PREDICTING CONDITION OF SMALL ESTUARINE SYSTEMS ALONG THE UNITED STATES' ATLANTIC COAST

BY

JEFFREY W. HOLLISTER

A DISSERTATION SUBMITTED IN PARTIAL FULLFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN

ENVIRONMENTAL SCIENCES

UNIVERSITY OF RHODE ISLAND

2004

DOCTOR OF PHILOSOPHY DISSERTATION

OF

JEFFREY WILLIAM HOLLISTER

APPROVED:

Dissertation Committee:

Major Professor__________________________________

____

____

____

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2004

Abstract

Anthropogenic activities are known to degrade the ecological condition of estuaries. Also, numerous studies find links between landscape structure and estuarine condition. From these relationships, it is possible to build reliable, predictive models of estuarine condition. In this study, I examined issues of data quality, scale, and statistical approach to developing predictive models of estuarine sediment metal concentrations in the Mid-Atlantic region of the United States. I assessed the accuracy of the National Land Cover Dataset (NLCD) at multiple scales by comparing it with reference data from Rhode Island and Massachusetts. Area of developed lands, agriculture, forest and water had acceptable accuracy at all but the finest scales (i.e., accurate for sampling units greater than $10km^2$). Rangeland, wetland, and barren were consistently, poorly classified. Appropriate scales for modeling landscape structure and estuarine sediment metal concentrations are often determined on a per-study basis. I varied the spatial extent used to calculate landscape structure and assessed linear relationships between estuarine sediment metal concentrations and the total area of developed and agricultural lands at each scale. Area of developed lands was consistently related to sediment metals while total agricultural land was not. Developed land had strongest associations with lead and copper; weakest with arsenic and chromium; and moderate with cadmium, mercury, and zinc. All metals showed negative scaling trends, but only lead and zinc were significant. Landscape structure within 15-20 km from a sampling station may be most important for describing the relationship between developed land and estuarine sediment metals. Lastly, with information-theoretic approaches I assessed multiple models containing pollution sources and estuarine characteristics and made predictions with model-averaging techniques. Total developed land and percent silt/clay of estuarine sediment were most important for all metals. Estuary area, hydrology, and total agricultural land varied in their importance. The model-averaged predictions explained 78.4%, 70.5%, 56.4%, and 50.3% of

the variation for copper, lead, mercury, and cadmium respectively. Overall prediction accuracies of the Effect's Range values were 83.9%, 84.8%, 78.6%, 92.0% for copper, lead, mercury, and cadmium respectively. I conclude that combining broad-scale data with information-theoretic approaches is an effective technique for predicting condition of small estuarine systems.

Acknowledgements

I am always struck by the inadequacy of acknowledgments. Trying to thank everyone that deserves it and doing so in a meaningful way seems impossible within the confines of just a few pages. Be that as it may, I will attempt to adequately acknowledge the many people and organizations that have helped with this study, helped me grow as a scientist, and helped me grow as a person. First, I must thank the various organizations that have funded my research and education. Funds from the U. S. Environmental Protection Agency, the Department of Natural Resources Science at the University of Rhode Island, the RI Chapter of the Surfrider Foundation, and the Rhode Island Natural History Survey made it possible for me to conduct this research. Second, I would like to thank all the members of my dissertation committee -- Y. Q. Wang, Howie Ginsberg, Art Gold, Liliana Gonzales, and John Paul -- for their insight and guidance. Third, Jane Copeland, Matt Nicholson, John Paul (again!) and all the people at the EPA's Atlantic Ecology Lab who have contributed in many ways to this research. Fourth, I would like to acknowledge everyone in the Department of Natural Resources Science, my academic home for the last 4 1/2 years and all the members of the Environmental Data Center. Through my experiences here, I now know the true meaning of collaboration and camaraderie and thanks to Arty and Mike I will always have a special appreciation for the Mt. Washington Hotel.

My friends and family back in Kansas also deserve mention. Thank you for letting me escape the safe confines of Kansas to explore my personal and educational goals and thanks for always giving me reasons to look forward to coming home. You all are partially responsible (for better or worse) for who I am today. Now, to all the friends who I have met over the last 5 years in Rhode Island (you know who you are): thank you! Without you, life in Rhode Island surely would have been dull. Thanks to you I now I have a bevy of stories to tell about flying rolls, drunken handstands, intoxicated bowling, and chocolate-eating dogs. Most importantly I now have a new group of supportive friends who have had more to do with me finishing my degree than they might appreciate.

Two last people need to be singled out. First, I must say thank you to my advisor, mentor, and friend, Peter August. Pete, your contributions to this research and to my growth as a scientist could never be summed up in a few sentences. Working with you over the last several years I have learned a considerable amount not just about GIS and Landscape Ecology, but also how to lead by example. If I end up being just half of who you are as a scientist, professional and person, I will have far exceeded even the grandest of my ambitions. Thank you.

And the last thank you goes to my fiancé, friend, confidant, and source of support, Lisa. Without you all this would be for naught. When I first came to Rhode Island, I expected that the Ph. D. would be the best thing I would leave Rhode Island with, now that is no longer the case. Your love and your support far outshine any degree. I certainly cannot fully express my feelings for you in one paragraph nor am I fully capable of letting you know how lucky I feel that I am to have you in my life.

Table of Contents

List of Tables

List of Figures

Chapter 1. Assessing the Accuracy of National Land Cover Dataset Area Estimates at Multiple Spatial Extents.

(This Chapter is written in the format required for publication in the journal *Photogrammetric Engineering and Remote Sensing* and was accepted for publication in April 2004.)

Introduction

Measurements of landscape pattern are linked with many key ecological processes and are often used to model a variety of parameters such as species diversity or water quality (Flather et al. 1992, Levine et al. 1993, Hunsaker and Levine 1995). These measurements are typically made from digital land use/land cover (LULC) data and are either calculated across an entire landscape or within sampling units of varying size such as a full watershed or a buffer around a sampling station (Comeleo et al. 1996). The accuracy of the estimates of landscape composition and pattern are dependent upon the accuracy of the source data. The standard method of assessing accuracy involves comparing the land cover class derived from the LULC data at various locations with the land cover class occurring in a reference dataset, then generating error matrices that show the degree to which land cover classes are correctly identified in the LULC dataset (Stehman and Czaplewski 1998, Congalton and Green 1999). These robust methods are effective at estimating LULC accuracy for an entire image. However, many ecological applications utilize measurements of the total area or percent of the total area of land cover classes for a range of spatial extents such as watersheds, bio-reserves, or political divisions like states or countries (McGarigal and McComb 1995, Comeleo et al. 1996, Jones et al. 1997, Paul et al. 2002). The accuracy of area measurements across a range of spatial extents cannot be fully inferred from accuracy assessments created using standard point-to-point methods, especially when site-specific accuracy is found to be poor.

The National Land Cover Dataset (Vogelmann and Wickham 2000) is a LULC data product for the contiguous United States developed by the Multi-Resolution Landscape Characteristics Consortium, a consortium of federal agencies, from Landsat Thematic Mapper Imagery acquired from 1990 to 1993 (http://landcover.usgs.gov/natllandcover.html). Using an unsupervised classification algorithm, along with ancillary data, a total of 21 thematic classes were derived (Lunetta et al. 1998, Vogelmann et al. 1998a, Vogelmann et al. 1998b, Vogelmann and Wickham 2000). The National Land Cover Dataset (NLCD) promises to serve as a rich source of information on landscape composition and pattern for ecological assessments. The accuracy of the NLCD within the New England region (Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, and Maine) has been assessed using point-to-point comparisons with reference data and the results of these studies are currently available (Yang et al. 2001, United States Geological Survey 2003). However, our use of this accuracy information was limited because the accuracy of the NLCD within subsets of the New England region and the accuracy of LULC area measurements within sampling units of varying extents has not been measured.

I addressed these limitations with three objectives. The first objective was to examine the site-specific accuracy of the NLCD within Rhode Island and Massachusetts and identify which classes exhibited poor site-specific accuracy. Second, I evaluated area estimates of NLCD classes using an accuracy assessment at multiple spatial extents and a variety of analysis techniques. Third, I determined if the accuracies changed as a function of spatial extent.

Study Area and Data

The study area consists of the states of Massachusetts and Rhode Island in southern New England. This region has four large urban centers (Boston, MA; Springfield, MA; Worcester, MA; and Providence, RI), and is dominated by forest, developed land, agriculture, wetlands, and water. Barren land, orchards/vineyards, and rangeland encompass a small portion of the overall landscape (Table 1-1). Our choice of study area was driven by availability of large scale, photo-interpreted reference data to compare to the NLCD.

I acquired the NLCD data for the states of Rhode Island and Massachusetts. The NLCD had not been filtered and exhibited a "salt and pepper" effect which is often seen in satellite derived LULC data (Loveland et al. 1999, United States Geological Survey 2003). It is common practice to use a small focal majority window to filter isolated pixels of land cover and merge them into the surrounding matrix cover (Lillesand and Kiefer 1995, DeMers 1997, Burrough and McDonnell 1998). Additionally, isolated pixels are smaller than the effective minimum mapping unit of satellite derived LULC data (Loveland et al. 1999). Following the recommendations of the United States Geological Survey, I filtered the NLCD with a 3 pixel by 3 pixel majority window with ties left unchanged, producing a LULC product with a 0.8 ha (2 acre) minimum mapping unit (MMU). All of our analyses were conducted on the resultant filtered data (United States Geological Survey 2003).

Reference data were acquired from the Massachusetts GIS (MassGIS, http://www.state.ma.us/mgis) and Rhode Island GIS (RIGIS, http://www.edc.uri.edu/rigis) Programs. The MassGIS LULC data were photo-interpreted from 1:25,000 scale aerial photography into a total of 37 classes, with acquisition dates ranging from 1985 to 1997 with a minimum polygon size of 0.4 ha (1 acre). The RIGIS LULC data consisted of 40 land use/ land cover classes that were photo-interpreted from 1:12,000 USGS digital orthophoto quadrangles (DOQ) obtained in 1995. The RIGIS data had a minimum polygon size of 0.2 ha (0.5 acre). Each dataset met National Map Accuracy Standards and underwent extensive quality assurance/quality control procedures (http://www.state.ma.us/mgis, http://www.edc.uri.edu/rigis, David W. Goodwin, personal communication). The RIGIS and MassGIS data were converted to the same format as the NLCD (raster format with 30 m pixels), recoded to Anderson Level I classification and appended together to create a single, seamless reference LULC data set (Anderson et al. 1976). The NLCD was recoded to Anderson Level I to be consistent with the Massachusetts and Rhode Island reference data (Table 1-2).

Methods

Site-Specific Accuracy Assessment

I assessed the NLCD within Rhode Island and Massachusetts to identify poorly classified classes, and to assess overall accuracy. This site-specific accuracy assessment provided a basis for comparing the multiple-extents accuracy assessment. The reference data represented complete coverage of the study area; therefore, it was unnecessary to sample and the assessment represents a complete census of the NLCD. All NLCD pixels within the study area were converted to points (points represent the center location of the pixel). For each point, the Anderson Level I LULC class was recorded for both the reference data and the NLCD. Agreement and disagreement between the two datasets were recorded in an error matrix, then user's accuracy, producer's accuracy, and overall accuracy were calculated (Stehman and Czaplewski 1998, Congalton and Green 1999).

Multiple-Extents Accuracy Assessment Sampling Scheme

To test the accuracy of NLCD-derived area measurements at multiple extents, I created 35 randomly chosen, non-overlapping circular sampling units with radii ranging from 0.178 to 7.98 kilometers (area of 0.1 km^2 to 200 km^2). I accomplished this by placing a random point within the study area and delineating a circle with the specified radius. If the resulting circle overlapped another, it was discarded. This process was repeated until 35 non-overlapping circles were selected for each of 40 radii for a total of 1,400 circles (40 radii x 35 circles). Each group of 35 circles are collectively referred to as "spatial extents." Each of the 35 circles of the 40 spatial extents was overlaid on the majority filtered NLCD and reference LULC data. Total area of the LULC classes within each circle was calculated, resulting in a total of 245 area measurements within each spatial extent (35 circles and 7 LULC classes).

It has been suggested that the NLCD not be used to characterize areas smaller than approximately 25 km^2 (United States Geological Survey 2003). Our range of spatial extents was intentionally chosen to span this suggested minimum functional area. I felt that the accuracy of area estimates from the NLCD may be acceptable within areas smaller than 25 km^2 and wanted to test this. Additionally, prior analysis indicated that accuracy stabilizes at larger extents. Consequently, I chose to evaluate more of the smaller spatial extents, examining 25 spatial extents smaller than 25 km^2. All 40 (25 less than 25 km^2 and 15 greater than 25 km^2) spatial extents were included in the final analyses.

Multiple-Extents Analysis 1: Testing for Differences in Mean Area

Because of the nature of the data (map and reference land cover measurements on the same sampling units) it is natural to consider using the well-known paired t test to test for a difference in the mean of the map and reference areas for each LULC class within each spatial extent. This test quantifies the magnitude and direction (i.e., positive or negative) of the bias of the area estimates derived from the NLCD. But, the paired t test cannot be used in isolation to evaluate accuracy. While equality in means implies no bias, it does not assure good accuracy. Large disagreements between the NLCD area and reference area may still exist because positive and negative differences may counteract and produce a mean difference near zero (i.e., similar mean but different variance). I address this limitation with the final analysis.

Multiple-Extents Analysis 2: Linear Regressions

I assessed accuracy of the NLCD at multiple spatial extents using linear regression. I fit a simple linear regression equation for each of the seven classes in each of forty spatial extents. I used the NLCD area as the dependent variable and the reference data area as the independent (NLCD = $ß_0$ + $ß_1$REFERENCE). An NLCD class can be considered accurate relative to the reference data when the regression equation displays a one-to-one relationship between the NLCD and the reference data (i.e., $ß_0 = 0$, $ß_1 = 1$, and high r^2) and indicates that an increase in area of the reference data results in an equivalent increase in area for the NLCD. I used these three values to determine the accuracy of the classes at all spatial extents and, in addition to the standard test for a regression equation (H_0: $ß_1 = 0$), I explicitly tested $ß_0$ and $ß_1$ separately to determine if the intercept was equal to zero (H_0: $ß_0 = 0$) and if the slope was equal to one (H_0: $ß_1 = 1$). Additionally, it is possible to have a curvilinear relationship that explains much of the variation in the NLCD area measurements; however, this situation would indicate that agreement was poor over portions of the range of values of the reference data. To ensure that this was not impacting our results, I constructed quadratic regressions for each class at each spatial extent, and tested for the significance of the quadratic terms.

Results

Site-specific Accuracy Assessment

Our complete census for Anderson Level I classes resulted in an overall accuracy of the NLCD within Rhode Island and Massachusetts was 71.4% (18,821,600 correct pixels / 26,367,557 total pixels). Only the producer's accuracy for the forest and water classes had an accuracy nearing 85% (Table 1-3). The poor accuracy of the rangeland class is likely a function of its rarity.

Multiple-Extents Accuracy Assessment: Testing for Differences in Mean Area

Forest and wetland classes showed significant, positive differences for mean area in more than half of the spatial extents, indicating higher mean values for the NLCD than for the reference data and the percent mean difference ranged from near 0 to 5% (Figures 1-1 and 1-2). Mean area of developed and barren showed a significant, negative mean differences for most (approximately 93%) of the spatial extents and the percent mean difference ranged from near 0 to -1%. These were the only LULC classes to consistently show lower mean values for the NLCD (Figures 1-1 and 1-2). Water and agriculture classes did not consistently display a significant difference between the two datasets. The magnitudes of the percent mean difference for all classes never exceeded 10% and rarely exceed 5% (Figure 1-2). Statistical tests on rangeland may provide confusing and unreliable results because no rangeland was found in many of the samples. Thus, paired *t* test and regression results for the rangeland class will not be presented.

Multiple-Extents Accuracy Assessment: Linear Regression Results

Significant linear relationships were found in 100% of the spatial extents for developed, agriculture, forest, and water and the linear component for these classes accounted for a significant portion of the variation (Figure 1-3). Wetlands had significant linear relationships for only 45% of the spatial extents, accounting for, on average, 29.0% of the variation (Figure 1-3). Barren land had significant linear relationships for 72.5% of the spatial extents and accounted for an average of only 27.0% of the variation (Figure 1-3).

For all classes except wetland and barren, significant curvilinear relationships existed for only a few spatial extents. For those that were significant, the inclusion of a quadratic term only accounted for a small amount of additional variation. Many spatial extents for the barren and wetland classes had significant curvilinear relationships, which accounted for nearly 20% of the variation in some cases. The low number of significant linear regressions and higher amount of variation accounted for by the quadratic term implies that a strong linear relationship does not exist for the wetland and barren classes. These classes appear to be non-linear (i.e., have poor accuracy) and assessment of the slope and intercept would not yield

additional information about accuracy; therefore, slope and intercept results for barren and wetland will not be discussed.

Intercepts ($ß_0$) for the developed class were less than 0 for 60.0% of the spatial extents. Agriculture and forest intercepts were greater than 0 in more than half of the spatial extents and water rarely had intercepts significantly different from 0. Having an intercept different than zero implies a bias in the estimate of the NLCD indicating that agriculture and forest were often overestimated while developed area was underestimated. Although testing slightly different parameters, these results and the results of the paired t-tests corroborate one another and show similar patterns of over- and underestimation. Therefore, it is not necessary for us to include a graph of the intercepts, as it illustrates the same pattern as seen in Figure 1-1.

The final parameter to be tested was slope of the regression line for each of the spatial extents. Slopes (β_1) were significantly less than one in 52.5% of the spatial extents for developed, in 70% of the spatial extents for agriculture, and in 42.5% for forest (Figure 1-4). The slope for area of water was greater than 1 in 12.5% of the spatial extents and less than 1 for 7.5% of the spatial extents (Figure 1-4).

Discussion and Conclusions

NLCD Area Accuracy

Two classes -- barren and wetland -- were shown to have the low r^2 values, significant differences in the mean, and non-linear relationships; therefore, the area measurements for these classes could be considered to have poor accuracy. The water class has high r^2 values, a slight difference in mean and intercept, and consistently exhibits a one-to-one relationship with the reference data. Thus, water appears to be well classified for area across most of the spatial extents. The remaining three classes, developed, agriculture, and forest had significant linear relationships, suggesting acceptable levels of accuracy, but also showed differences in the mean, intercepts not equal to 0, and a slope often not equal to one. Specifically, both the mean and intercept for developed area, as shown by the t-test and test of the intercept, are consistently underestimated in the NLCD, whereas agriculture and forest classes tend to be overestimated. Making a determination of poor accuracy for these classes is not entirely straightforward. Several of the tests do imply poor accuracy, but inherent differences in the two datasets (see *Sources of Error*) make it difficult to determine if the lack of a one-to-one relationship is due to misclassification or other errors.

In general, area measurements of the developed, agriculture, forest and water classes are fairly accurate. Area of rangeland, barren and wetland, are poorly classified and users of these data should either be cautious when using area estimates for these classes or, if necessary, explore ancillary sources of data. As with all accuracy assessments, making a final determination of the accuracy of area measurements must be done with caution. The accuracy and relevance of the reference data can have a profound impact on the determination of accuracy and it is always important to consider the use of the data and the level of detail in the classification scheme when making decisions regarding acceptable levels of misclassification (Loveland et al. 1999, Congalton and Green 1999).

Relationship Between Accuracy and Spatial Extent

One of the objectives of this study was to determine if accuracy changes as a function of spatial extent. Such a relationship would suggest that a minimum size for sampling units (i.e. a minimum functional extent) may exist, below which the accuracy of the NLCD would reach unacceptable levels and LULC area estimates based on those sampling units should be used cautiously. My analyses proved effective at exploring such a relationship. The r^2 values suggest that accuracy increases as spatial extent increases, with the r^2 leveling off at a radius between about 1.5 - 2.5 km (approx. 7 - 20 km^2 spatial extents) for developed, agriculture, forest, and water (Figure 1-3). However, r^2 is not ideal for assessing this relationship because of the difficulties in determining an "acceptable" accuracy value for r^2 as well as the considerable variation in r^2 from similar spatial extents due to sampling and other errors (see **Sources of Error**). If slope also displays a relationship with spatial extent and levels off around a slope of 1, then a minimum functional extent can be said to exist at that point. For the four classes with the highest accuracy, there is little variation in the slope and finding a point below which the slopes tended to be significantly different from 1 was not possible (Figure 1-4). Percent mean difference provides perhaps the best indication of a relationship between accuracy and spatial extent. Percent mean difference showed greater variation at smaller spatial extents and leveled off at a radius around 1 - 2 km (approx. 3 - 12 km^2 spatial extents) for all classes except water, which remained constant across most spatial extents (Figure 1-2). These results corroborate the suggestion by the USGS to exercise caution when using the NLCD within small, local areas; however, they also imply that the accuracy of the NLCD may be acceptable at extents smaller than the 25 km^2 originally suggested by the USGS (United States Geological Survey 2003).

Sources of Error

There are four principal reasons, besides misclassification, for the poor correspondence (i.e., low r^2, biases in the mean, $ß_0 \neq 0$, $ß_1 \neq 1$) between the NLCD and reference data and the large variation in accuracy sometimes shown between spatial extents of similar sizes. First, temporal discrepancies among the acquisition dates of the NLCD and reference data may add error (Congalton and Green 1999). NLCD data were obtained between 1990 and 1993. MassGIS data were created from imagery obtained between 1985 and 1997 and RIGIS data were obtained in 1995. Assessing the accuracy of a dataset using reference data from slightly different time periods may result in lowered accuracies due to landscape change, not misclassification (Congalton and Green 1999). It is also common, as is the case with the

NLCD, to have internal temporal inconsistencies in source data of regional scale LULC datasets (Vogelmann et al. 1998b, Loveland et al. 1999). The rates of landscape change are variable across classes, across period in history, and across regions. Thus, it is difficult to predict the effect that temporal discrepancies will have on estimation of classification accuracy (Dunn et al. 1990). Discrepancies such as these are, as Loveland *et al.* (1999) have suggested, an unavoidable part of characterizing landscapes over broad regions.

Second, comparing metrics derived from analog sources (i.e., aerial photos) to metrics with digital sources (i.e., satellite imagery) can be problematic (Loveland et al. 1999). The resultant land cover datasets derived from these different sources represent the landscape in two distinct ways, vector data from photo-interpretation of aerial photos and raster from the classification of satellite imagery (DeMers 1997, Burrough and McDonnell 1998) . Raster data suffer from a "stair-stepped" effect, noticeable along polygon perimeters, whereas vector data represents perimeters with clean, smooth lines (Congalton 1997). This will produce variation in the estimation of area totals based on data format and not necessarily misclassification.

Third, variation in the MMU of the three datasets may contribute to error in area estimation. The MMU for the three datasets ranged from 0.2 to 0.8 ha. The result of this is variation in detail, creating a different number of features for each reference dataset. Similarly, the minimum width of included features also differs. Although no specific mention of a consistent minimum width is included with any of the three datasets, the inclusion of roads (greater than 61 m (200 ft) for the RIGIS data) and power lines (greater than 30.5 m (100 ft) for the RIGIS data) in the reference datasets suggests that linear features smaller than the MMU of the NLCD were included. As there are currently no raster sources of reference data that are larger scale than NLCD, encompass a large area, and have relevant LULC classes, it is necessary that I use data of differing initial format and varying MMU.

Fourth, Congalton and Green (1999) have suggested that quantitative accuracy assessments not be conducted with existing sources of reference data because of additional error introduced from differing classification schemes. Typically, reference data are collected through field visits or with aerial photograph interpretation using the same classification protocols used to develop the LULC data being assessed and are assumed to represent "true" values (Congalton 1991, Congalton and Green 1999). Since our study was initiated several years after the completion of the NLCD, field data collection was not possible. Additionally, our method requires complete photo interpretation of large areas. Aerial photograph acquisition and interpretation for this research was not feasible. I followed Congalton and Green's (1999) recommendation for our site-specific accuracy assessment and only looked at producer’s, user’s, and overall accuracy. For the multiple-extent accuracy assessment, our need to have an estimate of the accuracy of area measurements, and the value added to our understanding of NLCD accuracy, overrides the potential problems associated with using existing reference data sources.

Poorly Classified Classes

The low accuracy of areas measured across multiple spatial extents for wetlands, barren, and rangeland is probably due to their low classification accuracy in the NLCD. These classes tend to have low overall classification accuracies in quality assurance testing conducted by the USGS (2003) and relative to our reference data (Table 1-3). The spectral signatures for wetland vegetation may be similar to forest (for forested wetlands) and rangeland (for shrub/scrub wetlands). This contributes to a high rate of misclassification for the wetland classes derived from the digitally processed NLCD. Our results and prior accuracy assessments suggest that it may not be advisable to use the NLCD to estimate wetland area. Although used as an ancillary source for labeling of clusters in the development of the NLCD, the National Wetlands Inventory (NWI, http://www.nwi.fws.gov) may provide better information as a sole source of data on wetlands because it is less likely to reflect the spectral confusion of the NLCD. The remaining poorly depicted classes, barren and rangeland, make up only 1.1% of the total study area. The rarity of these two classes likely contributes to the poor classification results.

The confusion identified by the site-specific accuracy assessment between the developed, forest, and agriculture classes is likely due to differences in classification methodologies of the reference and NLCD data and not necessarily due to misclassification. For instance, LULC data derived from satellite imagery often identifies individual trees within an urban or agricultural setting as forest. However, photo-interpretation of those same sites will include individual trees within the overall developed or agricultural land use class and thus, add to misclassification within the error matrix as well as the overestimation indicated in Figure 1-2 (Y.Q. Wang, personal communication, 2002)

Finally, my study only considers NLCD area accuracy within Massachusetts and Rhode Island. To fully understand the accuracy of NLCD-derived metrics within other geographic settings, similar multiple-extent assessment methods should be conducted in other regions. Published site-specific accuracy assessments show similar overall and individual LULC accuracies across a range of geographic locations; thus, I may expect a multiple-extent accuracy assessment to find similar results in these other regions as well (Yang et al. 2001, United States Geological Survey 2003). It is also reasonable to expect that the suite of poorly classified and correctly classified classes could change. For example, the rangeland class of the NLCD is exceedingly rare in Massachusetts and Rhode Island; however, in many western states rangeland would be a much more common class and it is possible that the NLCD would accurately depict it. The fact that rangeland was poorly depicted in this study probably reflects its rate of occurrence in the New England landscape.

Multiple-Extents Accuracy Assessment Methodology

Since I use areal sampling units (i.e., spatial extents), our accuracy evaluation protocol is similar to the non-site specific accuracy assessments which preceded the quantitative, site-specific assessments used today (Stehman and Czaplewski 1998). An important difference between our methodology and non-site specific assessments is our use of linear regression. Regression analysis allows us to utilize well developed statistical approaches for testing hypotheses and determining how well one classification agrees with another (Zar 1999). For instance, it would be possible to perform our analysis, even at a single extent, and compare two datasets developed with different classification methodologies to a common reference source and determine the relative performance of those two methods by assessing linearity, intercept, and slope. These analyses typically have not accompanied previous non-site specific assessments.

Furthermore, a site-specific assessment might identify several classes as being poorly classified but those same classes might be variable in their ability to correctly estimate area. Several classes do exhibit poor classifications with the site-specific accuracy assessment (Table 1-3). However, the ability of these classes to accurately depict area as indicated by the multiple-extent accuracy assessment is quite different. For instance, agricultural and wetland both displayed poor user's and producer's accuracies (Table 1-3), but area of agricultural lands was accurately estimated across many extents, while wetland area was poorly depicted (Figure 1-1 - 1-4).

Summary

Our primary goal for this study was to address the use of the NLCD as a source of landscape composition metrics by assessing the accuracy of areal summaries of LULC classes derived from the NLCD. Our general conclusion is that the NLCD provides reliable area estimates for LULC classes that are dominant and display unique spectral signatures. Additionally, sampling units used to subset the NLCD, must be large enough (i.e., 10 km^2 or greater in area or radii greater than 1.5 to 2 km) to ensure accurate measurements of area. Although our intent was not to develop an areal-based accuracy assessment method, I were forced to do so. I discovered that in cases where unacceptable accuracies are shown at fine scales with site-specific accuracy assessments, explicitly examining the accuracy of area measurements with a multiple-extents method is a logical additional step. Both site-specific and multiple-extents methods provide distinct information, which is important to the understanding of a dataset's thematic accuracy. Site-specific accuracy assessments provide unique information on estimates of overall accuracy (i.e., kappa statistics), user's and producer's accuracy, and sources of confusion among LULC classes (Stehman and Czaplewski 1998, Congalton and Green 1999). An accuracy assessment at multiple extents adds additional information regarding the accuracy of area measurements and, indirectly, the landscape metrics derived from those measurements.

Table 1-1. Area (km^2) and percent of total area of Anderson Level 1 classes for Rhode Island GIS (RIGIS) and Massachusetts GIS (MassGIS) reference data and NLCD data

Anderson	**RIGIS and MassGIS**		**NLCD**	
Level 1	**Area (km^2)**	**% Total**	**Area (km^2)**	**% Total**
Developed	5612.3	23.6%	4116.1	17.3%
Agriculture	1861.7	7.8%	2053.4	8.7%
Rangeland	46.3	0.2%	16.3	0.1%
Forest	13635.5	57.4%	14698.7	61.9%
Water	808.5	3.4%	943.3	4.0%
Wetland	1004.3	4.2%	1670.3	7.0%
Barren	788	3.3%	232.8	1.0%
TOTAL	23756.6	100.0%	23730.9	100.0%

Table 1-2. Modified Anderson Level 1 classification scheme used to standardize the cover classes between the NLCD and the reference data.

Anderson Level I	RIGIS Classes	Mass-GIS Classes	NLCD Classes
100 - Developed	All Residential Classes, Commercial, Commercial/ Industrial Mixed, Institutional, Airports, Other Transportation, Power Lines, Railroads, Roads, Waste Disposal, Water and Sewage Treatment, Developed Recreation, Cemeteries, Vacant Land	All Residential Classes, Commercial, Industrial, Transportation, Waste Disposal, All Recreation Classes, Urban/Open	Low Intensity Residential, High Intensity Residential, Commercial/Industrial /Transportation, Urban/Recreational
200 - Agriculture	Pasture, Cropland, Confined Feeding Operation, Idle Agriculture, Orchards, Groves, Nurseries	Cropland, Pasture, Woody Perennial	Pasture/Hay, Row Crops, Fallow, Orchard/Vineyards, Other
300 - Rangeland	Brushland	-NA-	Shrubland
400 - Forest	Deciduous Forest, Evergreen Forest, Mixed Deciduous Forest, Mixed Evergreen Forest	Forest	Deciduous Forest, Evergreen Forest, Mixed Forest
500 - Water	Water, Salt Water	Water, New Ocean	Open Water
600 - Wetlands	Wetland	All Wetland Classes	Wetlands
700 - Barren	Beaches, Sandy Areas, Rock Outcrops, Mines, Quarries and Gravel Pits, Transitional Areas, Mixed Barren Areas	Mining, Open Land	Bare Rock/Sand/Clay, Quarries/Strip Mines/Gravel Pits, Transitional

Table 1-3. Site-specific (i.e., pixel-to-pixel) error matrix for NLCD in Rhode Island and Massachusetts. Bold entries indicate agreement between the two datasets and off-diagonal values indicate disagreement.

Classified	**Reference Data - MassGIS and RIGIS**							
Data - MRLC	Developed	Agriculture	Rangeland	Forest	Water	Wetland	Barren	Row Total
Developed	**3,387,120**	221,426	7,561	710,965	20,159	89,266	136,936	4,573,433
Agriculture	749,086	**1,038,488**	3,833	323,435	3,596	40,797	122,280	2,281,515
Rangeland	1,283	1,972	**101**	4,144	38	510	10,010	18,058
Forest	1,616,502	647,596	28,187	**13,128,860**	52,162	467,195	391,409	16,331,911
Water	64,651	11,014	772	95,509	**758,576**	96,976	20,557	1,048,055
Wetland	331,274	118,056	4,625	849,219	58,822	**405,289**	88,582	1,855,867
Barren	78,190	25,415	953	33,210	3,618	14,166	**103,166**	258,718
Column Total	6,228,106	2,063,967	46,032	15,145,342	896,971	1,114,199	872,940	26,367,557

Overall Accuracy = 71.4%, **Producer's Accuracy** Developed = 54.4% Agriculture = 50.3% Rangeland = 0.2% Forest = 86.7% Water = 84.6% Wetland = 36.4% Barren = 11.8%, **User's Accuracy** Developed = 74.1% Agriculture = 45.5% Rangeland = 0.6% Forest = 80.4% Water = 72.4% Wetland = 21.8% Barren = 39.9%.

Figure 1-1. Student's t-statistic for the paired t-test testing difference between the NLCD and the Reference data plotted as a function of radius of the spatial extent. Reference lines represent critical value for t (α= 0.05, n = 35). Values outside of reference lines represent extents with significant mean differences.

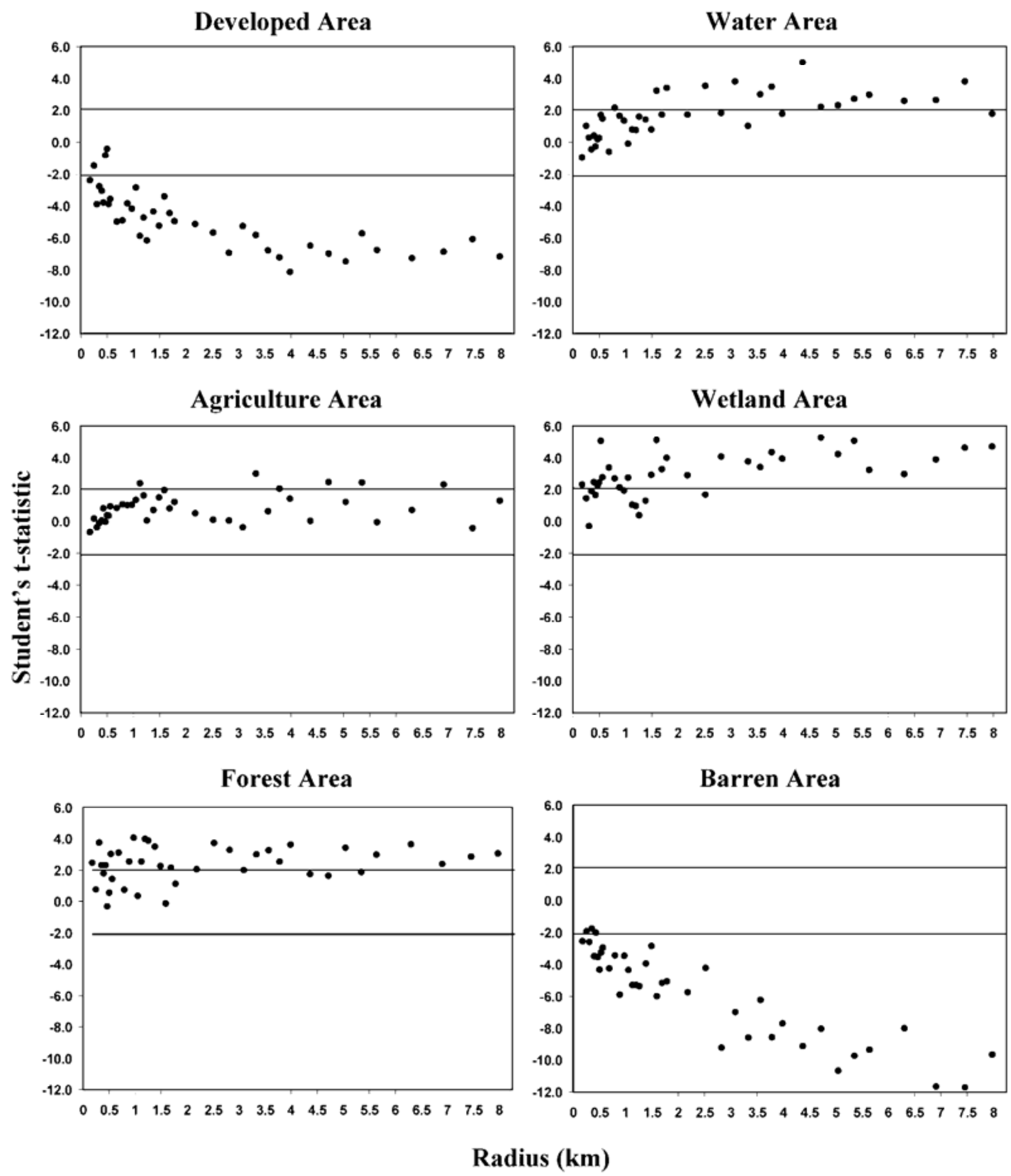


Figure 1-2. Percent mean difference ((NLCD - REFERENCE) / Area of Spatial Extent) between the NLCD and the Reference data plotted as a function of radius of the spatial extent.

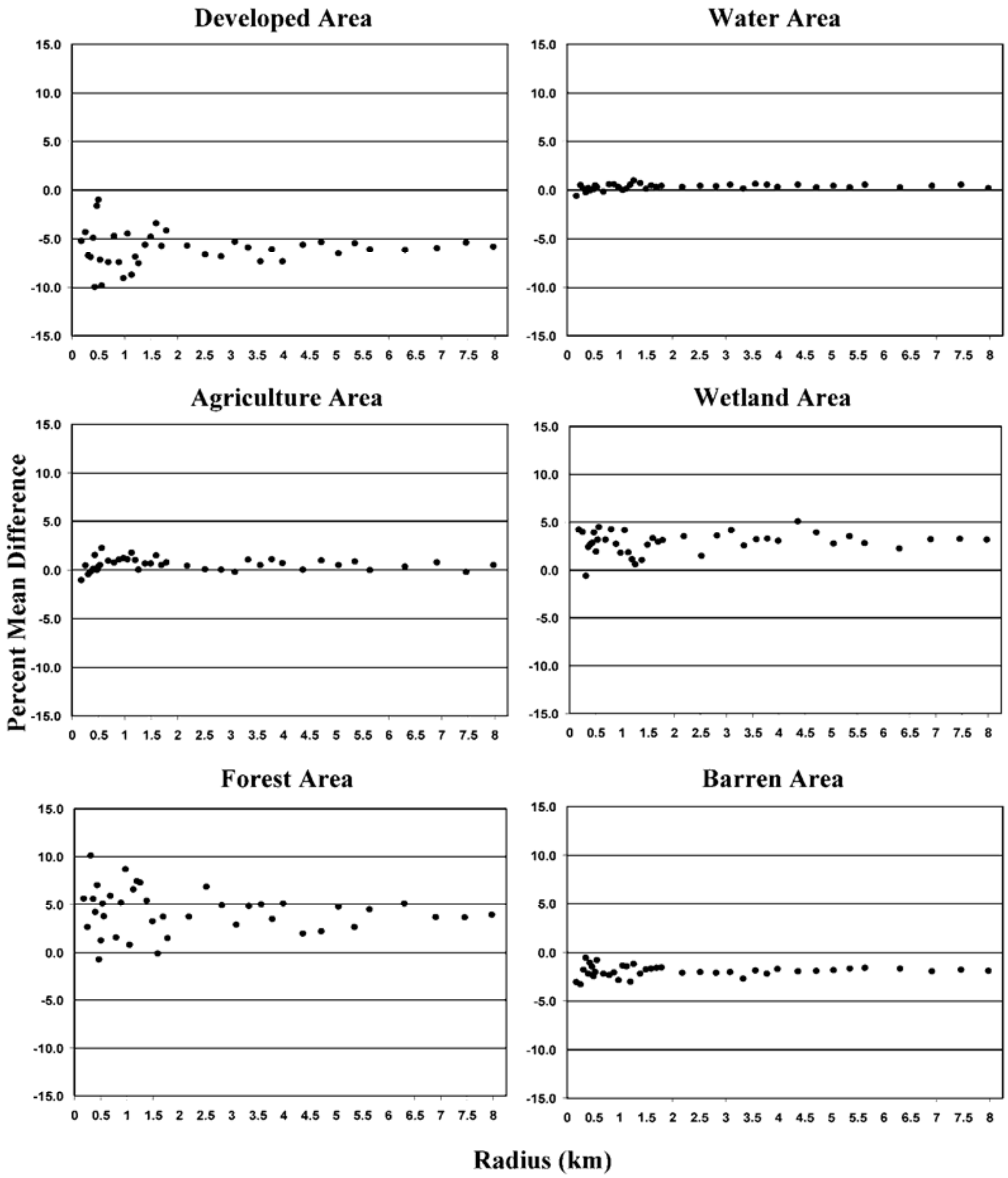


Figure 1-3. Coefficient of determination (r^2) of the regression for area (NLCD = $ß_0$ + $ß_1$REFERENCE) plotted as a function of radius of the spatial extent.

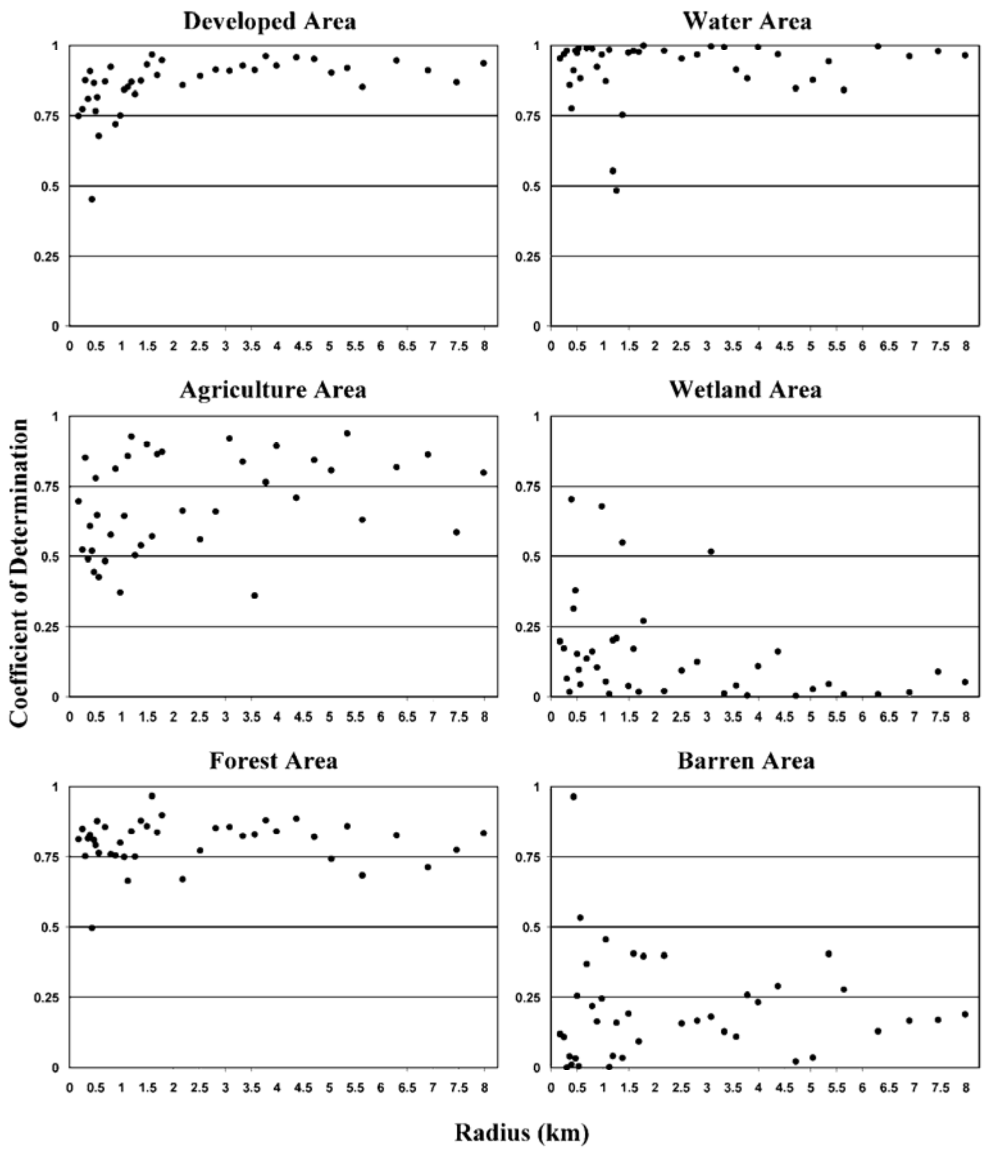
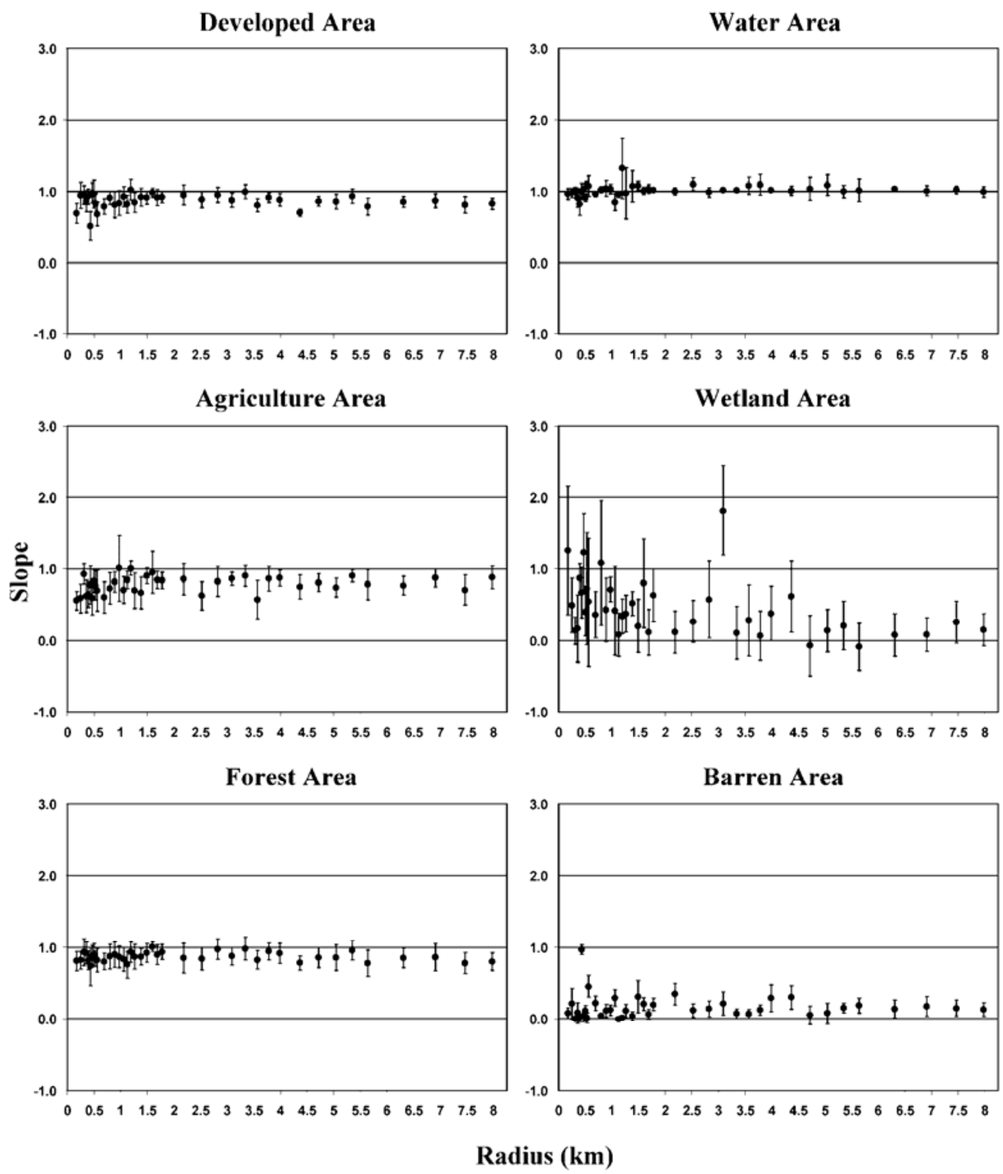


Figure 1-4. Slope of the regression for area (NLCD = $ß_0 + ß_1$REFERENCE) plotted as a function radius of the spatial extent. Error bars represent 95% confidence limits. Slopes not significantly different from 1 represent high accuracy.



Chapter 2. Assessing the Role of Scale in Landscape Structure/Estuarine Sediment Metal Concentration Relationships

(This Chapter is written in the format required for publication in the journal *Landscape Ecology*.)

Introduction

Using landscape pattern to predict the ecological condition of rivers, lakes and estuaries is a theme common to many recent ecological studies, all of which suggest that anthropogenic activities negatively impact ecological condition. One common approach is to correlate metrics of ecological condition (e.g., nutrients, biota, and sediment metals) with the amount of developed lands, agricultural lands, and/or wetlands in estuarine watersheds (Comeleo et al. 1996, Carpenter et al. 1998, Turner et al. 2001, Paul et al. 2002). These land use/land cover (LULC) classes are known to either be sources of pollutants (e.g., developed and agricultural lands) or sinks (e.g., wetlands) which act as filters and remove pollutants from surface flow of water before they enter estuaries (Kadlec 1979, Novotny and Olem 1994). These well-studied relationships suggest that predictive models of estuarine condition may be built from fairly general measures of landscape structure.

Although simple in concept, landscape structure is difficult to quantify. How one decides to calculate landscape structure metrics will have a profound impact on the relationships between ecological processes and the landscape. The scale of investigation also affects the relationships between landscape structure and ecological condition (Turner et al. 1989, Wickham and Riitters 1995, O'Neill et al. 1996). Furthermore, scale is seen as a fundamental characteristic of all ecological studies (Levin 1992, Turner et al. 2001) and how scale is defined may profoundly impact the results of an analysis.

The individual aspects of scale -- grain (i.e., the spatial resolution of a dataset) and extent (i.e., the size of a study area) -- have received considerable attention in the literature and both have been shown to affect landscape metrics as well as ecological phenomena (Urban and Shugart 1987, Wiens 1989, Turner et al. 1989, Wickham and Riitters 1995, O'Neill et al. 1996). Grain and extent often act independently on measures of landscape structure. From a practical standpoint, choosing an appropriate grain is often dictated by the data (e.g., 30 m pixels from LandSat Thematic Mapper imagery) thus researchers need only decide upon an appropriate extent. Research results on appropriate extents for analysis are ambiguous (Omernik et al. 1981).

The extent of a study area influences the relationships between landscape structure and the ecological condition of receiving waters (Omernik et al. 1981, Comeleo et al. 1996, Gergel et al. 1999a, daSilva 2004). Typically, studies measure an element of landscape pattern by sampling land use/land cover within sampling units of varying extent (i.e., a full watershed, a buffer around a water quality sampling point, or riparian buffers), generate metrics of landscape structure for the sampling units, and compare the structure-condition relationships at each extent. Some studies found that the riparian zone and land use/land cover close to the sampling site to be most important in describing water quality or biological condition (Gergel et al. 1999a, Paul et al. 2002, daSilva 2004). Whereas, other studies have found that land cover in full watersheds is a better predictor of water quality than local factors (Omernik et al. 1981). These seemingly contradictory results are likely a product of variation in the temporal scale of the data, spatial scale of the data, and variation in response variables (Turner et al.

2001). Consensus has not been reached on what constitutes an appropriate or "correct" scale to conduct regional studies of land use and ecological condition (Turner et al. 2001).

The objective of this study is to examine how changing spatial scale alters the relationships between landscape structure and contamination of estuarine sediments of the Mid-Atlantic region of the United States. I am specifically interested in quantifying how the relationship between sediment metal concentrations and land use varies among different sampling extents away from an estuary. Additionally, I am interested in understanding how land use relationships vary among toxic metals. The results of several studies suggest that stronger relationships between land use/land cover metrics and estuarine sediment metal concentrations will occur at more local scales (Gergel et al. 1999b, Paul et al. 2002, daSilva 2004).

Methods

I examined how the linear relationship between landscape structure and estuarine sediment metal concentrations is affected by changes in the scale (extent) of the sampling unit used to calculate the landscape metrics. My study included estuaries that extended from Narragansett Bay in Rhode Island to the North Carolina/South Carolina border. Sediment metal concentrations were obtained from the Environmental Protection Agency's Environmental Monitoring and Assessment Program - Estuaries dataset (EMAP-E). Landscape structure was derived from the National Land Cover Dataset (NLCD), point source estimates were taken from the National Coastal Pollutant Discharge Inventory, and I delineated watersheds using the National Elevation Dataset (NED). Each dataset contains data for the entire United States and were all collected in the 1990's, mostly from 1990-1993 (Table 2-1).

The EMAP-E datasets are monitoring results collected by the USEPA and other federal agencies and were designed to estimate status, trends and changes to our Nation's ecological resources (SAB 1988, Paul et al. 1999). The EPA uses a probabilistic sampling scheme to collect data on numerous metrics that describe the physical, chemical, and biological condition of estuaries throughout the United States. In each estuary, EMAP tested for a total of 15 metals, nine of which have been shown to have adverse affects on benthic biota. Of these nine, seven (As, Cd, Cr, Cu, Pb, Hg, and Zn) have available point source pollution estimates from the National Oceanic and Atmospheric Administration's National Coastal Pollution Discharge Inventory (Pacheco 1993, Long et al. 1995, Paul et al. 2002).

The NLCD is a Land Use/Land Cover (LULC) dataset developed by the Multi-Resolution Landscape Characteristics Consortium, a coalition of federal agencies, and is based on Landsat Thematic Mapper imagery of the conterminous United States acquired from 1990 to 1993 (http://landcover.usgs.gov/natllandcover.html). Using an unsupervised classification algorithm, along with ancillary data, a total of 21 thematic classes were derived (Lunetta et al. 1998, Vogelmann et al. 1998a, Vogelmann et al. 1998b, Vogelmann and Wickham 2000). The NLCD has been widely used and has been shown to have acceptable accuracy for broad scale studies of landscape structure (Hollister et al. 2004). The National Elevation Dataset, a 1:24,000 scale, 30 m resolution digital elevation model for the United States, was used to generate watersheds for each EMAP-E sampling station. At the time of this study, the NED was the best quality, widely available elevation. Since completion of analysis, the Shuttle Radar Topography Mission (SRTM) data has been completed and would provide a viable alternative source of data. Detailed descriptions of data compilation, data filtering and GIS methods used to subset and combine these datasets are provided in **Appendix 1**.

Defining Estuarine Condition

Many different measurements have been used to reflect ecological condition and include nitrogen, phosphorus, carbon, Indices of Biotic Integrity (IBI), and sediment metals (Hunsaker and Levine 1995, Comeleo et al. 1996, Paul et al. 2002, Morley and Karr 2002). The choice of variable is often a function of the goals of the study or availability of data rather than an individual metric's suitability for describing ecological condition. No single metric is capable of providing a clear picture of overall ecological condition and several metrics are required to describe different aspects of overall condition (Karr and Chu 1999). Because my research extends the work of Comeleo *et al*. (1996) and Paul *et al.* (1999), I have chosen to use similar dependent variables. They explored the relationships between landscape structure and sediment metal concentration data, among other variables, from EMAP-E. Their response variable was the result of a Principal Component Analysis and not the concentrations of individual metals. The result of the Principal Component Analysis is an excellent synthesis of the overall metals; however, it does not provide information on the individual metals. Since I am interested in how each individual metal relates to landscape structure, I used sediment concentrations of the seven individual metals shown to have biological affects to represent ecological condition.

Defining Landscape Structure

The structure of landscapes is defined by composition and configuration. Composition describes the numbers and abundance of individual landscape elements and configuration describes where the landscape elements occur in the landscape as well as their spatial relationships to other elements (McGarigal and Marks 1995, Gustafson 1998, Turner et al.

2001). Also important in describing landscape configuration, which is difficult to measure and its overall role in describing ecological and environmental phenomena is an active area of research (McGarigal and Marks 1995). In this study, I assessed landscape composition only and only explored the USGS Level I developed and agriculture LULC classes from the NLCD (Anderson et al. 1976, Yang et al. 2001). I use only these generalized (USGS Level I) land cover classes because they are more accurate than the modified USGS Level II classification used in the raw NLCD dataset, and they have been repeatedly shown to be related to various metrics of estuarine condition (Comeleo et al. 1996, Paul et al. 2002, Weller et al. 2003, Stehman 2003, Hollister et al. 2004).

Two commonly used metrics of landscape structure are total area and proportion of individual landscape elements. Where sampling units are of similar size, either may be used since the proportion is merely the total area divided by a constant (i.e., the area of the sampling unit). However, if sampling units vary in extent, proportion and total area reflect different landscape patterns. Proportion of land use may not accurately reflect estuarine condition because two watersheds may have similar proportions of a given land use; yet, the total area of that land use within each watershed may vary greatly. Therefore, I measured landscape composition by total area; however, I still use proportion of each land use to make comparisons within and among watersheds.

Defining Scale and Multi-scaled Watersheds

Spatial scale is defined by both the grain and extent. The grain of my datasets is 30 meters and is set by the NLCD and NED datasets. The extents of my analyses are variable and were set to 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 25, and 30 km from each EMAP-E sampling station that was within the watershed of each sampling station. I generated these extents by selecting all pixels less than or equal to the aforementioned distances, clipping the NED to just those pixels, and calculating the drainage area of the subset NED. These areas represented all area within a given distance that drained to a sampling station and were finally used to subset the LULC data. I refer to each of these distance subsetted watersheds as a multi-scaled watershed. Detailed explanation of the methods I used to create the multi-scale watersheds is given in **Appendix 1**.

The larger, multi-scaled watersheds had considerable overlap. This may seem contrary to the definition of a watershed; however, the methodology of delineating estuarine watersheds and specifically, delineating watersheds subset by distance has some unique aspects that may cause overlap. For instance, two sampling stations may occur in different bays of the same estuary and hence have very similar watersheds or the estuary may occur in a very flat area and the watersheds in those instances are difficult to delineate and are driven by the Euclidean distance from the sampling station and not hydrology. Because these multi-scaled watersheds may overlap, I wanted to assure independence in the data. Therefore, I removed all overlapping multi-scaled watersheds. Starting with the northernmost of the multi-scaled watersheds, I removed all other multi-scaled watersheds that overlapped the northernmost one. Once these were eliminated, I moved to the next remaining watersheds and repeated the process until a complete set of non-overlapping multi-scaled watersheds was obtained for each distance. This resulted in variation in the sample size for each of the different multi-scaled watersheds (Table 2-2).

Linear Relationships

My goal is to measure how relationships between land cover and sediment metals vary with scale; however, other factors also partially explain variation in sediment metals. Not accounting for these additional sources of variation would likely confound the relationship between sediment metals and landscape structure and mask any scaling functions. I de-trended the sediment data by regressing the raw sediment metal concentrations against point source estimates of metal loadings (accounts for an additional source of metals to estuarine sediments) and total area of the multi-scaled watershed (accounts for variation in the total area of a given land use that is caused by different multi-scaled watershed sizes). Point source estimates were taken from the NCPDI dataset and total estimated annual input for each metal was summed for all point sources that overlapped each of the multi-scaled watersheds. These data represent point sources from 1991 and prior. The residuals from this regression (i.e. the variation in sediment metal concentrations unaccounted for by point sources and total multi-scaled watershed area) were then correlated with the multi-scaled metrics of landscape composition. I will refer to this method as a de-trended correlation analysis(Zar 1999). The resulting correlation coefficients provide a measure of the strength of the linear relationship between estuarine sediment metal concentrations and landscape structure at different scales while holding constant the effect of point sources of metals and total watershed area. A t-test was used to calculate significance ($\alpha = 0.05$) of the de-trended correlation coefficients (Cohen and Cohen 1983, Zar 1999).

Scaling Relationships

I hypothesized that the strength of the relationship between landscape structure and estuarine sediment metals would exhibit a negative scaling trend. That is, landscape composition at smaller scales would better describe the variation in sediment metals than landscape structure at larger scales. I tested this on metals that showed significant de-trended correlation coefficients by regressing these correlation coefficients against the scale of the multi-scaled watershed (e.g. 2.5 km, 5 km, 7.5km, etc.) and tested the significance of the slope (H_0: $\beta_1 = 0$, $\alpha = 0.05$). A significant slope indicates that a trend exists between the strength of the landscape structure/estuarine sediment metal concentration relationship and scale.

Results

Estuarine Sediment Metal Concentration and Landscape Structure

Estuaries within the study area had a wide range of sediment metal concentrations with all metals at or near zero in some estuaries and other estuaries showing severely impacted conditions (Table 2-2). Distributions were positively skewed for all metals as indicated by the consistent differences between mean and median values of sediment metals, suggesting that higher concentrations of toxic metals in estuarine sediments are a relatively rare occurrence. Landscape composition of developed and agriculture classes in all multi-scaled watersheds ranged widely, as did the sizes of all multi-scaled watersheds (Table 2-3). The distribution of each LULC class remained positively skewed regardless of the size of the multi-scaled watershed. In general, the magnitude of mean and median of total area of developed and agriculture land increased as the size of the multi-scaled watersheds increase; however, proportion of developed land decreased as watershed size increased and the proportion of agricultural land increased as watershed size increased. This indicates that the relative proportions of different LULC classes change as the size of the watershed changes. Specifically, developed lands, on average, seem to be concentrated in coastal areas and agricultural lands are more concentrated in areas away from the coast (Table 2-3 and Figure 2-1). Figure 2-1b is an example of how drastically proportion and area can differ as distance from a sampling station is increased. Smaller multi-scaled watersheds for this specific example were almost entirely dominated by urban land while the larger multi-scaled watersheds were not.

Linear Relationships

Results for the regressions used to remove the effects of point sources of pollution and total area of the watershed are listed in **Appendix 2**. In all cases, point source inputs were significantly related to sediment metal concentrations. Several of these regressions also had adjusted r^2 considerably higher than expected. In those instances, a single outlier was usually driving the regression. Table 2-4 lists de-trended correlation coefficients between each of the seven metals and developed and agricultural LULC classes for all multi-scaled watersheds. The agricultural land cover classes showed weak, non-significant de-trended correlations with all metals (Table 2-4). Copper and mercury showed a single significant correlation at the 2.5 km distance and the 15 km distance, respectively. In both cases the magnitude of the correlation was relatively small. The amount of developed land showed significant correlations for many metals. Copper and lead showed the strongest average relationships with developed land. Cadmium, mercury, and zinc, were significantly related at many extents and showed only a moderate association with developed land. Arsenic and chromium were generally weakly related with total developed land.

Scaling Relationships

Copper showed significant relationships in all but the largest of the multi-scaled watersheds. Cadmium and mercury showed significant relationships in the smaller extents up to the 17.5 km and 15 km extents, respectively (Table 2-4). The regressions for these metals of de-trended correlation coefficients and size of multi-scaled watersheds showed a negative trend; however, these trends were not significant (Table 2-5 and Figure 2-2). Lead and zinc exhibited similar patterns of significant de-trended correlations and the regressions for these metals had significant, negative slopes (Table 2-5 and Figure 2-2).

Discussion and Conclusions

The goal of this research was to examine the relationships between the sediment concentration of a number of different metals and total area of developed and agricultural lands in watersheds of varying extent. Agricultural lands consistently showed weak correlations with all metals. Conversely, developed lands consistently showed significant, positive correlations with a number of metals. Numerous prior studies have also found strong relationships between developed land and condition of receiving waters (United States Environmental Protection Agency 1995, Comeleo et al. 1996, Wang et al. 1997, Dauer et al. 2000, Paul et al. 2002). The significant relationships seen between metals and developed land are as expected. As shown in Table 2-6, many of the known sources of metals are strongly associated with developed lands (Schroeder 1974, Nriagu and Pacyna 1988, United States Environmental Protection Agency 1990, Novotny and Olem 1994, Satarug et al. 2003, Monbet 2004, Blake et al. 2004). For instance, transportation- and automobile-related sources of toxic metals are often flushed in receiving waters directly from road runoff (Novotny and Olem 1994). Developed landscapes are presumed to have higher road densities; thus, the total amount of contaminated runoff would also be expected to be higher.

Another factor that may explain the positive relationship between toxic metals and developed land is that both toxic metals and developed land are long term impacts to an estuary and watershed. Turner et al. (1998, 2001) have shown that developed lands are unlikely to undergo change and represent a historic "footprint" on the landscape whereas, agricultural lands often change and may become either re-forested or developed. This suggests that the total amount of developed land in a watershed is unlikely to experience large changes and thus may be a good measure of the long-term cumulative impacts to that watershed. Similarly, estuarine sediments are known to be sinks of toxic metals and therefore, are also a measure of long-term impact (Duinker and Nolting 1978, Kennish 1986). Unlike ephemeral conditions such as dissolved oxygen and suspended sediments, concentrations of metals in estuarine sediment are conservative and persist for extended periods of time.

Individual metals varied in the strength of their relationship with landscape structure. None of the metals showed strong relationships with the total area of agricultural land (Table 2-4). For the most part this was expected, as the primary sources of many of these metals are more closely associated with developed land (Table 2-6). However, the lack of a relationship between arsenic and agriculture was unexpected. Arsenic is commonly found in many agricultural products (Table 2-6) and thus, was expected to relate with the total agricultural land. The lack of relationships does not necessarily suggest that agricultural lands are an unimportant non-point source of arsenic or other metals. It is likely that other factors associated with agricultural lands should be explored. For instance, vegetated riparian habitats between agricultural lands and watercourses can act as pollutant sinks and trap metals before they enter the river or stream (Jones et al. 1997, Weller et al. 1998). Also, particular agricultural practices are also likely to play a role. The type of agriculture being conducted, frequency and amount of pesticide and fertilizer application, and specific farming techniques will all cause variation in the amount of metals being introduced into estuarine waters and would not necessarily be captured by a single, simple metric of agricultural area.

With the exception of arsenic and chromium, which were weakly correlated with developed lands, all other metals exhibited significant, positive relationships with the total amount of developed land (Table 2-4). The weak, largely non-significant relationships between arsenic and developed land was expected because arsenic is introduced into a watershed through activities that are not usually associated with developed land (Table 2-6). The weak association between chromium and developed land was unexpected as many of these sources are associated with developed land. Variation in chromium inputs from industrial wastes may have been accounted for by the point source contributions, which I removed from the data. The strongest relationships were seen with copper and lead. Cadmium, mercury, and zinc were moderately associated with developed land. Some of this variation may be explained by unmeasured factors, such as historic point sources that are not included in the NCPDI. In general though, the different levels of association between each of these metals and developed land is somewhat confusing because the primary sources for these metals are all found in lands classified as developed (Table 2-6).

One potential source of variation that may explain the different levels of association is atmospheric deposition. Each of these metals has atmospheric sources and monitoring the magnitude and extent of atmospheric deposition is an active area of research and of interest to a wide variety of groups and agencies (see http://www.epa.gov/castnet for a review of national monitoring programs). Tracking the fate of metals that are delivered to estuarine watersheds by atmospheric deposition would be challenging because of the complex pathways taken by metal-bearing soils, sediments, and biological agents.

Finally, developed land is an aggregate land use class that contains a variety of land uses that range in intensity as well as relative proportion of impacts to the environment. For example, developed land may contain, high and low intensity residential lands, transportation, and industry. By aggregating all these classes into a single class, this source of variation is masked and thus may be confounding my ability to discern the relationships between developed land and estuarine sediment metals. Although most regional and national LULC data for the early 1990's provide these classes, the accuracy of the classes is relatively low (Vogelmann et al. 2001, United States Geological Survey 2003). Furthermore, since metal concentrations in estuarine sediments can accumulate over long periods of time, land-use history and specifically the time since development is likely an important factor. Simply examining the total area of developed land at one time does not provide insight in past land uses. For example, pesticides are a source of some metals. Extensive areas of agriculture are frequently converted to residential development, especially near urban and suburban regions (Turner et al. 2001). Low density residential development, which is not a significant source of metal pollutants, masks a previous land use that could have been a major source.

Scale Effects

Previous studies exploring the role of scale have developed independent models at a variety of scales (i.e. independent variables are not held constant across scales), compared the strength of the relationships for each scale specific model and evaluated the role of scale from these relationships (Comeleo et al. 1996, Gergel et al. 1999a). Developing each model independently may obscure the role of scale because changes in the strength of the relationship may be a function of changes in the independent variables as much as a function of changing scale. To test this, I compared the strength of the relationship for each metal across scales by regressing the de-trended correlation against the distance of the multi-scaled watershed.

For the five metals that demonstrated significant correlations with developed land (cadmium, copper, lead, mercury and zinc), all showed negative trends with watershed extent and lead and zinc showed statistically significant negative correlations (Figure 2-2, Table 2-5). If this pattern were manifest with only a single metal, then it would be possible to consider it a spurious result; however, all metals exhibit a similar pattern, suggesting land cover and anthropogenic activities in close proximity to sampling stations are better predictors of environmental condition. This conclusion is corroborated by several studies using similarly scaled data and a similar suite of predictors (Comeleo et al. 1996, Gergel et al. 1999a). All conclude that local scale landscape factors are more important than factors at broader scales.

One important caveat to this conclusion is that a majority of these studies have focused on broader scales. For instance, the smallest scale included in this study was 2.5 km from a sampling point. For many studies this may likely be considered "broad scale". Obviously, not all processes act on these same scales and conform to the same definitions of broad and fine scale. For instance, one study in Rhode Island explored the role that scale plays in describing variation in stream benthic condition and concluded that local scale factors were more important than full watersheds (daSilva 2004) and in this case, local scale factors were measured within 1000 meters of a stream sampling location and the full watersheds included land area that were approximately the same size as the smaller multi-scaled watersheds in my study.

Additionally, it may be tempting to assume that the linear scale effect exhibited here may exist across all scales and if so, one would expect even stronger relationships at finer scales. Regional LULC datasets, such as the NLCD were intended for regional use and have been proven very effective at these broader scales. However, regional LULC data such as the NLCD may not be sufficiently accurate at small spatial extents (Hollister et al. 2004). Local scale structure may in fact be very important in describing ecological condition, but the data used to characterize this must be applicable to those scales to find those relationships. Nevertheless, the trends are consistent; local factors are better predictors of ecological condition that distant factors.

Peripheral Issues

My research shows that metals in estuarine sediments have a strong relationship with the total amount of developed land and positive correlations are most pronounced when disturbance is close to sampling stations. I removed the effects of watershed area and major point source inputs from my measures of sediment contamination. Other factors such as atmospheric deposition, sediment grain size, and total estuarine area may influence the total concentration of metal observed at a givenlocation, but they are not likely to change with scale, and therefore were not considered in this analysis. Future analyses of these data should consider these additional factors. Second, the de-trended correlations reported here represent a single snapshot of this relationship. The variation in the correlations from one scale to another may partially be due to sampling variability. If a number of correlations were calculated at each scale (i.e. via a bootstrap) then the mean correlation would be a more stable statistic to compare with scale, but at this time that is merely speculation. Third, this analysis only considered linear relationships. For these types of studies, linear models and log-linear models have been successfully used (Comeleo et al. 1996, Gergel et al. 1999b, Paul et al. 2002). Plotting the raw values on the y-axis and total developed land on the x-axis for a variety of scales suggests that a linear relationship is present (Figure 2-3). If non-linearities are present, they are not readily apparent.

Summary

It has long been known that anthropogenic uses have a negative impact on the quality of riverine, lacustrine, and estuarine waters and sediments. It has also been recognized that scale plays an important role in accurately describing many ecological processes. The purpose of this research is to build on other studies and further explore the intersection of these two ideas. As such, I explored how varying scales used to calculate landscape structure altered the linear relationship with sediment metal concentrations. Additionally, I was interested in the relationships between specific metals and total area of two land use classes, developed land and agriculture. The measure of landscape structure that was consistently related to sediment metals was total amount of developed lands. Total agricultural land was not associated with metal concentrations of estuarine sediments. Of the seven metals explored (Zn, Hg, Pb, As, Cd, Cr, and Cu), lead and copper showed the strongest relationships with developed land. Arsenic and chromium showed relatively weak association with developed land. Cadmium, mercury, and zinc showed moderate associations. Finally, lead and zinc showed a significant negative scaling trend. Cadmium, copper, and mercury also showed negative scaling trends, but they were not significant. This suggests that landscape structure at distances less than approximately 15 - 20 km are more important for describing the relationship between developed land and estuarine sediment metals.

These results have implications for future modeling studies. First, developed land is an appropriate predictor of sediment metal concentrations in estuarine sediments, although other factors associated with developed land may also be important. This is corroborated in prior studies in both estuarine and fresh water systems (Comeleo et al. 1996, Paul et al. 2002, Hale et al. 2004). Second, for studies relating condition of receiving waters to human activities in watersheds an emerging consensus appears to be that landscape factors closer to a sampling point are more important than distant factors in the watershed. Since a range of "local" extents may potentially be appropriate and no single extent is necessarily "correct," other factors associated with a specific project's goals (i.e., needed sampling size or known scaling properties) should also be considered. As such, we are reminded that identifying a "correct" scale is often as much science as it is art (Levin 1992).

Table 2-1. Data sources used to explore scaling relationships between landscape composition and sediment metal concentrations.

Dataset Name	Dataset Source	Online Linkage	Temporal Extent
National Land Cover Dataset (NLCD)	USGS	http://landcover.usgs.gov/natllandcover.asp	1990-1993
Environmental Monitoring and Assessment Program - Estuaries (EMAP-E)	U.S. EPA	http://www.epa.gov/emap	1990-1997
National Elevation Dataset (NED)	USGS	http://ned.usgs.gov	-NA-
National Coastal Pollutant Discharge Inventory (NCPDI)	NOAA	NONE	1991 and prior

Table 2-2. Descriptive statistics for sediment metal concentrations. Concentrations are reported in ug/g.

Zinc	Mean	Median	S.D.	Mercury	Mean	Median	S.D.
2.5 km (n=105)	118.38	90.60	105.05	2.5 km (n=105)	0.21	0.08	0.44
5 km (n=90)	109.68	86.10	98.17	5 km (n=90)	0.15	0.07	0.28
7.5 km (n=75)	114.76	90.60	97.64	7.5 km (n=75)	0.21	0.07	0.47
10 km (n=64)	109.94	86.95	89.33	10 km (n=64)	0.14	0.07	0.23
12.5 km (n=53)	99.08	77.80	77.34	12.5 km (n=53)	0.11	0.05	0.19
15 km (n=47)	109.70	86.80	98.25	15 km (n=47)	0.14	0.05	0.23
17.5 km (n=41)	95.90	76.00	89.90	17.5 km (n=41)	0.16	0.04	0.36
20 km (n=33)	105.57	87.50	92.86	20 km (n=33)	0.11	0.04	0.14
25 km (n=27)	104.25	77.70	106.75	25 km (n=27)	0.11	0.04	0.15
30 km (n=27)	93.54	56.70	84.43	30 km (n=27)	0.12	0.03	0.25
Lead	**Mean**	**Median**	**S.D.**	**Arsenic**	**Mean**	**Median**	**S.D.**
2.5 km (n=105)	44.01	29.70	50.25	2.5 km (n=105)	8.89	7.40	7.41
5 km (n=90)	36.97	27.55	38.28	5 km (n=90)	7.79	6.48	6.46
7.5 km (n=75)	41.60	29.20	45.81	7.5 km (n=75)	8.26	6.60	6.72
10 km (n=64)	36.56	27.80	36.82	10 km (n=64)	7.52	6.91	5.11
12.5 km (n=53)	33.74	24.30	32.64	12.5 km (n=53)	7.19	6.32	5.33
15 km (n=47)	41.89	22.00	57.94	15 km (n=47)	7.20	6.32	5.97
17.5 km (n=41)	35.95	21.00	44.57	17.5 km (n=41)	6.32	5.34	5.33
20 km (n=33)	38.74	21.78	58.24	20 km (n=33)	7.29	5.80	5.78
25 km (n=27)	34.01	21.78	31.18	25 km (n=27)	6.82	5.64	6.73
30 km (n=27)	31.65	16.00	38.29	30 km (n=27)	6.30	4.60	6.14
Cadmium	**Mean**	**Median**	**S.D.**	**Chromium**	**Mean**	**Median**	**S.D.**
2.5 km (n=105)	0.48	0.24	0.80	2.5 km (n=105)	57.23	52.40	46.14
5 km (n=90)	0.42	0.23	0.78	5 km (n=90)	53.57	49.20	45.63
7.5 km (n=75)	0.49	0.27	0.86	7.5 km (n=75)	54.61	52.00	37.70
10 km (n=64)	0.37	0.24	0.40	10 km (n=64)	50.70	51.70	31.84
12.5 km (n=53)	0.36	0.24	0.41	12.5 km (n=53)	47.54	48.80	30.79
15 km (n=47)	0.41	0.23	0.58	15 km (n=47)	49.68	49.60	32.21
17.5 km (n=41)	0.47	0.21	1.06	17.5 km (n=41)	45.81	39.50	36.52
20 km (n=33)	0.37	0.21	0.45	20 km (n=33)	48.74	52.00	34.79
25 km (n=27)	0.39	0.27	0.55	25 km (n=27)	55.78	48.80	62.97
30 km (n=27)	0.32	0.21	0.44	30 km (n=27)	42.64	30.80	32.20
Copper	**Mean**	**Median**	**S.D.**				
2.5 km (n=105)	37.90	20.40	55.19				
5 km (n=90)	33.12	18.00	51.74				
7.5 km (n=75)	37.47	22.50	55.68				
10 km (n=64)	32.73	19.30	46.99				
12.5 km (n=53)	29.54	18.10	43.63				
15 km (n=47)	33.45	17.30	49.36				
17.5 km (n=41)	34.64	13.50	60.06				
20 km (n=33)	30.25	19.60	44.59				
25 km (n=27)	33.58	17.30	55.72				
30 km (n=27)	32.28	10.70	56.24				

Table 2-3. Descriptive statistics for LULC total area and proportion.

	2.5 km						5 km					
	Mean		Median		S.D.		Mean		Median		S.D.	
	ha	%	ha	%	ha	%	ha	%	ha	%	ha	%
Total Area	1,836.43	NA	1,889.10	NA	156.11	NA	7,074.94	NA	7,422.62	NA	752.38	NA
Developed	253.32	14.30%	94.77	5.10%	345.46	19.53%	1,071.91	15.79%	551.97	7.80%	1,248.96	18.94%
Agriculture	129.31	7.08%	49.86	2.74%	171.79	9.49%	720.38	10.17%	415.04	5.49%	788.44	11.10%
	7.5 km						**10 km**					
	Mean		**Median**		**S.D.**		**Mean**		**Median**		**S.D.**	
	ha	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**
Total Area	16,000.51	NA	16,430.31	NA	1,627.27	NA	28,120.91	NA	28,704.20	NA	2,706.99	NA
Developed	2,283.34	14.64%	997.02	6.19%	2,815.80	18.45%	3,488.93	12.90%	1,527.39	5.30%	4,400.12	16.99%
Agriculture	1,839.19	11.60%	1,204.65	7.65%	1,859.67	11.97%	3,432.40	12.15%	2,885.94	9.86%	3,250.62	11.39%
	12.5 km						**15 km**					
	Mean		**Median**		**S.D.**		**Mean**		**Median**		**S.D.**	
	ha	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**
Total Area	43,160.03	NA	44,587.35	NA	5,557.17	NA	61,404.85	NA	62,808.21	NA	7,980.40	NA
Developed	4,914.22	11.97%	1,775.07	3.97%	5,882.43	15.76%	7,448.88	12.13%	2,628.90	5.17%	9,712.89	15.96%
Agriculture	5,635.94	13.11%	3,900.06	8.61%	5,807.81	13.38%	7,402.23	12.30%	5,352.12	8.66%	6,682.24	10.98%
	17.5 km						**20 km**					
	Mean		**Median**		**S.D.**		**Mean**		**Median**		**S.D.**	
	ha	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**
Total Area	84,074.15	NA	88,169.04	NA	12,200.56	NA	106,900.67	NA	107,245.44	NA	15,485.38	NA
Developed	8,884.35	10.74%	2,962.26	3.50%	11,396.76	14.66%	8,538.46	8.02%	3,155.40	3.17%	11,948.97	11.41%
Agriculture	10,508.84	12.59%	7,665.75	8.27%	10,051.96	11.62%	14,576.55	13.68%	12,790.71	13.11%	11,951.79	10.84%
	25 km						**30 km**					
	Mean		**Median**		**S.D.**		**Mean**		**Median**		**S.D.**	
	ha	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**
Total Area	171,596.48	NA	182,475.18	NA	25,322.26	NA	240,129.48	NA	252,402.66	NA	39,351.06	NA
Developed	19,470.39	11.35%	6,416.28	3.96%	26,518.83	15.42%	22,474.93	9.42%	10,184.76	3.99%	28,499.42	12.17%
Agriculture	22,036.14	13.14%	15,665.49	8.51%	19,023.92	11.18%	30,600.95	13.08%	29,727.90	11.38%	27,654.16	11.39%

Table 2-4. De-trended correlation coefficients (*sr*) comparing sediment metal concentrations and Total area of LULC classes within a given distance. Significant de-trended correlation coefficients are bolded (α=0.05).

	2.5 km (n=105)	5 km (n=90)	7.5 km (n=75)	10 km (n=64)	12.5 km (n=53)	15 km (n=47)	17.5 km (n=41)	20 km (n=33)	25 km (n=27)	30 km (n=27)
Developed										
As	0.12	0.11	**0.25**	0.22	**0.27**	0.24	0.18	0.00	0.24	-0.09
Cd	**0.43**	**0.23**	**0.35**	**0.31**	**0.35**	**0.40**	0.10	0.24	0.36	0.22
Cr	**0.21**	**0.23**	**0.27**	0.17	0.14	0.27	0.27	0.18	0.32	0.00
Cu	**0.51**	**0.47**	**0.43**	**0.65**	**0.56**	**0.62**	**0.64**	**0.43**	**0.46**	0.29
Hg	**0.30**	**0.37**	**0.23**	**0.43**	**0.36**	**0.38**	**0.36**	0.17	0.33	0.09
Pb	**0.52**	**0.49**	**0.57**	**0.54**	**0.39**	**0.48**	**0.42**	0.27	0.37	0.16
Zn	**0.39**	**0.31**	**0.38**	**0.37**	0.25	**0.35**	0.25	0.18	0.34	0.04
Agriculture										
As	0.04	0.08	0.05	0.03	0.06	-0.06	0.01	0.12	0.09	0.31
Cd	-0.13	0.02	-0.08	-0.08	-0.06	-0.21	-0.03	-0.05	-0.10	0.00
Cr	-0.07	0.01	0.02	0.08	0.14	-0.06	0.06	0.11	-0.06	0.00
Cu	-0.18	-0.12	-0.10	-0.23	-0.20	**-0.30**	-0.23	-0.15	-0.23	-0.18
Hg	**-0.19**	-0.09	-0.18	-0.15	-0.07	-0.22	-0.20	-0.04	-0.30	-0.08
Pb	-0.12	-0.08	-0.16	-0.11	-0.05	-0.21	-0.15	0.02	-0.19	-0.05
Zn	-0.04	0.02	0.02	0.01	0.03	-0.11	0.03	0.13	-0.01	0.18

Table 2-5. Estimates and statistics for linear regression of de-trended correlation coefficients (sr) on distance of multi-scaled watershed (MSW). Significant regressions indicate a significant scaling relationship between toxic metals and scale of developed land estimates. General form of the regression equation is sr = β_1MSW + β_0.

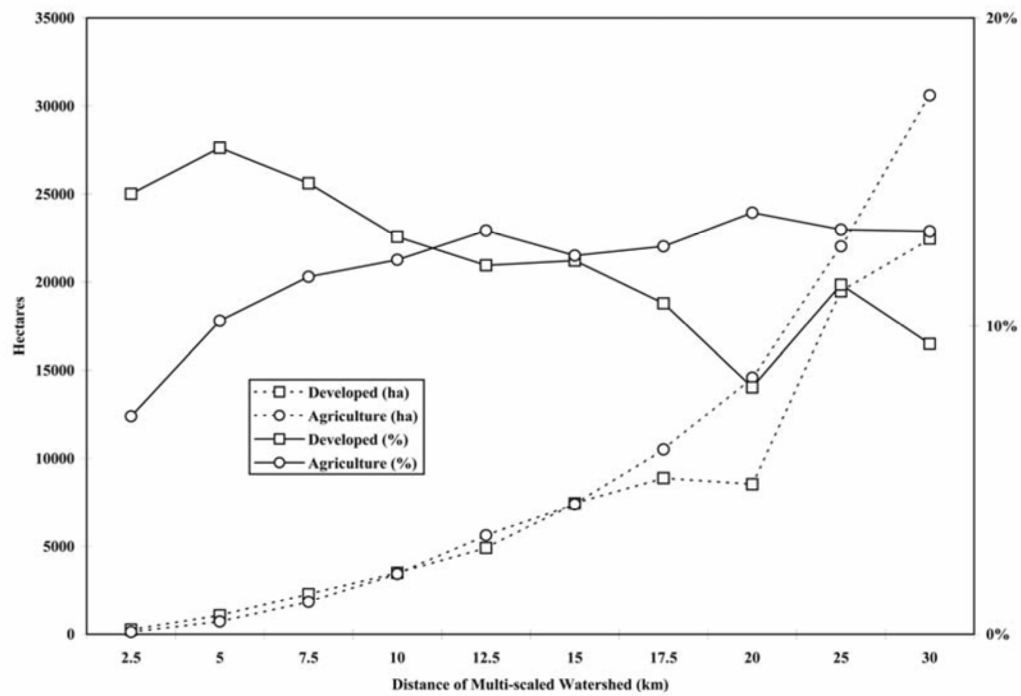
	β_0	β_1	R^2	p-value
Cd	0.357	-0.004	0.124	0.317
Cu	0.582	-0.005	0.162	0.248
Hg	0.391	-0.006	0.265	0.128
Pb	0.601	-0.013	0.734	0.001
Zn	0.417	-0.009	0.541	0.015

Table 2-6. Some common sources of metals.

Metal	Sources	Reference
Arsenic (As)	pesticides, defoliants, and algaecides, atmospheric deposition	Novotny and Olem (1994), Nriagu and Pacyna (1988)
Cadmium (Cd)	phosphate fertilizers, smelting, tires, plastics, and insecticides, atmospheric deposition	Schroeder (1974), Satarug et al. (2003), Monbet (2004), Nriagu and Pacyna (1988)
Chromium (Cr)	metal corrosion, paint, and electroplating, atmospheric deposition	United States Environmental Protection Agency (1990), Nriagu and Pacyna (1988)
Copper (Cu)	marine applications, metal corrosion, algaecides, paints, wood preservative, and electroplating, atmospheric deposition	Blake et al. (2004), United States Environmental Protection Agency (1990), Nriagu and Pacyna (1988)
Lead (Pb)	gasoline, batteries, mining, smelting, and insecticides, atmospheric deposition	United States Environmental Protection Agency (1990), Novotny and Olem (1994), Nriagu and Pacyna (1988)
Mercury (Hg)	batteries, electrical products, fungicide, medical products, mining, smelting, fossil fuels, and chloralkali industry, atmospheric deposition	Novotny and Olem (1994), Nriagu and Pacyna (1988)
Zinc (Zn)	metal corrosion, tires, road salt, wood preservatives, paint, metal corrosion, and roofing materials, atmospheric deposition	United States Environmental Protection Agency (1990), Novotny and Olem (1994), Nriagu and Pacyna (1988)

Figure 2-1. Total area and mean proportion of developed, agricultural, and forested land use/land cover classes plotted as a function of the distance of multi-scaled watersheds.

a. Mean total area and proportion for all samples



b. Total area and proportion for a single sampling station (VA92-527)

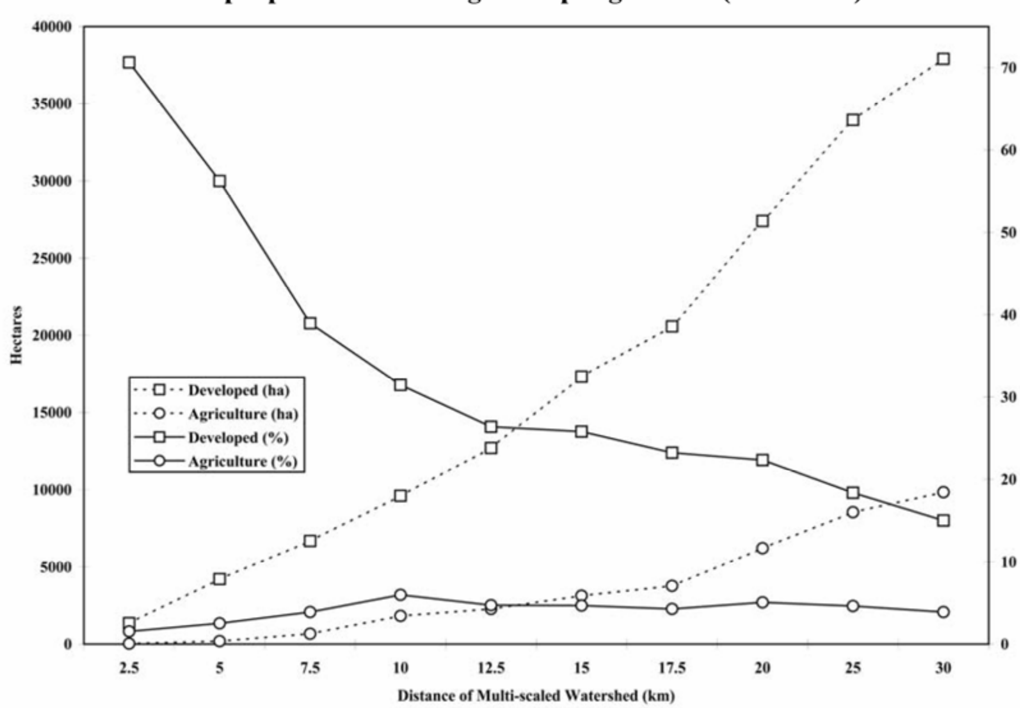


Figure 2-2. De-trended correlation coefficients (*sr)* plotted as a function of the distance of the multi-scaled watersheds (MSW). The lines represents a linear model with the general form $sr = \beta_0 + \beta_1$ MSW.

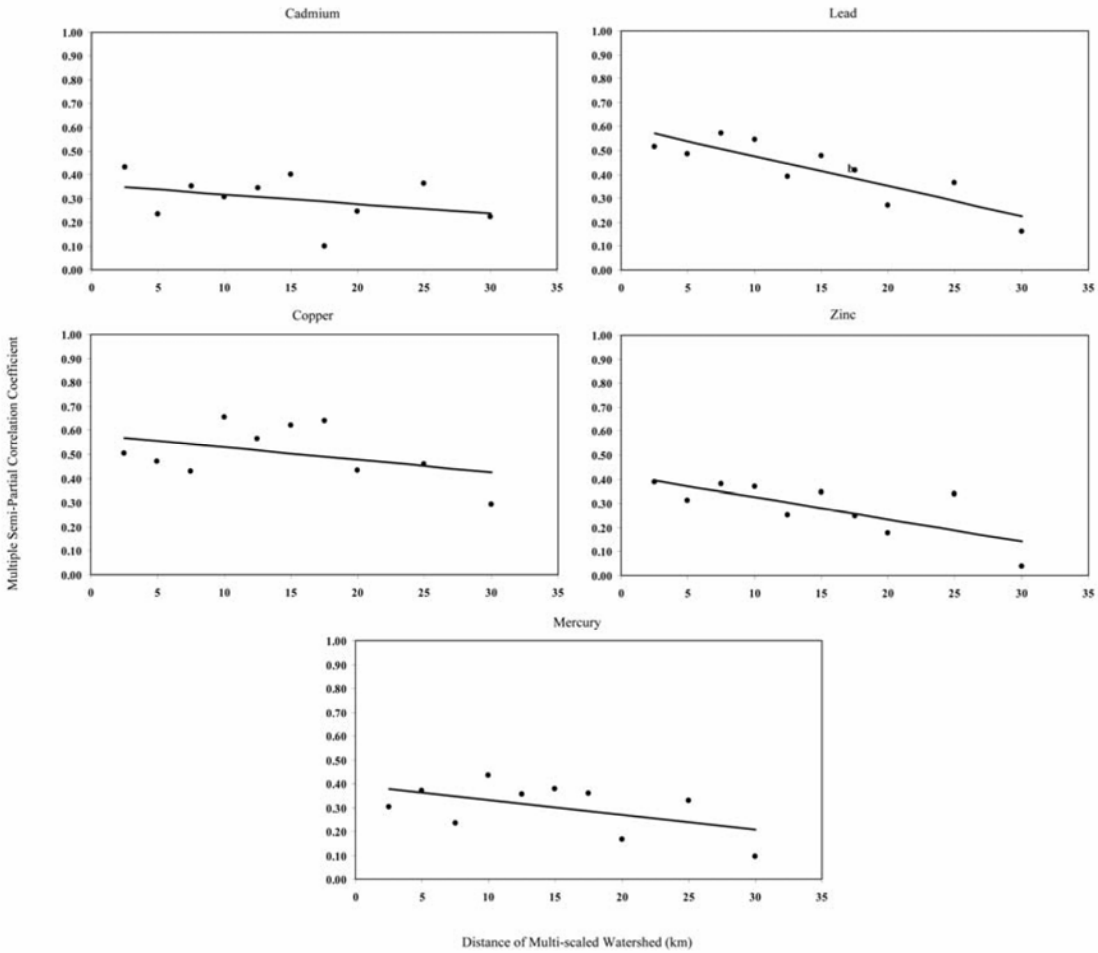
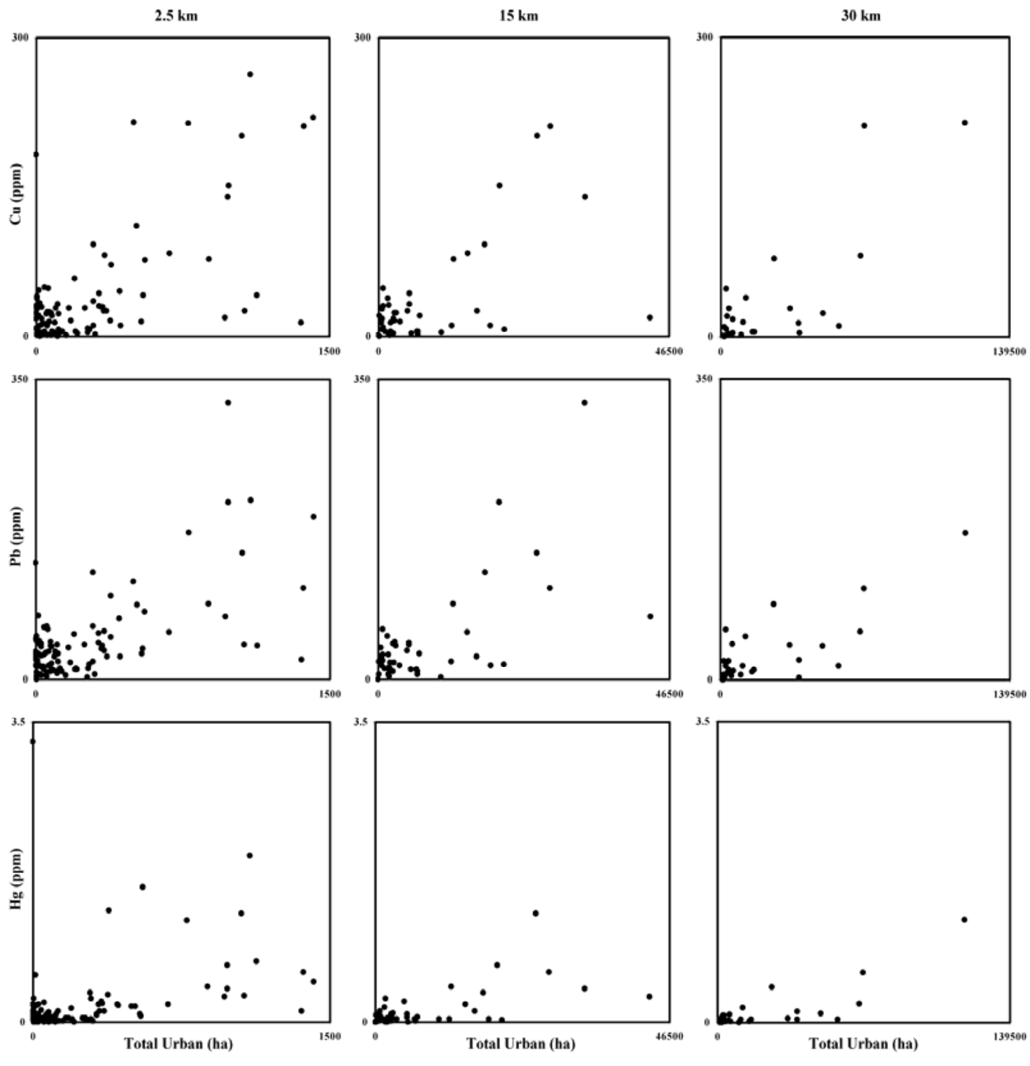


Figure 2-3. Scatterplots of raw data showing concentrations of mercury, lead, and copper versus total developed land at 2.5, 15, and 30 km from the sampling station.



Chapter 3. Predictive Modeling of Estuarine Sediment Metal Concentrations Along the Atlantic Coast of the United States: An Information Theoretic Approach

(This Chapter is written in the format required for publication in the *Journal of Environmental Quality*.)

Introduction

The Atlantic coast of the United States is an area characterized by diverse ecosystems with land use patterns dominated by intensive agriculture, forestry, and urbanization. The demands for food and timber production, recreational opportunities, and housing have had a lasting impact on the spatial patterns of the Atlantic coastal landscapes, which in turn have impacted ecological processes and conditions (Forman and Godron 1986, Turner 1989, Forman 1995). Many studies have linked anthropogenic disturbances, such as increased agriculture, loss of riparian areas, and decreased natural vegetation, to reduction in the surface water and groundwater filtering capacity of landscapes. Loss of natural filtering capacity frequently leads to declines in water quality (Likens et al. 1970, Peterjohn and Correll 1984, Levine et al. 1993, Hunsaker and Levine 1995, Turner et al. 2001, Weller et al. 2003)

Point and non-point pollution from anthropogenic land uses have added sediments, nutrients, pesticides, and toxic metals to fresh, salt and estuarine waterways (Novotny and Olem 1994, United States Environmental Protection Agency 1995, Basnyat et al. 1999). Fertilizer use, sewage and septic inputs, and livestock operations produce large amounts of nitrogen and phosphorus and can contaminate receiving waters (Novotny and Olem 1994, Carpenter et al. 1998). Certain anthropogenic activities are closely associated with land use types and thus, linkages have been found between specific pollutants and land use (Marsalek 1978, Beaulac and Reckhow 1982, Novotny and Olem 1994). For instance, industrial and commercial land uses are important sources of lead, copper, and zinc (Sonzogni et al. 1980). As loadings of nutrients and toxics into estuarine waters increase, the need for monitoring efforts also increase and are necessary to identify degraded and threatened waterways, to measure changes in condition, and to assess the success of mitigation or restoration activities. For example, the EPA's Environmental Monitoring and Assessment Program - Estuaries component (EMAP-E) provides quantitative data on the status and changing condition of our nation's estuaries (United States Environmental Protection Agency 1999). One challenge of large scale monitoring efforts, such as EMAP-E, is that it is impossible to monitor all estuaries all the time. Establishing the relationship between estuarine condition and landscape structure is important because it allows for the prediction of estuarine condition in estuaries that have not been monitored.

Turner et al. (2001) discusses several studies that have modeled the relationships between landscape structure and eutrophication and pollution for freshwater systems; however, a review of the literature reveals that less work has been done in estuarine systems. Comeleo et al. (1996) and Paul et al. (2002) have explored these relationships in small estuarine systems of the Mid-Atlantic region. Using the U.S. Geologic Survey's Land Use Data Analysis (LUDA) data set and the Chesapeake Bay Watershed Pilot Project land cover data set (CBWP), respectively, they found relationships between several measures of landscape composition (i.e., percent forest and percent urban) and sediment contamination. Additionally, in tributaries of Chesapeake Bay, urban land uses have been linked to a variety of measures of estuarine condition and within a variety of study sites in the northeastern U.S.

estuarine condition has also been linked to the amount of urbanized land near sampling stations (Dauer et al. 2000, Rodriguez 2003).

These studies show it is possible to develop predictive models of the ecological condition of estuarine waters. However, given the complexity of the relationships between sediment condition and anthropogenic impacts in coastal watersheds, there are a very large number of possible models that could be developed to predict sediment contamination. I used the relatively new modeling approach based upon information theoretic techniques to assess which model, or suite of models, provides the best predictive performance (Burnham and Anderson 2002). The ultimate goal of this analysis is to identify a simple set of readily available measures of watershed condition that can be used to identify degraded estuaries as indicated by metal pollutant loads in sediments. This process follows the recommendations of the National Research Council of avoiding reliance upon expensive, complex mechanistic models in favor of cheaper, simpler, statistical models (Reckhow et al. 2001).

Methods

Study Site and Data

My study area encompasses the coastal region from southern New England to the southeastern United States (Figure 3-1). This area has been intensively monitored for a number of years, supports a diversity of human activities and levels of disturbance, and contains a large number of estuaries. These factors combine to make the mid-Atlantic coast of the United States an excellent region to build and test predictive models of estuarine condition.

I used several readily available datasets to create the models of estuarine condition. The Environmental Protection Agency's Environmental Monitoring and Assessment Program - Estuaries dataset (EMAP-E) and the EPA mid-Atlantic integrated assessment (MAIA) data provided data on estuary condition and estuarine characteristics; the National Coastal Pollution Discharge Inventory (NCPDI) provided information on point sources of metal inputs; the National Land Cover Dataset (NLCD) was used to estimate all land use/land cover area estimates; and the United States Geological Survey river gauge data and the National Oceanic and Atmospheric Administration tide gauge data comprised estimates of estuarine hydrology. These formed the base datasets for my modeling. Additionally, the National Elevation Dataset (NED) was used to delineate watersheds for estuarine sampling sites (see **Appendix 1**). One of the driving forces behind this project was to develop a methodology that could be easily and quickly repeated; therefore, I only used public domain data so that these methods may be extended to other regions of the country with few requirements for additional data collection. Details on the individual datasets and the data acquisition, compilation, filtering and manipulation procedures used to develop the final dataset are provided in **Appendix 1**.

Predictor variables considered for developing the final models and reasons for inclusion or exclusion in the final models are listed in Table 3-1. In the preceding chapter**,** I determined that cadmium was most strongly related to urban lands within 2.5 km of sampling stations and inside their drainage basin. Copper and mercury showed the strongest relationship at 10 km, and lead at 7.5 km. In addition to generating estimates of developed and agricultural estimates with these distance-limited watersheds (2.5 km for cadmium, 7.5 km for lead, and 10 km for copper and mercury), I also used these to estimate the total annual point source loadings for each of the metals using the NCDPI data. For example, to estimate total loadings of lead, the total annual loading for all reported point sources that fell with 7.5 km of the sampling stations was summed. Although there is not a direct temporal relationship between the NCPDI and the EMAP-E and MAIA data -- the NCPDI is from 1991 and prior, EMAP-E and MAIA range in date from 1990 - 1997 -- it was assumed that the point loading estimate from 1991 would adequately represent total point sources for all years. The total area of the estuary, a hypothesized surrogate for atmospheric deposition, and sediment grain size (i.e. silt/clay percentage) were taken directly from the EMAP-E dataset. Estuarine hydrology is included in the models as freshwater inflow and tidal range. Freshwater inflow and tidal range were estimated from the closest USGS-monitored river or NOAA tide gauge. River gauges were limited to those that fell within the sampling stations' drainage area. Data for each of these variables (Table 3-1) were compiled from 112 EMAP-E and MAIA sampling stations (see **Appendix 1**) and formed the basis for all subsequent modeling.

Defining Estuarine Condition

It is not clear which metrics best quantify ecological condition of estuaries. Generally, ecological condition is best examined by collecting data across a gradient of human impact that assesses condition of the resource while accounting for spatial and temporal variation. This type of information allows for the separation of natural variation from human influenced variation. The EMAP and MAIA datasets were designed to characterize long-term and broad-scale status and trends of ecosystems; estuaries were randomly sampled with a single sampling location at a single point in time. As such, this sampling regime is inappropriate for characterizing fine-scale variation and would not provide reliable information for building mechanistic models of estuarine condition or for making spatially explicit predictions of estuarine condition (e.g., mapping an individual metric across an estuary). However, several studies have shown strong correlations between landscape structure and water quality metrics in both fresh and estuarine waters over broad regions and these correlations suggest that it may be possible to predict estuarine condition based on landscape characteristics (Hunsaker and Levine 1995, Comeleo et al. 1996, Paul et al. 2002, Morley and Karr 2002). For instance, Comeleo *et al.* (1996) and Paul *et al.* (2000) examined the utility of using the amount of urban land, non-forested wetlands, and point sources of pollution as independent variables for predicting overall sediment metal concentrations (i.e. they combined sediment metals with a principal components analysis). I expand on this by using the concentration of the individual metals as my dependent variables. I restricted my analyses to cadmium, copper, mercury, and lead because these metals showed the strongest correlations with developed land in coastal watersheds (chapter 2). Although defining estuarine condition with just sediment metal concentrations will not result in a true or "best" measure of estuarine condition, the models developed with these as dependent variables will have utility because they will predict values that are closely associated with anthropogenic impacts, will have links to existing broad scale data sets and will build on findings of previous studies.

Modeling Approach and Model Development

Prior studies of sediment contamination in estuaries used standard parametric modeling techniques, which provide a powerful means for creating predictive models; however, for these methods to remain robust several assumptions must be met -- equality of variances, multivariate normality, and linearity -- and the data must be replicated (Reckhow 1990, Carpenter 1990, Zar 1999). Due to practical and logistical limitations, these assumptions are rarely satisfied with broad scale environmental data and thus the results of statistical tests may be questionable (Reckhow 1990, Nester 1996, Anderson et al. 2000). Alternative methods to traditional parametric statistics include hierarchical Bayesian modeling and information theoretic approaches. Bayesian modeling methods have been successfully used in estuarine water quality monitoring (Borsuk et al. 2001) and promise to be powerful analytical tools. However, the software tools to perform Bayesian analyses are not yet well developed and are inaccessible to most researchers (Burnham and Anderson 2002). Information theoretic approaches for model selection are relatively new, easily implemented, and are just now beginning to see wider use in the biological and ecological sciences (Burnham and Anderson 2002).

The use of information theoretic methods to select models is straightforward and easy to conduct using standard ordinary least squares regression techniques. Information theoretic approaches, unlike standard regression model selection methods, do not rely upon p-values to measure the quality of a model. Instead, they attempt to directly estimate effects and effects sizes of a "set of candidate models" developed *a priori* and these models are ranked based on the Akaike Information Criteria (Burnham and Anderson 2002). The values of the Akaike Information Criteria (AIC) are unitless and can be converted to AIC differences ($\Delta_i = AIC_{min} - AIC_i$). The Δ_i is converted to an estimate of the model's relative likelihood, the Akaike weight w_i which is computed as follows:

$$w_i = \exp(\Delta_i/-2)/\Sigma \exp(\Delta_i/-2)$$

The Akaike weight may be directly interpreted as the probability that the model is the best model

Development of my set of candidate models was conducted independently of the data (i.e., no "data dredging") and was based upon the published literature and current scientific understanding of the effects of human impact on estuarine sediment metal concentrations. I followed common practice and developed a general global model of the system and created alternative models as nested subsets of the global model (Burnham and Anderson 2002).

Although the information theoretic approach is free from the assumptions required for hypothesis testing, the assumptions required for developing a linear model are still applicable. In order for any of the candidate models to provide a reasonable linear fit, the global model must be assessed for linearity and the residuals examined for homoscedasticity and normality. Adjusted r^2 values were examined for each of the four global models (Cu, Pb, Hg, and Cd). Residuals were examined for homoscedasticity with Levene's Test on the median of the residuals and a Shapiro-Wilk's test was used to assess normality (Brown and Forsythe 1974, Zar 1999). Standard transformations on both the dependent and independent variables were explored to resolve problems with the residuals and fit and the transformations used were determined by both residual analysis and ease of interpretation of the final model

Several factors were also taken into consideration when developing the set of candidate models. First, the number of parameters in any given model, should not be more that 1/10th of the sample size. Second, a combination of observational studies and critical *a priori* examination of potential variables should be used when selecting parameters (Burnham and Anderson 2002). I built models from the list of potential independent variables in Table 3-1. The variables retained in the final list of models were either found to be important in prior studies or theorized to be important. The process of *a priori* model development resulted in the inclusion of 7 parameters (Table 3-1). Several variables that were considered *a priori* were removed from consideration for several reasons (Table 3-1). My final 45 models (Table 3-2) were examined for major violations of multiple linear regression assumptions. It is important to recognize that the final goal of this analysis is development of predictive models of estuarine sediment metal concentrations and not necessarily identification of the mechanisms driving estuarine sediment contamination. As such, any conclusions regarding the underlying processes and mechanisms should be considered the result of a largely exploratory, and not confirmatory analyses.

Multimodel Inference

From the set of 45 parameterized candidate models, I calculated the AIC_c (an adjustment to AIC for small samples), AIC Δ_i, and AIC w_i. In a model selection framework, it is possible to interpret w_i as the probability that a given model is best; therefore the objective is to identify and select the model with the highest w_i. However, were a single "best" model to be used to make inferences, those inferences would be conditional upon that single model and model selection uncertainty would have to be explicitly considered. If it is not clear which of the models from the set of candidate models is "best" as measured by w_i (i.e., $w_i > 0.9$ for any given model), then that single "best" model is unlikely to provide reliable predictions.

Instead of relying on a single "best" model, I used information theoretic model averaging (Burnham and Anderson 2002). Each of the appropriate candidate models may be weighted by the models' Akaike weights and combined into a single, weighted model. This is accomplished by generating weighted parameter estimates for each parameter and making an "averaged" prediction from the combined model (Burnham and Anderson 2002). Model averaging is appropriate when the primary goal is prediction (Burnham and Anderson 2002). Models that had an Akaike w_i of essentially zero (i.e. less than 0.0001) were not included in the calculation of the model averaged parameter estimates. Additionally, the Akaike w_i may be also used to rank the relative importance of each of the variables in the candidate models. Summing the w_i for each model that contains a given variable results in an estimate of the relative importance of each variable.

Accuracy of predictions from the averaged model were assessed by the Mean Absolute Error (MAE) and by the percent of the total range of observed values (MAE / Max. Metal Conc. -

Min Metal Conc.); however, the ecological and biological relevance of these concentrations is not readily apparent from the raw numbers and further investigation is often required. One possibility is to use the Effects Range (ER) values (Long et al. 1995). The ER values are one way (of many) to examine the ecological and biological significance of a given toxic metal concentration. The ER values break down each metal concentration into a categorical variable indicating a low (ERL), medium (ERM), or high (ERH) potential for producing a biological impact. To provide a rough test of the ability of these models to accurately predict estuarine condition, I converted each of the predicted values to the ER values and calculated the accuracy of the ER value predictions with the use of a 3 x 3 error matrix similar to those used to assess the accuracy of thematic spatial data (Congalton and Green 1999).

Results

Descriptive Statistics and Effects Range Values

Many of the variables were positively skewed and in particular the four dependent variables (Cu, Cd, Hg, Pb) all showed fairly dramatic positive skewness indicating that high concentrations were rare (Table 3-3). Similarly, the distribution of the ER values indicates that for this particular sample of estuaries, highly polluted sediments are rare. For copper, 74% of the samples had a low potential for biological impact (ERL) and 26% had a medium risk of biological impact (ERM). None of the samples had an ER classification of high (ERH) for copper. Lead showed a similar pattern with 74% ERL, 24% ERM, and only 2% with ERH. Mercury had 69% ERL, 24% ERM, and 7% ERL; and cadmium had 89% ERL, 11% ERM and no ERH values.

Model Results

To determine if transformations were needed and if the set of independent variables were reasonable, the general fit of the global model was assessed and the residuals were examined for homogeneity of variance and normality. Although statistically significant, the Global models had adjusted r^2 values that were relatively low (0.23 to 0.55), had skewed residuals, and residual plots indicating heteroscedasticity. In all cases, transforming the dependent variables increased the adjusted r^2 and removed heteroscedasticity; however, the residuals were still significantly non-normal (Table 3-4). While transformations (log and square root) of many of the independent variables improved normality and made marginal improvements in the adjusted r^2, they would have added considerable complexity in the interpretation of the final models. Regression is fairly robust to deviations from normality in the residuals as long as the residuals are also homoscedastic (L. Gonzales, pers. comm.). Therefore, I only transformed dependent variables.

The adjusted r^2 of the 45 models ranged from 0 to a high of 0.79 (**Appendix 3)**. The w_i for each model ranged from essentially 0 (i.e. less than 0.0001) to a high of 0.409 (Table 3-2). Since a large portion of the models was found to have a w_i of near 0, meaning there is little if any support for those particular models, only the models with w_i greater than 0 are reported. These models are presented in Tables 3-5 a-d and are ranked according to their w_i. The best model for Cu, Pb, and Hg was model number 39 (metal = urb. + silt/clay). For cadmium the best model was model number 32 (metal = urb. + est. area + silt/clay).

Relative Variable Importance

All of the best models (i.e. $w_i > 0.0001$) contained the total amount of urban land and percent silt/clay content of the sediment. Using the w_i to compare the relative importance of each variable indicates that these two variables are the most important for all metals (Table 3-6). Point sources were the next most important for both copper and mercury. Area of the estuary was the next most important, by a large margin, for cadmium. Agriculture was the next most important for lead.

Model Averaging

The model averaged adjusted r^2 for the copper, lead, mercury, and cadmium models is 0.784, 0.705, 0.564, and 0.503 respectively; and the model averaged standard error of the residuals is 0.604, 0.540, 0.161, 0.242, respectively (Table 3-7). Tests of the assumptions (i.e. homoscedasticity and normality of residuals) for the model-averaged residuals are nearly identical to those of the global models, and thus are not presented.

Overall, the copper model predicted the correct value of ER 83.9% of the time, the lead model ER was correct 84.8% of the time, mercury ER was correctly predicted 78.6% of the time, and cadmium ER was correctly predicted 92.0% of the time (Table 3-8). In general each of the models correctly predicted the ERL values, but failed to accurately predict the ERM, and ERH conditions. Additionally, omission errors were generally lower than commission errors for the ERL values. This was reversed for the ERM values as the models more often classified actual ERM values as either ERL or ERH.

Further examination of the prediction accuracies of the actual concentrations indicates that on average (i.e. MAE), each of the models underestimated sediment metal concentrations. The model-averaged estimates of copper are on average 24.2 ug/g less than the actual concentration. Lead is underestimated by 29.6 ug/g, mercury is underestimated by 0.042 ug/g and, cadmium by 0.14 ug/g. As a percent of the total range of observed values, copper underestimates by 9.20%, lead by 9.16%, mercury by 1.28%, and cadmium by 2.13%.

Discussion

The overall fit for each of the best models was relatively strong and ranged from a low adjusted r^2 of 0.53 to a high of 0.78 and a greater degree of the variation was explained for the copper and lead models than for either mercury or cadmium. Not surprisingly, the measure of fit for the averaged model followed a similar pattern (Table 3-7). Of the complete set of 45 models, only those models that contained both the total hectares of urban land and the percent of silt/clay in the sediment had w_i values that were greater than 0. The adjusted r^2 for models without these variables was substantially lower. This result corroborates several prior studies and suggests that the total amount of urban land is indeed a useful predictor of sediment metal concentrations and that the physical conditions of sediments are important in determining how much of a given contaminant is likely to be present (Kennish 1986, Novotny and Olem 1994, Paul et al. 2002). There was variation in the importance of the remaining variables among the different metals and it was not readily apparent which were the most important. One notable exception was the high relative importance of the area of the estuary for cadmium. Area of the estuary was added to the models as a surrogate for atmospheric deposition. I hypothesized that the total area of the estuary was directly related to the total contaminants from the atmosphere. Total area of the estuarine watershed would likely also be related to atmospheric contaminants, but this simple assumption is possibly confounded by additional sources of metals in the land area of estuarine watersheds as well as sink landscape features (e.g., wetlands) that reduce the amount of metals reaching estuaries. Because of these confounding factors area of the watershed was not included. If my hypothesized relationships between estuarine area and atmospheric deposition are true, then this may suggest an important role for atmospheric deposition in the presence of cadmium in estuarine sediments. Since the link between estuarine area and atmospheric deposition is just a hypothesis, work is needed to further substantiate this claim.

The results from this study suggest promise for modeling with widely available monitoring data; however, for some of the metals, specifically mercury and cadmium, there was still substantial variation that was not accounted for by the models. There are several explanations for this. It is very likely that there are variables not included in the modeling that play a large role in controlling the amount of mercury and cadmium found in estuarine sediments. Furthermore, it is possible that atmospheric deposition of these metals is important, but simple estuarine area is a poor surrogate for atmospheric deposition.

Screening Models

One of the objectives for this analysis was that these models could easily be transferred to other areas and be used in a screening process to limit which estuaries should be included in a monitoring sampling scheme. This would make better use of environmental monitoring resources and allow focusing of data collection in estuaries that are most likely to be imperiled. The apparent importance of the silt/clay measures, or the general importance of sediment physical characteristics, would make it difficult to transfer these models to new estuaries without having specific information about the sediments. This does not necessarily imply that there is little use for these models. If samples had already been collected and resources scarce, then these models could be used to identify samples likely to have very low metal levels. Resources could then be better focused on the samples likely to have problematic levels of toxic metals and further lab analyses conducted on those samples only.

Data Accuracy and Consistency

Because of the expansive geographic extent of this study, it was not possible, nor necessarily desirable, to collect new data. A wide variety of existing broad scale datasets were collected, compiled, and analyzed to conduct this analysis. As with any broad scale study, compiling disparate data sources introduces its own set of problems. The accuracy of the NLCD has been well-studied and has been quantified (United States Geological Survey 2003, Hollister et al. 2004). Although there are a number of known problems with the NCPDI data, these data are currently the only broad scale dataset that contains estimates of total loadings of a variety of pollutants. Therefore, the NCPDI must be considered the best available data on point source loadings and, unfortunately must be used as is (Comeleo et al. 1996). Estimates of estuarine hydrology were derived from both the USGS river gauge data and the NOAA tide gauges. Both of these datasets have undergone extensive QA/QC procedures; however, with large datasets it is not unexpected to find some problems. That was the case with this analysis as one of the tide stations had a recorded mean tide range of 0 feet. When instances such as this are discovered it would be possible to either drop the data altogether from the analysis, remove the aberrant points, or retain the data and accept the errors as contributors to the overall error term of the models. For this analysis, it was assumed that these instances were rare and the data were used as received from NOAA. This policy was applied to all other datasets used in this study as well.

Methods for generating watershed boundaries from elevation data are well developed within GIS technology; however, delineating watersheds for estuaries presents unique challenges. First, the seaward boundary of the estuary has to be determined, as does the direction of flow within the estuary to the pour point (i.e., the point water flows to in a watershed). There are no standard protocols for making these decisions, which can have a profound impact on the definition of the watershed and in turn may affect how data is compiled and linked to a given estuarine sampling station. Thus, I found that the automatic generation of estuarine watershed boundaries from the NED to be difficult, due in part to reasons already mentioned as well as to the lack of relief in many of the coastal regions encompassed in my study. The forthcoming National Watersheds Boundary Dataset, under development by several agencies, should hopefully alleviate this problem (see http://www.ncgc.nrcs.usda.gov/branch/gdb/products/watershed/ for details).

Effects Range Prediction Accuracy

The models I developed do a fairly good job at predicting the actual metal concentrations and the ER values (Table 3-8). A recent study that used much of the same data, built logistic regression models to predict the Benthic Index of estuarine sediments (Hale et al. 2004). They were able to correctly predict the Benthic Index of nearly 80% of the estuaries in their study. Although my study and Hale et al. (2004) are predicting different aspects of estuarine condition they both point to the possibility that models of these types can accurately identify potentially degraded estuaries.

Overall measures of accuracy and class-specific measures of accuracy in an error matrix, however, can be somewhat misleading because they only show the probability of correct classification. Often it is more interesting what types of misclassifications are occurring. This information is contained in the commission and omission errors. Omission errors indicate the probability that the model incorrectly classified an estuary that was actually ER Low as an ER Medium or ER High estuary. Commission errors are the inverse of this and indicate that the model incorrectly classifies an estuary that is actually a Medium or High as a Low. If these models are to be used in a screening framework (see *Screening Models*), then the probability of commission error is the more important value. This is because a high commission error would result in potentially degraded estuaries being missed and not examined when in fact they should. Determining what level of commission error is unacceptable must be established when using predictive models such as these.

The data used to test the accuracy of my models was the same data used to generate the models. In order to get a better picture of the overall accuracy of the predictions of the ER values, an independent data set with a complete range of low, medium, and high ER estuaries would be required. I considered splitting my data into a training dataset and a validation dataset; however, with the relatively small sample size I determined that would have had a deleterious effect on the parameter estimates. Furthermore, estuaries with an ERH classification were very rare and it would have been impossible to ensure an equitable apportion of points between the classification and validation sub-samples.

Comments on the Information Theoretic Approach

An important objective of this research was to advance the modeling described in Paul *et al.* (2002) by using information theoretic approaches and the multi-model inference methods described by Burnham and Anderson (2002). This method is fundamentally different from widely used hypothesis methods which area most appropriate when experimental-control designs and the associated assumptions are met; however, in environmental applications such as mine, these assumptions are almost never satisfied and hypothesis based assessment of estuarine condition is logically impossible.

A number of statistical methodologies exist that accommodate messy data that are used in environmental applications and these include, among others, computer intensive methods (i.e. Monte Carlo simulation), information theoretic approaches and Hierarchical Bayesian methods (Reckhow 1990, Burnham and Anderson 2002). Most of these approaches are relatively new to ecology and it is not clear under which conditions one method is superior over the other. Currently, Bayesian methods are difficult to implement without a great deal of statistical expertise and custom programming. The information theoretic approaches are based on similar principles as the Bayesian approaches, but are much easier to implement. They provide an estimate of a model's suitability relative to competing models. No arbitrary probability of acceptance or rejection is required and the importance of a given model or set of models is left up to the interpretation of the analyst.

Using these methods requires the assumption that the Akaike w_i are indeed a reliable measure of the likelihood of a model. Final judgment on that question is still a mater of debate, as such, environmental scientists are left with a decision to either use the traditional methods that have known shortcomings for many environmental applications or to implement more contemporary methods that seem to be a significant (pardon the pun) improvement on traditional p-value based hypothesis testing methods. Until superior methods are developed, it seems wise to choose the best available statistical methodology that can also be readily implemented. I believe that information theoretic approaches are such a method. They represent a step forward in the analysis of environmental data and advance the science of understanding how human impacts alter ecological condition.

Summary

There is a long history in ecological research of linking human activities on land to water quality. Much of the original work was conducted in freshwater systems, but more recent work has seen these ideas extended into estuaries and coastal waters. Many of these studies have uncovered quantitative relationships between various metrics of landscape structure, known sources of degradation, hydrologic parameters, and physical controls. The existence of these relationships points to the possibility of building predictive models. Using updated LULC data and an information theoretic approach to model selection and model averaging; I developed statistical models that could be used to predict the condition of small estuarine systems along the Atlantic Coast. My analyses were designed to address two questions. First, is there utility in using freely available, broad scale datasets to develop models of estuarine condition and second, how are accurate are these models?

Several studies including this one have determined that indeed broad scale data can be used effectively to model the condition of fresh and estuarine waters (Comeleo et al. 1996, Turner et al. 2001, Paul et al. 2002). Using information theoretic based model averaging I created models that correctly predict categorical measures of estuarine condition at least 80% of the time. Total area of urban land and the percent silt/clay in estuarine sediments were extremely important in describing the variation in estuarine sediment metal concentrations. This builds on prior studies by showing how these relationships vary with individual metals and also quantifying these relationships with higher resolution and more recent LULC data. It is now possible to use these results to protect and manage estuarine resources by identifying relatively un-impacted estuaries with the models and focusing resources on more highly impacted sites. Our ability to do this will only improve as the quality of data available for environmental modeling and ecological forecasting improves and the accuracy and reliability of the methods used to conduct the modeling also improves.

Table 3-1. Predictor Variables Examined during *a priori* modeling process. All variables are included here to highlight breadth of variables considered during the modeling process.

Predictor	Description	Inclusion/Exclusion Reason
Urban	Total hectares of urban in an estuarine drainage within a given distance	Included - Hypothesized source of metals
Agriculture	Total hectares of agricultural in an estuarine drainage within a given distance	Included - Hypothesized source of some metals, may act as filter on others
Wetland	Total hectares of wetland in an estuarine drainage within a given distance	Excluded - Shown to be inaccurate (USGS 2003, Hollister et al. 2004)
Forest	Total hectares of forest in an estuarine drainage within a given distance	Excluded - Negatively correlated with human land uses. Not needed in models with urban and agriculture (Weller et al. 2003)
Barren	Total hectares of barren land in an estuarine drainage within a given distance	Excluded - Shown to be inaccurate (USGS 2003, Hollister et al. 2004)
Scrub/shrub	Total hectares of scrub/shrub in an estuarine drainage within a given distance	Excluded - Shown to be inaccurate (USGS 2003, Hollister et al. 2004)
Total sinks	Total hectares of sink LULC types (i.e. wetland, forest, scrub, shrub) in an estuarine drainage within a given distance	Excluded - Redundancy with individual classes, accuracy issues, not always clear which LULC classes are sinks
Total sources	Total hectares of source LULC types (i.e. urban and agriculture) in an estuarine drainage within a given distance	Excluded - Redundancy with individual classes
Imp. surface	Total hectares of impervious surface within a given distance in an estuarine drainage within a given distance	Excluded - Often correlated with urban land, and need raw remotely sensed data. No broad scale source currently exists
Tidal volume	Tidal range multiplied by area of estuary	Excluded - Correlated with tidal range and est. area
Tidal range	Tidal range in feet as measured by the nearest NOAA tide gauges	Included - Estimate of Estuarine Hydrology
River flow	Annual average flow as measured by the closest USGS river gauge (cfs.)	Included - Estimate of Estuarine Hydrology
River/Tidal ratio	Ratio of River Flow per tidal cycle to tidal volume	Excluded - Used individual values instead
Est. area	Area of estuary as determined by EMAP	Included - Hypothesized surrogate of atmospheric deposition
Watershed slope	Slope within watershed	Excluded - Summary statistic not adequate descriptor. Full watersheds were computationally inefficient to derive.
Watershed area	Total watershed area	Excluded - Full watersheds were computationally inefficient to derive.
Silt/clay content	Percent of silt/clay in estuarine sediments	Included - Important for determining how metals bind to sediment
TOC	Total organic carbon in estuarine sediments	Excluded - Strong correlation with silt/clay content
Point sources	Estimate of total annual loading (in lbs.) of each metal from known point sources.	Included - Known source of metal inputs into estuaries

Table 3-2. Specification for the 45 Models examined in this study with the associated Akaike weights (*w*) for each of the dependent variables. The transformations of the dependent variable were used to account heteroscedasticity in the residuals. Bolded and italicized entries are models which had a *w* greater than 0.0000.

	w				
Model #	**ln(Cu+1)**	**ln(Pb+1)**	**Hg $^{0.2}$**	**Cd $^{0.3}$**	**Model**
1 (Global)	***0.0045***	***0.0043***	***0.0061***	***0.0052***	***urb.+pt. src.+est. area+ag.+riv. flow+tide+silt/clay***
2	***0.0136***	***0.0110***	***0.0193***	***0.0017***	***urb.+pt. src.+ag.+riv. flow+tidal range+silt/clay***
3	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.+est. area+ag.+riv. flow+tidal range
4	***0.0142***	***0.0130***	***0.0093***	***0.0139***	***urb.+est. area+ag.+riv. flow+tidal range+silt/clay***
5	0.0000	0.0000	0.0000	0.0000	pt. src.+est. area+ag.+riv. flow+tidal range+silt/clay
6	***0.0416***	***0.0321***	***0.0292***	***0.0039***	***urb.+ag.+riv. flow+tidal range+silt/clay***
7	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.+ag.+riv. flow+tidal range
8	***0.0425***	***0.0244***	***0.0580***	***0.0051***	***urb.+pt. src.+riv. flow+tidal range+silt/clay***
9	***0.0123***	***0.0242***	***0.0184***	***0.0521***	***urb.+pt. src.+est. area+ag.+silt/clay***
10	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.+est. area+riv. flow+tidal range
11	0.0000	0.0000	0.0000	0.0000	urb.+est. area+ag.+riv. flow+tidal range
12	0.0000	0.0000	0.0000	0.0000	pt. src.+ag.+riv. flow+tidal range+silt/clay
13	0.0000	0.0000	0.0000	0.0000	pt. src.+est. area+ag.+riv. flow+tidal range
14	0.0000	0.0000	0.0000	0.0000	urb.+ag.+riv. flow+tidal range
15	***0.0376***	***0.0520***	***0.0534***	***0.0164***	***urb.+pt. src.+ag.+silt/clay***
16	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.+riv. flow+tidal range
17	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.+est. area+ag.
18	***0.0379***	***0.0449***	***0.0575***	***0.1626***	***urb.+pt. src.+est. area+silt/clay***
19	***0.1267***	***0.0676***	***0.0893***	***0.0111***	***urb.+riv. flow+tidal range+silt/clay***
20	***0.0349***	***0.0727***	***0.0320***	***0.1345***	***urb.+est. area+ag.+silt/clay***
21	0.0000	0.0000	0.0000	0.0000	urb.+est. area+riv. flow+tidal range
22	0.0000	0.0000	0.0000	0.0000	pt. src.+ag.+riv. flow+tidal range
23	0.0000	0.0000	0.0000	0.0000	pt. src.+riv. flow+tidal range+silt/clay
24	0.0000	0.0000	0.0000	0.0000	pt. src.+est. area+ag.+silt/clay
25	0.0000	0.0000	0.0000	0.0000	pt. src.+est. area+riv. flow+tidal range
26	***0.1044***	***0.1533***	***0.0928***	***0.0361***	***urb.+ag.+silt/clay***
27	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.+ag.
28	***0.1137***	***0.0967***	***0.1630***	***0.0477***	***urb.+pt. src.+silt/clay***
29	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.+est. area
30	0.0000	0.0000	0.0000	0.0000	urb.+riv. flow+tidal range
31	0.0000	0.0000	0.0000	0.0000	urb.+est. area+ag.
32	***0.1060***	***0.1295***	***0.0969***	***0.4093***	***urb.+est. area+silt/clay***
33	0.0000	0.0000	0.0000	0.0000	pt. src.+ag.+silt/clay
34	0.0000	0.0000	0.0000	0.0000	pt. src.+riv. flow+tidal range
35	0.0000	0.0000	0.0000	0.0000	pt. src.+est. area+ag.
36	0.0000	0.0000	0.0000	0.0000	pt. src.+est. area+silt/clay
37	0.0000	0.0000	0.0000	0.0000	urb.+ag.
38	0.0000	0.0000	0.0000	0.0000	urb.+pt. src.
39	***0.3103***	***0.2744***	***0.2749***	***0.1004***	***urb.+silt/clay***
40	0.0000	0.0000	0.0000	0.0000	urb.+est. area
41	0.0000	0.0000	0.0000	0.0000	pt. src.+ag.
42	0.0000	0.0000	0.0000	0.0000	pt. src.+silt/clay
43	0.0000	0.0000	0.0000	0.0000	pt. src.+est. area
44	0.0000	0.0000	0.0000	0.0000	urb.
45	0.0000	0.0000	0.0000	0.0000	pt. src.

Table 3-3. Descriptive statistics of un-transformed dependent variables and independent variables.

Dep. Variables	Mean	Median	Min	Max	S.D.
Cu (ug/g)	39.75	20.55	0.00	263.00	58.10
Cd (ug/g)	0.54	0.24	0.00	6.58	0.97
Hg (ug/g)	0.22	0.08	0.00	3.27	0.45
Pb (ug/g)	46.13	29.50	0.00	323.00	55.34
Ind. Variables	**Mean**	**Median**	**Min**	**Max**	**S.D.**
Urban (ha)	4029.06	1652.49	14.22	20239.47	4796.50
Cu point sources (lbs/year)	3903.18	59.53	0.00	151886.20	15894.68
Cd point sources (lbs/year)	38.08	0.00	0.00	1288.14	161.79
Hg point sources (lbs/year)	56.93	1.22	0.00	1101.91	179.93
Pb point sources (lbs/year)	3545.29	27.91	0.00	99815.64	13268.69
Estuarine Area (km^2)	34.77	22.27	1.91	219.21	41.51
Agriculture (ha)	3128.60	1943.06	0.00	15304.14	3182.58
River Flow (cfs)	299.69	12.33	0.96	5839.31	1207.48
Tide Range (ft)	3.08	3.00	0.00	7.30	1.58
Silt/Clay (%)	51.86	62.40	0.80	99.60	35.78

Table 3-4. Global model residual analysis statistics for the untransformed and transformed dependent variables.

	Dependent Variables							
	Raw				Transformed			
	Cu	**Pb**	**Hg**	**Cd**	**ln(Cu + 1)**	**ln(Pb + 1)**	**$Hg^{0.2}$**	**$Cd^{0.3}$**
Levene's								
F	13.963	11.664	10.768	8.773	0.120	1.066	0.346	0.448
p-value	0.000	0.001	0.001	0.004	0.730	0.304	0.557	0.505
Shapiro-Wilk								
W	0.957	0.738	0.597	0.674	0.988	0.934	0.959	0.980
p-value	0.001	0.000	0.000	0.000	0.412	0.000	0.002	0.095

Table 3-5a. Basic Model Results and Akaike Information Criteria (AIC) statistics for ln(Cu+1). Only the models with an AIC *w* greater than 0 are included. In the calculation of AIC, K is used to represent the number of parameters and includes the parameters, intercept and an error term.

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC *w*
39	0.793	4	-98.82	206.02	0.00	0.3103
19	0.794	6	-97.50	207.81	1.79	0.1267
28	0.792	5	-98.73	208.02	2.01	0.1137
32	0.791	5	-98.80	208.16	2.15	0.1060
26	0.791	5	-98.81	208.19	2.18	0.1044
8	0.792	7	-97.46	209.99	3.98	0.0425
6	0.792	7	-97.48	210.04	4.02	0.0416
18	0.790	6	-98.71	210.22	4.20	0.0379
15	0.790	6	-98.72	210.24	4.22	0.0376
20	0.789	6	-98.79	210.39	4.37	0.0349
4	0.791	8	-97.39	212.19	6.17	0.0142
2	0.790	8	-97.44	212.27	6.26	0.0136
9	0.788	7	-98.70	212.48	6.46	0.0123
1 (global)	0.789	9	-97.36	214.48	8.47	0.0045

Table 3-5b. Basic Model Results and Akaike Information Criteria (AIC) statistics for ln(Pb+1). Only the models with an AIC *w* greater than 0 are included. In the calculation of AIC, K is used to represent the number of parameters and includes the parameters, intercept and an error term.

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC *w*
39	0.719	4	-86.36	181.10	0.00	0.2744
26	0.719	5	-85.85	182.27	1.16	0.1533
32	0.718	5	-86.02	182.60	1.50	0.1295
28	0.717	5	-86.31	183.19	2.09	0.0967
20	0.718	6	-85.48	183.76	2.66	0.0727
19	0.718	6	-85.55	183.90	2.80	0.0676
15	0.717	6	-85.81	184.43	3.33	0.0520
18	0.716	6	-85.96	184.72	3.62	0.0449
6	0.717	7	-85.16	185.40	4.29	0.0321
8	0.716	7	-85.43	185.94	4.84	0.0244
9	0.716	7	-85.44	185.96	4.85	0.0242
4	0.716	8	-84.90	187.21	6.11	0.0130
2	0.715	8	-85.07	187.54	6.43	0.0110
1 (global)	0.714	9	-84.81	189.39	8.29	0.0043

Table 3-5c. Basic Model Results and Akaike Information Criteria (AIC) statistics for ($Hg^{0.2}$). Only the models with an AIC *w* greater than 0 are included. In the calculation of AIC, K is used to represent the number of parameters and includes the parameters, intercept and an error term.

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC *w*
39	0.582	4	49.08	-89.78	0.00	0.2749
28	0.583	5	49.65	-88.73	1.05	0.1630
32	0.579	5	49.13	-87.69	2.09	0.0969
26	0.579	5	49.08	-87.60	2.17	0.0928
19	0.583	6	50.16	-87.53	2.25	0.0893
8	0.584	7	50.87	-86.66	3.11	0.0580
18	0.579	6	49.72	-86.65	3.13	0.0575
15	0.579	6	49.65	-86.50	3.28	0.0534
20	0.575	6	49.14	-85.48	4.30	0.0320
6	0.579	7	50.18	-85.29	4.48	0.0292
2	0.581	8	50.93	-84.46	5.32	0.0193
9	0.575	7	49.72	-84.37	5.41	0.0184
4	0.575	8	50.20	-83.01	6.77	0.0093
1 (global)	0.577	9	50.95	-82.15	7.63	0.0061

Table 3-5d. Basic Model Results and Akaike Information Criteria (AIC) statistics for ($Cd^{0.3}$). Only the models with an AIC *w* greater than 0 are included. In the calculation of AIC, K is used to represent the number of parameters and includes the parameters, intercept and an error term.

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC *w*
32	0.530	5	4.02	2.53	0.00	0.4093
18	0.527	6	4.21	4.37	1.85	0.1626
20	0.525	6	4.02	4.75	2.23	0.1345
39	0.513	4	1.52	5.34	2.81	0.1004
9	0.522	7	4.21	6.65	4.12	0.0521
28	0.511	5	1.87	6.82	4.30	0.0477
26	0.509	5	1.59	7.38	4.86	0.0361
15	0.507	6	1.92	8.96	6.43	0.0164
4	0.517	8	4.06	9.29	6.76	0.0139
19	0.504	6	1.53	9.73	7.21	0.0111
1 (global)	0.514	9	4.25	11.26	8.73	0.0052
8	0.502	7	1.89	11.31	8.78	0.0051
6	0.500	7	1.61	11.86	9.33	0.0039
2	0.498	8	1.93	13.53	11.00	0.0017

Table 3-6. Relative importance of each model parameter as measured by the AIC weight for all models with AIC weight greater than 0. Importance calculated by summing the AIC weight for each model that contained the variable (Burnham and Anderson 2002).

Dependent	Urban	Point Source	Est. Area	Agriculture	Riv. Fl. & Td. Rg.	SiltClay
ln(Cu+1)	1	0.262	0.210	0.221	0.243	1
ln(Pb+1)	1	0.258	0.289	0.331	0.152	1
$Hg^{0.2}$	1	0.376	0.220	0.231	0.211	1
$Cd^{0.3}$	1	0.291	0.778	0.260	0.041	1

Table 3-7. Model averaged parameter estimates and parameter standard errors averaged from all models with w_i greater than 0. These parameters constitute the final averaged model and may be directly used to predict concentrations of the metals.

	ln(Cu+1)		ln(Pb+1)		$Hg^{0.2}$		$Cd^{0.3}$	
	Theta	**S.E.**	**Theta**	**S.E.**	**Theta**	**S.E.**	**Theta**	**S.E.**
Int	1.049E+00	1.469E-01	1.972E+00	1.293E-01	3.005E-01	3.758E-02	3.140E-01	5.590E-02
Urb.	1.669E-04	1.329E-05	2.104E-04	2.060E-05	2.878E-05	4.070E-06	4.947E-04	7.070E-05
Pt. Src.	-4.116E-07	1.404E-06	3.890E-07	1.517E-06	4.662E-05	8.121E-05	-2.880E-05	6.884E-05
Area	-6.297E-05	3.431E-04	2.905E-04	6.173E-04	2.680E-05	1.044E-04	-9.526E-04	6.964E-04
Ag	1.437E-07	5.689E-06	1.159E-05	2.138E-05	1.804E-08	1.491E-06	4.912E-06	4.221E-05
Flow	1.036E-05	2.191E-05	-1.832E-07	6.753E-06	2.928E-06	5.873E-06	1.132E-07	9.115E-07
Tide	-1.390E-02	2.484E-02	-6.482E-03	1.325E-02	-2.605E-03	5.110E-03	-6.863E-05	7.388E-04
SiltClay	2.397E-02	1.614E-03	1.701E-02	1.451E-03	3.470E-03	4.287E-04	4.854E-03	6.425E-04

Model average adjusted r^2: Cu 0.784, Pb 0.705, Hg, 0.564, Cd 0.503
Model average S. E.: Cu 0.604, Pb 0.540, Hg, 0.161, Cd 0.242

Table 3-8. Error matrices comparing predictions of Effects Ranges to Effects Ranges from the raw data. The Effects Ranges were converted directly from the back-transformed predictions of the model averaged sediment metal concentrations.

Cu

Model Averaged Effects Range	Raw Data Effects Range			
	Low	**Medium**	**High**	**Row Total**
Low	76	9	0	85
Medium	7	18	0	25
High	0	2	0	2
Column Total	83	29	0	112

Overall Accuracy = 84%, **Percent Correct** Low = 89% Medium = 72% High = NA, **Commission Error** Low = 11% Medium = 38% High = NA, **Ommission Error** Low = 8% Medium = 38% High = NA

Pb

Model Averaged Effects Range	Raw Data Effects Range			
	Low	**Medium**	**High**	**Row Total**
Low	78	8	0	86
Medium	5	17	2	24
High	0	2	0	2
Column Total	83	27	2	112

Overall Accuracy = 85%, **Percent Correct** Low = 91% Medium = 71% High = NA, **Commission Error** Low = 9% Medium = 29% High = NA, **Ommission Error** Low = 6% Medium = 37% High = NA

Hg

Model Averaged Effects Range	Raw Data Effects Range			
	Low	**Medium**	**High**	**Row Total**
Low	73	14	0	87
Medium	4	11	4	19
High	0	2	4	6
Column Total	77	27	8	112

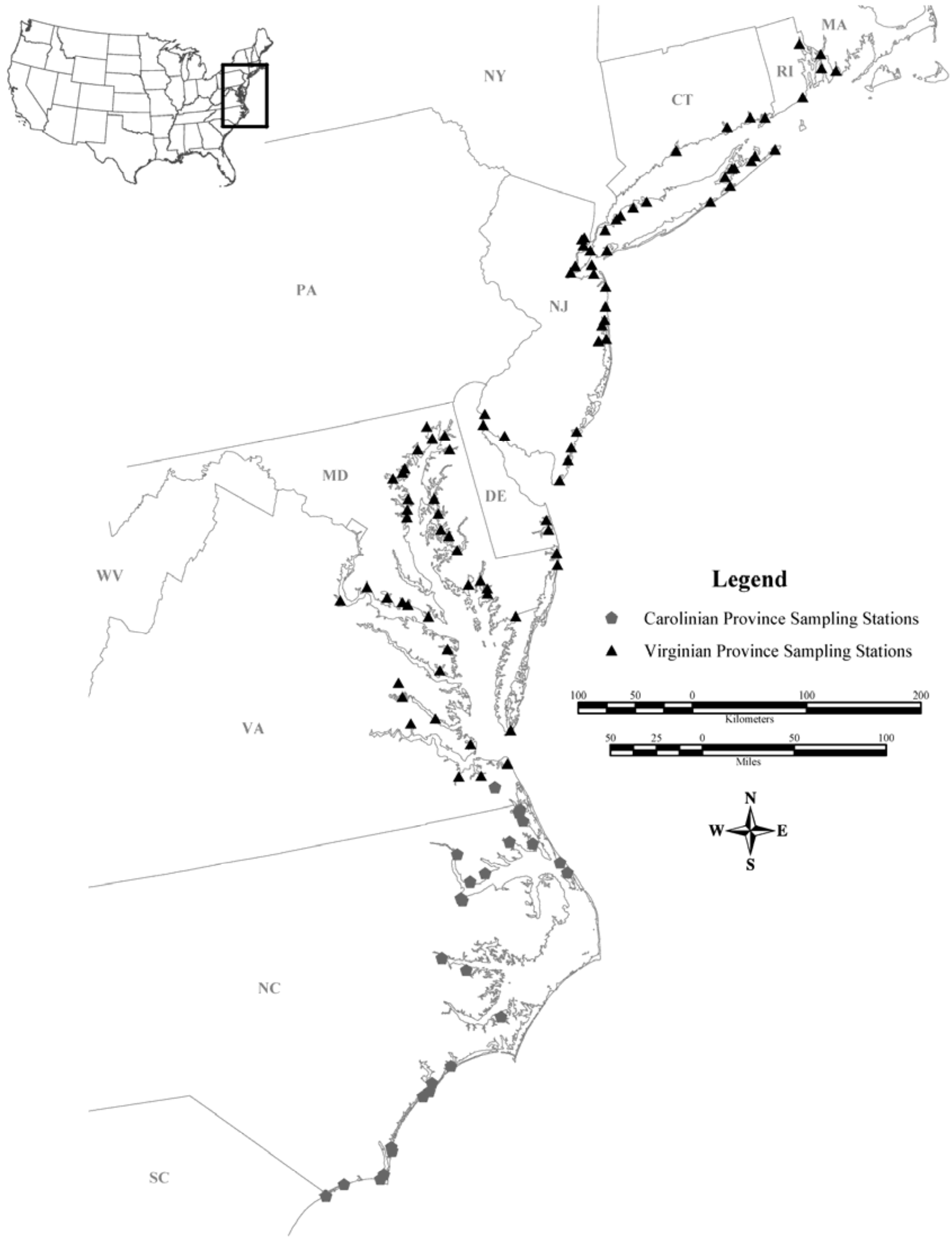
Overall Accuracy = 79%, **Percent Correct** Low = 84% Medium = 58% High = 67%, **Commission Error** Low = 16% Medium = 42% High = 33%, **Ommission Error** Low = 5% Medium = 59% High = 50%

Cd

Model Averaged Effects Range	Raw Data Effects Range			
	Low	**Medium**	**High**	**Row Total**
Low	98	7	0	105
Medium	2	5	0	7
High	0	0	0	0
Column Total	100	12	0	112

Overall Accuracy = 92%, **Percent Correct** Low = 93% Medium = 71% High = NA, **Commission Error** Low = 7% Medium = 29% High = NA, **Ommission Error** Low = 2.00% Medium = 58% High = NA

Figure 3-1. Map of study area with locations of EMAP-E sampling stations from the Virginian and Carolinian provinces.



Appendices

Appendix 1. Data Files and Metadata for Chapters 2 and 3

Data Files

Data files and this metadata document are available from

http://www.edc.uri.edu/personal/jeff/dissertation/data

There is a single zip file (jwh_diss_data.zip) which contains the following files

chp2_data/

- data2.5txt - data30.txt - These files contain the raw data used in Chapter 2. The data dictionary in this document as well as the methods contains further details. For each file, the number in the name indicates the multi-scaled watershed used to calculate the area of developed and agricultural land as well as the summed point source loadings.

chp3_data/

jwh_diss_data.txt

This file contains the EMAP station data, sediment metal concentrations, land use/land cover proportions, point sources total, river flow, and tide gauge data are available as a single flat, comma delimited ASCII text file. This is the raw data file for all analyses described in Chapter 3 of Predicting Estuarine Condition for Small Estuarine Systems Along the United States' Atlantic Coast.

ms_watersheds/

This folder contains the multi-scaled watersheds used to subset the land use/land cover data to generate the landscape metrics are available in a zip file with each coverage in ESRI Arc/Info export format. The watersheds are named with the associated EMAP-E sampling station name.

draft_app1.pdf - This appendix.

README.txt - A short file briefly describing the contents of the zip file.

Metadata

Introduction

The analysis of ecological processes on the landscape scale is a relatively recent capability. The initial mode of inquiry for landscape scale studies of ecological processes was often focused on a single study, at a single site or small suite of sites. Thinking about monitoring the environment at spatial and temporal scales appropriate to the questions was often a secondary consideration. Over the last several decades this had begun to change and now several large, broad scale datasets exist. Many of these datasets are national in scope and provide a broad range of information that may be used in a variety of ecological and environmental modeling applications.

The Environmental Monitoring and Assessment Program - Estuaries (EMAP-E) dataset, the National Land Cover Dataset (NLCD), the National Coastal Pollution Discharge Inventory (NCPDI), the USGS River Gauge data, the NOAA Tide Gauge data, and the National Elevation Dataset (NED) are three datasets I compiled for use in modeling metal concentrations in estuarine sediment. Although these datasets cover the same general geographic extent and are publicly available via the web, in their native formats they are not readily modeled. Considerable processing and compilation steps are required to form the EMAP Virginian and Carolinian Province Station Data and the Multi-scale watersheds used to subset the NLCD. Although these datasets do not represent new data *per se*, the manipulation and compilation of the data into a form ready for landscape analyses is novel and provides added benefit to the use of these data.

The EMAP-E data were collected using a probabilistic sampling scheme which allows for estimation of spatial distribution, extent and condition of estuarine resources (Paul et al. 1999). Estuaries within the study area were divided into three classes, Large Estuarine Systems, Small Estuarine Systems, and Tidal Rivers, and each class was sampled separately. A subset of small estuarine systems, the focus of this research, was randomly chosen and sampled by EMAP. For further details of the sampling see Overton et al. (1991) and Paul et al. (1992). All EMAP-E data in the Virginian and Carolinian biogeographic provinces stations were downloaded from the National Coastal Assessment web page (http://www.epa.gov/emap/nca).

Both the NLCD and the NED is available via the seamless data distribution interface of the National Map (http://nationalmap.gov);however, the large geographic area of my study exceeded this. Each dataset was ordered and delivered via CD. The final datasets used in these modeling exercises are not as readily available. The NCPDI provides rough estimates of point source pollution loads into coastal waters and date back to the early 1990's. Although there are known issues with these estimates, the NCPDI remains the best continuous, national dataset on point source loadings. The river and tide gauge data were used to estimate estuarine hydrology. These data are currently available via the NOAA Coastal Assessment and Data Synthesis website; however, this website is no longer being supported and its continued availability is questionable (Percy Pacheco, Personal comm.). Regardless, the river and tide gauge, as of this writing, were available in a geographic format directly from the CA&DS website.

Methods

<u>EMAP Station Locations, Sediment Analyte Concentrations, and Virginian Benthic Index:</u> Description of the EMAP-E data collection procedures, quality assurance, and quality control is provided in detail by several publications and is not repeated here. Should additional information be required on how the EMAP-E data was collected, see Paul et al. (1992) for more detail. In addition to the original data collection procedures described elsewhere, I applied additional queries and compiled the EMAP-E data to develop a base set of EMAP stations with sediment metal concentrations and benthic index scores.

The EMAP-E datasets are all currently maintained by the Environmental Protection Agency and are available at http://www.epa.gov/emap/nca/html/data/index.html. Using the web based query tools, I queried the dataset and selected EMAP datasets from the Carolinian and Virginian Provinces. I downloaded metadata and associated datasets for Station Locations Information, Sediment Analyte Concentration, and Virginian Benthic Index. I stored each of these locally as individual text files. Second, I imported each text file into an MS Access database. This provided an efficient means to re-compile, store, and manipulate the data.

The Station Location information was converted directly into an Access table with the "Station Name" as the primary key. The Benthic Index and Sediment Analyte datasets were combined together to form another table. Some of the sampling stations had multiple visits thus, the "Station Name" and "Sampling Collection Date" fields were merged to create a unique Primary Key. Station Name was used to build a one-to-many relate between the Station Location Information table and the combined benthic indices and sediment analyte tables. This particular database structure, while simple, was adequate for all subsequent queries.

I used four primary queries to subset the data. First, since I focused my modeling efforts on only small estuarine systems, the data collected in "Large Estuaries" and "Large Tidal Rivers" were discarded. Second, EPA uses careful quality assurance and quality control procedures and when problems with individual samples are identified, they are labeled as such in the database. A variety of QA codes indicate problems and suggest caution when using those data, thus problematic samples were removed from the analysis. The following QA codes were selected and the associated samples removed: CH-CC, CH-G, CH-O, CH-Q, and CH-T.

EMAP-E tested for a total of 15 metals; however, only nine of these metals are known to have adverse affects on benthic biota. Data on point sources exist for seven (As, Cd, Cr, Cu, Pb, Hg, and Zn) of these nine metals and since many modeling efforts will likely include point source information, I used only stations where these seven metals were analyzed (Pacheco 1993, Long et al. 1995, Paul et al. 2002). Additionally, not all stations examined each of these seven metals. Thus, the third query accounted for this and I selected only points that contained analyte concentrations for all seven metals.

Finally, a number of estuaries contain more than one sampling location. Stations were classified as a base sampling station, dissolved oxygen station, index station, supplemental station, or an indicator station. Only those sites classified as base stations were part of the probabilistic sampling scheme. Thus, all non-base stations were removed from the dataset. Additionally, a few small estuaries contained replicate stations, which were used to estimate variance; these replicate stations were also removed. After these filtering steps, a small number of estuaries still contained more than a single sampling location. In those instances (only 9 stations) it was not possible to identify differences, other than sampling date, between the samples. Because I was unable to logically choose which sampling date to use, these 9 stations were removed from the dataset.

Stations were converted to a vector coverage that contained the station location information for all the EMAP sampling stations in the Virginian and Carolinian Provinces, the concentrations of As, Cd, Cr, Cu, Hg, Pb, Zn, and the EMAP Virginian Province Benthic Index. The coverage was converted to an Albers Equal-Area projection to match the NLCD data projection. Since stations from both the Carolinian and Virginian provinces were downloaded, several of the southern stations fell outside of my study area. Stations south of the North Carolina border were selected and removed from the coverage. The final coverage contained 209 stations. On of these stations, VA92-521, had a recorded value of 13,600 ug/g. The next highest amount recorded for this dataset was approximately 300 ug/g. It was assumed that the value recorded for VA92-521 was in error and this stations was removed. These 208 stations formed the basis for all subsequent data manipulation, generation and analysis.

<u>Multi-Scaled Watersheds:</u> Many landscape modeling efforts are based on the assumption that land use and land cover in contributing watersheds leads to degraded water quality and ecological condition in receiving waters. Generally, this assumption holds true; however, using a "watershed" as a unit of comparison introduces a number of problems. First, as of this writing (Oct 2004), no consistent, national spatial datasets at a scale appropriate for landscape analysis yet exist that identify watershed boundaries. Development of the Watersheds Boundary Dataset (WBD) should rectify this. Until that is available, researchers must rely on

incomplete watershed datasets, use broad scale watershed data or, delineate them on a case-by-case basis and sacrificing comparability across studies. Second, studies done at the "watershed scale" are not necessarily comparable because watersheds may be of vastly different sizes. These different sized watersheds result in a sampling unit with varying extent and land area at varying distances from a sampling point. Finally, generating watersheds for estuaries presents a special set of circumstances. Watersheds for upland water bodies are relatively straightforward since the area represented by the waterbody may be discounted and all area draining to that point can be readily identified. However, in an estuarine setting there are no obvious boundaries since estuaries drain to the ocean or obvious direction of flow because of the sometimes large area of the estuary and tidal interactions.

Because of these varied issues, I chose to develop a consistent method to generate hydrologically based sampling units for estuaries from widely available elevation data. A consistent repeatable method allows for comparability across studies and allows for defining hydrologically based sampling units with consistent extent, consistent distances from a sampling station, and logically defined estuarine boundaries. To accomplish this, I acquired the NED on DVD directly from the USGS EROS data center. These data may also be obtained online at http://seamless.usgs.gov. The data acquisition area ranged from North Carolina to Maine and from Cape Cod to West Virginia. As delivered, the NED was stored in a number of separate Arc/Info GRIDS. In order to facilitate future analysis, I "tiled" the NED GRIDS by USGS 4-digit Hydrologic Unit Codes or sub-regions (see http://water.usgs.gov/GIS/huc.html). To prevent loss of data along the boundaries of sub-regions, each sub-region was buffered by 1 km, resulting in a 2 km overlap between any two adjacent sub-regions.

Using the buffered NED sub-regions as the base elevation data, I generated hydrologically based sampling units for estuaries of two types. First, for each of the 209 EMAP sampling stations described above, I generated a sampling unit that included all land area draining into the estuary that was a set Euclidean distance from the sampling station. I refer to these as "multi-scaled watersheds". The distances used to generate the multi-scaled watersheds in this study were 2.5, 5, 7.5 10, 12.5, 15, 17.5, 20, 25, and 30 km.

I used Arc/Info and AML to generate the hydrologically based sampling units. The general steps are as follows:

- Extract EMAP sampling station
- Buffer sampling station (2.5, 5, 7.5 10, 12.5, 15, 17.5, 20, 25, and 30 km)
- Extract NED for buffered area
- Fill NED
- Generate flow direction
    - Flow direction on land is straightforward and driven by elevation
    - Flow direction in estuary is based on shortest distances from a pixel to the sampling locations. Land area is set as NODATA.
    - I defined estuaries as all pixels with an elevation less than 3 meters. I chose 3 meters as the cut-off by examining all the elevation of all 209 EMAP sampling stations. Since it is known that all sampling stations occur in estuaries, I made the assumptions that all sampling stations should have an elevation of 0 and forced into the NED by assigning all areas less than 3 meters to an elevation of 0.
    - I combined Flow direction on land and in the estuary.

- These processing steps provided the basis for generating multi-scaled watersheds. I used the WATERSHED command in Arc/INFO GRID with the combined flow direction grids described above. To generate the final multi-scaled watershed.

The processing steps above work well for areas with significant relief; however, generating hydrologically defined sampling units was not possible in areas of low relief. In cases where the relief was very low it is possible to either use a simple Euclidean buffer as a sampling unit or to not include those stations in any analysis. As suggested, the landscape analyses these data support are based on the assumption that land use and land cover in contributing watersheds leads to degraded water quality and impacted ecological condition and is therefore dependent upon hydrologic processes. Estuaries with no obvious watersheds and very flat hydrology likely violate this assumption. Thus, the condition of estuaries surrounded by exceedingly flat terrain is less a function of upstream land use/land cover and more a function of other processes. I chose to remove sampling stations that are surrounded by flat terrain. An EMAP sampling station was classified as having flat terrain if the land area with 5 km of the sampling station had a maximum elevation of 3 meters or less. Removing these data from a dataset with the sole purpose of exploring landscape linkages to receiving estuarine waters is a legitimate step. Removing these extraordinarily flat watersheds resulted in a total of 112 sampling stations.

<u>Land Use/Land Cover Area and Proportions:</u> The NLCD, which may be acquired online at http://seamless.usgs.gov, was initially classified with a Modified USGS Level II classification; however, at this level of classification, the accuracy has been shown to be questionable (Yang et al. 2001, United States Geological Survey 2003). Combining these classes into a more generalized USGS Level I classification improves the classification accuracy considerably and these more general classes are still appropriate for use in broad scale landscape analysis (Yang

et al. 2001, United States Geological Survey 2003, Hollister et al. 2004). I clipped out the NLCD for each of the 10 multi-scaled watershed distances for each EMAP sampling station. I summed the total area and calculated proportions for each of the USGS Level I classes.

Tidal Range and River Flow: Data for tidal range and estimates of freshwater inflow into each estuary were taken from the NOAA Tide Gauge and USGS River Gauge data that was compiled by NOAA as part of the Coastal Assessment and Data Synthesis (CA&DS) program. The goals of this program were to compile a wide variety of data for used by the coastal management community. Unfortunately, this program has been discontinued, but the results of their efforts are still available on-line, albeit unsupported. Further details on the data and some metadata are also available via the CA&DS website at http://cads.nos.noaa.gov/.

I downloaded both the tide and river data from CA&DS and converted the ESRI shapefiles in Arc/INFO coverages. The river data was matched to the EMAP sampling station through a two-step process. First all gauges that fell within the largest of the multi-scaled watersheds were selected. Second from the set of selected gauges, the closest gauge was selected by using the Arc/INFO NEAR command and the identifying the gauge with the minimum distance. Similarly, the closest tide gauge was also chosen using the NEAR command. For the closest selected gauges, the mean of all daily values and the mean range were selected for river flow and tidal range, respectively. A text file with the 112 EMAP-E station names and the river flow and tidal range data was created and merged with the other datasets using the EMAP-E station name as the related item. The resulting merged file was imported into R version 1.9.1 and maintained as an R data object. It was from this data object that the final EMAP Virginian and Carolinian Province Station Data text file was created.

Data Dictionaries

The list below provides the definitions of each of the fields in the jwh_diss_data_chp2.zip and jwh_diss_data_chp3.txt files.

Station.Name - Name of the estuarine sampling station. These are the same as the stations names in the EMAP-E and MAIA datasets.

Long - Longitude of the estuarine sampling station in decimal degrees.

Lat -Latitude of the estuarine sampling station in decimal degrees.

x-coord - The x-coordinated of the estuarine sampling station in Albers NAD83, Meters with a first standard parallel of 29 30 00, second standard parallel of 45 30 00, central meridian of -96 0 0, and a latitude of origin of 23 0 0.

y-coord -The x-coordinated of the estuarine sampling station in Albers NAD83, Meters with a first standard parallel of 29 30 00, second standard parallel of 45 30 00, central meridian of -96 0 0, and a latitude of origin of 23 0 0.

Zn - Concentration of Zinc in ug/g in estuarine sediment. These measurements are from either the EMAP-E or the MAIA datasets.

Hg - Concentration of Mercury in ug/g in estuarine sediment. These measurements are from either the EMAP-E or the MAIA datasets.

Pb - Concentration of Lead in ug/g in estuarine sediment. These measurements are from either the EMAP-E or the MAIA datasets.

As - Concentration of Arsenic in ug/g in estuarine sediment. These measurements are from either the EMAP-E or the MAIA datasets.

Cd - Concentration of Cadmium in ug/g in estuarine sediment. These measurements are from either the EMAP-E or the MAIA datasets.

Cr - Concentration of Chromium in ug/g in estuarine sediment. These measurements are from either the EMAP-E or the MAIA datasets.

Cu- Concentration of Copper in ug/g in estuarine sediment. These measurements are from either the EMAP-E or the MAIA datasets.

dist2_5ktotal - Total area of the 2.5 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist2_5k100 - Total area of developed land from the NLCD in the 2.5 km multi-scaled watershed.

dist2_5k200 - Total area of agricultural land from the NLCD in the 2.5 km multi-scaled watershed.

dist5ktotal - Total area of the 5 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist5k100 - Total area of developed land from the NLCD in the 5 km multi-scaled watershed.

dist5k200 - Total area of agricultural land from the NLCD in the 5 km multi-scaled watershed.

dist7_5ktotal - Total area of the 7.5 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist7_5k100 - Total area of developed land from the NLCD in the 7.5 km multi-scaled watershed.

dist7_5k200 - Total area of agricultural land from the NLCD in the 7.5 km multi-scaled watershed.

dist10ktotal - Total area of the 10 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist10k100 - Total area of developed land from the NLCD in the 10 km multi-scaled watershed.

dist10k200 - Total area of agricultural land from the NLCD in the 10 km multi-scaled watershed.

dist12_5ktotal - Total area of the 12.5 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist12_5k100 - Total area of developed land from the NLCD in the 12.5 km multi-scaled watershed.

dist12_5k200 - Total area of agricultural land from the NLCD in the 12.5 km multi-scaled watershed.

dist15ktotal - Total area of the 15 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist15k100 - Total area of developed land from the NLCD in the 15 km multi-scaled watershed.

dist15k200 - Total area of agricultural land from the NLCD in the 15 km multi-scaled watershed.

dist17_5ktotal - Total area of the 17.5 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist17_5k100 - Total area of developed land from the NLCD in the 17.5 km multi-scaled watershed.

dist17_5k200 - Total area of agricultural land from the NLCD in the 17.5 km multi-scaled watershed.

dist20ktotal - Total area of the 20 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist20k100 - Total area of developed land from the NLCD in the 20 km multi-scaled watershed.

dist20k200 - Total area of agricultural land from the NLCD in the 20 km multi-scaled watershed.

dist25ktotal - Total area of the 25 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist25k100 - Total area of developed land from the NLCD in the 25 km multi-scaled watershed.

dist25k200 - Total area of agricultural land from the NLCD in the 25 km multi-scaled watershed.

dist30ktotal - Total area of the 30 km multi-scaled watershed. This area corresponds to the multi-scaled watershed with the same Station Name. These are located in the multiscale_ws.zip file**.**

dist30k100 - Total area of developed land from the NLCD in the 30 km multi-scaled watershed.

dist30k200 - Total area of agricultural land from the NLCD in the 30 km multi-scaled watershed.

siltclay - Percent of silt clay in estuarine sediments. This is from either the EMAP-E of the MAIA datasets.

TOC - Percent Total Organic Carbon of estuarine sediments. This is from either the EMAP-E of the MAIA datasets.

pointsrc2.5.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 2.5 km from the sampling station were summed to create these estimates.

pointsrc2.5.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 2.5 km from the sampling station were summed to create these estimates.

pointsrc2.5.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 2.5 km from the sampling station were summed to create these estimates.

pointsrc2.5.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 2.5 km from the sampling station were summed to create these estimates.

pointsrc2.5.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 2.5 km from the sampling station were summed to create these estimates.

pointsrc2.5.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 2.5 km from the sampling station were summed to create these estimates.

pointsrc2.5.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 2.5 km from the sampling station were summed to create these estimates.

pointsrc5.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 5 km from the sampling station were summed to create these estimates.

pointsrc5.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 5 km from the sampling station were summed to create these estimates.

pointsrc5.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 5 km from the sampling station were summed to create these estimates.

pointsrc5.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 5 km from the sampling station were summed to create these estimates.

pointsrc5.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 5 km from the sampling station were summed to create these estimates.

pointsrc5.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 5 km from the sampling station were summed to create these estimates.

pointsrc5.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 5 km from the sampling station were summed to create these estimates.

pointsrc7.5.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 7.5 km from the sampling station were summed to create these estimates.

pointsrc7.5.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 7.5 km from the sampling station were summed to create these estimates.

pointsrc7.5.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 7.5 km from the sampling station were summed to create these estimates.

pointsrc7.5.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 7.5 km from the sampling station were summed to create these estimates.

pointsrc7.5.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 7.5 km from the sampling station were summed to create these estimates.

pointsrc7.5.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 7.5 km from the sampling station were summed to create these estimates.

pointsrc7.5.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 7.5 km from the sampling station were summed to create these estimates.

pointsrc10.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 10 km from the sampling station were summed to create these estimates.

pointsrc10.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 10 km from the sampling station were summed to create these estimates.

pointsrc10.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 10 km from the sampling station were summed to create these estimates.

pointsrc10.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 210 km from the sampling station were summed to create these estimates.

pointsrc10.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 10 km from the sampling station were summed to create these estimates.

pointsrc10.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 10 km from the sampling station were summed to create these estimates.

pointsrc10.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 10 km from the sampling station were summed to create these estimates.

pointsrc12.5.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 12.5 km from the sampling station were summed to create these estimates.

pointsrc12.5.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 12.5 km from the sampling station were summed to create these estimates.

pointsrc12.5.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 12.5 km from the sampling station were summed to create these estimates.

pointsrc12.5.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 12.5 km from the sampling station were summed to create these estimates.

pointsrc12.5.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 12.5 km from the sampling station were summed to create these estimates.

pointsrc12.5.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 12.5 km from the sampling station were summed to create these estimates.

pointsrc12.5.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 12.5 km from the sampling station were summed to create these estimates.

pointsrc15.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 15 km from the sampling station were summed to create these estimates.

pointsrc15.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 15 km from the sampling station were summed to create these estimates.

pointsrc15.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 15 km from the sampling station were summed to create these estimates.

pointsrc15.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 15 km from the sampling station were summed to create these estimates.

pointsrc15.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 15 km from the sampling station were summed to create these estimates.

pointsrc15.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 15 km from the sampling station were summed to create these estimates.

pointsrc15.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 15 km from the sampling station were summed to create these estimates.

pointsrc17.5.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 17.5 km from the sampling station were summed to create these estimates.

pointsrc17.5.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 17.5 km from the sampling station were summed to create these estimates.

pointsrc17.5.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 17.5 km from the sampling station were summed to create these estimates.

pointsrc17.5.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 17.5 km from the sampling station were summed to create these estimates.

pointsrc17.5.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 17.5 km from the sampling station were summed to create these estimates.

pointsrc17.5.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 17.5 km from the sampling station were summed to create these estimates.

pointsrc17.5.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 17.5 km from the sampling station were summed to create these estimates.

pointsrc20.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 20 km from the sampling station were summed to create these estimates.

pointsrc20.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 20 km from the sampling station were summed to create these estimates.

pointsrc20.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 20 km from the sampling station were summed to create these estimates.

pointsrc20.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 20 km from the sampling station were summed to create these estimates.

pointsrc20.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 20 km from the sampling station were summed to create these estimates.

pointsrc20.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 20 km from the sampling station were summed to create these estimates.

pointsrc20.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 20 km from the sampling station were summed to create these estimates.

pointsrc25.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 25 km from the sampling station were summed to create these estimates.

pointsrc25.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 25 km from the sampling station were summed to create these estimates.

pointsrc25.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 25 km from the sampling station were summed to create these estimates.

pointsrc25.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 25 km from the sampling station were summed to create these estimates.

pointsrc25.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 25 km from the sampling station were summed to create these estimates.

pointsrc25.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 25 km from the sampling station were summed to create these estimates.

pointsrc25.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 25 km from the sampling station were summed to create these estimates.

pointsrc30.asann - Estimate of the annual loading of arsenic. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 30 km from the sampling station were summed to create these estimates.

pointsrc30.cdann - Estimate of the annual loading of cadmium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 30 km from the sampling station were summed to create these estimates.

pointsrc30.crann - Estimate of the annual loading of chromium. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 30 km from the sampling station were summed to create these estimates.

pointsrc30.cuann - Estimate of the annual loading of copper. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 30 km from the sampling station were summed to create these estimates.

pointsrc30.hgann - Estimate of the annual loading of mercury. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 30 km from the sampling station were summed to create these estimates.

pointsrc30.pbann - Estimate of the annual loading of lead. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 30 km from the sampling station were summed to create these estimates.

pointsrc30.znann - Estimate of the annual loading of zinc. All known point sources in NCPDI data that fell in the estuarine watershed and were less than 30 km from the sampling station were summed to create these estimates.

station.area - Estimated area of the estuary. These estimates are from the EMAP-E or MAIA datasets.

RiverFlow - Estimate of the annual freshwater input into the estuary. These values are taken from the closest USGS river gauge to the EMAP-E or MAIA sampling station.

TideRange - Estimate of the mean Tidal Range for each estuary. These values are taken from the closest NOAA tide gauge to the EMAP-E or MAIA sampling station.

stations_no_overlap.zip This file contains the station name of the non-overlaping multi-scaled watersheds used in Chapter 2. The only field in these files is the EMAP-E or MAIA station name.

Appendix 2. De-trending Regression Results for Chapter 2

Table A2-1. Adjusted r^2 and significance for regressions used to remove the effects of point sources and total watershed area.

		Distance of Multi-Scaled Watershed									
		2.5 km	**5 km**	**7.5 km**	**10 km**	**12.5 km**	**15 km**	**17.5 km**	**20 km**	**25 km**	**30 km**
As											
	Adj. r^2	0.0075	-0.0025	-0.0070	-0.0228	-0.0371	-0.0443	0.0123	0.0641	-0.0786	*0.7006*
	p-value	0.2529	0.4140	0.4787	0.7444	0.9332	0.9768	0.2985	0.1407	0.9488	*0.0000*
Cd											
	Adj. r^2	*0.0617*	*0.7986*	*0.7443*	*0.1538*	*0.1610*	0.0053	*0.9253*	*0.3703*	-0.0712	*0.3466*
	p-value	*0.0145*	*0.0000*	*0.0000*	*0.0023*	*0.0047*	0.3345	*0.0000*	*0.0004*	0.8732	*0.0023*
Cr											
	Adj. r^2	0.0023	*0.0902*	*0.0655*	0.0371	0.0522	-0.0165	*0.3508*	0.0089	-0.0728	*0.3168*
	p-value	0.3308	*0.0061*	*0.0325*	0.1181	0.0982	0.5391	*0.0001*	0.3324	0.8894	*0.0040*
Cu											
	Adj. r^2	*0.1326*	*0.3599*	*0.5065*	0.0391	0.0406	-0.0359	*0.3733*	*0.4231*	-0.0738	*0.5212*
	p-value	*0.0003*	*0.0000*	*0.0000*	0.1109	0.1331	0.8169	*0.0001*	*0.0001*	0.8994	*0.0001*
Hg											
	Adj. r^2	0.0229	*0.5422*	*0.2165*	*0.1175*	*0.2535*	*0.1000*	*0.8349*	*0.4161*	0.0263	*0.7947*
	p-value	0.1142	*0.0000*	*0.0001*	*0.0083*	*0.0003*	*0.0371*	*0.0000*	*0.0001*	0.2778	*0.0000*
Pb											
	Adj. r^2	*0.1082*	*0.3914*	*0.2714*	*0.0988*	*0.1478*	0.0347	*0.5949*	*0.6119*	0.0338	*0.5977*
	p-value	*0.0011*	*0.0000*	*0.0000*	*0.0156*	*0.0069*	0.1728	*0.0000*	*0.0000*	0.2532	*0.0000*
Zn											
	Adj. r^2	*0.0541*	*0.1579*	*0.1122*	0.0322	0.0281	-0.0140	*0.2855*	*0.3054*	-0.0525	*0.3640*
	p-value	*0.0218*	*0.0002*	*0.0051*	0.1376	0.1840	0.5107	*0.0006*	*0.0016*	0.7072	*0.0017*

Table A2-2a. Parameter estimates and significance of the parameters for regressions used to remove the effects of point sources and total watershed area.

	Distance of Multi-Scaled Watershed									
	2.5 km		5 km		7.5 km		10 km		12.5 km	
As	**Beta**	**p**	**Beta**	**p**	**Beta**	**p**	**Beta**	**p**	**Beta**	**p**
Intercept	7.7E+00	3.7E-01	1.6E+01	1.9E-02	-1.1E+00	8.9E-01	1.1E+01	1.5E-01	9.0E+00	1.7E-01
Watershed Area	4.6E-04	9.2E-01	-1.1E-03	2.3E-01	5.9E-04	2.3E-01	-1.2E-04	6.5E-01	-4.1E-05	7.8E-01
Point Sources	3.6E-03	9.8E-02	1.5E-04	6.7E-01	-1.6E-05	9.5E-01	-1.1E-04	4.5E-01	-4.7E-05	7.4E-01
Cd										
Intercept	2.6E+00	5.8E-03	1.6E+00	3.2E-05	2.2E-01	6.6E-01	1.1E+00	3.1E-02	7.9E-01	6.6E-02
Watershed Area	-1.1E-03	2.1E-02	-1.8E-04	4.7E-04	6.6E-06	8.3E-01	-2.7E-05	1.2E-01	-1.1E-05	2.4E-01
Point Sources	7.0E-04	1.3E-01	7.5E-04	4.3E-31	4.9E-04	1.9E-23	1.5E-04	5.2E-03	1.3E-04	5.8E-03
Cr										
Intercept	1.0E+02	5.4E-02	1.1E+02	1.8E-02	7.5E+01	9.2E-02	9.8E+01	3.3E-02	7.9E+01	3.0E-02
Watershed Area	-2.6E-02	3.7E-01	-8.2E-03	2.0E-01	-1.4E-03	6.0E-01	-1.7E-03	2.8E-01	-7.9E-04	3.3E-01
Point Sources	1.6E-03	2.5E-01	3.3E-03	1.1E-02	2.5E-03	2.1E-02	6.9E-04	2.1E-01	7.7E-04	1.5E-01
Cu										
Intercept	1.4E+02	2.6E-02	1.1E+02	1.1E-02	-6.1E+01	2.1E-01	1.5E+02	2.0E-02	1.1E+02	2.5E-02
Watershed Area	-5.5E-02	9.3E-02	-1.2E-02	4.4E-02	5.2E-03	8.0E-02	-4.1E-03	6.5E-02	-1.9E-03	8.9E-02
Point Sources	8.8E-03	2.6E-04	1.1E-02	3.1E-08	1.1E-02	5.8E-13	2.3E-04	4.5E-01	2.4E-04	3.9E-01
Hg										
Intercept	9.5E-01	6.5E-02	6.7E-01	8.6E-04	2.8E-01	5.6E-01	4.4E-01	1.7E-01	9.7E-02	6.1E-01
Watershed Area	-4.1E-04	1.4E-01	-7.9E-05	4.7E-03	-7.6E-06	8.0E-01	-1.1E-05	3.0E-01	-4.0E-07	9.3E-01
Point Sources	3.3E-03	1.6E-01	2.6E-03	4.4E-15	1.7E-03	1.2E-05	5.9E-04	1.8E-02	6.8E-04	1.4E-04
Pb										
Intercept	1.0E+02	7.1E-02	1.4E+02	6.4E-06	5.1E+01	2.6E-01	8.2E+01	1.0E-01	5.4E+01	1.3E-01
Watershed Area	-3.3E-02	2.7E-01	-1.6E-02	3.3E-04	-9.0E-04	7.5E-01	-1.7E-03	3.2E-01	-5.5E-04	4.9E-01
Point Sources	1.7E-02	4.7E-04	3.0E-03	9.3E-09	2.0E-03	7.7E-07	1.1E-03	2.5E-02	1.0E-03	6.5E-03
Zn										
Intercept	2.5E+02	3.8E-02	3.4E+02	3.1E-04	1.4E+02	2.0E-01	2.8E+02	3.0E-02	2.0E+02	3.0E-02
Watershed Area	-7.4E-02	2.5E-01	-3.4E-02	9.6E-03	-1.9E-03	7.7E-01	-6.2E-03	1.7E-01	-2.4E-03	2.4E-01
Point Sources	4.6E-03	1.4E-02	1.6E-03	3.2E-03	1.1E-03	1.4E-03	3.5E-04	4.1E-01	3.0E-04	3.7E-01

Table A2-2b. Parameter estimates and significance of the parameters for regressions used to remove the effects of point sources and total watershed area.

	Distance of Multi-Scaled Watershed									
	15 km		17.5 km		20 km		25 km		30 km	
As	**Beta**	**p**	**Beta**	**p**	**Beta**	**p**	**Beta**	**p**	**Beta**	**p**
Intercept	6.4E+00	3.7E-01	3.1E-01	9.6E-01	3.1E+00	6.6E-01	6.1E+00	5.2E-01	2.3E+00	5.8E-01
Watershed Area	1.4E-05	9.0E-01	6.7E-05	3.5E-01	3.1E-05	6.3E-01	5.1E-06	9.3E-01	9.1E-06	5.9E-01
Point Sources	-1.7E-05	8.6E-01	1.8E-04	1.6E-01	7.8E-04	6.3E-02	-2.2E-05	7.6E-01	5.0E-04	4.0E-08
Cd										
Intercept	5.8E-01	3.9E-01	7.4E-01	3.2E-02	4.8E-01	2.9E-01	6.8E-01	3.8E-01	2.6E-01	5.6E-01
Watershed Area	-3.4E-06	7.6E-01	-6.5E-06	1.0E-01	-2.1E-06	6.1E-01	-1.7E-06	7.0E-01	-2.9E-07	8.7E-01
Point Sources	4.1E-05	1.5E-01	2.6E-04	1.0E-22	2.4E-04	8.0E-05	5.3E-06	7.3E-01	1.1E-04	5.7E-04
Cr										
Intercept	7.4E+01	5.3E-02	3.4E+01	3.3E-01	4.1E+01	3.5E-01	8.9E+01	3.2E-01	1.2E+01	7.2E-01
Watershed Area	-4.1E-04	5.0E-01	7.1E-05	8.6E-01	3.9E-05	9.2E-01	-2.0E-04	7.0E-01	1.0E-04	4.6E-01
Point Sources	1.5E-04	3.7E-01	8.9E-04	3.0E-05	1.2E-03	1.4E-01	7.9E-05	7.7E-01	9.7E-04	1.3E-03
Cu										
Intercept	7.0E+01	2.3E-01	4.0E+01	4.8E-01	5.2E+01	2.2E-01	6.9E+01	3.9E-01	1.9E+00	9.7E-01
Watershed Area	-5.9E-04	5.3E-01	-1.4E-04	8.2E-01	-3.6E-04	3.6E-01	-2.0E-04	6.5E-01	6.8E-05	7.3E-01
Point Sources	-4.5E-06	9.5E-01	3.3E-04	2.3E-05	4.4E-03	2.1E-05	-5.3E-06	9.4E-01	1.1E-03	1.3E-05
Hg										
Intercept	2.6E-01	2.9E-01	9.4E-02	5.7E-01	6.8E-02	6.2E-01	5.2E-02	7.9E-01	-8.3E-02	5.5E-01
Watershed Area	-2.4E-06	5.5E-01	-3.6E-07	8.5E-01	1.7E-08	9.9E-01	2.8E-07	8.1E-01	4.8E-07	4.0E-01
Point Sources	2.5E-04	1.3E-02	9.4E-04	1.7E-16	7.6E-04	2.6E-05	6.3E-05	1.2E-01	6.8E-04	4.3E-10
Pb										
Intercept	3.9E+01	5.5E-01	1.3E+01	7.0E-01	6.6E+01	1.5E-01	2.9E+01	4.8E-01	-1.0E+01	7.3E-01
Watershed Area	-1.7E-05	9.9E-01	1.6E-04	6.8E-01	-4.5E-04	2.9E-01	1.0E-05	9.7E-01	1.2E-04	3.2E-01
Point Sources	5.5E-04	6.2E-02	1.1E-03	3.1E-09	5.0E-03	4.6E-08	1.6E-04	1.0E-01	1.1E-03	1.7E-06
Zn										
Intercept	1.8E+02	1.2E-01	1.2E+02	1.8E-01	1.8E+02	7.3E-02	2.2E+02	1.4E-01	7.0E+01	4.0E-01
Watershed Area	-1.2E-03	5.1E-01	-4.4E-04	6.6E-01	-9.3E-04	3.1E-01	-7.0E-04	4.2E-01	8.9E-07	1.0E+00
Point Sources	1.8E-04	3.4E-01	6.1E-04	2.9E-04	2.0E-03	4.2E-04	2.1E-05	8.7E-01	7.4E-04	4.0E-04

Appendix 3. Model Results for Chapter 3

Table A3-1. Full List of Models.

Model Number	Model
1 (Global Model)	metal = urban+point source+est. area+agriculture+river flow+tide range+silt/clay
2	metal = urban+point source+agriculture+river flow+tidal range+silt/clay
3	metal = urban+point source+est. area+agriculture+river flow+tidal range
4	metal = urban+est. area+agriculture+river flow+tidal range+silt/clay
5	metal = point source+est. area+agriculture+river flow+tidal range+silt/clay
6	metal = urban+agriculture+river flow+tidal range+silt/clay
7	metal = urban+point source+agriculture+river flow+tidal range
8	metal = urban+point source+river flow+tidal range+silt/clay
9	metal = urban+point source+est. area+agriculture+silt/clay
10	metal = urban+point source+est. area+river flow+tidal range
11	metal = urban+est. area+agriculture+river flow+tidal range
12	metal = point source+agriculture+river flow+tidal range+silt/clay
13	metal = point source+est. area+agriculture+river flow+tidal range
14	metal = urban+agriculture+river flow+tidal range
15	metal = urban+point source+agriculture+silt/clay
16	metal = urban+point source+river flow+tidal range
17	metal = urban+point source+est. area+agriculture
18	metal = urban+point source+est. area+silt/clay
19	metal = urban+river flow+tidal range+silt/clay
20	metal = urban+est. area+agriculture+silt/clay
21	metal = urban+est. area+river flow+tidal range
22	metal = point source+agriculture+river flow+tidal range
23	metal = point source+river flow+tidal range+silt/clay
24	metal = point source+est. area+agriculture+silt/clay
25	metal = point source+est. area+river flow+tidal range
26	metal = urban+agriculture+silt/clay
27	metal = urban+point source+agriculture
28	metal = urban+point source+silt/clay
29	metal = urban+point source+est. area
30	metal = urban+river flow+tidal range
31	metal = urban+est. area+agriculture
32	metal = urban+est. area+silt/clay
33	metal = point source+agriculture+silt/clay
34	metal = point source+river flow+tidal range
35	metal = point source+est. area+agriculture
36	metal = point source+est. area+silt/clay
37	metal = urban+agriculture
38	metal = urban+point source
39	metal = urban+silt/clay
40	metal = urban+est. area
41	metal = point source+agriculture
42	metal = point source+silt/clay
43	metal = point source+est. area
44	metal = urban
45	metal = point source

Table A3-2a Parameter Estimates of models built for ln(Cu + 1).

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
1(Global)	1.1E+00	2.1E-01	1.7E-04	1.6E-05	-1.4E-06	4.1E-06	-4.0E-04	1.4E-03	-2.9E-06	2.2E-05	3.9E-05	5.0E-05	-3.1E-02	4.2E-02	2.4E-02	1.7E-03	6.0E-01
2	1.1E+00	1.9E-01	1.7E-04	1.5E-05	-1.5E-06	4.1E-06	NA	NA	-2.7E-06	2.2E-05	3.9E-05	5.0E-05	-2.9E-02	4.1E-02	2.4E-02	1.7E-03	6.0E-01
3	2.4E+00	3.2E-01	2.0E-04	2.6E-05	-3.5E-06	7.0E-06	-2.2E-03	2.4E-03	6.0E-05	3.7E-05	1.6E-05	8.5E-05	-1.4E-01	7.1E-02	NA	NA	1.0E+00
4	1.1E+00	2.1E-01	1.7E-04	1.4E-05	NA	NA	-4.2E-04	1.4E-03	-3.5E-06	2.2E-05	3.9E-05	5.0E-05	-3.3E-02	4.2E-02	2.4E-02	1.7E-03	6.0E-01
5	1.6E+00	3.0E-01	NA	NA	1.7E-05	5.5E-06	-1.4E-03	2.0E-03	-9.5E-05	2.9E-05	-1.2E-05	7.2E-05	6.1E-02	6.0E-02	2.6E-02	2.4E-03	8.8E-01
6	1.1E+00	1.9E-01	1.7E-04	1.4E-05	NA	NA	NA	NA	-3.2E-06	2.2E-05	3.9E-05	4.9E-05	-3.1E-02	4.1E-02	2.4E-02	1.7E-03	6.0E-01
7	2.3E+00	2.9E-01	2.0E-04	2.6E-05	-3.9E-06	7.0E-06	NA	NA	6.2E-05	3.6E-05	1.5E-05	8.5E-05	-1.3E-01	7.0E-02	NA	NA	1.0E+00
8	1.1E+00	1.7E-01	1.7E-04	1.4E-05	-1.5E-06	4.1E-06	NA	NA	NA	NA	3.8E-05	4.8E-05	-2.8E-02	4.1E-02	2.4E-02	1.6E-03	6.0E-01
9	1.0E+00	1.5E-01	1.7E-04	1.5E-05	-1.7E-06	4.1E-06	-2.6E-04	1.4E-03	2.8E-06	2.1E-05	NA	NA	NA	NA	2.4E-02	1.6E-03	6.0E-01
10	2.8E+00	2.5E-01	1.9E-04	2.5E-05	-2.8E-06	7.0E-06	-2.5E-03	2.4E-03	NA	NA	4.6E-05	8.3E-05	-1.6E-01	6.9E-02	NA	NA	1.0E+00
11	2.5E+00	3.2E-01	2.0E-04	2.4E-05	NA	NA	-2.3E-03	2.4E-03	5.9E-05	3.6E-05	1.5E-05	8.4E-05	-1.4E-01	7.0E-02	NA	NA	1.0E+00
12	1.6E+00	2.7E-01	NA	NA	1.7E-05	5.5E-06	NA	NA	-9.5E-05	2.9E-05	-1.3E-05	7.2E-05	6.8E-02	5.9E-02	2.7E-02	2.4E-03	8.8E-01
13	3.2E+00	3.8E-01	NA	NA	1.8E-05	8.0E-06	-3.7E-03	3.0E-03	-4.4E-05	4.2E-05	-4.9E-05	1.0E-04	-3.8E-02	8.7E-02	NA	NA	1.3E+00
14	2.3E+00	2.9E-01	2.0E-04	2.4E-05	NA	NA	NA	NA	6.1E-05	3.6E-05	1.5E-05	8.4E-05	-1.3E-01	6.9E-02	NA	NA	1.0E+00
15	1.0E+00	1.3E-01	1.7E-04	1.5E-05	-1.7E-06	4.0E-06	NA	NA	2.9E-06	2.1E-05	NA	NA	NA	NA	2.4E-02	1.6E-03	6.0E-01
16	2.6E+00	2.2E-01	1.9E-04	2.5E-05	-3.1E-06	7.0E-06	NA	NA	NA	NA	4.6E-05	8.3E-05	-1.5E-01	6.9E-02	NA	NA	1.0E+00
17	2.0E+00	2.2E-01	1.9E-04	2.6E-05	-5.3E-06	7.0E-06	-1.6E-03	2.4E-03	7.5E-05	3.5E-05	NA	NA	NA	NA	NA	NA	1.0E+00
18	1.0E+00	1.3E-01	1.7E-04	1.3E-05	-1.7E-06	4.0E-06	-2.6E-04	1.4E-03	NA	NA	NA	NA	NA	NA	2.4E-02	1.6E-03	6.0E-01
19	1.1E+00	1.7E-01	1.7E-04	1.3E-05	NA	NA	NA	NA	NA	NA	3.7E-05	4.8E-05	-3.0E-02	4.0E-02	2.4E-02	1.6E-03	6.0E-01
20	1.0E+00	1.5E-01	1.6E-04	1.4E-05	NA	NA	-2.8E-04	1.4E-03	2.3E-06	2.1E-05	NA	NA	NA	NA	2.4E-02	1.6E-03	6.0E-01
21	2.8E+00	2.5E-01	1.8E-04	2.2E-05	NA	NA	-2.5E-03	2.4E-03	NA	NA	4.5E-05	8.3E-05	-1.6E-01	6.9E-02	NA	NA	1.0E+00
22	3.1E+00	3.5E-01	NA	NA	1.8E-05	8.0E-06	NA	NA	-4.2E-05	4.2E-05	-5.1E-05	1.1E-04	-2.2E-02	8.6E-02	NA	NA	1.3E+00
23	1.2E+00	2.6E-01	NA	NA	1.8E-05	5.7E-06	NA	NA	NA	NA	-7.7E-05	7.2E-05	1.2E-01	5.9E-02	2.5E-02	2.5E-03	9.2E-01
24	1.9E+00	1.8E-01	NA	NA	1.8E-05	5.3E-06	-1.8E-03	2.0E-03	-1.0E-04	2.7E-05	NA	NA	NA	NA	2.6E-02	2.4E-03	8.8E-01
25	3.0E+00	3.1E-01	NA	NA	1.9E-05	7.9E-06	-3.6E-03	3.0E-03	NA	NA	-7.8E-05	1.0E-04	-9.7E-03	8.2E-02	NA	NA	1.3E+00

Table A3-2a (cont.). Parameter Estimates of models built for ln(Cu + 1).

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
26	1.0E+00	1.3E-01	1.6E-04	1.3E-05	NA	NA	NA	NA	2.4E-06	2.1E-05	NA	NA	NA	NA	2.4E-02	1.6E-03	6.0E-01
27	1.9E+00	2.0E-01	1.9E-04	2.6E-05	-5.5E-06	7.0E-06	NA	NA	7.5E-05	3.5E-05	NA	NA	NA	NA	NA	NA	1.0E+00
28	1.0E+00	1.1E-01	1.7E-04	1.3E-05	-1.7E-06	4.0E-06	NA	NA	NA	NA	NA	NA	NA	NA	2.4E-02	1.6E-03	5.9E-01
29	2.3E+00	1.6E-01	1.7E-04	2.4E-05	-4.7E-06	7.1E-06	-1.8E-03	2.4E-03	NA	NA	NA	NA	NA	NA	NA	NA	1.1E+00
30	2.7E+00	2.2E-01	1.8E-04	2.2E-05	NA	NA	NA	NA	NA	NA	4.5E-05	8.3E-05	-1.6E-01	6.8E-02	NA	NA	1.0E+00
31	2.0E+00	2.2E-01	1.8E-04	2.3E-05	NA	NA	-1.7E-03	2.4E-03	7.3E-05	3.5E-05	NA	NA	NA	NA	NA	NA	1.0E+00
32	1.0E+00	1.2E-01	1.6E-04	1.2E-05	NA	NA	-2.8E-04	1.4E-03	NA	NA	NA	NA	NA	NA	2.4E-02	1.6E-03	6.0E-01
33	1.8E+00	1.6E-01	NA	NA	1.8E-05	5.3E-06	NA	NA	-1.0E-04	2.7E-05	NA	NA	NA	NA	2.6E-02	2.4E-03	8.7E-01
34	2.9E+00	2.8E-01	NA	NA	1.9E-05	8.0E-06	NA	NA	NA	NA	-7.9E-05	1.0E-04	5.2E-03	8.1E-02	NA	NA	1.3E+00
35	3.1E+00	2.0E-01	NA	NA	1.7E-05	7.7E-06	-3.5E-03	2.9E-03	-4.3E-05	3.9E-05	NA	NA	NA	NA	NA	NA	1.3E+00
36	1.6E+00	1.8E-01	NA	NA	2.2E-05	5.6E-06	-2.0E-03	2.1E-03	NA	NA	NA	NA	NA	NA	2.4E-02	2.5E-03	9.3E-01
37	1.9E+00	2.0E-01	1.9E-04	2.3E-05	NA	NA	NA	NA	7.4E-05	3.5E-05	NA	NA	NA	NA	NA	NA	1.0E+00
38	2.3E+00	1.3E-01	1.7E-04	2.3E-05	-4.8E-06	7.1E-06	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.1E+00
39	1.0E+00	1.1E-01	1.6E-04	1.2E-05	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.4E-02	1.6E-03	5.9E-01
40	2.3E+00	1.6E-01	1.6E-04	2.1E-05	NA	NA	-1.8E-03	2.4E-03	NA	NA	NA	NA	NA	NA	NA	NA	1.1E+00
41	3.0E+00	1.8E-01	NA	NA	1.7E-05	7.7E-06	NA	NA	-4.3E-05	3.9E-05	NA	NA	NA	NA	NA	NA	1.3E+00
42	1.6E+00	1.6E-01	NA	NA	2.2E-05	5.6E-06	NA	NA	NA	NA	NA	NA	NA	NA	2.4E-02	2.5E-03	9.3E-01
43	3.0E+00	1.6E-01	NA	NA	1.9E-05	7.6E-06	-3.5E-03	2.9E-03	NA	NA	NA	NA	NA	NA	NA	NA	1.3E+00
44	2.3E+00	1.3E-01	1.6E-04	2.1E-05	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.0E+00
45	2.8E+00	1.2E-01	NA	NA	1.9E-05	7.6E-06	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.3E+00

Table A3-2b Parameter Estimates of models built for ln(Pb + 1).

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
1(Global)	2.0E+00	1.9E-01	2.2E-04	2.5E-05	1.9E-06	4.7E-06	8.6E-04	1.2E-03	2.9E-05	3.4E-05	-5.6E-06	4.4E-05	-3.8E-02	3.7E-02	1.7E-02	1.5E-03	5.4E-01
2	2.1E+00	1.7E-01	2.2E-04	2.4E-05	1.9E-06	4.7E-06	NA	NA	2.8E-05	3.4E-05	-5.5E-06	4.4E-05	-4.2E-02	3.6E-02	1.6E-02	1.5E-03	5.3E-01
3	3.0E+00	2.4E-01	2.4E-04	3.6E-05	7.8E-06	6.8E-06	-3.9E-04	1.8E-03	9.8E-05	4.8E-05	-1.5E-05	6.4E-05	-1.3E-01	5.2E-02	NA	NA	7.8E-01
4	2.0E+00	1.8E-01	2.2E-04	2.1E-05	NA	NA	8.6E-04	1.2E-03	3.0E-05	3.4E-05	-5.4E-06	4.4E-05	-3.6E-02	3.6E-02	1.7E-02	1.5E-03	5.3E-01
5	2.5E+00	2.4E-01	NA	NA	2.2E-05	5.4E-06	1.3E-04	1.6E-03	-8.5E-05	4.1E-05	-4.5E-05	5.8E-05	1.9E-02	4.8E-02	1.8E-02	2.0E-03	7.1E-01
6	2.1E+00	1.7E-01	2.2E-04	2.1E-05	NA	NA	NA	NA	2.9E-05	3.3E-05	-5.4E-06	4.4E-05	-4.0E-02	3.6E-02	1.6E-02	1.5E-03	5.3E-01
7	3.0E+00	2.2E-01	2.4E-04	3.5E-05	7.8E-06	6.8E-06	NA	NA	9.9E-05	4.8E-05	-1.5E-05	6.4E-05	-1.3E-01	5.2E-02	NA	NA	7.8E-01
8	2.1E+00	1.5E-01	2.1E-04	2.3E-05	2.2E-06	4.7E-06	NA	NA	NA	NA	1.2E-06	4.3E-05	-4.6E-02	3.6E-02	1.7E-02	1.5E-03	5.3E-01
9	1.9E+00	1.3E-01	2.1E-04	2.4E-05	1.3E-06	4.6E-06	1.0E-03	1.2E-03	3.3E-05	3.3E-05	NA	NA	NA	NA	1.7E-02	1.5E-03	5.3E-01
10	3.3E+00	1.9E-01	2.1E-04	3.4E-05	9.2E-06	6.9E-06	-6.1E-04	1.8E-03	NA	NA	8.8E-06	6.4E-05	-1.5E-01	5.2E-02	NA	NA	7.9E-01
11	3.0E+00	2.3E-01	2.6E-04	3.1E-05	NA	NA	-4.2E-04	1.8E-03	1.0E-04	4.8E-05	-1.5E-05	6.4E-05	-1.2E-01	5.2E-02	NA	NA	7.8E-01
12	2.5E+00	2.2E-01	NA	NA	2.2E-05	5.4E-06	NA	NA	-8.5E-05	4.1E-05	-4.5E-05	5.7E-05	1.8E-02	4.7E-02	1.8E-02	2.0E-03	7.0E-01
13	3.6E+00	2.6E-01	NA	NA	3.1E-05	7.0E-06	-1.3E-03	2.2E-03	-2.3E-05	5.3E-05	-5.9E-05	7.6E-05	-7.4E-02	6.1E-02	NA	NA	9.3E-01
14	2.9E+00	2.1E-01	2.6E-04	3.1E-05	NA	NA	NA	NA	1.0E-04	4.8E-05	-1.5E-05	6.4E-05	-1.2E-01	5.1E-02	NA	NA	7.8E-01
15	1.9E+00	1.2E-01	2.1E-04	2.4E-05	1.2E-06	4.6E-06	NA	NA	3.2E-05	3.3E-05	NA	NA	NA	NA	1.7E-02	1.5E-03	5.3E-01
16	3.3E+00	1.7E-01	2.1E-04	3.3E-05	9.2E-06	6.8E-06	NA	NA	NA	NA	9.0E-06	6.3E-05	-1.5E-01	5.1E-02	NA	NA	7.9E-01
17	2.6E+00	1.7E-01	2.3E-04	3.6E-05	5.9E-06	6.9E-06	1.5E-04	1.8E-03	1.2E-04	4.8E-05	NA	NA	NA	NA	NA	NA	8.0E-01
18	1.9E+00	1.1E-01	2.0E-04	2.2E-05	1.6E-06	4.6E-06	1.0E-03	1.2E-03	NA	NA	NA	NA	NA	NA	1.7E-02	1.4E-03	5.3E-01
19	2.1E+00	1.5E-01	2.1E-04	1.9E-05	NA	NA	NA	NA	NA	NA	1.7E-06	4.3E-05	-4.4E-02	3.6E-02	1.7E-02	1.5E-03	5.3E-01
20	1.9E+00	1.3E-01	2.2E-04	2.0E-05	NA	NA	1.0E-03	1.2E-03	3.3E-05	3.3E-05	NA	NA	NA	NA	1.7E-02	1.5E-03	5.3E-01
21	3.3E+00	1.9E-01	2.4E-04	2.9E-05	NA	NA	-6.6E-04	1.8E-03	NA	NA	1.1E-05	6.4E-05	-1.5E-01	5.2E-02	NA	NA	7.9E-01
22	3.5E+00	2.4E-01	NA	NA	3.1E-05	6.9E-06	NA	NA	-2.2E-05	5.3E-05	-6.0E-05	7.5E-05	-6.9E-02	6.1E-02	NA	NA	9.3E-01
23	2.3E+00	2.0E-01	NA	NA	2.4E-05	5.4E-06	NA	NA	NA	NA	-7.4E-05	5.6E-05	4.2E-02	4.6E-02	1.7E-02	2.0E-03	7.1E-01
24	2.5E+00	1.4E-01	NA	NA	2.3E-05	5.2E-06	5.1E-05	1.6E-03	-9.6E-05	3.9E-05	NA	NA	NA	NA	1.8E-02	1.9E-03	7.0E-01
25	3.5E+00	2.2E-01	NA	NA	3.1E-05	6.9E-06	-1.3E-03	2.1E-03	NA	NA	-6.7E-05	7.3E-05	-6.6E-02	5.8E-02	NA	NA	9.3E-01

Table A3-2b (cont.) Parameter Estimates of models built for ln(Pb + 1).

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
26	1.9E+00	1.2E-01	2.2E-04	2.0E-05	NA	NA	NA	NA	3.2E-05	3.3E-05	NA	NA	NA	NA	1.7E-02	1.4E-03	5.3E-01
27	2.6E+00	1.5E-01	2.3E-04	3.6E-05	5.9E-06	6.9E-06	NA	NA	1.2E-04	4.7E-05	NA	NA	NA	NA	NA	NA	7.9E-01
28	2.0E+00	1.0E-01	2.0E-04	2.1E-05	1.5E-06	4.6E-06	NA	NA	NA	NA	NA	NA	NA	NA	1.7E-02	1.4E-03	5.3E-01
29	2.9E+00	1.3E-01	1.9E-04	3.3E-05	7.4E-06	7.0E-06	-3.5E-05	1.9E-03	NA	NA	NA	NA	NA	NA	NA	NA	8.2E-01
30	3.2E+00	1.7E-01	2.4E-04	2.9E-05	NA	NA	NA	NA	NA	NA	1.1E-05	6.4E-05	-1.4E-01	5.1E-02	NA	NA	7.9E-01
31	2.6E+00	1.7E-01	2.4E-04	3.0E-05	NA	NA	1.1E-04	1.8E-03	1.2E-04	4.8E-05	NA	NA	NA	NA	NA	NA	8.0E-01
32	1.9E+00	1.1E-01	2.1E-04	1.8E-05	NA	NA	1.0E-03	1.2E-03	NA	NA	NA	NA	NA	NA	1.7E-02	1.4E-03	5.3E-01
33	2.5E+00	1.3E-01	NA	NA	2.3E-05	5.2E-06	NA	NA	-9.6E-05	3.9E-05	NA	NA	NA	NA	1.8E-02	1.9E-03	7.0E-01
34	3.5E+00	1.9E-01	NA	NA	3.1E-05	6.9E-06	NA	NA	NA	NA	-6.7E-05	7.3E-05	-6.2E-02	5.8E-02	NA	NA	9.2E-01
35	3.3E+00	1.5E-01	NA	NA	2.9E-05	6.8E-06	-9.4E-04	2.1E-03	-1.5E-05	5.0E-05	NA	NA	NA	NA	NA	NA	9.3E-01
36	2.4E+00	1.4E-01	NA	NA	2.6E-05	5.2E-06	-1.4E-05	1.7E-03	NA	NA	NA	NA	NA	NA	1.6E-02	1.9E-03	7.2E-01
37	2.6E+00	1.5E-01	2.4E-04	3.0E-05	NA	NA	NA	NA	1.2E-04	4.7E-05	NA	NA	NA	NA	NA	NA	7.9E-01
38	2.9E+00	1.0E-01	1.9E-04	3.3E-05	7.4E-06	7.0E-06	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	8.1E-01
39	2.0E+00	9.8E-02	2.1E-04	1.8E-05	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.7E-02	1.4E-03	5.3E-01
40	2.9E+00	1.2E-01	2.1E-04	2.8E-05	NA	NA	-8.9E-05	1.9E-03	NA	NA	NA	NA	NA	NA	NA	NA	8.2E-01
41	3.3E+00	1.3E-01	NA	NA	2.9E-05	6.8E-06	NA	NA	-1.5E-05	5.0E-05	NA	NA	NA	NA	NA	NA	9.3E-01
42	2.4E+00	1.2E-01	NA	NA	2.6E-05	5.1E-06	NA	NA	NA	NA	NA	NA	NA	NA	1.6E-02	1.9E-03	7.2E-01
43	3.3E+00	1.2E-01	NA	NA	3.0E-05	6.7E-06	-9.4E-04	2.1E-03	NA	NA	NA	NA	NA	NA	NA	NA	9.3E-01
44	2.9E+00	1.0E-01	2.1E-04	2.7E-05	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	8.1E-01
45	3.3E+00	9.0E-02	NA	NA	3.0E-05	6.6E-06	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	9.2E-01

Table A3-2c Parameter Estimates of models built for $Hg^{0.2}$.

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
1(Global)	3.4E-01	5.6E-02	2.8E-05	4.8E-06	1.4E-04	1.2E-04	8.2E-05	3.7E-04	-1.9E-06	5.8E-06	1.4E-05	1.3E-05	-1.4E-02	1.1E-02	3.4E-03	4.5E-04	1.6E-01
2	3.4E-01	5.1E-02	2.8E-05	4.8E-06	1.4E-04	1.2E-04	NA	NA	-1.9E-06	5.7E-06	1.4E-05	1.3E-05	-1.4E-02	1.1E-02	3.4E-03	4.4E-04	1.6E-01
3	5.4E-01	6.1E-02	3.1E-05	6.0E-06	1.6E-04	1.5E-04	-1.9E-04	4.6E-04	6.8E-06	7.0E-06	1.1E-05	1.6E-05	-3.2E-02	1.3E-02	NA	NA	2.0E-01
4	3.2E-01	5.4E-02	3.1E-05	3.8E-06	NA	NA	6.9E-05	3.7E-04	-1.1E-06	5.7E-06	1.5E-05	1.3E-05	-1.2E-02	1.1E-02	3.4E-03	4.5E-04	1.6E-01
5	4.3E-01	6.1E-02	NA	NA	5.6E-04	1.0E-04	3.9E-05	4.2E-04	-1.4E-05	6.1E-06	7.0E-06	1.5E-05	-5.9E-03	1.2E-02	3.7E-03	5.1E-04	1.8E-01
6	3.3E-01	5.0E-02	3.1E-05	3.8E-06	NA	NA	NA	NA	-1.1E-06	5.7E-06	1.5E-05	1.3E-05	-1.2E-02	1.1E-02	3.4E-03	4.4E-04	1.6E-01
7	5.3E-01	5.6E-02	3.1E-05	5.9E-06	1.6E-04	1.5E-04	NA	NA	7.0E-06	7.0E-06	1.1E-05	1.6E-05	-3.1E-02	1.3E-02	NA	NA	2.0E-01
8	3.3E-01	4.5E-02	2.8E-05	4.4E-06	1.3E-04	1.2E-04	NA	NA	NA	NA	1.3E-05	1.3E-05	-1.3E-02	1.1E-02	3.4E-03	4.3E-04	1.6E-01
9	2.9E-01	3.9E-02	2.7E-05	4.8E-06	1.2E-04	1.2E-04	1.4E-04	3.7E-04	2.0E-07	5.6E-06	NA	NA	NA	NA	3.5E-03	4.4E-04	1.6E-01
10	5.7E-01	4.8E-02	2.9E-05	5.5E-06	1.8E-04	1.4E-04	-2.1E-04	4.6E-04	NA	NA	1.5E-05	1.6E-05	-3.4E-02	1.3E-02	NA	NA	2.0E-01
11	5.2E-01	5.9E-02	3.6E-05	4.7E-06	NA	NA	-2.1E-04	4.6E-04	7.7E-06	7.0E-06	1.2E-05	1.6E-05	-2.9E-02	1.3E-02	NA	NA	2.0E-01
12	4.3E-01	5.6E-02	NA	NA	5.6E-04	1.0E-04	NA	NA	-1.4E-05	6.0E-06	7.0E-06	1.5E-05	-6.0E-03	1.2E-02	3.7E-03	5.0E-04	1.8E-01
13	6.6E-01	6.3E-02	NA	NA	6.4E-04	1.3E-04	-2.6E-04	5.1E-04	-6.8E-06	7.3E-06	2.9E-06	1.8E-05	-2.4E-02	1.5E-02	NA	NA	2.2E-01
14	5.1E-01	5.4E-02	3.6E-05	4.6E-06	NA	NA	NA	NA	7.9E-06	6.9E-06	1.2E-05	1.6E-05	-2.9E-02	1.3E-02	NA	NA	2.0E-01
15	3.0E-01	3.6E-02	2.7E-05	4.8E-06	1.2E-04	1.2E-04	NA	NA	1.7E-07	5.6E-06	NA	NA	NA	NA	3.5E-03	4.3E-04	1.6E-01
16	5.6E-01	4.3E-02	2.9E-05	5.5E-06	1.8E-04	1.4E-04	NA	NA	NA	NA	1.5E-05	1.6E-05	-3.4E-02	1.3E-02	NA	NA	2.0E-01
17	4.4E-01	4.4E-02	3.0E-05	6.0E-06	1.1E-04	1.5E-04	-6.4E-05	4.6E-04	1.1E-05	6.9E-06	NA	NA	NA	NA	NA	NA	2.0E-01
18	2.9E-01	3.4E-02	2.6E-05	4.3E-06	1.2E-04	1.1E-04	1.4E-04	3.7E-04	NA	NA	NA	NA	NA	NA	3.5E-03	4.2E-04	1.6E-01
19	3.2E-01	4.4E-02	3.1E-05	3.4E-06	NA	NA	NA	NA	NA	NA	1.4E-05	1.3E-05	-1.2E-02	1.1E-02	3.4E-03	4.3E-04	1.6E-01
20	2.8E-01	3.9E-02	3.0E-05	3.6E-06	NA	NA	1.2E-04	3.7E-04	7.9E-07	5.6E-06	NA	NA	NA	NA	3.5E-03	4.4E-04	1.6E-01
21	5.6E-01	4.7E-02	3.4E-05	4.3E-06	NA	NA	-2.3E-04	4.6E-04	NA	NA	1.6E-05	1.6E-05	-3.2E-02	1.3E-02	NA	NA	2.0E-01
22	6.5E-01	5.8E-02	NA	NA	6.5E-04	1.3E-04	NA	NA	-6.7E-06	7.3E-06	2.8E-06	1.8E-05	-2.3E-02	1.5E-02	NA	NA	2.2E-01
23	3.8E-01	5.2E-02	NA	NA	6.0E-04	1.0E-04	NA	NA	NA	NA	-2.4E-06	1.5E-05	3.0E-04	1.2E-02	3.4E-03	5.1E-04	1.8E-01
24	4.1E-01	3.8E-02	NA	NA	5.4E-04	9.9E-05	6.5E-05	4.2E-04	-1.3E-05	5.7E-06	NA	NA	NA	NA	3.7E-03	4.9E-04	1.8E-01
25	6.3E-01	5.2E-02	NA	NA	6.6E-04	1.3E-04	-2.4E-04	5.1E-04	NA	NA	-1.6E-06	1.7E-05	-2.1E-02	1.4E-02	NA	NA	2.2E-01

Table A3-2c (cont.) Parameter Estimates of models built for $Hg^{0.2}$.

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
26	2.9E-01	3.5E-02	3.0E-05	3.6E-06	NA	NA	NA	NA	7.6E-07	5.5E-06	NA	NA	NA	NA	3.5E-03	4.3E-04	1.6E-01
27	4.3E-01	4.0E-02	3.0E-05	6.0E-06	1.1E-04	1.5E-04	NA	NA	1.1E-05	6.8E-06	NA	NA	NA	NA	NA	NA	2.0E-01
28	3.0E-01	3.0E-02	2.6E-05	4.3E-06	1.2E-04	1.1E-04	NA	NA	NA	NA	NA	NA	NA	NA	3.5E-03	4.2E-04	1.6E-01
29	4.8E-01	3.1E-02	2.6E-05	5.5E-06	1.3E-04	1.5E-04	-8.4E-05	4.7E-04	NA	NA	NA	NA	NA	NA	NA	NA	2.0E-01
30	5.5E-01	4.1E-02	3.4E-05	4.3E-06	NA	NA	NA	NA	NA	NA	1.6E-05	1.6E-05	-3.1E-02	1.3E-02	NA	NA	2.0E-01
31	4.3E-01	4.3E-02	3.3E-05	4.5E-06	NA	NA	-8.3E-05	4.6E-04	1.1E-05	6.8E-06	NA	NA	NA	NA	NA	NA	2.0E-01
32	2.9E-01	3.3E-02	3.0E-05	3.2E-06	NA	NA	1.2E-04	3.7E-04	NA	NA	NA	NA	NA	NA	3.5E-03	4.2E-04	1.6E-01
33	4.1E-01	3.3E-02	NA	NA	5.4E-04	9.8E-05	NA	NA	-1.3E-05	5.7E-06	NA	NA	NA	NA	3.7E-03	4.9E-04	1.8E-01
34	6.2E-01	4.7E-02	NA	NA	6.6E-04	1.2E-04	NA	NA	NA	NA	-1.6E-06	1.7E-05	-2.0E-02	1.4E-02	NA	NA	2.2E-01
35	5.8E-01	3.7E-02	NA	NA	5.8E-04	1.2E-04	-1.6E-04	5.1E-04	-3.9E-06	6.9E-06	NA	NA	NA	NA	NA	NA	2.2E-01
36	3.8E-01	3.6E-02	NA	NA	6.0E-04	9.8E-05	6.1E-05	4.2E-04	NA	NA	NA	NA	NA	NA	3.4E-03	4.9E-04	1.8E-01
37	4.3E-01	3.9E-02	3.3E-05	4.5E-06	NA	NA	NA	NA	1.1E-05	6.8E-06	NA	NA	NA	NA	NA	NA	2.0E-01
38	4.8E-01	2.6E-02	2.6E-05	5.5E-06	1.3E-04	1.5E-04	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.0E-01
39	2.9E-01	2.9E-02	3.0E-05	3.1E-06	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.5E-03	4.2E-04	1.6E-01
40	4.8E-01	3.1E-02	2.9E-05	4.0E-06	NA	NA	-1.1E-04	4.6E-04	NA	NA	NA	NA	NA	NA	NA	NA	2.0E-01
41	5.7E-01	3.2E-02	NA	NA	5.9E-04	1.2E-04	NA	NA	-3.8E-06	6.8E-06	NA	NA	NA	NA	NA	NA	2.2E-01
42	3.8E-01	3.1E-02	NA	NA	6.0E-04	9.7E-05	NA	NA	NA	NA	NA	NA	NA	NA	3.4E-03	4.9E-04	1.8E-01
43	5.6E-01	2.9E-02	NA	NA	6.0E-04	1.2E-04	-1.6E-04	5.1E-04	NA	NA	NA	NA	NA	NA	NA	NA	2.2E-01
44	4.7E-01	2.5E-02	2.9E-05	4.0E-06	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.0E-01
45	5.6E-01	2.2E-02	NA	NA	6.1E-04	1.2E-04	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.2E-01

Table A3-2d Parameter Estimates of models built for $Cd^{0.3}$

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
1 (Global)	3.4E-01	8.4E-02	5.0E-04	7.7E-05	-9.3E-05	1.5E-04	-1.2E-03	5.8E-04	-6.5E-06	1.5E-04	2.4E-06	1.9E-05	-4.3E-03	1.7E-02	4.8E-03	6.7E-04	2.4E-01
2	2.6E-01	7.6E-02	5.4E-04	7.6E-05	-1.2E-04	1.5E-04	NA	NA	4.6E-05	1.5E-04	2.8E-06	2.0E-05	9.4E-04	1.7E-02	5.0E-03	6.8E-04	2.5E-01
3	6.5E-01	8.7E-02	5.2E-04	9.4E-05	5.9E-05	1.8E-04	-1.6E-03	7.0E-04	9.1E-05	1.8E-04	2.9E-06	2.4E-05	-2.8E-02	2.0E-02	NA	NA	2.9E-01
4	3.4E-01	8.3E-02	4.9E-04	7.3E-05	NA	NA	-1.2E-03	5.8E-04	1.5E-06	1.5E-04	2.3E-06	1.9E-05	-3.8E-03	1.6E-02	4.8E-03	6.6E-04	2.4E-01
5	4.4E-01	9.7E-02	NA	NA	2.1E-04	1.7E-04	-2.0E-03	6.7E-04	-1.9E-04	1.8E-04	-1.6E-05	2.3E-05	1.7E-02	1.9E-02	5.0E-03	7.9E-04	2.9E-01
6	2.6E-01	7.6E-02	5.2E-04	7.3E-05	NA	NA	NA	NA	5.8E-05	1.5E-04	2.7E-06	2.0E-05	1.7E-03	1.7E-02	4.9E-03	6.7E-04	2.5E-01
7	5.6E-01	7.8E-02	5.6E-04	9.3E-05	2.7E-05	1.9E-04	NA	NA	1.7E-04	1.9E-04	3.5E-06	2.4E-05	-2.2E-02	2.0E-02	NA	NA	3.0E-01
8	2.7E-01	6.8E-02	5.3E-04	7.5E-05	-1.2E-04	1.5E-04	NA	NA	NA	NA	3.2E-06	2.0E-05	-4.5E-04	1.6E-02	5.0E-03	6.7E-04	2.4E-01
9	3.2E-01	5.4E-02	5.0E-04	7.4E-05	-9.1E-05	1.5E-04	-1.2E-03	5.7E-04	5.6E-06	1.4E-04	NA	NA	NA	NA	4.9E-03	6.5E-04	2.4E-01
10	6.8E-01	7.0E-02	5.1E-04	9.2E-05	5.2E-05	1.8E-04	-1.7E-03	6.9E-04	NA	NA	3.7E-06	2.3E-05	-3.1E-02	1.9E-02	NA	NA	2.9E-01
11	6.6E-01	8.6E-02	5.3E-04	8.9E-05	NA	NA	-1.6E-03	7.0E-04	8.7E-05	1.8E-04	3.0E-06	2.3E-05	-2.9E-02	2.0E-02	NA	NA	2.9E-01
12	3.1E-01	9.1E-02	NA	NA	2.0E-04	1.8E-04	NA	NA	-1.3E-04	1.8E-04	-1.7E-05	2.3E-05	2.9E-02	1.9E-02	5.2E-03	8.2E-04	3.0E-01
13	7.7E-01	9.5E-02	NA	NA	3.7E-04	2.0E-04	-2.5E-03	7.8E-04	-1.0E-04	2.1E-04	-1.6E-05	2.6E-05	-7.0E-03	2.2E-02	NA	NA	3.3E-01
14	5.6E-01	7.7E-02	5.7E-04	8.9E-05	NA	NA	NA	NA	1.6E-04	1.8E-04	3.5E-06	2.4E-05	-2.2E-02	2.0E-02	NA	NA	3.0E-01
15	2.6E-01	4.7E-02	5.4E-04	7.3E-05	-1.2E-04	1.5E-04	NA	NA	4.4E-05	1.5E-04	NA	NA	NA	NA	5.0E-03	6.6E-04	2.4E-01
16	6.0E-01	6.3E-02	5.5E-04	9.2E-05	1.3E-05	1.9E-04	NA	NA	NA	NA	5.0E-06	2.4E-05	-2.8E-02	1.9E-02	NA	NA	3.0E-01
17	5.6E-01	5.5E-02	4.9E-04	9.1E-05	7.9E-05	1.8E-04	-1.5E-03	7.0E-04	1.8E-04	1.8E-04	NA	NA	NA	NA	NA	NA	2.9E-01
18	3.2E-01	5.0E-02	5.0E-04	7.1E-05	-9.1E-05	1.5E-04	-1.2E-03	5.6E-04	NA	NA	NA	NA	NA	NA	4.9E-03	6.4E-04	2.4E-01
19	2.7E-01	6.8E-02	5.1E-04	7.1E-05	NA	NA	NA	NA	NA	NA	3.2E-06	2.0E-05	1.1E-05	1.6E-02	4.9E-03	6.7E-04	2.4E-01
20	3.3E-01	5.4E-02	4.8E-04	7.1E-05	NA	NA	-1.2E-03	5.6E-04	1.2E-05	1.4E-04	NA	NA	NA	NA	4.8E-03	6.4E-04	2.4E-01
21	6.8E-01	6.9E-02	5.2E-04	8.6E-05	NA	NA	-1.7E-03	6.9E-04	NA	NA	3.7E-06	2.3E-05	-3.2E-02	1.8E-02	NA	NA	2.9E-01
22	6.3E-01	8.9E-02	NA	NA	3.7E-04	2.1E-04	NA	NA	-9.7E-06	2.1E-04	-1.8E-05	2.7E-05	5.8E-03	2.2E-02	NA	NA	3.5E-01
23	2.8E-01	8.2E-02	NA	NA	2.1E-04	1.8E-04	NA	NA	NA	NA	-1.9E-05	2.3E-05	3.3E-02	1.8E-02	5.1E-03	8.1E-04	3.0E-01
24	5.0E-01	5.6E-02	NA	NA	2.1E-04	1.7E-04	-2.1E-03	6.5E-04	-2.6E-04	1.6E-04	NA	NA	NA	NA	4.8E-03	7.7E-04	2.8E-01
25	7.4E-01	7.8E-02	NA	NA	3.9E-04	2.0E-04	-2.4E-03	7.7E-04	NA	NA	-1.7E-05	2.6E-05	-3.1E-03	2.0E-02	NA	NA	3.3E-01

Table A3-2d (cont.) Parameter Estimates of models built for Cd $^{0.3}$

	Intercept		Urban		Point Sources		Estuarine Area		Agriculture		River Flow		Tidal Range		Silt/Clay		Model
Model #	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	Beta	S.E.	S.E.
26	2.7E-01	4.7E-02	5.2E-04	7.0E-05	NA	NA	NA	NA	5.5E-05	1.4E-04	NA	NA	NA	NA	4.9E-03	6.5E-04	2.4E-01
27	4.9E-01	4.5E-02	5.4E-04	9.0E-05	4.6E-05	1.9E-04	NA	NA	2.3E-04	1.8E-04	NA	NA	NA	NA	NA	NA	3.0E-01
28	2.7E-01	4.4E-02	5.3E-04	7.0E-05	-1.2E-04	1.5E-04	NA	NA	NA	NA	NA	NA	NA	NA	5.0E-03	6.5E-04	2.4E-01
29	5.9E-01	4.4E-02	4.7E-04	8.8E-05	7.0E-05	1.8E-04	-1.6E-03	6.9E-04	NA	NA	NA	NA	NA	NA	NA	NA	2.9E-01
30	6.0E-01	6.3E-02	5.5E-04	8.7E-05	NA	NA	NA	NA	NA	NA	5.0E-06	2.4E-05	-2.8E-02	1.9E-02	NA	NA	3.0E-01
31	5.6E-01	5.4E-02	5.0E-04	8.7E-05	NA	NA	-1.4E-03	6.9E-04	1.7E-04	1.7E-04	NA	NA	NA	NA	NA	NA	2.9E-01
32	3.3E-01	5.0E-02	4.8E-04	6.7E-05	NA	NA	-1.2E-03	5.6E-04	NA	NA	NA	NA	NA	NA	4.8E-03	6.3E-04	2.4E-01
33	4.2E-01	5.2E-02	NA	NA	2.0E-04	1.8E-04	NA	NA	-2.2E-04	1.7E-04	NA	NA	NA	NA	5.0E-03	8.1E-04	3.0E-01
34	6.3E-01	7.2E-02	NA	NA	3.7E-04	2.0E-04	NA	NA	NA	NA	-1.8E-05	2.7E-05	6.1E-03	2.1E-02	NA	NA	3.5E-01
35	7.3E-01	4.9E-02	NA	NA	3.8E-04	2.0E-04	-2.4E-03	7.6E-04	-8.3E-05	1.9E-04	NA	NA	NA	NA	NA	NA	3.3E-01
36	4.8E-01	5.4E-02	NA	NA	2.6E-04	1.7E-04	-2.1E-03	6.6E-04	NA	NA	NA	NA	NA	NA	4.6E-03	7.7E-04	2.9E-01
37	4.9E-01	4.4E-02	5.5E-04	8.5E-05	NA	NA	NA	NA	2.3E-04	1.7E-04	NA	NA	NA	NA	NA	NA	3.0E-01
38	5.3E-01	3.5E-02	5.1E-04	8.7E-05	3.0E-05	1.9E-04	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.0E-01
39	2.7E-01	4.4E-02	5.1E-04	6.7E-05	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4.9E-03	6.4E-04	2.4E-01
40	5.9E-01	4.4E-02	4.8E-04	8.2E-05	NA	NA	-1.5E-03	6.9E-04	NA	NA	NA	NA	NA	NA	NA	NA	2.9E-01
41	6.5E-01	4.2E-02	NA	NA	3.7E-04	2.0E-04	NA	NA	-3.7E-05	2.0E-04	NA	NA	NA	NA	NA	NA	3.4E-01
42	4.0E-01	5.0E-02	NA	NA	2.4E-04	1.8E-04	NA	NA	NA	NA	NA	NA	NA	NA	4.8E-03	8.0E-04	3.0E-01
43	7.2E-01	4.1E-02	NA	NA	3.9E-04	1.9E-04	-2.4E-03	7.5E-04	NA	NA	NA	NA	NA	NA	NA	NA	3.3E-01
44	5.3E-01	3.5E-02	5.2E-04	8.2E-05	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.0E-01
45	6.4E-01	3.3E-02	NA	NA	3.8E-04	2.0E-04	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.4E-01

Table A3-3a. Model fit and AIC statistics of Each Model ln(Cu + 1)

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC w
39	*0.793*	*4*	*-98.82*	*206.02*	*0.00*	*0.3103*
19	*0.794*	*6*	*-97.50*	*207.81*	*1.79*	*0.1267*
28	*0.792*	*5*	*-98.73*	*208.02*	*2.01*	*0.1137*
32	*0.791*	*5*	*-98.80*	*208.16*	*2.15*	*0.1060*
26	*0.791*	*5*	*-98.81*	*208.19*	*2.18*	*0.1044*
8	*0.792*	*7*	*-97.46*	*209.99*	*3.98*	*0.0425*
6	*0.792*	*7*	*-97.48*	*210.04*	*4.02*	*0.0416*
18	*0.790*	*6*	*-98.71*	*210.22*	*4.20*	*0.0379*
15	*0.790*	*6*	*-98.72*	*210.24*	*4.22*	*0.0376*
20	*0.789*	*6*	*-98.79*	*210.39*	*4.37*	*0.0349*
4	*0.791*	*8*	*-97.39*	*212.19*	*6.17*	*0.0142*
2	*0.790*	*8*	*-97.44*	*212.27*	*6.26*	*0.0136*
9	*0.788*	*7*	*-98.70*	*212.48*	*6.46*	*0.0123*
1 (global)	*0.789*	*9*	*-97.36*	*214.48*	*8.47*	*0.0045*
33	0.549	5	-141.92	294.41	88.40	0.0000
24	0.548	6	-141.52	295.85	89.83	0.0000
12	0.543	7	-141.63	298.34	92.32	0.0000
5	0.541	8	-141.33	300.05	94.04	0.0000
42	0.492	4	-149.13	306.63	100.62	0.0000
23	0.498	6	-147.41	307.62	101.61	0.0000
36	0.491	5	-148.68	307.93	101.91	0.0000
14	0.399	6	-157.48	327.77	121.75	0.0000
30	0.390	5	-158.87	328.30	122.29	0.0000
11	0.400	7	-156.90	328.89	122.87	0.0000
21	0.391	6	-158.21	329.22	123.20	0.0000
7	0.395	7	-157.37	329.82	123.80	0.0000
37	0.373	4	-160.96	330.29	124.27	0.0000
16	0.385	6	-158.80	330.40	124.38	0.0000
3	0.395	8	-156.81	331.03	125.01	0.0000
10	0.386	7	-158.16	331.39	125.37	0.0000
27	0.370	5	-160.64	331.85	125.83	0.0000
31	0.370	5	-160.71	331.98	125.96	0.0000
44	0.353	3	-163.22	332.65	126.64	0.0000
17	0.367	6	-160.41	333.61	127.59	0.0000
40	0.350	4	-162.92	334.22	128.20	0.0000
38	0.349	4	-162.98	334.33	128.31	0.0000
29	0.347	5	-162.70	335.96	129.95	0.0000
45	0.045	3	-185.00	376.23	170.21	0.0000
43	0.049	4	-184.25	376.88	170.86	0.0000
41	0.047	4	-184.36	377.09	171.08	0.0000
35	0.051	5	-183.62	377.80	171.78	0.0000
34	0.034	5	-184.58	379.72	173.70	0.0000
25	0.040	6	-183.71	380.22	174.20	0.0000
22	0.037	6	-183.90	380.60	174.58	0.0000
13	0.044	7	-182.98	381.03	175.01	0.0000

Table A3-3b. Model fit and AIC statistics of Each Model ln(Pb + 1).

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC *w*
39	*0.719*	*4*	*-86.36*	*181.10*	*0.00*	*0.2744*
26	*0.719*	*5*	*-85.85*	*182.27*	*1.16*	*0.1533*
32	*0.718*	*5*	*-86.02*	*182.60*	*1.50*	*0.1295*
28	*0.717*	*5*	*-86.31*	*183.19*	*2.09*	*0.0967*
20	*0.718*	*6*	*-85.48*	*183.76*	*2.66*	*0.0727*
19	*0.718*	*6*	*-85.55*	*183.90*	*2.80*	*0.0676*
15	*0.717*	*6*	*-85.81*	*184.43*	*3.33*	*0.0520*
18	*0.716*	*6*	*-85.96*	*184.72*	*3.62*	*0.0449*
6	*0.717*	*7*	*-85.16*	*185.40*	*4.29*	*0.0321*
8	*0.716*	*7*	*-85.43*	*185.94*	*4.84*	*0.0244*
9	*0.716*	*7*	*-85.44*	*185.96*	*4.85*	*0.0242*
4	*0.716*	*8*	*-84.90*	*187.21*	*6.11*	*0.0130*
2	*0.715*	*8*	*-85.07*	*187.54*	*6.43*	*0.0110*
1 (global)	*0.714*	*9*	*-84.81*	*189.39*	*8.29*	*0.0043*
33	0.512	5	-116.76	244.09	62.99	0.0000
24	0.508	6	-116.76	246.33	65.22	0.0000
12	0.506	7	-116.40	247.88	66.78	0.0000
42	0.489	4	-119.86	248.09	66.99	0.0000
23	0.491	6	-118.65	250.10	68.99	0.0000
5	0.501	8	-116.40	250.20	69.10	0.0000
36	0.484	5	-119.86	250.28	69.18	0.0000
14	0.395	6	-128.28	269.37	88.27	0.0000
7	0.397	7	-127.58	270.24	89.14	0.0000
37	0.373	4	-131.38	271.13	90.02	0.0000
11	0.390	7	-128.26	271.59	90.49	0.0000
30	0.374	5	-130.71	271.98	90.87	0.0000
16	0.379	6	-129.76	272.33	91.23	0.0000
3	0.392	8	-127.56	272.51	91.41	0.0000
27	0.371	5	-130.99	272.55	91.45	0.0000
31	0.367	5	-131.37	273.31	92.21	0.0000
21	0.369	6	-130.64	274.08	92.98	0.0000
10	0.374	7	-129.70	274.49	93.38	0.0000
17	0.365	6	-130.99	274.78	93.68	0.0000
44	0.341	3	-134.66	275.55	94.45	0.0000
38	0.341	4	-134.10	276.57	95.47	0.0000
40	0.335	4	-134.66	277.70	96.60	0.0000
29	0.335	5	-134.10	278.76	97.66	0.0000
45	0.149	3	-148.97	304.16	123.05	0.0000
43	0.142	4	-148.87	306.11	125.01	0.0000
41	0.142	4	-148.92	306.21	125.11	0.0000
34	0.150	5	-147.85	306.27	125.16	0.0000
25	0.145	6	-147.67	308.14	127.04	0.0000
35	0.135	5	-148.82	308.21	127.10	0.0000
22	0.144	6	-147.76	308.32	127.22	0.0000
13	0.138	7	-147.57	310.21	129.11	0.0000

Table A3-3c. Model fit and AIC statistics of Each Model (Hg $^{0.2}$).

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC *w*
39	*0.582*	*4*	*49.08*	*-89.78*	*0.00*	*0.2749*
28	*0.583*	*5*	*49.65*	*-88.73*	*1.05*	*0.1630*
32	*0.579*	*5*	*49.13*	*-87.69*	*2.09*	*0.0969*
26	*0.579*	*5*	*49.08*	*-87.60*	*2.17*	*0.0928*
19	*0.583*	*6*	*50.16*	*-87.53*	*2.25*	*0.0893*
8	*0.584*	*7*	*50.87*	*-86.66*	*3.11*	*0.0580*
18	*0.579*	*6*	*49.72*	*-86.65*	*3.13*	*0.0575*
15	*0.579*	*6*	*49.65*	*-86.50*	*3.28*	*0.0534*
20	*0.575*	*6*	*49.14*	*-85.48*	*4.30*	*0.0320*
6	*0.579*	*7*	*50.18*	*-85.29*	*4.48*	*0.0292*
2	*0.581*	*8*	*50.93*	*-84.46*	*5.32*	*0.0193*
9	*0.575*	*7*	*49.72*	*-84.37*	*5.41*	*0.0184*
4	*0.575*	*8*	*50.20*	*-83.01*	*6.77*	*0.0093*
1 (global)	*0.577*	*9*	*50.95*	*-82.15*	*7.63*	*0.0061*
33	0.462	5	35.46	-60.35	29.42	0.0000
24	0.458	6	35.47	-58.14	31.63	0.0000
42	0.441	4	32.73	-57.09	32.68	0.0000
12	0.454	7	35.67	-56.27	33.51	0.0000
36	0.436	5	32.74	-54.92	34.86	0.0000
5	0.449	8	35.68	-53.96	35.82	0.0000
23	0.431	6	32.75	-52.69	37.08	0.0000
30	0.349	5	24.73	-38.90	50.88	0.0000
16	0.352	6	25.55	-38.29	51.48	0.0000
14	0.351	6	25.41	-38.01	51.76	0.0000
37	0.333	4	22.89	-37.41	52.37	0.0000
7	0.352	7	26.07	-37.06	52.71	0.0000
21	0.344	6	24.87	-36.93	52.84	0.0000
44	0.323	3	21.52	-36.82	52.95	0.0000
10	0.348	7	25.66	-36.24	53.54	0.0000
11	0.346	7	25.51	-35.95	53.83	0.0000
27	0.331	5	23.20	-35.83	53.95	0.0000
38	0.322	4	21.96	-35.55	54.23	0.0000
31	0.327	5	22.91	-35.25	54.53	0.0000
3	0.347	8	26.16	-34.92	54.85	0.0000
40	0.317	4	21.55	-34.73	55.05	0.0000
17	0.325	6	23.21	-33.61	56.16	0.0000
29	0.316	5	21.98	-33.39	56.38	0.0000
45	0.190	3	11.49	-16.76	73.01	0.0000
41	0.185	4	11.65	-14.94	74.84	0.0000
43	0.184	4	11.54	-14.71	75.06	0.0000
34	0.191	5	12.53	-14.50	75.27	0.0000
22	0.189	6	12.97	-13.15	76.63	0.0000
35	0.179	5	11.71	-12.85	76.93	0.0000
25	0.185	6	12.65	-12.50	77.27	0.0000
13	0.184	7	13.11	-11.15	78.63	0.0000

Table A3-3d. Model fit and AIC statistics (Cd $^{0.3}$).

Model #	Adj. r^2	K	log Likelihood	AIC_C	AIC Delta	AIC *w*
32	*0.530*	*5*	*4.02*	*2.53*	*0.00*	*0.4093*
18	*0.527*	*6*	*4.21*	*4.37*	*1.85*	*0.1626*
20	*0.525*	*6*	*4.02*	*4.75*	*2.23*	*0.1345*
39	*0.513*	*4*	*1.52*	*5.34*	*2.81*	*0.1004*
9	*0.522*	*7*	*4.21*	*6.65*	*4.12*	*0.0521*
28	*0.511*	*5*	*1.87*	*6.82*	*4.30*	*0.0477*
26	*0.509*	*5*	*1.59*	*7.38*	*4.86*	*0.0361*
15	*0.507*	*6*	*1.92*	*8.96*	*6.43*	*0.0164*
4	*0.517*	*8*	*4.06*	*9.29*	*6.76*	*0.0139*
19	*0.504*	*6*	*1.53*	*9.73*	*7.21*	*0.0111*
1 (global)	*0.514*	*9*	*4.25*	*11.26*	*8.73*	*0.0052*
8	*0.502*	*7*	*1.89*	*11.31*	*8.78*	*0.0051*
6	*0.500*	*7*	*1.61*	*11.86*	*9.33*	*0.0039*
2	*0.498*	*8*	*1.93*	*13.53*	*11.00*	*0.0017*
24	0.326	6	-15.63	44.06	41.53	0.0000
36	0.317	5	-16.87	44.31	41.78	0.0000
5	0.320	8	-15.01	47.42	44.90	0.0000
40	0.283	4	-20.07	48.52	45.99	0.0000
31	0.283	5	-19.57	49.70	47.17	0.0000
21	0.290	6	-18.56	49.91	47.39	0.0000
29	0.278	5	-20.00	50.56	48.04	0.0000
44	0.257	3	-22.63	51.49	48.96	0.0000
17	0.278	6	-19.47	51.74	49.21	0.0000
37	0.261	4	-21.79	51.95	49.42	0.0000
11	0.284	7	-18.44	51.95	49.43	0.0000
42	0.260	4	-21.85	52.07	49.54	0.0000
10	0.283	7	-18.51	52.10	49.58	0.0000
33	0.265	5	-21.00	52.57	50.04	0.0000
23	0.272	6	-19.94	52.68	50.15	0.0000
30	0.258	5	-21.52	53.61	51.09	0.0000
38	0.250	4	-22.62	53.61	51.09	0.0000
27	0.255	5	-21.76	54.08	51.55	0.0000
3	0.278	8	-18.38	54.16	51.64	0.0000
12	0.268	7	-19.69	54.46	51.93	0.0000
14	0.256	6	-21.12	55.03	52.51	0.0000
16	0.251	6	-21.52	55.84	53.32	0.0000
7	0.249	7	-21.11	57.29	54.76	0.0000
43	0.096	4	-33.07	74.51	71.99	0.0000
35	0.089	5	-32.97	76.50	73.98	0.0000
25	0.083	6	-32.82	78.45	75.92	0.0000
13	0.077	7	-32.70	80.48	77.95	0.0000
45	0.022	3	-37.99	82.21	79.68	0.0000
41	0.013	4	-37.97	84.32	81.80	0.0000
34	0.009	5	-37.74	86.04	83.51	0.0000
22	-0.001	6	-37.74	88.27	85.75	0.0000

Bibliography

Anderson, D. R., Burnham, K. P., and Thompson, W. L. 2000. Null Hypothesis Testing: Problems, Prevalence and An Alternative. *Journal of Wildlife Management* 64: 912-923.

Anderson, J. R., Hardy, E. E., Roach, J. T., and Witmer, R. E. 1976. *USGS Professional Paper 964*. U.S. Geological Survey, Washington, D.C.

Basnyat, P., Teeter, L. D., Flynn, K. M., and Lockaby, B. G. 1999. Relationships Between Landscape Characteristics and Nonpoint Source Pollution Inputs to Coastal Estuaries. *Environmental Management* 23: 539-549.

Beaulac, M. N. and Reckhow, K. H. 1982. An Examination of Land Use - Nutrient Export Relationships. *Water Resources Bulleting* 18: 1013-1024.

Blake, A. C., Chadwick, D. B., Zirino, A., and Rivera-Duarte, I. 2004. Spatial and Temporal Variations in Copper Speciation in San Diego Bay. *Estuaries* 27: 437-447.

Borsuk, M. E., Higdon, D., Stow, C. A., and Reckhow, K. H. 2001. A Bayesian Hierarchical Model to Predict Benthic Oxygen Demand from Organic Matter Loading in Estuaries and Coastal Zones. *Ecological Modeling* 143: 165-181.

Brown, M. B. and Forsythe, A. B. 1974. Robust tests for the equaity of variances. *Journal of the American Statistical Association* 69: 364-387.

Burnham, K. P. and Anderson, D. R. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York.

Burrough, P. A. and McDonnell, R. A. 1998. *Principles of Geographical Information Systems*. Oxford University Press, New York.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., and Smith, V. H. 1998. Nonpoint Pollution of Surface Waters With Phosphorus and Nitrogen. *Ecological Applications* 8: 559-568.

Carpenter, S. R. 1990. Large-Scale Perturbations: Opportunities for Innovation. *Ecology* 71: 2038-2043.

Cohen, J. and Cohen, P. 1983. *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Comeleo, R. L., Paul, J. F., August, P. V., Copeland, J. L., Baker, C., Hale, S. S., and Latimer, R. W. 1996. Relationships Between Watershed Stressors and Sediment Contamination in Chesapeake Bay Estuaries. *Landscape Ecology* 11: 307-319.

Congalton, R. G. 1991. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment* 37: 35-46.

Congalton, R. G. 1997. Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion. *Photogrammetric Engineering and Remote Sensing* 63: 425-434.

Congalton, R. G. and Green, K. 1999. *Assessing the Accuracy of Remotely Sensed Data*. Lewis Publishers, New York.

daSilva, S. 2004. *A multiple scale approach to assessing the biological integrity of Rhode Island streams*. Ph.D. Thesis, University of Rhode Island: Kingston, RI.

Dauer, D. M., Ranasinghe, J. A., and Weisber, S. B. 2000. Relationships between benthic community condition, water quality, sediment quality, nutrient loads, and land use patterns in Chesapeake Bay. *Estuaries* 23: 80-96.

DeMers, M. N. 1997. *Fundamentals of Geographic Information Systems*. John Wiley and Sons, New York.

Duinker, J. C. and Nolting, R. F. 1978. Mixing, removal, and mobilizationof trace metals in the Rhine Estaury. *Netherlands Journal of Sea Research* 12: 205.

Dunn, C. P., Sharpe, D. M., Guntenspergen, G. R., Stearns, F., and Yang, Z. 1990. Methods for Analyzing Temporal Changes in Landscape Pattern. *In* Turner, M. G. and Gardner, R. H. (ed.), *Quantitative Methods in Landscape Ecology*, pp. 173-198. Springer-Verlag, New York.

Flather, C. H., Bradley, S. J., and Inkley, D. B. 1992. Regional Habitat Appraisals of Wildlife Communities: A Landscape-level Evaluation of a Resource Planning Model Using Avian Distribution Data. *Landscape Ecology* 7: 137-147.

Forman, R. T. and Godron, M. 1986. *Landscape Ecology*. John Wiley & Sons, New York.

Forman, R. T. T. 1995. *Land Mosaics: The Ecology of Landscapes and Regions*. Cambridge University Press , Cambridge.

Gergel, S. E., Turner, M. G., and Kratz, T. K. 1999a. Scale-dependent Landscape Effects on North Temperate Lakes and Rivers. *Ecological Applications* 9: 1377-1390.

Gergel, S. E., Turner, M. G., and Kratz, T. K. 1999b. Dissolved organic carbon as an indicator of the scale of watershed influence on lakes and rivers. *Ecological Applications* 9: 1377-1390.

Gustafson, E. J. 1998. Quantifying Landscape Spatial Pattern: What Is the State of the Art? *Ecosystems* 1: 143-156.

Hale, S. S., Paul, J. F., and Heltshe, J. F. 2004. Watershed Landscape Indicators of Estuarine Benthic Condition. *Estuaries* 27: 283-295.

Hollister, J. W., Gonzalez, L. M., Paul, J. F., August, P. V., and Copeland, J. 2004. Assessing the accuracy of the National Land Cover Dataset and Multiple Spatial Extents. *Photogrammetric Engineering and Remote Sensing* 70: 405-414.

Hunsaker, C. T. and Levine, D. A. 1995. Hierarchical Approaches to the Study of Water Quality in Rivers. *BioScience* 45: 193-203.

Jones, K. B., Riitters, K. H., Wickham, J. D., Tankersley, R. D., O'Neill, R. V., Chaloud, D. J., Smith, E. R., and Neale, A. C. 1997. *An Ecological Assessment of the United States Mid-Atlantic Region: A Landscape Atlas*. EPA/600/R-97/130. Office of Research and Development, U. S. Environmental Protection Agency, Washington, D. C.

Kadlec, R. H. K. J. A. 1979. Wetlands and Water Quality. *In* Gresson, P. E. C. J. R. a. C. J. E. (ed.), *Wetland Functions and Values: the State of Our Understanding*, pp. 436-456. American Water Resources Association, Proceedings of the National Symposium on Wetlands, Lake Buena Vista, Florida.

Karr, J. R. and Chu, E. W. 1999. *Restoring Life in Running Waters:Better Biological Monitoring*. Island Press, Washington, D. C.

Kennish, M. J. 1986. *Ecology of estuaries. Volume I: physical and chemical aspects*. CRC Press, Boca Raton, Florida.

Levin, S. A. 1992. The Problem of Pattern and Scale in Ecology. *Ecology* 73: 1943-1983.

Levine, D. A., Hunsaker, C. T., Timms, S. P., and Beauchamp, J. J. 1993. *Geographic Information System Approach to Modeling Nutrient and Sediment Transport*. Report 6736. Oak Ridge National Laboratory.

Likens, G. E., Borman, F. H., Johnson, N. M., Fisher, D. W., and Pierce, R. S. 1970. Effects of Forest Cutting and Herbicide Treatment on Nutrient Budgets in the Hubbard Brook Watershed-Ecosystem. *Ecological Monographs* 40: 23-47.

Lillesand, T. M. and Kiefer, R. W. 1995. *Remote Sensing and Image Interpretation*. John Wiley and Sons, New York.

Long, E. R., MacDonald, D. D., Smith, S. L., and Calder, F. D. 1995. Incidence of Adverse Biological Effect Within Ranges of Chemical Concentration in Marine and Estuarine Sediment. *Environmental Management* 19: 81-97.

Loveland, T. R., Gallant, A. L., and Vogelmann, J. E. 1999. Perspectives on the Use of Land Cover Data for Ecological Investigations. *In Issues in Landscape Ecology*, International Association for Landscape Ecology, Guelph, Ontario.

Lunetta, R. L., Lyon, J. G., Guidon, B., and Elvidge, C. D. 1998. North American Landscape Characterization:Data Set Development and Data Fusion Issues. *Photogrammetric Engineering and Remote Sensing* 64: 821-829.

Marsalek, J. 1978. *Pollution Due to Urban Runoff: Unit Loads and Abatement Measures*. Pollution from Land Use Activities Reference Group, International Joint Commission, Windsor, Ontario.

McGarigal, K. and Marks, B. J. 1995. *FRAGSTATS: spatial pattern analysis program for quantifying landscape structure* . General Technical Report PNW-GTR-351. USDA Forest Service, Pacific Northwest Research Station, Portland OR.

McGarigal, K. and McComb, W. C. 1995. Relationships Between Landscape Structure and Breeding Birds in the Oregon Coast Range. *Ecological Monographs* 65: 235-260.

Monbet, P. 2004. Seasonal Cycle and Mass Balance of Cadmium in an Estuary with an Agricultural Catchment: The Morlaix Tiver Estuary (Brittany, France). *Estuaries* 27: 448-459.

Morley, S. A. and Karr, J. R. 2002. Assessing and restoring the health of urban streams. *Conservation Biology* 16: 1498-1509.

Nester, M. R. 1996. An Applied Statistician's Creed. *Applied Statistics* 45: 401-410.

Novotny, V. and Olem, H. 1994. *Water Quality: Prevention, Identification, and Management of Diffuse Pollution*. Van Nostrand Reinhold, New York.

Nriagu, J. O. and Pacyna, J. M. 1988. Quantitative assessment of worldwide contamination of air, water, and soils by trace metals. *Nature* 333: 134-139.

O'Neill, R. V., Hunsaker, C. T., Timmins, S. P., Jackson, B. L., Jones, B. K., Riitters, K. H. , and Wickham, J. D. 1996. Scale Problems in Reporting Landscape Pattern at the Regional Scale. *Landscape Ecology* 11: 169-180.

Omernik, J. M., Abernathy, A. R., and Male, L. M. 1981. Stream Nutrient Levels and Proximity of Agricultural and Forest Land to Streams:Some Relationships. *Journal of Soil and Water Conservation* 36.

Overton, W. S., White, D., and Stevens, D. L. 1991. *Design Report of the Environmental Monitoring and Assessment Program* . EPA/600/3-91/053. US Environmental Protection Agency, Environmental Research Laboratory, Corvallis, OR.

Pacheco, P. A. 1993. *Point Source Methods Document for the Virginian Province*. National Oceanic and Atmospheric Administration, Silver Spring, MD.

Paul, J. F., Comeleo, R. L., and Copeland, J. L. 2002. Quantitative Relationships Between Landscape Metrics and Sediment Contamination for Small Estuarine Systems in the Mid-Atlantic and Southern New England Regions of the United States. *Journal of Environmental Quality* 31: 836-845.

Paul, J. F., Gentile, K. J., Scott, K. J., Schimmel, S. C., Campbell, D. E., and Latimer, R. W. 1999. *EMAP-Virginian Province Four Year Assessment (1990-93)*. EPA/620/R-99/004. US Environmental Protection Agency, Office of Research and Development, Washington, D. C.

Paul, J. F., Scott, K. J., Holland, A. F., Weisberg, S. B., Summers, J. K., and Robertson, A. 1992. The Estuarine Component of the US E. P. A. 's Environmental Monitoring and Assessment Program. *Chemistry and Ecology* 7: 93-116.

Peterjohn, W. T. and Correll, I. 1984. Nutrient Dynamics in an Agricultural Watershed: Observations on the Role of a Riparian Forest. *Ecology* 65: 1466-1475.

Reckhow, K. H. 1990. Bayesian Inference in Non-Replicated Ecological Studies. *Ecology* 71: 2053-2059.

Reckhow, K. H., Donigian Jr., A. S., Karr, J. R. , Mandrup-Poulsen, J., McDonald, H. S., Novotny, V., Smith, R. A., and Yoder, C. O. 2001. *Assessing the TMDL Approach to Water Quality Management*. National Academy Press, Washington, D. C.

Rodriguez, W. 2003. *Remote Sensing and Landscape Analysis of Indicators of Estuarine Condition in the Northeastern United States*. Ph.D. Thesis, University of Rhode Island: Kingston, RI.

SAB 1988. *Future risk: Research Strategiew for the 1990s. The report of the Research Strategies Committe*. Science Advisory Board, Washington, D.C.

Satarug, S., Baker, J. R., Urbenjapol, S., Haswell-Elkins, P., Reilly, E. B., Williams, D. J., and Moore, M. R. 2003. A golbal perspective on cadmium pollution and toxicity in non-occupationally exposed population. *Toxicity Letters* 137: 65-83.

Schroeder, H. A. 1974. *The Poisons Around Us: toxic metals in food, air, and water*. Indiana University Press, Bloomington, IN.

Sonzogni, W. C., Chesters, G., Coote, D. R., Jeffs, D. N., Konrad, J. C., Ostry, R. C., and Robinson, J. B. 1980. Pollution from Land Runoff. *Environmental Science and Technology* 14: 148-153.

Stehman, S. V. and Czaplewski, R. L. 1998. Design and Analysis for Thematic Map Accuracy Assessment: Fundamental Principles. *Remote Sensing of Environment* 64: 331-344.

Stehman, S. V., Wickham J. D.,Yang L., Smith J. H. 2003. Thematic accuracy of the 1992 National Land-Cover Data (NLCD) for the eastern United States: statistical methodology and regional results. *Remote Sensing of the Environment* 86: 500-516.

Turner, M. G. 1989. Landscape Ecology: The Effect of Pattern on Process. *Annual Review of Ecology and Systematics* 20: 171-197.

Turner, M. G., Gardner, R. H., and O'Neill, R. V. 2001. *Landscape Ecology In Theory and Practice*. Springer-Verlag, New York.

Turner, M. G., O'Neill, R. V., Gardner, R. G., and Milne, B. T. 1989. Effects of Changing Spatial Scale on the Analysis of Landscape Pattern. *Landscape Ecology* 3: 153-162.

United States Environmental Protection Agency 1990. *Urban targeting and BMP selection. An information and guidance manual for state NPS program staff*. EPA 660/3-74-02C. U.S. EPA Office of Water, Washington, D.C.

United States Environmental Protection Agency 1995. *National Water Quality Inventory: 1994 Report to Congress*. 841-R-95-005. USEPA, Washington, D. C.

United States Environmental Protection Agency 1999. *EMAP-Virginian Province Four-Year Assessment (1990-1993)*. EPA/620/R-99/004. United States Environmental Protection Agency, Washington, D. C.

United States Geological Survey. 2003. *Accuracy Assessment of 1992 National Land Cover Data*. Retrieved on 12/12/2004 from http://landcover.usgs.gov/accuracy/.

Urban, D. L. and Shugart, H. H. 1987. Landscape Ecology. *Bioscience* 37.

Vogelmann, J. E., Sohl, T., and Howard, S. M. 1998a. Regional Characterization of Land Cover Using Multiple Sources of Data. *Photogrammetric Engineering and Remote Sensing* 64: 45-57.

Vogelmann, J. E., Sohl, T., Howard, S. M., and Shaw, D. M. 1998b. Regional Land Cover Characterization Using Landsat Thematic Mapper Data and Ancillary Data. *Environmental Monitoring and Assessment* 51: 415-428.

Vogelmann, J. E. and Wickham, J. D. 2000. *Implementation Strategy for Production of National Land Cover Data (NLCD) from the Landsat 7 Thematic Mapper Satellite*. EPA 600/R-00/051. Office of Research and Development, U. S. Environmental Protection Agency, Washington, D. C.

Vogelmann, J. E., Howard, S. M., Yang, L., Larson, C. R., Wylie, B. K., and Van Driel, N. 2001. Completion of the 1990's National Land Cover Data Set for the Conterminous United States from Landsat Thematic Mapper Data and Ancillary Data Sources. *Photogrammetric Engineering and Remote Sensing* 67: 650-662.

Wang, L., Lyons, J., Kanehl, P., and Gatti, R. 1997. Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. *Fisheries* 22: 6-12.

Weller, D. E., Jordan, T. E., and Correll, D. L. 1998. Heuristic models for material discharge from landscapes with riparian buffers. *Ecological Applications* 8: 1156-1169.

Weller, D. E., Jordan, T. E., Correll, D. L., and Liu, Z.-J. 2003. Effects of land-use change on Nutrient Discharges from the Patuxent River Watershed . *Estuaries* 26: 244-266.

Wickham, J. D. and Riitters, K. H. 1995. Sensitivity of Landscape Metrics to Pixel Size. *International Journal of Remote Sensing* 16: 3585-3594.

Wiens, J. A. 1989. Spatial scaling in ecology. *Functional Ecology* 3.

Yang, L., Stehman, S. V., Smith, J. H., and Wickham, J. D. 2001. Thematic Accuracy of MRLC Land Cover for the Eastern United States. *Remote Sensing of Environment* 76: 418-422.

Zar, J. H. 1999. *Biostatistical Analysis*. Prentice Hall, New Jersey.