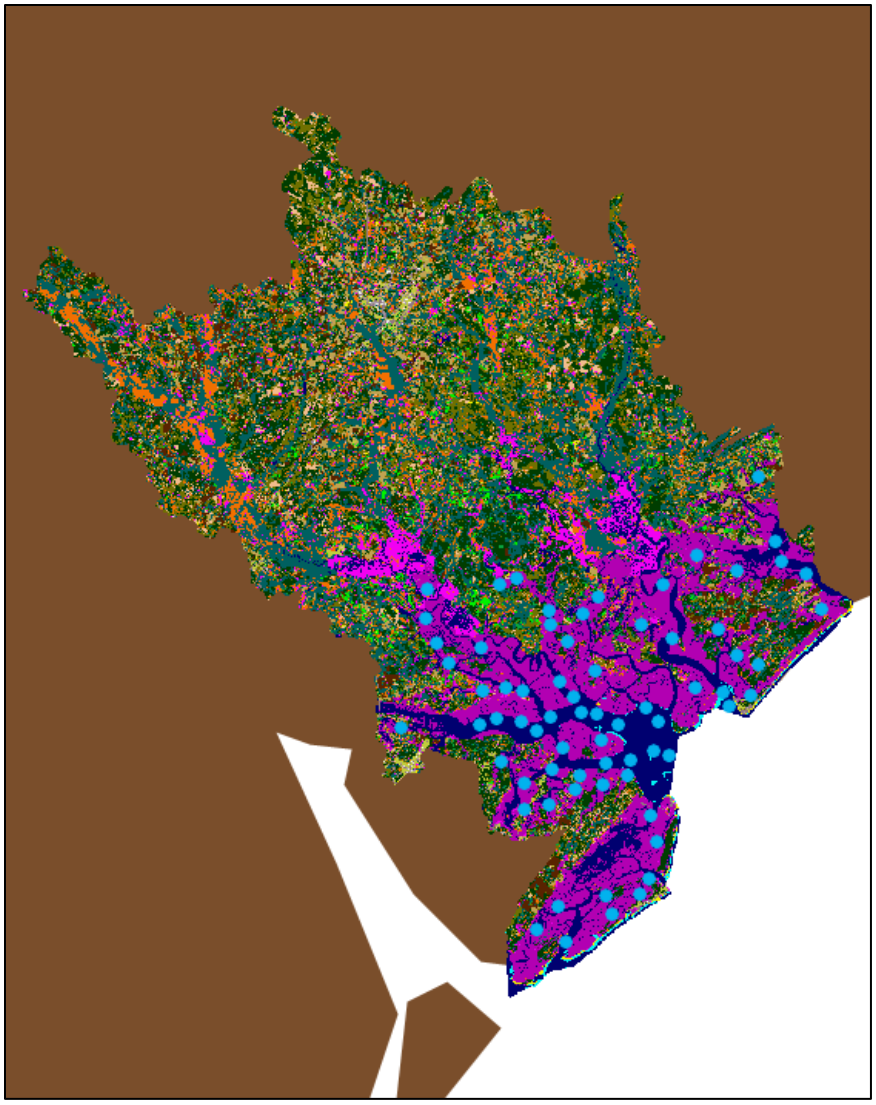

Figure 5. Locations of sampling sites in the ACE Basin NERR.

Data were available for Charleston Harbor, which also has a similar tidal regime, but has many of the sampling locations in much closer proximity to land (Figure 6). Unfortunately, because the low topography and relatively dense network of rivers and creeks in the Charleston area, we could not use all of the sampling points since we needed sites to be located at the pour point of a watershed (i.e., where the water leaves the watershed). The geographic technique used in OpenNSPECT defines a watershed as the area that contributes accumulated runoff into a defined stream. There is a threshold value set for how many grid cells must be accumulated to define a stream. Due to the low topography in the area, there is not much flow accumulation driven by elevation, and because of the network of water bodies the length of any potential stream can be limited. Therefore, after the Charleston runoff calculations, only 49 of the 119 possible sampling points ended up as "pour points." The excluded points are shown in red, and the included points in blue, in Figure 6.

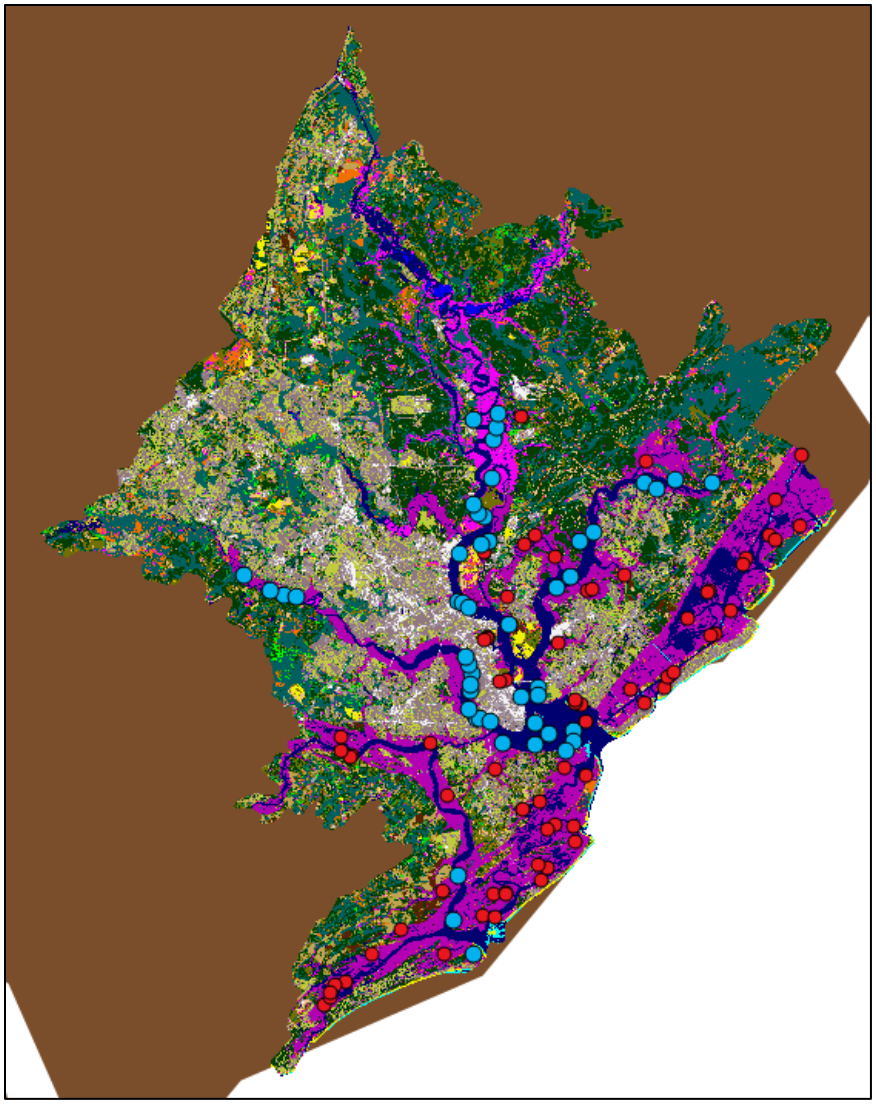

Figure 6. Sampling sites retained (blue circles) or excluded (red circles) for use in OpenNSPECT modeling runs in Charleston, SC watersheds.

Even with the smaller data set, however, there was a significant association between sediment total PAH concentrations and both the simple class counts and runoff-weighted counts for Charleston Harbor land cover (Table 2). Regressions for all tested groupings were significant, with adjusted R^2 values ranging from 0.42 to 0.65. Interestingly, the highest adjusted R^2 value was in the completely ungrouped data. Complete regression results are given in Table 2.

Table 2. Regression results (overall model p-value and R^2) for total PAH for different geographies and land cover groupings.

Land Cover Groupings	ACE		Charleston Harbor		Chesapeake Bay	
	Runoff	**LC Counts**	**Runoff**	**LC Counts**	**Runoff**	**LC Counts**
Ungrouped (19 classes)	p = 0.69 R^2 = 0.25 R^2_{adj} = -0.06	p = 0.61 R^2 = 0.26 R^2_{adj} = -0.04	p<0.0001 R^2 = 0.79 R^2_{adj} = 0.65	p<0.0001 R^2 = 0.76 R^2_{adj} = 0.61	p = 0.21 R^2 = 0.54 R^2_{adj} = 0 .16	p = 0.31 R^2 = 0.50 R^2_{adj} = 0 .10
Group A (5 classes)	p = 0.41 R^2 = 0.08 R^2_{adj} = 0.00	p = 0.46 R^2 = 0.08 R^2_{adj} = -0.01	p<0.00001 R^2 = 0.51 R^2_{adj} = 0.45	p<0.00001 R^2 = 0.55 R^2_{adj} = 0.50	p = 0.53 R^2 = 0.11 R^2_{adj} = -0.02	p = 0.66 R^2 = 0.09 R^2_{adj} = -0.04
Group B (6 classes)	p = 0.41 R^2 = 0.10 R^2_{adj} = 0.00	p = 0.51 R^2 = 0.09 R^2_{adj} = -0.01	p<0.0001 R^2 = 0.49 R^2_{adj} = 0.42	p<0.0001 R^2 = 0.50 R^2_{adj} = 0.43	p = 0.67 R^2 = 0.11 R^2_{adj} = -0.05	p = 0.86 R^2 = 0.07 R^2_{adj} = -0.09
Group C (7 classes)	p = 0.41 R^2 = 0.12 R^2_{adj} = 0.01	p = 0.50 R^2 = 0.10 R^2_{adj} = -0.01	p<0.0001 R^2 = 0.50 R^2_{adj} = 0.42	p<0.0001 R^2 = 0.54 R^2_{adj} = 0.46	p = 0.62 R^2 = 0.14 R^2_{adj} = -0.04	p = 0.88 R^2 = 0.08 R^2_{adj} = -0.11

The statistically significant regression coefficients for the individual classes from the runoff-weighted results were used as OpenNSPECT pollutant coefficient values. Pollutant coefficients were set to zero for non-significant land cover classes. Results of the OpenNSPECT run are shown in Figure 7 and Figure 8. Figure 7 shows the accumulated runoff over the entire study area and Figure 8 shows a detailed view of the area surrounding the Charleston peninsula. In Figure 8, the accumulation is seen best in the streamlines running down the center of the estuaries. Most of the land cover area appears the same salmon color, since there is very little accumulation outside of the streams.

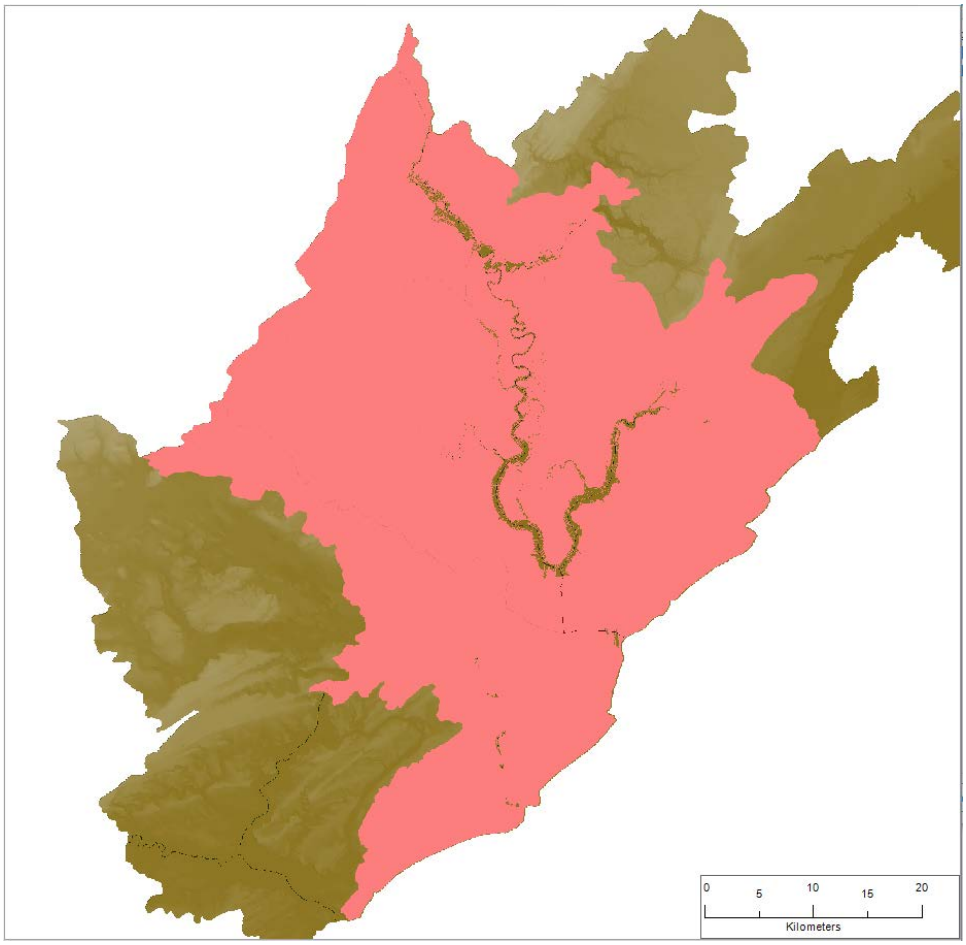

Figure 7. Accumulated runoff obtained from OpenNSPECT runs using pollutant coefficients derived from multiple linear regression of total PAH concentrations versus runoff-weighted land cover counts in Charleston, SC watersheds. Salmon-colored land cover areas represent low (near zero) accumulated runoff. Most of the accumulated runoff occurs along stream flow centerlines.

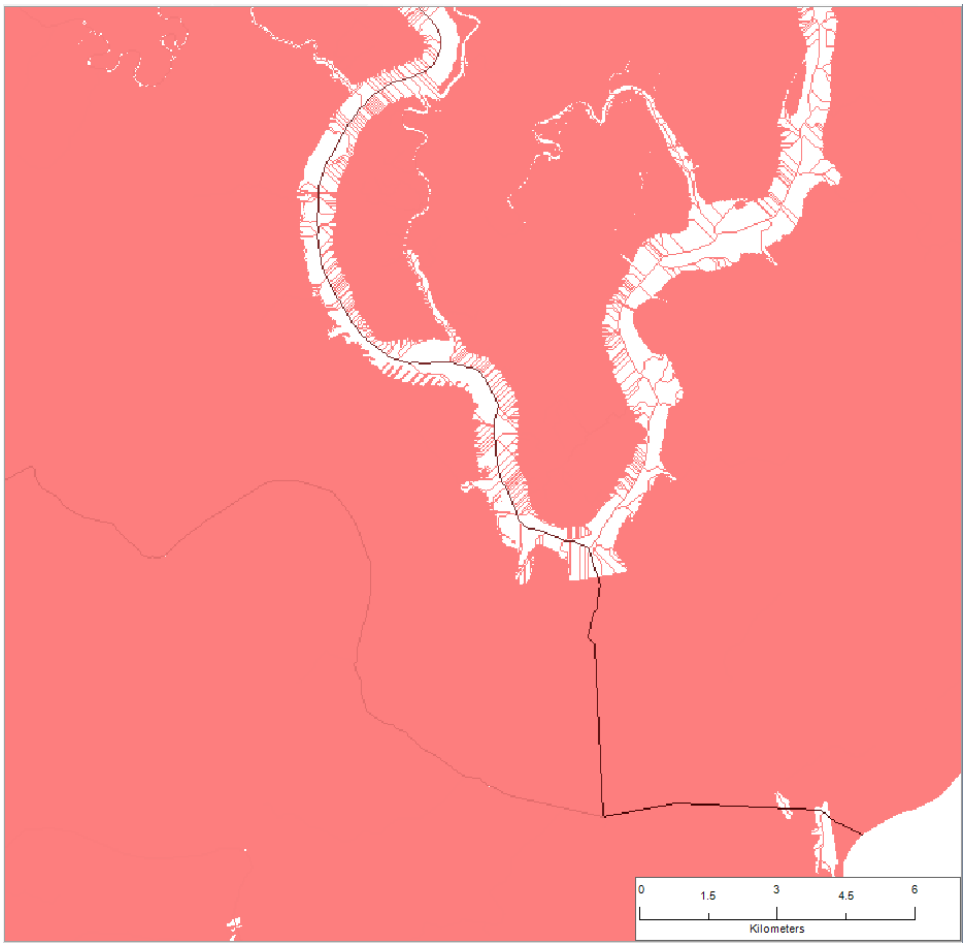

Figure 8. Detailed view of Figure 7, showing accumulated runoff obtained from OpenNSPECT using coefficients from multiple linear regression of total PAH concentrations versus runoff-weighted land cover counts as pollutant coefficients. Salmon-colored land cover areas represent low (near zero) accumulated runoff. Most of the accumulated runoff occurs along stream flow centerlines.

Figure 9 shows the results of extracting the point estimates from the modeled results compared to the initial field measurements. Although there is some modest amount of agreement, it is clear that the model does a poor job at estimating the field concentrations. In looking at Figure 9, there are a couple of observations that help understand the poor fit to a 1:1 line. The predicted PAH concentrations have a bimodal distribution with a clear separation between predicted concentrations above and below 100 ng/g. The higher predicted concentrations all correspond to the main flowline of the Cooper River, as indicated in the darker red observations in Figure 10. The highest field-measured concentrations generally occur in the Ashley River, and especially for those which have predicted values less than 100 ng/g. This is seen in Figure 11, which shows the C-CAP land cover classes and field measured values at both the points they moved to on the imposed flow (the red circles) and the corresponding actual location (the red squares). More discussion of the relevance and interpretation of this figure is provided in the discussion section.

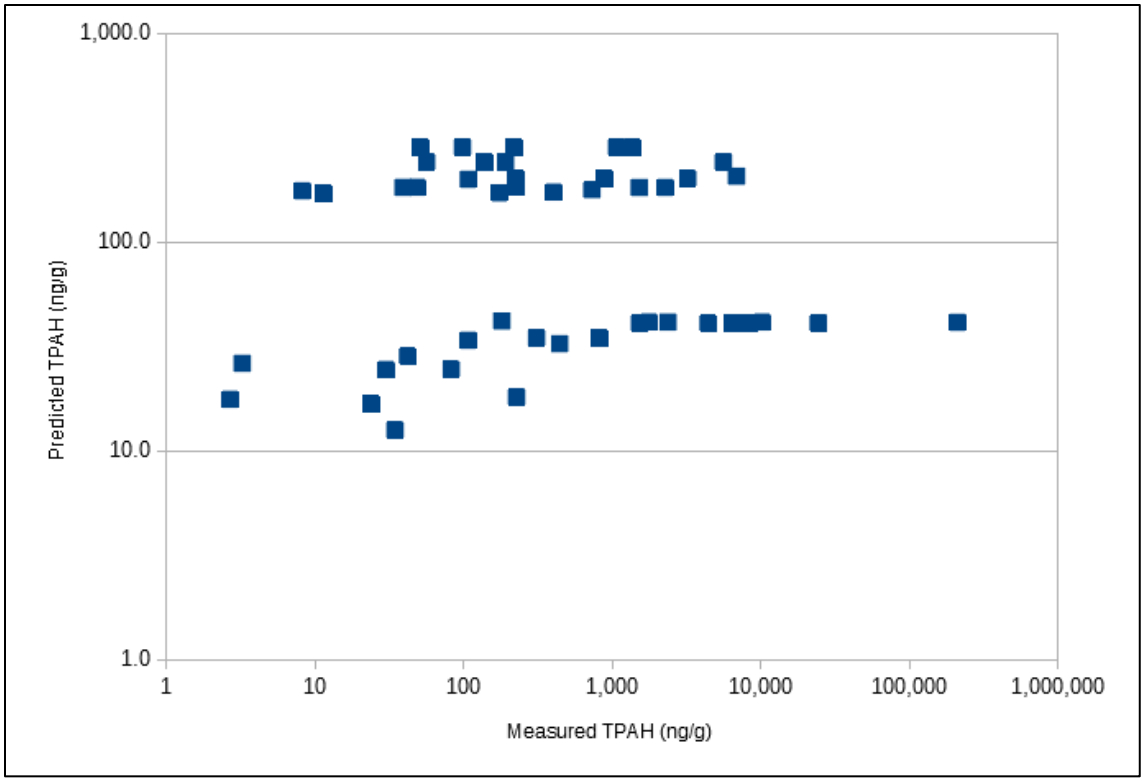

Figure 9. Predicted versus measured total PAH (TPAH) concentrations. Predicted values were obtained from a multiple linear regression of total PAH versus runoff-weighted land cover counts for Charleston, SC watersheds.

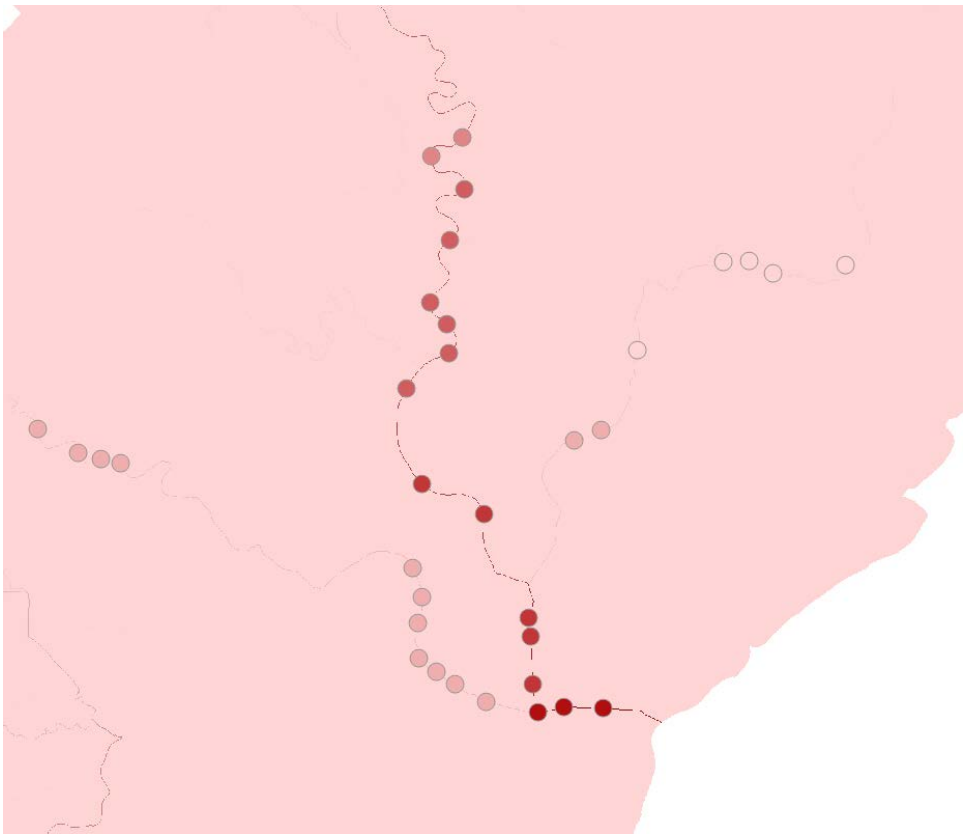

Figure 10. Predicted total PAH concentrations along stream flow lines in Charleston, SC watersheds (darker symbols represent higher concentrations).

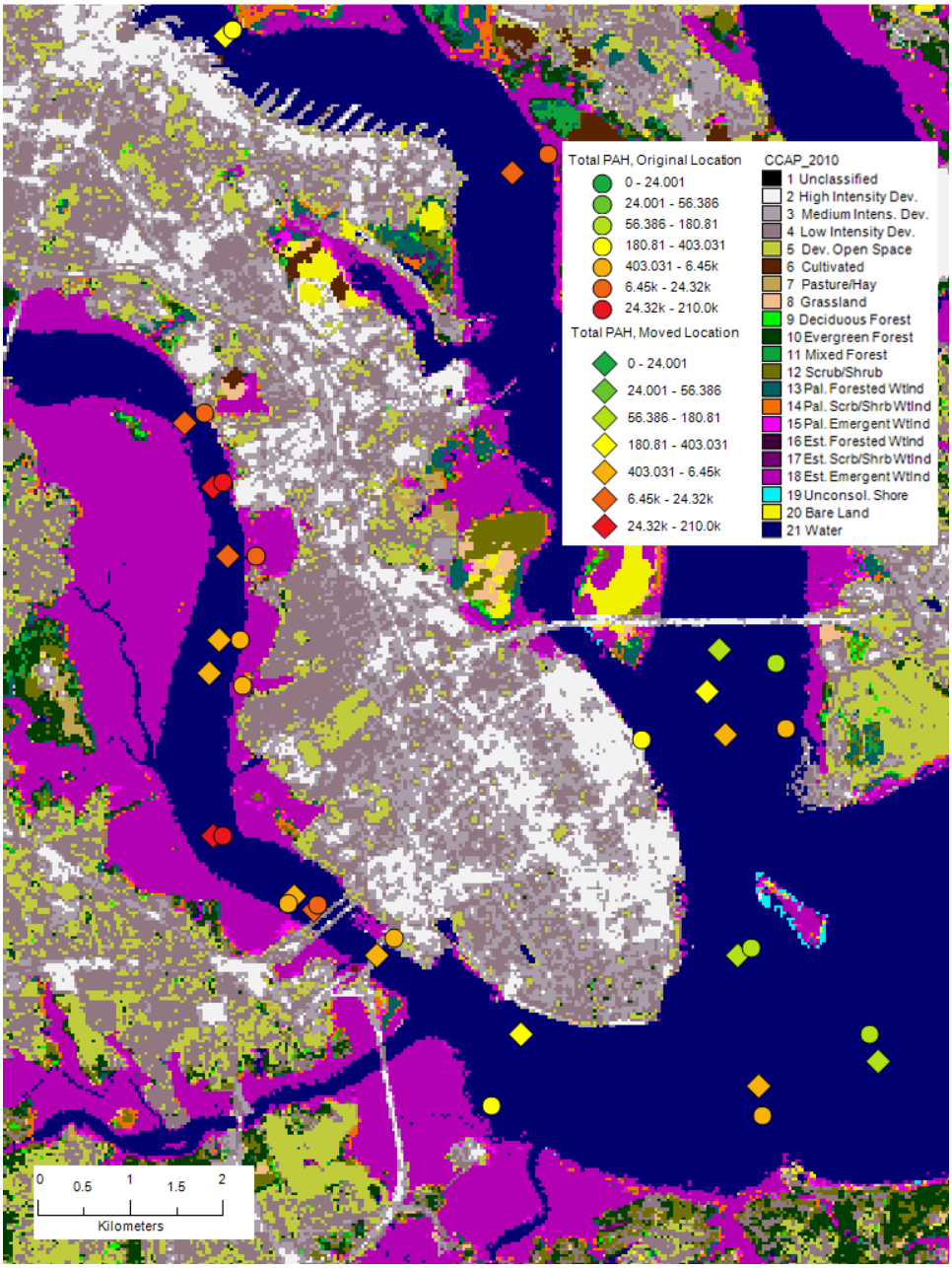

Figure 11. Locations of original field measurements (square symbols) of sediment total PAH and relocated positions (filled circles) along stream flow center lines in Charleston, SC watersheds. Lighter colors represent lower concentrations; darker symbols represent higher concentrations.

3.1 Land cover proximity analysis

Results of land cover proximity analysis varied by region. In regressions of log-transformed total PAH versus land cover counts (i.e., the number of 30x30 m grid cells of each land cover type) for the Charleston area, model R^2 increased with buffer distance initially, then tended either to level out or taper off after 5 - 10 km (Figure 12). All of the models tested were highly significant ($p << 0.05$) irrespective of buffer distance. In regressions using counts from ungrouped land cover classes, the significance of each of the 19 land cover types varied widely depending on the selected buffer distance (Figure 13), making it difficult to discern a consistent pattern of significance of any particular land cover type(s). This may be due, at least in part, to linear dependencies among the predictors (i.e., land cover types) in the multiple regression. For example, some of the developed land cover classes appeared to be linearly related (Figure 14), as were several forested and scrub/shrub land cover types (Figure 15). Calculation of a variance inflation factor (VIF), a widely used method of detecting the presence of multicollinearity, suggested linear dependencies among predictors (VIF > 10), particularly at buffer distances greater than about 2 km (Figure 16).

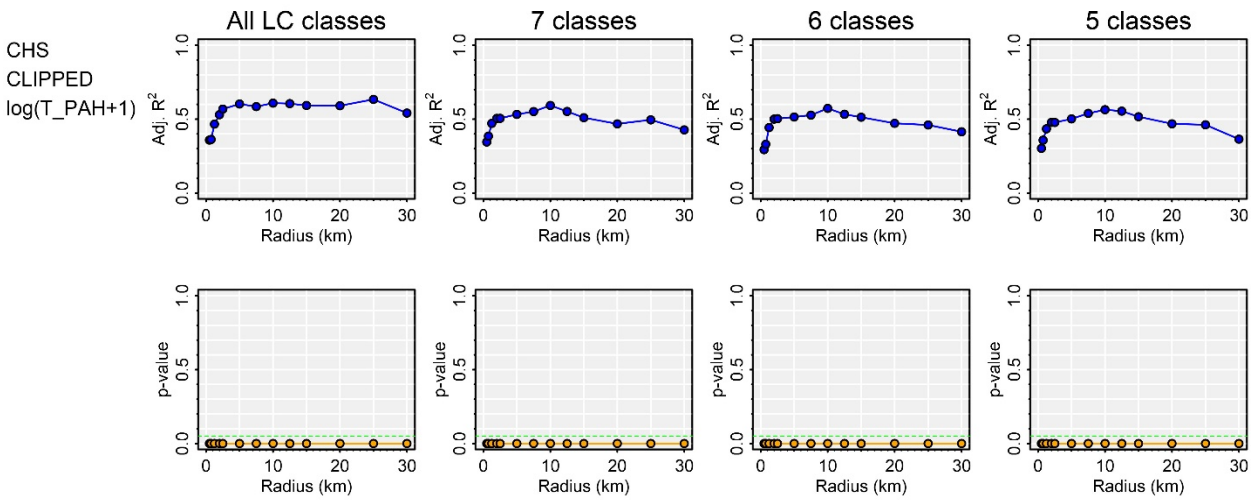

Figure 12. Model R^2 and p value from multiple regressions of total PAH versus land cover counts at varying buffer radii for Charleston, SC watersheds.

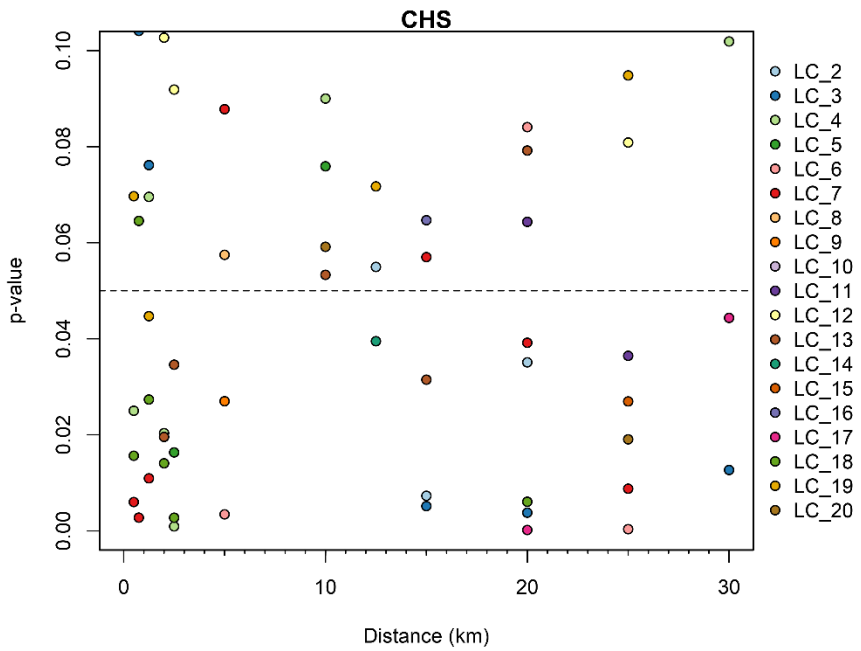

Figure 13. Individual p-values associated with each individual predictor (ungrouped land cover types) in multiple linear regressions of sediment total PAH versus land cover at varying buffer radii for Charleston, SC watersheds. See Table 1 for land cover classifications.

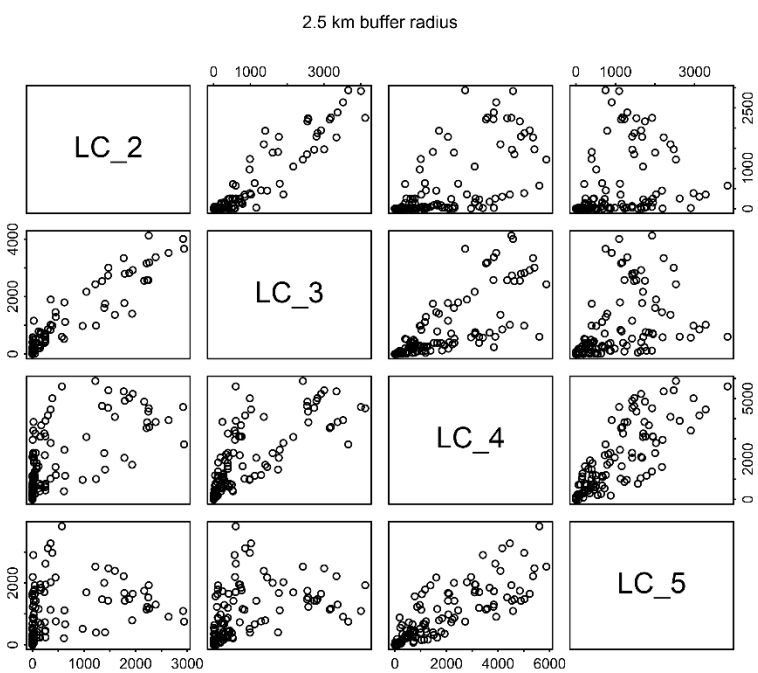

Figure 14. Dependencies among counts of 'developed' land cover classes for Charleston, SC watersheds (2.5 km buffer radius).

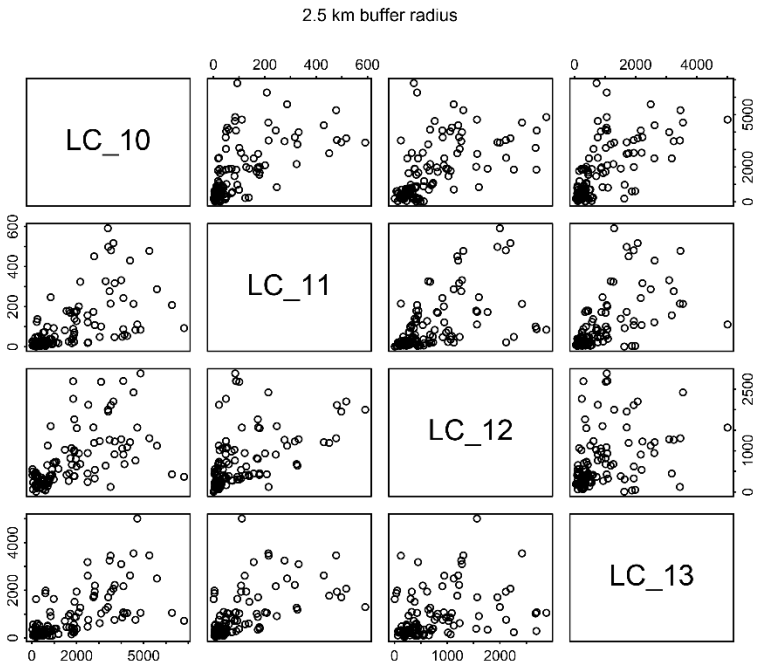

Figure 15. Dependencies among forested and scrub/shrub wetland types in Charleston, SC watersheds.

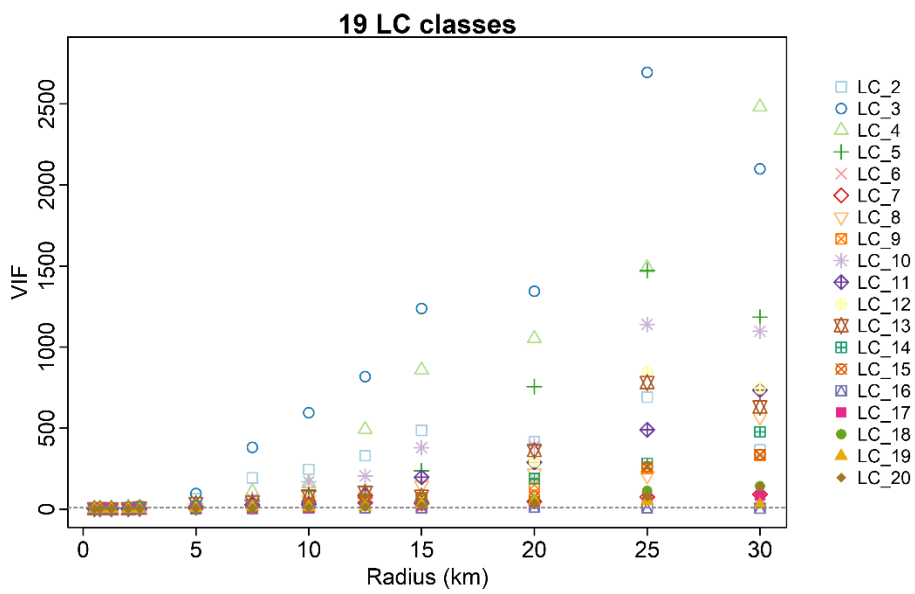

Figure 16. Variance inflation factors (VIFs) associated with multiple linear regressions of total PAH versus ungrouped land cover types at varying buffer radii for Charleston, SC watersheds. VIFs > 10 are suggestive of linear dependencies (multicollinearity) among predictors.

Aggregating land cover types to a smaller number of classes appeared to reduce the observed multicollinearity noted above. Figure 17 shows VIFs calculated for each of seven aggregated land cover types over the range of buffer distances tested. Compared with regressions using ungrouped land cover classes, VIFs at larger buffer distances were reduced by an order of magnitude, and most VIF values for radii < 15 km were typically less than 10. Aggregating further to six classes by combining the crop and pasture land cover types (i.e., farmed) also improved VIFs, although the difference between seven and six classes was small. Using only five land cover groups also resulted in lower VIF values at buffer distances from 2 - 15 km, but elevated values below 2 km suggested the presence of linear dependencies among classes combined as 'other'.

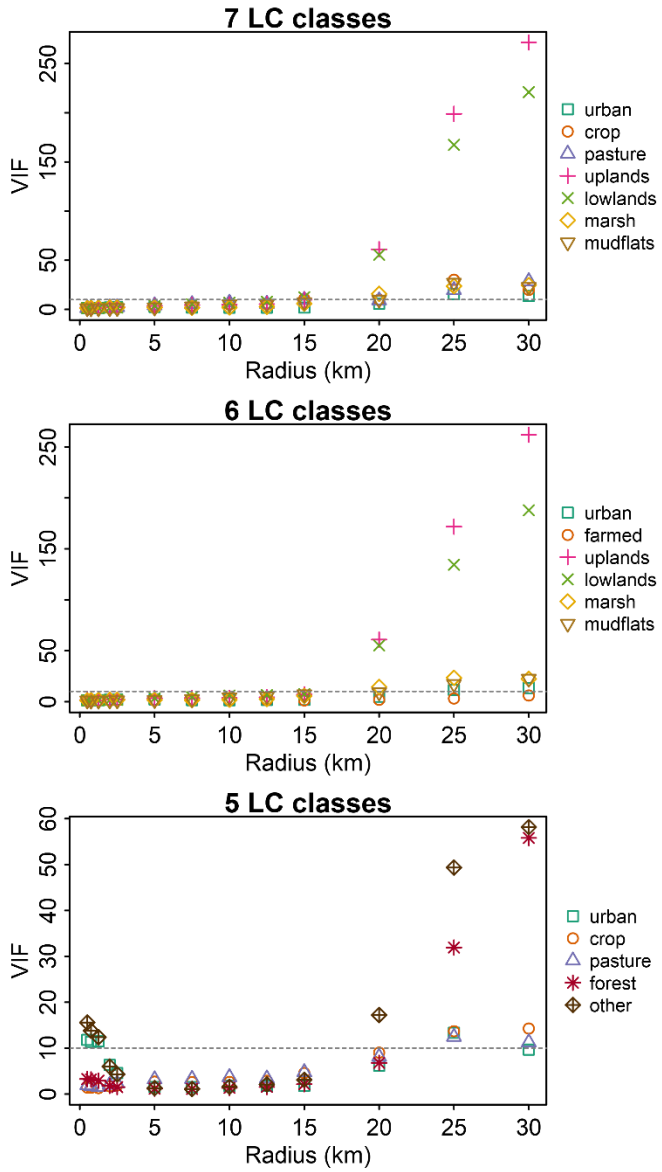

Figure 17. Variance inflation factors (VIFs) associated with multiple linear regressions of total PAH versus land cover aggregated to seven, six, and five categories for Charleston, SC watersheds.

In all Charleston area regressions, the urban land cover types were statistically significant consistently at buffer distances between 1.25 and 25 km. R^2 values exhibited an initial peak at around 2 - 2.5 km, then continued to increase to a second maximum before eventually tapering off (see previous Figure 15). The greatest level of agreement among all regressions (i.e., urban land cover types significant) was at 2 and 2.5 km. Using ungrouped land cover types and a buffer distance of 2.5 km, for example, the low-intensity developed land cover type (LC_4) was the most highly significant term, as well as developed open space (LC_5), palustrine forested wetland (LC_13), and estuarine emergent wetland (LC_18; Table 3). Similarly, at a buffer distance of 2.5 km, all of the Charleston area regressions using aggregated land cover types (7, 6, and 5 classes, respectively) found the urban land cover class to be highly significant ($p < 0.001$ in all cases; Table 4 - Table 6), with the marsh and upland types marginally significant (or significant at $\alpha=0.10$) in some regressions using aggregated land cover. The strong association between sediment total PAH levels and urban land cover types in the Charleston area can be seen in the bivariate plot of PAH vs. urban cell counts (Figure 18).

Table 3. Results of multiple regression (total PAH versus land cover) for the Charleston area using ungrouped land cover types and a buffer distance of 2.5 km. (Overall model R^2: 0.57; F-statistic: 9.16 on 19 and 99 DF; p-value < 0.0001.)

Effect	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	4.19345	0.91422	4.59	<0.0001
LC_2	0.00023	0.00077	0.30	0.7641
LC_3	0.00036	0.00064	0.57	0.5707
LC_4	0.00112	0.00033	3.41	**0.0009**
LC_5	-0.00107	0.00044	-2.44	**0.0163**
LC_6	0.00007	0.00060	0.11	0.9128
LC_7	-0.00061	0.00054	-1.14	0.2588
LC_8	-0.00258	0.00173	-1.49	0.1387
LC_9	-0.00615	0.00492	-1.25	0.2141
LC_10	0.00020	0.00022	0.90	0.3681
LC_11	0.00440	0.00284	1.55	0.1246
LC_12	0.00064	0.00038	1.70	0.0919
LC_13	-0.00087	0.00041	-2.14	**0.0346**
LC_14	0.00163	0.00163	1.00	0.3204
LC_15	0.00009	0.00013	0.68	0.4955
LC_16	-0.00718	0.00951	-0.75	0.4524
LC_17	0.00303	0.00286	1.06	0.2919
LC_18	-0.00025	0.00008	-3.07	**0.0027**
LC_19	-0.00089	0.00093	-0.95	0.3420
LC_20	0.00008	0.00080	0.10	0.9214

Table 4. Results of regression (total PAH versus land cover) for the Charleston area using land cover types aggregated to seven classes (urban, crop, pasture, uplands, lowlands, marsh, mudflats) and a buffer distance of 2.5 km. (Overall model R^2: 0.51; F-statistic: 18.23 on 7 and 111 DF; p-value < 0.0001.)

Effect	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.84576	0.90874	4.23	<0.0001
urban	0.00038	0.00006	6.03	**<0.0001**
crop	0.00047	0.00060	0.78	0.4353
pasture	-0.00080	0.00051	-1.58	0.1173
uplands	0.00019	0.00011	1.76	0.0813
lowlands	-0.00031	0.00020	-1.54	0.1270
marsh	-0.00015	0.00007	-1.95	0.0535
mudflats	-0.00002	0.00042	-0.04	0.9698

Table 5. Results of regression (total PAH versus land cover) for the Charleston area using land cover types aggregated to six classes (urban, farmed, uplands, lowlands, marsh, mudflats) and a buffer distance of 2.5 km. (Overall model R^2: 0.50; F-statistic: 20.89 on 6 and 112 DF; p-value <0.0001.)

Effect	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.81393	0.91073	4.19	0.0001
urban	0.00039	0.00006	6.17	**<0.0001**
farmed	-0.00023	0.00023	-1.02	0.3121
uplands	0.00017	0.00011	1.60	0.1121
lowlands	-0.00033	0.00020	-1.61	0.1102
marsh	-0.00014	0.00007	-1.90	0.0599
mudflats	0.00005	0.00041	0.12	0.9057

Table 6. Results of regression (total PAH versus land cover) for the Charleston area using land cover types aggregated to five classes (urban, crop, pasture, forest, other) and a buffer distance of 2.5 km. (Overall model R^2: 0.48; F-statistic: 22.56 on 5 and 113 DF; p-value < 0.0001.)

Effect	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.17492	2.03664	0.58	0.5652
urban	0.00052	0.00010	5.41	**<0.0001**
crop	0.00044	0.00062	0.71	0.4779
pasture	-0.00060	0.00053	-1.14	0.2554
forest	0.00016	0.00013	1.27	0.2074
other	0.00006	0.00010	0.60	0.5518

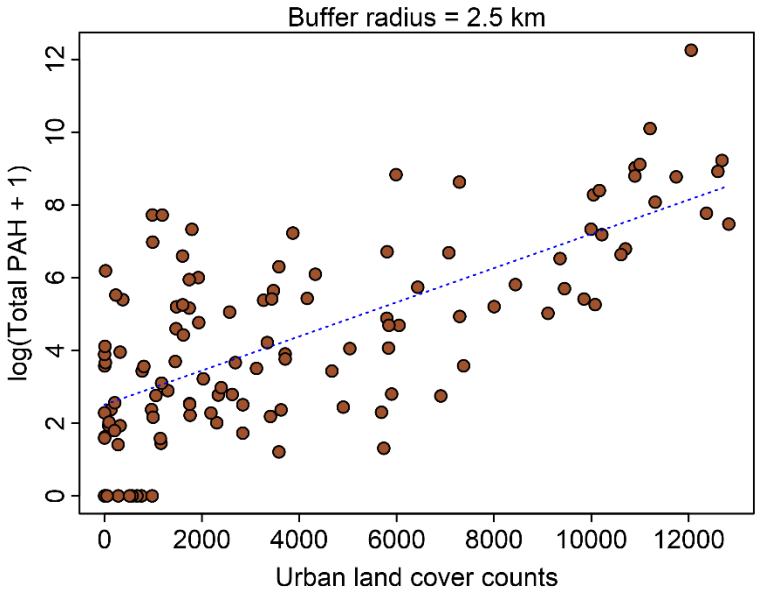

Figure 18. Plot of log-transformed total PAH concentrations versus urban land cover counts for Charleston, SC watersheds.

Results from regressions of total PAH vs. land cover counts for ACE Basin sites were highly variable with buffer distance. Model R^2 values were very low (typically <= 0.1) and most model p-values were non-significant (Figure 19). VIFs were suggestive of multicollinearity effects involving LC_3 and LC_4 (Medium- and Low-intensity Developed, respectively) and LC_13 and LC_14 (Palustrine Wetland: forested and scrub/shrub, respectively), particularly at larger buffer radii. In regressions using all land cover types (19 classes), VIFs for the medium and low-intensity developed classes were elevated even at smaller buffer distances (Figure 20). Aggregating to seven (or six) classes resulted in some improvement, although at larger buffer distances VIFs indicated multicollinearity effects involving uplands and lowlands (Figure 21). At buffer distances from 5 to 15 km VIFs were very low for regressions using land cover data aggregated to five classes, but at radii < 2 km VIFs were also elevated for the 'forest' and 'other' categories (Figure 21).

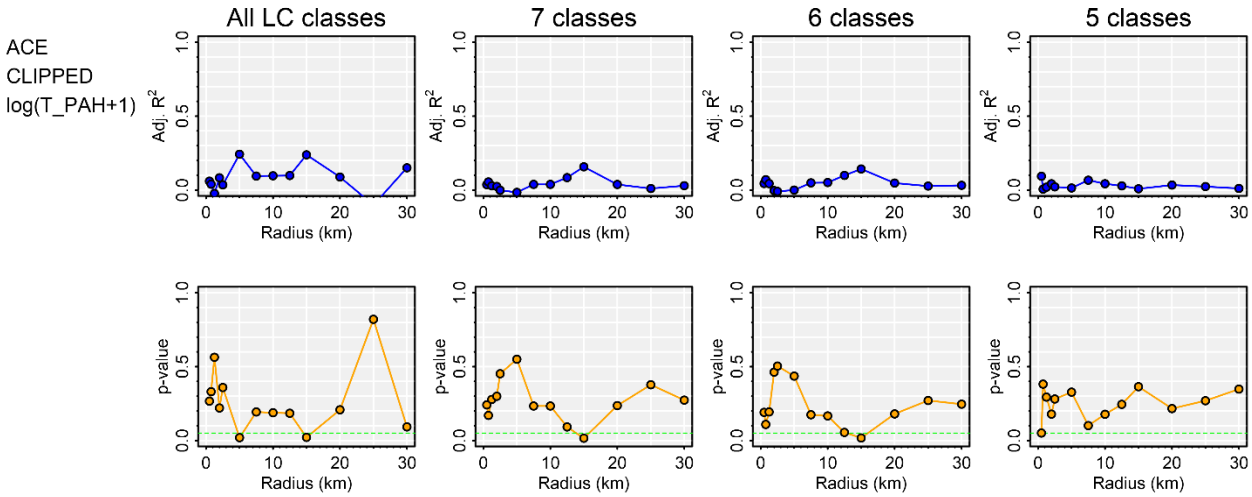

Figure 19. Model R^2 and p values from multiple regressions of total PAH versus land cover counts at varying buffer radii for ACE Basin NERR.

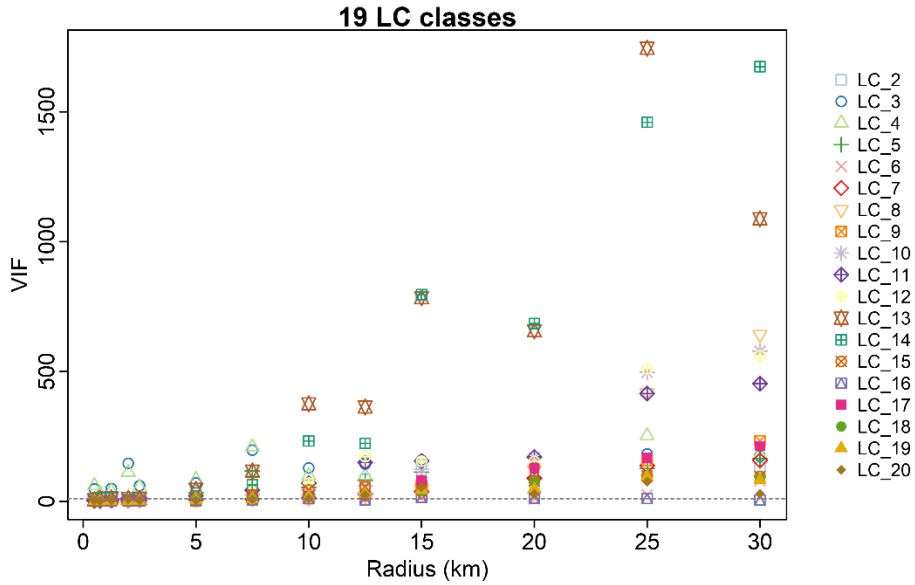

Figure 20. Variance inflation factors (VIFs) associated with multiple linear regressions of total PAH versus ungrouped land cover types at varying buffer radii for ACE Basin NERR. VIFs > 10 are suggestive of linear dependencies (multicollinearity) among predictors.

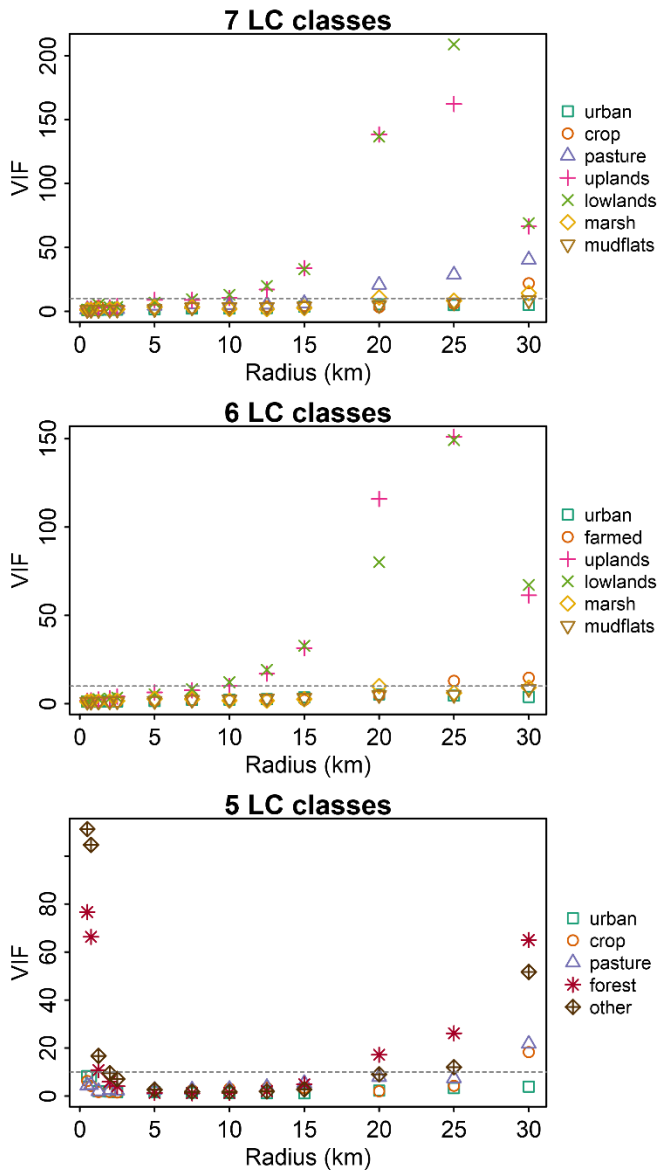

Figure 21. Variance inflation factors (VIFs) associated with multiple linear regressions of total PAH versus land cover aggregated to seven, six, and five categories for ACE Basin NERR.

In some ACE Basin regressions where model p-values attained significance (buffer distances of 5 and/or 15 km), the urban and/or upland land cover types were significant, although as noted above R^2 values typically were low (< 0.30). Furthermore, when using ungrouped land cover types VIFs were unacceptably high at nearly all buffer radii for some land cover classes (e.g., some 'developed' and/or 'palustrine wetland' types), as noted above. Aggregation of land cover classes improved VIFs, but overall model R^2 values were very low (< 0.20), suggesting that each of the respective models did a poor job of explaining the observed variability in sediment total PAH levels.

Regressions for Chesapeake Bay sites exhibited a high degree of variability across buffer radii (Figure 22). Regressions were significant for a limited number of buffer distances (e.g., 7.5 and 12.5 km), with maximum R^2 values approaching 0.50, but VIFs tended to be extremely high for most radii (Figure 23 and Figure 24; VIFs > 2000 not shown on plots).

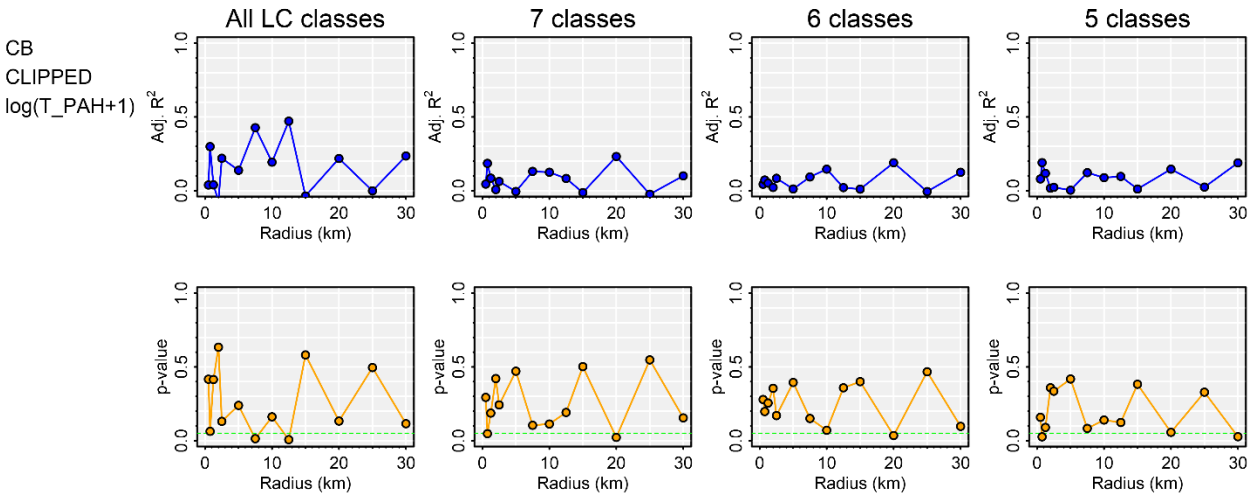

Figure 22. Model R^2 and p values from multiple regressions of total PAH versus land cover counts at varying buffer radii for Chesapeake Bay watersheds.

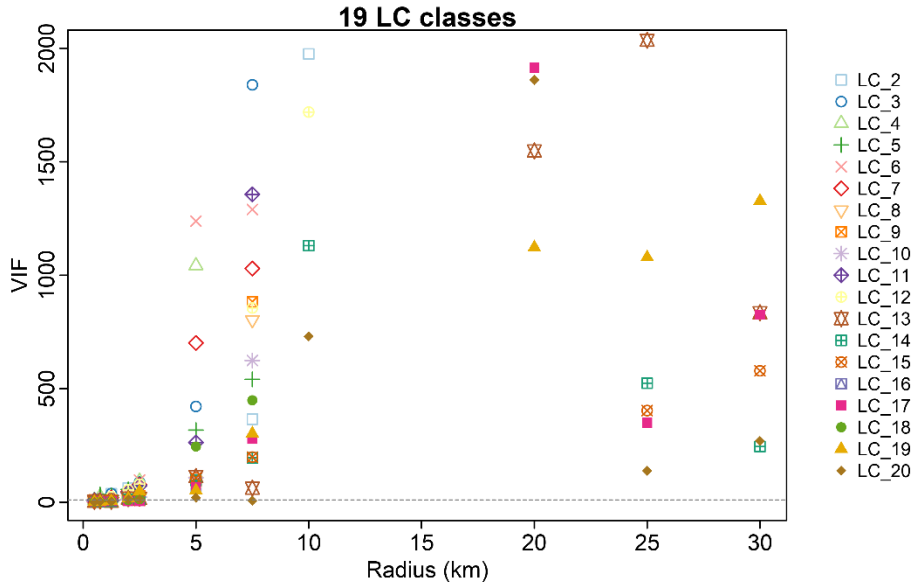

Figure 23. Variance inflation factors (VIFs) associated with multiple linear regressions of total PAH versus ungrouped land cover types at varying buffer radii for Chesapeake Bay watersheds. VIFs > 10 are suggestive of linear dependencies (multicollinearity) among predictors.

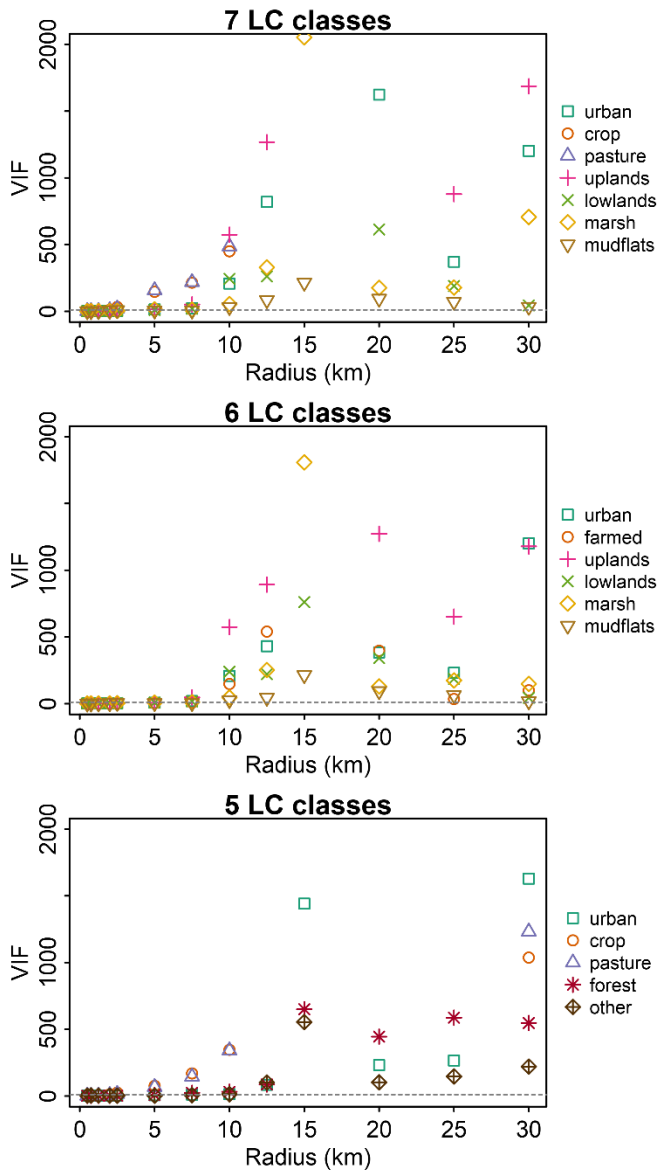

Figure 24. Variance inflation factors (VIFs) associated with multiple linear regressions of total PAH versus land cover aggregated to seven, six, and five categories for Chesapeake Bay watersheds.

4 Discussion

Our results show the challenges in trying to relate nearby land cover directly to sediment pollutant concentrations in the shallow coastal waters we tested. It appears from the proximity analysis that there is a relationship to nearby land cover, especially when land cover classes are aggregated. In particular, there was a consistent relationship between the urban land cover classes and total PAH in the Charleston study area. Although this relationship appears in the proximity analysis, attempts to impose a flow-driven watershed model, such as OpenNSPECT, highlighted some of the difficulties in trying to

apply an accumulated flow model to predict sediment pollutant concentrations. In Figure 10, it is clear that burning streamlines into the watershed DEM, which is required to get actual flow through the broad Charleston estuaries, resulted in an unrealistic accumulation of high pollutant values in one part of the stream (the upper Cooper River, Figure 10), which then propagated downstream due to the highly constrained flow. This leads to the bimodal distribution of predicted total PAH values in Figure 9, where almost all of the over-predicted values are found downstream from one high value in the upper Cooper River. The artificial creation of the stream flow lines was required because of the way in which "watersheds" were defined in the simple downhill-flow runoff model. Without that constraint, there would be almost no accumulation at the actual field locations, which would then lead to erroneously high loading estimates for each land cover class.

Another level of inaccuracy that was introduced by the stream-flow constraint was due to the artificial relocation of sampling points onto the nearest stream flow line (a requirement imposed by the model to get flow and concentration estimates at the location of the sampling points). In Figure 11, it is clear that the highest field concentrations were located in the Ashley River, quite near the urbanized shore of downtown Charleston. This spatial arrangement is what drives the relationship found in the proximity analysis and is what confounds the more complex, but in this case, highly artificial, flow model. It is important to apply models in a manner consistent with the assumptions that are made within them. In this present study, the assumption of unidirectional downhill flow is a very poor assumption for the tidal areas in which we tested it. However, one of the initial goals of this effort was to get a better understanding of how well OpenNSPECT would work in such areas. We have a clear answer: poorly.

Some additional areas that might be considered in future work are the consequences of using an annual model to estimate accumulated sediment pollutants. One of our ideas in this work was that the sediments would act as an integrator of pollution events and might, therefore, be a better proxy for comparing with the time-insensitive OpenNSPECT model than individual water column estimates. However, using it to estimate accumulated benthic pollutants makes an implicit assumption that all of those pollutants are a result of one year's runoff. That appears to be an unreasonable assumption in this case, but it might be interesting to understand more about the distribution of the pollutants within the sediments and if there is a way to estimate the annual pollutant load.

5 References

Balthis, L., J. Hyland, C. Cooksey, E. Wirth, M. Fulton, J. Moore, D. Hurley. 2012. *Support for Integrated Ecosystem Assessments of NOAA's National Estuarine Research Reserve System (NERRS): Assessment of Ecological Condition and Stressor Impacts in Subtidal Waters of the Sapelo Island National Estuarine Research Reserve.* NOAA Technical Memorandum NOS NCCOS 150. NOAA Center for Coastal Environmental Health and Biomolecular Research, Charleston, SC. 94 p. [Internet]. [Cited 2018 Mar 05]. Available from: NOAA Institutional Repository - 2614

[EPA] U.S. Environmental Protection Agency. 2015. *National Coastal Condition Assessment 2010 (EPA 841-R-15-006).* Washington, DC. Office of Water and Office of Research and Development. [Internet]. [Cited 2016 Sep 28]. Available from: National Coastal Condition Assessment

Eslinger, D.L., H.J. Carter, M. Pendleton, S. Burkhalter, M. Allen. 2012. *OpenNSPECT: The Open-source Nonpoint Source Pollution and Erosion Comparison Tool.* NOAA Office for Coastal Management, Charleston, South Carolina. [Internet] [Cited 2018 Jan 01]. Available from: OpenNSPECT.

Homer, C.G., J.A. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N.D. Herold, J.D. Wickham, and K. Megown. 2015. Completion of the 2011 National Land Cover Database for the conterminous United States-Representing a decade of land cover change information. *Photogrammetric Engineering and Remote Sensing* 81(5):345-354. [Internet]. [Cited 2018 Mar 05]. Available from: Multi-Resolution Land Characteristics Consortium - National Land Cover Database

Leight, A., J. Jacobs, L. Gonsalves, G. Messick, S. McLaughlin, J. Lewis, J. Brush, E. Daniels, M. Rhodes, L. Collier, B. Wood. 2014. *Coastal Ecosystem Assessment of Chesapeake Bay Watersheds: A Story of Three Rivers - the Corsica, Magothy, and Rhode.* NOAA Technical Memorandum NOS NCCOS 189. 93 p. [Internet]. [Cited 2018 Mar 05]. Available from: NOAA Institutional Repository - 2722

[NOAA] National Oceanic and Atmospheric Administration. 2013. *National Coastal Population Report: Population Trends from 1970 to 2020*. [Internet]. [Cited 2016 Sep 28]. Available from: National Coastal Population Report

Schenk, M., N. Garfield, D. Eslinger, J. Carter. 2008. *Development of Regional Pollution Export Coefficients for the N-SPECT Model: Use of USGS National Water Quality Assessment (NAWQA) Summary Data to Prototype a Total Nitrogen Regional Coefficient.* Poster presentation. [Internet]. [Cited 2018 Mar 05]. Available from: OpenNSPECT Pollution Export Coefficients Development

Van Dolah, R.F., D.M. Sanger, G.H.M. Riekerk, S.E. Crowe, M.V. Levisen, D.C. Bergquist, D.E. Chestnut, W. McDermott, M.H. Fulton, E. Wirth. 2013. *The Condition of South Carolina's Estuarine and Coastal Habitats During 2009-2010: Technical Report.* Charleston, SC: South Carolina Marine Resources Division. Technical Report No. 107. 54 p. [Internet]. [Cited 2018 Mar 05]. Available from: SCECAP Report 2013