TOWSON UNIVERSITY
COLLEGE OF GRADUATE STUDIES AND RESEARCH

PREDICTING THE EFFECTS OF LAND COVER AND
STORMWATER MANAGEMENT ON STREAM HEALTH
IN BALTIMORE COUNTY, MARYLAND

by

Robert H. Hirsch, IV

A thesis

Presented to the faculty of

Towson University

in partial fulfillment

of the requirements for the degree

Master of Arts

August, 2011

Towson University
Towson, Maryland 21252
TOWSON UNIVERSITY
COLLEGE OF GRADUATE STUDIES AND RESEARCH

THESIS APPROVAL PAGE

This is to certify that the thesis prepared by Robert H. Hirsch, IV, entitled PREDICTING THE EFFECTS OF LAND COVER AND STORMWATER MANAGEMENT ON STREAM HEALTH IN BALTIMORE COUNTY, MARYLAND has been approved by the thesis committee as satisfactorily completing the thesis requirements for the degree Master of Arts.

Dr. Kang Shou Lu, Chair, Thesis Committee

Dr. Kent Barnes, Committee Member

Dr. Martin Roberge, Committee Member

7/29/2011 Date

7/29/2011 Date

7/29/2011 Date

Associate Dean, College of Graduate Studies and Research

29 July 2011 Date
Acknowledgements

This work would not have been possible without the contributions of many. First, I must thank Steven Stewart and Dennis Genito at the Baltimore County Department of Environmental Protection and Sustainability for sharing their stream health data, discussing their environmental management needs, and encouraging this research. I am indebted to the creators and stewards of the geographic and administrative data used in this research, including Karen Coffman at Maryland State Highways Administration, Hugh Murphy at Carroll County, Fred Irani at the USGS Chesapeake Bay Program Office, Jarlath O’Neil-Dunne at University of Vermont, and my coworkers at Baltimore County in the Office of Information Technology and the Department of Environmental Protection and Sustainability.

My instructors and advisors at Towson University introduced me to many of the concepts applied in this research. In particular, I wish to thank Dr. Martin Roberge and Dr. Susan Gresens for their thoughtful comments, and Dr. Kent Barnes for his last-minute heroism. My many discussions with Dr. Kang Shou Lu were invaluable. Dr. Lu challenged me to explore new methods and polish my research writing.

Finally, I would like to dedicate this work to my wife, Elaina. I am grateful for her patience, proof reading, and project management. I hope this was worth the wait.
ABSTRACT

PREDICTING THE EFFECTS OF LAND COVER AND STORMWATER MANAGEMENT ON STREAM HEALTH IN BALTIMORE COUNTY, MARYLAND

Robert H. Hirsch, IV

Maintaining stream health in urban and agricultural riverscapes is a challenge for environmental managers, and requires a better understanding of stream ecology and biogeography. This study develops a spatial generalized least squares model for predicting the effects of land cover and stormwater management (SWM) on Piedmont stream health in Baltimore County, Maryland. Such models are needed to better integrate stream health management with local land use planning and SWM programs. The model predicts stream health impairment with >80% accuracy. Model performance is suitable for supporting a precautionary policy for preventing water quality impairment. Results show how SWM reshapes catchments, and interacts with stream health through coupled human-environment systems. Riparian buffers had little impact on stream health, while large effects were found for impervious surfaces and forest cover. Continuing development of SWM data resources and refinements to the model promise new knowledge and more nuanced and precise predictions.
TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. viii
LIST OF FIGURES ............................................................................................................... ix
KEY TO ABBREVIATIONS ................................................................................................. xi
CHAPTER 1: INTRODUCTION .............................................................................................. 1
  Problem Statement ............................................................................................................. 1
  Context and Motivation .................................................................................................... 3
  Goals and Objectives ....................................................................................................... 5
CHAPTER 2: LITERATURE REVIEW ...................................................................................... 6
  Understanding Stream Health ......................................................................................... 6
  Measuring Stream Health ............................................................................................... 9
  Predicting Stream Health .............................................................................................. 12
    Collinearity .................................................................................................................. 12
    Independent Samples and Autocorrelation .................................................................. 13
    Variography ................................................................................................................ 16
    Regression Kriging ...................................................................................................... 17
    Variography and Residual Drift .................................................................................. 18
    Variable Selection and Information Criteria ............................................................. 19
    Stationarity .................................................................................................................. 21
    Categorical Predictions .............................................................................................. 22
CHAPTER 3: METHODS ....................................................................................................... 23
  Study Area ...................................................................................................................... 23
  Sampling Design ............................................................................................................ 24
    Physiographic Province ............................................................................................... 25
    Completeness of SWM and Tree Canopy Data ......................................................... 26
    Division of Data into Fitting and Evaluation Sets ....................................................... 26
  Geographic Units: Catchments, Riparian Buffers, And Reaches ................................. 27
  Distribution of BIBI ....................................................................................................... 29
  Measures of Stormwater Management ......................................................................... 31
    Description of Stormwater Management Facilities ................................................... 32
    Stormwater Management and Site Catchments ......................................................... 35
  Measures of Land Cover ............................................................................................... 36
  Statistical Methods ....................................................................................................... 38
    Specification of the GLS Model .................................................................................... 41
      Heteroscedasticity ...................................................................................................... 41
      Spatial Autocorrelation and Drift ............................................................................. 41
      Predictor Variable Selection ..................................................................................... 42
    Validation of Model Assumptions .............................................................................. 42
Evaluation of Model Performance ..................................................43
CHAPTER 4: RESULTS ........................................................................47
  Description of Data .......................................................................47
  Effect of Stormwater Management on Catchments .........................50
  Multicollinearity in Land Cover ....................................................50
  Specifying the Error Terms ..........................................................53
  Specifying the Fixed Terms ..........................................................59
  Validating Model Assumptions ......................................................64
  Estimates of the Model Parameters ..............................................71
    Fixed Terms .............................................................................71
    Random Terms .........................................................................74
  Model Performance ......................................................................76
    Predictions of BIBI ...................................................................76
    Predicting Regulatory Categories Using Naïve Cutoffs ....................78
    Predicting Regulatory Categories Using ROC ..............................79
CHAPTER 5: DISCUSSION .................................................................84
  Effects of Land Cover on Stream Health .......................................84
    Impervious Surfaces Land Cover ................................................84
    Agricultural Land Cover ............................................................85
    Forest Land Cover ..................................................................86
    Tree Canopy Cover ..................................................................86
    Riparian Canopy Cover .............................................................87
    Reach Canopy Cover ................................................................87
  Effects of Stormwater Management on Stream Health ...................88
  Suitability of Model for Applied Use ..........................................91
    Continuous BIBI Predictions .....................................................91
      Model Accuracy .....................................................................91
      Model Precision ....................................................................91
    Regulatory Category Predictions ..............................................92
  Limitations of Final Model .........................................................93
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS ......................97
  Land Cover and Stream Health ....................................................97
  Stormwater Management and Stream Health ...............................98
  Predictive Modeling and Decision Making ....................................99
  Further Work ..............................................................................100
  Summary ...................................................................................103
APPENDICES ..................................................................................105
  Appendix A: Maps of Variables ...............................................106
  Appendix B: Additional Bivariate Scatterplots ...............................112
  Appendix C: Partial Residual Plots of Final Model .........................114
  Appendix D: Confusion Matrices for Categorical Predictions ............115
LITERATURE CITED ........................................................................116
CURRICULUM VITA .........................................................................127
LIST OF TABLES

Table 1. Description of variables used ........................................................................................................29
Table 2. Distribution of BIBI by regulatory category and sample .................................................................30
Table 3. Sequence of model fitting rounds ........................................................................................................39
Table 4. Predictor variables used in saturated model .........................................................................................40
Table 5. Distributions of variables by sample ....................................................................................................48
Table 6. Bivariate correlations ..........................................................................................................................51
Table 7. Sequential model fitting rounds ...........................................................................................................53
Table 8. AICc comparison of several variance-covariance functions in the saturated model ............................54
Table 9. AICc comparison of several theoretical variograms on the saturated model ........................................57
Table 10. AICc comparison of models with imperviousness ..............................................................................59
Table 11A. AICc comparison of models with agriculture ..................................................................................59
Table 11B. AICc comparison of models with forest ..........................................................................................60
Table 11C. AICc comparison of models with canopy .......................................................................................60
Table 11D. AICc comparison of models with land covers .................................................................................60
Table 12. AICc comparison of models with SWM ............................................................................................61
Table 13. AICc comparison of models with riparian canopy ............................................................................61
Table 14. AICc comparison of models with reach canopy .................................................................................62
Table 15. AICc comparison of best models from each variable selection round ................................................62
Table 16. AICc comparison of best agriculture models .......................................................................................63
Table 17. AICc comparison of variance-covariance functions for residuals ......................................................68
Table 18. AICc comparison of variance-covariance functions for prediction errors .........................................69
Table 19. Effects of watershed on variance of residuals ...................................................................................69
Table 20. Effects of watershed on variance of prediction errors .....................................................................69
Table 21. AICc comparison of models predicting standardized residuals .........................................................69
Table 22. Effects of year on raw residuals .........................................................................................................70
Table 23. AICc comparison of models predicting RK errors .............................................................................70
Table 24. Effects of watershed on RK errors ..................................................................................................70
Table 25. Effects of individual variables on BIBI: canopy model .................................................................71
Table 26. Effects of individual variables on BIBI: agriculture model ..............................................................73
Table 27. Pseudo $R^2$ of final model .................................................................................................................77
Table 28. Accuracy and precision of model predictions ....................................................................................77
Table 29. Classification accuracy for 3-category predictions using naïve cutoffs .............................................78
Table 30. Classification accuracy for predicting impaired and passing stream health, using naïve cutoffs ..........................79
Table 31. Error rates observed when ROC-informed cutoffs are used to predict impaired stream health ..................81
Table 32. Error rates observed when ROC-informed cutoffs are used to predict high quality stream health ..........83
LIST OF FIGURES

Figure 1. Study location maps.........................................................................................................24
Figure 2. Physiographic provinces and watersheds ...........................................................................25
Figure 3. Geographic units used in this study ....................................................................................27
Figure 4. Physiographic provinces and watersheds ...........................................................................25
Figure 5. Numerical distribution of BIBI ..........................................................................................29
Figure 6. Geographic distribution of BIBI..........................................................................................31
Figure 7. Geographic distribution of dry detention pond and wet pond SWM facilities ..................34
Figure 8. Geographic distribution of agricultural and tree canopy land covers ..............................37
Figure 9. Moran’s spatial correlograms of BIBI and predictor variables ...........................................49
Figure 10. Changes in catchment areas due to SWM .......................................................................50
Figure 11. Bivariate scatterplots for BIBI and catchment land cover variables ...............................52
Figure 12. Scatterplots illustrating relationship of variance of saturated model residuals with Fitted Values and Impervious .................................................................54
Figure 13. Scatterplots illustrating homoscedasticity after adding the variance covariate function to the saturated model .................................................................55
Figure 14. Scatterplots illustrating spatial trends in saturated model residuals ..............................55
Figure 15. Correlogram and variogram illustrating spatial autocorrelation of saturated model residuals ............................................................................................................56
Figure 16. Variograms from saturated GLS and GLS-DT models ..................................................57
Figure 17. Correlogram and variogram illustrating diminished spatial autocorrelation in the saturated GLS-DT model residuals .................................................................58
Figure 18. Validation plots of residuals from the saturated GLS-DT model ......................................58
Figure 19. Distribution of final model residuals and errors ...............................................................64
Figure 20. Patterns in final model residuals and errors .................................................................65
Figure 21. Global spatial trends in residuals and errors of final model .........................................66
Figure 22. Autocorrelation in residuals and errors of final model ...................................................67
Figure 23. LISA maps illustrating the clustering of residuals and errors .......................................68
Figure 24. Scatterplots illustrating the estimated variance covariate function ..............................74
Figure 25. Semi-variogram of model residuals, with the estimated variogram model .................75
Figure 26. Filled-contour maps of kriged residuals and kriging variance .......................................76
Figure 27. Receiver operating characteristic curves for predicting impairment ..........................80
Figure 28. Receiver operating characteristic curves for predicting high quality .........................82
### KEY TO ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICc</td>
<td>Akaike information criterion, small sample size adjusted</td>
</tr>
<tr>
<td>AUC</td>
<td>area under curve</td>
</tr>
<tr>
<td>BIBI</td>
<td>benthic index of biotic integrity</td>
</tr>
<tr>
<td>BMP</td>
<td>best management practice</td>
</tr>
<tr>
<td>DDP</td>
<td>dry detention pond stormwater management facility</td>
</tr>
<tr>
<td>EDP</td>
<td>extended detention pond stormwater management facility</td>
</tr>
<tr>
<td>F</td>
<td>filtration stormwater management facility</td>
</tr>
<tr>
<td>GIS</td>
<td>geographic information systems</td>
</tr>
<tr>
<td>GLS</td>
<td>generalized least squares regression</td>
</tr>
<tr>
<td>GLS-DT</td>
<td>spatial generalized least squares regression with detrending</td>
</tr>
<tr>
<td>I</td>
<td>infiltration stormwater management facility</td>
</tr>
<tr>
<td>IBI</td>
<td>index of biotic integrity</td>
</tr>
<tr>
<td>LiDAR</td>
<td>light detection and ranging</td>
</tr>
<tr>
<td>LISA</td>
<td>local indicators of spatial association</td>
</tr>
<tr>
<td>LOESS</td>
<td>locally weighted scatterplot smoothing</td>
</tr>
<tr>
<td>MAE</td>
<td>mean absolute error</td>
</tr>
<tr>
<td>MBSS</td>
<td>Maryland Biological Stream Survey</td>
</tr>
<tr>
<td>MDE</td>
<td>Maryland Department of the Environment</td>
</tr>
<tr>
<td>MDNR</td>
<td>Maryland Department of Natural Resources</td>
</tr>
<tr>
<td>MDP</td>
<td>Maryland Department of Planning</td>
</tr>
<tr>
<td>NLCD</td>
<td>National Land Cover Dataset</td>
</tr>
<tr>
<td>NPDES</td>
<td>National Pollutant Discharge Elimination System</td>
</tr>
<tr>
<td>OLS</td>
<td>ordinary least squares regression</td>
</tr>
<tr>
<td>RK</td>
<td>regression kriging</td>
</tr>
<tr>
<td>RMSE</td>
<td>root mean square error</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operating characteristics</td>
</tr>
<tr>
<td>SWM</td>
<td>stormwater management</td>
</tr>
<tr>
<td>TMDL</td>
<td>total maximum daily load</td>
</tr>
<tr>
<td>UTC</td>
<td>urban tree canopy land cover data</td>
</tr>
<tr>
<td>VIF</td>
<td>variance inflation factor</td>
</tr>
<tr>
<td>WP</td>
<td>wet pond stormwater management facility</td>
</tr>
<tr>
<td>WRE</td>
<td>Water Resources Element of the Master Plan</td>
</tr>
<tr>
<td>Ŷ</td>
<td>fitted values</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

Stream health is impaired and threatened throughout the United States, despite the existence of legal and regulatory frameworks designed to protect water quality. Environmental management and planning tools cannot be used effectively without a good understanding of the geography and ecology of streams. The goal of this study is to enable better management decisions by understanding the effects of land cover and stormwater management (SWM) on stream health. I will accomplish this goal by developing a statistical model for predicting stream health that is accurate enough to support decision making in land use planning, development regulation, and other areas of environmental management.

Problem Statement

Maintaining stream health is both a challenging problem for environmental managers, and an opportunity to apply and advance knowledge of ecology and human-environment interactions.

The United States Environmental Protection Agency (EPA) has estimated that 41.9% of streams in the lower 48 states are in poor health (US EPA 2006). Most of these streams are impaired by non-point source pollution, a pervasive form of pollution that the US Clean Water Act has historically failed to redress (Andreen 2004). Recently, enforcement of the Clean Water Act has been escalated (e.g., the Chesapeake Bay Total Maximum Daily Load (Copeland 2010)), and new tools for protecting water resources
have been introduced into local land use planning (e.g., the Water Resources Element (WRE) (Maryland Department of Planning (MDP), Maryland Department of Environment (MDE), Maryland Department of Natural Resources (MDNR) 2007)). The success of these efforts depends on predictions of how planning, restoration, and other actions will impact stream health. Unfortunately, models for predicting impacts to biological stream health are not widely available.

An incomplete understanding of how land cover and stormwater management (SWM) impact streams impedes developing and applying the predictive models that would support stream health management (Allan 2004). Complex causal mechanisms contribute to this incomplete understanding (e.g., Moglen et al. 2004; Novotny et al. 2005; King et al. 2005). Streams exist within riverscapes (Ward 1998), which are complex and multi-scalar social-ecological systems (Redman, Grove, and Kuby 2004), and are the locus of coupled human-environment interactions (Groffman et al. 2003; Walsh et al. 2005). Social processes such as land cover change and stormwater management (SWM) implementation alter stream health. In turn, stream health alters these same social processes through environmental governance (e.g., environmental law, land use planning, and land development regulations). Therefore, understanding stream health requires knowledge of local environmental governance and regional stream ecology.

Another challenge for understanding and predicting stream health is the heterogeneity inherent among riverscapes, in terms of their biogeophysical context (Frimpong et al. 2005; Roy et al. 2006), stormwater and agricultural best management practices, and environmental management interventions (Moore and Palmer 2005).
Heterogeneity makes it difficult to generalize across regions (e.g., Utz, Hilderbrand, and Boward 2009), and requires regional data collection and model development efforts.

Additionally, the predictor variables that can be used in predictive models of stream health are constrained by limitations in the planning and environmental management settings. For example, the time and resources required to obtain data in the field may make use of certain predictor variables impractical.

A final difficulty is the challenge of developing statistically valid models. Ecological and geographic data tend to exhibit spatial trends, collinearity, non-linearity, regional variation in relationships, and spatial and temporal autocorrelation (Legendre 1993; King et al. 2005; Utz, Hilderbrand, and Boward 2009; Dodds et al. 2010). Statistical methods that can accommodate these properties are available (e.g., reviews in Dormann et al. 2007; Beale et al. 2010). However, even a valid model may include features that are difficult to use in applied settings (e.g., nonparametric smoothers, local or regional error adjustments, local or regional regression coefficients, etc.)

**Context and Motivation**

The United States’ Clean Water Act of 1972 requires that surface waters maintain their chemical, physical and biological integrity (Novotny et al. 2005). Current estimates fault nonpoint pollution for the majority of water quality problems in the US. Determining the sources of water quality impairment (loss of integrity or quality) and finding remedies for nonpoint pollution sources are the major activities under the Clean Water Act (Karr 1991; Novotny et al. 2005).

Reactive approaches to remedying water quality degradation focus on taking corrective actions (enforcement of regulations, execution of restoration projects) when
degradation is made apparent by affected stakeholders or through monitoring programs. Proactive approaches to water quality management seek to prevent degradation before it occurs. Proactive methods include conservation planning practices, such as zoning, conservation easements and land acquisition, and development regulations like stream buffer setbacks and stormwater management requirements. Predictive models are often used to inform proactive rational planning and defend decisions made during implementation.

In Maryland, a new tool for proactive water resource management is being implemented in the master planning process. With the passing of Maryland House Bill 1141 (2006 Maryland general assembly session), water resource elements (WRE) are now required in all county and municipal master plans. WREs must anticipate the impacts of growth and land cover changes on water resources, project the impacts out 25 years (15 years beyond the typical 10-year master planning horizon), and pass review by MDE staff. Thus, the WRE transforms the master planning process into a tool for proactive and rational stream health management.

A guidance document prepared by state agencies (MDP, MDE, MDNR 2007) outlines the components and analytic methods expected in the WREs. The document steers counties and municipalities towards a proactive, rational planning approach by focusing on prediction of water quality status and stressors. The current WRE guidance emphasizes chemical water quality, in particular the nitrogen and phosphate reduction goals of the Chesapeake Bay program. The guidance document advises planners to anticipate the effect of growth and Master Plan components on compliance with Total Maximum Daily Loads (TMDLs) set for impaired water bodies. It does not focus
attention on stream health, despite the fact that biological impairment listings are an integral part of Maryland’s clean water act reporting and implementation strategies (MDE 2008). This is likely due to the fact that MDE and MDNR are still developing mechanisms for setting TMDLs to remedy biological impairment (MDE 2009a). However, the WRE legislation and guidance do not prevent counties and municipalities from including biological or physical integrity in their plans.

One challenge to using the WRE as a tool for stream health management is a lack of effective tools for predicting the impacts of development and water management actions on stream health. Sophisticated tools for predicting the nutrient impacts of land cover change, SWM, and agricultural best management practices (BMPs) have been created for and adopted by planners and environmental managers (e.g., Smith, Schwarz, and Alexander 1997; Borah and Bera 2004; Simpson and Weammert 2009). Similar predictive tools for stream health are not as widely available or utilized by planners and environmental managers (Horn et al. 2004).

**Goals and Objectives**

The goal of this study is to enable better management decisions by understanding the effects of land cover and SWM on stream health. I will accomplish this goal by developing a statistical model for predicting stream health using land cover and SWM variables. Land cover and SWM are important riverscape elements that planners, developers, and environmental managers can alter to affect stream health outcomes. Will this model be useful for environmental management, planning, and development regulation? Success will require model predictions that are accurate and consistent across space, time, and region.
CHAPTER 2
LITERATURE REVIEW

Successful stream health management depends on knowledge of stream ecology and geography, and statistical tools capable of modeling stream health outcomes. This chapter summarizes current knowledge of the effects of land cover and stormwater management (SWM) on stream health. Statistical modeling issues are discussed, and the powerful methods of spatial generalized least squares (GLS) and regression kriging (RK) are introduced.

Understanding Stream Health

Stream health is a complex, multi-dimensional construct. The US Clean Water Act requires protection and restoration of the physical, chemical, and biological integrity of streams (Novotny et al. 2005). Since J. R. Karr focused attention on biological integrity (Karr 1991), a large body of literature has developed around the effects of land cover and SWM on biological stream health. Several review articles have been published (e.g., Paul and Meyer 2001; Snyder et al. 2003; Allan 2004; Novotny et al. 2005). These authors have shown that stream health emerges as a property of riverscapes. Riverscapes are a series of nested and hierarchical landscapes. Riverscape levels range from the watershed and drainage network, down to channel geomorphology and microhabitats (Ward 1998). Levels are physically coupled: processes at one level drive processes at others, so that examining the properties of each level individually will not capture the functioning of the whole stream. Riverscapes must be viewed holistically as social-
ecological systems (Redman, Grove, and Kuby 2004): physical, biological, and social processes interact to create the unique form and function of each riverscape (Allan 2004). Therefore, understanding stream health requires the consideration of physical, biological, and social processes operating at the multiple scales of riverscapes.

Furthermore, idiosyncrasies of each unique riverscape interact with land cover change and SWM practices, causing relationships between land cover, SWM, and stream biological health to vary among riverscapes. For example, the impact of agriculture on stream health is highly variable between studies: Genito, Gburek, and Sharpley (2002) found a negative correlation, Moore and Palmer (2005) found a positive correlation, while others (Maryland Department of the Environment (MDE) 2009a; Utz, Hilderbrand, and Boward 2009) found no useful correlation between agriculture and stream dwelling invertebrates in Maryland’s Piedmont region. Allan’s 2004 literature review identified variation in stream response to agriculture due to thresholds, diversity of agricultural uses (e.g., pasture versus cropland), and variation in management practices.

In contrast to agricultural land cover, urban land cover reliably shows negative correlations to stream health. Urban land affects streams by diminishing infiltration rates, altering stream flow regimes and increasing stream flashiness. Additional urban impacts include increased quantity and diversity of chemical pollutants, loss of riparian vegetation, sediment inputs, bank destabilization, channelization, and restriction of riparian and hyporheic zones.

It is clear that the intensity of urban development (amount of impervious surface, industrial and stormwater discharges to the streams, etc.) is critical to understanding how urban land covers influence stream water quality (e.g., Cuffney et al. 2005). Urban
impacts to streams include hydrologic alteration, sedimentation, toxic chemicals, nutrient enrichment, and temperature perturbations (Allan 2004; Walsh et al. 2005). The relative importance of these impacts appears to depend upon intensity of urban development and other contextual information. For example, Gresens et al. (2007) found evidence that toxic chemicals and nutrients in stormwater runoff have a larger effect on macroinvertebrates than physical disturbance due to hydrologic alterations. Allan (2004) and Walton et al. (2007) note that physical habitat changes due to urbanization do not appear as important to stream health as chemical or hydrologic alterations, possibly due to past degradation from agriculture, habitat protection by parks and greenways, or uniform habitat degradation in single studies.

Land cover and stream health often exhibit non-linear relationships. Many studies find thresholds of biological response to impervious surface and urban land cover thresholds at about 10% impervious surfaces in the catchment area (Allan 2004; King and Baker 2010). However, there is disagreement as to a global threshold for impervious surfaces and urban land covers, and due to interactions among stressors (Novotny et al. 2005), geographic variation, and diversity of stream health metrics, there probably is no global threshold (Allan 2004). In addition, measurement of both impervious surfaces and urban land can be variable (Booth and Jackson 1997, McMahon and Cuffney 2000) and measurement method has a documented effect on observed relationships (King et al. 2005).

SWM facilities are important hydrologic features in urbanized areas, affecting hydrology (Booth and Jackson 1997), nutrients and sediment (Simpson and Weammert 2009), and toxics (Pitt et al. 1995; Pitt, Clark, and Field 1999). Therefore, their presence
should change the effects of land cover on stream health. However, empirical studies of this hypothesis are rare (Walsh et al. 2005; Schueler et al. 2009), and results are likely to vary among jurisdictions, due to heterogeneity of biogeophysical characteristics, development regulations, and SWM design and maintenance. SWM often directs stormflows through pipes, revising the physical borders of catchments and watersheds. When SWM is present, the borders of catchments delineated from surface elevations should be revised to match the new borders imposed by SWM drainage systems.

Researchers have studied the relative importance of different levels of the riverscape hierarchy in affecting stream health. The evidence indicates that variation in one level of the hierarchy alters the stream health impact of other levels of the hierarchy (Urban et al. 2006). Open questions remain, for example, the relative effects of catchment and riparian land cover on water quality (Sponseller, Benfield, and Valett 2001; Moore and Palmer 2005; Snyder, Goetz, and Wright 2005; Walsh et al. 2005; Baker, Weller, and Jordan 2006; Roy et al. 2006; Urban et al. 2006; Walsh et al. 2007).

**Measuring Stream Health**

Stream health is a complex construct with many dimensions. Thus, stream health cannot be directly measured. Instead, indicators representing one or more dimensions are used. Indices of Biotic Integrity (IBIs) reflect a range of water quality conditions, and are thus suitable indicators of stream health (Karr 1999). Macroinvertebrate communities act as long-term water quality indicators, as the presence or absence of particular groups of species is attributable to water quality variables (e.g., Tullos et al. 2006). Macroinvertebrates are particularly valuable as indicators because their responses to a pollutant register only if the pollution level is high enough to impair the biology of the
stream and the function of its ecosystem (Bonada et al. 2006). Macroinvertebrates also play important roles in stream ecosystems and are appropriate indicators of stream health (Karr 1999; Barbour et al. 2000; Bonada et al. 2006). IBIs have been adopted for environmental management and governance in many locations (e.g., Yoder and Rankin 1998; US EPA 2006).

In Maryland, IBIs have been developed for the purpose of assessing biological integrity and listing biological impairments as required under the Clean Water Act. Benthic Index of Biotic Integrity (BIBI) is a quantitative multi-metric indicator of biological water quality, calculated using field sampling of stream-dwelling macroinvertebrate communities. The BIBI, developed for use throughout Maryland, summarizes overall stream water and habitat quality (Stribling et al. 1998; Southerland et al. 2005). It is a continuous variable, and ranges have been assigned to categories of stream health, using the following cutoffs: BIBI ≥ 4 indicates high quality stream health; 4 > BIBI ≥ 3 indicates fair stream health; BIBI < 3 indicates poor or impaired stream health (Southerland et al. 2005). These categories are helpful for communicating assessment results, and are the basis for clean water act implementation. BIBI < 3 is used to identify non-tidal streams that are biologically impaired and require restoration. BIBI ≥ 4 identifies streams that are high quality (“Tier 2” waters) and require special protections (MDE 2008).

There are limitations to using IBIs and other synthetic variables to understand stream health. As synthetic variables, a change in the variable cannot be tied directly to conditions in the stream. The change could indicate one or more changed conditions. The magnitude of change is also difficult to interpret. However, BIBI is used by MDE to
make regulatory decisions that structure local government land use planning and require financial outlays for restoration efforts. Therefore, changes in BIBI have very real impacts on local governments that are simple to interpret. Predicting changes in BIBI, even incremental changes that do not cross a regulatory threshold, is important for local governments and environmental regulators.

From a statistical perspective, the distribution of synthetic variables may not be well defined. In the case of the Maryland BIBI, scores have a substantial range of variation, but are ultimately ordinal and not truly continuous variables. Observations of BIBI are unlikely to be normally distributed: the range of variation is truncated at 1 and 5. This should create negative excess kurtosis, meaning that extreme values are less common and values near the mean are more common than expected for a normal distribution. These statistical issues may adversely affect statistical models of BIBI that assume a normal distribution. First, errors are likely to be biased towards the mean and away from the limiting values of 1 and 5. Such errors may affect suitability of the model for predictions. Second, as predictions move away from the mean towards the limiting values, error variance should decrease. This heteroscedasticity can bias estimates of statistical significance and confidence intervals of regression parameters, leading to erroneous conclusions.

Modeling BIBI as a set of regulatory categories is one response, but ordinal and logistic models offer less power than continuous regression models. Despite the distribution issues discussed above, others have modeled BIBI as a continuous variable (e.g., Morely and Karr 2002; Goetz and Fiske 2008).

Predicting Stream Health
Deterministic approaches, like those used in hydrologic simulation modeling, are unlikely to succeed due to the aforementioned complexity of linkages between stressors and biota (Horn et al. 2004). Statistical or probabilistic approaches incorporating mechanistic relationships are appealing due to their transferability and ecological logic (Poff 1997; Moglen et al. 2004; Novotny et al. 2005). Several of the metrics used to calculate Benethic Index of Biotic Integrity (BIBI; see methods section) have mechanistic relationships to catchment and riparian land cover, for example trophic groups and habit groups should respond to riparian land cover, while tolerance groups should respond to catchment urban land covers (Poff 1997). Parametric multiple regression is ideal for an applied predictive model, in that the method is widely known and linear coefficients are easy to use for making predictions and explaining decisions.

Ordinary least squares (OLS) is a familiar form of regression, but its assumptions are too strict for most geographical analyses. These assumptions include residuals distributed normally with constant mean and variance, independence of residuals, spatial and temporal stationarity, and independence of predictor variables.

Collinearity

Land cover variables are usually collinear and violate the assumption of independence of predictor variables (Graham 2003; Allan 2004; King et al. 2005). Land cover is usually measured as a proportion of an area. Because the total proportion is always 1, land cover variables are never completely independent: as one land cover increases, the sum of the others must decrease. The more land cover variables are included in an analysis, the stronger this multicollinearity becomes. Collinearity confounds correlative analyses: the observed correlations may be due to true causal
relationships between the variables, or due to spurious correlations (Graham 2003). Ignoring multicollinearity during multiple regression analysis can result in model misspecification and reduced statistical power (Graham 2003).

The easiest way to avoid the problems of collinearity is to measure it and filter out models that include collinear combinations of variables (Graham 2003). Detecting collinearity is usually accomplished through bivariate scatterplots and correlation matrices. A useful measure of collinearity is the variance inflation factor (VIF: Fox and Monette 1992). Combinations of variables may be filtered if VIF exceeds a particular threshold. Zuur, Ieno, and Elphick (2010) recommend using a VIF threshold as low as 2 in studies where relationships between the dependent and predictor variables are very noisy: in these situations, variance of regression parameters are already quite large, and inflation of variance is likely to make important parameters appear to be insignificant. Note however that eliminating one collinear variable may inflate the estimated effect of the retained variables (O’Brien 2007). Therefore, if variables are eliminated from a model due to collinearity, changes to effect sizes should be examined, as for example in King et al. 2005.

Independent Samples and Autocorrelation

Geographical studies must contend with autocorrelation, or the dependence of observations at one location on observations at nearby locations. This is a common property of ecological data (Sokal and Oden 1978; Legendre 1993; Legendre and Legendre 1998; Dormann et al. 2007). Autocorrelation in stream biota may be generated by the spatial pattern of habitat niches (environmental control), or the biological and hydraulic processes of reproduction and dispersal (neutral theory), or both (e.g.,
Thompson and Townsend 2006). Statistically, autocorrelation violates the assumption of independent observations, and therefore skews the results of statistical tests and biases regression coefficients (Legendre and Legendre 1998). Other studies of stream health have used a variety of correlative methods that account for autocorrelation. These include Mantel tests (Legendre and Fortin 1989; e.g., King et al. 2005), eigenfunction methods (Griffith and Peres-Neto 2006) such as principal coordinates of neighbor matrices (Borcard and Legendre 2002; e.g., Urban et al. 2006; Blanchet, Legendre, and Borcard 2008), and spatial generalized least squares (GLS: Pinheiro et al. 2010; e.g., Peterson et al. 2006).

Of the many methods for estimating regression equations using autocorrelated ecological data, GLS appears to be particularly robust and accurate (Dormann et al. 2007; Beale et al. 2010). Spatial GLS models are also easily used for predictions (Dormann et al. 2007), through geostatistical methods such as regression-kriging (RK: Odeh, McBratney, and Chittleborough 1995; Hengl, Heuvelink, and Rossiter 2007; Karl 2010).

Advantages of generalized least squares

Previously, I showed that models of BIBI are likely to violate the OLS assumptions of constant residual variance (homoscedasticity) and independence of residuals. OLS regression is sensitive to violation of these assumptions, and models fit using OLS are likely to suffer from bias in the estimation of coefficients, precision, and the selection of variables (Legendre 1993; Beale et al. 2010; Zuur, Ieno and Elphick 2010). A different regression method should be used for this study, and spatial GLS appears to be well-suited to this research.

Spatial GLS models may incorporate heteroscedasticity and spatial
autocorrelation into the error or random term of the regression equation. Coefficients describing the slope, intercept, autocorrelation, and heteroscedasticity are estimated simultaneously. This results in unbiased estimates of slope and intercept, and parametric models describing the pattern of spatial autocorrelation and heteroscedasticity.

Spatial GLS models can be described by equation 1 (after Hengl, Huevelink, and Rossiter 2007):

\[ Y(s_0) = \sum_{k=0}^{p} \beta_k \cdot x_k(s_0) + e(s_0) + \varepsilon(s_0) \]  

where \( Y(s_0) \) is the value of the dependent variable at location \( s_0 \), \( \beta_k \) are the global regression coefficients, \( x_k(s_0) \) is the value of predictor variable \( k \), \( e(s_0) \) is the spatially structured or local component of the error, and \( \varepsilon(s_0) \) is the unstructured error. The unstructured error should be normally distributed with a mean of zero, but it may exhibit heteroscedasticity, represented in equation 2:

\[ \varepsilon \sim N(0, \sigma^2 \times f(\ldots)) \]  

where \( f(\ldots) \) is a function describing the pattern of variance. This function may be arbitrarily complex, incorporating one or more dependent or predictor variables (a variance-covariate function). The GLS fitting process estimates any unknown parameters in the variance-covariate function. The correlation and variance-covariate terms of GLS models affect the fitting of fixed regression coefficients. For this reason, it is important to select the form and variables for these random terms before selecting among fixed regression variables (Zuur et al. 2009).

GLS parameters are estimated using likelihood methods. Likelihood measures
how likely a set of parameter values are, given a set of observations and assumptions about probability distributions. In other words, likelihood tells us how well the model fits the data. Maximum likelihood estimation finds the set of coefficient estimates that maximize model likelihood. Restricted maximum likelihood provides unbiased estimates of variance-covariate and autocorrelation functions. Both methods of estimation are appropriate in different situations (Zuur et al. 2009).

**Variography**

Typically, the pattern of spatial autocorrelation is not *a priori* known, and must be modeled and estimated. Variograms are used to measure and model spatial autocorrelation. In variography, autocorrelation is quantified as the semivariance. Semivariance is the variance among a subset of sites, where subsets are defined by pairwise distance intervals (lags). Lower values of semivariance indicate autocorrelation, and distance decay of autocorrelation is the increase of semivariance with lag. Empirical variograms plot the observed semivarance versus the lag distances. Theoretical variograms and variogram models represent the distance decay of autocorrelation using a function of one or more parameters.

Parametric variograms can be described using three variables. Sill is the semivariance value at which the variogram function levels out, indicating that autocorrelation is no longer present. Range is the maximum lag where autocorrelation is detected, or where the variogram function approximates the sill. Nugget is the amount of semivariance observed for pairs of sites separated by the smallest lag considered, and represents the amount of variance that is not autocorrelated (random error) (de Smith, Goodchild and Longley 2011). Variogram parameters may vary with direction.
(anisotropy), or be invariant to direction (isotropy). Anisotropic variogram models contain parameters that describe how the variogram changes with direction. GLS models may use parametric variograms to model and estimate the spatial autocorrelation among samples. The variogram parameters are estimated along with the other GLS parameters during GLS fitting (Pinheiro et al. 2010). However, the analyst must choose what theoretical variogram model to use (e.g., Gaussian versus exponential distance decay, isotropic versus anisotropic).

Regression Kriging

Kriging is a form of geostatistical interpolation that predicts values at unsampled locations using a variogram model of autocorrelation among sampled locations. Kriging provides surfaces of predicted values and estimates of prediction variance (kriging variance) (Cressie 1986). Regression kriging (RK) is a method of interpolation that combines the global predictions of a regression equation with local kriging predictions of regression errors (Odeh, McBratney, and Chittleborough 1995). Predictions at new locations are made by adding the kriged residuals to the regression predictions. RK leverages the spatial dependence of the regression model residuals to increase prediction accuracy and remove spatial patterns of bias. The global regression equation removes spatial trends from the residuals, allowing the use of mathematically simple kriging methods (Hengl, Heuvelink, and Rossiter 2007). RK has been used extensively in soil science (e.g., Hengl, Heuvelink, and Stein 2004), and is beginning to see application in studies of wildlife and vegetation (e.g., Hengl et al. 2009; Karl 2010).

When using RK to make predictions at new locations, the spatially structured
errors are estimated as:

\[
\hat{e}(s_0) = \sum_{i=1}^{n} \lambda_i \cdot e(s_i)
\]  

(3)

where \( \lambda_i \) are the kriging weights as determined by the estimated variogram model, and \( e(s_i) \) are the residuals at calibration location \( s_i \) (Hengl, Heuvelink, and Rossiter 2007). Regression kriging is simple to use for prediction of new data with spatial GLS models. The spatial GLS model results include a variogram fitted to the GLS regression residuals. This variogram model can be used to krige the GLS regression residuals at the new data locations.

Variography and Residual Drift

Geostatistical techniques, such as spatial GLS and regression kriging, share the assumptions of geostatistics. One of these assumptions, the “intrinsic hypothesis” of Matheron (Cressie 1986), requires that the mean is constant across space, and the variance depends only on the relative position of sampling sites. Thus, variography is sensitive to trends (“drift”) in the modeled random field (in this case, the model residuals) (Cressie 1986; Legendre and Fortin 1989). The variogram will not attain a constant sill, and coefficients for the nugget and range may be unstable. The logic of spatial GLS, RK, and many other methods in geostatistics is to account for global trends in the fixed terms of the regression equation, and account for local autocorrelation through local error terms, as specified in the correlation structure (Cressie 1986).

Thus, drift in regression residuals is problematic for GLS models, in that it confounds the correlation structure: the variogram coefficients will reflect a combination of strong, local autocorrelation (of interest) and weak autocorrelation created by the drift.
(not of interest), and therefore will be biased. In GLS, this problem will reduce the effective sample size, limiting statistical power in the regression analysis. It may also result in biased beta coefficients and variances, and biased model residual variance (Beale et al. 2010). These biases can affect the outcome of the variable selection method, as well as model interpretation and prediction. Furthermore, this may bias the kriged residuals, reducing the accuracy and precision of predictions.

One accepted method for addressing drift is to detrend the residuals before estimating the variogram (Cressie 1986; Gringarten and Deutsch 2001). This can be accomplished by fitting a linear regression of the form residuals = bX + bY, where X and Y are Easting and Northing coordinates of the study sites. The residuals of this regression are detrended, and used to estimate the variogram parameters. Spatial GLS fitting then continues using the estimates as fixed parameters. While detrending in this manner also results in biased variogram parameters (Cressie 1986), any bias created by detrending will likely be less than bias created by drift (Gringarten and Deutsch 2001).

Variable Selection and Information Criteria

Multiple regression, including GLS, is sensitive to model specification error. Including the unimportant or spurious predictors can bias coefficients and precision estimates. Variable selection is a key step in regression modeling in order to avoid these biases. When important predictor variables are excluded from a model, or unimportant variables are included, estimated coefficients may be biased. Automated variable selection procedures, such as stepwise and exhaustive search methods, can help researchers avoid excluding important variables, but increase the risk that spurious correlations will influence variable selection and bias coefficients (e.g., Anderson and
Burnham 2002; Johnson and Omland 2004). To avoid these problems, the set of models that will be compared can be fixed prior to the analysis. Models representing specific hypotheses are constructed \textit{a priori} and compared. The hypotheses address the importance of variables and the form of their relationships with the dependent variable. The result of these comparisons provides evidence for each hypothesis, and this evidence is used to inform variable selection (Johnson and Omland 2004).

Information theoretic approaches to model comparison balance improved model fit with increasing model complexity. Information criteria measure how efficiently models use information (variables) to make predictions. Models that are more efficient (make better predictions with fewer variables) receive better information criteria scores. In this way, using information criteria to compare models helps us include important predictors while excluding unimportant predictors. Information criteria are used to compare the information efficiency of two or more models. Results may identify clearly superior models, but also identify situations where support for several models is equivocal. This property allows for some reflexivity in the statistical analysis, and can be used to inform alternative model development or further study.

Akaike information criteria (AIC) is a well known information criterion. It represents the amount of information lost when using a model to describe reality. Smaller AIC scores indicate less information loss and a more information efficient model. When the sample size is small relative to the number of parameters being estimated, it is best to use the small sample corrected Akaike information criteria (AICc: Equation 4) (Burnham and Anderson 2002):

\[
AICc = 2k - 2\ln(L) + \frac{2k(k + 1)}{n - k - 1}
\] (4)
where \( k \) is the number of parameters estimated, \( L \) is the likelihood of the fitted model, and \( n \) is the sample size.

\[
AIC = -2 \ln(L) + 2k
\]

AIC is only meaningful when comparing competing models. Therefore, the delta AIC (\( \Delta AICc \)) is useful to present. \( \Delta AICc \) is the difference between a model's AICc and the smallest AICc in the comparison set. Models with \( \Delta AICc \) values greater than 10 are clearly inferior in information efficiency. \( \Delta AICc \) greater than 6 is strong evidence of inferiority, and greater than 3 is moderate evidence. When \( \Delta AICc \) is less than 3, there is not clear evidence that the model is less information efficient than the best model in the set. In these situations, comparison of model likelihood may be helpful (Johnson and Omland 2004).

Stationarity

Because the number of factors influencing stream health is large, it is unlikely that statistical models for study areas covering heterogeneous riverscapes can incorporate all relevant variables, and some degree of model misspecification is unavoidable. Missing variables will reduce predictive accuracy, and may produce temporal, spatial, or regional non-stationarity. Non-stationarity is patterns of bias in prediction accuracy and precision. Checking for and describing any patterns of non-stationarity is an important part of validating that model assumptions were met. It also provides useful information for users applying the model in their work and researchers seeking to further improve knowledge and predictive abilities. The GLS modeling methods discussed earlier provide powerful methods for testing and measuring non-stationarity. Bias in precision can be modeled using variance-covariate functions. Residual autocorrelation patterns can be modeled using variogram correlation structures. Bias in prediction accuracy is modeled using the
fixed regression terms. AICc model comparisons are used to compare the evidence for
different patterns of non-stationarity.

Categorical Predictions

Predicted levels of BIBI correspond to regulatory categories. Continuous
variables are converted to categories by assigning thresholds or cutoffs: observations
above or below each cutoff is assigned to a different category. Choice of cutoff affects
the probability of different kinds of error, and therefore the frequency and associated cost
of incorrect predictions. One method for evaluating cutoff options and choosing optimal
cutoffs is receiver operating characteristics (ROC). ROC methods have been used
extensively in medicine, for example to optimize error rates of diagnostic tests (Zweig
and Campbell 1993), and are also used in conservation biology (Fielding and Bell 1997;
Drechsler and Burgman 2004), geography (Pontius and Schneider 2001), and stream
ecology (Wenger et al. 2008). Area under curve (AUC) is helpful as a measure of
categorical prediction quality: AUC = 0.5 is equivalent to random guessing, and AUC =
1.0 is a perfect classification. The ROC graphs themselves are useful tools for managers,
allowing one to pick cutoff values to match desired accuracy or error rates. Confusion
matrices, classification accuracies, and the kappa statistic are also useful measures of
categorical prediction performance (Congalton 1991; Fielding and Bell 1997).
CHAPTER 3

METHODS

The literature indicates that land cover and stormwater management (SWM) throughout the riverscape affect stream health. This chapter introduces the study area, and describes the variables and data we use to measure land cover, SWM, and the benthic index of biotic integrity (BIBI). I describe the statistical methods used, and develop the analysis plan.

Study Area

Baltimore County lies on the western shore of the Chesapeake Bay in the State of Maryland, in the Mid-Atlantic region of the United States. The county’s 1570 square kilometers enclose the City of Baltimore, and extend from the Patapsco River north to Pennsylvania and east to the Little Gunpowder Falls (Figure 1). Baltimore County has a history of strong land use controls, manifested in the present distribution of land cover and population. About one-third of the land area is forest, one-third is agriculture, and one-third is developed. Approximately 85% of the population lives inside the Urban Rural Demarcation Line, which delineates the county into two-thirds rural and one-third urban (Duerkson and Snyder 2005).

Watersheds contributing to Baltimore County streams extend into neighboring jurisdictions, including portions of Carroll County to the west, Harford County to the east, the City of Baltimore to the south, and York County, Pennsylvania to the North.
Figure 1. Study location maps. From left to right: Maryland within the Mid-Atlantic region, Baltimore County within Maryland, and field study sites within Baltimore County.

Sampling Design

Baltimore County samples invertebrates in wadeable, perennial streams with catchment areas > 121.4 hectares (300 acres), using methods similar to the Maryland Biological Stream Survey (MBSS) (Klauda et al. 1998; Southerland et al. 2005). Each year approximately 100 sites are chosen at random along perennial streams. In odd years, streams in the Patapsco and Back River basins are sampled. In even years, sampling occurs in the Gunpowder River and Deer Creek basins (Figure 2). Baltimore County provided tables of macroinvertebrate data, site locations, and site catchment polygons for 610 samples collected during 2003 and 2008. One study catchment is smaller than 0.5 hectares, and this site is removed from the analysis.

Baltimore County returns to a few “sentinel” sites every two years, providing longitudinal BIBI data between 2003 and 2008. However, the number of longitudinal sites is small, and the majority of samples are from unique, randomly located sites. Although each sample is dated, multiple years of data for predictor variables are not
available, making it impossible to conduct a temporal analysis. Therefore, this study takes a purely spatial or cross-sectional approach. Only the earliest sample from each sentinel sites is included in the study: samples from subsequent years were excluded.

**Figure 2.** Physiographic provinces and watersheds
Watersheds are coded: BckR is Back River, DrCr is Deer Creek, GwyF is Gwynns Falls, JnsF is Jones Falls, LbrR is Liberty Reservoir, LcRR is Loch Raven Reservoir, LtGF is Little Gunpowder Falls, LwGF is Lower Gunpowder Falls, PatR is Patapsco River, and PrtR is Prettyboy Reservoir.
Grid is in meters, Maryland State Plan 1983 HARN

Physiographic Province

This study examines only data from sites located in the Piedmont physiographic province (Figure 2). The BIBI statistic is derived differently in each of the three
physiographic provinces in Maryland. Additionally, differences in hydrogeochemistry, geomorphology, and biology should result in different relationships between stream health, land cover, and SWM. Any sites attributed as Coastal Plain by Baltimore County, or located in the Coastal Plain physiographic province described in the US Level III Ecoregions shapefile (U.S. Environmental Protection Agency 2010), were removed from the sample.

Completeness of SWM and Tree Canopy Data

Data for tree canopy and SWM were not available for all the study sites. Three types of sites with incomplete data were removed before analysis. They include: 1) sites without complete urban tree canopy land cover data (UTC) coverage; 2) sites with drainage from York PA, Carroll County, and Harford County, and 3) sites where more than 5% of SWM facilities did not have a drainage area attribute.

Division of Data into Fitting and Evaluation Sets

As one goal of this study is to create a predictive model, it is preferable to split the data into a calibration or fitting sample, and a separate validation or evaluation sample. The large number of covariates of interest demands a large sample size for the fitting sample, and a 50%/50% random split provided too few sites. Instead, the first four years of data (2003 to 2006) are used for fitting, and the last two years (2007 and 2008) are used for evaluation. This is similar to the way the regression models would be used: past data for calibration, future conditions predicted. Although the model calibration and validation are spatially explicit in nature, the results could be used as an assessment of temporal and spatial variation of stream health. The final dataset contained 428 field sites (293 in fitting sample, 135 in evaluation sample).
Geographic Units: Catchments, Riparian Buffers, and Reaches

The scale and method of measuring geographic variables such as land cover has a demonstrable effect on the results of stream health studies (e.g., Frimpong et al. 2005; King et al. 2005; Barker et al. 2006). This section explains the geographic units used in this study. Predictor variables are summarized at the catchment, riparian, and reach levels of the riverscape. Each study site has a unique catchment, riparian buffer, and reach (Figure 3).

Figure 3. Geographic units used in this study.

Baltimore County delineated catchments (drainage areas to each study site) using heads-up digitizing from local 2 ft interval elevation contours. Riparian zones are delineated as 30.5 m (100 ft) buffers from both banks of all perennial streams within the catchment. A narrow riparian zone width is selected for two reasons. First, the Piedmont region has many narrow stream valleys with incised channels and correspondingly narrow riparian zones. Second, larger buffer widths resulted in riparian zones occupying large fractions of catchments, diminishing the contrast between riparian and catchment
land covers and increasing the probability that riparian and catchment variables would be collinear.

Reach zones capture the environment in the immediate vicinity of the sampling site. Each MBSS study reach is 75 m long (Stribling et al. 1998). Reach zones are estimated by extracting the catchment riparian buffer within a 100m buffer of the sample site point location.

Baltimore County provided a GIS feature class representing these riparian buffers. Buffers are derived from a local hydrology geodatabases, compiled at 1:2400 scale by the Sanborn mapping company using photogrammetric methods. A Python script was written to iterate over all study sites, clipping the stream buffers to the catchment, creating the reach zone, and summarizing SWM and land cover variables within the catchment, riparian buffers, and reach zone.

The variables used in this study are summarized in Table 1 and are further described in the following sections.
Table 1. Description of variables used

<table>
<thead>
<tr>
<th>Category</th>
<th>Variable</th>
<th>Abbreviation</th>
<th>Level of Measurement</th>
<th>Source Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground Condition</td>
<td>Benthic Index of Biotic Integrity</td>
<td>BIBI</td>
<td>interval</td>
<td>2003 - 2008 point = 75m reach</td>
</tr>
<tr>
<td>Regulatory Category</td>
<td>Stream Health Regulatory Category</td>
<td>Regulatory Category</td>
<td>ordinal</td>
<td>&quot;</td>
</tr>
<tr>
<td>Land Cover</td>
<td>Catchment % Impervious Surfaces</td>
<td>Impervious</td>
<td>ratio</td>
<td>2005 1:2400</td>
</tr>
<tr>
<td>Land Cover</td>
<td>Catchment % Agriculture</td>
<td>Ag</td>
<td>&quot;</td>
<td>2006 30m pixel</td>
</tr>
<tr>
<td>Land Cover</td>
<td>Catchment % Forest</td>
<td>Forest</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>Land Cover</td>
<td>Catchment % Tree Canopy</td>
<td>Canopy</td>
<td>&quot;</td>
<td>2007 0.61m pixel</td>
</tr>
<tr>
<td>Land Cover</td>
<td>Riparian Buffer % Tree Canopy</td>
<td>Riparian Canopy</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>Land Cover</td>
<td>Reach Buffer % Tree Canopy</td>
<td>Reach Canopy</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>SWM</td>
<td>% of Catchment Area treated by SWM</td>
<td>SWM All</td>
<td>ratio</td>
<td>2007 point = drainage area</td>
</tr>
<tr>
<td>SWM</td>
<td>% of Catchment Area treated by EDP</td>
<td>SWM EDP</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>SWM</td>
<td>% of Catchment Area treated by DDP</td>
<td>SWM DDP</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>SWM</td>
<td>% of Catchment Area treated by F</td>
<td>SWM F</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>SWM</td>
<td>% of Catchment Area treated by I</td>
<td>SWM I</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>SWM</td>
<td>% of Catchment Area treated by WP</td>
<td>SWM WP</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

**Distribution of BIBI**

There are 12 observed levels of BIBI (Figure 5). Modeling BIBI as a continuous variable is reasonable with this number of ordinal levels. BIBI exhibits a bimodal distribution in both the fitting and evaluation samples (Figure 5). Note that GLS regression assumes normally distributed errors, not a normally distributed dependent variable (Zuur et al. 2009). The evaluation sample has a greater proportion of sites with BIBI > 3.5 than the fitting data. Mean BIBI is slightly higher in the evaluation sample (3.15) than the fitting sample (2.92). Standard deviation of BIBI is nearly equivalent (1.05 in fitting sample, 1.08 in evaluation sample).
Figure 5. Numerical distribution of BIBI
Lines indicate proportion of observations at each possible BIBI value. Dark blue lines are the fitting sample; light green lines are the evaluation sample.

Table 2 illustrates the distribution of regulatory categories. The proportion of sites in the fair category is similar between the fitting and evaluation samples. However, there are more impaired and fewer high quality observations in the fitting sample.

<table>
<thead>
<tr>
<th>Regulatory Category</th>
<th>Fitting Sample</th>
<th>Evaluation Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>High Quality</td>
<td>77</td>
<td>26.3%</td>
</tr>
<tr>
<td>Fair</td>
<td>80</td>
<td>27.3%</td>
</tr>
<tr>
<td>Impaired</td>
<td>136</td>
<td>46.4%</td>
</tr>
</tbody>
</table>

Figure 6 shows strong clustering of BIBI. BIBI also shows a spatial trend, increasing from south to north. In this pattern of autocorrelation, sites with similar BIBI scores tend to be closer to each other than they are to sites with dissimilar scores. Some areas contain a cluster of similar BIBI scores and with one or several outliers bucking the trend. If land cover and SWM predictors do not explain all of the clustering, the residuals of the model will remain clustered and autocorrelated. BIBI varies among Maryland Department of the Environment (MDE) 8 digit watersheds (compare Figure 6 with Figure 2).
Figure 6. Geographic distribution of BIBI
Symbols reflect the distribution of BIBI: each class is one standard deviation in width. The last circle symbol ranges from $(\mu - 0.5\sigma)$ to $(\mu + 0.5\sigma)$. Lighter colored circles are below the mean and the regulatory threshold for impairment, while darker colored squares are above than the mean and are unimpaired. Study area is the combination of all individual study site catchments. Grid is in meters, Maryland State Plan 1983, HARN.

**Measures of Stormwater Management**

Carroll County, Baltimore County, and the Maryland State Highways Administration provided SWM facility GIS data. These entities are required to obtain National Pollutant Discharge Elimination System (NPDES) permits for stormwater discharge systems. The Maryland Department of the Environment has stimulated the creation and maintenance of SWM geodatabases throughout Maryland by phasing in NPDES permit requirements for SWM GIS data submission. However, SWM geodatabases maintained by different government agencies are (at the time of study)
heterogeneous in design and completeness. The Maryland State Highways Administration had a complex and robust geodatabase design, including facility locations, drainage areas, conveyances, inlets, and outfalls. Baltimore County provided two shapefiles, one with facility locations, and another with drainage areas. Carroll County provided a facility location shapefile. Each geodatabase is able to store land area draining to each facility. Facilities stored in these geodatabases were approved for construction between 1958 and 2009. Facilities approved after 2007 were removed from the data.

After reviewing attribute definitions, database table schemas, and geographic projections, translation scripts were written to bring the data into a shared geodatabases schema. The combined data supports gross area treated by SWM facilities, by type of SWM facility, for most of the study area. As drainage area polygon data is not complete throughout the study area, it is not possible to use more nuanced approaches to measuring SWM (e.g., land cover treated by each SWM, accounting for SWM facility “trains” treating nested drainage areas, etc.).

SWM is operationalized as the proportion of catchment treated by SWM (area treated by SWM facilities ÷ catchment area). Because SWM facility drainage areas can nest inside each other, and we count the drainage area of each facility separately, there is no upper limit on this number.

Description of Stormwater Management Facilities

SWM facilities include a diversity of stormwater best management practices (BMPs). Each BMP operates differently and is designed to perform different functions. Different types of BMPs may have different effects on stream health. SWM facilities are
assigned types using the definitions employed by the Chesapeake Bay Program that include: dry detention ponds, extended detention ponds, wetponds and wetlands, filtration, and infiltration (Chesapeake Bay Program, Urban Stormwater Working Group 2003; Simpson and Weammert 2009).

Dry detention ponds (DDP) slow the flow of storm runoff, storing excess flow for a short period of time while it is slowly released downstream. They are meant to dry up completely between storms. This detention effect reduces the maximum flood volume, protecting streams from bank erosion and reducing flooding. However, these ponds do not retain water long enough for sediment deposition, toxin uptake by plants, or nutrient removal. This limits their water quality benefits (Simpson and Weammert 2009).

Extended detention ponds (EDP) are similar: they dry out between storms, but retain water long and release it more slowly, allowing time for sediment, nutrient, and toxin removal.

Wetponds and wetlands have permanent pools of water and often feature naturalized vegetation. The permanent pools and wetlands provide sediment removal and nutrient cycling, and the vegetation fixes nutrients and toxins. These facilities also detain stormwater, protecting stream channels from massive peak flows (Simpson and Weammert 2009). Unfortunately, wet ponds can increase the temperature of stormwater runoff, resulting in reduced D.O. levels (Galli 1990). This has been shown to be harmful to trout and other coldwater fish species.

Filtration BMPs include a diversity of practices. In all of these, stormwater passes through a filtering medium, such as sand or soil. The stormwater then enters a pipe and is routed elsewhere, either to a stormdrain, outfall, or a detention facility. There
is no treatment of water quantity, but the filter does remove sediment, nutrients, and some toxins (Simpson and Weammert 2009).

Infiltration practices store a volume of stormwater, preventing large stormflows from damaging stream channels. They are designed so water infiltrates the soil and enters groundwater flows, rather than moving directly to a stream channel, stormdrain, or surface outfall. Sediment, nutrients, and toxins are removed and processed during the passage through the infiltration media, the subsoil, and the groundwater (Simpson and Weammert 2009). Ideally, the groundwater will flow through the riparian and hyporheic zones before entering a stream channel, where further nutrient and toxin processing and uptake work is done by the unique microbial communities of these soil zones (Groffman et al. 2003).
Figure 7. Geographic distribution of dry detention pond and wet pond SWM facilities. Points represent SWM facility locations. Study area is the combination of all individual study site catchments. Grid is in meters, Maryland State Plane 1983 HARN

SWM facilities are clustered in the urbanized portions of the study area (Figure 7). Some types of SWM (e.g., dry detention ponds) are relatively common and evenly distributed. Other types (e.g., wet ponds and wetlands) are rare and concentrated in small areas. Additional maps illustrating the geographic distribution of SWM facilities and proportion of catchment treated by SWM are in Appendix 1.

It is possible that SWM treatment mechanisms are more important than the five types described above. Therefore, SWM facilities are grouped by treatment mechanism: Detention (dry detention ponds, extended detention ponds), Filtration, and Both (infiltration, and wetponds and wetlands). A third measure of SWM is total area treated by all SWM types.

I expect that all SWM facility types will have a positive, statistically significant
effect on BIBI. The effect size will be large enough to have regulatory and biological importance. Effect sizes will be smallest for BMPs using only detention treatment, larger for BMPs using filtration treatment, and largest for BMPs using both detention and filtration.

Stormwater Management and Site Catchments

While reviewing the SWM drainage areas, it became clear that many SWM drainage areas modify the catchments of the study sites: the drainage areas cross topographic ridges, piping stormwater from one topographic catchment into another. SWM drainage area polygons that touched study site catchment boundaries are identified using spatial queries, and a relationship table is constructed to link individual SWM facilities to the study sites they drain to. A SWM facility drains to a study site if the SWM facility outfall discharges into the study site catchment. A variety of ancillary data were consulted to identify outfall locations, including aerial photography, elevation contours and Light Detection and Ranging (LiDAR)-derived digital elevation models, and SWM facility plans when available.

Once the relationship table is completely populated, a geoprocessing script is written to loop through each study site catchment, adding SWM facility drainage polygons discharging to the study site catchments, and erasing SWM facility drainage polygons not discharging to the site. These SWM-corrected catchments are used for the rest of the study.

The impact of SWM drainage areas on the catchment data is quantified using GIS overlay functions to determine the area gained, lost, and retained between the original and SWM revised catchments.
Measures of Land Cover

Impervious surfaces are estimated using GIS feature classes representing building footprints, road surfaces, and large parking lots and driveways, provided by Baltimore County Office of Information Technology (Baltimore County 2007a, 2007b). These feature classes were compiled at 1:2400 scale by the Sanborn mapping company using photogrammetric methods.

Urban tree canopy land cover (UTC) is provided by the University of Vermont (UVM) (UVM 2009a, 2009b). This data is ideal for measuring tree canopy in heterogeneous environments such as urban, suburban, or agricultural landscapes (e.g., Smith et al. 2010). Given the UTC’s high spatial resolution (0.61 meter pixels), it is also ideal for measuring tree canopy in small areas such as riparian and reach buffers (e.g., Goetz et al. 2003). The UTC data was creating by classifying leaf-on 4-band aerial photography collected in 2006 and a digital surface model derived from LiDAR collected in 2005, using an object-based image analysis method (OBIA).

Datasets for forest and agricultural land cover are extracted from a Chesapeake Bay watershed land cover map, ground conditions 2006 (U.S. Geological Survey 2009). This data provides estimates of agriculture and forest that are comparable to the 2001 National Land Cover Dataset (NLCD, Homer et al. 2004), but at a more recent date. This land cover map was created by detecting changes in LANDSAT scenes relative to a 2001 scene, and then modifying the 2001 NLCD where changes are detected (similar to Xian, Homer, and Fry 2009).
Figure 8. Geographic distribution of agricultural and tree canopy land covers.
Points represent study site locations, with symbols reflecting the distribution of the variable of interest: each class is one standard deviation in width. The last circle symbol ranges from $(\mu - 0.5\sigma)$ to $(\mu + 0.5\sigma)$. 
Lighter colored circles are below the mean and the regulatory threshold for impairment, while darker colored squares are greater than the mean and are unimpaired. Study area is the combination of all individual study site catchments. Grid is in meters, Maryland State Plane 1983 HARN.

Most land cover variables exhibited clustering of similar values. Imperviousness and agriculture exhibited large spatial gradients, while canopy variables displayed smaller clusters (Figure 8). Additional maps illustrating the geographic distribution of land cover variables are in Appendix A.

I expect impervious surfaces will have a large, negative, and statistically significant effect on BIBI. The effect will be non-linear, and the impact of increasing imperviousness will diminish as the level of imperviousness increases. I expect agriculture will have a statistically significant effect on BIBI, but the effect size will depend on the level of imperviousness: the effect of increasing agriculture will lessen as imperviousness increases. Forest land cover, riparian, and reach scale tree canopy are all expected to have positive, statistically significant effects on BIBI. These effects will depend on imperviousness in the same manner described above for agriculture.

**Statistical Methods**

When predictor variables exhibit outliers, transformations are applied to improve regression model validity. SWM variables are square-root transformed. Reach scale tree canopy is angular transformed (arcsine(square root(x))).

A sequence of comparisons among *a priori* specified models is used to build the final model. The comparisons are designed to support the research objectives by identifying important predictors and the form of their relationships with BIBI. In each round of comparisons, a new predictor is added to the best model from the prior round,
starting with a null model with no predictors. Several models are constructed and compared at once. Each of these *a priori* models represents a specific hypothesized relationship between the new predictor and BIBI by using a different set of variables (e.g., linear term, quadratic term, interaction term) for the new predictor. The order in which predictors are added reflects the expected importance of predictor variables, based on literature review and prior studies (e.g., Goetz and Fisk 2008). Variables known to have a large effect on stream health are added first, and less certain predictors are added later (Table 3).

<table>
<thead>
<tr>
<th>Round</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Impervious</td>
</tr>
<tr>
<td>2</td>
<td>Catchment Ag, Forest, and Canopy</td>
</tr>
<tr>
<td>3</td>
<td>SWM</td>
</tr>
<tr>
<td>4</td>
<td>Riparian Canopy</td>
</tr>
<tr>
<td>5</td>
<td>Reach Canopy</td>
</tr>
</tbody>
</table>

To avoid collinearity issues, variance inflation statistics are computed for combinations of predictor variables. Any combinations where variance inflation factor (VIF) exceeded 3 are avoided when constructing the *a priori* models.

Regression model assumptions must be met to make valid inferences and maximize statistical efficiency. Following the examples of Zuur et al. (2009), selection of an optimal regression model began with a saturated model fit via ordinary least squares. A saturated model is used to ensure that any problems with model assumptions are not caused by missing predictor variables (Zuur et al. 2009). The saturated model included quadratic and interaction terms, but avoided collinear sets of variables (Table 4).
Table 4. Predictor variables used in saturated model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Form</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Impervious</td>
<td>quadratic</td>
<td></td>
</tr>
<tr>
<td>Canopy</td>
<td>quadratic</td>
<td>Impervious</td>
</tr>
<tr>
<td>Riparian Canopy</td>
<td>quadratic</td>
<td>Impervious</td>
</tr>
<tr>
<td>Reach Canopy</td>
<td>linear</td>
<td>Impervious</td>
</tr>
<tr>
<td>SWM EDP</td>
<td>linear</td>
<td>Impervious</td>
</tr>
<tr>
<td>SWM DDP</td>
<td>linear</td>
<td>Impervious</td>
</tr>
<tr>
<td>SWM F</td>
<td>linear</td>
<td>Impervious</td>
</tr>
<tr>
<td>SWM I</td>
<td>linear</td>
<td>Impervious</td>
</tr>
<tr>
<td>SWM WP</td>
<td>linear</td>
<td>Impervious</td>
</tr>
</tbody>
</table>

Residuals are examined for violations of assumptions. Normality of residuals is examined using normal probability plots and histograms. Residual constant mean and variance are examined using scatterplots of residuals against fitted values and continuous predictors, and boxplots against factors. Variograms, correlograms, and Local Indicators of Spatial Association (LISA) maps (Anselin 1995) are used to detect and examine spatial autocorrelation patterns. The ncf package for R (Bjornstad 2009) is used to calculate correlograms and LISA statistics. The geoR package for R (Ribeiro and Diggle 2001) is used to calculate and plot isotropic variograms. Statistical significance in correlograms is evaluated at individual lags using ninety-nine Monte Carlo permutations of residuals on spatial locations. The observed Moran’s I is compared to the 95% percentile from the Monte Carlo permutations. Statistical significance of semi-variogram points is determined using 99% confidence envelopes, produced using the same Monte Carlo method. LISA neighborhoods are defined as circular areas with radii equal to the range of global spatial autocorrelation, estimated from the correlogram. Due to issues of multiple testing, statistical significance of individual LISA clusters is best evaluated

**Specification of the GLS Model**

**Heteroscedasticity**

Heteroscedasticity is addressed by fitting a saturated GLS model, containing all the potential fixed effects terms. Residuals of this saturated model are reviewed for patterns of heteroscedasticity, guiding the selection of candidate variance-covariate functions. Several GLS models, one for each candidate variance-covariate function, are fitted using restricted maximum likelihood. The most likely variance covariance function is identified by comparing small-sample corrected Akaike information criterion (AICc), examining normalized model residuals to confirm that assumptions are met (Zuur et al. 2009), and considering whether the interpretation of the fitted variance structure is plausible. The chosen variance structure is included in subsequent rounds of model fitting.

Our covariates can take values of zero, therefore exponential variance-covariance functions are suitable (Zuur et al. 2009). The distribution of residuals with an exponential variance function is described by Equation 5,

\[\epsilon_i = N(0, \sigma^2 \times e^{2\delta_1 x_{i1}} \times \cdots \times e^{2\delta_p x_{ip}})\]  

(5)

where \(x_1\) to \(x_p\) are the \(p\) covariates, \(\delta_1\) to \(\delta_p\) are the \(p\) coefficients, and each residual \(\epsilon_i\) may have a unique variance.

**Spatial Autocorrelation and Drift**

The nlme package for R (Pinheiro et al. 2010) offers a limited set of theoretical variograms, including isotropic spherical, exponential, Gaussian, rational quadratic, and a
linear type model, with or without nuggets. Residuals from each model are detrended via a first-order linear regression on site coordinates, of the form Residuals = bX + bY, where X and Y are Easting and Northing coordinates of the study sites. Variogram models are fit to the detrended residuals, then applied to the original model. I refer to this method as spatial generalized least squares detrended (GLS-DT). An R language function was written to automate this process and return a fitted GLS-DT model.

AICc model comparison is used to identify the variogram model best supported by the data. Normalized model residuals are examined to confirm residual autocorrelation is removed, and variogram fit is visually examined (Zuur et al. 2009).

**Predictor Variable Selection**

After the form of the random term is specified, the fixed terms are specified, using the sequential comparison of a priori models described earlier. In each sequential round, models are compared using AICc. The best model is used as a ‘null’ model for the next round in the sequence. A final ‘best model’ is chosen for validation and subsequent inference and predictive use.

**Validation of Model Assumptions**

Residuals are extracted from the final model, and regression kriging (RK) predictions are made using the final model and the evaluation sample. The assumption of constant mean and variance are validated by examining plots of residuals and square root absolute residuals against fitted values and predictor variables.

The independence of residuals is evaluated by looking for spatial autocorrelation in correlograms and variograms. Spatial stationarity can be assessed by looking for spatial trends, autocorrelation, LISA clusters, and variation among regions (e.g.,
watersheds) in the residuals. Temporal stationarity is examined by looking for variation in residuals among sampling years.

When patterns among years or watersheds are detected, GLS regression is used to determine if the patterns are statistically significant and to estimate the effect sizes. First, to avoid problems with small sample sizes, watersheds with < 10 sites are removed from the residuals (estimating the variance of a population from a sample smaller than 10 is likely to produce erroneous estimates). Then, variance-covariate functions are compared in a model predicting residuals using watershed. After the variance covariates are selected, a null model is compared to models predicting residuals using watershed and year. If the null model is not the AICc best, I conclude that there is statistically significant regional or temporal structure in the residuals. The beta coefficients of the AICc best model are interpreted to determine the effect of year and watershed on mean error. If watershed or year is a statistically significant variance covariate, the fitted variance function is interpreted to determine the effect of year and watershed on residual variance.

Note that year and watershed cannot be combined in the same model: watersheds are sampled in alternating years, and thus year is identical to a collection of watersheds. Model comparison is carried out for models predicting normalized residuals, to ensure that any patterns detected cannot be explained away by the spatial correlation structure and variance covariates used in the model fitting process. For interpreting the coefficients, raw residuals are used.

Evaluation of Model Performance

After validation, the effects of SWM and land cover variables are described using
beta coefficients and effect sizes, and partial-residual plots. The variance structure is described by plotting the function over the range of variance covariate values, and the correlation structure is described with a variogram model plot, and maps of kriged residuals and kriging variance.

RK predictions and model fitting predictions are evaluated using a variety of performance measures. Regulatory category classification accuracy, BIBI prediction accuracy and precision, and proportion of variance explained (pseudo-R$^2$) are the primary measures used. Large differences in prediction performance between the fitting and evaluation data is evidence of over-fitting, missing covariates, or non-stationary in the processes generating stream health.

Predictive performance and precision of continuous BIBI predictions are described using measures of precision (standard deviation of errors, mean absolute error (MAE), root mean squared error (RMSE)), and two pseudo-R$^2$ measures: a coefficient of determination and the proportion of variance explained. The coefficient of determination is simply the square of the Pearson correlation of the predicted and true values. The proportion of variance explained is the ratio of the sum of squared residuals divided by the sum of squared deviations from the mean. These measures are described in Equations 5 and 6,

$$\text{Coeff. of Determination} = r^2_{Y, \hat{Y}}$$

$$\% \text{Variance Explained} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

where $Y$ is the true BIBI, $\hat{Y}$ is the fitted or predicted BIBI, $\bar{Y}$ is the sample mean.
BIBI, and \( n \) is the sample size.

Predictions are classified into the regulatory categories (high quality, fair, and impaired), and categorical accuracy and precision are evaluated. First, receiver operating characteristics (ROC) graphs are constructed, and the area under curve (AUC) statistic extracted using the ROCR package for R (Sing et al. 2009). These “ROC-informed” cutoffs are then used to make predictions. The accuracy of these predictions is compared to predictions made using “naïve” cutoffs, which are the cutoffs for field observations of BIBI specified by Maryland regulations.

As an example of this use of ROC curves, two error goals are chosen that might correspond to policy goals for a development management agency. The decision considered is whether to permit a land development project in a catchment when the project might cause stream health impairment. The first goal is a precautionary approach to protecting stream health: the rate of false predictions of unimpaired stream health (\( \beta \) error) should be held below 5%. The second goal is a permissive approach that implies that the cost of denying the project is greater than the cost of stream impairment. In this scenario, we require a high level of certainty that the stream health will be impaired by the project under review, and so the goal is to hold the rate of false predictions of impaired stream health (\( \alpha \) error) below 5%.

Performance of these ROC cutoffs is evaluated by using them to convert the RK predictions to regulatory categories, and comparing the error rates to the policy goal scenarios. Confusion matrices are constructed using the regulatory cutoffs and the “ROC-informed” cutoffs, and accuracy and kappa statistics are calculated. Fitting and
RK prediction accuracy and kappa statistics are compared: consistency is desirable.

Statistical analyses are performed in the R language for statistical computing (R Development Core Team 2010). There is a large and growing code base for spatial statistical analysis in R. Moreover, the open-source R language affords flexibility and customization of analysis functions and graphs.
CHAPTER 4

RESULTS

In this chapter, the results of analyses described in chapter 3 are presented. Evidence for collinearity, autocorrelation, and heteroscedasticity is presented. Model comparisons provide evidence for competing hypotheses regarding the relationship of land cover and stormwater management (SWM) to the benthic index of biotic integrity (BIBI). A single best model is selected, and is validated against the assumptions of generalized least squares models. Patterns of residual variation across space, time, and watersheds are described. The model is used to estimate the effects of land cover and SWM on BIBI, and to predict BIBI at the independent evaluation sites. Predictions of BIBI and regulatory categories made using this model are evaluated for accuracy. The use of receiver operating characteristics (ROC) methods to optimize category predictions is demonstrated.

Description of Data

The distribution of each variable used in the regression model is described in Table 5. Percent reach canopy was skewed towards high values, while percent of catchment treated by stormwater management facilities was skewed towards low values. To reduce skew and avoid outlier effects, reach canopy was arcsine transformed, and SWM was square-root transformed. The distribution of variables was similar in the sample and evaluation samples (Table 5). The evaluation sample shows less agriculture,
more forest, and more tree canopy. The range of tree canopy and infiltration SWM was larger in the evaluation sample.

**Table 5.** Distributions of variables by sample

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fitting Sample (n = 293)</th>
<th>Evaluation Sample (n = 135)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Quartiles</td>
<td>Quartiles</td>
</tr>
<tr>
<td></td>
<td>Min. 1st 2nd 3rd Max. Mean SD</td>
<td>Min. 1st 2nd 3rd Max. Mean SD</td>
</tr>
<tr>
<td>BIBI</td>
<td>1.00 2.00 3.00 4.00 4.67 2.92 1.05</td>
<td>1.00 2.00 3.33 4.00 4.67 3.15 1.08</td>
</tr>
<tr>
<td>Impervious</td>
<td>0.000 0.031 0.069 0.181 0.526 0.115 0.112</td>
<td>0.012 0.034 0.062 0.138 0.440 0.113 0.112</td>
</tr>
<tr>
<td>Ag</td>
<td>0.000 0.055 0.214 0.492 0.924 0.292 0.259</td>
<td>0.000 0.035 0.193 0.412 0.923 0.255 0.236</td>
</tr>
<tr>
<td>Forest</td>
<td>0.020 0.250 0.374 0.508 0.962 0.385 0.186</td>
<td>0.000 0.270 0.418 0.577 0.919 0.424 0.202</td>
</tr>
<tr>
<td>Canopy</td>
<td>0.093 0.375 0.454 0.550 0.925 0.460 0.136</td>
<td>0.011 0.386 0.480 0.578 0.965 0.489 0.146</td>
</tr>
<tr>
<td>Riparian Canopy</td>
<td>0.000 0.604 0.741 0.797 0.988 0.698 0.163</td>
<td>0.020 0.658 0.744 0.829 0.996 0.731 0.142</td>
</tr>
<tr>
<td>Reach Canopy</td>
<td>0.000 0.568 0.821 0.956 0.998 0.734 0.262</td>
<td>0.100 0.728 0.903 0.982 0.997 0.810 0.229</td>
</tr>
<tr>
<td>SWM All</td>
<td>0.000 0.000 0.057 0.148 1.125 0.116 0.182</td>
<td>0.000 0.000 0.042 0.122 1.763 0.105 0.208</td>
</tr>
<tr>
<td>SWM EDP</td>
<td>0.000 0.000 0.007 0.053 0.628 0.040 0.077</td>
<td>0.000 0.000 0.010 0.049 0.252 0.031 0.044</td>
</tr>
<tr>
<td>SWM DDP</td>
<td>0.000 0.000 0.001 0.053 0.901 0.047 0.101</td>
<td>0.000 0.000 0.000 0.031 0.904 0.047 0.111</td>
</tr>
<tr>
<td>SWM F</td>
<td>0.000 0.000 0.000 0.005 0.082 0.005 0.011</td>
<td>0.000 0.000 0.000 0.002 0.031 0.003 0.006</td>
</tr>
<tr>
<td>SWM I</td>
<td>0.000 0.000 0.000 0.004 0.133 0.005 0.015</td>
<td>0.000 0.000 0.000 0.001 0.707 0.010 0.062</td>
</tr>
<tr>
<td>SWM WP</td>
<td>0.000 0.000 0.000 0.006 0.938 0.018 0.085</td>
<td>0.000 0.000 0.000 0.000 0.613 0.015 0.073</td>
</tr>
</tbody>
</table>

**Note:** EDP = extended detention pond, DDP = dry detention pond, F = filtering, I = infiltration, and WP = wet pond

Autocorrelation of each variable is described using Moran’s correlograms (Figure 9). Most predictor variables exhibit strong spatial autocorrelation, with Moran’s I exceeding 0.5 at small lags (Figure 9). Reach canopy and filtering SWM are exceptions, with weak autocorrelation. BIBI exhibits an autocorrelation pattern similar to several predictor variables, including agriculture, imperviousness, and dry and extended detention pond SWM. If spatial generalized least squares (GLS) methods are not used to address these shared patterns of autocorrelation, the effects of these variables would be inflated.
Figure 9. Moran’s spatial correlograms of BIBI and predictor variables
Filled dots have $I \neq 0$ at $p < 0.05$. $p$-values calculated using 99 monte-carlo permutations.
Nbr dist is the farthest lag distance at which autocorrelation is statistically significant and positive, and is indicated with a box and dashed line.
Effect of Stormwater Management on Catchments

As shown by the horizontal lines in Figure 10, nearly 75% of study site catchment boundaries were modified by SWM drainage areas. Some catchments experienced large alterations: for example, in one catchment, SWM drainage areas added and removed land equivalent to 19% of the area of the original topographic catchment. However, most changes were small compared to the size of the catchments: in 90% of study site catchments, gross change area was less than 4% of the original catchment area (Figure 10). Moreover, gains and losses tended to offset each other: net change fell between -3% and 2% for 90% of catchments (Figure 10).

\[
\begin{align*}
\text{Gross } \Delta \text{ Catchment Area} &= \frac{\text{Area Removed} + \text{Area Added}}{\text{Original Catchment Area}} \\
\text{Net } \Delta \text{ Catchment Area} &= \frac{\text{Area Gained} - \text{Area Removed}}{\text{Original Catchment Area}}
\end{align*}
\]

\( \text{Rug plot shows individual observations.} \)

\textbf{Figure 10. Changes in catchment areas due to SWM}  
Gross \( \Delta \) Catchment Area = (Area Removed + Area Added) / Original Catchment Area  
Net \( \Delta \) Catchment Area = (Area Gained – Area Removed) / Original Catchment Area  
Rug plot shows individual observations.

\textbf{Multicollinearity in Land Cover}

Table 6 presents bivariate correlations among the variables. Bivariate collinearity was particularly strong for Forest and Canopy (Pearson’s \( r = 0.91 \)). This is to be expected, because they measure very similar phenomena. Correlations were high or
moderate among many other pairs of variables, including: canopy and riparian canopy; agriculture and impervious; forest and impervious; and riparian canopy and agriculture.

Table 6. Bivariate correlations

<table>
<thead>
<tr>
<th></th>
<th>BIBI</th>
<th>Impervious</th>
<th>Agriculture</th>
<th>Forest</th>
<th>Canopy</th>
<th>Rip. Canopy</th>
<th>Reach Canopy</th>
<th>SWM EDP</th>
<th>SWM DDP</th>
<th>SWM F</th>
<th>SWM I</th>
<th>SWM WP</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIBI</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impervious</td>
<td>-0.69</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>0.43</td>
<td>-0.66</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>0.53</td>
<td>-0.59</td>
<td>-0.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canopy</td>
<td>0.32</td>
<td>-0.31</td>
<td>-0.40</td>
<td>0.91</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rip. Canopy</td>
<td>0.16</td>
<td>-0.08</td>
<td>-0.48</td>
<td>0.65</td>
<td>0.75</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reach Canopy</td>
<td>0.14</td>
<td>-0.12</td>
<td>-0.15</td>
<td>0.32</td>
<td>0.34</td>
<td>0.44</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWM EDP</td>
<td>-0.39</td>
<td>0.51</td>
<td>-0.55</td>
<td>-0.16</td>
<td>-0.05</td>
<td>0.17</td>
<td>0.06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWM DDP</td>
<td>-0.52</td>
<td>0.66</td>
<td>-0.61</td>
<td>-0.29</td>
<td>-0.10</td>
<td>0.09</td>
<td>-0.02</td>
<td>0.52</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWM F</td>
<td>-0.11</td>
<td>0.21</td>
<td>-0.23</td>
<td>-0.07</td>
<td>-0.02</td>
<td>0.14</td>
<td>-0.02</td>
<td>0.34</td>
<td>0.21</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWM I</td>
<td>-0.14</td>
<td>0.19</td>
<td>-0.22</td>
<td>0.01</td>
<td>0.02</td>
<td>0.09</td>
<td>-0.03</td>
<td>0.42</td>
<td>0.29</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>SWM WP</td>
<td>-0.27</td>
<td>0.24</td>
<td>-0.24</td>
<td>-0.13</td>
<td>-0.07</td>
<td>-0.03</td>
<td>0.03</td>
<td>0.29</td>
<td>0.33</td>
<td>0.15</td>
<td>0.07</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Bivariate scatterplots showed that Impervious, Agriculture, and Forest or Canopy have strong non-linear correlations, as shown by the scatterplots and locally weighted scatterplot smoothing (LOESS) lines in Figure 11. This kind of multicollinearity is to be expected among land cover and other proportional variables (King et al. 2005). Additional scatter plots are provided in the appendix.

**Figure 11.** Bivariate scatterplots for BIBI and catchment land cover variables
Darkness of points is proportional to density of points. Black line is LOESS smoother. Diagonals show histogram for each variable.

These patterns of collinearity caused variance inflation factor (VIF) to exceed 3 for all models containing impervious surfaces and more than one of forest, canopy, or agriculture. Any models containing Riparian Canopy and either Agriculture or Forest also had VIF > 3. Surprisingly, models with Canopy, Riparian Canopy, and Reach
Canopy had VIF < 3. All models where VIF exceeded 3 are likely to have biased \( p \)-values for regression coefficients (Zuur, Ieno, and Elphick 2010), and were therefore excluded from the set of \( a \ priori \) models. The sequence of model fitting rounds, adjusted to avoid high VIF models, is shown in Table 7.

<table>
<thead>
<tr>
<th>Round</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Impervious</td>
</tr>
<tr>
<td>2</td>
<td>Ag vs. Forest vs. Canopy</td>
</tr>
<tr>
<td>3</td>
<td>SWM</td>
</tr>
<tr>
<td>4</td>
<td>Riparian Canopy*</td>
</tr>
<tr>
<td>5</td>
<td>Reach Canopy*</td>
</tr>
</tbody>
</table>

*Note:* Riparian and Reach Canopy do not enter models with Ag or forest.

**Specifying the Error Terms**

We began the model selection by checking for heteroscedasticity and spatial autocorrelation in a saturated model. The saturated model contained the maximum number of potential predictor variables: Impervious, Impervious\(^2\), Canopy, Canopy\(^2\), Riparian Canopy, Riparian Canopy\(^2\), \( \sin^{-1}(\text{Reach Canopy}) \), \( \sqrt{(SWM\ DDP)} \), \( \sqrt{(SWM\ EDP)} \), \( \sqrt{(SWM\ F)} \), \( \sqrt{(SWM\ I)} \), \( \sqrt{(SWM\ WP)} \), and interaction terms between all variables and Impervious. Concerns about collinearity prevented the inclusion of agriculture in the saturated model.

Figure 12 shows that standardized residuals from the saturated model are heteroscedastic. It is clear that residual variance varies systematically with fitted values (\( \hat{Y} \)) and Impervious. Therefore, fitted value and impervious are variance-covariates.
Figure 12. Scatterplots illustrating relationship of variance of saturated model residuals with Fitted Values and Impervious Residuals are Observed BIBI – Fitted BIBI, adjusted using the variance covariance function. Dot darkness increases with scatterplot density. Thick black lines are LOESS smoothers.

A variety of variance functions were added to the saturated model, and these models were compared. The results are presented in Table 8. The best model, as determined by small-sample corrected Akaike information criterion (AICc), used impervious, Ŷ, and Ŷ² as variance covariates.

Table 8. AICc comparison of several variance covariance functions in the saturated model

<table>
<thead>
<tr>
<th>Variance Covariates</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imp., Ŷ, Ŷ²</td>
<td>25</td>
<td>-238.0</td>
<td>530.8</td>
<td>0.0</td>
<td>0.712</td>
</tr>
<tr>
<td>Imp., Imp.², Ŷ, Ŷ²</td>
<td>26</td>
<td>-237.9</td>
<td>533.0</td>
<td>2.3</td>
<td>0.228</td>
</tr>
<tr>
<td>Ŷ, Ŷ²</td>
<td>24</td>
<td>-242.3</td>
<td>537.1</td>
<td>6.3</td>
<td>0.030</td>
</tr>
<tr>
<td>Imp.²</td>
<td>23</td>
<td>-244.0</td>
<td>538.1</td>
<td>7.4</td>
<td>0.018</td>
</tr>
<tr>
<td>Imp., Imp.²</td>
<td>24</td>
<td>-243.8</td>
<td>540.1</td>
<td>9.3</td>
<td>0.007</td>
</tr>
<tr>
<td>Imp.</td>
<td>23</td>
<td>-246.2</td>
<td>542.4</td>
<td>11.6</td>
<td>0.002</td>
</tr>
<tr>
<td>Imp., Imp.², Ŷ</td>
<td>25</td>
<td>-243.8</td>
<td>542.5</td>
<td>11.7</td>
<td>0.002</td>
</tr>
<tr>
<td>Imp., Ŷ</td>
<td>24</td>
<td>-245.9</td>
<td>544.2</td>
<td>13.4</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Ŷ</td>
<td>23</td>
<td>-251.8</td>
<td>553.7</td>
<td>22.9</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>constant variance</td>
<td>22</td>
<td>-262.4</td>
<td>572.6</td>
<td>41.8</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Minimum of 11.3 samples per coefficient estimated

Notes: Imp. = % Catchment Impervious Surfaces

The standardized residuals from this model are homoscedastic (Figure 13).

Therefore, this variance function was added to the saturated model and used in all further model development.
Figure 13. Scatterplots illustrating homoscedasticity after adding the variance covariate function to the saturated model
Residuals are Observed BIBI – Fitted BIBI, adjusted using the variance covariance function. Dot darkness increases with scatterplot density. Thick black lines are LOESS smoothers.

Next, spatial autocorrelation and stationarity were evaluated in the saturated model. There is a clear first order spatial trend in standardized residual values and variance: residuals increase with Northing, and residual variance increases with Easting (Figure 14). These first order or “global” spatial trends operate on the scale of the entire study area.

Figure 14. Scatterplots illustrating spatial trends in saturated model residuals
Coordinates are in NAD 1983 HARN, Maryland State Plane, meters. Residuals are Observed BIBI – Fitted BIBI, adjusted using the variance covariance function. Dot darkness increases with scatterplot density. Thick black lines are LOESS smoothers.
Residual autocorrelation is detected in correlograms and semi-viograms (Figure 15). The correlogram indicates statistically significant correlations (filled dots) among the residuals for sites located less than 1,500 meters apart.

**Figure 15.** Correlogram and variogram illustrating spatial autocorrelation of saturated model residuals
Filled dots in the correlogram indicate statistically significant correlation ($p < 0.05$).
Dashed lines in the semi-viogram illustrate 99% confidence envelopes.

To address this autocorrelation, correlation structures (theoretical variograms) were added to the saturated model. However, using the standard GLS method, the variogram fails to attain a sill, and the range is unrealistically large (Figure 16). Using the GLS-DT method results in a variogram with a sill, reasonable range, and larger nugget (Figure 16). The degree of this problem increased with the number of fixed variables added to the model, suggesting that variance inflation or over-fitting compound this drift-related problem. This demonstrates why the GLS-DT method was developed and used in this study.
A variety of theoretical variograms were added to the saturated model, and compared via AICc (Table 9). AICc provides substantial evidence that models with spatial correlation structures are better supported by the data. The exponential variogram is the AICc best model. Including the variogram correlation structure diminished the amount and statistical significance of autocorrelation in the saturated GLS-DT model residuals (Figure 17).

**Table 9.** AICc comparison of several theoretical variograms on the saturated model

<table>
<thead>
<tr>
<th>Variogram model used</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>weight</th>
<th>range (m)</th>
<th>effective range (m)</th>
<th>nugget</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>29</td>
<td>-229.4</td>
<td>523.4</td>
<td>0.0</td>
<td>0.450</td>
<td>1,387</td>
<td>4,160</td>
<td>0.760</td>
</tr>
<tr>
<td>Gaussian</td>
<td>29</td>
<td>-229.8</td>
<td>524.2</td>
<td>0.9</td>
<td>0.289</td>
<td>1,725</td>
<td>2,987</td>
<td>0.799</td>
</tr>
<tr>
<td>Spherical</td>
<td>29</td>
<td>-230.0</td>
<td>524.5</td>
<td>1.2</td>
<td>0.250</td>
<td>3,605</td>
<td>3,605</td>
<td>0.780</td>
</tr>
<tr>
<td>No correlation structure</td>
<td>25</td>
<td>-238.0</td>
<td>530.8</td>
<td>7.4</td>
<td>0.011</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Minimum of 10.1 samples per coefficient estimated
Figure 17. Correlogram and variogram illustrating diminished spatial autocorrelation in the saturated GLS-DT model residuals. Filled dots in the correlogram indicate statistically significant correlation ($p < 0.05$). Dashed lines in the semi-variogram illustrate 99% confidence envelopes.

The saturated GLS-DT model appears to meet the assumptions of GLS regression: normalized residuals are normally distributed and show no signs of heteroscedasticity (Figure 18), and are not affected by spatial dependence (Figure 18).

Figure 18. Validation plots of residuals from the saturated GLS-DT model. A: Histograms of residuals, B: Normal Quantile-Quantile plots of residuals, C: Plot of observed versus fitted BIBI. Dot darkness increases with scatterplot density, dark line is a LOESS smoother. D: Plot of square root of absolute value of residuals and errors versus fitted and predicted BIBI. Dot darkness increases with scatterplot density, dark line is a LOESS smoother.
**Specifying the Fixed Terms**

The model with a quadratic relationship between BIBI and % Impervious Surfaces is small sample corrected Akaike information criteria (AICc) best, and was chosen for the next step in model fitting (Table 10). The next round is catchment land cover.

### Table 10. AICc comparison of models with imperviousness

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Impervious quadratic</td>
<td>11</td>
<td>-278.0</td>
<td>579.0</td>
<td>0.0</td>
<td>0.696</td>
<td>1.000</td>
</tr>
<tr>
<td>+ Impervious linear</td>
<td>10</td>
<td>-310.0</td>
<td>640.8</td>
<td>61.8</td>
<td>0.777</td>
<td>0.000</td>
</tr>
<tr>
<td>null model</td>
<td>9</td>
<td>-327.7</td>
<td>674.0</td>
<td>95.0</td>
<td>1.453</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Minimum of 26.6 samples per coefficient estimated

Agricultural land cover clearly improves the model, but there is ambivalence regarding the form of the relationship between agriculture and BIBI: the three AICc best models have nearly identical likelihood, AICc, and residual root mean square error (RMSE) (Table 11A). Examination of partial residual plots did not support a strong curvilinear relationship. In the interest of parsimony, statistical power, and interpretability, the linear model was selected for further work.

### Table 11A. AICc comparison of models with agriculture

<table>
<thead>
<tr>
<th>Model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Ag linear</td>
<td>12</td>
<td>-269.6</td>
<td>564.3</td>
<td>0.0</td>
<td>0.669</td>
<td>0.314</td>
</tr>
<tr>
<td>+ Ag quadratic</td>
<td>13</td>
<td>-268.5</td>
<td>564.3</td>
<td>0.1</td>
<td>0.668</td>
<td>0.302</td>
</tr>
<tr>
<td>+ Ag quadratic * Imp</td>
<td>14</td>
<td>-267.5</td>
<td>564.5</td>
<td>0.2</td>
<td>0.667</td>
<td>0.280</td>
</tr>
<tr>
<td>+ Ag linear * Imp</td>
<td>13</td>
<td>-269.6</td>
<td>566.5</td>
<td>2.2</td>
<td>0.669</td>
<td>0.104</td>
</tr>
<tr>
<td>Impervious quadratic</td>
<td>11</td>
<td>-278.0</td>
<td>579.0</td>
<td>14.7</td>
<td>0.696</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Minimum of 20.9 samples per coefficient estimated

Forest land cover improves the model, but AICc does not provide an obvious choice between the linear, quadratic, or impervious interaction forms (Table 11B). Once again, the linear (simplest) model was chosen to maximize parsimony, power, and interpretability. Tree canopy land cover results were similar to forest land cover, but AICc evidence was stronger (Table 11C).
Table 11B. AICc comparison of models with forest

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Forest linear</td>
<td>12</td>
<td>-262.7</td>
<td>550.5</td>
<td>0.0</td>
<td>0.671</td>
<td>0.484</td>
</tr>
<tr>
<td>+ Forest quadratic</td>
<td>13</td>
<td>-262.2</td>
<td>551.8</td>
<td>1.3</td>
<td>0.673</td>
<td>0.251</td>
</tr>
<tr>
<td>+ Forest linear * Imp</td>
<td>13</td>
<td>-262.3</td>
<td>552.0</td>
<td>1.5</td>
<td>0.667</td>
<td>0.232</td>
</tr>
<tr>
<td>+ Forest quadratic * Imp</td>
<td>14</td>
<td>-263.2</td>
<td>555.9</td>
<td>5.4</td>
<td>0.664</td>
<td>0.033</td>
</tr>
<tr>
<td>Impervious quadratic</td>
<td>11</td>
<td>-278.0</td>
<td>579.0</td>
<td>28.5</td>
<td>0.696</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Minimum of 20.9 samples per coefficient estimated

Table 11C. AICc comparison of models with canopy

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Canopy linear</td>
<td>12</td>
<td>-262.5</td>
<td>550.0</td>
<td>0.0</td>
<td>0.668</td>
<td>0.705</td>
</tr>
<tr>
<td>+ Canopy quadratic</td>
<td>13</td>
<td>-262.6</td>
<td>552.4</td>
<td>2.4</td>
<td>0.665</td>
<td>0.216</td>
</tr>
<tr>
<td>+ Canopy linear * Imp</td>
<td>13</td>
<td>-263.7</td>
<td>554.7</td>
<td>4.6</td>
<td>0.665</td>
<td>0.070</td>
</tr>
<tr>
<td>+ Canopy quadratic * Imp</td>
<td>14</td>
<td>-264.8</td>
<td>559.1</td>
<td>9.0</td>
<td>0.658</td>
<td>0.008</td>
</tr>
<tr>
<td>Impervious quadratic</td>
<td>11</td>
<td>-278.0</td>
<td>579.0</td>
<td>28.9</td>
<td>0.696</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Minimum of 20.9 samples per coefficient estimated

Models with forest or canopy have higher likelihoods than those with agriculture. However, models with Canopy and Forest are nearly AICc equivalent (Table 11D).

Guided by the VIF results presented earlier, we chose to do further model development with canopy, as we are interested in the effects of riparian canopy cover, and models with forest and riparian canopy have stronger multicollinearity and will be more difficult to interpret. The next round is stormwater management.

Table 11D. AICc comparison of models with land covers

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Canopy linear</td>
<td>12</td>
<td>-262.5</td>
<td>550.0</td>
<td>0.0</td>
<td>0.668</td>
<td>0.554</td>
</tr>
<tr>
<td>+ Forest linear</td>
<td>12</td>
<td>-262.7</td>
<td>550.5</td>
<td>0.4</td>
<td>0.671</td>
<td>0.445</td>
</tr>
<tr>
<td>+ Ag linear</td>
<td>12</td>
<td>-269.6</td>
<td>564.3</td>
<td>14.2</td>
<td>0.669</td>
<td>0.000</td>
</tr>
<tr>
<td>Impervious quadratic</td>
<td>11</td>
<td>-278.0</td>
<td>579.0</td>
<td>28.9</td>
<td>0.696</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Minimum of 24.4 samples per coefficient estimated
The model using SWM facilities grouped by mechanism of action was AICc best (Table 12). However, the coefficients for SWM by mechanism were unexpected: SWM facilities using both detention and filtering of stormwater had a negative relationship with BIBI. This reveals the hypothesis regarding SWM facilities by mechanism to be incorrect. The next best model uses the SWM by type variables (Table 12).

**Table 12.** AICc comparison of models with SWM

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ SWM by Mechanism</td>
<td>15</td>
<td>-255.0</td>
<td>541.8</td>
<td>0.0</td>
<td>0.659</td>
<td>0.843</td>
</tr>
<tr>
<td>+ SWM by Type</td>
<td>17</td>
<td>-254.6</td>
<td>545.3</td>
<td>3.6</td>
<td>0.659</td>
<td>0.140</td>
</tr>
<tr>
<td>Canopy linear</td>
<td>12</td>
<td>-262.5</td>
<td>550.0</td>
<td>8.3</td>
<td>0.668</td>
<td>0.013</td>
</tr>
<tr>
<td>+ SWM Total</td>
<td>13</td>
<td>-262.7</td>
<td>552.7</td>
<td>11.0</td>
<td>0.667</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Minimum of 17.2 samples per coefficient estimated

As shown in Table 13, models with quadratic terms for riparian canopy have higher likelihood than models without riparian canopy or with linear terms. However, the improvement is small relative to the number of additional variables, as indicated by AICc values. Interactions between riparian canopy and impervious surfaces decrease the likelihood of the models, and these can be ruled out. There is ambivalence between models with linear and quadratic terms for buffer canopy, and models with no buffer canopy terms.

**Table 13.** AICc comparison of models with riparian canopy

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ Buffer Canopy quadratic</td>
<td>19</td>
<td>-252.2</td>
<td>545.3</td>
<td>0.0</td>
<td>0.657</td>
<td>0.388</td>
</tr>
<tr>
<td>SWM by Type</td>
<td>17</td>
<td>-254.6</td>
<td>545.3</td>
<td>0.1</td>
<td>0.659</td>
<td>0.373</td>
</tr>
<tr>
<td>+ Buffer Canopy linear</td>
<td>18</td>
<td>-254.4</td>
<td>547.3</td>
<td>2.0</td>
<td>0.659</td>
<td>0.142</td>
</tr>
<tr>
<td>+ Buffer Canopy quadratic * Imp</td>
<td>20</td>
<td>-252.6</td>
<td>548.2</td>
<td>2.9</td>
<td>0.644</td>
<td>0.089</td>
</tr>
<tr>
<td>+ Buffer Canopy linear * Imp</td>
<td>19</td>
<td>-256.1</td>
<td>552.9</td>
<td>7.7</td>
<td>0.653</td>
<td>0.008</td>
</tr>
</tbody>
</table>

Minimum of 14.7 samples per coefficient estimated

Similar to the results for riparian buffer canopy, there is much ambivalence between different models with reach canopy variables, and AICc for models without reach canopy appear just as likely as models with linear and quadratic terms for reach
canopy (Table 14). Again, including an interaction with imperviousness causes models to be less likely.

**Table 14. AICc comparison of models with reach canopy**

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buffer Canopy quadratic</td>
<td>19</td>
<td>-252.2</td>
<td>545.3</td>
<td>0.0</td>
<td>0.657</td>
<td>0.396</td>
</tr>
<tr>
<td>+ Reach Canopy quadratic</td>
<td>21</td>
<td>-250.0</td>
<td>545.5</td>
<td>0.2</td>
<td>0.651</td>
<td>0.361</td>
</tr>
<tr>
<td>+ Reach Canopy linear</td>
<td>20</td>
<td>-252.2</td>
<td>547.6</td>
<td>2.3</td>
<td>0.657</td>
<td>0.126</td>
</tr>
<tr>
<td>+ Reach Canopy quadratic * Imp</td>
<td>22</td>
<td>-250.4</td>
<td>548.5</td>
<td>3.2</td>
<td>0.650</td>
<td>0.079</td>
</tr>
<tr>
<td>+ Reach Canopy linear * Imp</td>
<td>21</td>
<td>-252.3</td>
<td>549.9</td>
<td>4.7</td>
<td>0.655</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Minimum of 13.3 samples per coefficient estimated

Comparing the best models from each round, we see that large AICc improvements end with the addition of SWM terms (Table 15). Models with and without riparian and reach canopy have equal support from the data. Because we are interested in the effect sizes of riparian and reach canopy, the model containing both these variables was selected for validation and interpretation.

**Table 15. AICc comparison of best models from each variable selection round**

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imp² + Canopy + SWM Type + Riparian Can.²</td>
<td>19</td>
<td>-252.2</td>
<td>545.3</td>
<td>0.0</td>
<td>0.657</td>
<td>0.338</td>
</tr>
<tr>
<td>Imp² + Canopy + SWM Type</td>
<td>17</td>
<td>-254.6</td>
<td>545.3</td>
<td>0.1</td>
<td>0.659</td>
<td>0.324</td>
</tr>
<tr>
<td>Imp² + Canopy + SWM Type + Riparian Can.² + Reach Can.²</td>
<td>21</td>
<td>-250.0</td>
<td>545.5</td>
<td>0.2</td>
<td>0.651</td>
<td>0.307</td>
</tr>
<tr>
<td>Imp² + Canopy</td>
<td>12</td>
<td>-262.5</td>
<td>550.0</td>
<td>4.8</td>
<td>0.668</td>
<td>0.031</td>
</tr>
<tr>
<td>Imp²</td>
<td>11</td>
<td>-278.0</td>
<td>579.0</td>
<td>33.7</td>
<td>0.696</td>
<td>0.000</td>
</tr>
<tr>
<td>null model</td>
<td>9</td>
<td>-327.7</td>
<td>674.0</td>
<td>128.8</td>
<td>1.453</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Minimum of 14 samples per coefficient estimated


An objective of this work is to study the effects of agricultural land cover on stream health. Towards this end, models containing agriculture were examined. Adding a linear agriculture term to the model with imperviousness, canopy, SWM type, riparian canopy, and reach canopy (in Table 15) increased AICc (AICc = 547.6, df = 22, n = 293, log likelihood = -249.9). This also provided insubstantial improvements to accuracy (RMSE = 0.650). Comparing this new model and the corresponding model in Table 15,
we see that $dAICc = 2.1$. This indicates uncertainty regarding which model has greater information efficiency. Likelihood values are essentially identical.

Additional models were fit by adding agriculture and removing canopy. The AICc best of these agriculture models predicts BIBI using imperviousness, agriculture, and SWM by type (Table 16). The highest likelihood also included riparian and reach canopy. All of the agriculture models exhibited worse AICc scores (Table 16: min AICc = 560.0) than models with canopy (Table 15: min AICc = 545.3). The $dAICc$ between AICc best canopy and agriculture models is 14.7, providing strong evidence that models with canopy are more information efficient than models with agriculture.

Table 16. AICc comparison of best agriculture models

<table>
<thead>
<tr>
<th>model</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>$dAICc$</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imp$^2$ + Ag + SWM Type</td>
<td>17</td>
<td>-261.9</td>
<td>560.0</td>
<td>0.0</td>
<td>0.659</td>
<td>0.663</td>
</tr>
<tr>
<td>Imp$^2$ + Ag + SWM Type + Riparian Can.$^2$</td>
<td>19</td>
<td>-261.0</td>
<td>562.7</td>
<td>2.8</td>
<td>0.656</td>
<td>0.164</td>
</tr>
<tr>
<td>Imp$^2$ + Ag + SWM Type + Riparian Can.$^2$ + Reach Can.$^2$</td>
<td>21</td>
<td>-259.2</td>
<td>563.8</td>
<td>3.9</td>
<td>0.652</td>
<td>0.095</td>
</tr>
<tr>
<td>Imp$^2$ + Ag</td>
<td>12</td>
<td>-269.6</td>
<td>564.3</td>
<td>4.3</td>
<td>0.669</td>
<td>0.077</td>
</tr>
<tr>
<td>Imp$^2$</td>
<td>11</td>
<td>-278.0</td>
<td>579.0</td>
<td>19.0</td>
<td>0.696</td>
<td>0.000</td>
</tr>
<tr>
<td>null model</td>
<td>9</td>
<td>-327.7</td>
<td>674.0</td>
<td>114.1</td>
<td>1.453</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Minimum of 14 samples per coefficient estimated

Validating Model Assumptions

While the residuals are normally distributed, the distribution of regression kriging (RK) errors clearly diverges from normal (Figure 19A, Figure 19B). In addition, Figure 19A shows that RK errors have an overall under-prediction bias.

![Fitting Sample Standardized Residuals](image)

![Evaluation Sample RK Prediction Error](image)

**Figure 19.** Distribution of final model residuals and errors

RK prediction errors are Observed BIBI - RK prediction. Standardized residuals are Observed BIBI – Fitted BIBI, adjusted using the variance covariance function and error correlation matrix. A: Histograms of residuals and errors, used to assess bias and normality. B: Normal Quantile-Quantile plots of residuals and errors, used to assess normality.

Mean error is not constant: residuals from the fitted model and the RK errors exhibit a slight curvilinear relationship with fitted value (Figure 20A). This non-linearity results in under-prediction of BIBI for Ŷ greater than about 3, and over-prediction of BIBI for lower Ŷ. It appears that our model’s predictions are biased towards the mean, and away from the extremes. This pattern is also seen in Goetz and Fisk 2010, Figure 7A.
The assumption of homoscedasticity is met for the fitted model. RK errors exhibit a pattern of heteroscedasticity similar to that described by the variance covariate function of the fitted model (Figure 20B).

**Figure 20.** Patterns in final model residuals and errors
RK prediction errors are Observed BIBI - RK prediction. Standardized residuals are Observed BIBI – Fitted BIBI, adjusted using the variance covariance function and error correlation matrix. A: Plot of observed versus fitted and predicted BIBI, used to assess linearity and bias. Dot darkness increases with scatterplot density, dark line is a LOESS smoother. B: Plot of square root of absolute value of residuals and errors versus fitted and predicted BIBI, used to assess homoskedasticity. Dot darkness increases with scatterplot density.
The mean and variance of the residuals and errors exhibit a global trend. Residuals and errors increase from south to north (Figure 21A). Variance of residuals increases from west to east, however this pattern does not exist in the RK errors. (Figure 21B).

**Figure 21.** Global spatial trends in residuals and errors of final model
(A) Mean residual/error versus northing. (B) Variance of residual/error versus Easting. RK prediction errors are Observed BIBI - RK prediction. Standardized residuals are Observed BIBI – Fitted BIBI, adjusted using the variance covariance function and error correlation matrix. Dot darkness increases with scatterplot density. Thick black lines are LOESS smoothers.
Spatial autocorrelation is not found in standardized residuals, but is found in the evaluation RK errors (Figure 22). Some evidence of clustering is detected in the residuals and errors via local indicators of spatial association (LISA) (Figure 23). However, no clusters are statistically significant after adjusting for multiple testing. The pattern of clustering differs between residuals and RK errors: the RK errors exhibit two clusters of over-prediction, and one of under-prediction, while the residuals exhibit two clusters of over-prediction.

**Figure 22.** Autocorrelation in residuals and errors of final model
B: Correlogram. Filled dots indicate statistically significant correlation ($p < 0.05$).
Patterns in the mean and variance of residuals and RK errors were apparent across years and watersheds. Watersheds structure the variance of residuals and errors (Table 17, Table 18). Some watersheds exhibit 2 to 3 times the variance of others (Table 19, Table 20).

**Figure 23.** LISA maps illustrating the clustering of residuals and errors
Black line length is equal to the neighborhood distance used to compute LISA statistics. Sites marked with a + are clusters of high values, - are clusters of low values, and * are local outliers. Significance levels include ns ($p \geq 0.05$), raw ($p < 0.05$), and FDR (Benjamini-Hochberg false discovery rate adjusted $p < 0.05$).

A: Local Moran’s I for standardized residuals. B. Local Moran’s I for RK prediction errors.

**Table 17.** AICc comparison of variance-covariance functions for residuals

<table>
<thead>
<tr>
<th>Variance Covariates</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>watershed</td>
<td>14</td>
<td>-377.3</td>
<td>786.0</td>
<td>0.0</td>
<td>0.946</td>
<td>0.864</td>
</tr>
<tr>
<td>constant variance</td>
<td>8</td>
<td>-386.6</td>
<td>790.3</td>
<td>4.3</td>
<td>0.946</td>
<td>0.101</td>
</tr>
<tr>
<td>year</td>
<td>11</td>
<td>-384.1</td>
<td>792.4</td>
<td>6.4</td>
<td>0.946</td>
<td>0.036</td>
</tr>
</tbody>
</table>

Minimum of 19.93 samples per coefficient estimated
**Table 18.** AICc comparison of variance-covariance functions for prediction errors

<table>
<thead>
<tr>
<th>Variance Covariates</th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>watershed</td>
<td>14</td>
<td>-107.7</td>
<td>246.9</td>
<td>0.0</td>
<td>0.575</td>
<td>0.883</td>
</tr>
<tr>
<td>constant variance</td>
<td>8</td>
<td>-117.2</td>
<td>251.6</td>
<td>4.7</td>
<td>0.575</td>
<td>0.085</td>
</tr>
<tr>
<td>year</td>
<td>9</td>
<td>-117.1</td>
<td>253.6</td>
<td>6.7</td>
<td>0.575</td>
<td>0.032</td>
</tr>
</tbody>
</table>

Minimum of 9.07 samples per coefficient estimated

**Table 19.** Effects of watershed on variance of residuals

<table>
<thead>
<tr>
<th>Watershed</th>
<th>n</th>
<th>σ</th>
<th>5% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prettyboy Res.</td>
<td>14</td>
<td>0.71</td>
<td>0.49, 1.03</td>
</tr>
<tr>
<td>Liberty Res.</td>
<td>30</td>
<td>0.78</td>
<td>0.50, 1.23</td>
</tr>
<tr>
<td>Loch Raven Res.</td>
<td>85</td>
<td>0.80</td>
<td>0.54, 1.20</td>
</tr>
<tr>
<td>Gwynns Falls</td>
<td>50</td>
<td>0.87</td>
<td>0.57, 1.33</td>
</tr>
<tr>
<td>Patapsco River</td>
<td>26</td>
<td>1.05</td>
<td>0.66, 1.66</td>
</tr>
<tr>
<td>Jones Falls</td>
<td>54</td>
<td>1.12</td>
<td>0.74, 1.70</td>
</tr>
<tr>
<td>Lower Gunpowder Falls</td>
<td>20</td>
<td>1.46</td>
<td>0.88, 2.39</td>
</tr>
</tbody>
</table>

**Table 20.** Effects of watershed on variance of prediction errors

<table>
<thead>
<tr>
<th>Watershed</th>
<th>n</th>
<th>σ</th>
<th>5% C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Little Gunpowder Falls</td>
<td>10</td>
<td>0.23</td>
<td>0.14, 0.38</td>
</tr>
<tr>
<td>Liberty Res.</td>
<td>17</td>
<td>0.38</td>
<td>0.25, 0.58</td>
</tr>
<tr>
<td>Gwynns Falls</td>
<td>18</td>
<td>0.45</td>
<td>0.29, 0.67</td>
</tr>
<tr>
<td>Loch Raven Res.</td>
<td>30</td>
<td>0.58</td>
<td>0.45, 0.74</td>
</tr>
<tr>
<td>Lower Gunpowder Falls</td>
<td>10</td>
<td>0.67</td>
<td>0.40, 1.11</td>
</tr>
<tr>
<td>Patapsco River</td>
<td>18</td>
<td>0.69</td>
<td>0.46, 1.05</td>
</tr>
<tr>
<td>Jones Falls</td>
<td>24</td>
<td>0.72</td>
<td>0.49, 1.05</td>
</tr>
</tbody>
</table>

Year also has a significant effect on residuals (Table 21, Table 22), but not on RK error (Table 23). Residuals of the 2003 sites are significantly < 0, and significantly lower than residuals of the 2004, 2005, and 2006 sites at $\alpha < 0.05$.

**Table 21.** AICc comparison of models predicting standardized residuals

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ year</td>
<td>11</td>
<td>-373.1</td>
<td>770.4</td>
<td>0.0</td>
<td>0.960</td>
<td>0.654</td>
</tr>
<tr>
<td>Null model</td>
<td>8</td>
<td>-377.9</td>
<td>772.9</td>
<td>2.5</td>
<td>0.976</td>
<td>0.184</td>
</tr>
<tr>
<td>+ wshed</td>
<td>14</td>
<td>-370.8</td>
<td>773.2</td>
<td>2.8</td>
<td>0.946</td>
<td>0.162</td>
</tr>
</tbody>
</table>

Minimum of 19.93 samples per coefficient estimated
**Table 22. Effects of year on raw residuals**

<table>
<thead>
<tr>
<th>Year</th>
<th>n</th>
<th>( \beta )</th>
<th>SE</th>
<th>p-value</th>
<th>fdr p-value</th>
<th>95% CI</th>
<th>( \hat{Y} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2003*</td>
<td>78</td>
<td>-0.23</td>
<td>0.06</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>-0.34 , -0.12</td>
<td>-0.23</td>
</tr>
<tr>
<td>2004</td>
<td>68</td>
<td>0.40</td>
<td>0.09</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.22 , 0.58</td>
<td>0.17</td>
</tr>
<tr>
<td>2005</td>
<td>82</td>
<td>0.21</td>
<td>0.08</td>
<td>0.013</td>
<td>0.013</td>
<td>0.04 , 0.37</td>
<td>-0.02</td>
</tr>
<tr>
<td>2006</td>
<td>51</td>
<td>0.39</td>
<td>0.10</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>0.18 , 0.59</td>
<td>0.16</td>
</tr>
</tbody>
</table>

**Notes:** * 2003 is the reference year, and its beta coefficient is the expected value of residuals in 2003.
Raw residuals were used to estimate these coefficients, so they are on the BIBI measurement scale, and not the scale of standardized residuals. For the remaining years, beta coefficients and p-values reflect difference from the reference year. FDR p-value is the Benjamini-Hochberg false discovery rate corrected p-value.

Watersheds have a significant effect on RK errors (Table 23, Table 24). RK errors in three watersheds are significantly lower than those in the Lower Gunpowder Falls watershed, at \( \alpha < 0.05 \). Effect size approaches 0.5 BIBI points in the Liberty Reservoir and Little Gunpowder Falls watersheds.

**Table 23. AICc comparison of models predicting RK errors**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>loglike</th>
<th>AICc</th>
<th>dAICc</th>
<th>rmse</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>+ washed</td>
<td>14</td>
<td>-99.5</td>
<td>230.4</td>
<td>0.0</td>
<td>0.575</td>
<td>0.839</td>
</tr>
<tr>
<td>Null model</td>
<td>8</td>
<td>-108.6</td>
<td>234.2</td>
<td>3.8</td>
<td>0.625</td>
<td>0.123</td>
</tr>
<tr>
<td>+ year</td>
<td>9</td>
<td>-108.6</td>
<td>236.5</td>
<td>6.1</td>
<td>0.625</td>
<td>0.039</td>
</tr>
</tbody>
</table>

Minimum of 9.07 samples per coefficient estimated

**Table 24. Effects of watershed on RK errors**

<table>
<thead>
<tr>
<th>Watershed</th>
<th>n</th>
<th>( \beta )</th>
<th>SE</th>
<th>p-value</th>
<th>fdr p-value</th>
<th>95% CI</th>
<th>( \hat{Y} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lower Gunpowder Falls*</td>
<td>10</td>
<td>-0.27</td>
<td>0.22</td>
<td>0.217</td>
<td>0.217</td>
<td>-0.70 , 0.16</td>
<td>-0.27</td>
</tr>
<tr>
<td>Patapsco River</td>
<td>18</td>
<td>0.34</td>
<td>0.27</td>
<td>0.213</td>
<td>0.217</td>
<td>-0.20 , 0.89</td>
<td>0.07</td>
</tr>
<tr>
<td>Jones Falls</td>
<td>24</td>
<td>0.35</td>
<td>0.26</td>
<td>0.192</td>
<td>0.217</td>
<td>-0.18 , 0.87</td>
<td>0.08</td>
</tr>
<tr>
<td>Loch Raven Res.</td>
<td>30</td>
<td>0.47</td>
<td>0.24</td>
<td>0.056</td>
<td>0.098</td>
<td>-0.01 , 0.95</td>
<td>0.20</td>
</tr>
<tr>
<td>Gwynns Falls</td>
<td>18</td>
<td>0.66</td>
<td>0.24</td>
<td>0.008</td>
<td>0.018</td>
<td>0.18 , 1.14</td>
<td>0.39</td>
</tr>
<tr>
<td>Liberty Res.</td>
<td>17</td>
<td>0.73</td>
<td>0.24</td>
<td>0.003</td>
<td>0.009</td>
<td>0.26 , 1.20</td>
<td>0.46</td>
</tr>
<tr>
<td>Little Gunpowder Falls</td>
<td>10</td>
<td>0.75</td>
<td>0.23</td>
<td>0.002</td>
<td>0.009</td>
<td>0.29 , 1.20</td>
<td>0.48</td>
</tr>
</tbody>
</table>

**Notes:** * Lower Gunpowder Falls is the reference watershed: its beta coefficient is the expected value of RK error for sites located in the Lower Gunpowder Falls watershed.
For the remaining watersheds, beta coefficients and p-values reflect difference from the reference watershed.
FDR p-value is the Benjamini-Hochberg false discovery rate corrected p-value.
Estimates of the Model Parameters

Fixed Terms

Table 25 describes the fitted model’s beta coefficients and the maximum effects sizes of each variable. Partial residual plots (in appendix) illustrate the relationship of each variable to BIBI.

### Table 25. Effects of individual variables on BIBI: canopy model

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>Beta Coefficients</th>
<th>Max. Observed Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>SE</td>
</tr>
<tr>
<td>Intercept</td>
<td>2.65</td>
<td>0.37</td>
</tr>
<tr>
<td>% impervious</td>
<td>-14.46</td>
<td>1.11</td>
</tr>
<tr>
<td>% impervious²</td>
<td>22.56</td>
<td>1.98</td>
</tr>
<tr>
<td>% canopy</td>
<td>2.11</td>
<td>0.51</td>
</tr>
<tr>
<td>% riparian canopy</td>
<td>2.29</td>
<td>0.96</td>
</tr>
<tr>
<td>% riparian canopy²</td>
<td>-2.15</td>
<td>0.88</td>
</tr>
<tr>
<td>sin⁻¹(% reach canopy)</td>
<td>-0.73</td>
<td>0.36</td>
</tr>
<tr>
<td>sin⁻¹(% reach canopy)²</td>
<td>0.47</td>
<td>0.21</td>
</tr>
<tr>
<td>√(% treated DDP)</td>
<td>0.38</td>
<td>0.28</td>
</tr>
<tr>
<td>√(% treated EDP)</td>
<td>0.55</td>
<td>0.33</td>
</tr>
<tr>
<td>√(% treated Filt)</td>
<td>1.77</td>
<td>0.64</td>
</tr>
<tr>
<td>√(% treated Inf)</td>
<td>-0.25</td>
<td>0.66</td>
</tr>
<tr>
<td>√(% treated WP)</td>
<td>-1.27</td>
<td>0.35</td>
</tr>
</tbody>
</table>

Max. Observed Effect is the range of observed effects on BIBI. For linear variables, this is the range of $\beta \cdot x$. For polynomial variables, this is the range of $(\beta_1 \cdot x + \ldots + \beta_p \cdot x^p)$, evaluated for all sites.

Impervious surfaces have a large and statistically significant negative effect on BIBI (Table 25). As impervious surfaces increase from 0% to approximately 30% of the catchment area, BIBI falls by as much as 2.3 points. Through this range, the marginal effect on BIBI of adding impervious surfaces decreases. The model also predicts a positive marginal effect as impervious surfaces increase beyond 30%, but the estimated curve diverges from the data points in this region (appendix).

Tree canopy (and by correlation, forest cover), have a statistically significant, positive effect on BIBI. BIBI scores increase by 0.211 points for each 10% step in tree
canopy cover (Table 25). This effect is large enough to alter clean water act listing categories. Riparian canopy appears to have a significant, positive effect on BIBI. In the final model, as riparian canopy cover rises from 0% to approximately 50%, average BIBI increases by 0.61 BIBI points (Table 25). However, the marginal improvement of adding riparian canopy diminishes through this range, with effect reaching a maximum at approximately 50%. As riparian canopy increases past 50%, adding riparian canopy has a negative marginal effect on BIBI. For sites with 100% riparian canopy, there appears to be no meaningful improvement in BIBI. Reach canopy has a small negative effect on BIBI, which is not statistically significant (Table 25). Although both of reach canopy’s beta coefficients are significantly different from 0 at alpha < 0.05, the 95% confidence interval for the effect size crosses zero.

Filtration SWM facilities have a significant, positive effect on BIBI. In contrast, infiltration facilities appear to have no effect on BIBI at all: the effect size is very close to zero. Dry and Extended Detention Ponds (DDP and EDP) have a positive effect on BIBI, and this effect is statistically significant at alpha ≤ 0.20. In the agriculture model, these variables have reduced significance, and EDP has a large reduction in effect size. Wetpond facilities had a significant, negative effect on BIBI. There are three sites with dramatically high levels of wetponds. However, fitting the model without these sites did not change the sign or significance of the wetpond beta coefficient, rather it increased them (β = -1.53, SE = 0.43).

Adding agriculture to this model resulted in a statistically insignificant coefficient for agriculture (β = 0.203, p = 0.700). The coefficients for impervious surface and
canopy were not significantly altered by the addition of agriculture (canopy $\beta = 2.31, p = 0.002$: impervious $\beta = -13.84, p < 0.001$: impervious$^2$ $\beta = -21.77, p < 0.001$).

I now turn to the models that replaced canopy with agriculture (“agriculture models”). In the agriculture model with highest likelihood, agriculture had a significant negative effect on BIBI, the effect size of impervious surfaces was larger, and the effects of DDP and EDP SWM are reduced in magnitude and statistical significance (Table 26, compare Table 25).

**Table 26.** Effects of individual variables on BIBI: agriculture model

<table>
<thead>
<tr>
<th>Predictor Variable</th>
<th>$\beta$</th>
<th>SE</th>
<th>$p$-value</th>
<th>95% CI</th>
<th>$\Delta$BIBI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>4.15</td>
<td>0.35</td>
<td>&lt; 0.01</td>
<td>3.47, 4.84</td>
<td></td>
</tr>
<tr>
<td>% impervious</td>
<td>-17.67</td>
<td>1.47</td>
<td>&lt; 0.01</td>
<td>-20.57, -14.77</td>
<td>-3.03, -3.49, -2.57</td>
</tr>
<tr>
<td>% impervious$^2$</td>
<td>25.76</td>
<td>2.31</td>
<td>&lt; 0.01</td>
<td>21.21, 30.30</td>
<td></td>
</tr>
<tr>
<td>% agriculture</td>
<td>-0.95</td>
<td>0.36</td>
<td>0.01</td>
<td>-1.67, -0.24</td>
<td>-0.88, -1.54, -0.22</td>
</tr>
<tr>
<td>% riparian canopy</td>
<td>1.65</td>
<td>0.95</td>
<td>0.09</td>
<td>-0.23, 3.53</td>
<td>0.59, 0.22, 1.11</td>
</tr>
<tr>
<td>% riparian canopy$^2$</td>
<td>-1.15</td>
<td>0.83</td>
<td>0.17</td>
<td>-2.79, 0.49</td>
<td></td>
</tr>
<tr>
<td>sin$^{-1}$(% reach canopy)</td>
<td>-0.67</td>
<td>0.38</td>
<td>0.08</td>
<td>-1.43, 0.08</td>
<td>-0.26, -0.58, 0.11</td>
</tr>
<tr>
<td>sin$^{-1}$(% reach canopy)$^2$</td>
<td>0.44</td>
<td>0.22</td>
<td>0.05</td>
<td>0.00, 0.87</td>
<td></td>
</tr>
<tr>
<td>$\sqrt{(% treated DDP)}$</td>
<td>0.34</td>
<td>0.29</td>
<td>0.24</td>
<td>-0.23, 0.90</td>
<td>0.32, -0.22, 0.86</td>
</tr>
<tr>
<td>$\sqrt{(% treated EDP)}$</td>
<td>0.10</td>
<td>0.33</td>
<td>0.77</td>
<td>-0.56, 0.75</td>
<td>0.08, -0.44, 0.60</td>
</tr>
<tr>
<td>$\sqrt{(% treated Filt)}$</td>
<td>1.49</td>
<td>0.66</td>
<td>0.02</td>
<td>0.20, 2.79</td>
<td>0.43, 0.06, 0.80</td>
</tr>
<tr>
<td>$\sqrt{(% treated Inf)}$</td>
<td>0.10</td>
<td>0.67</td>
<td>0.88</td>
<td>-1.21, 1.41</td>
<td>0.04, -0.44, 0.51</td>
</tr>
<tr>
<td>$\sqrt{(% treated WP)}$</td>
<td>-1.31</td>
<td>0.35</td>
<td>&lt; 0.01</td>
<td>-1.99, -0.63</td>
<td>-1.27, -1.93, -0.61</td>
</tr>
</tbody>
</table>

**Notes:** Max. Observed Effect is the range of observed effects on BIBI. For linear variables, this is the range of $\beta \cdot x$. For polynomial variables, this is the range of $(\beta_1 + \beta_2 x + \ldots + \beta_p x^p)$, evaluated for all sites.

Information criteria and likelihood evidence strongly favor models with canopy over models with agriculture. In models with both tree canopy and agriculture, agriculture does not have a significant effect, and these models offer no improvements in AICc or accuracy over models with canopy alone. For these reasons, the remainder of the research utilized the canopy model described in Table 25.
Statistically significant variation in residual variance was identified, and is best explained by using impervious surface and predicted values as variance-covariates. The fitted variance-covariate function is described by Equation 6. This function describes the distribution of model residuals. Figure 24 illustrates the relationship of the variance covariates to the residual variance.

\[ \varepsilon_i = N\left(0, \sigma^2 \times e^{2x\cdot-3.772\cdot imp_i} \times e^{2x\cdot0.744\cdot \hat{Y}_i} \times e^{2x\cdot-0.186\cdot \hat{Y}_i^2}\right) \]  

where \( imp = \text{Impervious} \) and \( \hat{Y} = \text{fitted values} \).

**Figure 24.** Scatterplots illustrating the estimated variance covariate function

Statistically significant autocorrelation of model residuals was identified, and controlled using an error correlation matrix corresponding to an exponential variogram model. Figure 25 illustrates the variogram model describing the pattern of spatial autocorrelation in the residuals. Autocorrelation was strongest among sites located < 2,000 meters apart.
Figure 25. Semi-variogram of model residuals, with the estimated variogram model
Range = 2,093 m, effective range* = 6,278 m, and nugget = 0.662
Variogram model is isotropic.
* Effective range of exponential variogram is 3*Range (Pebesma 2001)

The map of kriged residuals (Figure 26) illustrates the spatial pattern of over- and under-prediction of BIBI described by the fitted variogram. This is the local component of the error term, and is used in RK to adjust predictions from the fixed terms of the final model. Kriging variance is illustrated in a map (Figure 26) and illustrates the precision of the kriged residuals. Kriging variance is an estimate of the mean square error of the kriged residual. 95% confidence intervals can be estimated for kriged residuals as +/- 2*sqrt(kriging variance), and these intervals range from +/- 1.1 to +/- 1.3 BIBI points for out model.
**Figure 26.** Filled-contour maps of kriged residuals and kriging variance

**Model Performance**
Predictions of BIBI

The model used land cover, SWM, and the autocorrelation among study sites to explain more than 60% of the variability in BIBI scores (Table 27). Pseudo $R^2$ measures were similar for the fitted model and evaluation RK predictions. While this is good performance, it also indicates that further variation in BIBI might be explained by other predictor variables or more efficient statistical models. The agriculture and canopy models exhibited similar pseudo $R^2$ measures.

<table>
<thead>
<tr>
<th>Model</th>
<th>Sample</th>
<th>Coef of Determ</th>
<th>% Var Explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canopy</td>
<td>Fitted Model</td>
<td>61.9%</td>
<td>61.9%</td>
</tr>
<tr>
<td>Canopy</td>
<td>Evaluation Predictions</td>
<td>63.0%</td>
<td>60.0%</td>
</tr>
<tr>
<td>Canopy</td>
<td>Change (&quot;Shrinkage&quot;)</td>
<td>1.8%</td>
<td>-2.3%</td>
</tr>
<tr>
<td>Ag</td>
<td>Fitted Model</td>
<td>61.4%</td>
<td>61.2%</td>
</tr>
<tr>
<td>Ag</td>
<td>Evaluation Predictions</td>
<td>64.9%</td>
<td>61.8%</td>
</tr>
<tr>
<td>Ag</td>
<td>Change (&quot;Shrinkage&quot;)</td>
<td>5.6%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

Table 27. Pseudo $R^2$ of final model

Table 28 presents a variety of accuracy and precision measurements for the model’s predictions. Predictions for the fitting dataset were more accurate than predictions for the evaluation data. Mean absolute error (MAE) increased more than root mean square error, indicating that while the evaluation errors were on average larger, there were fewer large errors in the evaluation predictions. The evaluation predictions were noticeably biased: mean and median errors were substantially greater than zero, particularly in comparison to the fitting data predictions.

Table 28. Accuracy and precision of model predictions
Notes: RMSE is root mean squared error. MAE is mean absolute error. Pr(|ε| ≤ 0.5) is the proportion of absolute errors less than or equal to 0.5 BIBI points. 80th P(|ε|) is the 80th percentile of absolute error. 95th P(|ε|) is the 95th percentile of absolute error.

In the fitting data, 61.4% of predictions had error less than +/- 0.5 BIBI points. Only 47.4% of evaluation predictions met this level of accuracy. Eighty percent of predictions had error less than +/- 0.77 and +/- 0.81 BIBI points for fitting and evaluation predictions, respectively. Precision of predictions was consistent between the fitting and evaluation predictions. Standard deviation from the mean error is ~0.65 BIBI points.

Predictions using the agriculture and canopy models showed similar patterns of accuracy and precision. Predictions on the evaluation sample using the agriculture model were slightly more accurate and precise (Table 28).

Predicting Regulatory Categories Using Naïve Cutoffs

Confusion matrices for categorical predictions are found in the appendix. Table 29 shows that overall accuracy for 3-category predictions is low, between 50% and 65%. Kappa statistics show that these results are 30% to 45% more accurate than random guessing (Table 29). Prediction accuracies show large variations among the categories. There is substantial variation among fitting and evaluation predictions, and between user’s and producer’s accuracy measures. Accuracies are particularly low and erratic for predictions of high quality stream health, ranging from 8.0% to 92.3%.

Table 29. Classification accuracy for 3-category predictions using naïve cutoffs
In contrast, Table 30 demonstrates high accuracy for predicting two categories of stream health (impaired and passing). Overall accuracies exceed 80%, and Kappa indicates these results are about 70% better than random guessing (Table 30). Accuracy measurements are consistently near 80% between the fitting and evaluation predictions, among predictions of impaired and passing stream health, and between user’s and producer’s measures.

**Table 30. Classification accuracy for predicting impaired and passing stream health, using naïve cutoffs**
Predicting Regulatory Categories Using ROC

ROC graphs (Figure 27) illustrate that the model can be used to separate binary regulatory categories with reasonable accuracy. However, they also illustrate that naively applying the regulatory thresholds for BIBI scores to the model predictions may not produce desirable or consistent error rates.

![ROC Graphs](image)

**Figure 27.** Receiver operating characteristic curves for predicting impairment. ROC curves are labeled with the cutoffs corresponding to the indicated error rates. The diagonal line is the classification accuracy curve, using naïve cutoff (Impaired = BIBI < 3).
ROC of random guessing. Dashed lines highlight the error rates observed if the regulatory cutoff is used naively. These charts can be used to find cutoffs achieving desired error rates.

In the ROC graphs in Figure 27, the hypothesis being tested is that a stream is not impaired. The null hypothesis is that the stream is impaired. Type II error is the probability of a false positive, or the probability of erroneously accepting the conclusion that a stream is not impaired. This is the same as predicting that the stream is not impaired when it actually is. Type I error is the probability if a false negative. This is the same as erroneously concluding that the stream is impaired. By choosing cutoffs with low Type II errors, regulators can support a precautionary policy that is very protective of stream health. A contrasting policy would choose cutoffs with low Type I errors, so that regulatory actions have a high level of evidence supporting them. I refer to this as a permissive policy.

The ROC curve for predicting poor stream health using the fitting predictions (Figure 27) indicates that Type II error rate of 5% is attained when the cutoff for poor stream health is set at predicted BIBI = 3.437. An agency using the precautionary goal would deny a project when the model predicts BIBI ≤ 3.437. Applying this cutoff to the evaluation predictions, we see that the Type II error rates are similar to those for the fitting predictions (Table 31). This suggests that the model could be relied upon to support this precautionary development review approach.

Table 31. Error rates observed when ROC-informed cutoffs are used to predict impaired stream health
Type I errors are false positive predictions, where stream health is predicted to be impaired but is in reality acceptable. Using the permissive policy goal, an agency would deny a project when the model predicts $\text{BIBI} \leq 2.908$ (Figure 28). The Type I error rates for this cutoff in the evaluation predictions are higher than the fitting predictions (Table 31). This inconsistency between fitting and evaluation predictions suggests we cannot rely upon the model to support this permissive development review approach.

The same evaluation is applied for a decision when the development project might degrade a high quality stream. Figure 28 presents ROC graphs for the hypothesis that a stream is not high quality. Type II error is the probability of erroneously accepting the conclusion that a stream is not high quality. Committing a type II error in this case is predicting less than high quality when the quality is in fact high. Type I error is the probability that we predict high quality when the stream is not high quality. As before, greater protection for stream health comes by limiting Type II errors. The permissive policy seeks to limit Type I errors.

<table>
<thead>
<tr>
<th>Cutoff Goal</th>
<th>Cutoff</th>
<th>Type I Error</th>
<th>Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive (match regulations)</td>
<td>3.000</td>
<td>6.4%  3.1%, 11.4%</td>
<td>22.1%, 15.4%, 30.0%</td>
</tr>
<tr>
<td>Precautionary: $\beta \leq 5%$</td>
<td>3.437</td>
<td>46.5% 38.5%, 54.6%</td>
<td>4.4%, 1.6%, 9.4%</td>
</tr>
<tr>
<td>Permissive: $\alpha \leq 5%$</td>
<td>2.908</td>
<td>4.5% 1.8%, 9.0%</td>
<td>13.1%, 6.7%, 22.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cutoff Goal</th>
<th>Cutoff</th>
<th>Type I Error</th>
<th>Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive (match regulations)</td>
<td>3.000</td>
<td>6.4%  3.1%, 11.4%</td>
<td>22.1%, 15.4%, 30.0%</td>
</tr>
<tr>
<td>Precautionary: $\beta \leq 5%$</td>
<td>3.437</td>
<td>46.5% 38.5%, 54.6%</td>
<td>4.4%, 1.6%, 9.4%</td>
</tr>
<tr>
<td>Permissive: $\alpha \leq 5%$</td>
<td>2.908</td>
<td>4.5% 1.8%, 9.0%</td>
<td>13.1%, 6.7%, 22.2%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cutoff Goal</th>
<th>Cutoff</th>
<th>Type I Error</th>
<th>Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naive (match regulations)</td>
<td>3.000</td>
<td>6.4%  3.1%, 11.4%</td>
<td>22.1%, 15.4%, 30.0%</td>
</tr>
<tr>
<td>Precautionary: $\beta \leq 5%$</td>
<td>3.437</td>
<td>46.5% 38.5%, 54.6%</td>
<td>4.4%, 1.6%, 9.4%</td>
</tr>
<tr>
<td>Permissive: $\alpha \leq 5%$</td>
<td>2.908</td>
<td>4.5% 1.8%, 9.0%</td>
<td>13.1%, 6.7%, 22.2%</td>
</tr>
</tbody>
</table>
Figure 28. Receiver Operating Characteristics (ROC) curves for predicting high quality ROC curves are labeled with the cutoffs corresponding to the indicated error rates. The diagonal line is the ROC of random guessing. Dashed lines highlight the error rates observed if the regulatory cutoff is used naïvely. These charts can be used to find cutoffs achieving desired error rates.

Here, the precautionary cutoff is predicted BIBI = 3.264, and if the project would cause the model to predict BIBI ≤ 3.264, we would conclude there is a > 5% chance that the project would degrade the high quality stream below the regulatory threshold for high quality streams (Figure 28). This cutoff does not perform well in the evaluation predictions (Table 32). The permissive cutoff is predicted BIBI = 3.732, and the fitting and evaluation predictions have similar error rates. Based on these performance figures, the model could be used to support a permissive review process for high quality streams, but should not be used to support a precautionary review process.

Table 32. Error rates observed when ROC-informed cutoffs are used to predict high quality stream health
### Predicting High Quality Conditions (BIBI ≥ 4)

<table>
<thead>
<tr>
<th>Cutoff Goal</th>
<th>Cutoff</th>
<th>Population</th>
<th>Type I Error</th>
<th>Type II Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve (match regulations)</td>
<td>4.000</td>
<td>Fitting</td>
<td>0.5%</td>
<td>84.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Evaluation</td>
<td>0.0%</td>
<td>74.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.6%</td>
<td>91.7%</td>
</tr>
<tr>
<td>Precautionary:</td>
<td>3.264</td>
<td>Fitting</td>
<td>33.3%</td>
<td>3.9%</td>
</tr>
<tr>
<td>$\beta \leq 5%$</td>
<td></td>
<td>Evaluation</td>
<td>27.1%</td>
<td>0.8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>40.0%</td>
<td>11.0%</td>
</tr>
<tr>
<td>Permissive:</td>
<td>3.732</td>
<td>Fitting</td>
<td>4.6%</td>
<td>72.7%</td>
</tr>
<tr>
<td>$\alpha \leq 5%$</td>
<td></td>
<td>Evaluation</td>
<td>2.2%</td>
<td>61.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.3%</td>
<td>82.3%</td>
</tr>
</tbody>
</table>

- Precautionary: $\beta \leq 5\%$
- Permissive: $\alpha \leq 5\%$

#### Notes:
- Precautionary: $\beta \leq 5\%$
- Permissive: $\alpha \leq 5\%$

#### Naïve (match regulations):
- $\alpha$ = 95% CI
- $\beta \leq 5\%$

#### Permissive:
- $\alpha$ = 95% CI

#### Naïve (match regulations):
- $\alpha$ = 95% CI
- $\beta \leq 5\%$

#### Evaluation:
- Precautionary: $\beta \leq 5\%$
- Permissive: $\alpha \leq 5\%$

#### Precautionary: $\beta \leq 5\%$
- $\alpha$ = 95% CI
- $\beta \leq 5\%$

#### Permissive: $\alpha \leq 5\%$
- $\alpha$ = 95% CI
- $\beta \leq 5\%$
CHAPTER 5
DISCUSSION

The analysis produced a model that predicts the effects of land cover and stormwater management (SWM) on the benthic index of biotic integrity (BIBI). This model and the model comparisons used to construct it provide new insight into the relationships between land cover, SWM, and stream health. What do the results mean for environmental managers? Does the model perform well enough to support policy makers and proactive planning?

Effects of Land Cover on Stream Health

Impervious Surfaces Land Cover

Impervious surfaces are clearly an important predictor of stream health (Table 10) and we found strong information criteria evidence favoring a quadratic over a linear relationship to BIBI (Table 10). Impervious surfaces have a large and statistically significant negative effect on stream health (Table 25 and Table 26).

BIBI was calibrated against reference sites along a gradient of urban and forest land covers (Southerland et al. 2005, 4-6). Therefore, the strong negative relationship between BIBI and impervious surfaces is not an important finding: it is almost a tautology. However, the shape of the relationship and estimates of effect sizes are of interest.

In the final model, BIBI falls rapidly as impervious surfaces increase from 0%. The rate of BIBI loss decreases as the level of imperviousness increases, reaching a slope
of 0 near 30% imperviousness. As imperviousness increased, the precision of BIBI predictions increased, with a similar curvature. These results are quite similar to the model of impervious cover and stream quality recently proposed by Schueler, Fraley-McNeal, and Cappiella (2009).

When imperviousness increases beyond 30%, the model predicts that BIBI will also increase. This shift to a positive slope is unexpected. It may be a statistical artifact of the quadratic relationship used in the modeling, combined with the small proportion of samples with >30% impervious surfaces. Note in particular the discrepancies between the model’s quadratic function and the loess smoother curve in the partial residual plot (Appendix C): the loess curve suggests a slope near zero past 30% impervious, not a hyperbolic slope. Testing a range of other relationship forms, for example segmented regression or smoothing splines, might identify more tenable relationship forms, and could result in a more accurate and precise model.

Agricultural Land Cover

Agricultural land cover appears to be an important predictor of stream health. Models with agricultural land cover were superior to models with only impervious surfaces (Table 11A). The form of agriculture’s relationship to BIBI remains uncertain: small-sample corrected Akaike information criteria (AICc) was nearly identical for linear, quadratic, and quadratic with impervious interaction (Table 11A).

Agriculture has a negative correlation with BIBI after controlling for impervious surfaces, riparian tree canopy, and SWM (Table 26). However, when catchment scale tree canopy is included in the model, agriculture has no statistically significant effect on BIBI. Moreover, the data provide better support (lower AICc) for models with forest or
canopy than models with agriculture (Table 11D). This suggests that the effects of agriculture on BIBI are not as strong as the effects of tree canopy, and that correlations observed have more to do with agriculture displacing tree canopy, and less to do with any water quality effects unique to agriculture.

Forest Land Cover

Forest land cover appears to be an important predictor of stream health: models with forest were dramatically better than models without forest land cover (Table 11B), and models with forest or tree canopy have nearly identical support from the data (Table 11D). Information criteria show that a linear relationship is better supported than a quadratic, or an interaction with impervious surfaces, although the AICc evidence is weak (Table 11B). Because tree canopy cover was selected over forest land cover for inclusion in the final model, we cannot interpret beta coefficients to make conclusions regarding the direction and magnitude of forest’s effects on stream health. However, given the strong linear correlation between forest and tree canopy land covers, it is not unreasonable to assume that forest land cover has similar direction and strength to tree canopy.

Tree Canopy Cover

Tree canopy cover is also an important predictor of stream health (Table 11C). A linear relationship with BIBI is best supported by the data: AICc provides moderate evidence for this (Table 11D). Tree canopy (and by correlation, forest cover), have a large, statistically significant, positive effect on BIBI (Table 25 and Table 26). We did not identify a statistically important interaction with impervious surfaces, so this effect appears to be independent of the level of impervious surfaces in the catchment.
Riparian Canopy Cover

The results indicate uncertainty regarding the importance of riparian tree canopy cover: the model without riparian canopy is AICc equivalent to the model with quadratic terms for riparian canopy (Table 13). The data provide more support for models lacking an interaction between riparian canopy and impervious surfaces (Table 13).

A surprising result is that average BIBI is maximized at riparian canopy of 50%, and BIBI declines with riparian canopy increases beyond this level (Appendix C). This result should be interpreted with caution, for two reasons. First, AICc evidence for including riparian canopy in the model is weak. Second, few sites with less than 50% riparian canopy were sampled (30 of 293, 10.2%). Both these facts suggest that details of the observed relationship might be statistical anomalies.

Results show that riparian buffer vegetation does not have a large effect on stream health (Table 26, Table 26). This result seems to support recent theories of urban stream (dys)function (Groffman et al. 2003). There is a growing body of evidence in support of this conclusion (Goetz et al. 2003; Baker, Weller, and Jordan 2006; Walsh et al. 2007; Schueler, Fraley-McNeal, and Cappiella 2009), but other studies have contradictory results (e.g., Moore and Palmer 2005; Snyder, Goetz, and Wright 2005).

Reach Canopy Cover

There is uncertainty regarding the importance of reach canopy coverage as well. Models without reach canopy and with quadratic terms for reach canopy are nearly AICc equivalent (Table 14). The data provide substantially more support for models without interactions between reach canopy and impervious surfaces (Table 14). Reach canopy was included in further model fitting to meet research goals. In the final fitted model,
reach canopy has a small negative effect on BIBI, which is not statistically significant (Table 25 and Table 26).

**Effects of Stormwater Management on Stream Health**

It is apparent that SWM best management practices (BMPs) are an important predictor of stream health. Models with different categories of SWM are more likely than models without SWM, or models where all SWM BMPs are combined in a single variable (Table 12). There is ambivalence between models using SWM by type of facility, and by mechanism of facility: AICc indicates more support for SWM by Mechanism, but SWM by Type has higher likelihood (Table 12).

The effect of SWM BMPs on BIBI varies by SWM type (Table 25 and Table 26). In the agriculture model, these variables have reduced significance, and extended detention pond (EDP) treatment has a large reduction in effect size. Dry detention pond (DDP) and EDP SWM exhibit moderate negative correlation with agriculture (Table 6). These results suggest that the observed effects of DDP and EDP SWM and agriculture are not fully independent. Some portion of the effect attributed to agriculture may be due to reduced prevalence of DDP and EDP SWM, and vice-versa.

The result that wetpond facilities have a negative effect on stream health is unexpected. Subsequent discussions with staff at Baltimore County revealed that SWM BMP selection is structured by surface water use designations. The designations are defined and assigned to particular stream segments in the Code of Maryland Regulations (COMAR: Maryland 2011a, Maryland 2011b). These use designations include Use III and III-P, which “shall be given the additional protection required for […] natural trout waters” (Maryland 2011a). SWM BMPs that increase stream water temperatures, e.g.,
wetponds (Galli 1990), are actively discouraged in such streams (Maryland Department of the Environment (MDE) 2009b). The MDE has created maps to help communicate these regulations (MDE 2010). Comparison of SWM data with these maps shows wetpond BMPs are indeed rare in Use III watersheds.

Trout and other cold water fishes are present only in streams of high quality and good health. Therefore, wetpond SWM BMP facilities are deployed primarily in the watersheds of streams that had poor health before the SWM facility was constructed. Stream health, through environmental governance, causes the distribution and abundance of wetpond SWM facilities. This likely explains much of the negative correlation between the area treated by wetpond SWM facilities and stream health.

However, the metrics used in the BIBI are sensitive to stream temperature regime changes (e.g., species richness, Ephemeroptera, Plecoptera, and Trichoptera (Vannote and Sweeney 1980; Sponseller, Benfield, and Valett 2001)). Therefore, wetponds, through thermal pollution, can conversely cause BIBI scores to fall (Galli 1990).

Thus, the regression results are not sufficient to estimate the impact on BIBI of installing wetponds. The results are confounded: the regression coefficient describes an unknown mixture of the effect of wetponds on BIBI, and the effect of stream health on wetponds.

A corollary to this explanation is that higher quality streams cause installation of SWM BMPs other than wetponds in urbanizing watersheds. This threatens the validity of drawing conclusions about the effects of other SWM facility types on stream health. The causal mechanism (high stream health causes BMPs other than wetponds to be installed when urban land is developed) should manifest as a categorical effect on stream health:
the presence of BMPs other than wetponds should be correlated with higher stream health.

The mechanism theorized above would not explain continuous effects, where increasing amounts of BMPs are correlated with continuous changes in stream health. Therefore, any continuous effects observed in the model may be a combination of a categorical effect (stream health, through environmental governance, determining the diversity of SWM facilities) and a continuous effect (amount of SWM treatment, through environmental alteration, affecting stream health). The model examined only a continuous effect, not a categorical effect, and found large, statistically significant continuous effects of BMPs on BIBI (Table 25, Table 26). Therefore, I cautiously conclude that these results do reflect the impact of the amount of SWM BMPs on stream health, rather than the mere presence or absence of SWM BMPs. However, future studies should address this question more directly, by comparing the quality of models with and without both categorical and continuous effects from SWM BMPs, in a population of sites that all have some SWM.

The results for SWM BMPs do not match the *a priori* hypothesis regarding mechanism of action. This is due in part to the wetpond issue discussed above: including wetponds with infiltration devices confounds the analysis of SWM-by-mechanism, because the correlation between wetponds and BIBI is due not to the water quality mechanisms of wetponds but rather due to active discouragement of wetponds in the watersheds of high quality streams by the Maryland Stormwater Design manual. The question of mechanism should be revisited, with wetponds treated separately from questions of SWM BMP mechanism. The interdependence of agriculture and DDP/EDP
SWM effects on BIBI should be investigated. Further study should explore and evaluate causal mechanisms and/or theoretical expectations for SWM BMP effects on stream health.

**Suitability of Model for Applied Use**

**Continuous BIBI Predictions**

**Model Accuracy**

It was hoped that the model would achieve an accuracy of +/- 0.5 BIBI points with ≥ 80% confidence. However, more than 50% of predictions had errors larger than +/- 0.5 BIBI. The median bias of the regression kriging (RK) predictions (+0.3 BIBI points) is large enough to be meaningful for stream ecology and regulations (Table 28). The model can be recommended for applications tolerant of these biases and errors. Addressing the temporal non-stationarity and other issues discussed later in this chapter might improve prediction accuracy and precision.

**Model Precision**

We know that the prediction variance varies across space (map of kriging variance, Figure 26), by amount of impervious surfaces (Figure 24), by fitted BIBI value (Figure 24), by watershed (Table 17, Table 18, Table 19, and Table 20), and across the range of predictor variables (Hengl, Heuvelink, and Stein 2004). Combining all these variance structures to construct prediction confidence intervals is not trivial and is beyond the scope of this study. If accurate measurements of prediction variance are needed, further study with a statistician is recommended. However, we can get a sense for levels of precision by examining the variance-covariate functions.
Prediction variance tends to decrease as impervious surfaces increase. This is particularly true when impervious surfaces are greater than approximately 15% (Figure 24). Model precision is likely to be lowest in the least urbanized areas. Prediction variance is shown to be maximized at middle values of BIBI, and to have minima at extremely low and high BIBI scores. This is likely due to the limits of BIBI at 1 and 5.

Finally, model validation showed that watersheds are a covariate for prediction error variance. Heterogeneity among watersheds results in differences in the precision of the model predictions. Identifying what differences among watersheds generate this pattern may prove difficult, as it is not trivial to separate the effect of watershed from the effect of other predictors that vary among watersheds. Indeed, we did not include watersheds in the model fitting effort out of concern for variance inflation effects and small sample sizes.

Regulatory Category Predictions

Despite bias in predictions of BIBI, predictions of impaired versus passing stream health are highly accurate and consistent. Accuracy measures are consistently near or above 80%, in both the fitting and evaluation predictions. This level of accuracy should be suitable for many applied purposes. In contrast, predictions of 3-category stream health categories were accurate less than 65% of the time, and accuracies were inconsistent. Prediction of high quality stream health was particularly inaccurate. These categorical accuracies correspond to the pattern of bias observed in the BIBI predictions: bias was at a minimum near BIBI = 3, which is the cutoff for impaired and fair quality. Under-prediction was largest near BIBI = 4, which is the cutoff for fair and high quality (Figure 28).
Use of receiver operating characteristics (ROC) curves to select cutoffs to support a precautionary policy for preventing impairment was successful: false positive and false negative predictions were maintained at acceptable levels in both the fitting and evaluation samples. Stable results were also achieved after selecting cutoffs for a permissive policy on high quality stream degradation.

Notably, the precautionary cutoffs for predicting impaired stream health and high quality stream health are incompatible. The impaired cutoff is higher than the high quality cutoff, meaning that some sites would be predicted as both impaired and high quality. The model’s predictions are not precise enough to support this precautionary policy scenario for both regulatory categories.

The ROC analysis is clearly useful when policy goals require the control of errors in binary classifications. Potential users of this model should carefully consider what costs should be assigned to different types of errors, choose a cutoff that results in the lowest total error cost, and evaluate the stability of the cutoff by testing it against evaluation data.

**Limitations of Final Model**

The results suggest that some of the assumptions of GLS regression are violated to varying degrees. This is not surprising given the complexity of biogeographical systems. Careful attention to deviations from GLS assumptions can provide further information about the relationships between stream health, land cover, and SWM, and suggest avenues for further work.

Residuals and RK prediction errors exhibited a slight curvilinear relationship with fitted and predicted BIBI, indicating bias in predicted BIBI. This sort of bias can arise
when there are missing fixed terms, or the fixed terms do not adequately model non-linearities in the relationships between the predictor and dependent variables (Osborne and Waters 2002). While the pattern of bias is not particularly strong, it does appear to impact other GLS assumptions (as discussed below), and indicates that the model could be improved.

Model residuals are normally distributed with a mean of zero. In contrast, the RK prediction errors have a skewed distribution and the mean error is greater than zero (Table 28 and Figure 18). Some of this bias is created by the non-linearity (described above) interacting with a difference in BIBI distribution between the fitting and evaluation samples. The evaluation sample contains a larger proportion of observations with BIBI > 3 (58.5%) than the fitting sample (47.4%). The model tends to underpredict BIBI when true BIBI > 3, and so the errors from the fitting sample should be expected to have a positive bias. This may indicate other issues, including: over-fitting, misspecification of predictor variables, or misspecification of the variogram model used to krig the residuals.

RK prediction errors exhibited spatial autocorrelation, in contrast with the residuals from the fitted model (Figure 22). Similarly, there are different patterns of clustering in residuals and RK errors (Figure 23). This suggests one or more of the following: the spatial autocorrelation structure is misspecified, the sampling is too sparse or clustered for accurate modeling of spatial dependence throughout the study area, or there is temporal variation in the processes generating spatial dependence of BIBI among proximal stream reaches. The kriging variance map (Figure 26) highlights several regions where sampling was too sparse to create a precise kriged residual surface. The
model may be accurate for the fitting time period, or may be biased due to a misspecified spatial correlation structure. Model coefficients should be interpreted cautiously.

Residuals from the fitted model were significantly structured by year (Table 21). The effect of year was small (Table 22), but could affect management decisions. This temporal non-stationarity may indicate model misspecification, or reflect normal temporal variation in BIBI. As the number of years available for study is limited, it would be difficult to represent year as a continuous variable in the model. At the same time, a factor for year is not desirable for environmental management applications. A random effects term could be appropriate, but due to the geographic stratification of sampling by year and the strong spatial trends in both BIBI and the predictor variables, this would interfere with fixed term estimation. A simpler approach might be to examine the sentinel site data for temporal trends. If a trend is found, the BIBI scores can be “annually adjusted” to account for inter-annual variability.

Residuals were further structured by a global spatial trend, increasing from South to North. This spatial trend can be viewed as a spatial expression of the relationship between fitted value and error: BIBI values > 3 are more prevalent in the North than in the South of the study area, and so under-prediction should be more prevalent in the North. Such trends may result in biased coefficients, particularly when they accompany strong spatial autocorrelation in the dependent and predictor variables. Therefore, model coefficients should be interpreted with caution.

Residual mean and variance were also structured by watersheds (Table 19, Table 24). Watershed effects in the RK errors were slightly larger than the effects of year on the fitting residuals, and are large enough to affect management outcomes (Table 24). As
factors, watersheds may structure the variables and thus interact with the nonlinearity discussed above. As hydrologic and biogeographic regions, watersheds may present different relationships between environmental variables and stream health. Watersheds are distributed through space, and thus spatial trends and autocorrelation can create or be created by patterns among watersheds. Untangling the relationship of watershed regional non-stationarity with the other forms of non-stationarity is beyond the scope of this study.
CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

By using the methods of generalized least squares (GLS), regression kriging (RK), and receiver operating characteristics (ROC), I was able to construct an accurate and useful model from “messy” real-world data. The model is capable of supporting some regulatory applications. This research explored the important relationships between stream health and stormwater management (SWM). Careful study of environmental governance and access to SWM administrative records generates important insights into these relationships. The results contribute to the growing body of knowledge regarding land cover and stream health by demonstrating the minor effect of riparian forest and confirming a common model of impervious. Further work can yield models that are more accurate, and provide a yet deeper understanding of the relationships between land cover, SWM, and stream health.

Land Cover and Stream Health

This study adds to the large body of evidence indicating that catchment scale impervious surfaces and forest land cover have large effects on stream health. The results confirm a non-linear relationship of impervious surfaces to stream health. The variance-covariate results provide formal empirical confirmation of a well-established idea: the precision and certainty of stream health degradation increases as impervious surfaces increase.

Results do not provide evidence that interaction between impervious surfaces and
other land covers or SWM impact stream health. However, model validation shows that nonlinear responses remain in the final model, and missing interaction terms might account for this. It is possible that interaction effects exist, and the sampling strategy and statistical methods were not sensitive enough to detect them. Further investigation of impervious surface interactions might begin with identification of a study area possessing a more homogeneous riverscape context and sufficient variation in impervious surfaces and the covariates of interest.

This study found that riparian tree canopy has only a small effect on the health of urban and agricultural streams in the Piedmont region of Baltimore County, Maryland. This is an area of uncertainty in the field. Recent work developing alternative means of measuring stream buffers and their effects on stream health may provide evidence and methods to clarify the situation (e.g., King et al. 2005; Baker, Weller, and Jordan 2006; Goetz 2006; Goetz and Fisk 2008; Van Sickle and Johnson 2008).

Agricultural land cover was shown to correlate with declining stream health, but the observed effect appears to be caused by the displacement of tree canopy land cover by agriculture, and not due to any unique water quality impacts from agriculture. Further study of agriculture and stream health is needed.

**Stormwater Management and Stream Health**

SWM was shown to have several impacts on stream health. At a very basic level, SWM practices are shown to alter the natural boundaries of catchments, expanding and contracting riverscapes. This may alter their proportional land covers, a topic for further study. Without accurate and complete stormwater management and storm drain geodatabases, measurement of the riverscape variables is imprecise. This reduces our
ability to understand and predict stream health. Further development, integration, and maintenance of SWM tracking systems will contribute to our ability to understand and manage stream health in urban riverscapes.

SWM facilities have significant and unique correlations with stream health. The results suggest that filtering SWM facilities improve stream health, and other SWM types create smaller stream health improvements.

The relationship between wetpond facilities and stream health is particularly strong, and is the product of coupled human-environment interactions: wetponds act to degrade stream health, and stormwater management regulations act to restrict wetponds in the catchments of healthy streams while encouraging wetponds in the catchments of degraded streams. Sensitivity of researchers to such interactions is vital when designing studies of stream health. As a precautionary example, this study attempted to evaluate the effects of SWM treatment mechanisms, but results were confounded by this unanticipated human-environment interaction.

**Predictive Modeling and Decision Making**

The accuracy of the models predictions of the benthic index of biotic integrity (BIBI) could be improved. Regression kriging predictions on the independent evaluation sample were significantly biased. Prediction error and precision were correlated with year, a north-south trend, and watersheds, despite the use of geostatistical methods to account for local autocorrelation. However, the patterns of bias may prove useful in identifying ways to improve the model and our understanding of relationships between stream health, land cover, and SWM.

Translating predictions from BIBI to regulatory categories resulted in excellent
performance. Predictions of passing and failing stream health were accurate 80% of the
time and consistent between the fitting and evaluation data. Customization of cutoff
thresholds using ROC curves was shown to be an effective strategy for regulators
interested in taking a precautionary approach to preventing stream health impairments.
However, predictions of high quality stream health proved to be inaccurate and
inconsistent, and the model should not be used to predict high quality streams.

It remains to be seen how land use planners and environmental managers will
accept regression kriging predictions. For those who are unfamiliar with regression
kriging, the kriged local error term may be viewed as a ‘fudge-factor’ or black box.
Some amount of educational work will be required if regression kriging prediction
models are proposed for use in regulatory settings.

**Further Work**

The model developed in this study can be improved by further work. Several
avenues can be pursued. The bias in RK predictions may be a result of overfitting.
Alternative models without reach and riparian buffer tree canopy variables should be
evaluated for accuracy improvements and model validity. Models without these terms
would be more broadly transferable, as they would not require the use of scarce high
resolution land cover data. Multi-model averaging approaches to inference and
prediction (Anderson and Burnham 2002) work around the issues of multicollinearity,
allowing more variables to be included in prediction (e.g., forest and agriculture), which
might act to reduce bias. The geographic patterns of agriculture (Figure 8) and bias
(Figure 26) are similar, suggesting that incorporating agriculture in predictions may
remove the bias in predictions.

Agriculture encompasses a diverse and heterogeneous set of land use practices. Further study should examine the stream health effects of different types of agriculture (e.g., cropland, pasture, animal feeding operations, etc.). This may help reduce issues of collinearity, and should provide a more nuanced understanding of land cover effects on stream health.

Non-linearity was apparent in the residuals and prediction errors. This might be eliminated with the use of smoother functions, higher order polynomial terms, or transformations of BIBI.

Temporal information was not incorporated into the model, and year was found to have an effect on prediction accuracy. Use of a random effects term, or adjustment of BIBI scores using trends from sentinel sites, might improve model precision and accuracy.

Environmental variables in this study area exhibit strong spatial trends and clusters. Moreover, model accuracy varied across space and among regions. For this reason, interactions in the effects of land cover and SWM variables may be more readily identified using regionalized (e.g., Smith, Schwarz, and Alexander 1997; Robertson, Saad, and Heisey 2006; Tsang 2008) or geographically weighted methods (e.g., Tu and Xia 2007; Lu, Thebpanya, and White 2010). This may help elaborate the stream health effects of correlated variables, such as agriculture, tree canopy, and extended detention pond SWM. Such methods can also prove to be more useful in applied settings than global regression models (e.g., Griffis and Stedinger 2007).

Different specifications of the spatial autocorrelation function can be explored.
Replacing or augmenting euclidean distance with stream flow path distances may provide a more realistic model for spatial autocorrelation of stream health (Peterson et al. 2007). Anisotrophic variograms could also be valuable, but would require more R programming to implement for GLS model fitting. Dormann et al. (2007) argue that multiple spatial regression methods should be applied and compared. Methods based in spatial filtering (Borcard and Legendre 2002; Getis and Griffith 2002; Griffith and Peres-Neto 2006) may provide a useful comparison to the GLS model, and can be adapted for stream flow distances (e.g., Urban et al. 2006; Blanchet, Legendre, and Borcard 2008).

Given the high performance of categorical predictions, and lower performance of continuous BIBI predictions, it is worth considering a logistic regression model for regulatory categories. Such a model would provide coefficients with direct interpretations for inference and prediction of passing, failing, and high quality stream health. Additionally, logistic regression requires fewer assumptions, and it could be easier to attain a statistically valid model.

To be useful for applied purposes, predictive models must address predictor variables that can by affected by environmental planners and managers, for example land cover and SWM. However, the effects of land cover and SWM on stream health are indirect. Land cover and SWM alter stream biota stressors through complex biogeochemical, hydrologic, and ecological systems. Hierarchical approaches that model the cascade of effects through these systems may provide better predictions (Novotny et al. 2005). Methods such as neural networks (Kralisch et al. 2003; Lu, Thebpanya, and White 2010), classification and regression trees (Hudy and Thieling 2008), or structural equation modeling (Graham 2003; King et al. 2005; Hung, Asaeda, and Manatunge 2007)
might be pursued. Additionally, these modeling approaches may be more effective for exploring interactions and thresholds among land cover and SWM variables.

Finally, BIBI is one measure of stream health, and there are many others for example species richness, tolerance, habit, and composition metrics (Barbour et al. 1999; Southerland et al. 2005), measurements of distance from biological potential (Paul et al 2009), and presence/absence of individual species (King and Baker 2010). Further work should model other measures of stream health, and compare results to those obtained for BIBI. This will enable more robust and nuanced management of stream health, and deepen our knowledge of stream ecology and human-environment interactions.

Summary

This study produced new knowledge on the relationship of stream health to land cover and SWM, and created a spatial GLS model useful for predicting passing and failing stream health. This model was shown to account for more than 60% of variation in BIBI scores. The model accurately predicts stream health impairment for more than 80% of evaluation sites. This level of performance is suitable for supporting precautionary policies for the prevention of water quality impairment. The model can be used to help enable rational, proactive environmental management and planning.

SWM altered the boundaries of perennial stream catchments, fundamentally altering catchment characteristics. Human-environment interactions shape the relationship of stream health and SWM. Specifically, Maryland’s stormwater design guidelines cause stormwater management decisions to react to and reinforce existing stream health conditions by concentrating biologically harmful wet ponds in the catchments of biologically impaired streams. The GLS model showed that filtration
facilities, and with less certainty detention ponds, provide small but significant improvements to stream health, increasing BIBI scores by as much as 0.5 points.

Riparian tree canopy has a small and variable effect on stream health. In contrast, preventing declines in catchment tree canopy avoids large declines in BIBI, regardless of the existing amount of tree canopy. Variation in catchment scale tree canopy alters BIBI by as much as 1.75 points. Tree canopy at the stream reach scale does not appear to be an important factor in stream health.

Impervious surfaces were confirmed to have the greatest effect on stream health. It was demonstrated that the effect of increasing impervious surfaces is greatest in catchments with the least existing impervious surfaces. The impact of additional impervious surfaces declines as catchment imperviousness increases. A nadir is reached at 30% impervious, where stream health averages 2.31 BIBI points less than sites with 0% impervious. Precision of BIBI predictions increases as impervious surfaces increase, reflecting the increasing certainty of biological impairment.

Model prediction errors were patterned by space, year, and watershed. This shows there is room for refinements to the model, and several avenues for further work were discussed. Continued development of predictive stream health models will further improve our knowledge of stream ecology and biogeography, and enable more effective environmental management and governance.
Appendix A: Maps of Variables

[Map showing impervious surfaces, study area, and open water]
Geographic distribution of impervious surfaces

- 0 - 6%
- 6 - 17%
- 17 - 28%
- 28 - 39%
- 39 - 53%

Agriculture

Study Area

Open Water

Streams
Geographic distribution of agricultural land cover

Tree Canopy

% Agriculture
- 0%
- 0 - 15%
- 15 - 41%
- 41 - 66%
- 66 - 92%

Study Area
Open Water
Streams
Geographic distribution of tree canopy cover.
Geographic distribution of riparian buffer tree canopy cover.

Geographic distribution of reach buffer tree canopy cover.
Geographic distribution of dry detention pond SWM facilities.
Geographic distribution of dry extended detention pond SWM facilities.
Geographic distribution of wet pond and wetland SWM facilities.
Geographic distribution of filtration SWM facilities.
Geographic distribution of infiltration SWM facilities.
Appendix B: Additional Bivariate Scatterplots

Bivariate scatterplots for BIBI and tree canopy variables
Darkness of points is proportional to density of points. Black line is LOESS smoother. Diagonals show histogram for each variable.
Bivariate scatterplots for BIBI and SWM variables

Darkness of points is proportional to density of points. Black line is LOESS smoother. Diagonals show histogram for each variable.
Appendix C: Partial Residual Plots of Final Model

Partial residual plots illustrating relationship of land cover and SWM to BIBI. Thick black lines illustrate the estimated relationship. Thin black lines are loess smoothers.
Appendix D: Confusion Matrices for Categorical Predictions

Confusion matrices for 3-category predictions of stream health, using naïve cutoffs.

<table>
<thead>
<tr>
<th>Fitting Population</th>
<th>Evaluation Population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True Class</strong></td>
<td><strong>True Class</strong></td>
</tr>
<tr>
<td><strong>Predicted Class</strong></td>
<td><strong>Predicted Class</strong></td>
</tr>
<tr>
<td>high</td>
<td>high</td>
</tr>
<tr>
<td>fair</td>
<td>fair</td>
</tr>
<tr>
<td>impaired</td>
<td>impaired</td>
</tr>
<tr>
<td>high</td>
<td>12</td>
</tr>
<tr>
<td>fair</td>
<td>65</td>
</tr>
<tr>
<td>impaired</td>
<td>0</td>
</tr>
<tr>
<td>total</td>
<td>13</td>
</tr>
<tr>
<td>high</td>
<td>4</td>
</tr>
<tr>
<td>fair</td>
<td>42</td>
</tr>
<tr>
<td>impaired</td>
<td>4</td>
</tr>
<tr>
<td>total</td>
<td>73</td>
</tr>
<tr>
<td>high</td>
<td>1</td>
</tr>
<tr>
<td>fair</td>
<td>24</td>
</tr>
<tr>
<td>impaired</td>
<td>9</td>
</tr>
<tr>
<td>total</td>
<td>56</td>
</tr>
<tr>
<td>total</td>
<td>293</td>
</tr>
<tr>
<td>total</td>
<td>135</td>
</tr>
</tbody>
</table>

Confusion matrices for predicting poor and passing stream health, using naïve cutoffs.

<table>
<thead>
<tr>
<th>Fitting Population</th>
<th>Evaluation Population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>True Class</strong></td>
<td><strong>True Class</strong></td>
</tr>
<tr>
<td><strong>Predicted Class</strong></td>
<td><strong>Predicted Class</strong></td>
</tr>
<tr>
<td>Passing</td>
<td>Passing</td>
</tr>
<tr>
<td>Impaired</td>
<td>Impaired</td>
</tr>
<tr>
<td>passing</td>
<td>147</td>
</tr>
<tr>
<td>Impaired</td>
<td>10</td>
</tr>
<tr>
<td>total</td>
<td>177</td>
</tr>
<tr>
<td>Passing</td>
<td>71</td>
</tr>
<tr>
<td>Impaired</td>
<td>13</td>
</tr>
<tr>
<td>total</td>
<td>84</td>
</tr>
<tr>
<td>total</td>
<td>293</td>
</tr>
<tr>
<td>total</td>
<td>135</td>
</tr>
</tbody>
</table>


———. 2007b. *TRANSPORTATION.ROAD_POLY* [computer file]. Towson, MD.


http://www.jstatsoft.org/v28/c01/ (last accessed 18 April 2011).


[http://www.dsd.state.md.us/comar/comarhtml/26/26.08.02.07.htm](http://www.dsd.state.md.us/comar/comarhtml/26/26.08.02.07.htm) (last accessed 17 April 2011)

[http://www.dsd.state.md.us/comar/comarhtml/26/26.08.02.08.htm](http://www.dsd.state.md.us/comar/comarhtml/26/26.08.02.08.htm) (last accessed 17 April 2011)


CURRICULUM VITA

NAME: Robert H. Hirsch, IV
PERMANENT ADDRESS: 1357 E Northern Pkwy, Baltimore MD, 21239
PROGRAM OF STUDY: Geography and Environmental Planning
DEGREE AND DATE TO BE CONFERRED: Master of Arts, 2011


Collegiate institutions attended:
   Major: Biology
Towson University, Towson, MD. 2006-2011. Master of Arts. 2011
   Major: Geography and Environmental Planning

Professional publications:

   Poster presented at the Towson University Geographic Information Sciences Conference 2005, Towson, MD.


Professional positions held:

2004-Present. Geographic Information Systems Analyst. Baltimore County Department of Environmental Protection and Sustainability. 105 W Chesapeake Ave, Room 400, Towson, MD 21204.

