DISSERTATION

PREDICTING THE LIKELIHOOD OF WATER QUALITY IMPAIRED STREAM SEGMENTS USING LANDSCAPE-SCALE DATA AND A HIERARCHICAL METHODOLOGY

Submitted by

Erin Elizabeth Peterson

Department of Geosciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2005

COLORADO STATE UNIVERSITY

August 23, 2005

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY ERIN ELIZABETH PETERSON ENTITLED PREDICTING THE LIKELIHOOD OF WATER QUALITY IMPAIRED STREAM SEGMENTS USING LANDSCAPE SCALE DATA AND A HIERARCHICAL METHODOLOGY BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Brian P. Bledsoe

Advisor

Co-Advisor

Department Head/Director

ABSTRACT OF DISSERTATION

PREDICTING THE LIKELIHOOD OF WATER QUALITY IMPAIRED STREAM SEGMENTS USING LANDSCAPE SCALE DATA AND A HIERARCHICAL METHODOLOGY

The purpose of my dissertation research was to develop a methodology that can be used by states and tribes to comply with two requirements of the Clean Water Act (1972): 1) to obtain an estimate of regional water quality condition and 2) to identify water quality impaired stream segments. To meet this goal, I developed a predictive model that incorporated spatial autocorrelation and combined coarse-scale geographic information system (GIS) data, such as mean elevation, with data collected using field-survey methods.

Geostatistical models are typically based on straight-line distance (SLD), which fails to represent the spatial configuration, connectivity, and directionality of sites in a stream network and may not be ecologically valid for studies in freshwater streams. Instead, hydrologic distances may represent the transfer of organisms, material, and energy through stream networks more accurately.

I developed GIS tools that generate the spatial data necessary for geostatistical modeling in stream networks. I quantified patterns of spatial autocorrelation in pH, conductivity, nitrate, sulfate, acid neutralizing capacity, dissolved organic carbon (DOC), temperature, and dissolved oxygen using three distance measures: SLD, symmetric hydrologic distance, and weighted asymmetric hydrologic distance. My results indicated that spatial autocorrelation exists in stream chemistry data at a relatively coarse scale and that geostatistical models consistently improved the accuracy of model predictions. SLD appeared to be the most suitable distance measure for regional geostatistical modeling of water chemistry in Maryland due to the extensive pre-processing time required for hydrologic distance measures and the inability of the survey design to adequately represent hydrologic relationships in a stream network. I developed a geostatistical model and used it to predict DOC at 3083 unobserved stream segments in Maryland. DOC estimates were categorized using ecological thresholds and reported in kilometers. The predictions and prediction variances were displayed using a GIS, which provided a simple way to recognize and communicate regional patterns in DOC. This methodology has clear advantages related to regional water quality monitoring because additional field sampling is not necessary, inferences about regional stream condition are generated, and it can be used to locate potentially impaired segments in a rapid and cost-efficient manner.

Erin E. Peterson
Department of Geosciences
Colorado State University
Fort Collins, CO 80523
Fall 2005

DISCLAIMER

The work reported here was developed under STAR Research Assistance Agreement CR-829095 awarded by the U.S. Environmental Protection Agency (EPA) to the Space Time Aquatic Resource Modeling and Analysis Program (STARMAP) at Colorado State University. This dissertation has not been formally reviewed by the EPA. The views expressed here are solely those of the author. The EPA does not endorse any products or commercial services mentioned in this dissertation.

DEDICATION

I dedicate this research to my family: my parents, Fred and Charlotte, my sister, Lindsay, and my husband, Nate. I am grateful for your patience, support, encouragement, and love throughout my graduate experience.

ACKNOWLEDGEMENTS

Completing my dissertation has been a tremendous personal achievement and I would like to take this opportunity to acknowledge some of the people who helped me reach this goal. First, I would like to thank my advisor Melinda Laituri and co-advisor Dave Theobald for their support and guidance throughout this process. Drs. Laituri and Theobald were enthusiastic about my research and gave me the freedom to shape it into something of my own. I am grateful to Dr. N. Scott Urquhart who also acted as a mentor to me. Dr. Urquhart spent an enormous amount of time answering my questions and discussing my ideas. In addition, he helped me to broaden my understanding of statistics.

I feel fortunate that I had the opportunity to collaborate with Dr. Jay Ver Hoef and Andrew Merton. Their creativity, knowledge of spatial statistics, and assistance in building and implementing innovative statistical tools allowed me to incorporate hydrologic distances into geostatistical models.

I would like to thank my office mate, Kirk Sherrill, for making me laugh and helping me to keep a positive attitude. Hadie and Mabel made sure that I spent enough time outside and kept me company during late nights and weekends at the office. Most importantly, I would like to thank my family, who has provided emotional and financial support throughout my education. I could not have done this without you all!

Funding for my research was provided by the STAR Research Assistance Agreement CR-829095 awarded by the U.S. Environmental Protection Agency to the

Space Time Aquatic Resource Modeling and Analysis Program (STARMAP) at Colorado State University.

TABLE OF CONTENTS

Page

TITLE...i

SIGNATURE PAGE...ii

ABSTRACT OF DISSERTATION...iii

DISCLAIMER...v

DEDICATION...vi

ACKNOWLEDGEMENTS...vii

LIST OF TABLES...xii

LIST OF FIGURES...xiv

CHAPTER ONE: INTRODUCTION TO DISSERTATION
- PROBLEM STATEMENT...1
- INTRODUCTION...2
- BACKGROUND...6
  - Heterogeneity in stream networks...7
  - Hierarchical frameworks in aquatic systems...8
  - pH...9
  - Acid neutralizing capacity...11
  - Nitrate...12
  - Sulfate...14
  - Dissolved organic carbon...15
  - Conductivity...17
  - Temperature...18
  - Dissolved oxygen...19
  - Classification systems...20
  - Local habitat classification...20
  - Landscape classification...21
  - Hierarchical classification...22
  - Geostatistical modeling...22
  - Geostatistical modeling in aquatic systems...24
- LITERATURE CITED...27

TABLES....39
FIGURES....40

CHAPTER TWO: SUPPORT FOR GEOSTATISTICAL MODELING ON STREAM NETWORKS: DEVELOPING VALID COVARIANCE MATRICES BASED ON HYDROLOGIC DISTANCE AND STREAM FLOW
INTRODUCTION....43
METHODS....53
Preprocessing streams data....53
Calculating hydrologic distance measures....54
Calculating the spatial weights....55
Developing a valid covariance matrix based on hydrologic distance and flow....56
Example....58
DISCUSSION....60
LITERATURE CITED....62
FIGURES....66

CHAPTER THREE: PATTERNS OF SPATIAL AUTOCORRELATION IN STREAM WATER CHEMISTRY
INTRODUCTION....70
BACKGROUND....73
METHODOLOGY....78
Data....78
Statistical analyses....82
RESULTS....86
DISCUSSION....89
CONCLUSIONS....97
LITERATURE CITED....99
TABLES....104
FIGURES....113

CHAPTER FOUR: PREDICTING WATER QUALITY IMPAIRED STREAM SEGMENTS USING LANDSCAPE-SCALE DATA AND A REGIONAL GEOSTATISTICAL MODEL: A CASE STUDY IN MARYLAND
INTRODUCTION....117
METHODS....122
Study area and design....122
GIS pre-processing....123
Statistical analyses....125
RESULTS....130
DISCUSSION....133
CONCLUSIONS....140
LITERATURE CITED....141
TABLES....147

FIGURES.....154

CONCLUSIONS
GEOSTATISTICAL MODELING AND IMPLICATIONS FOR WATER QUALITY MONITORING.....160
Spatial autocorrelation in stream water chemistry.....160
Scale of analysis.....161
Geostatistical modeling and regional water quality monitoring.....162
FUTURE RESEARCH.....164
Functional distance measures.....164
Survey designs for stream networks.....165

ADDITIONAL EXPERIENCES, CHALLENGES, AND THOUGHTS
TECHNICAL CHALLENGES.....167
INTERDISCIPLINARY COMMUNICATION.....168

APPENDIX I: FLOWCHART OF METHODOLOGY FOR SPATIAL ANALYSIS.....171

APPENDIX II: GIS CODE.....174
SNAP POINTS TO LINES PROGRAM.....174
CALCULATE PROPORTIONAL INFLUENCE PROGRAM.....179
ACCUMULATE CATCHMENT ATTRIBUTE PROGRAM.....188
ACCUMULATE MEAN CATCHMENT ATTRIBUTE PROGRAM.....207
CALCULATE WATERSHED ATTRIBUTES.....231
ESTIMATE MEAN WATERSHED ATTRIBUTE.....240
OUTPUT MATRIX.....251
CALCULATE HYDROLOGIC DISTANCE PROGRAM.....255
CREATE THE NETWORK ELEMENT OBJECT.....280

APPENDIX III: STATISTICS CODE.....282
MODEL SELECTION USING SPATIAL AICC.....282
R FUNCTIONS FOR SPATIAL AICC MODEL SELECTION.....283
MAKE PREDICTIONS AT UNOBSERVED SITES.....285
R FUNCTIONS FOR MAKE PREDICTIONS AT UNOBSERVED SITES.....287

APPENDIX IV: CLASSIFICATION OF PRIMARY LITHOLOGIES INTO FIVE CLASSES OF GEOLOGICAL SENSITIVITY.....292

LIST OF TABLES

Page

Table 1.1 - Functional linkage between the landscape and benthic macroinvertebrates.....39

Table 3.1 - A summary of studies that have explored patterns of spatial autocorrelation in stream chemistry data using straight-line distance and symmetric hydrologic distance.....104

Table 3.2 - Potential watershed covariates used in the initial model selection.....106

Table 3.3 - Summary statistics for chemical response variables and significant covariates with the exception of NORTHING.....107

Table 3.4 - Summary statistics for distance measures in kilometers using dissolved oxygen (n=826).....109

Table 3.5 - Five significant covariates for each chemical response variable selected using leaps and bounds regression.....110

Table 3.6 - Autocorrelation parameter estimates, mean square prediction error, and r^2 for the general linear model, straight-line distance, symmetric hydrologic distance, and weighted asymmetric hydrologic distance model for each unscaled response variable.....111

Table 4.1 - Potential watershed covariates.....147

Table 4.2 - Summary statistics for dissolved organic carbon and model covariates.....149

Table 4.3 - Model results.....150

Table 4.4 - Cross-validation intervals for straight-line distance Mariah regression coefficients.....151

Table 4.5 - Unstandardized parameter estimates for the straight-line distance Mariah model that was used to make predictions at unobserved stream segments.....152

Table 4.6 - Summary statistics for dissolved organic carbon predictions and prediction variances produced by the straight-line distance Mariah model....................153

LIST OF FIGURES

Page

Figure 1.1 - Heterogeneity in terrestrial and aquatic ecosystems....................................40

Figure 1.2 - The semivariogram..41

Figure 1.3 - Symmetric and asymmetric distance classes and their associated connectivity matrices..42

Figure 2.1 - Symmetric and asymmetric distance classes..66

Figure 2.2 - The semivariogram..67

Figure 2.3 - Hydrologic distance and proportional influence matrices..........................68

Figure 2.4 - Calculating the proportional influence..69

Figure 2.5 - Example data...70

Figure 3.1 - Symmetric and asymmetric distance classes..113

Figure 3.2 - The Maryland Biological Stream Survey dataset.......................................114

Figure 3.3 - Spatial relationships between neighboring survey sites.............................115

Figure 3.4 - Difference between observed values and general linear model predictions or observed values and weighted asymmetric hydrologic distance predictions according to the number of neighbors.............................116

Figure 4.1 - Straight-line distance and asymmetric hydrologic distance........................154

Figure 4.2 - The Maryland Biological Stream Survey sites and their dissolved organic values in mg/l collected by the Maryland Department of Natural Resources in 1996..155

Figure 4.3 - The segment proportional influence (PI) represents a segment's PI on the segment directly downstream..156

Figure 4.4 - r^2 for the observed versus predicted values of dissolved organic carbon using the straight-line distance Mariah model..157

Figure 4.5 - Square prediction error values for the straight-line distance Mariah model...158

Figure 4.6 - Map of the dissolved organic carbon predictions and prediction variances for 3083 first, second, and third order non-tidal streams in seven interbasins throughout Maryland sampled in 1996...................................159

CHAPTER ONE

INTRODUCTION TO DISSERTATION

PROBLEM STATEMENT

The Clean Water Act (CWA) (1972) requires states and tribes to identify water quality impaired stream segments, to create a priority ranking of those segments, and to calculate the Total Maximum Daily Load (TMDL) for each impaired segment based upon chemical and physical water quality standards. States and tribes are also required to create a biennial inventory that characterizes regional water quality based on the attainment of designated use standards assigned to individual stream segments. Yet, it is impossible to physically sample every stream within a large area due to the immense number of segments, limited personnel, and the cost associated with sampling (Olsen and Ivanovich, 1993; Herlihy et al., 2000; USEPA, 2001). Despite difficulties related to sampling, lawsuits have been filed in 38 states by environmental groups demanding that the requirements of the CWA be met in a timely fashion (Copeland, 2002). This results in increased pressure to develop a rapid and cost-efficient sampling method that has the ability to identify potentially impaired stream segments in large areas.

The purpose of my dissertation research was to develop a methodology using geostatistical models to predict water quality characteristics for stream segments found throughout a large geographic area (e.g., state). I developed a predictive model that incorporated spatial autocorrelation and that combined coarse-scale geographic information system (GIS) data, such as percent agricultural land cover or mean elevation, with data collected using field-survey methods (e.g., Environmental Protection Agency (EPA) Environmental Monitoring and Assessment Program (EMAP)). These geostatistical models were used to generate water chemistry predictions, which were assigned to stream segments within a GIS. This approach allowed me to identify potentially impaired stream segments and to estimate regional water quality conditions.

INTRODUCTION

It is generally accepted that stream water chemistry is spatially and temporally heterogeneous at multiple scales (Pringle, 1991; Chambers et al., 1992; Dawson et al., 2001). Conditions at observed sites result from the collective influence of multi-scale filters acting within and between the terrestrial and aquatic ecosystems (Frissell et al., 1986; Poff, 1997; Fausch et al., 2002). For example, Quinn and Hickey (1990) found that percent development in the watershed influenced total stream phosphorus, which acted as a filter determining benthic macroinvertebrate (BMI) taxonomic richness in New Zealand rivers. A hierarchical constraint exists among filters (Frissell et al., 1986; Davies et al., 2000) and the strength of the linkage varies between filter scales (Mykra et al., 2004; Sandin and Johnson, 2004). Consequently, the ability of individual stream segments to withstand potentially degrading conditions should be influenced by coarse-scale conditions in the watershed, such as geology or vegetation type (Richards and Host, 1994; Richards et al., 1996).

The heterogeneity found in stream segments at multiple filter scales makes it difficult to distinguish between the natural variability in the ecosystem and changes caused by human disturbance. Strahler stream order (Strahler, 1957) is commonly used to categorize stream segments so that inferences can be made about the condition of streams within an order (Herlihy et al., 2000). However, two segments of the same stream order may have dissimilar characteristics if conditions at coarse filter scales are also dissimilar. For instance, two 3rd order streams could have dramatically different heavy metal concentrations resulting from unique coarse-scale watershed conditions, such as the presence of a mine. Conversely, these differences might result from local chemical conditions, such as in stream pH or acid neutralizing capacity, which both influence metal solubility. The hierarchical structure of filter scales makes it necessary to model and analyze stream processes at multiple scales to gain a better understanding of stream condition.

One of the most prevalent methods of identifying water quality impaired stream segments is to perform a bioassessment using biological data (Lenat, 1988; Barbour et al., 1999; Mercurio et al., 1999). Benthic macroinvertebrates, such as mayflies (Ephemeroptera) and stoneflies (Plecoptera), are commonly used (Kiffney and Clements, 1996) because they are sensitive to hydrologic, chemical, and physical degradation at all life stages (Barbour et al., 1999). The BMI community structure of a segment provides a measure of degradation resulting from background, point, and non-point source pollution, as well as episodic events that might not be detected using traditional sampling techniques (Hughes, 1990). This integrative approach makes it possible to measure all of the spatial and temporal changes in the hydrologic, chemical, and physical properties of the stream segment using one sample.

Classification systems based on BMI species structure, species abundance, or functional group composition are an essential component of bioassessments (Barbour et al., 1999). They are used to explain variability found in BMI community structure resulting from conditions at coarser filter scales, such as the watershed and/or stream segment (Poff and Ward, 1989; Richards et al., 1996; Richards et al., 1997; Davies et al., 2000; Hawkins et al., 2000; Sandin, 2003; Mykra et al., 2004). This allows researchers to distinguish between natural spatial and temporal variability in the system and degradation resulting from human disturbance (Barbour et al., 1999).

Bioassessments based on classification systems provide a cumulative approach to evaluating stream conditions, but are not a rapid or cost-effective means for assessing regional stream condition when used alone. Landscape-scale data can be remotely derived using GIS, satellite images, or aerial photography, but biological and reach-scale data must be collected by field crews, which is prohibitively expensive and time consuming if large areas are surveyed. Bioassessments only provide estimates of stream condition at observed sites and cannot be used to identify potentially impaired stream segments that were not surveyed. In addition, it is challenging to derive a regional estimate of stream condition using data collected for bioassessments because of the large amount of natural variability found in streams.

Geostatistical models have the potential to improve the efficiency of regional water quality monitoring because they can be used to estimate stream condition throughout a large area based on digitally derived landscape variables and the distribution of the observed data. Geostatistical modeling is specifically designed to analyze spatially correlated data. A statistical model is used to explain the variability in the data associated with the mean and the spatial variability of the observed data is used to predict values at unobserved sites based on their distribution through space (Cressie, 1993). Geostatistical models are typically able to explain more variability in the data and provide more accurate predictions at unobserved sites when spatial autocorrelation is present in the data (Isaaks and Srivastava, 1989).

Geostatistical model predictions provide a cost-efficient alternative to extensive field sampling that can be used by states and tribes to gather information concerning regional stream condition. Compliance with the CWA would be more efficient if potentially impaired stream segments could be identified more rapidly. More intense field sampling efforts could focus on potentially impaired sites, making additional resources available for the TMDL calculation for a specific segment. A regional geostatistical model could also help to provide an estimate of regional stream condition. The model predictions themselves would provide a regional estimate of specific segment scale water quality parameters regulated by the CWA, such as nitrate or pH. In the future, it might also be possible to use the predicted values as input into a hierarchical classification system, which could be used to generate predictions of BMI community structure.

Few geostatistical models have been developed for stream networks (see Kellum, 2003; Yuan, 2004a). One reason may be that they are generally based on symmetric straight-line distance, which fails to represent the spatial configuration, connectivity, directionality, and relative position of sites in a stream network. However, freshwater ecologists have begun to explore spatial patterns in stream networks using hydrologic distance measures (Dent and Grimm, 1999; Gardner et al., 2003; Ganio et al., 2005; Cressie et al., in press) and new geostatistical methodologies have recently been developed that enable directional hydrologic distance measures to be considered (Ver Hoef et al., in press).

The purpose of my dissertation research was to explore patterns of spatial autocorrelation in stream water chemistry using a GIS-based geostatistical methodology (Appendices I, II, and III) and to demonstrate its value for regional water quality monitoring. Below, I provide background information concerning the ecological relationship between landscape variables, several water chemistry variables, and community composition of BMI. In addition, I discuss the applicability of various types of classification systems for regional-scale bioassessments and provide an overview of geostatistical modeling. Chapter One describes the GIS tools that I developed to calculate the hydrologic distances and spatial weights necessary for geostatistical modeling in stream networks (Appendices I and II). In Chapter Two, I quantified patterns of spatial autocorrelation in eight stream water chemistry variables using three distance measures: straight-line distance, symmetric hydrologic distance, and weighted asymmetric hydrologic distance (Appendices I, II, and II). I developed a geostatistical model in Chapter Three and used it to generate dissolved organic carbon predictions and prediction variances for 3,083 stream segments throughout Maryland (Appendices I, II, and III). I conclude in Chapter Four with a discussion about my research and its implications for water quality monitoring. I also include a short discussion in Chapter Five about my thoughts and experiences as a student working on an interdisciplinary project.

BACKGROUND

Scale is a term with many definitions between and even within disciplines. For the purposes of my dissertation, scale refers to the grain rather than the extent of the data. Coarse-scale or landscape-scale data represent lumped watershed attributes, such as percent wetlands or mean elevation in the watershed, which are typically calculated using GIS or remotely sensed data. Fine-scale or local-scale data refers to measurements at the segment or microhabitat scale that are typically collected in the field (Figure 1.1).

Heterogeneity in stream networks

The River Continuum Concept (RCC) proposed by Vannote and others (1980) suggests that biotic communities and ecological processes change predictably downstream along a continuous gradient of physical conditions. The RCC represents the longitudinal connectivity in a river, but fails to represent abrupt differences in stream condition related to tributary confluences or differences in the spatial context. In contrast, Frissell and others (1986) proposed a spatially nested hierarchical framework for stream habitat classification that is similar to hierarchical classification systems developed for terrestrial ecosystems (Lotspeich and Platt, 1982; Bailey, 1983) because it includes both context and scale. The hierarchical framework developed by Frissell and others (1986) is restricted to aquatic filter scales, which include the stream, segment, reach, pool/riffle, and microhabitat systems. Yet, a stream network is also strongly influenced by the context of the terrestrial surroundings (Hynes, 1975).

More recently, stream ecologists have developed hierarchical frameworks that include both terrestrial and aquatic filter scales (Poff, 1997; Richards et al., 1997; Fausch et al., 2002). These frameworks generally represent the relationship between the watershed, reach, and microhabitat scales (Figure 1.1). However, processes and interactions are not restricted to these common filter scales. Instead, they occur between and within multiple aquatic and terrestrial filter scales.

The Network Dynamics Hypothesis (NDH) developed by Benda and others (2004) is a geomorphic rather than hierarchical framework. It is unlike the RCC and hierarchical frameworks because it can be used to investigate heterogeneity resulting from the spatial arrangement of tributaries in a stream network or stochastic disturbances, such as floods. It is based on seven structural frameworks, which include the basin size, basin shape, network configuration, size differences between tributary channels and the main stem, drainage density, confluence density, and network geometry.

The RCC, hierarchical classification systems, and the NDH are valuable because they can be used to better understand and describe physical, chemical, and biological heterogeneity in stream networks. None of these frameworks have the ability to explain all heterogeneity in stream networks. Instead, they compliment one another and should be used in concert.

Hierarchical frameworks in aquatic systems

Despite their simplicity, hierarchical frameworks have been used to explain variability in BMI species structure, species abundance, or functional group composition. Chemical habitat characteristics strongly influence BMI community structure because species have evolved to live in specific habitat conditions, such as a cold temperature range (Hynes, 1960) and have developed unique behavioral and physiological mechanisms suited to specific aquatic habitat types (Angelier, 2003). Landscape variables, such as geology, vegetative composition, and topography, influence chemical habitat characteristics in the stream (Nelson et al., 1992; Richards et al., 1996; Davies et al., 2000), which essentially act as a functional linkage between the watershed and BMI community structure (Lanka et al., 1987). The following summary describes how eight chemical response variables: pH, acid neutralizing capacity, nitrate, conductivity, sulfate, dissolved organic carbon, dissolved oxygen, and temperature, act as a functional linkage between the landscape and BMI community structure (Table 1.1).

pH

The pH (*potential hydrogenii*) is a measure of the hydrogen ion concentration in a solution and is used to describe the relative strength of alkaline and acidic conditions of stream water. pH ranges between 0 and 14, where a value of 7 is neutral. A decrease in the hydrogen ion concentration produces alkaline conditions ($pH > 7$) and an increase in the hydrogen ion concentration produces acidic conditions ($pH < 7$). When the system is dominated by carbonate, the hydrogen ions bind with other substances, and the pH remains stable. When the system is dominated by carbon dioxide, the number of free hydrogen ions increases, which causes the pH value to fall. Acidification of stream water can result from anthropogenic stresses such as acid mine drainage (Herlihy et al., 1990) or the atmospheric deposition of nitric and sulfuric acids (Angelier, 2003). However, naturally acidic streams can also be found in areas with considerable humic inputs (Allan, 1995).

Although BMI sensitivities to pH vary (Yuan, 2004b), values below 5.0 and greater than 9.0 are considered harmful (Leidy, 1980). Low pH values are associated with lower diversity (Thomsen and Friberg, 2002), decreased emergence rates (Hall et al., 1980), and egg failure in BMI (Willoughby and Mappin, 1988). Physiological problems occur because it is difficult for BMI to regulate ions within their bodies and to absorb the calcium needed for exoskeletons (Hall et al., 1980). In addition, a limited amount of evidence indicates that acidic conditions decrease microorganism productivity and consequently food availability for BMI (Willoughby and Mappin, 1988; Groom and Hildrew, 1989). Lower growth rates have been observed in some species of BMI and attributed to a decrease food quality (Groom and Hildrew, 1989; Thomsen and Friberg, 2002). A decrease in the pH of stream water can trigger the release of heavy metals, which are toxic to BMI (Clements, 1994; Peiffer et al. 1997; Ramsey and Brannon, 1988). Non-toxic metal concentrations also have a long-term effect due to physiological and behavioral changes in BMI (EMCBC, 1999), as well as, a change in species abundance and dominance (Clements et al., 2000).

The pH value of stream water lies at the equilibrium point between acidic inputs and the acid neutralizing capacity of the system (Allan, 1995). Nitrogen and sulfur emissions are secondary products of industrial fuel combustion and atmospheric deposition of these constituents can cause significant acidification of soil and stream water (Driscoll et al., 2001). Nitrate may also contribute to acidification and is produced by natural sources such as coniferous forests (Thomsen and Friberg, 2002) or from anthropogenic sources such as agricultural areas (EHMP, 2004). As a result, streams that drain catchments dominated by these landuses tend to be acidic (Kawakami et al., 2001;

Thomsen and Friberg, 2002). Waters affected by mining effluent generally have a low pH (Herlihy et al., 1990) and may in turn decrease the pH of the stream water. The weathering of parent material in the catchment is the primary source of buffering substances found in the stream, which could indicate that geology and climate significantly influence pH.

Acid neutralizing capacity

Acid neutralizing capacity (ANC) is a quantitative measure of a stream's ability to withstand changes in pH. The carbonate system, which consists primarily of calcium and magnesium bearing carbonates and soluble nonresistant silicate minerals, is the main factor affecting the ANC (Kang et al., 2001). If a sufficient amount of base cations, such as calcium, are present in the system, the hydrogen ions are removed from the solution, and the pH of the water remains stable (Kang et al., 2001). ANC plays a critical role in the regulation of chemical ions because it affects the chemical form, concentration, biological availability, and toxicity of many ions (Feldman and Conner, 1985). Low ANC values are detrimental to the health of BMI (Baker and Schofield, 1982) and cause declines in BMI fertility, growth, and emergence (Angelier, 2003).

A substantial amount of research indicates that ANC is strongly correlated to landscape variables and that statistical models can be used to make predictions in streams (Herlihy et al., 1998; Shirazi et al., 2001; Kellum, 2003). Shirazi and others (2001) investigated the relationship between catchment soil characteristics and water quality in the Mid-Atlantic region and found that the mean ANC increased as the soil particle size decreased. The geology (Shirazi et al., 2001; Kellum, 2003; Cooper et al., 2004) and soil type (Shirazi et al., 2001) are related to ANC because the parent material in the catchment is a significant source of carbonates (Jenny, 1941), while the climate, elevation (Kellum, 2003), and aspect of the catchment affect the rate at which the parent material weathers. ANC tends to be lower in headwaters compared to larger streams that drain the watershed (Ward, 1992) and Strahler stream order (Strahler, 1957) could be used to represent the position of the segment in the stream network. Increased levels of ANC have been associated with catchments containing agricultural and urban land uses (Johnson et al., 1997; Herlihy et al., 1998; Kellum, 2003), while forested catchments tend to produce lower ANC values (Herlihy et al., 1998; Cooper et al., 2004).

Nitrate

Nitrate (NO_3) is a nutrient commonly found in streams, which has both natural and anthropogenic sources. Organic nitrogen (ON) is converted to ammonium (NH_4^+) during decomposition. Ammonium that escapes rapid uptake by microbes is converted to nitrate via nitrification (Equation 1.1).

$$ON \rightarrow NH_4^+ \rightarrow NO_2^- \rightarrow NO_3^- \tag{1.1}$$

Nitrate and ammonia are the only biologically available forms of nitrogen for aquatic plants. Nitrates that originate in the terrestrial system may be taken up by terrestrial plants or leach from the soil and move to the stream or the aquifer. Nitrate is removed from stream water by aquatic plants or is eliminated during the denitrification process (Equation 1.2).

$$N_2 \leftarrow N_2O \leftarrow NO \leftarrow NO_2^- \leftarrow NO_3^- \qquad (1.2)$$

To some extent, increased levels of nitrate have been associated with increased species richness (Hynes, 1975). However, chronic nitrate exposure has also been shown to reduce growth and increase mortality in BMI (Canadian Council of Ministers of the Environment, 2003). Camargo and Ward (1992) estimated acute toxicity levels of NO_3 on BMI to range between 290 and 527 mg/L^{-1}. In addition, elevated nitrate levels can increase algae growth, promote bacterial decomposition, cause oxygen depletion, and increase the likelihood of eutrophication (Nordin and Pommen, 1986; Meade and Watts, 1995).

NO_3 is acidic and has been associated with decreased pH and ANC levels (Kawakami et al, 2001). Natural concentrations of NO_3 in stream water are low compared to streams affected by anthropogenic inputs (Meybeck, 1982), which are generally responsible for elevated NO_3 levels in stream water (Chapin et al., 2002). Agricultural fertilizers may be flushed from fields during storm events and are a source of NO_3 in stream water. Feed lots also act as agricultural point sources because animal manure contains NO_3 (Sheets, 1980). Urban areas contribute NO_3 rich municipal waste water (Allan, 1995) that comes from residential fertilizers, septic systems, and garbage dumps (Sheets, 1980). Therefore, it is not surprising that NO_3 is correlated with percent agriculture (Osborne and Wiley, 1988; Jordan et al., 1997; Herlihy et al., 1998) and percent urban landuse in the watershed (Field et al., 1996; Herlihy et al., 1998).

Sulfate

Sulfur is an essential plant nutrient that is naturally present in the environment. Sulfur, in the form of sulfur dioxide or sulfite, is emitted to the atmosphere, where it comes in contact with water vapor. An oxidation reaction occurs and droplets of sulfuric acid are produced. Sulfates (SO_4) are formed when one or more hydrogen ion in sulfuric acid is replaced by a metal or a radical. Sulfates move from the atmosphere to the land surface via wet and dry deposition. In addition to atmospheric SO_4 production, SO_4 may be created by the oxidation of sulfide minerals (Blowes et al., 2003) and subsequently transported to the stream via overland or sub-surface water flow.

$$H_2S + \left\{\begin{matrix} O \\ O_2 \\ O_3 \end{matrix}\right\} \rightarrow SO_2 + \left\{\begin{matrix} O \\ O_2 \end{matrix}\right\} \rightarrow SO_3 + H_2O \rightarrow H_2SO_4 + X \rightarrow XSO_4 \qquad (1.3)$$

Natural sources of sulfur include mineral weathering (Alewell et al., 1999), volcanoes, decomposition of organic matter, and sea salt from the ocean surface (Kellogg et al., 1972). However, sulfur is also produced by industrial sources, such as mining, paper mills, textile mills, and tanneries (Singleton, 2000). Atmospheric emissions of sulfur from the burning of fossil fuels result in bulk precipitation and dry deposition of SO_4 (Alewell et al., 1999). Agricultural fertilizers have also been shown to produce elevated SO_4 levels in stream water (Kellogg et al., 1972).

Elevated SO_4 concentrations can be toxic in the form of metal-sulfate compounds. They are strongly acidic and have the ability to reduce the pH of stream water when the ANC is insufficient (Kellogg et al., 1972). In addition, most metal-sulfates dissolve in stream water, which causes metals to be released. Low pH values and elevated heavy metal concentrations are toxic to benthic organisms (Willoughby and Mappin, 1988; Clements, 1994; Peiffer et al., 1997; Ramsey and Brannon, 1999; Thomsen and Friberg, 2002).

The atmosphere is the most common transport mechanism for industrial emissions of sulfur (Kellogg et al. 1972) and the proximity of large urban areas may be correlated to sulfur inputs. SO_4 is also a secondary product of mining operations (EMCBC, 1999), which could be represented by the presence of mines in the watershed. In addition, geology types that contain sulfides may be correlated with stream SO_4.

Dissolved organic carbon

Dissolved organic carbon (DOC) is a significant source of energy for lotic systems (Allan, 1995). The majority of DOC comes from allochthonous sources (Wetzel, 1992) of organic matter (OM) such as soil, groundwater, and dead terrestrial plant material. DOC is transported from the watershed via overland, sub-surface, or base flow and the flow path of water affects the stream concentration. Shallow sub-surface paths and overland flow through wetlands, organic soil layers, and shallow soils tend to produce water with relatively high concentrations of DOC (Mulholland, 2003). Conversely, sub-surface or base flow moving through deeper soil horizons may lose DOC, which is adsorped by the mineral soils (Qualls and Haines, 1992). When OM enters the stream, chemicals are leached from the detritus within a few days (Allan, 1995) and are generally processed close to their point of entry (Hall and Meyer, 1998).

DOC provides energy to BMI through a number of trophic pathways. Initially, it is taken up by microorganisms, such as algae and bacteria, and assimilated into microbial biomass. Microbial consumers, such as flagellate and ciliates, feed on microorganisms and are ingested by larger BMI (Hall and Meyer, 1998). Organic microlayers located on rocks or other microorganisms are consumed by BMI grazers and deposit feeders (Rounick and Winterbourn, 1983). Microorganisms growing on detritus are also inadvertently consumed by detritivores (Meyer, 1994). Clearly, the energy linkage between the microbial population and the BMI is not always direct (Fuller et al., 2004). Therefore, the strength of the linkage between DOC and BMI will vary between stream ecosystems and in some cases a substantial amount of energy may dissipate without reaching the BMI community (Allan, 1995). In addition, DOC concentration indirectly affects BMI because metals bind with OM and make them less available for uptake by aquatic organisms (Prusha and Clements, 2004). DOC also physically limits the amount of ultraviolet rays that penetrate the water column, which have been shown to negatively impact BMI and reduce their abundance (Kiffney et al., 1997).

The dominant watershed processes affecting DOC production are associated with the production of organic matter. For example, coarse-scale watershed attributes, such as wetlands (Eckhardt and Moore, 1990; Wetzel, 1992; Gergel et al., 1999), forest cover (Guyot and Wasson, 1994; Canham et al., 2004; Prusha and Clements, 2004), temperature (Hejzlar et al., 2003), and precipitation (Wetzel, 1992; Canham et al., 2004) have been shown to influence DOC concentrations. However, the flow path between the watershed and the stream is also important. It would be difficult to explicitly represent water transport for every watershed in a region using coarse-scale GIS data. However, studies have shown that DOC is related to topography (Eckhardt and Moore, 1990; Houle et al., 1995), soil texture (Shirazi et al., 2001), the adsorptive capacity of the soil (Nelson et al., 1996), and precipitation (Wetzel, 1992; Canham et al., 2004), which may represent variability associated with the flow path of water.

Conductivity

Conductivity is a measure of the ability of a solution to carry an electrical charge based on ion concentration and temperature (Allan, 1995). Inorganic dissolved solids, such as carbonate or sulfate, increase the conductivity of water and organic compounds, such as sugars and oils, decrease the conductivity. In addition, conductivity is positively correlated to temperature and therefore measurements must be recorded at 25 degrees Celsius.

BMI differ in their conductivity tolerance, but extreme values are generally considered harmful (Krogh, 1965). Low conductivity levels cause BMI to expend excessive amounts of energy on osmoregulation. Conversely, high conductivity levels indicate that the water has an elevated ion concentration, which can be toxic to BMI. Conductivity may also indirectly affect BMI because it represents ion concentrations that influence nutrient cycling, rates of production and respiration, and the condition of riparian vegetation (Allan, 1995).

Conductivity is strongly influenced by landscape scale conditions. The geology in the catchment is the source of the ions that act as conductors of electricity (Golterman, 1975). Granite bedrock tends to produce lower conductivity values, while clays produce higher levels of conductivity. Urban and agricultural land uses have also been shown to increase conductivity levels (Gray, 2004). In addition, water affected by mining generally has a high conductivity (Herlihy et al., 1990). Urban areas may decrease conductivity due to inputs of total dissolved solids (Gray, 2004) and agriculture may increase conductivity because of nitrate inputs. Furthermore, there are also seasonal differences in conductivity that result from a negative relationship with discharge volume (Caruso, 2002; Gray, 2004).

Temperature

Stream temperature is spatially and temporally variable (Hynes, 1960; Biggs et al., 1990) and is a function of the source water temperature and its transport time (Angelier, 2003). Temperatures may be relatively stable in large rivers with low flow velocities, but can fluctuate quickly in steep shallow streams. Seasonal variation also results from changes in the hydrologic regime (Angelier, 2003) and air temperature (Smith, 1981).

BMI have evolved to live within a specific temperature range, which limits their distribution and affects the community structure (Hynes, 1960; Biggs et al., 1990). Temperature affects emergence patterns (Lehmkuhl, 1974), growth rates (Sweeney and Schnack, 1977), metabolism (Angelier, 2003), reproduction (Vannote and Sweeny, 1980), and body size (Sweeney and Schnack, 1977). Species vary in their tolerance to temperature ranges, but few are able to tolerate temperatures beyond their upper tolerance limit (Angelier, 2003).

There is evidence that stream temperatures can be predicted using landscape variables. Smith (1981) found that stream temperatures in Great Britain were highly correlated to air temperature. In addition, other studies show that elevation, riparian vegetation, and channel width influence stream temperature (Osborne and Wiley, 1988; Gregory et al., 1991). These results indicate that readily available landscape variables, such as elevation, air temperature, and riparian condition (Platts, 1979; Vannote and Sweeney, 1980), may explain some variability in stream temperature.

Dissolved oxygen

Dissolved oxygen is critical to BMI (Hynes, 1960; Dauer et al., 2000) and is related to the biological oxygen demand in the stream. As organic matter increases, decomposition rates rise. Microbial biomass increases in response the addition of nutrients and more oxygen is consumed (Hynes, 1960). If the oxygen deficit in the stream becomes severe, it can cause a dead zone downstream that is free of aquatic life. Decomposition rates can be limited by nutrient supply (Chapin et al., 2002) and are sensitive to nutrient additions resulting from agriculture or urban runoff (Dauer et al., 2000). Oxygen is slowly replenished by atmospheric uptake, photosynthetic additions, and the turbulent mixing of oxygen and water (Hynes, 1960). It varies diurnally in productive waters due to photosynthetic inputs during the day and respiratory losses at night. In unpolluted headwater streams, DO is inversely related to water temperature (Hynes, 1960) and the opposite diurnal pattern may be observed (Allan, 1995).

DO is critical to BMI survival because they are aerobic organisms. When diurnal fluctuations of DO become extreme, the oxygen saturation drops to critically low levels, which BMI cannot tolerate. DO is indirectly affected by natural and anthropogenic inputs of OM from the watershed because they affect the biological oxygen demand in the water. Inputs from agricultural areas, such as fertilizers and animal manure, as well as, sewage effluents from urban areas, are a common source of nutrients in the stream (Sheets, 1980). Dauer and others (2000) found that population density, percent forested area, and percent urban area are significantly correlated to DO. In addition, stream size may be significant since the water in steep rocky streams is constantly mixing with the atmosphere, which increases oxygenation rates (Allan, 1995). Although these coarse filter scales may influence DO to some degree, it is probable that it is more strongly influenced by finer scale filters, such as substrate type, stream temperature, local nutrient levels, net primary production, and DOC (Hynes, 1960; Allan, 1995; Angelier, 2003).

Classification systems

Clearly, BMI community structure is affected by various chemical, physical, and biological conditions acting across multiple aquatic and terrestrial filter scales. Many classification systems have been developed to partition natural and anthropogenically derived variability found in BMI community structure. I will describe three general systems here: local habitat classification, landscape classification, and hierarchical classification. I will also discuss the suitability of each system for regional assessment of stream condition.

Local habitat classification

Previous studies have shown that BMI community structure is somewhat dependent upon the local habitat characteristics that exist at the reach and microhabitat scale (Richards et al., 1997; Maddock, 1999; Dauer et al., 2000; Prusha and Clements, 2004). Consequently, it is not surprising that classification systems that include local habitat characteristics have been proposed (Poff and Ward, 1989; Biggs et al., 1990; Hawkins et al., 1993). The use of these data generally improve estimates of BMI community structure, but they require intensive field surveys, which are expensive and time consuming to physically collect. This can be prohibitive if the study area is very large, a rapid assessment is necessary, or the project budget is restrictive (Smart et al., 1981; Richards and Host, 1994; Roth et al., 1996; Davies et al., 2000; Mykra et al., 2004). In addition, accessibility issues related to rugged terrain and private land ownership make it challenging to survey in some areas despite adequate time and funding (Herlihy et al., 2000).

Landscape classification

Landscape classification is based on readily available landscape-scale variables, such as geology and elevation, which are used to assess stream health (Hawkins and Vinson, 2000; Rabeni and Doisy, 2000; Sandin et al., 2000; Waite et al., 2000). Landscape classification is appealing because it is not dependent upon intensive field surveys. Reach and microhabitat-scale data is typically collected using field surveys, but landscape-scale data must be collected remotely using GIS, aerial photography, or satellite images. This clearly reduces the need for intensive field sampling (Nelson et al., 1992; Richards and Host, 1994; Davies et al., 2000). Unfortunately, results indicate that landscape-scale data are correlated to aquatic community composition, but not strongly (Hawkins et al., 2000; Sandin, 2003). A hierarchy exists between the landscape, reach, and microhabitat filter scales, but landscape variables alone cannot account for the variability in BMI community structure because their influence is indirect and the difference in scale is too great (Hawkins et al., 2000).

Hierarchical classification

Evidence provided by previous studies suggests that hierarchical classification systems have the ability to explain a significant portion of the variability found in BMI community structure (Poff and Ward, 1989; Davies et al., 2000; Hawkins et al., 2000; Richards et al., 1996; Richards et al., 1997; Sandin, 2003; Mykra et al., 2004). The most important landscape variables appear to be those related to hydrologic (Statzner and Higler, 1986; Poff and Ward, 1989), physical (Nelson et al., 1992; Fox et al., 1996), and chemical (Johnson et al., 1997; Dauer et al., 2000) characteristics of the stream segment. For instance, Sandin and Johnson (2004) found a hierarchical relationship between agricultural land use, stream nutrient concentrations, and variance in BMI species data. This evidence indicates that a hierarchy of filter scales exists in stream ecosystems. Although a hierarchical classification system appears to be the most suitable for regional assessment of stream condition, it also requires local habitat data, which again must be collected using field sampling. Therefore, it seems necessary to develop a rapid and cost-efficient method that can be used to predict segment scale habitat conditions before a hierarchical classification system can be developed and applied over large areas.

Geostatistical modeling

Traditional statistical models represent broad-scale trends in the mean of the data (Bailey and Gatrell, 1995). The model errors are usually assumed to be normally distributed, homoscedastic (constant variance), and independent, so that the value at one site is not influenced by the values at other sites. Yet, without randomized designs these assumptions may be violated. Geostatistical models are similar to traditional statistical models because they can also represent the broad scale trend in the data. However, they relax the assumption of independence and allow spatial autocorrelation in the errors. Local deviations from the mean are modeled by the covariance between neighboring sites. Heterogeneity in the large scale mean and variance are permitted, but the mean, variance, and autocorrelation structure of the error term are assumed to be stationary or similar across a study area (Bailey and Gatrell, 1995).

The covariance represents the strength of spatial autocorrelation between two sites given their separation distance (h) (Olea, 1991). The separation distance is simply the distance traveled from one location to a second location. It is typically measured using straight-line distance, but can be calculated using a variety of distance measures. The covariance is the joint variation of the two values (z) at two locations (x and $x+h$) about the global mean ($\bar{z}$) and provides a quantitative measure of the way sites co-vary in space (Bailey and Gatrell, 1995). An empirical estimator of the covariance between all data pairs (n of them) at a separation distance h is an average of the cross-products.

$$Cov(h) = \frac{\sum\left[z(x)z(x+h) - \bar{z}^2\right]}{n_h} \tag{1.4}$$

An autocovariance function, such as the exponential function, can be fit to the estimated autocovariances using several moment-based and likelihood-based approaches (Cressie, 1993). The autocovariance has three parameters (θ), the nugget, sill, and range, which must be estimated by fitting the autocovariance function to the data (Figure 1.2). The nugget represents the variation between sites as their separation distance approaches zero. It can result from experimental error or could indicate that a substantial amount of variation occurs at a scale finer than the minimum separation distance in the survey. The sill is delineated where the autocovariance becomes asymptotic and represents variance found among independent data. The range parameter controls how fast the autocovariance decays with distance.

The mathematical structure of the fitted autocovariance function provides a way to estimate the local deviation from the mean value at an unobserved site using observed values at neighboring sites. Thus, geostatistical models are typically able to model more variability in the data and provide more accurate predictions at unobserved sites when spatial autocorrelation is present (Isaaks and Srivastava, 1989).

Geostatistical modeling in aquatic systems

Geostatistical modeling (Cressie, 1993) has been widely applied in terrestrial ecosystems to predict characteristics related to wildlife (Reich et al., 2000; Pleydell et al., 2004), agriculture (Jurado-Exposito et al, 2003; Dobermann and Ping, 2004), vegetation (Chong et al., 2001; Hudak et al., 2002), fire (Robichaud and Miller, 2000; Flores-Garnica and Omi, 2003), and snow (Bales et al., 2001; Josberger and Mognard, 2002). Geostatistical models have been used less frequently in aquatic systems such as lakes (Altunkaynak et al., 2003) and estuaries (Little et al., 1997; Rathbun, 1998), and rarely in stream networks (e.g. Kellum, 2003; Yuan, 2004a). Freshwater ecologists might have found little utility in geostatistical models because they are typically based on straight-line (also commonly called Euclidean distance) and symmetric distance. Straight-line distance is not ecologically representative of stream ecosystems because it fails to represent the spatial configuration, connectivity, directionality, and relative position of sites in a stream network. Therefore, it may not be a suitable distance measure for most chemical, physical, and biological studies of freshwater streams (Olden et al., 2001; Benda et al., 2004; Ganio et al., 2005). As a result, freshwater ecologists have begun to explore spatial patterns in stream networks using hydrologic distance measures and new geostatistical methodologies have recently been developed that enable directional hydrologic distance measures to be considered (Ver Hoef et al., in press).

Two types of distances can be used to represent physical and ecological processes in stream ecosystems: symmetric and asymmetric distance classes. These distances are used to characterize the spatial neighborhood for each site. A spatial neighborhood includes sites that are nearby and have a quantifiable influence upon one another. Sites outside of the spatial neighborhood are considered spatially independent. Symmetric distance is directionless (isotropic) and has equal correlation in all directions (or both on a stream). Straight-line distance (Figure 1.3a) is symmetric and all locations in a study area can be considered neighbors. *Hydrologic distance* can be either symmetric or asymmetric and is simply the distance between two locations when movement is restricted to the stream network. Symmetric hydrologic distance is the total upstream and downstream hydrologic distance between two sites (Figure 1.3b). Thus, all sites located within a stream network are neighboring sites (assuming all basins in the study area share a common outlet) because flow direction is disregarded. Asymmetric distances include unidirectional measures that are restricted to either the upstream or downstream flow direction. Water must flow from one location to another to be considered neighbors (Figure 1.3c). Metrics that represent relative network position, such as stream order or watershed area, can also be used to weight hydrologic distance measures to make them more ecologically representative.

Asymmetric hydrologic distance measures are unidirectional, but there is a symmetric correlation between flow-connected sites. This may seem counter intuitive because the measures represent asymmetric flow relationships. Even though a downstream site does not affect upstream sites, the conditions at the downstream site are, in part, a result of those found upstream. This concept also applies to autoregressive models in time series, where the model is constructed with later time events depending on earlier ones. Although time flows in one direction, the correlation between two time events is symmetric. If I wanted to predict forward in one unit of time from a single observed event the prediction would be the same as if we went back one unit of time (Barnett, 2004). However, there is an added twist to stream networks because the spatial neighborhood only includes flow-connected sites (Figure 1.3c). This unique characteristic makes a model based on asymmetric hydrologic distance dramatically different from models based on straight-line or symmetric hydrologic distance.

Stream processes must be modeled and analyzed at multiple scales (Frissell et al., 1986; Petts, 1994) and patterns of spatial autocorrelation provide additional information about spatial relationships in stream networks. If we ignore seasonal differences, stream conditions are influenced by the surrounding landscape (Davies et al., 2000), the geographic location of the study area (Hill et al., 2000), the spatial configuration, connectivity, and directionality of the stream network, and the relative position of the site in a stream network. However, the degree to which each of these factors affects conditions at a specific site is unclear. The identification and examination of spatial autocorrelation in streams data using a variety of distance measures will provide quantitative answers concerning the relative effect of ecosystem processes and spatial relationships occurring across multiple aquatic and terrestrial filter scales. Geostatistical models based on hydrologic distance measures may provide information about heterogeneity resulting from differences in tributary size or network connectivity. In addition, geostatistical modeling may reduce the need for additional field surveys required to collect local habitat data for input into a hierarchical classification system or for assessing regional compliance with water quality standards.

LITERATURE CITED

Alewell C., Mitchell M.J., Likens G.E., & Krouse H.R. (1999) Sources of stream sulfate at the Hubbard Brook Experimental Forest: Long-term analyses using stable isotopes. *Biogeochemistry*, 44, 281-299.

Allan J.D. (1995) *Stream ecology: Structure and function of running waters, 1st Edition.* Kluwer Academic Publishers, Dordrecht, The Netherlands.

Altunkaynak A., Ozger M., & Sen Z. (2003) Triple diagram model of level fluctuations in Lake Van, Turkey. *Hydrology and Earth System Sciences,* 7, 235-244.

Angelier E. (2003) *Ecology of streams and rivers.* Science Publishers, Inc., Enfield, NH, USA.

Bailey R.G. (1983) Delineation of ecosystem regions. *Environmental Management,* 7, 365-373.

Bailey T.C. and Gatrell A.C. (1995) *Interactive spatial data analysis*. Pearson Education Limited, Essex, England.

Baker J.P. and Schofield C.L. (1982) Aluminum toxicity to fish in acidic waters. *Water, Air and Soil Pollution*, 18, 289-309.

Bales R.C., McConnell J.R., Mosley-Thompson E., & Lamorey G. (2001) Accumulation map for the Greenland Ice Sheet. *Geophysical Research Letters,* 28, 2967-2970.

Barbour M.T., Gerritsen J., Snyder B.D., & Stribling J.B. (1999) *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, Second edition.* PA 841-B-99-002, U.S. Environmental Protection Agency, Office of Water, Washington, D.C.,

Barnett V. (2004) *Environmental statistics methods and applications.* John Wiley and Sons, Chichester, West Sussex, England.

Benda L., Poff N.L., Miller D., Dunne T., Reeves G., Pess G., & Pollock M. (2004) The network dynamics hypothesis: How channel networks structure riverine habitats. *BioScience*, 54, 413-427.

Biggs B.J., Jowett I.G., Quinn J.M., Hickey C.W., Davies-Colley R.J., & Close M. (1990) Ecological characterisation, classification, and modelling of New Zealand rivers: an introduction and synthesis. *New Zealand Journal of Marine and Freshwater Research*, 24, 277-304.

Blowes D.W., Ptacek C.J., Jambor J.L., & Weisener C.G. (2003) The geochemistry of acid mine drainage. In: *Anonymous treatise on geochemistry*, pp. 149-204, Elsevier-Pergamon, Oxford, England.

Camargo J.A. & Ward J.V. (1992) Short-term toxicity of sodium nitrate ($NaNO_3$) to non-target freshwater invertebrates. *Chemosphere*, 24, 23-28.

Canadian Council of Ministers and the Environment (2003) Canadian water quality guidelines for the protection of aquatic life: Nitrate Ion. *In: Canadian environmental quality guidelines, 1999*, Canadian Council of Ministers of the Environment, Winnipeg, Canada.

Canham C.D., Pace M.L., Papaik M.J., Primack A.G.B., Roy K.M., Maranger R.J., Curran R.P., & Spada D.M. (2004) A spatially explicit watershed-scale analysis of dissolved organic carbon in Adirondack lakes. *Ecological Applications*, 14, 839-854.

Caruso B.S. (2002) Temporal and spatial patterns of extreme low flows and effects on stream ecosystems in Otago, New Zealand, *Journal of Hydrology*, 257, 115-133.

Chambers P.A., Prepas E.E., & Gibson K. (1992) Temporal and spatial dynamics in riverbed chemistry - The influence of flow and sediment composition. *Canadian Journal of Fisheries and Aquatic Sciences*, 49, 2128-2140.

Chapin III F.S., Matson P.A., & Mooney H.A. (2002) *Principles of terrestrial ecosystem ecology*, Springer-Verlag New York, Inc., New York, NY, USA.

Chong G.W., Reich R.M., Kalkhan M.A., & Stohlgren T.J. (2001) New approaches for sampling and modeling native and exotic plant species richness. *Western North American Naturalist*, 61, 328-335.

Clements W.H. (1994) Benthic invertebrate community responses to heavy metals in the Upper Arkansas River Basin, Colorado. *Journal of the North American Benthological Society*, 13, 30-44.

Clements W.H., Carlisle D.M., Lazorchak J.M., & Johnson P.C. (2000) Heavy metals structure benthic communities in Colorado mountain streams. *Ecological Applications*, 10, 626-638.

Cooper D.M., Helliwell R.C., & Coull M.C. (2004) Predicting acid neutralizing capacity from landscape classification: application to Galloway, south-west Scotland. *Hydrological Processes*, 18, 455-471.

Copeland C. (2002) *Water quality: Implementing the Clean Water Act.* IB89102, Congressional Research Service, Resources, Science, and Industry Division, Washington, D.C.

Cressie N. (1993) *Statistics for spatial data, Revised edition.* John Wiley and Sons, New York, NY, USA.

Cressie N., Frey J., Harch B., & Smith M. (submitted) Spatial prediction on a river network. Department of Statistics Preprint No. 747, The Ohio State University, Submitted to *Journal of Agriculture, Biological and Environmental Statistics.*

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

Davies N.M., Norris R.H., & Thoms M.C. (2000) Prediction and assessment of local stream habitat features using large-scale catchment characteristics. *Freshwater Biology*, 45, 343-369.

Dawson J.J.C., Billett M.F., & Hope D. (2001) Diurnal variations in the carbon chemistry of two acidic peatland streams in north-east Scotland. *Freshwater Biology*, 46,1309-1322.

Dent C.L. & Grimm N.B. (1999) Spatial heterogeneity of stream water nutrient concentrations over successional time. *Ecology*, 80, 2283-2298.

Dobermann A. & Ping J.L. (2004) Geostatistical integration of yield monitor data and remote sensing improves yield maps. *Agronomy Journal*, 96, 285-297.

Driscoll C.T., Lawrence G.B., Bulger A.J., Butler T.J., Cronan C.S., Eagar C., Lambert K.F., Likens G.E., Stoddard J.L., & Weathers K.C. (2001) *Acid rain revisited: Advances in scientific understanding since the passage of the 1970 and 1990 Clean Water Act Amendments*. Hubbard Brook Research Foundation, Science Links™ Publication.

Eckhardt B.W. & Moore T.R. (1990) Controls on dissolved organic carbon concentrations in streams, Southern Quebec. *Canadian Journal of Fisheries and Aquatic Sciences*, 47, 1537-1544.

Ecosystem Health and Monitoring Program (EHMP) (2004) *Ecosystem Health and Monitoring Program 2002-2003 annual technical report.* Ecosystem Health and Monitoring Program, Morton Bay Waterways and Catchments Partnership Brisbane, Australia.

Environmental Mining Council of British Columbia (EMCBC) (1999) *Acid mine drainage: Mining and water pollution issues in BC.* Environmental Mining Council of British Columbia, Victoria, BC, Canada.

Fausch K.D., Torgersen C.E., Baxter C.V., & Li H.W. (2002) Landscapes to riverscapes: Bridging the gap between research and conservation of stream reaches. *BioScience*, 52, 483-498.

Feldman R.S. & Conner E.F. (1985) The relationship between pH and community structure of invertebrates in streams of the Shenandoah National Park, Virginia, USA. *Freshwater Biology*, 27, 261-276.

Field C.K., Siver P.A., & Lott A. (1996) Estimating the effects of changing land use patterns on Connecticut lakes. *Journal of Environmental Quality,* 25, 325-333.

Flores-Garnica J.G. and Omi P.N. (2003) Mapping forest fuels for spatial fire behavior simulations using geomatic strategies. *Agrociencia*, 37, 65-72.

Fox P.J.A., Naura M., & Raven P. (1996) Predicting habitat components for semi-natural rivers in the United Kingdom. In: Leclerc M., Capra H., Valentin S., Boudreault A., & Cote Y. (eds). *Ecohydaulics 2000. Proceedings of the 2nd International Symposium on Habitat Hydraulics.* INRS-Eau, Quebec City, Canada, 227-237.

Frissell C.A., Liss W.J., Warren C.E. & Hurley M.D. (1986) A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management*, 10, 199-214.

Fuller R.L., Kennedy B.P., & Nielson C. (2004) Macroinvertebrate responses to algal and bacterial manipulations in streams. *Hydrobiologia*, 523, 113-126.

Ganio L.M., Torgersen C.E., & Gresswell R.E. (2005) A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and the Environment*, 3, 138-144.

Gardner B., Sullivan P.J., & Lembo A.J. (2003) Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. *Canadian Journal of Aquatic Science*, 60, 344-351.

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

Golterman H.L. (1975) Chemistry. In: Whitton B.A. (ed.), *River Ecology*, Blackwell Scientific Publications, Oxford, London, pp. 39-80.

Gray L. (2004) Changes in water quality and macroinvertebrate communities resulting from urban stormflows in the Provo River, Utah, USA. *Hydrobiologia*, 518, 33-46.

Gregory S.V., Swanson F.J., McKee W.A., & Cummins K.W. (1991) An ecosystem perspective of riparian zones: Focus on links between land and water. *BioScience*, 41, 540-551.

Groom A.P. & Hildrew A.G. (1989) Food quality for detritivores in streams of contrasting pH. *Journal of Animal Ecology*, 58, 863-881.

Guyot J.L. & Wasson J.G. (1994) Regional pattern of riverine dissolved organic carbon in the Amazon Drainage Basin of Bolivia. *Limnology and Oceanography*, 39, 452-458.

Hall R.J., Likens G.E., Fiance S.B., & Hendrey G.R. (1980) Experimental acidification of a stream in the Hubbard Brook experimental forest, New Hampshire. *Ecology*, 61, 976-989.

Hall R.O. & Meyer J.L. (1998) The trophic significance of bacteria in a detritus based stream food web. *Ecology*, 79, 1995-2012.

Hawkins C.P., Kershner J.L., Bisson P.A., Bryant M.D., Decker L.M., Gregory S.V., McCullough D.A., Overton C.K., Reeves G.H., Steedman R.J. & Young M.K. (1993) A hierarchical approach to classifying stream habitat features. *Fisheries*, 18, 3-12.

Hawkins C.P. & Vinson M.R. (2000) Weak correspondence between landscape classifications and stream invertebrate assemblages: implications for bioassessment. *Journal of the North American Benthological Society*, 19, 501-517.

Hawkins C.P., Norris R.H., Gerritsen J., Hughs R.M., Jackson S.K., Johnson R.K. & Stevenson R.J. (2000) Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society*, 19, 541-556.

Hejzlar J., Dubrovsky M., Buchtele J., & Ruzicka M. (2003) The apparent and potential effects of climate change on the inferred concentration of dissolved organic matter in a temperate stream (the Malse River, South Bohemia). *The Science of the Total Environment*, 310, 143-152.

Herlihy, A.T., Kaufmann P.R., & Mitch M.E. (1990) Regional estimates of acid mine drainage impact on streams in the Mid-Atlantic and southeastern United States. *Water, Air, and Soil Pollution*, 50, 91-107.

Herlihy A.T., Stoddard J.L., & Johnson C.B. (1998) The relationship between stream chemistry and watershed land cover data in the Mid-Atlantic region, U.S. *Water, Air, and Soil Pollution*, 105, 377-386.

Herlihy A.T., Larsen D.P., Paulsen S.G., Urquhart N.S. & Rosenbaum B.J. (2000) Designing a spatially balanced, randomized site selection process for regional stream surveys: The EMAP Mid-Atlantic pilot study. *Environmental Monitoring and Assessment*, 63, 95-113.

Hill B.H., Hall R.K., Husby P., Herlihy A.T. & Dunne M. (2000) Interregional comparisons of sediment microbial respiration in streams. *Freshwater Biology*, 44, 213-222.

Houle D., Carignan R., Lachance M., & Dupont J. (1995) Dissolved organic carbon and sulfur in Southwestern Quebec Lakes: Relationships with catchment and lake properties. *Limnology and Oceanography*, 40, 710-717.

Hudak A.T., Lefsky M.A., Cohen W.B., & Berterretche M. (2002) Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sensing of the Environment*, 82, 397-416.

Hughes R. M. (1990) What can biological monitoring tell us about the environmental health of aquatic ecosystems? *Proceedings of the International Symposium on the Design of Water Quality Information Systems*, June 7-9, 1989, Fort Collins, CO, Colorado Water Resources Research Institute, 447-460.

Hynes H.B. (1960) *The biology of polluted waters*. Liverpool University Press, Liverpool, England.

Hynes H.B.N. (1975) The stream and its valley. *Verhandlugen der Internationalen Vereinigung für Theoretische und Angewandte Limnologie*, 19, 1-15.

Isaaks E.H. & Srivastava R.M. (1989) *An introduction to applied geostatistics*. Oxford University Press, Inc., New York, NY.

Jenny, H. (1941) *Factors of soil formation*. McGraw-Hill, New York, NY, USA.

Johnson L.B., Richards C., Host G.E., & Arthur J.W. (1997) Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshwater Biology*, 37, 193-208.

Jordan T.E., Correll D.L., & Weller D.E. (1997) Effects of agriculture on discharges of nutrients from coastal plain watersheds of Chesapeake Bay. *Journal of Environmental Quality*, 26, 836-848.

Josberger E.G. & Mognard N.M. (2002) A passive microwave snow depth algorithm with a proxy for snow metamorphism. *Hydrological Processes Special Issue*, 16, 1557-1568.

Jurado-Exposito M., Lopez-Granados F., Garcia-Torres L., Garcia-Ferrer A., de la Orden M.S., & Atenciano S. (2003) Multi-species weed spatial variability and site-specific management maps in cultivated sunflower. *Weed Science*, 51, 319-328.

Kang J.K., Song Y., Moon J.W., & Moon H.S. (2001) Water quality impact of mining in the Wolmyoung area of Korea, and its short-term changes. *Water, Air, and Soil Pollution*, 129, 349-367.

Kawakami T., Honoki H., & Yasuda H. (2001) Acidification of a small stream on Kureha Hill caused by nitrate leached from a forested watershed. *Water, Air, and Soil Pollution*, 130, 1097-1102.

Kellogg W.W., Cadle R.D., Allen E.R., Lazrus A.L., & Martell E.A. (1972) The sulfur cycle. *Science*, 175, 587-596.

Kellum B. (2003) *Analysis and modeling of acid neutralizing capacity in the Mid-Atlantic Highlands area*. MS Thesis, Colorado State University, 69 p.

Kiffney P.M. & Clements W.H. (1996) Effects of metals on stream macroinvertebrate assemblages from different altitudes. *Ecological Applications*, 6, 472-481.

Kiffney P.M., Clements W.H., & Cady T.A. (1997) Influence of ultraviolet radiation on the colonization dynamics of a Rocky Mountain stream benthic community. *Journal of the North American Benthological Society*, 16, 520-530.

Krogh A. (1965) *Osmotic regulation in aquatic animals*. Dover Publications, Inc., New York, NY, USA.

Lanka R.P., Hurbert W.A., & Wesche T.A. (1987) Relations of geomorphology to stream habitat and trout standing stock in small Rocky Mountain streams. *Transactions of the American Fisheries Society*, 116, 21-28.

Lehmkuhl D.M. (1974) Thermal regime alteration and vital environmental physiological signals in aquatic organisms. In: Gibbons J.W & Scharits R.R. (eds.), *Thermal Ecology*, AEC Symposium Series (CONF 730505), 216-222.

Leidy R.B. (1980) Aquatic organisms. In: Guthrie F.E. & Perry J.J. (eds.), *Introduction to environmental toxicology*, Elsevier North Holland, Inc., New York, NY, USA, pp. 120-134.

Lenat D.R. (1988) Water quality assessment of streams using a qualitative collection method for benthic macroinvertebrates. *Journal of the North American Benthological Society*, 7, 222-233.

Little L.S., Edwards D., & Porter D.E. (1997) Kriging in estuaries: as the crow flies, or as the fish swims? *Journal of Experimental Marine Biology and Ecology*, 213, 1-11.

Lotspeich F.B. & Platts W.S. (1982) An integrated land-aquatic classification system. *North American Journal of Fisheries Management,* 2, 138-149.

Maddock I. (1999) The importance of physical habitat assessment for evaluating river health. *Freshwater Biology*, 41, 373-391.

Meade M.E. & Watts S.A. (1995) Toxicity of ammonia, nitrite, and nitrate to juvenile Australian crayfish, Cherax quadricarinatus. *Journal of Shellfish Research.* 14, 341-346.

Mercurio G., Chaillou J.C., & Roth N.E. (1999) *Guide to using 1995-1997 Maryland Biological Stream Survey Data, Maryland Department of Natural Resources.* Maryland Department of Natural Resources, Annapolis, MD.

Meybeck M. (1982) Carbon, nitrogen, and phosphorus transport by world rivers. *American Journal of Science*, 282, 401-450.

Meyer J.L (1994) The microbial loop in flowing waters. *Microbial Ecology*, 28, 195-199.

Mulholland P.J. (2003) Large-scale patterns in dissolved organic carbon concentration, flux, and sources. In: Findlay, S.E.G. & Sinsabough R.L. (eds.) *Aquatic ecosystems: Interactivity of dissolved organic matter*, Academic Press, San Diego, CA, USA, 139-159.

Mykra H., Heino J., & Muotka T. (2004) Variability of lotic macroinvertebrate assemblages and stream habitat characteristics across hierarchical landscape classifications. *Environmental Management*, 34, 341-352.

Nelson R.L., Platts W.S., Larsen D.P., & Jenson S.E. (1992) Trout distributions and habitat in relation to geology and geomorphology in the North Fork Humbolt River Drainage, Northeastern Nevada. *Transactions of the American Fisheries Society*, 121, 405-426.

Nelson P.N., Cosaris E., & Oades J.M. (1996) Nitrogen, phosphorus, and organic carbon in streams draining two grazed catchments. *Journal of Environmental Quality*, 25, 1221-1229.

Nordin R.N. & Pommen L.W. (1986) Water quality criteria for nitrogen (nitrate, nitrite, and ammonia): Technical appendix. *In: Water Quality Criteria.* Water Management Branch, British Columbia Ministry of Environment and Parks. Victoria, BC.

Olden J.D., Jackson D.A., & Peres-Neto P.R. (2001) Spatial isolation and fish communities in drainage lakes. *Oecologia*, 127, 572-585.

Olea R.A. (1991) *Geostatistical glossary and multilingual dictionary.* Oxford University Press, New York, NY, USA.

Olsen A.R. & Ivanovich M. (1993) EMAP monitoring strategy and sampling design. *Office of Research and Development Environmental Research Laboratory Corvallis (OR): U.S. Environmental Protection Agency*, Corvallis, OR, video.

Osborne L.L. & Wiley M.J. (1988) Empirical relationships between land use/cover and stream water quality in an agricultural watershed. *Journal of Environmental Management*, 26, 9-27.

Peiffer S., Beierkuhnlein C., Sandhage-Hofmann A., Kaupenjohann M., & Bar S. (1997) Impact of high aluminum loading on a small catchment area (Thuringia slate mining area) - Geochemical transformations and hydrological transport. *Water, Air, and Soil Pollution*, 94, 401-416.

Petts G.E. (1994) Rivers: Dynamic components of catchment ecosystems. In: *The rivers handbook, Vol. 2*, Blackwell Scientific Publications, Oxford, England, 3-22.

Platts W.S. (1979) Relationships among stream order, fish populations, and aquatic geomorphology in an Idaho river drainage. *Fisheries*, 4, 5-9.

Pleydell D.R.J., Raoul F., Tourneux F., Danson F.M., Graham A.J., Craig P.S. & Giraudoux P. (2004) Modeling the spatial distribution of Echinococcus multilocularis infection in foxes. *Acta Tropica*, 91, 253-265.

Poff N.L. & Ward J.V. (1989) Implications of streamflow variability and predictabiliy for lotic community structure: A regional analysis of streamflow patterns. *Canadian Journal of Fisheries and Aquatic Sciences*, 46, 1805-1817.

Poff N.L. (1997) Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society,* 16, 391-409.

Pringle C.M. (1991) Geothermally modified waters surface at La Selva Biological Station, Costa Rica: Volcanic processes introduce chemical discontinuities into lowland tropical streams. *Biotropica,* 23, 523-529.

Prusha B.A. & Clements W.H. (2004) Landscape attributes, dissolved organic C, and metal bioaccumulation in aquatic macroinvertebrates (Arkansas River Basin, Colorado). *Journal of the North American Benthological Society*, 23, 327-339.

Qualls R.G. & Haines B.L. (1992) Biodegradability of dissolved organic matter in forest throughfall, soil solution, and stream water. *Soil Science Society of American Journal*, 56, 578-586.

Quinn J.M. & Hickey C.W. (1990) Characterisation and classification of benthic macroinvertebrate communities in 88 New Zealand rivers in relation to environmental factors. *New Zealand Journal of Marine and Freshwater Research*, 24, 387-409.

Rabeni C.F. & Doisy K.E. (2000) Correspondence of stream benthic invertebrate assemblages to regional classification schemes in Missouri. *Journal of the North American Benthological Society*, 19, 419-428.

Ramsey D.L. & Brannon D.G. (1988) Predicted acid mine drainage impacts to the Buckhannon River, W.V., USA. *Water, Air, and Soil Pollution*, 39, 1-14.

Rathbun S.L. (1998) Spatial modeling in irregularly shaped regions: Kriging estuaries. *Environmetrics*, 9, 109-129.

Reich R.M., Lundquist J., & Bravo V.A. (2000) Spatial relationship of resident and migratory birds and canopy openings in diseased ponderosa pine forests. *Environmental Modeling and Software*, 15, 189-197.

Richards C. & Host G. (1994) Examining land use influences on stream habitats and macroinvertebrates: A GIS approach. *Water Resources Bulletin*, 30, 729-738.

Richards C., Johnson L.B. & Host G.E. (1996) Landscape-scale influences on stream habitats and biota. *Canadian Journal of Fisheries and Aquatic Science*, 53, 295-311.

Richards C., Haro R.J., Johnson L.B., & Host G.E. (1997) Catchment and reach-scale properties as indicators of macroinvertebrate species traits. *Freshwater Biology*, 37, 219-230.

Robichaud P.R. & Miller S.M. (2000) Spatial interpolation and simulation of post-burn duff thickness after prescribed fire. *International Journal of Wildland Fire*, 9, 137-143.

Roth N.E., Allan J.D., & Erickson D.L. (1996) Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape Ecology*, 11, 141-156.

Rounick J.S. & Winterbourn M.J. (1983) The formation, structure and utilization of stone surface organic layers in two New Zealand streams. *Freshwater Biology*, 13, 57-72.

Sandin L. & Johnson R.K. (2000) Ecoregions and benthic macroinvertebrate assemblages of Swedish streams. *Journal of the North American Benthological Society*, 19, 462-474.

Sandin L. (2003) Benthic macroinvertebrates in Swedish streams: community structure, taxon richness, and environmental relations. *Ecography*, 26, 269-282.

Sandin L. & Johnson R.K. (2004) Local, landscape and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. *Landscape Ecology*, 19, 501-514.

Sheets T.J. (1980) Agricultural pollutants. In: Guthrie F.E. and Perry J.J. (eds), *Introduction to environmental toxicology*, Elsevier North Holland, Inc., New York, NY, USA, 24-33.

Shirazi M.A., Boersma L., Johnson C.B., & Haggerty P.K. (2001) Predicting physical and chemical water properties from relationships with watershed soil characteristics. *Journal of Environmental Quality*, 30, 112-120.

Singleton H. (2000) *Ambient water quality guidelines for sulphate*. Ministry of Environment, Lands and Parks, Victoria, BC, Canada.

Smart M.M., Barney T.W., & Jones J.R. (1981) Watershed impact on stream water quality: A technique for regional assessment. *Journal of Soil and Water Conservation*, 36, 297-300.

Smith, K. (1981) The prediction of river water temperatures. *Hydrological Sciences Bulletin*, 26, 19-32.

Statzner B. & Higler B. (1986) Stream hydraulics as a major determinant of benthic invertebrate zonation patterns. *Freshwater Biology*, 16, 127-139.

Strahler A.N. (1957) Quantitative analysis of watershed geomorphology. *Transactions of the American Geophysical Union*, 21, 913-920.

Sweeney B.W. & Schnack J.A. (1977) Egg development, growth, and metabolism of Sigara alternata (say) (Hemiptera: Corixidae) in flucturating thermal environments. *Ecology*, 58, 265-277.

Thomsen A.G. & Friberg N. (2002) Growth and emergence of the stonefly Leuctra nigra in coniferous forest streams with contrasting pH. *Freshwater Biology*, 47, 1159-1172.

U.S. Environmental Protection Agency (USEPA) (2001) *Survey designs for sampling surface water condition in the West.* A620/R-01/004c. USEPA, Office of Research and Development, Washington, DC.

Vannote R.L. & Sweeney B.W. (1980) Geographic analysis of thermal equilibria: A conceptual model for evaluating the effect of natural and modified thermal regimes on aquatic insect communities. *American Naturalist*, 115, 667-695.

Ver Hoef J.M., Peterson E.E., & Theobald D.M. (In Press) Some new spatial statistical models for stream networks. *Environmental and Ecological Statistics.*

Waite I.R., Herlihy A.T., Larsen D.P., & Klemm D.J. (2000) Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. *Journal of the North American Benthological Society*, 19, 429-441.

Ward J.V. (1992) *Aquatic insect ecology.* John Wiley & Sons, Inc., New York, NY, USA.

Wetzel R.G. (1992) Gradient-dominated ecosystems: sources and regulatory functions of dissolved organic matter in freshwater ecosystems. *Hydrobiologia*, 229, 181-198.

Willoughby L.G. & Mappin R.G. (1988) The distribution of Ephemerella ignita (Ephemeroptera) in streams: the role of pH and food resources. *Freshwater Biology*, 19, 145-155.

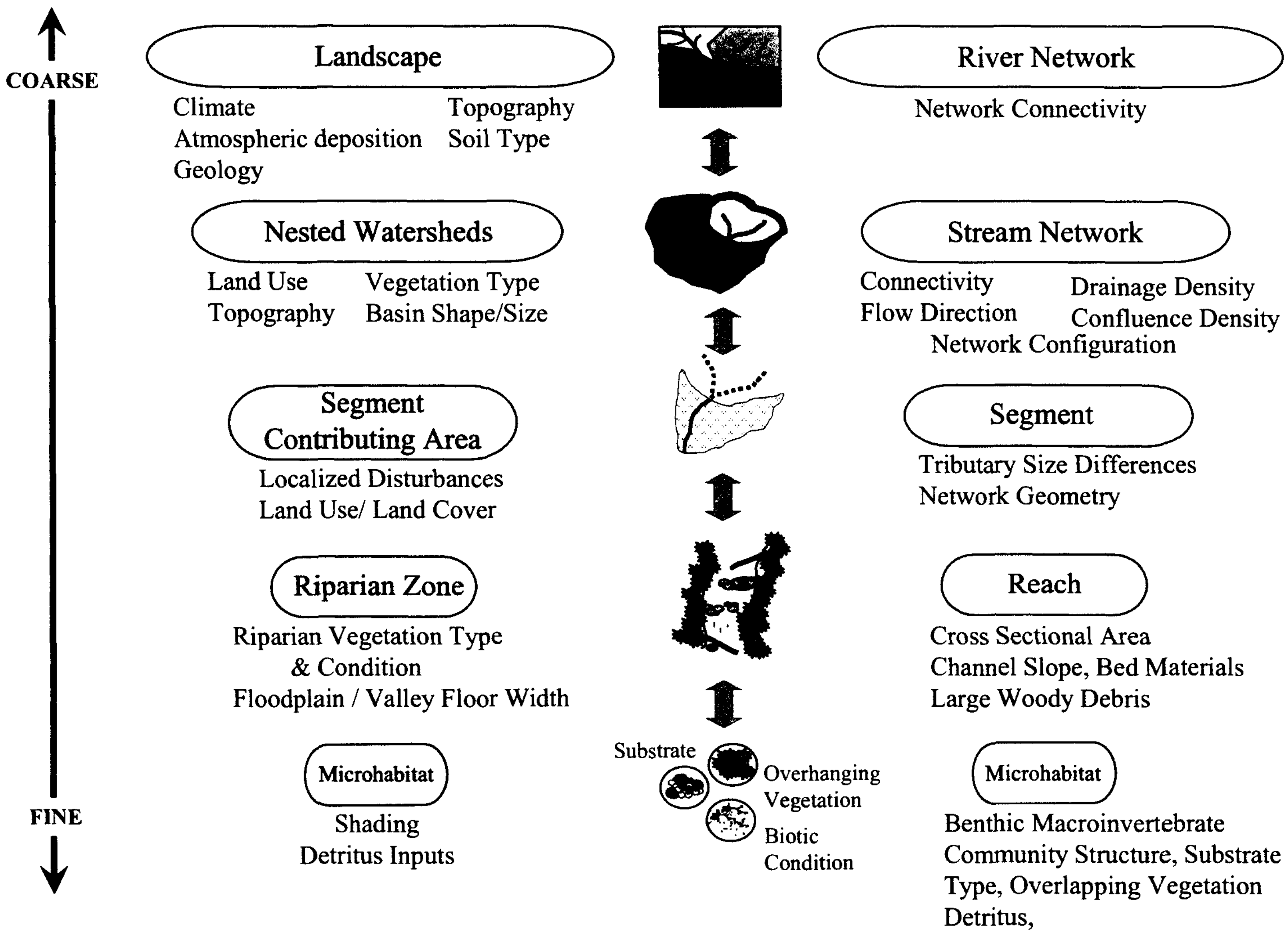
Yuan L.L. (2004a) Using spatial interpolation to estimate stressor levels in unsampled streams. *Environmental Monitoring and Assessment*, 94, 23-38.

Yuan L.L. (2004b) Assigning macroinvertebrate tolerance classifications using generalized additive models. *Freshwater Biology*, 49, 662-677.

Table 1.1 Functional linkage between the landscape and benthic macroinvertebrates (BMI). The following summary gives examples of how eight chemical response variables act as a function linkage between the landscape and BMI community structure.

Chemical response variable	Effects on BMI	Influential coarse-scale landscape variables
pH	emergence rates, egg failure, physiological problems, availability & quality, bioavailability of metals	atmospheric deposition of nitrogen & sulfur, agriculture, mining, geology, climate
acid neutralizing capacity	fertility, growth & emergence	soils, geology, climate, elevation, stream order, agriculture, urban areas, forest cover
nitrate	increased species richness, biological oxygen demand, growth & mortality rates	agriculture, urban areas
sulfate	bioavailability of metals	atmospheric deposition of sulfur, urban areas, mining, geology
dissolved organic carbon	energy, bioavailability of metals, ultraviolet rays	wetlands, forest cover, temperature, predcipitation, flow path of water, soil properties
conductivity	osmoregulation	geology, urban areas, agriculture, mining
temperature	emergence patterns, growth rates, metabolism, reproduction, body size	air temperature, elevation, stream order
dissolved oxygen	aerobic respiration	agriculture, urban areas, forests

SCALE: Grain


COARSE
FINE
Landscape
Climate
Atmospheric deposition
Geology
Topography
Soil Type
River Network
Network Connectivity
Nested Watersheds
Land Use
Topography
Vegetation Type
Basin Shape/Size
Stream Network
Connectivity
Flow Direction
Drainage Density
Confluence Density
Network Configuration
Segment Contributing Area
Localized Disturbances
Land Use/ Land Cover
Segment
Tributary Size Differences
Network Geometry
Riparian Zone
Riparian Vegetation Type & Condition
Floodplain / Valley Floor Width
Reach
Cross Sectional Area
Channel Slope, Bed Materials
Large Woody Debris
Microhabitat
Shading
Detritus Inputs
Substrate
Overhanging Vegetation
Biotic Condition
Microhabitat
Benthic Macroinvertebrate Community Structure, Substrate Type, Overlapping Vegetation Detritus,

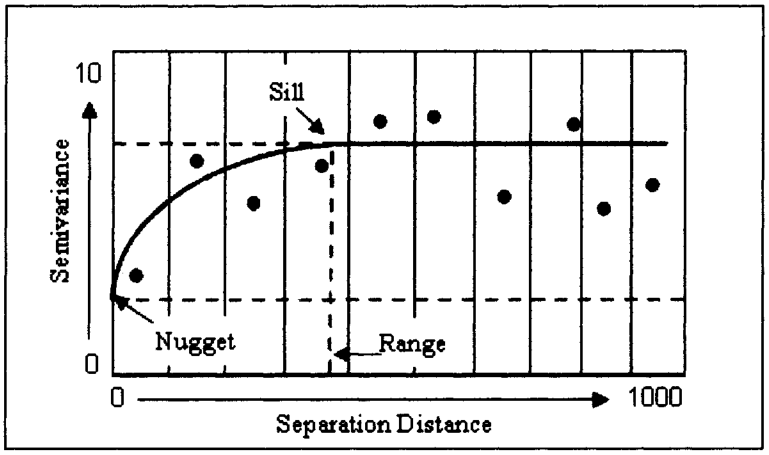
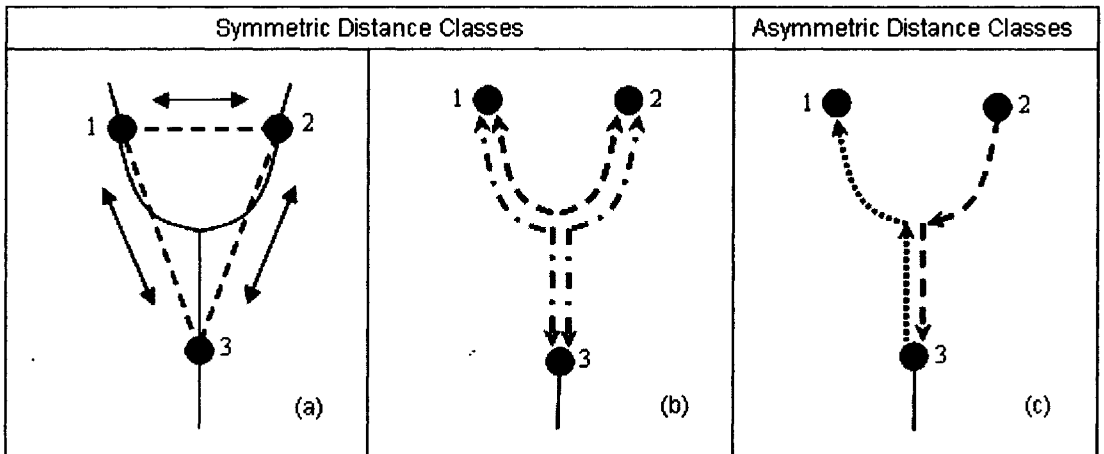


Figure 1.2 - The semivariogram. Semivariograms are generated by grouping the separation distances into bins, calculating a mean semivariance for each bin, and plotting them in ascending order. Each mean semivariance is represented by a single point. The semivariogram is used to derive visual estimates of the nugget, range, and sill.



From

(a)

To	1	2	3
1	1	1	1
2	1	1	1
3	1	1	1

(b)

	1	2	3
1	1	1	1
2	1	1	1
3	1	1	1

(c)

	1	2	3
1	1	0	0
2	0	1	0
3	1	1	1

Figure 1.3 - Symmetric and asymmetric distance classes and their associated connectivity matrices (1 = connected, 0 = not connected). The stream network is represented by a solid line, while distance measurements are represented with dashed lines. Symmetric hydrologic distance measures include straight line distance (a) and symmetric hydrologic distance (b). Sites 1, 2, and 3 are all neighbors to one another when these distance measures are used. Asymmetric distance classes include upstream and downstream asymmetric hydrologic distance (c). Sites 1 and 2 are neighbors to site 3, but not to each other.

CHAPTER TWO

SUPPORT FOR GEOSTATISTICAL MODELING ON STREAM NETWORKS: DEVELOPING VALID COVARIANCE MATRICES BASED ON HYDROLOGIC DISTANCE AND STREAM FLOW[1]

INTRODUCTION

Geostatistical modeling (Cressie, 1993) has been widely applied in terrestrial ecosystems. It has been used to predict characteristics related to wildlife (Reich, Lundquist, and Bravo, 2000; Pleydell et al., 2004), agriculture (Jurado-Exposito et al, 2003; Dobermann and Ping, 2004), vegetation (Chong et al., 2001; Hudak et al., 2002), fire (Robichaud and Miller, 2000; Flores-Garnica and Omi, 2003), and snow (Bales et al., 2001; Josberger and Mognard, 2002). Geostatistical models are not usually applied in aquatic systems such as lakes (Altunkaynak, Ozger, and Sen, 2003) and estuaries (Little, Edwards, and Porter, 1997; Rathbun, 1998), and have rarely been applied to stream networks (e.g. Kellum, 2003; Yuan, 2004). Freshwater ecologists may not have found geostatistical models useful because they are typically based on straight-line (also commonly called Euclidean) distance and symmetric distance. Straight-line distance is not ecologically representative because it does not characterize the spatial configuration,

---

[1] This paper was submitted to Freshwater Biology and my co-authors were David M. Theobald and Jay M. Ver Hoef.

connectivity, directionality, and relative position of sites in a stream network. Therefore, it may not be a suitable distance measure for most chemical, physical, and biological studies of freshwater streams (Olden, Jackson, and Peres-Neto, 2001; Benda et al., 2004; Ganio, Torgersen, and Gresswell, 2005). Rather, hydrologic distances that reflect ecological processes of stream ecosystems are more likely to be useful to freshwater ecologists.

Our objective is to discuss and demonstrate methods used to create valid covariance matrices based on hydrologic distance measures. Geostatistical modeling assumptions are reviewed and the statistical consequences of substituting hydrologic distance measures for straight-line distance will be explained. We also describe a methodology to generate a valid covariance matrix based on asymmetric hydrologic distance weighted by discharge volume.

Two types of distances are often used to represent physical and ecological processes in stream ecosystems: symmetric and asymmetric distance classes. The purpose of a distance measure is to characterize the spatial neighborhood for each site. A spatial neighborhood includes nearby sites that have a quantifiable influence upon one another. Sites outside of the spatial neighborhood are considered uncorrelated. Symmetric distance is directionless (isotropic) and has equal correlation in all directions (or both on a stream). Straight-line distance (Figure 2.1a) is one example of a symmetric distance measure. When straight-line distance is used all locations in a study area can be considered neighbors. *Hydrologic distance* can be either symmetric or asymmetric and is simply the distance between two locations when movement is restricted to the stream network. Symmetric hydrologic distance ignores flow direction and includes both upstream and downstream hydrologic distance between two sites (Figure 2.1b). All sites located on a single stream network are neighboring sites if all basins in the study area share a common outlet. Asymmetric distances include unidirectional measures that are restricted to either the upstream or downstream flow direction. Sites are considered neighbors if water flows from one location to another (Figure 2.1c). There is a symmetric correlation between flow-connected sites, but the unique spatial neighborhood makes a model based on asymmetric hydrologic distance significantly different from models based on straight-line or symmetric hydrologic distance. Metrics, such as stream order or watershed area, which represent relative network position can also be used to weight hydrologic distance measures to make them more ecologically representative.

Terrestrial processes may be better represented with straight-line distance because the terrestrial landscape is also represented as a two-dimensional surface where any two sites may be connected. At times, it may also be appropriate to apply straight-line distance to stream ecosystems. For example, an aquatic response variable may be significantly influenced by a continuous landscape variable, such as geology type (Kellum, 2003) or by broad scale factors, such as acid precipitation (Driscoll et al., 2001). However, straight-line distance may not be as useful in freshwater riverine systems, which differ from terrestrial ecosystems because they are linear and typically the movement of material is restricted to the stream network. Although freshwater systems have four dimensions: longitudinal, lateral, vertical, and temporal (Ward,1989), we focus solely on the longitudinal dimension. For instance, some fish move both up and downstream, but cannot move across the terrestrial landscape. Other materials, such as seeds or chemicals, are passive movers primarily affected by longitudinal transport. In these cases, movement is restricted to the network, but occurs primarily in the downstream direction. The relative position in the network also affects the condition of a site (Pringle, 2001; Benda et al., 2004) and reflects the influence that it will have on other sites (Cumming, 2002). For example, a site located on a small tributary may have little influence on a downstream site located on the mainstem due to substantial differences in discharge volume (Benda et al., 2004). Clearly, physical characteristics of the stream network provide a vast amount of information about conditions at unobserved sites and therefore, functional distances based on hydrology should be considered.

Traditional statistical models represent global trends in the mean of the data (Bailey and Gatrell, 1995). The model errors must be normally distributed, homoscedastic (constant variance), and independent, so that values at sites are considered spatially uncorrelated. Yet, these assumptions may be violated if randomized designs are not used. Geostatistical models are similar to traditional statistical models since they can also represent the global trend in the data. However, the assumption of independence is relaxed and they allow spatial autocorrelation in the errors. Local variations from the mean are modeled using the covariance between neighboring sites. Heterogeneity in the global mean and variance are permitted, but the mean, variance, and autocorrelation structure of the error term are assumed to be stationary or similar across a study area (Bailey and Gatrell, 1995).

The covariance represents the strength of spatial autocorrelation between two sites given their separation distance (h) (Olea, 1991). The separation distance is simply the distance traveled from one location to a second location and can be calculated using a variety of distance measures. The covariance is the joint variation of the values (x_i, x_j) at two locations (i and j) about their means (m_i, m_j) and provides a quantitative measure of the way sites co-vary in space (Bailey and Gatrell, 1995). An empirical estimator of the covariance between all data pairs (n of them) at a separation distance h is an average of the cross-products.

$$Cov(h) = \frac{1}{n}\sum\left[\left(x_i - m_i\right)\left(x_j - m_j\right)\right] \tag{2.1}$$

An autocovariance function, such as the exponential function, can be fit to the estimated autocovariances using several moment-based and likelihood-based approaches (Cressie, 1993). There are three autocovariance has three parameters (θ): the nugget, sill, and range. They are estimated by fitting the autocovariance function to the data (Figure 2.2). The nugget represents the variation between sites as their separation distance approaches zero. It may result from experimental error introduced during sampling or could suggest that a considerable amount of variation occurs at a scale finer than the survey scale. The sill is delineated where the autocovariance asymptotes and represents variance found among independent data. The range parameter controls how fast the autocovariance decays with distance.

The local deviation from the mean value at an unobserved site is estimated using the mathematical structure of the fitted autocovariance function, which is based on the observed values at neighboring sites. Thus, geostatistical models are typically able to provide more accurate predictions of unobserved sites when spatial autocorrelation is present in the data (Isaaks and Srivastava, 1989).

In general, covariance matrices contain the covariance between each site and every other site. They have n rows and n columns, where n is the total number of sample sites. In geostatistics, these covariances are obtained from the fitted autocovariance function. Not all functions can be used because the covariance matrix must be symmetric, positive definite, and all diagonal elements must be non-negative (Cressie, 1993). However, a review of methods used to examine these assumptions is beyond the scope of this manuscript. See Isaaks and Srivastava (1989) for a detailed discussion of positive definite matrices and rules used to test them.

A few notable efforts have used symmetric hydrologic distance to explore spatial autocorrelation in stream networks. Their results indicate that patterns of spatial autocorrelation in biological (Torgersen et al., 2004; Ganio et al., 2005), chemical (Dent and Grimm, 1999; Gardner, Sullivan, and Lembo, 2003), and physical (Legleiter et al., 2003) stream data can be represented using symmetric hydrologic distance. However, predictions at unobserved locations using covariance matrices based solely on symmetric hydrologic distance are only valid when the exponential autocovariance function is used (Ver Hoef, Peterson, and Theobald, in press). When symmetric hydrologic distance is substituted for straight-line distance in other commonly-used geostatistical autocovariance functions, the covariance matrix may contain negative eigenvalues, it may produce negative variance estimates, it is not guaranteed to be positive definite, and therefore it is not valid. To our knowledge, there have been no valid predictions generated using geostatistical models based on symmetric hydrologic distance measures.

Asymmetric hydrologic distance measures (Figure 2.1c) represent network configuration and flow direction, which strongly influence ecological and physical processes in stream networks (Olden et al., 2001; Fagan, 2002). However, no valid geostatistical models have been developed that are based purely on asymmetric hydrologic distance. Covariance matrices must be symmetric and therefore valid covariance matrices based purely on asymmetric hydrologic distance cannot be generated. In spite of these statistical challenges, we argue that asymmetric distance measures are ecologically valid because they more truthfully represent the connectivity and flow relationships in stream networks. We also believe that they would provide a useful tool for geostatistical modeling in stream networks. To address this issue, we developed spatial autocovariance models using a moving average construction and based on asymmetric hydrologic distance. These autocovariance models produce covariance matrices that are both statistically and ecologically valid for stream networks (Ver Hoef et al., in press).

Barry and Ver Hoef (1996) show that a large class of autocovariances can be developed by creating random variables as the integration of a moving-average function over white-noise random process.

$$Z(s) = \int_{-\infty}^{\infty} g(x - s \mid \theta) W(x) dx, \tag{2.2}$$

where $Z(s)$ is the random variable Z at a spatial location s, x is the distance from location s, θ represents the covariance parameters, $W(x)$ is a white noise process and $g(x|\theta)$ is the moving average function. The moving average construction allows a valid autocovariance to be expressed as,

$$C(h \mid \theta) = \begin{cases} \int_{-\infty}^{\infty} (g(x \mid \theta))^2 dx + \theta_0 & \text{if h=0} \\ \int_{-\infty}^{\infty} g(x \mid \theta) g(x-h \mid \theta) dx & \text{if h >0} \end{cases} \tag{2.3}$$

where θ_0 is the nugget and h is the separation distance. We also assume that the integral exists.

The moving average would be a standard construction for a single line segment that was continuous from $-\infty$ to ∞ on the real line, such as for time series models. However, for stream networks, the line segments split at stream confluences. Hence, spatial weights are assigned to each stream segment in the network and are used to split the moving average function at confluer $+\sum_{j \in U_{s_i}} \prod_{k \in B_{s_i \downarrow j}} \sqrt{\omega_k} \int_{l_j}^{u_j} g(x_j - s_i \mid \theta) W(x_j) dx_j,$ be equally divided among the upstream stream segments. The weights could also be based on a more ecologically relevant measure, such as stream order or discharge. However, the weights must sum to one at each stream confluence in order to maintain stationarity in the variances.

For a stream network, the construction that is equivalent to Equation 2.2 is

$$Z(s_i) = \int_{s_i}^{u_i} g(x_i - s_i \mid \theta) W(x_i) dx_{i}x_{i})dx_{i} \tag{2.4}$$

$$+\sum_{j \in U_{s_i}} \prod_{k \in B_{s_i \downarrow j}} \sqrt{\omega_k} \int_{l_j}^{u_j} g(x_j - s_i \mid \theta) W(x_j) dx_j,$$

where

$Z(s_i)$ is the random variable Z at a spatial location on the stream segment s_i,

x_i and x_j are the distance upstream on the i^{th} and j^{th} stream segment,

u_i and u_j are the most upstream locations on the i^{th} and j^{th} stream segment , where u can be ∞ if there are no stream segments upstream of a segment,

l_i and l_j are the most downstream locations on the i^{th} and j^{th} stream segment,

$\prod_{k \in B_{i,j}} \sqrt{\omega_k}$ is the square root of the percentage of watershed area (used as a surrogate for discharge volume) that the downstream location receives from the upstream location, and

U_{x_i} is an index set of stream segments upstream of x_i, excluding i.

We use the definition given in Equation 2.4 to build valid autocovariance models for stream networks.

$$C(s_i, s_j \mid \theta) = \begin{cases} 0 & \text{locations are not flow connected,} \\ C_1(0) + \theta_0 & \text{if location 1 = location 2,} \\ \Pi_{j \in B_D} \sqrt{w_j}\, C_1(h) & \text{otherwise.} \end{cases} \tag{2.5}$$

and

$$C_1(h) = \int_{-\infty}^{\infty} g(x \mid \theta) g(x - h \mid \theta) dx .$$

where s_i and s_j represent two spatial locations on the stream network. The exponential, linear with sill, spherical, and Mariah autocovariance functions can be fit to the asymmetric hydrologic distance data once the distances are weighted appropriately (Ver Hoef et al., in press). A valid covariance matrix produced using the moving average autocovariance models can be used with common kriging equations (Cressie, 1993).

This advance makes it possible to create valid geostatistical models, but calculating distances and spatial weights along stream networks remains challenging. A handful of geographic information system (GIS) tools have been developed to calculate hydrologic distance, including the National Hydrography Dataset (NHD) ArcView Toolkit version 7.0 (USGS, 2004a) from the United States Geological Survey (USGS) and the Environmental Systems Research Institute (ESRI) Arc Hydro Tools (ESRI, 2001a). However, each of these tools has practical and theoretical limitations. The NHD ArcView Toolkit allows the user to calculate the hydrologic distance between two sites based on flow direction, but does not provide a tool to calculate the spatial weights. In addition, the code is encrypted, which makes it impossible to automate the tool so that processing large datasets is difficult and time consuming. The Arc Hydro Tools do not calculate the hydrologic distance between sites or the spatial weights without extensive data pre-processing and programmatic modification. Other researchers have developed their own scripts, written in Avenue, Arc Macro Language, or Visual Basic, to calculate the hydrologic distance between sample sites (Rathbun, 1998; Theobald, 2002; Gardner et al., 2003; Dussault and Brochu, 2003; Torgersen et al., 2004), but they were not designed to calculate the spatial weights for the stream network. We believe that some practical and accessible tools are needed to generate the hydrologic distance matrices and spatial weights matrices in a cost efficient manner. The purpose of this manuscript is to make the methodology and tools readily available to others in order to facilitate valid geostatistical modeling in freshwater riverine systems.

METHODS

We detail how to calculate a valid covariance matrix based on asymmetric hydrologic distance and weighted by stream discharge. There are three steps: 1) data pre-processing, 2) generating the hydrologic distance and spatial weights matrices, and 3) creating a statistically valid covariance matrix. We also provide an example to illustrate these methods.

Preprocessing streams data

One challenge of working with GIS data is that sample sites collected within a stream are not always located directly on a stream segment, even though they should be. This is a common phenomenon that can result for a variety of reasons. Though GPS-based points are differentially corrected, they still have some error and do not always fall directly on a vertex or line segment representing a stream. Some stretches of river can move (e.g. meander) slightly from their mapped position. Streams are often represented on a map by lines and so samples collected on the banks of a large river may not fall directly on a line segment. When streams are represented at coarser scales the digital streams datasets may contain mapping errors and generalizations, such as the absence of small tributaries and the homogenization of form. Regardless of the error source, the sample sites must fall exactly on a stream line. Our solution is to "snap" the sites to the nearest stream segment, but a refinement might be to move the site to the nearest stream segment and compare attributes, such as stream name or upstream watershed area, to ensure that it lies in the correct location (Mixon, 2002). In contrast, the NHD Reach

Indexing Tool (USEPA, 2002) uses dynamic segmentation to relate features to NHD reaches without moving the feature or altering the reach.

Calculating hydrologic distance measures

We developed a program in Visual Basic Applications for ArcGIS version 8.3 (ESRI, 2002) to locate the path between sample sites and to compute the hydrologic distance between geographic locations. The total distance traversed in the downstream direction is recorded in a distance table (Figure 2.3b), which provides sufficient information to calculate symmetric and asymmetric hydrologic distance measures. Flow direction is retained by recording the downstream distance in both directions (e.g. downstream from A to B and downstream from B to A). The upstream distance between two sites is found by switching the direction of the path and the symmetric hydrologic distance is calculated by summing the two downstream distances. Asymmetric hydrologic distance is restricted to flow-connected sites, which are identified by comparing the downstream distances between sites. If the distance is greater than zero in one direction and equal to zero in the other, then the two sites are connected by flow. They are not connected if the downstream distance is equal to zero in both directions. The symmetric or asymmetric hydrologic distance measures are calculated and recorded in an *n* by *n* hydrologic distance matrix and output as a comma delimited text file, which is compatible with most statistical software.

Calculating the spatial weights

The proportional influence (PI) is defined as the influence of an upstream location on a downstream location and is used to create a spatial weights matrix. It is based on discharge volume, which can be calculated using regression equations (Vogel, Wilson, and Daly, 1999; USGS, 2004b) or process-based models such as the Soil and Water Assessment Tool (Neitsch et al., 2002). We use watershed area as a surrogate for discharge, which appears to be a viable alternative. Vogel and others (1999) evaluated discharge in 18 regions through the United States and found that drainage area and mean annual discharge were correlated in every region of the United States. The adjusted r^2 value ranged between 0.273 to 0.991 and the average was 0.714. We calculated the upstream watershed area for each stream segment in the network using a GIS. For our purposes, a stream segment is defined as a single line feature in a vector dataset. The area was stored as a segment attribute in the streams dataset and represents the watershed area for the downstream node of the stream segment.

Calculating the PI of one sample site on another is a two-step process. First, the PI of each stream segment on the segment directly downstream must be calculated and recorded in the streams attribute table (Figure 2.4). This was accomplished using a VBA program implemented in ArcGIS version 8.3 (ESRI, 2002). At each node in the network, the incoming segment(s) are identified. The total incoming area is calculated by summing the cumulative watershed area for the incoming stream segments. Then, the PI for each incoming segment is calculated by dividing its cumulative watershed area by the total incoming area. The PIs of the incoming segments always sum to one because they are proportions. When the program is complete, each segment in the streams dataset has a PI value stored in its attribute table.

The second step is to use the segment PIs to calculate the PI for each pair of flow-connected sites (Figure 2.3c). The PI for a pair of sites is equal to the product of the segment PIs found in the downstream path between them. The PI of a site to itself is equal to one and two sites that are not flow connected receive a PI value equal to zero. The site PI for each pair of sample sites in the dataset are output as an n by n PI matrix, which is also in comma delimited text format.

Developing a valid covariance matrix based on hydrologic distance and flow

The asymmetric hydrologic distance and PI matrices must be reformatted before they can be used to create a statistically valid covariance matrix. A matrix, W, is computed by taking the square root of the PI matrix. A symmetric spatial weights matrix is created by taking $A = W + W'$. Then, the asymmetric hydrologic distance matrix, D, is forced into symmetry by computing the symmetric hydrologic distance between all flow-connected sites. Unconnected sites retain a distance or spatial weight equal to zero, while pairs of flow-connected sites are assigned an identical distance or spatial weight in both directions. This may seem counter intuitive because the matrices are intended to represent asymmetric flow relationships. Nevertheless, there is a symmetric correlation between flow-connected sites. Even though a downstream site does not affect upstream sites, the conditions at the downstream site are, in part, a result of those found upstream. This concept also applies to autoregressive models in time series, where the model is constructed with later time events depending on earlier ones. Although time flows in one direction, the correlation between two time events is symmetric. If we wanted to predict forward in one unit of time from a single observed event the prediction would be the same as if we went back one unit of time (Barnett, 2004). However, there is an added twist to stream networks because the spatial neighborhood only includes flow-connected sites (Figure 2.1c). This unique characteristic makes a model based on asymmetric hydrologic distance dramatically different from models based on straight-line or symmetric hydrologic distance. Flow connectivity is preserved in the symmetric distance matrix, while the strength of the spatial autocorrelation between flow-connected sites is represented by the spatial weights.

To our knowledge, the exponential autocorrelation function is the *only* function that can be used to create a statistically valid covariance matrix based on symmetric hydrologic distance. However, the exponential, spherical, linear with sill, and Mariah autocorrelation functions (Ver Hoef et al., in press) can be fit to the asymmetric hydrologic distance data (which has been forced into symmetry) because the spatial weights ensure its validity. In order to generate a covariance matrix, V, from an autocorrelation function, $\rho(h)$, the distance (h) may be scaled by the range parameter (θ_2), multiplied by the sill (θ_1), and the nugget effect (θ_0) added when the distance is equal to zero (Equation 2.6). We recorded a distance equal to zero when two sites were not flow-connected, but a true distance of zero only occurs on the diagonal of the distance matrix when we measure the distance between a site and itself. We use the distance matrix D to compute a matrix V (Equation 2.6), where each element of V uses the hydrologic distance h (from D) between each pair of locations.

$$C_1(h \mid \theta) = \begin{cases} \theta_0 + \theta_1 & \text{if } h = 0 \\ \theta_1 \rho(h / \theta_2) & \text{if } h > 0 \end{cases} \tag{2.6}$$

The Hadamard (element-wise) product, $\Sigma = \mathrm{A} \odot \mathrm{V}$, is applied to the two matrices and the product is a covariance matrix that meets the statistical assumptions necessary for geostatistical modeling (Ver Hoef et al., in press). After these conversions, the covariance matrix can be used with the standard geostatistical models (Cressie, 1993; Isaaks and Srivastava, 1989).

Example

We provide an example using data from a hypothetical basin to illustrate the methods used to calculate a valid covariance matrix based on asymmetric hydrologic distance weighted by discharge volume. Figure 2.5 contains information about a small stream network with five sample sites and eleven stream segments, which will be referred to throughout this example. The segment lengths and watershed areas were calculated using a GIS and are included in Figure 2.5. We generated a five by five asymmetric hydrologic distance matrix and forced it into symmetry by assigning the one directional asymmetric distance to both upstream and downstream flow directions for sites s1-s5.

$$\begin{pmatrix} 0 & 730 & 1190 & 830 & 900 \\ 730 & 0 & 460 & 0 & 0 \\ 1190 & 460 & 0 & 0 & 0 \\ 830 & 0 & 0 & 0 & 0 \\ 900 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{2.7}$$

The segment PI was calculated for each of the 11 stream segments based on the watershed areas in the network (Figure 2.5). The PI for each pair of sites was computed and recorded in the PI matrix, which was also forced into symmetry.

$$\begin{pmatrix} 1 & 0.73 & 0.36 & 0.06 & 0.21 \\ 0.73 & 1 & 0.49 & 0 & 0 \\ 0.36 & 0.49 & 1 & 0 & 0 \\ 0.06 & 0 & 0 & 1 & 0 \\ 0.21 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.8}$$

We set the covariance parameters, $\theta_0 = 0.1,\ \theta_1 = 0.7,\ \theta_2 = 750,$ rather than using likelihood estimates to maintain the simplicity of the example. Matrix V was generated by using the exponential autocorrelation function for the data contained in the hydrologic distance matrix, D.

$$C(D;\theta_0,\theta_1,\theta_2) = \begin{cases} \theta_0 + \theta_1 & \text{if } h = 0, \\ \theta_1 \exp(-D/\theta_2) & \text{if } 0 < D \end{cases} \tag{2.9}$$

We created matrix A by taking the square root of the PI matrix (Equation 2.8) and then applied the Hadamard (element-wise) product to the two matrices, $\Sigma = \mathrm{A} \odot \mathrm{V}$, to obtain a valid covariance matrix (Equation 2.10).

$$\begin{pmatrix} 0.800 & 0.226 & 0.086 & 0.057 & 0.097 \\ 0.226 & 0.800 & 0.265 & 0 & 0 \\ 0.086 & 0.265 & 0.800 & 0 & 0 \\ 0.057 & 0 & 0 & 0.800 & 0 \\ 0.097 & 0 & 0 & 0 & 0.800 \end{pmatrix} \tag{2.10}$$

DISCUSSION

We developed the methodology described here to provide a cost-efficient means of implementing geostatistical models for stream networks in both small and large study areas. To maintain cost efficiency, it was important to automate data processing. Automation may not be necessary for modeling in a small basin with only a few stream segments, but the requirements for regional analyses, such as the Mid-Atlantic Region (Herlihy, Stoddard, and Johnson, 1998), can be quite large. Regional datasets can also be costly and time consuming to obtain and process.

The ability to efficiently calculate hydrologic distances and spatial weights provides the opportunity to create more ecologically valid distance measures for geostatistical modeling in stream networks. Until recently, straight-line distance has been the primary distance measure used in geostatistical models. However, the physical characteristics of streams, such as network configuration, connectivity, flow direction, and position within the network, demand more functional, process-based measures. Stream ecologists will be able to choose models that are more appropriate for testing ecological hypotheses. In addition, different patterns of spatial correlation may occur at coarse and fine scales, which could warrant modeling each pattern using a different distance measure.

Current models are based on straight-line or hydrologic distance, but it may be possible to create other more ecologically relevant distance measures that incorporate physical characteristics such as flow velocity, stream gradient, or structures, and better reflect the energy an organism expends to move from one location to another. Network connectivity could also include chemical, physical, and biological barriers, such as pH, waterfalls, and predators, to make the potential movement of organisms and material more realistic. Given the complexity of stream ecosystems, there is unlikely to be one measure of distance and connectivity that is most appropriate for all situations. Instead, providing a variety of functional measures will facilitate exploration so that stream ecologists can select or develop a measure appropriate for their data.

As new functional distance measures are developed, it is imperative to be aware of the statistical assumptions on which geostatistical models are based. We encourage stream ecologists to work with their statistician colleagues to conceive of and calculate functional distance measures. These interdisciplinary collaborations will help to ensure that geostatistical models for stream networks are both ecologically and statistically valid.

The tools and methodologies presented here provide an example of how to calculate the hydrologic distances and spatial weights needed for geostatistical modeling in stream networks. We feel that it is important to present this methodology so that valid geostatistical models are applied to stream networks. Given this objective, the tools described here will be made freely available to the public as an ESRI ArcGIS extension (http://www.stat.colostate.edu/%7Ensu/starmap/program.html).

LITERATURE CITED

Altunkaynak A., Ozger M., & Sen Z. (2003) Triple diagram model of level fluctuations in Lake Van, Turkey. *Hydrology and Earth System Sciences*, 7, 235-244.

Bailey T.C. & Gatrell A.C. (1995) *Interactive spatial data analysis.* Pearson Education Limited, Essex, England.

Bales R.C., McConnell J.R., Mosley-Thompson E., Lamorey G. (2001) Accumulation map for the Greenland Ice Sheet. *Geophysical Research Letters*, 28, 2967-2970.

Barnett V. (2004) *Environmental Statistics Methods and applications.* John Wiley and Sons, Chichester, West Sussex, England.

Barry R.P. and Ver Hoef J.M. (1996) Blackbox krigng: Spatial prediction without specifying the variogram. *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297-322.

Benda L., Poff N.L., Miller D., Dunne T., Reeves G., Pess G., & Pollock M. (2004) The network dynamics hypothesis: How channel networks structure riverine habitats. *BioScience*, 54, 413-427.

Chong G.W., Reich R.M., Kalkhan M.A., & Stohlgren T.J. (2001) New approaches for sampling and modeling native and exotic plant species richness. *Western North American Naturalist*, 61, 328-335.

Cressie N. (1993) *Statistics for spatial data, Revised Edition.* John Wiley and Sons, New York.

Cumming G.S. (2002) Habitat shape, species invasions, and reserve design: insights from simple models. *Conservation Ecology*, 6.

Dent C.L. & Grimm N.B. (1999) Spatial heterogeneity of stream water nutrient concentrations over successional time. *Ecology*, 80, 2283-2298.

Dobermann A. & Ping J.L. (2004) Geostatistical integration of yield monitor data and remote sensing improves yield maps. *Agronomy Journal*, 96, 285-297.

Dussault G. & Brochu M. (2003) Distance matrix calculation. http://arcscripts.esri.com. Viewed 17 March 2005. Quebec, Canada: Institut Nationale de la Recherche Scientifique, Urbanisation, Culture et Société (INRS-UCS), Université du Québec.

ESRI (Environmental Systems Research Institute, Inc) (2002) *ArcGIS version 8.3.* Redlands, CA.

ESRI (2001a) *Hydro Data Model.* http://support.esri.com/index.cfm?fa=downloads.dataModels.filteredGateway&dmid=15. Viewed 17 January 2005. Redlands, CA.

Fagan W.F. (2002) Connectivity, fragmentation, and extinction risk in dendritic metapopulations. *Ecology,* 83, 3243-3249.

Flores-Garnica J.G. & Omi P.N. (2003) Mapping forest fuels for spatial fire behavior simulations using geomatic strategies. *Agrociencia*, 37, 65-72.

Ganio L.M., Torgersen C.E., & Gresswell R.E. (2005) A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and the Environment*, 3(3), 138-144.

Gardner B., Sullivan P.J., & Lembo A.J. (2003) Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. *Canadian Journal of Aquatic Science*, 60, 344-351.

Hudak A.T., Lefsky M.A., Cohen W.B., & Berterretche, M. (2002) Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sensing of the Environment*, 82, 397-416.

Isaaks E.H. & Srivastava R.M. (1989) *An introduction to applied geostatistics*. Oxford University Press, Inc., New York, NY.

Josberger E.G. & Mognard N.M. (2002) A passive microwave snow depth algorithm with a proxy for snow metamorphism. *Hydrological Processes Special Issue*, 16, 1557-1568.

Jurado-Exposito M., Lopez-Granados F., Garcia-Torres L., Garcia-Ferrer A., de la Orden M.S., & Atenciano S. (2003) Multi-species weed spatial variability and site-specific management maps in cultivated sunflower. *Weed Science*, 51, 319-328.

Kellum B. (2002) *Analysis and modeling of acid neutralizing capacity in the Mid-Atlantic Highlands area.* MS thesis, Colorado State University, 69 pp.

Legleiter C.J., Lawrence R.L., Fonstad M.A., Marcus W.A., & Aspinall R. (2003) Fluvial response a decade after wildfire in the northern Yellowstone ecosystem: A spatially explicit analysis. *Geomorphology*, 54, 119-136.

Little L.S., Edwards D., & Porter D.E. (1997) Kriging in estuaries: As the crow flies, or as the fish swims? *Journal of Experimental Marine Biology and Ecology*, 213, 1-11.

Mixon D.M. (2002) *Automatic watershed location and characterization with GIS for an analysis of reservoir sediment patterns*. University of Colorado, 117 pp.

Neitsch S.L., Arnold J.G., Kiniry J.R., Srinivasan R., and & Williams J.R. (2002) *Soil and water assessment tool user's manual.* Texas Water Resources Institute, College Station, TX.

Olden J.D., Jackson D.A., & Peres-Neto P.R. (2001) Spatial isolation and fish communities in drainage lakes. *Oecologia*, 127, 572-585.

Olea R.A. (1991) *Geostatistical glossary and multilingual dictionary.* Oxford University Press, New York, NY.

Pleydell D.R.J., Raoul F., Tourneux F., Danson F.M., Graham A.J., Craig P.S., & Giraudoux P. (2004) Modeling the spatial distribution of *Echinococcus multilocularis* infection in foxes. *Acta Tropica*, 91, 253-265.

Pringle C.M. (2001) Hydrologic connectivity and the management of biological reserves: A global perspective. *Ecological Applications*, 11, 981-998.

Rathbun S.L. (1998) Spatial modeling in irregularly shaped regions: Kriging estuaries. *Environmetrics*, 9, 109-129.

Reich R.M., Lundquist J., & Bravo V.A. (2000) Spatial relationship of resident and migratory birds and canopy openings n diseased ponderosa pine forests. *Environmental Modeling and Software*, 15, 189-197.

Robichaud P.R. & Miller S.M. (2000) Spatial interpolation and simulation of post-burn duff thickness after prescribed fire. *International Journal of Wildland Fire*, 9, 137-143.

Theobald D.M. (2002) *RW Tools for ArcView Version 3.* Unpublished program. Natural Resource Ecology Laboratory, Fort Collins, CO.

Torgersen C.E., Gresswell R.E., & Bateman D.S. (2004) Pattern detection in stream networks: Quantifying spatial variability in fish distribution. *Proceedings of the Second Annual International Symposium on GIS/Spatial Analyses in Fishery and Aquatic Sciences.* (eds. Nishida, T., Kailola, P. J. , and Hollingworth, C. E.) pp. 405-420. Fishery GIS Research Group, Saitama, Japan.

USEPA (United States Environmental Protection Agency) (2002) *NHD Reach Indexing Tool user's guide.* USEPA Office of Water, Washington D.C.

USGS (United States Geological Survey). (2004a) *National Hydrography Dataset.* http://nhd.usgs.gov. Viewed 27 January 2005. U.S. Department of the Interior, Rolla, MO.

United States Geological Survey (2004b) *StreamStats.* http://streamstats.usgs.gov. Viewed 3 March 2005. U.S. Department of the Interior, Washington D.C.

Ver Hoef J.M., Peterson E.E., & Theobald D.M. Some new spatial statistical models for stream networks. *Environmental and Ecological Statistics.* In Press.

Vogel R.M., Wilson I., Daly C. (1999) Regional regression models of annual streamflow for the United States. *Journal of Irrigation and Drainage Engineering*, May/June, 148-157.

Ward J.V. (1989) The four-dimensional nature of lotic ecosystems. *Journal of the North American Benthological Society,* 8, 2-8.

Yuan L.L. (2004) Using spatial interpolation to estimate stressor levels in unsampled streams. *Environmental Monitoring and Assessment*, 94, 23-38.

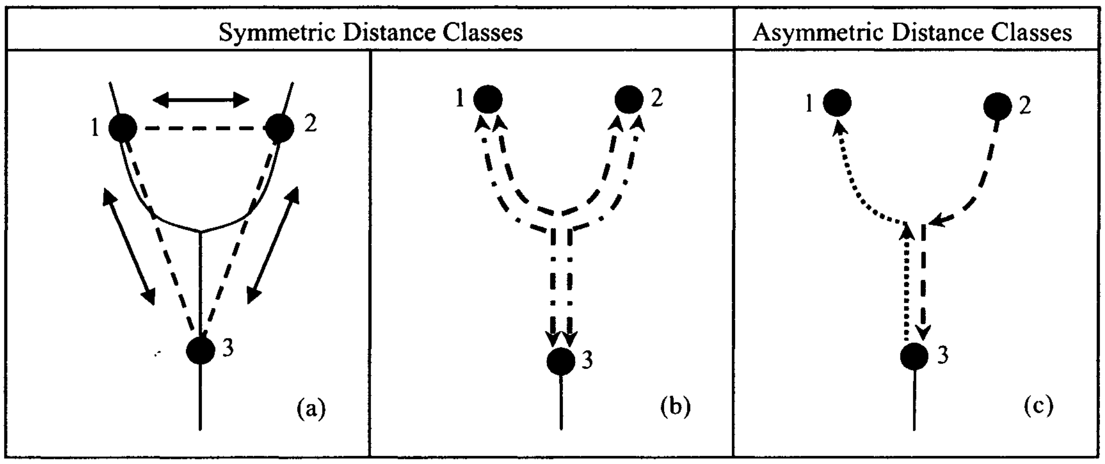


Figure 2.1. Symmetric and asymmetric distance classes. The stream network is represented by a solid line, while distance measurements are represented with dashed lines. Symmetric hydrologic distance measures include straight-line distance (a) and symmetric hydrologic distance (b). Sites 1, 2, and 3 are all neighbors to one another when these distance measures are used. Asymmetric distance classes include upstream and downstream asymmetric hydrologic distance (c). Sites 1 and 2 are neighbors to site 3, but not to each other.

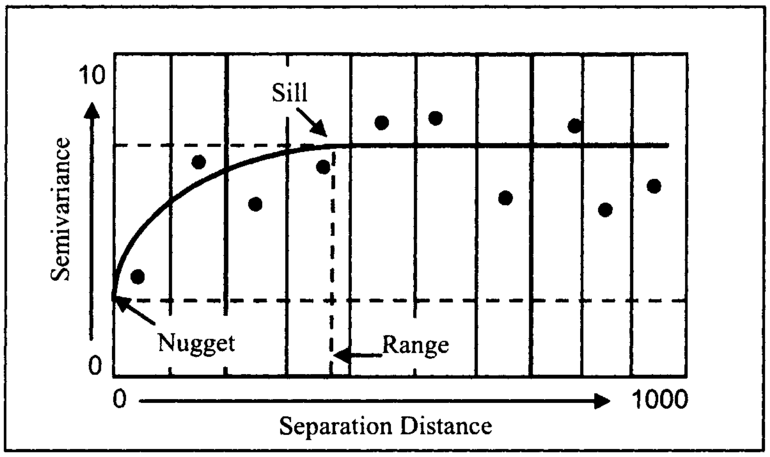
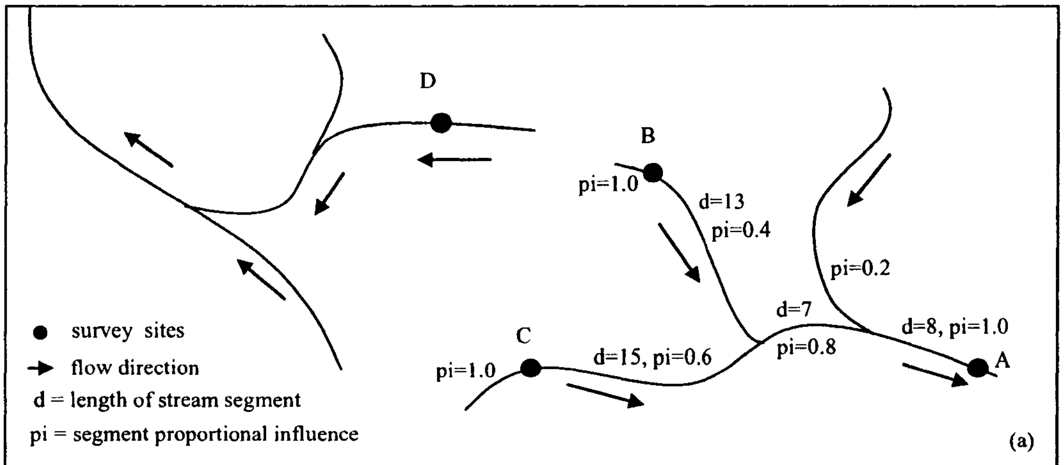


Figure 2.2. The semivariogram. Semivariograms are generated by grouping the separation distances into bins, calculating a mean semivariance for each bin, and plotting them in ascending order. Each mean semivariance is represented by a single point. The semivariogram is used to derive visual estimates of the nugget, range, and sill.



FROM Site

TO Site	A	B	C	D
A	0	28	30	0
B	0	0	15	0
C	0	13	0	0
D	0	0	0	0

Distance Matrix (b)

FROM Site

TO Site	A	B	C	D
A	1	0.32	0.48	0
B	0	1	0	0
C	0	0	1	0
D	0	0	0	1

PI Matrix (c)

Figure 2.3. Hydrologic distance and proportional influence (PI) matrices. The *n* by *n* hydrologic distance and PI matrices represent the spatial connectivity and neighborhood relationships in a stream network (a). The hydrologic distance matrix (b) only contains the downstream distance between sites (B→C=13), but contains sufficient information to calculate a variety of hydrologic distance measures. The upstream distance between the same two sites can be found by reversing the direction of the path (C→B=15). The total hydrologic distance, regardless of flow direction, is calculated by summing the two downstream distances between sites (B↔C=28). A pair of sites is connected by flow if the downstream distance between two sites is greater than zero in one direction and equal to zero in the other direction (C→A=30, A→C=0). Sites are not flow-connected if the downstream distance is equal to zero in both directions (D→A=0, A→D=0). The PI matrix (c) represents the proportion of water at a downstream site that comes from an upstream site. The PI for a pair of sites is equal to the product of the segment PIs found in the path between them (PI B→A = 0.4*0.8*1.0 = 0.32).

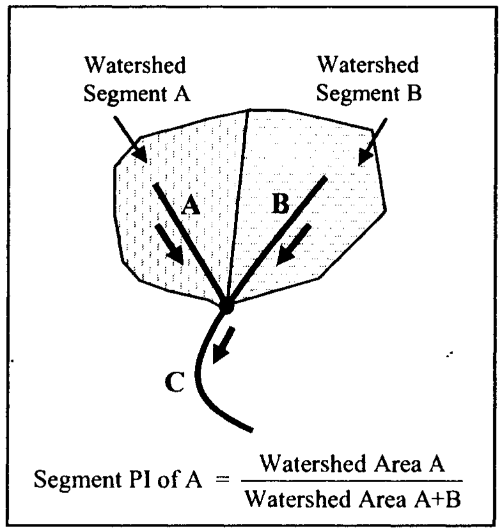
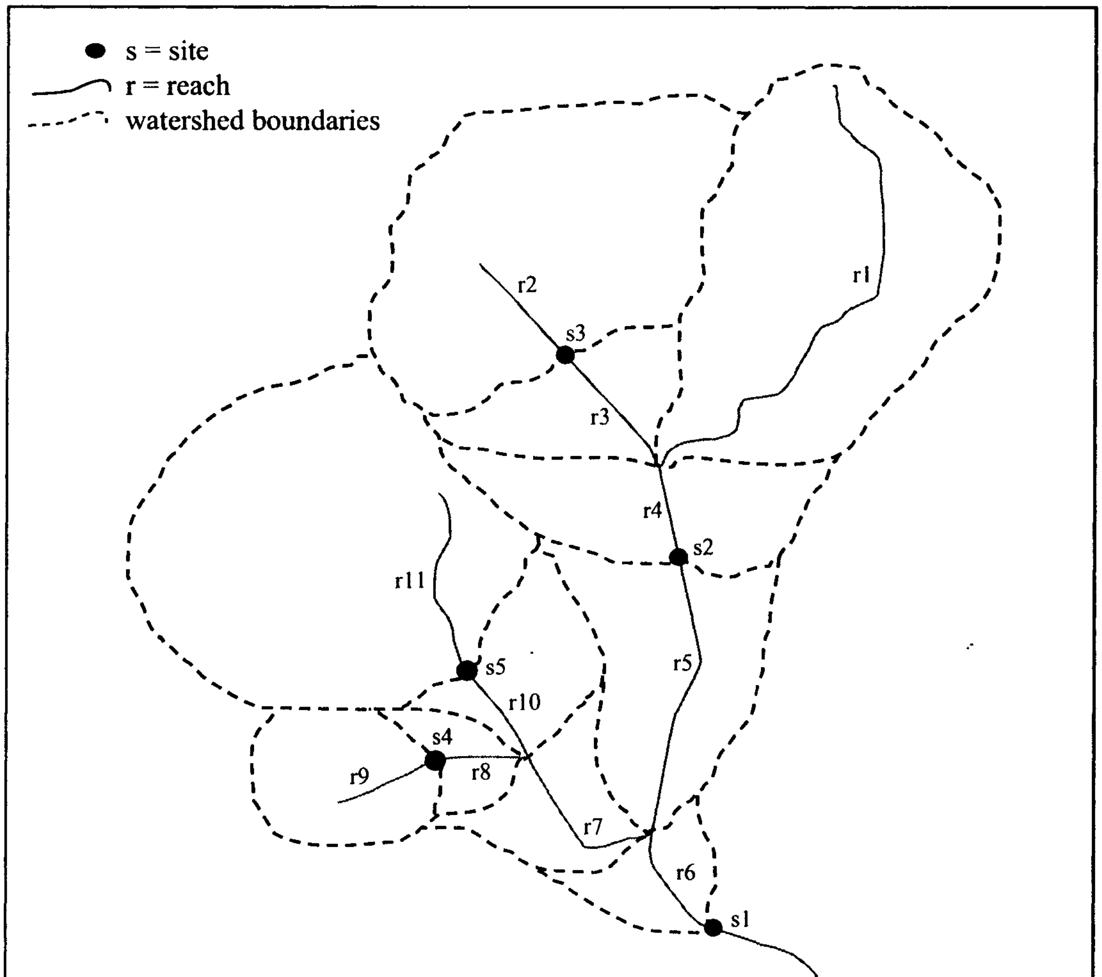


Figure 2.4. Calculating the proportional influence (PI). The segment PI represents a segment's PI on the segment directly downstream. It is calculated by dividing the segment's cumulative watershed area by the total incoming area at its downstream node.



Segment	r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11
Length (m)	1000	230	280	180	540	190	450	190	290	260	520
Watershed Area (km^2)	270	220	270	600	679	1000	291	48	28	192	160
Segment PI	0.51	1.00	0.49	1.00	0.73	1.00	0.27	0.22	1.00	0.78	1.00

Figure 2.5. Example data. The data derived from this small stream network with 5 sample sites (s1-s5) and 11 stream segments (r1-r11) was used to calculate a valid covariance matrix based on asymmetric hydrologic distance weighted by watershed area (Equation 2.10), which is used as a surrogate for discharge volume. The watershed boundaries are delineated for the downstream node of each stream segment. They are nested and so the cumulative watershed area includes the entire upstream drainage area. The length and true watershed area are recorded in the table above. The segment proportional influence (PI) values are computed using watershed area and are also recorded above.

CHAPTER THREE

PATTERNS OF SPATIAL AUTOCORRELATION IN STREAM WATER CHEMISTRY[2]

INTRODUCTION

Stream water chemistry is spatially and temporally heterogeneous at multiple scales (Pringle, 1991; Chambers et al., 1992; Dawson et al., 2001) and the conditions observed at survey sites result from the collective influence of multi-scale landscape filters (Frissell et al., 1986; Poff, 1997). A hierarchical constraint exists among filters (Frissell et al., 1986; Davies et al., 2000), but the strength of the linkage varies between filter scales. Processes acting across spatial and temporal filter scales produce spatial patterns in water chemistry (Poff, 1997). The quantification of these patterns provides information concerning the importance of ecosystem processes and spatial relationships occurring across multiple scales.

Geostatistical models are commonly used to quantify spatial patterns in the terrestrial environment, but have been applied less frequently to aquatic systems such as lakes (Altunkaynak et al., 2003), estuaries (Little et al., 1997; Rathbun, 1998), and streams (Kellum, 2003; Yuan, 2004). Geostatistical models are typically based on

[2] This chapter was submitted to Ecological Monitoring and Assessment and my co-authors were Andrew A. Merton, David M. Theobald, and N. Scott Urquhart.

symmetric straight-line distance, which is not an ecologically representative distance measure because it fails to represent the spatial configuration, connectivity, directionality, and relative position of sites in a stream network (Olden et al., 2001; Benda et al., 2004; Ganio et al., 2005). Recently, freshwater ecologists have begun to explore spatial patterns in stream networks using hydrologic distance measures (Dent and Grimm, 1999; Gardner et al., 2003; Legleiter et al., 2003; Torgersen et al., 2004; Ganio et al., 2005). In addition, new geostatistical methodologies have recently been developed that enable directional hydrologic distance measures to be considered (Ver Hoef et al., in press). This provides freshwater ecologists with a variety of distance measures to choose from, but it is not obvious which measure is most appropriate for water chemistry data.

Our ability to detect patterns of spatial correlation depends on the grain of the survey design, the extent of the study area, and the configuration of survey locations (Levin, 1992; Cooper et al., 1997). It is likely that the distance measure will also affect the spatial patterns that we observe. Different patterns are likely to occur within and between filter scales (Poff, 1997) and freshwater ecologists must choose the survey scale, design, and distance measure that is most appropriate for their research questions. Little information is available concerning the effect of the distance measure on observed patterns of spatial correlation in stream networks (see Gardner et al., 2003). Therefore, the purpose of this study is to explore and quantify patterns of spatial correlation in chemical response variables using three distance measures: straight-line distance (SLD), symmetric hydrologic distance (SHD), and weighted asymmetric hydrologic distance (WAHD).

BACKGROUND

Symmetric and asymmetric distances can be used to represent physical and ecological processes in stream ecosystems. These distances are used to characterize the spatial neighborhood for each site. A spatial neighborhood includes sites that are relatively close in space and have a measurable influence upon one another, i.e., are correlated with one another. Sites that do not fall within the spatial neighborhood are considered uncorrelated. Symmetric distance is directionless (isotropic) and has equal correlation in all directions (or both on a stream). SLD is symmetric and all locations in a study area have the potential to be neighbors (Figure 3.1a). *Hydrologic distance* can be either symmetric or asymmetric and is merely the distance between two locations when movement is limited to the stream network. SHD is the total upstream and downstream hydrologic distance between two sites (Figure 3.1b). When all basins in the study are share a common outlet all sites within a stream network are considered neighboring sites since flow direction is disregarded. Asymmetric distances include one-directional measures that are restricted to either the upstream or downstream flow direction. Water must flow from one location to another to be considered neighbors (Figure 3.1c). Spatial weights can also be generated using metrics that represent relative network position, such as watershed area or stream order, and used to create more ecologically representative WAHD (Ver Hoef et al., in press).

Terrestrial processes may be better represented with SLD because the terrestrial landscape can be represented as a two-dimensional plane where any pair of sites may be neighbors. For example, simplistic models designed to represent terrestrial transport mechanisms, such as seed dispersal, provide few restrictions to the direction of movement. In some cases it may also be appropriate to apply SLD to stream ecosystems. For example, a chemical response variable may be significantly influenced by a continuous landscape variable, such as elevation (Kellum, 2003) or by broad-scale factors, such as acid precipitation (Driscoll et al., 2001). However, SLD may not be as useful when instream processes dominate water chemistry conditions. Freshwater systems have four dimensions: longitudinal, lateral, vertical, and temporal (Ward, 1989), but we represent streams as one-dimensional features and focus on the longitudinal or upstream-downstream dimension. Freshwater riverine systems differ from terrestrial ecosystems because they are linear and typically the movement of material is restricted to the stream network. Some aquatic fauna (e.g. fish) move both up and downstream, but cannot move outside of the stream network. Water chemistry is generally characterized by passive movement and is primarily affected by longitudinal transport. In this case, movement is restricted to the network, but occurs primarily in the downstream direction.

The relative position in the stream network also has an effect on the condition of a site (Pringle, 2001; Benda et al., 2004) and reflects the influence that it will have on neighboring sites (Cumming, 2002). For example, substantial differences in discharge volume may cause a site located on a minor tributary to have little influence on a downstream site located on the mainstem (Benda et al., 2004). Clearly, both topographical and topological characteristics of the stream network provide information about chemical conditions at unobserved locations. Therefore, functional distances that represent hydrologic connectivity should also be considered for geostatistical modeling in stream networks.

Patterns of spatial correlation are visualized using a graphical representation called an empirical semivariogram, which is a plot of the semivariance between sites given their separation distance. The semivariance represents the strength of spatial correlation between two sites (Olea, 1991) and the separation distance is simply the distance traveled from one location to a second location. Semivariograms are generated by dividing the separation distances into groups, or bins, calculating the mean semivariance for each bin (Equation 3.1), and plotting the semivariances for the bins in ascending order (i.e. 100, 200, 300….). The semivariance is given by

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z_i - Z_{i+h} \right]^2, \tag{3.1}$$

where h is the mean separation distance between sites within a bin, $\gamma(h)$ is the semivariance for the bin, Z_i is the observed sample value at site i, Z_{i+h} is the observed sample value at $i+h$, and $N(h)$ is the total number of sample pairs for the bin.

The semivariogram is a powerful diagnostic tool for identifying whether there is correlation among sites. It is common to fit an autocorrelation function, such as the exponential function, to the semivariogram in order to estimate the three autocorrelation parameters (θ): the nugget, sill, and range (see Gardner et al., 2003; Kellum, 2003; Yuan, 2004). The nugget represents the variation in the data as the separation distance between sites approaches zero. It can result from experimental error or could indicate that a substantial amount of variation is present at a scale finer than the minimum separation distance. The sill is delineated where the autocorrelation function becomes asymptotic and represents the variance found between uncorrelated data. The range parameter controls how fast the autocorrelation decays with distance.

A drawback to using the empirical semivariogram for estimating the autocorrelation parameters is that the investigator must select bin size in order to generate a useful semivariogram. The fitted values of the autocorrelation parameters are, therefore, dependent on the bin size selected. Thus, parameter estimates can vary from investigator to investigator as a function of bin size.

Likelihood-based approaches are a robust method of obtaining estimates of the autocorrelation parameters (Kitanidis, 1983). The log likelihood function is used to estimate unknown parameter values based on the data and is derived from a specific probability distribution function, such as the exponential autocorrelation function (Hoeting et al., in press). The data values and the probability distribution function are fixed while the parameter values are allowed to vary. Maximizing the (log) likelihood with respect to the unknown parameters provides maximum likelihood (ML) estimates. ML estimation is an efficient method of parameter estimation for large sample sizes and provides a means for estimating uncertainty in the estimates (Pardo-Igúzquiza, 1998). In addition, competing models can be readily compared using, for example, Akaike's Information Corrected Criterion (AICC) which is itself likelihood based (Akaike, 1973).

Few studies have explored patterns of spatial correlation in stream chemistry data (but see Dent and Grimm, 1999; Gardner et al., 2003; Kellum, 2003; Yuan, 2004), yet findings suggest that the range varies with respect to the distance measure, spatial correlation differs between chemical response variables, and that patterns of spatial correlation change over time (Table 3.1). Gardner and others (2003) compared spatial correlation in temperature using three distance measures: SLD, SHD, and SHD weighted by Strahler stream order. They found that hydrologic distance measures led to semivariograms with a larger range than SLD and that weighting hydrologic distance further increased the range (Table 3.1). This was surprising since theoretically SLD would explain patterns of spatial correlation produced by broad-scale ecological processes that are not constrained to one watershed, such as the weathering of geological parent material. We presumed that hydrologic distance measures would better represent finer-grain processes related to flow connectivity. Dent and Grimm (1999) investigated the effect of flood events on spatial correlation in three chemical response variables using SHD and found that their range was affected by flood frequency. The range increased immediately following a flood event and then decreased over time, which created heterogeneous patterns of chemical concentration. The observed temporal pattern of correlation was similar for all response variables, but the range of their spatial correlation differed (Table 3.1).

We found only one study where a WAHD measure was used to explore spatial patterns of correlation in water chemistry. Cressie and others (in review) developed an asymmetric hydrologic distance measure weighted by stream order, which was based on the work of Ver Hoef and others (in press). They considered a mixture of autocorrelation functions based on SLD and WAHD and determined that spatial dependence between instream dissolved oxygen measurements was better represented by SLD.

More than one distance measure can be used to explain variability in stream chemistry data (Gardner et al., 2003), which is useful because we expect spatial correlation to differ between temporal and spatial filter scales (Levin, 1992; Poff, 1997).

The distance measure and scale must be appropriate for the ecological process being studied. However, applying geostatistical techniques to stream networks is a relatively new field of research and the limited findings to date do not clearly indicate which distance measure to use. To our knowledge, Cressie and others (in review) is the only study to quantitatively fit a variety of distance measures to one dataset to determine which best explains the variability in stream chemistry data. Gardner and others (2003) also compared predictions made using three distance measures, but used the spherical autocorrelation function to generate correlations based on hydrologic distances. The spherical autocorrelation function is invalid for pure hydrologic distance measures because the correlation matrices may contain negative eigenvalues, may produce negative variance estimates, and are not guaranteed to be positive definite (Ver Hoef et al., in press).

Our goal is to provide a detailed investigation into patterns of spatial correlation in eight chemical response variables collected throughout the state of Maryland, USA. We develop simple geostatistical models based on SLD, SHD, and WAHD. We generate predictions and make a statistical comparison to determine which distance measure best explains the variability in each chemical response variable.

METHODOLOGY

Data

The Maryland Biological Stream Survey (MBSS) data (Figure 3.2) were collected throughout Maryland by the Department of Natural Resources (DNR) (Mercurio et al., 1999). Maryland is a geographically diverse state that can be divided into three general provinces: the Coastal Plain, the Piedmont, and the Appalachian Plateau (Boward et al., 1999). The Coastal Plain borders Chesapeake Bay and produces low gradient streams with sandy gravel substrates. Elevation increases from east to west and Piedmont streams are characterized by steeper slopes and rock or bedrock substrates. The Appalachian Plateau is the westernmost region and is a diverse composition of valleys, sloping mountains, and steep ridges. Streams generally have rocky substrates, but range from low gradient meandering streams to steep cascading streams.

The Maryland DNR used a probability-based survey to collect chemical, physical, and biological data from first, second, and third order non-tidal streams in 17 interbasins throughout the state (Mercurio et al., 1999). An interbasin differs from a true watershed because it does not include the entire area that flows to an outlet point. A stratified random sample was collected from each interbasin based on Strahler stream order. The number of samples collected per stream order was proportional to the number of stream order miles within the interbasin. Ten chemical variables were collected: acid neutralizing capacity (ANC), in-situ conductivity, conductivity measured in the lab (CONDLAB), dissolved organic carbon (DOC), dissolved oxygen (DO), nitrate-nitrogen (NO_3), in-situ pH, pH measured in the lab (PHLAB), sulfate (SO_4), and temperature (TEMP). However, we did not include in-situ pH and conductivity in our analysis. In total, 955 sites were visited during 1995, 1996, and 1997.

The stream network and survey sites were pre-processed in a geographical information system (GIS) to ensure that sites were positioned on the correct stream segment. There are a variety of reasons why it is rare for GIS data collected within a stream to fall directly on a line segment representing a stream. Though the spatial accuracy of points collected using a global positioning system are becoming more precise, they still have some error. Some stretches of river can move (e.g. meander) slightly from their mapped position. Streams are often represented by center lines and so samples collected on the banks of a large river may not fall directly on a line segment. Digital streams datasets may contain mapping errors and generalizations, such as the absence of small tributaries and the generalization of form, which are found when streams are represented at coarser scales. As a result of these data problems, we could not identify the survey stream and it was necessary to discard 74 sites. In addition, there were a minimal number of missing data values for each chemical response variable. These sites were not completely eliminated from further analysis since more than one variable was collected at each site. Instead, they were temporarily removed from the analysis when a response variable contained no data.

Distance matrices were generated for SLD, SHD, and WAHD measures (Appendices I and II). We projected the data from latitude/longitude to Albers Equal Area projection (North American Datum 1983 based on the GRS1980 spheroid) before calculating the distance measurements. Projecting these data were necessary because the latitude/longitude coordinates have a known, systematic bias associated with increasing latitude. The SLD matrix was calculated in R statistical software package (Ihaka and Gentleman, 1996) using the easting and northing values as x, y coordinates. The SHD, asymmetric hydrologic distance, and spatial weights matrices were calculated in a GIS using programs written in Visual Basic for Applications for ArcGIS version 8.3 (ESRI, 2002) (Appendices I and II).

The spatial weights were used to develop the WAHD measure and represent the relative influence of one site on another. The weights were based on watershed area, which we use as a rough proxy for discharge volume. The spatial weights were generated by calculating the upstream watershed area for the downstream node of each segment in the stream network using a GIS. We defined a stream segment as the portion of a stream located between two confluences. When survey sites fell midway along a segment it was split into two separate stream segments. At each confluence or survey site in the network, the total upstream watershed area was calculated by summing the watershed area for the incoming stream segments. The proportional influence for each incoming segment was calculated by dividing its watershed area by the total upstream watershed area at the confluence or survey site. Every stream segment in the network contained its proportional influence on the segment directly downstream when this process was complete. We located the path between flow connected sites and calculated the influence of one site on another, which was equal to the product of the segment proportional influences found in the path.

The spatial weights matrix was simply an n by n matrix that contained the square root of the influence for all pairs of sites, which we used to maintain stationarity of the variances (Ver Hoef et al., in press). If two sites were not connected by flow the spatial weight was equal to zero and a site's influence on itself was equal to one. The GIS methods used to generate the hydrologic distance matrices and spatial weights matrix were lengthy, but are not the focal point of this manuscript (but see Peterson et al., in review).

Statistical analyses

The MBSS dataset contains lumped watershed attributes for each survey site (Mercurio et al., 1999), which the Maryland DNR derived using a GIS and the 1992 National Land Cover Data (NLCD) (MRLC Consortium, 2003). These were used as potential covariates (Table 3.2). In addition, we included the sample year, Level III Omernik's ecoregion (Omernik, 1987), mean elevation in the watershed, geology type (Appendix IV), and geographic location. The variance inflation factor (VIF) collinearity statistic (Helsel and Hirsch, 1992) indicated that five potential covariates were significantly correlated with other covariates (VIF > 10) so we removed them from further analysis. In addition, we made certain that the correlation coefficients for the remaining covariates were not greater than 0.8. We also created a validation set for each chemical response variable, which contained a unique set of 100 randomly selected sites (without replacement). These data were set aside in order to assess the accuracy of the final models.

We reduced the field of potential covariates due to the considerable number of response variables and distance measures and the processing time required for model selection. We used a Leaps and Bounds algorithm (Furnival and Wilson, 1974) to find the "best" set of five covariates for each chemical response variable and used them to develop a general linear model (GLM). We checked the model residuals for signs of non-normality and transformed ANC, CONDLAB, DOC, and NO_3 using a log10(x+n) transformation (Table 3.3). The studentized residuals were used to identify extreme outliers at a significance level less than 0.001, which we removed from further analysis.

We also calculated summary statistics for the response variables and the significant covariates.

We restricted the model space to all possible linear models using the five "best" explanatory variables determined by the initial covariate selection process described above. For a given distance measure there were 32 ($2^5 = 32$) competing models. Hence, four sets of 32 models were developed for each chemical response variable. The first set consisted of non-spatial models developed using GLM as a baseline to determine whether geostatistical models provided additional predictive ability. Each of the remaining sets assumed a geostatistical model using one of three distance measures: SLD, SHD, or WAHD. We also examined the model residuals to ensure that they were normally distributed with mean zero and variance-covariance matrix $\Sigma = \sigma^2\Omega$, where σ^2 is the variance and $\Omega = \Omega(d;\theta)$ is the correlation matrix. Note that Ω is a function of the distance between sites, *d*, given the autocorrelation parameter vector, θ. Therefore, the model for response variable Z is written in matrix notation as $Z = X\beta + \varepsilon$ where $\varepsilon \sim N(0,\sigma^2\Omega)$. Here X is the $n \times p$ design matrix of covariates, β is a vector of coefficients of length p, and ε is a vector of n (correlated) errors.

The log-likelihood function of the parameters $\left(\theta,\beta,\sigma^2\right)$ given the observed data, Z, is

$$\ell(\theta,\beta,\sigma^2;Z) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left|\sigma^2\Omega\right| - \frac{1}{2\sigma^2}(Z - X\beta)'\Omega^{-1}(Z - X\beta). \qquad (3.2)$$

Maximizing the log-likelihood (Equation 3.2) with respect to β and σ^2 yields

$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z$ and $\hat{\sigma}^2 = \dfrac{(Z - X\hat{\beta})'\Omega^{-1}(Z - X\hat{\beta})}{n}$. Both maximum likelihood estimators (MLE) can be written as functions of θ alone. Thus, we derive the *profile* log-likelihood function by substituting the MLEs back into (Equation 3.2):

$$\ell_{profile}(\theta; \hat{\beta}, \hat{\sigma}^2, Z) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{1}{2}\log|\Omega| - \frac{n}{2}\ . \tag{3.3}$$

The primary advantage to using the profile log-likelihood is that it reduces the dimensionality of the problem, which can reduce the amount of time required to find a numerical solution. This is especially important when there are a large number of models to compare.

The correlation matrix, Ω, is computed using the exponential autocorrelation function defined as

$$C_1(d; \theta_1, \theta_2) = \begin{cases} 1 & \text{if } d = 0 \\ (1 - \theta_1)\exp(-d/\theta_2) & \text{if } d > 0 \end{cases}, \tag{3.4}$$

where θ_1 is the *proportion* of nugget effect and θ_2 is the range parameter. The nugget is estimated by $\hat{\theta}_1\hat{\sigma}^2$ where θ_1 is restricted between zero and one. The approximate range over which sites are considered to be correlated is $3\theta_2$ (Cressie, 1993). The value d represents the distance between any two sites relative to the distance measure, e.g., SLD or SHD. The correlation matrix for the WAHD measure is generated by taking the

Hadamard (element-wise) product of (Equation 3.4) and the spatial weights matrix (Ver Hoef et al., in press). We restricted our analysis to the exponential autocorrelation function because it is currently the single known valid autocorrelation function for SHD (Ver Hoef et al., in press). We assumed that the errors were independent for the GLM model, i.e., Ω can be replaced with the identity matrix, I_n.

The MLE for θ is found by maximizing the profile log-likelihood (Equation 3.3) using a quasi-Newton method (Byrd et al., 1995), which is in turn used to compute the MLEs for the model parameters: β and σ^2 (Appendix III). To promote numerical stability, we standardized the response and explanatory variables to have mean zero and unit variance and scaled the distances to fall between zero and one.

We used the spatial AICC statistic (Hoeting et al., in press) to select the "best" GLM and the three "best" geostatistical models for each response variable: one for each of the three distance measures. The spatial AICC statistic is defined as

$$AICC = -2\ell_{profile}(\theta;\beta,\sigma^2,Z) + 2n\frac{p+k+1}{n-p-k-2}, \tag{3.5}$$

where n is the number of observations, $p-1$ is the number of covariates, and k is the number of autocorrelation parameters. To select the "best" GLM model we set $k=0$. The parameter k was set to two for the remaining three distance measures.

The model with the smallest AICC from each set was used to generate predictions using the Universal Kriging algorithm (Cressie, 1993). We used a split-sample approach to calculate the mean square prediction error (MSPE) for each model using validation sets that were set aside at the beginning of the analysis. The MSPE is defined as

$$MSPE = \frac{\sum_{i=1}^{n_p}(Z_i - \hat{Z}_i)^2}{n_p}, \tag{3.6}$$

where Z_i is the observed value at site *i*, $\hat{Z}_i$ is the predicted value at site *i*, and n_p is the total number of predictions. The MSPE was computed using a unique validation set for each response variable, but the four models within a response variable were tested using the same validation set. Models with small MSPE are desirable. The MSPE provided a way to compare models constructed using different distance measures and to determine which measure, if any, was more able to account for the variability in the response variable. In addition, we calculated the r^2 between the predictions and observations.

RESULTS

Summary statistics for the chemical response variable and significant covariate distributions are provided in Table 3.3. The spatial neighborhood produced by each distance measure differs, which affects the number of neighboring sites, as well as, the median, mean, and maximum separation distance between sites. Table 3.4 provides an example using the sites included in the DO model. Asymmetric hydrologic distance had considerably fewer pairs of neighboring sites compared to SLD and SHD. The minimum separation distance between neighboring sites was similar for all distance measures, but the asymmetric hydrologic distance measure had a shorter median, mean, and maximum value than the other distance measures. SHD consistently had the largest median, mean, and maximum separation distance.

We used five covariates in the model selection process (Table 3.5) and our results show that the models with the lowest spatial AICC value tended to be complex, meaning that they included a large number of covariates (Table 3.6). Covariates were systematically added during model selection so that model 1 represents the null model (no covariates) and model 32 the full model (all five covariates). The full model was selected for every GLM and for 17 of the 24 geostatistical models (Table 3.6).

The models for DO, PHLAB, and SO_4 displayed differences in complexity. Model 31 was selected for the DO and PHLAB models based on SLD and SHD. The DO model included all of the significant covariates except DECIDFOR, while the PHLAB model excluded PROBCROP (Table 3.5). This was not surprising because exploratory data analysis indicated that DO and PHLAB were weakly correlated with DECIDFOR and PROBCROP ($r^2 = 5\%$ and $r^2 = 1\%$, respectively). However, EASTING explained 44% of the variability in PROBCROP and 63% of the variability in DECIDFOR. The simplification of the model suggests that the correlations produced using SLD and SHD successfully represented the relatively small amount of variability in the response that was also explained by DECIDFOR and PROBCROP. In contrast, the full model was selected for the WAHD. Although models 31 and 32 are similar, the difference in model complexity indicates that the correlations based on WAHD contained less information about spatial correlation in the response variable compared to those produced using the SLD and SHD measures. SO_4 was the only chemical response variable that consistently selected a simpler model for all distance measures (Table 3.6). The models based on SHD and WAHD models omitted the NORTHING variable and the SLD model also omitted ER67 (Table 3.5). We believe that NORTHING and ER67 also represent broad-scale trends on the landscape related to spatial location.

The nugget estimates produced using ML differed between response variables. ANC, CONDLAB, NO_3, and PHLAB models had small nugget estimates, which included less than 7.5% of the variability in the data (Table 3.6). The nugget estimates for the DOC, SO_4, and TEMP models were larger and represented between 12.5% and 42% of the variability. Nugget estimates for DO were noticeably larger for all three distance measures compared to other chemical response variables and included between 39% and 70.4% of the variability in the data.

The ML estimates for the range parameter varied greatly with respect to distance measure and response variable (Table 3.6). We found that SLD produced the shortest range for every chemical response variable. The mean range for SLD, SHD, and WAHD were 28.2 km, 88.03 km, and 57.8 km, respectively. TEMP consistently produced the shortest range values and DO generated the largest range values for every distance measure. It should be noted that the ML range parameter estimates for the DOC, DO, and SO_4 models based on WAHD fell at the (user-defined) upper bounds of the optimization method.

It is inappropriate to use the r^2 for model selection because it continues to increase as the number of covariates in the model increase. However, I believe it is useful because it is familiar to most researchers and it provides a general estimate about the predictive ability of the model. The geostatistical models consistently explained more variability in the chemical response variables than the GLM (Table 3.6). The most dramatic differences were found in ANC, CONDLAB, and SO_4 where the r^2 value increased by 49%, 24%, and 21%, respectively. The geostatistical models for ANC, CONDLAB, DOC, NO_3, and PHLAB models generated predictions that had a strong correlation to the observed values (r^2= 0.63 to 0.95). However, the correlation between the geostatistical model predictions generated by the DO, SO_4, and TEMP models and the observations was weaker (r^2 = 0.28 to 0.41).

A comparison of the MSPE values suggests that more than one distance measure performed well for most chemical response variables (Table 3.6). The MSPE values for the DOC, TEMP, and PHLAB models did not exhibit clear differences relative to distance measure. SLD and SHD models developed using NO_3, ANC, and DO had similar MSPE values and produced more accurate predictions than the equivalent models using WAHD. In contrast, the SHD and WAHD models provided more accurate CONDLAB predictions than the SLD model. SO_4 was the only chemical response variable that displayed obvious differences between the models using different distance measures. The SLD model had the smallest MSPE value and produced more accurate predictions compared to the other distance measures.

DISCUSSION

The summary statistics for the distance measures (Table 3.4) demonstrate how a distance measure can significantly influence the way that spatial relationships are represented in a stream network. Not only does it affect the distance between neighbors and their relative influence, but it also dictates the form and size of the spatial neighborhood. The spatial neighborhood is important because it has a profound impact on the accuracy of the predictions produced by the geostatistical model. Furthermore, decisions concerning which distance measure to use are not just statistical choices, but should also be founded on the specific characteristics of the ecological process of interest and the research questions. For example, patterns of spatial autocorrelation in stream chemistry at the regional scale may be better described using straight-line distance because they are influenced by coarse-scale landscape variables, such as geology type, that are not constrained to watershed boundaries. However, regional conditions may not differ throughout a study area if the spatial extent is reduced to a small stream network. In this case, differences in stream chemistry may be dominated by instream processes and observable patterns of spatial autocorrelation may be better described using a hydrologic distance measure.

It was difficult to locate and obtain a dataset that was suitable for a regional analysis of spatial correlation in stream networks. We examined over 35 Environmental Protection Agency (EPA) datasets and most had large minimum separation distances between sites or few neighboring sites when a WAHD measure was used (e.g. Colorado Regional Environmental Mapping and Assessment Program dataset (USEPA, 1993)). The EPA uses a probability-based random survey design based on stream order to estimate regional stream conditions (Herlihy et al., 2000). This method is useful because it provides a statistical inference about the entire population of streams, within stream order, over a large area. However, it was designed to maximize spatial independence of survey sites, and consequently does not adequately represent spatial relationships in stream networks based on hydrologic distance measures.

The MBSS data were also collected using a probability-based design (Mercurio et al., 1999), but the sampling density was greater than other water chemistry datasets. Despite our large sample size (n=881), 244 sites did not have neighbors (Figure 3.3). The validation sites were randomly selected and it is likely that a considerable number did not have neighbors using the WAHD.

The models based on WAHD consistently produced more accurate predictions than the GLM models (Table 3.6). However, when a spatial neighborhood was deficient or absent for a specific site, the WAHD model performed in a manner similar to the GLM (Figure 3.4). This is a common feature for geostatistical models. The associated standard error for prediction sites with many observed neighbors is small compared to sites that have few (or no) neighbors. Thus, the WAHD model had the ability to explain the broad-scale mean in the data, but did not provide additional predictive ability at that site. The WAHD model generally explained more variability in the data as the number of neighboring sites increased (Figure 3.4). However, notable exceptions occurred when a site had neighbors with significantly different water chemistry values. Although the WAHD models were comparable to the SLD and SHD models, we believe that their performance may have been hindered by the survey design used to collect the raw data and the consequent lack of neighboring sites.

The GLM predictions also improved as the number of neighbors increased (Figure 3.4) because clusters of sites in space tend to have similar covariate values, i.e., they are positively correlated. Even though spatial location was not included in the model, the statistical regression was pulled towards the cluster of similar values. Thus, the GLM contained hidden spatial information, which allowed the non-spatial model to explain additional variability in the data when a greater number of neighbors were present.

Our results provide evidence that patterns of spatial correlation exist in stream water chemistry. The predictive ability of every geostatistical model was considerably greater than the GLM (Table 3.6) and in most cases more than one distance measure could be used to quantify the additional variability in stream chemistry data. SLD clearly does not represent the flow connectivity between sites in a stream network. SHD is similar in this respect because sites need not be connected by flow to be neighbors. The SHD separation distances and range values are consistently larger than those produced using SLD (Table 3.4). This is intuitive since the distance traveled between two sites increases when movement is restricted to the sinuous network. If the models based on SLD had performed poorly, we could assume that water chemistry was dominated by instream processes at a regional scale. However, the SLD models were never substantially inferior, which leads us to believe that the SLD, SHD, and WAHD measures are representing patterns of spatial correlation in continuous coarse-scale variables, such as geology type, that influence stream chemistry rather than the movement of chemicals through a stream network.

The chemical response variables that produced models with the greatest amount of predictive ability tend to be strongly related to coarse-scale landscape variables. ANC, PHLAB, NO_3, and CONDLAB are all significantly affected by landscape variables that are not restricted to watershed boundaries, such as geology type (Kellum, 2003), agricultural and urban areas (Herlihy et al., 1990; Gray, 2004), and the atmospheric deposition of nitric and sulfuric acids (Angelier, 2003). The small nuggets estimated by the geostatistical models indicate that the survey scale was fine enough to capture the majority of variability in the data.

The range values differ between ANC, NO_3, PHLAB, and CONDLAB and we believe that they represent the coarse scale of heterogeneity in the ecological processes that control them. Covariance parameters, such as the range value, cannot be used to identify which ecosystem processes are producing patterns of spatial correlation. Therefore, what follows is simply knowledgeable speculation about some, but not all, potential sources of spatial correlation in the data.

ANC and NO_3 models contained coarse-scale agricultural and urban covariates and produced similar range values. Thus, runoff or leaching from agricultural and urban areas may strongly influence ANC and NO_3 concentrations in the stream. The PHLAB model contained an agricultural covariate, but the range estimates were slightly shorter. PHLAB was also correlated to forest and wetland watershed area, which would be expected to exhibit more heterogeneous patterns on the landscape. CONDLAB produced the shortest range values of the models with strong predictive ability and was affected by urban land uses. However, percent coal mine area in the watershed was another influential covariate, which could actually be considered a point source of pollution. Coal mines are generally located in the western portion of Maryland, but would not produce a continuous pattern on the landscape. For example, the presence of one coal mine does not indicate that another coal mine will be located in a neighboring watershed. The relatively unpredictable distribution of coal mines may explain the short range values for CONDLAB. It is also not surprising that the models using SHD and WAHD performed slightly better than the SLD model (Table 3.6) since the effects of coal mines are found downstream rather than in the adjacent watershed. Although the magnitude of the range values differ for ANC, NO_3, PHLAB and CONDLAB, they are all strongly influenced by the coarse-scale condition on the landscape, which reduces the effect of in stream processes at this survey scale.

The predictive ability of our models suggests that DOC is influenced by ecological processes that produce coarse-scale patterns of spatial correlation on the landscape. This is not surprising since the majority of stream DOC comes from allochthonous sources of organic matter such as dead terrestrial plant material, soil, groundwater inputs, and wetlands (Wetzel, 1992). In addition, DOC is transported from the watershed to the stream via overland, sub-surface, or base flow and the flow path of water affects the DOC concentration of stream water (Qualls and Haines, 1992; Mulholland, 2003). The ML range estimate for the WAHD model was set to the upper limit in the optimization, but we feel that this is reasonable because the range values for SLD and SHD were also quite large. Interestingly, models generated to predict DOC were the only models with considerable predictive ability and relatively large nugget estimates. The large nugget estimate may result from our failure to represent the terrestrial water flow in the model, point sources of organic pollution, or fine scale instream processes that significantly affect stream DOC.

SO_4 is also influenced by coarse-scale processes, such as atmospheric inputs of sulfur and the weathering of sulfate minerals, and finer-scale processes related to the mineralization of organic sulfur and the adsorption and desorption of sulfate (e.g. Alewell et al., 1999). However, our results were inconclusive and were similar to other studies that have failed to establish strong correlations between SO_4 and watershed landcover categories (Herlihy et al., 1998).

The SO_4 model using SLD had the smallest MSPE but still had little predictive capability. All models produced large nugget values and the range value for SO_4 based on WAHD was set at the maximum separation distance. These results could indicate that finer-scale processes dominate patterns of spatial correlation in SO_4 and that coarse-scale processes are inconsequential. However, the ecological literature does not support that conclusion since the effects of atmospheric deposition of sulfuric acid on stream water chemistry have been well documented (Driscoll et al., 2001). A more plausible explanation is that the model fails to represent a SO_4 input that does not produce a consistent spatial pattern on the landscape at the survey scale. Also, we may have failed to include an important covariate during model selection or information may have been lost by using coarse-scale lumped watershed covariates. Nevertheless, the unexplained variability in SO_4 would significantly impair the predictive ability of the model and would essentially make coarse and fine scale patterns of spatial correlation indiscernible. The magnitude of the unexplained variability in the data suggests that more work is required to identify a better model.

The geostatistical models produced for TEMP and DO also had relatively little predictive ability. We concede that we are unable to conclusively determine the source of the residual error in these models, but we speculate that TEMP and DO are spatially and temporally variable over short distances (Hynes, 1960; Biggs et al., 1990). They are somewhat influenced by coarse-scale factors, such as air temperature or nutrient inputs (Smith, 1981), but are most likely dominated by instream processes related to the biological oxygen demand or water depth. In addition, diurnal fluctuations in TEMP and DO resulting from changes in climate and the biological oxygen demand in the water may have contributed to the variability in the data. TEMP consistently produced the smallest range values and the nugget estimates were large, which provided further evidence that patterns of spatial correlation were occurring at a fine scale. All three distance measures performed equally well for TEMP, which again lead us to believe that the WAHD measure represented coarse-scale patterns of spatial correlation across the landscape. Although the geostatistical models for DO showed a slight improvement over the GLM model in the accuracy of predictions, the model fit was extremely poor. The nugget estimates for DO were consistently large, which indicated that there was little to no spatial correlation between sites. The range values were also unrealistically large (Table 3.6) given the ecological behavior of DO, which provided further evidence that the form of the selected model was inappropriate. Given the large nugget estimates and the poor predictive ability of the models, we believe that the dominant processes controlling TEMP and DO occur at a scale finer than our minimum separation distance.

It is difficult to compare the results of our study to those of other studies because there are regional differences in the ecological processes that affect water chemistry (Hill et al., 2000). However, Yuan (2004) also used the MBSS dataset and found strikingly different range estimates for NO_3 and SO_4 based on SLD (Tables 3.1 and 3.6). We suspect that the differences result from the methods that were used to fit the autocorrelation functions. Yuan (2004) used a weighted least squares method to fit the autocorrelation function to the mean semivariance at each separation distance. Maximizing the likelihood function prevents inconsistencies in the autocorrelation parameters related to (empirical) semivariogram bin size. The methodology proposed by Hoeting and others (in press) allows for simultaneous fitting of the model and the error process. In addition, it can be used to compare different models using the spatial AICC.

CONCLUSIONS

The results of our study clearly demonstrate that spatial correlation exists in stream chemistry data at a relatively coarse scale and that geostatistical models improve the accuracy of predictions. The ranges and patterns of spatial correlation differ between chemical response variables, which are influenced by ecological processes acting at different spatial and temporal filter scales. We believe that the coarse-scale patterns we identified reflect the effects of regional scale ecological processes that are better described using SLD. Yet, spatial patterns are likely to change with the grain of the survey scale and the configuration of survey sites. Therefore, we inevitably impose bias related to the minimum separation distance and spatial neighborhood (Levin, 1992). For example, coarse-scale patterns may not be easily discernable if the extent of our survey area were smaller. In addition, a dataset with shorter separation distances and a more dense spatial arrangement would likely reveal spatial patterns related to finer-scale processes. Our results provide information about the ability of SLD, SHD, and WAHD measures to account for additional variability in stream water chemistry at a regional scale. Further research is needed to assess the capacity of these distance measures to explain additional variability at finer scales, such as at the watershed or reach scale.

Agencies have invested substantial resources collecting datasets using probability-based random survey designs and our study demonstrates that these data can be used to predict water quality conditions throughout large areas. At present, SLD appears to be the most suitable distance measure for regional geostatistical modeling of CONDLAB, ANC, NO_3, SO^4, PHLAB, DO, DOC, and TEMP in Maryland when the data are collected using a probability-based random survey design. SLD matrices are simple to calculate in statistical or GIS software, while the pre-processing time for hydrologic distance measures is greater. To fully explore the possible advantages of hydrologic distances, easy-to-use tools need to be widely available. At present, a regional geostatistical model based on SLD could be used by managers to predict water quality conditions for every stream segment within a large area. This methodology could potentially help states and tribes identify water quality impaired stream segments, which would allow agencies to focus additional field sampling efforts on potentially impaired sites.

It is likely that water chemistry contains both coarse and fine scales of spatial correlation and they may be better quantified using different distance measures. It is possible to describe multiple spatial patterns using different distance measures and to incorporate them into one geostatistical model (Cressie et al., in review). However, it is doubtful that datasets collected using a probability-based random survey design will be suitable for these types of models. Multi-scale geostatistical models may require new survey designs that more fully capture spatial correlation at a variety of scales using multiple distance measures. Survey designs that explain variability at multiple scales have been developed for terrestrial ecosystems (Shmida, 1984; Stohlgren et al., 1995) and it should also be possible to develop them for stream networks.

LITERATURE CITED

Akaike H. (1973) Information theory and an extension of the maximum likelihood principle. In: *Second International Symposium on Information Theory*, Budapest, Hungary, 267-281.

Alewell C., Mitchell M.J., Likens G.E., & Krouse H.R. (1999) Sources of stream sulfate at the Hubbard Brook Experimental Forest: Long-term analyses using stable isotopes. *Biogeochemistry*, 44, 281-299.

Altunkaynak A., Ozger M., & Sen Z. (2003) Triple diagram model of level fluctuations in Lake Van, Turkey, *Hydrology and Earth System Sciences*, 7, 235-244.

Angelier E. (2003) *Ecology of streams and rivers*. Science Publishers, Inc, Enfield, NH, USA.

Benda L., Poff N.L., Miller D., Dunne T., Reeves G., Pess G., & Pollock M. (2004) The network dynamics hypothesis: How channel networks structure riverine habitats, *BioScience,* 54, 413-427.

Biggs B.J., Jowett I.J., Quinn J.M., Hickey C. W., Davies-Colley R. J., & Close M. (1990) Ecological characterisation, classification, and modeling of New Zealand rivers: An introduction and synthesis. *New Zealand Journal of Marine and Freshwater Research,* 24, 277-304.

Boward D., Kayzak P., Stranko S., Hurd M., & Prochaska, A. (1999) *From the mountains to the sea: The state of Maryland's freshwater streams*. EPA/903/R-99/023, Maryland Department of Natural Resources, Monitoring and Non-tidal Assessment Division, Annapolis, MD, USA.

Byrd R.H., Lu P., Nocedal J.l., Zhu C., & Siam J. (1995) A limited memory algorithm for bound constrained optimization. *Scientific Computing,* 16, 1190-1208.

Chambers P.A., Prepas E.E., & Gibson K. (1992) Temporal and spatial dynamics in riverbed chemistry - The influence of flow and sediment composition. *Canadian Journal of Fisheries and Aquatic Sciences,* 49, 2128-2140.

Cooper S.D., Barmuta L., Sarnelle O., Kratz K., & Diehl S. (1997) Quantifying spatial heterogeneity in streams. *Journal of the North American Benthological Society,* 16, 174-188.

Cressie N. (1993) *Statistics for Spatial Data Revised Edition*. John Wiley and Sons, New York, NY, USA.

Cressie N., Frey J., Harch B., & Smith M. (in review) Spatial prediction on a river network. Department of Statistics, Preprint No. 747, The Ohio State University. http://www.stat.ohio-state.edu/~sses/papers.html.

Cumming G.S. (2002) Habitat shape, species invasions, and reserve design: Insights from simple models. *Conservation Ecology,* 6, 3. [online] URL: http://www.consecol.org/vol6/iss1/art3/

Davies N.M., Norris R.H., & Thoms M.C. (2000) Prediction and assessment of local stream habitat features using large-scale catchment characteristics. *Freshwater Biology,* 45, 343-369.

Dawson J.J.C., Billett M.F., & Hope D. (2001) Diurnal variations in the carbon chemistry of two acidic peatland streams in north-east Scotland. *Freshwater Biology*, 46, 1309-1322.

Dent C.L. & Grimm N.B. (1999) Spatial heterogeneity of stream water nutrient concentrations over successional time. *Ecology*, 80, 2283-2298.

Driscoll C.T., Lawrence G.B., Bulger A.J., Butler T.J., Cronan C.S., Eagar C., Lambert K.F., Likens G.E., Stoddard J.L., & Weathers K.C. (2001) *Acid rain revisited: Advances in scientific understanding since the passage of the 1970 and 1990 Clean Air Act Amendments.* Science Links™ Publication, Hubbard Brook Research Foundation.

Environmental Systems Research Institute (ESRI) (2002) *ArcGIS version 8.3.* Redlands, CA.

Frissell C.A., Liss W.J., Warren C.E., & Hurley M.D. (1986) A hierarchical framework for stream habitat classification: Viewing streams in a watershed context. *Environmental Management,* 10, 199-214.

Furnival G. & Wilson, R. (1974) Regression by leaps and bounds. *Technometrics*, 16, 499-511.

Ganio L.M., Torgersen C.E., & Gresswell R.E. (2005) A geostatistical approach for describing spatial pattern in stream networks. *Frontiers in Ecology and the Environment*, 3, 138-144.

Gardner B., Sullivan P.J., & Lembo A.J. (2003) Predicting stream temperatures: Geostatistical model comparison using alternative distance metrics. *Canadian Journal of Aquatic Science*, 60, 344-351.

Gray L. (2004) Changes in water quality and macroinvertebrate communities resulting from urban stormflows in the Provo River, Utah, USA. *Hydrobiologia*, 518, 33-46.

Helsel D.R. & Hirsch R.M. (1992) *Statistical methods in water resources*. Elsevier Science Publishing Co., New York, NY, USA.

Herlihy A.T., Kaufmann P.R., & Mitch M.E. (1990) Regional estimates of acid mine drainage impact on streams in the Mid-Atlantic and southeastern United States. *Water, Air, and Soil Pollution*, 50, 91-107.

Herlihy A.T., Stoddard J.L., & Johnson C.B. (1998) The relationship between stream chemistry and watershed land cover data in the Mid-Atlantic region, USA. *Water, Air, and Soil Pollution,* 105, 377-386.

Herlihy A.T., Larsen D.P., Paulsen S.G., Urquhart N.S., & Rosenbaum B.J. (2000) Designing a spatially balanced, randomized site selection process for regional stream surveys: The EMAP Mid-Atlantic Pilot Study. *Environmental Monitoring and Assessment,* 63, 95-113.

Hill B.H., Hall R.K., Husby P., Herlihy A.T., & Dunne M. (2000) Interregional comparisons of sediment microbial respiration in streams. *Freshwater Biology*, 44, 213-222.

Hoeting J.A., Davis R.A., Merton A.A., & Thompson S.E. (in press) Model selection for geostatistical models. *Ecological Applications*.

Hynes H. B. (1960) *The biology of polluted waters*. Liverpool University Press, Liverpool, England.

Ihaka R. & Gentleman R. (1996) R: a language for data analysis and graphics. *Journal of the Computational and Graphical Statistics,* 5, 299-314.

Kellum B. (2003) *Analysis and modeling of acid neutralizing capacity in the Mid-Atlantic Highlands area*. MS Thesis, Colorado State University, Fort Collins, CO.

Kitanidis P.K. (1983) Statistical estimation of polynomial generalized covariance functions and hydrologic applications. *Water Resources Research,* 19, 909-921.

Legleiter C.J., Lawrence R.L., Fonstad M.A., Marcus W.A., & Aspinall R. (2003) Fluvial response a decade after wildfire in the northern Yellowstone ecosystem: A spatially explicit analysis. *Geomorphology*, 54, 119-136.

Levin S.A. (1992) The problem of pattern and scale in ecology, *Ecology*, 73, 1943-1967.

Little L.S., Edwards D., & Porter D.E. (1997) Kriging in estuaries: As the crow flies, or as the fish swims? *Journal of Experimental Marine Biology and Ecology,* 213, 1-11.

Mercurio G., Chaillou J.C., & Roth N.E. (1999) *Guide to using 1995-1997 Maryland Biological Stream Survey Data.* Maryland Department of Natural Resources, Maryland Department of Natural Resources, Annapolis, MD, USA.

Mulholland P.J. (2003) Large-scale patterns in dissolved organic carbon concentration, flux, and sources. *In: Aquatic Ecosystems: Interactivity of Dissolved Organic Matter*, Academic Press, San Diego, CA, 139-159.

Multi-Resolution Land Characteristics Consortium (MRLC) (2003) U.S. Federal Region III land cover data set. Web page, Accessed June 17, 2005, http://www.epa.gov/mrlc/.

Olden J.D., Jackson D.A., & Peres-Neto P.R. (2001) Spatial isolation and fish communities in drainage lakes. *Oecologia*, 127, 572-585.

Olea R.A. (1991) *Geostatistical glossary and multilingual dictionary.* Oxford University Press, New York, NY, USA.

Omernik J. M. (1987) Ecoregions of the conterminous United States. *Annals of the Association of American Geographers*, 77, 118-125.

Pardo-Iǵuzquiza E. (1998) Maximum likelihood estimation of spatial covariance parameters. *Mathematical Geology*, 40, 95-108.

Peterson E.E., Theobald D.M., & Ver Hoef J.M. (in review) 'Support for geostatistical modeling on stream networks: Developing valid covariance matrices based on hydrologic distance and stream flow. *Freshwater Biology.*

Poff N.L. (1997) Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society,* 16, 391-409.

Pringle C.M. (1991) Geothermally modified waters surface at La Selva Biological Station, Costa Rica: Volcanic processes introduce chemical discontinuities into lowland tropical streams. *Biotropica,* 40, 523-529.

Pringle C.M. (2001) Hydrologic connectivity and the management of biological reserves: A global perspective. *Ecological Applications*, 11, 981-998.

Qualls R.G. & Haines B.L. (1992) Biodegradability of dissolved organic matter in forest throughfall, soil solution, and stream water. *Soil Science Society of American Journal,* 56, 578-586.

Rathbun S.L. (1998) Spatial modeling in irregularly shaped regions: Kriging estuaries. *Environmetrics*, 9, 109-129.

Shmida A. (1984) Whittaker plant diversity sampling method. *Israel Journal of Botany*, 33, 41-46.

Smith K. (1981) The prediction of river water temperatures. *Hydrological Sciences Bulletin,* 26, 19-32.

Stohlgren T.J., Falkner M.B., & Schell L.D. (1995) A modified-whittaker nested vegetation sampling method. *Vegetatio*, 117, 113-121.

Torgersen C.E., Gresswell R.E., & Bateman D.S. (2004) Pattern detection in stream networks: Quantifying spatial variability in fish distribution. *In: Proceedings of the Second Annual International Symposium on GIS/Spatial Analyses in Fishery and Aquatic Sciences,* Fishery GIS Research Group, Saitama, Japan.

U.S. Environmental Protection Agency (USEPA) (1993) *Regional Environmental Monitoring and Assessment Program.* EPA 625-R-93-012, USEPA, Office of Research and Development, Washington, D.C.

Ver Hoef J.M., Peterson E.E., & Theobald D.M. (in press) Some new spatial statistical models for stream networks. *Environmental and Ecological Statistics.*

Ward J.V. (1989) The four-dimensional nature of lotic ecosystems. *Journal of the North American Benthological Society*, 8, 2-8.

Wetzel R.G. (1992) Gradient-dominated ecosystems: Sources and regulatory functions of dissolved organic matter in freshwater ecosystems. *Hydrobiologia,* 229, 181-198.

Yuan L.L. (2004) Using spatial interpolation to estimate stressor levels in unsampled streams. *Environmental Monitoring and Assessment,* 94, 23-38.

Table 3.1. A summary of studies that have explored patterns of spatial autocorrelation in stream chemistry data using straight-line distance (SLD) and symmetric hydrologic distance (SHD).

Response Variable	Geographic Location	Distance Measure [a]	Nugget	Sill	Range (km)	Autocorrelation Function
Yuan, 2004	Maryland					
Nitrate (NO3)		SLD	N/A	N/A	49	Exponential
Sulfate (SO4)		SLD	N/A	N/A	68	Exponential
Kellum, 2003	Mid-Atlantic Highlands					
Acid nuetralizing		Anisotrophic SLD	0.104	0.21	160.93	Exponential
Gardner et al., 2003	Catskill Mountains, New York					
Temperature		SLD	2.0	3.5	6	Spherical
Temperature		SHD	2.0	3.5	7.5	Spherical
Temperature		SHD weighted by stream order	0	5.2	10	Spherical

Table 3.1. continued.

Response Variable	Geographic Location	Distance Measure [a]	Nugget	Sill	Range (km)	Autocorrelation Function
Dent and Grimm, 1999	Central					
Nitrate-nitrogen	Arizona					
2 week post flood		SHD	N/A	N/A	> 3	Spherical
2 month post flood		SHD	4.70	31.5	0.401	Spherical
9 month post flood		SHD	195.00	2265	0.359	Spherical
Soluble Reactive Phosphorus						
2 week post flood		SHD	N/A	N/A	> 3	Spherical
2 month post flood		SHD	N/A	N/A	>3	Spherical
9 month post flood		SHD	9.50	120	1.068	Spherical
Conductivity						
2 week post flood		SHD	N/A	N/A	> 3	Spherical
2 month post flood		SHD	N/A	N/A	> 3	Spherical
9 month post flood		SHD	1.00	973	1.025	Spherical
Cressie et al., in press	Queensland,					
Dissolved Oxygen	Australia	SLD	0.88	1.04	6.07	Spherical

[a] Unless otherwise noted, covariance parameters for straight-line distance are isotropic.

Table 3.2. Potential watershed covariates used in the initial model selection.

Covariate	Description
AREA	Catchment area (ha)
BARREN	% Barren landcover
WATER	% Open Water
HIGHURB	% High intensity urban
LOWURB	% Low intensity urban
PASTUR	% Hay/pasture/grass
PROBCROP	% Probable row crop
ROWCROP	% Row crop
CONIFER	% Conifer or evergreen forest type
DECIDFOR	% Deciduous forest type
MIXEDFOR	% Mixed forest type
EMERGWET	% Herbacious Emergent Wetlands
WOODYWET	% Woody wetlands
COALMINE	% Watershed area in coalmine landuse
EASTING	Easting - Albers Equal Area Conic
NORTHING	Northing - Albers Equal Area Conic
ELEV	Mean elevation in the watershed
YR96	Sample Year 1996
YR97	Sample Year 1997
ER67	Omernik's Level 3 Ecoregion 67
ER69	Omernik's Level 3 Ecoregion 69
ARGPERC	% Argillaceous rock type
CARPERC	% Carbonic rock type
FELPERC	% Felsic rock type
MAFPERC	% Mafic rock type
SILPERC	% Siliceous rock type

Table 3.3. Summary statistics for chemical response variables and significant covariates with the exception of NORTHING.

Response	Transformation	N	Min	1st Qu.	Median	Mean	3rd Qu.	Max
ANC (μeq/l)	log10(x+320)	874	2.358	2.727	2.847	2.881	2.996	3.752
CONDLAB (μmho/cm)	log10(x+1)	874	1.449	2.055	2.196	2.204	2.335	3.516
DOC (mg/l)	log10(x+1)	877	0	0.3979	0.4914	0.5643	0.699	1.53
DO (mg/l)	none	826	1.1	7.5	8.4	8.221	9.3	12.6
NO3 (mg/l)	log10(x+1)	873	0	0.2175	0.42	0.4382	0.6405	1.069
PHLAB	none	866	4.4	6.79	7.16	7.115	7.448	8.9
SO4 (mg/l)	none	870	0.3201	0.927	1.095	1.101	1.244	2.716
TEMP (°C)	none	842	11.4	17.2	19.1	19.36	21.5	28.9

Table 3.3. continued.

Covariate									
LOWURB	none	NA	0.00	0.10	0.97	6.70	5.05	76.74
HIGHURB	none	NA	0.00	0.00	0.12	1.12	0.63	22.90
WOODYWET	none	NA	0.00	0.00	0.80	2.91	3.13	34.96
PROBCROP	none	NA	0.00	10.34	19.15	20.27	29.05	85.47
ROWCROP	none	NA	0.00	2.51	6.72	9.46	13.36	50.44
PASTUR	none	NA	0.00	4.00	11.90	15.00	23.17	62.07
DECIDFOR	none	NA	1.74	16.52	26.05	32.66	45.06	98.43
CONIFER	none	NA	0.00	0.81	1.98	3.92	4.60	37.85
MIXEDFOR	none	NA	0.00	2.33	4.97	6.31	8.56	29.45
COALMINE	none	NA	0.00	0.00	0.00	0.18	0.00	12.91
WATER	none	NA	0.00	0.05	0.17	0.28	0.34	4.64
AREA (ha)	none	NA	11.63	385.46	1264.64	2602.13	3407.05	29068.57

*YR96, YR97, ER67, and ER69 were also significant covariates, but were not included the table since they are categorical data.

Table 3.4. Summary statistics for distance measures in kilometers using dissolved oxygen (n=826). We chose dissolved oxygen from several variables to provide an example.

Distance Measure	N Pairs	Min	Median	Mean	Max
Straight-line Distance	340725	0.05	101.02	118.16	385.53
Symmetric Hydrologic	62625	0.05	156.29	187.10	611.74
Pure Asymmetric[a] Hydrologic	1117	0.05	4.49	5.83	27.44

[a] Asymmetric hydrologic distance was not weighted in the summary statistics.

Table 3.5. Five significant covariates for each chemical response variable selected using leaps and bounds regression. The covariates are not listed in order of importance.

Response	Significant Covariates
ANC (μeq/l)	PASTUR, LOWURB, WOODYWET, YR96, YR97
CONDLAB (μmho/cm)	HIGHURB, LOWURB, COALMINE, YR96, NORTHING
DOC (mg/l)	WOODYWET, CONIFER, MIXEDFOR, LOWURB, NORTHING
DO (mg/l)	DECIDFOR, HIGHURB, WOODYWET, YR96, YR97
NO3 (mg/l)	PASTUR, PROBCROP, ROWCROP, LOWURB, WATER
PHLAB	PROBCROP, DECIDFOR, WOODYWET, AREA, CONIFER
SO4 (mg/l)	LOWURB, COALMINE, NORTHING, ER67, ER69
TEMP (°C)	PROBCROP, LOWURB, WATER, YR96, YR97

Table 3.6. Autocorrelation parameter estimates, mean square prediction error (MSPE), and r^2 for the general linear model (GLM), straight-line distance (SLD), symmetric hydrologic distance (SHD), and weighted asymmetric hydrologic distance (WAHD) model for each unscaled response variable. The MSPE and r^2 values were calculated using the observed and predicted values contained in the validation set.

Response Variable	Distance Measure	Model Complexity	MSPE	r^2	Nugget %	Sill	Range (km)
ANC	GLM	32	293792.34	0.411	NA	NA	NA
	SLD	32	50672.47	0.899	1.90	0.388	26.85
	SHD	32	55861.61	0.877	2.00	0.286	57.57
	WAHD	32	87203.41	0.843	0.90	0.644	47.76
CONDLAB	GLM	32	34969.41	0.712	NA	NA	NA
	SLD	32	11623.06	0.921	0.90	0.961	12.47
	SHD	32	3239.63	0.959	1.30	0.573	27.67
	WAHD	32	4014.98	0.948	1.10	0.569	45.27
DOC	GLM	32	7.85	0.520	NA	NA	NA
	SLD	32	5.46	0.644	15.40	0.282	56.39
	SHD	32	5.37	0.656	28.90	0.693	180.79
	WAHD	32	5.47	0.649	19.90	0.734	82.32
DO	GLM	32	1.91	0.294	NA	NA	NA
	SLD	31	1.58	0.414	54.50	0.202	62.77
	SHD	31	1.64	0.392	70.40	0.283	301.76
	WAHD	32	1.74	0.355	39.80	0.263	82.32

Table 3.6. continued.

Response Variable	Distance Measure	Model Complexity	MSPE	r^2	Nugget %	Sill	Range (km)
NO3	GLM	32	1.14	0.671	NA	NA	NA
	SLD	32	0.82	0.772	3.60	0.593	20.78
	SHD	32	0.75	0.783	7.40	0.957	45.13
	WAHD	32	0.95	0.725	6.50	0.937	73.30
PHLAB	GLM	32	0.16	0.504	NA	NA	NA
	SLD	31	0.11	0.663	4.70	0.647	16.35
	SHD	31	0.10	0.679	6.40	0.500	36.46
	WAHD	32	0.11	0.663	3.50	0.503	33.99
SO4	GLM	32	363.78	0.190	NA	NA	NA
	SLD	20	210.14	0.400	18.10	0.271	23.46
	SHD	28	259.42	0.360	30.60	0.443	40.84
	WAHD	28	292.21	0.286	17.60	0.922	82.32
TEMP	GLM	32	8.81	0.177	NA	NA	NA
	SLD	32	7.72	0.278	12.50	0.310	6.90
	SHD	32	7.49	0.298	42.00	0.702	14.03
	WAHD	32	7.37	0.309	18.80	0.473	15.49

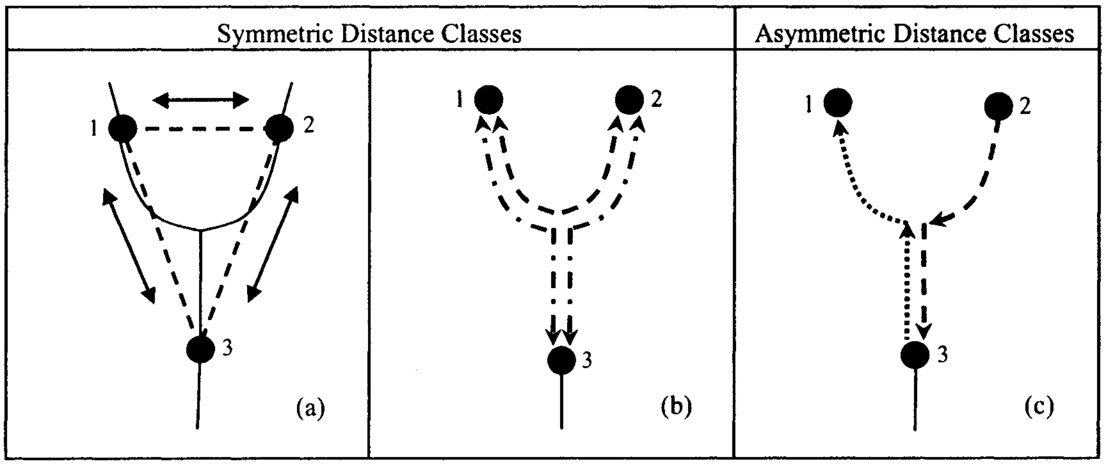


Figure 3.1. Symmetric and asymmetric distance classes. The stream network is represented by a solid line, while distance measurements are represented with dashed lines. Symmetric hydrologic distance measures include straight-line distance (a) and symmetric hydrologic distance (b). Sites 1, 2, and 3 are all neighbors to one another when these distance measures are used. Asymmetric distance classes include upstream and downstream asymmetric hydrologic distance (c). Sites 1 and 2 are neighbors to site 3, but not to each other.

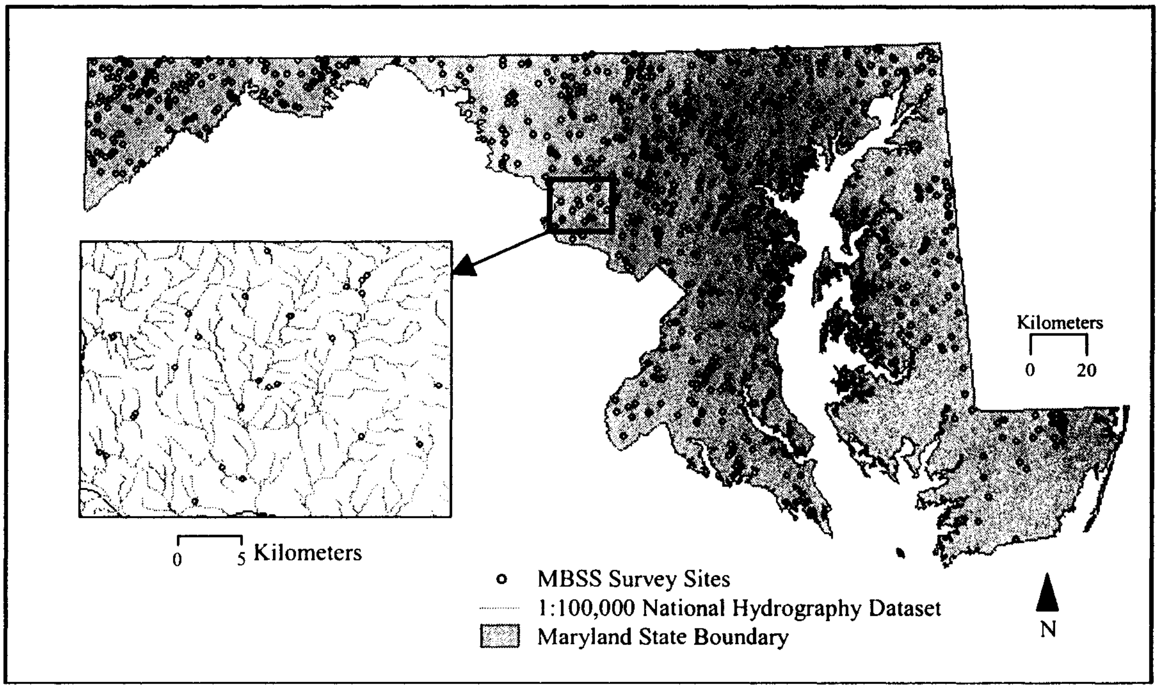


Figure 3.2. The Maryland Biological Stream Survey (MBSS) dataset. The MBSS data includes chemical, physical, and biological data, which were collected throughout Maryland by the Maryland Department of Natural Resources during 1995, 1996, and 1997.

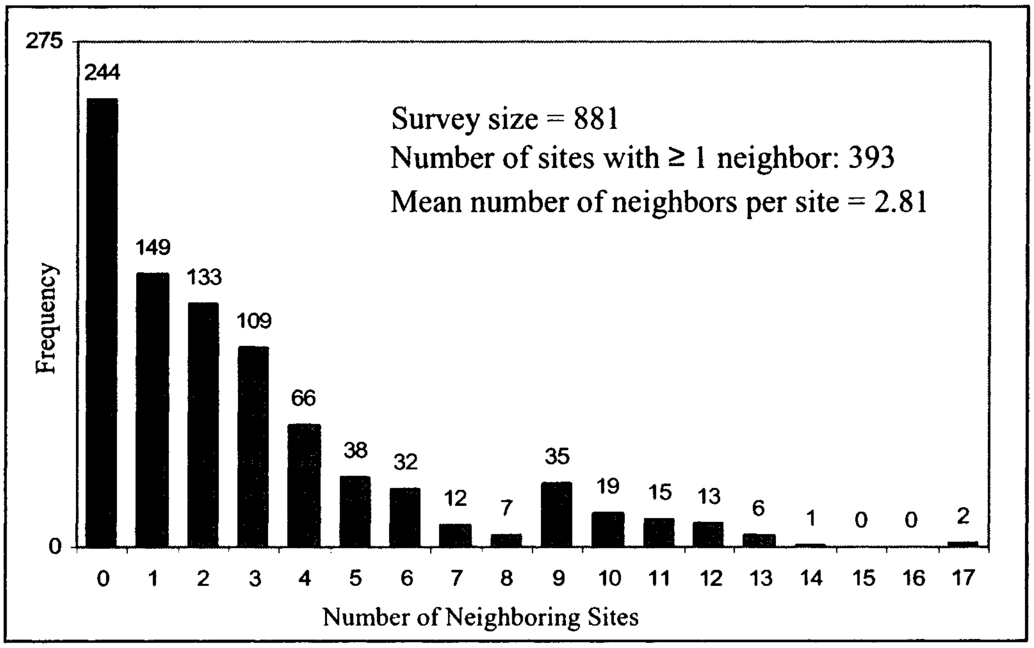


Figure 3.3. Spatial relationships between neighboring survey sites. Summary of spatial relationships between neighboring survey sites for acid neutralizing capacity using the weighted asymmetric hydrologic distance measure.

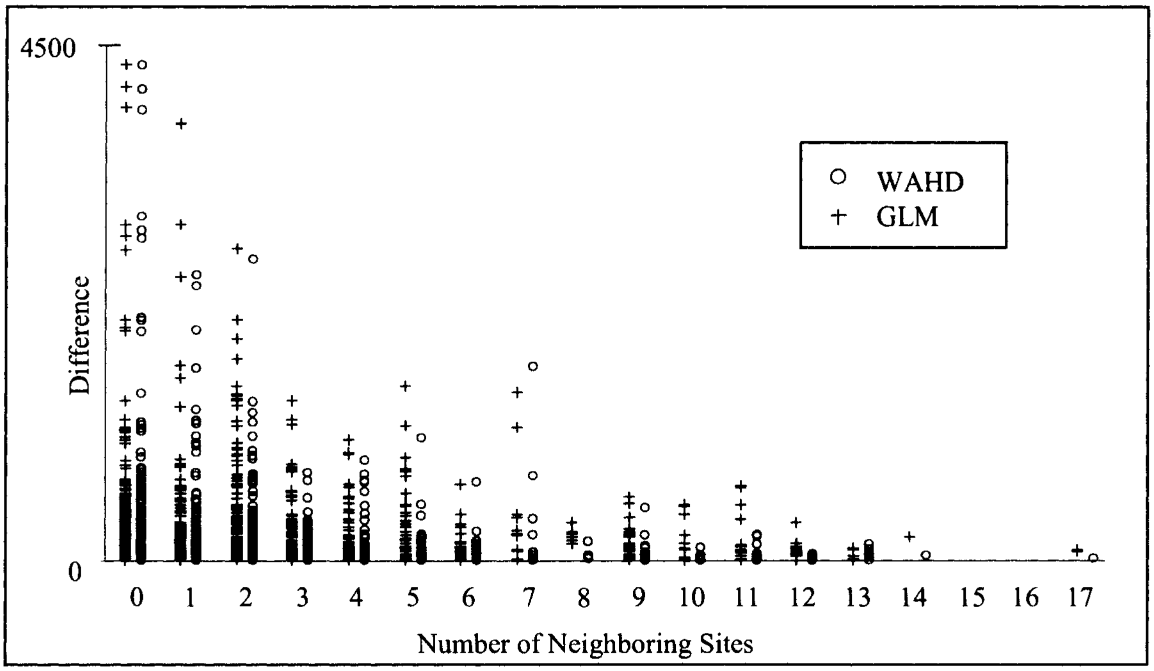


Figure 3.4. Difference between observed values and general linear model (GLM) predictions or observed values and weighted asymmetric hydrologic distance (WAHD) predictions according to the number of neighbors.

CHAPTER FOUR

PREDICTING WATER QUALITY IMPAIRED STREAM SEGMENTS USING LANDSCAPE-SCALE DATA AND A REGIONAL GEOSTATISTICAL MODEL: A CASE STUDY IN MARYLAND[3]

INTRODUCTION

The Clean Water Act (CWA) of 1972 requires states and tribes to identify water quality impaired stream segments, to create a priority ranking of those segments, and to calculate the Total Maximum Daily Load (TMDL) for each impaired segment based upon chemical and physical water quality standards. States and tribes are also required to create a biennial water quality inventory that characterizes regional water quality based on the attainment of designated-use standards assigned to individual stream segments. Yet, it is impossible to physically survey every stream within a large area due to the immense number of segments, limited personnel, and the cost associated with sampling (Olsen and Ivanovich, 1993; Herlihy et al., 2000; USEPA, 2001). In addition, lawsuits have been filed in 38 states by environmental groups demanding that the requirements of the CWA be met in a timely fashion (Copeland, 2002). These increased pressures have led to the need to develop a rapid and cost-efficient survey method that has the ability to statistically predict potentially impaired stream segments in large areas. To address this

---

[3] This chapter was submitted to Environmental Monitoring and Assessment and my co-author was N. Scott Urquhart.

issue, we developed a geostatistical model based on coarse-scale geographical information system (GIS) data and used it to make predictions for every stream segment in our study area. We predicted dissolved organic carbon (DOC) because it can be used to identify waters that may be biologically or ecologically stressed (MDDNR, 1999), but these methods could also be applied to other water quality constituents.

Dissolved organic carbon (DOC) is an important constituent of water quality because it affects the chemical and biological condition of freshwater ecosystems. DOC is a significant energy source in aquatic food webs (Wetzel, 1992), absorbs biologically harmful ultraviolet rays that penetrate the water column (Williamson et al., 1996; Kiffney et al., 1997), acts as a weak acid (Sullivan et al., 1989), and binds dissolved substances, such as metals, making them temporarily less bioavailable (Driscoll et al., 1995; Prusha and Clements, 2004).

The concentration of DOC in headwater streams is strongly influenced by the production and transport of organic matter from the terrestrial environment. The main sources are soil, groundwater, and dead terrestrial plant material (Wetzel, 1992). DOC is transported from the watershed via overland, sub-surface, or base flow and the flow path of water affects the stream concentration. Shallow sub-surface paths and overland flow through wetlands, organic soil layers, and shallow soils tend to produce water with relatively high concentrations of DOC (Mulholland, 2003). Conversely, sub-surface or base flow moving through deeper soil horizons may lose DOC, which is adsorped by the mineral soils (Qualls and Haines, 1992).

Models based on both local and coarse-scale input data have been created to explain variability in lake or stream DOC (Rasmussen et al., 1989; Houle et al., 1995;

Currie and Aber, 1997; Neff and Asner, 2001; Ouyang, 2003). For our purposes, local-scale data refers to fine-scale measurements, such as depth of litter mass, which must be collected in the field. Coarse-scale or landscape-scale data represent lumped watershed attributes, such as percent wetlands or mean elevation in the watershed, which are typically calculated using GIS or remotely sensed data. Models that include locally-derived input data are not suitable for regional DOC estimation because they require extensive and expensive field sampling (Herlihy et al., 2000; USEPA, 2001). Models based solely on remotely derived coarse-scale data have also been generated to explain variability in DOC concentration. There appears to be a significant tradeoff between the cost, in both time and money, of input data and the accuracy of the model predictions. For example, Creed and others (2003) delineated wetlands using LIDAR (light detection and ranging) data and were able to explain 91% of the variance in stream and lake DOC in 12 watersheds. Their results are promising, but their input data are currently expensive (Haneberg, 2005) and therefore unsuitable for regional monitoring.

Other models have been generated using coarse-scale remotely sensed data, such as United States Geological Survey (USGS) Land Use Land Cover data (USGS, 2005) or USGS digital elevation models (DEM) (USGS, 2003), which are easily accessible and available at no cost. Yet, these models explained less variance in DOC (Eckhardt and Moore, 1990; Herlihy et al., 1998; Gergel et al., 1999; Prusha and Clements, 2004). For example, Kortelainen (1993) used latitude, catchment area/lake area ratio, and percent of watershed area covered by peatlands, fields, and upstream lakes to explain 55% of the variance in DOC collected in 978 lakes throughout Finland. Canham and others (2004) built a process-based model to better understand the production, transport, and loss of

DOC in 428 Adirondack lakes based on landscape data and were also able to explain 55% of the variance in DOC. Although these models are useful, there is a need for more accurate regional models based on accessible and inexpensive input data.

We recently completed research that indicates that geostatistical models based on coarse-scale data have the ability to produce more accurate stream predictions than non-spatial models based on coarse-scale landscape data (Peterson et al., in review[2]). Geostatistical models are similar to traditional statistical models, in that they represent the broad-scale trend in the mean of the data, but relax the assumption of independence and allow spatial autocorrelation in the residuals. Local deviations from the mean are modeled using the covariance between neighboring sites. Heterogeneity in the broad-scale mean and variance are permitted, but the mean, variance, and autocorrelation structure of the error term are assumed to be stationary or similar across a study area (Bailey and Gatrell, 1995).

The covariance represents the strength of spatial autocorrelation between pairs of sites within a spatial neighborhood given their separation distance (Olea, 1991). A spatial neighborhood includes sites that are nearby and have a quantifiable influence upon one another. Sites outside of the spatial neighborhood are not considered to be spatially correlated. The separation distance is simply the distance traveled from one location to a second location.

The separation distance can be calculated using a variety of distance measures (Peterson et al., in review[2]), but we only concern ourselves with straight-line distance (SLD) (aka Euclidean distance) and asymmetric hydrologic distance weighted by discharge volume. SLD (Figure 4.1a) is directionless (isotropic) and has equal correlation in all directions. In addition, all locations in a study area are considered potential neighbors. Hydrologic distance is simply the distance between two locations when movement is restricted to the stream network. Asymmetric hydrologic distance (Figure 4.1b) is a unidirectional measure and water must flow from one location to another for two sites to be neighbors. Spatial weights are generated using metrics that represent relative network position, such as watershed area, and used to represent discharge volume (Ver Hoef et al., in press; Peterson et al., in review[1]).

Geostatistical models have not typically been used to model water chemistry in stream networks (but see Gardner et al., 2003; Kellum, 2003; Yuan, 2004; Peterson et al., in review[2]), but they have been shown to improve the accuracy of water chemistry predictions (Yuan, 2004; Peterson et al., in review[2]). It is common for researchers to obtain an estimate of the model fit by removing observed sites from the dataset and making predictions at those sites (Gardner et al., 2003; Yuan, 2004). To our knowledge, no one has used a fitted geostatistical model to make water chemistry predictions for every unobserved stream segment in their study area. We created a set of geostatistical models using coarse-scale GIS data and two distance measures, SLD and weighted asymmetric hydrologic distance (WAHD). We compared the models to determine which distance measure more accurately represented the spatial relationship of stream DOC. We selected the model with the best fit and used it to generate predictions and prediction variances for 3083 stream segments located in seven interbasins throughout Maryland.

METHODS

Study area and design

The Maryland Biological Stream Survey (MBSS) data (Figure 4.2) were collected throughout Maryland by the Department of Natural Resources (DNR) (Mercurio et al., 1999). Maryland is a geographically diverse state that can be divided into five physiographic provinces: the Appalachian Plateau, the Blue Ridge, the Coastal Plain, the Piedmont, and the Valley and Ridge (Boward et al., 1999). Elevation increases from the eastern Coastal Plain to the Appalachian Plateau in the west.

The Maryland DNR used a probability-based survey to collect chemical, physical, and biological data from first, second, and third order non-tidal streams in 17 interbasins throughout the state (Mercurio et al., 1999). An interbasin differs from a true watershed because it does not include the entire area that flows to an outlet point. A stratified random sample was collected from each interbasin based on Strahler stream order (Strahler, 1957) and the number of samples collected per stream order was proportional to the number of stream order miles within the interbasin. Sampling was conducted during 1995, 1996, and 1997, but we restricted our analyses to data collected in 1996 for a number of reasons. We needed to reduce the processing time for geostatistical model selection and wanted to reduce differences in DOC resulting from interannual variation. In addition, the WAHD measure performed better using the DOC data collected in 1996 than the data from 1995 and 1997.

Seven interbasins were visited and 343 DOC samples were collected at individual locations between March 1 and May 1, 1996. Grab samples were collected in one liter bottles and stored and shipped to the laboratory on wet ice within 48 hours. The samples had a holding time of 14 days and were analyzed using a Doorman DC-80 carbon analyzer, which had a minimum detection limit of 1.0 mg per liter (Mercurio et al., 1999). The samples represent the carbon that remained after filtration with a 0.45 µm filter (USEPA, 1987).

GIS preprocessing

The stream network and survey sites were preprocessed in a GIS to ensure that sites were positioned on the correct stream segment. There are a variety of reasons why it is rare for GIS data collected within a stream to fall directly on a line segment representing a stream. Though points collected using global positioning systems are differentially corrected, they still have some error. Some stretches of river can move (e.g. meander) slightly from their mapped position. Streams are often represented by lines and so samples collected on the banks of a large river may not fall directly on a line segment. Digital streams datasets may contain mapping errors and generalizations, such as the absence of small tributaries and the homogenization of form, which are found when streams are represented at coarser scales. As a result of these data problems, we discarded 23 sites because the survey stream could not be identified.

Distance matrices were generated for SLD and asymmetric hydrologic distance measures (Figures 1a and 1b) (Appendices I and II). We projected the data from latitude/longitude to Albers Equal Area projection (North American Datum 1983 based on the GRS1980 spheroid) before calculating the distance measurements. Projecting the data was necessary because the latitude/longitude coordinates have a known, systematic bias associated with increasing latitude. The SLD matrix was calculated in R statistical software package (Ihaka and Gentleman, 1996) using northing and easting values as x, y coordinates. The asymmetric hydrologic distance and spatial weights matrices were calculated in a GIS using programs written in Visual Basic for Applications for ArcGIS version 8.3 (ESRI, 2002) (Appendices I and II).

The spatial weights are used to develop the WAHD measure and represent the relative influence of one site on another. If two sites are not connected by flow the spatial weight is equal to zero and a sites influence on itself is equal to one. The spatial weights for flow-connected sites are based on watershed area, which we use as a surrogate for discharge volume. The spatial weights are generated by calculating the upstream watershed area for the downstream node of each segment in the stream network using a GIS. We define a stream segment as the portion of a stream located between two confluences. When survey sites fall midway along a segment it is split into two separate stream segments. At each confluence or survey site in the network, the total upstream watershed area is calculated by summing the watershed area for the incoming stream segments. The proportional influence for each incoming segment is calculated by dividing its watershed area by the total upstream watershed area at the confluence or survey site (Figure 4.3). Every stream segment in the network contains its proportional influence on the segment directly downstream when this process is complete. Then, we locate the path between flow-connected sites and calculate the influence of one site on another, which is equal to the product of the segment proportional influences found in the path. The spatial weights matrix is simply an n by n matrix that contains the square root of the proportional influence for all pairs of sites, which ensures that stationarity in the variances is maintained (Ver Hoef et al., in press). The GIS methods used to generate the hydrologic distance matrices and spatial weights matrix are lengthy and are not the focal point of this manuscript (but see Peterson et al., in review[1]).

Statistical analyses (Appendix III)

The MBSS dataset contains coarse-scale watershed data for each survey site (Mercurio et al., 1999), which the Maryland DNR derived using a GIS and the 1992 National Land Cover Data (NLCD) (MRLC Consortium, 2003) (Table 4.1). In addition, we calculated Level III Omernik's ecoregion (Omernik, 1987), geographic location, temperature, elevation, slope, precipitation, geology type, soil pH, and a soil erodability factor (Table 4.1). The variance inflation factor (VIF) collinearity statistic (Helsel and Hirsch, 1992) indicated that some potential covariates were significantly correlated with other covariates (VIF > 10) and we removed them from further analysis.

We reduced the list of potential covariates (Table 4.1) due to the processing time required for model selection. We used a Leaps and Bounds algorithm (Furnival and Wilson, 1974) to find the "best" set of covariates and used them to develop a linear model. Ten watershed attributes were selected as potential covariates: WATER, WOODYWET, EMERGWET, FELPERC, MINTEMP, ER64, ER65, ER66, ER67, and ER69 (Table 4.1). The WATER, WOODYWET, EMERGWET, AND FELPERC covariates represent percent open water, forested or shrubby wetland areas, perennial herbaceous wetland areas, and felsic rock type in the watershed (Appendix IV), respectively (Table 4.1). MINTEMP is the mean minimum temperature in the watershed for the first four months of 1996. ER64, ER65, ER66, ER67, and ER69 represent the Omernik's Level III ecoregion (Omernik, 1987) where the site is located. We checked the model residuals for signs of non-normality and transformed DOC using a $\log_{10}$ transformation. The studentized residuals were used to identify eight extreme outliers at a significance level less than 0.01. These sites were not spatially clustered and there were no evident patterns related to DOC or watershed attributes. Therefore, the survey sites were removed from further analysis.

We restricted the model space to all possible linear models using the 10 "best" explanatory variables determined by the initial covariate selection process described above. For a given distance measure there were 1024 ($2^{10} = 1024$) competing models. Five sets of 1024 models were generated using the SLD measure and one of five common autocovariance models: the exponential, spherical, Mariah, hole effect, and rational quadratic. Moving average representations of common autocovariance functions were required to generate the four sets of WAHD models and we used the exponential, spherical, linear with sill, and Mariah functions, which were presented by Ver Hoef and others (in press). We also assumed that the model residuals were normally distributed with mean zero and variance-covariance matrix $\Sigma = \sigma^2\Omega$, where σ^2 is the variance and $\Omega = \Omega(d;\theta)$ is the correlation matrix. Note that Ω is a function of the distance between sites, d, given the autocorrelation parameter vector, θ. Therefore, the model for response variable Z is written in matrix notation as $Z = X\beta + \varepsilon$, where $\varepsilon \sim N(0, \sigma^2\Omega)$. Here X is the $n \times p$ design matrix of covariates, β is a vector of coefficients of length p, and ε is a vector of n (correlated) errors.

The log-likelihood function of the parameters $\left(\theta, \beta, \sigma^2\right)$ given the observed data, Z, is

$$\ell(\theta,\beta,\sigma^2;Z) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left|\sigma^2\Omega\right| - \frac{1}{2\sigma^2}(Z - X\beta)'\Omega^{-1}(Z - X\beta). \quad (4.1)$$

Maximizing the log-likelihood (Equation 4.1) with respect to β and σ^2 yields

$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Z$ and $\hat{\sigma}^2 = \dfrac{(Z - X\hat{\beta})'\Omega^{-1}(Z - X\hat{\beta})}{n}$. Both maximum likelihood estimators (MLE) can be written as functions of θ alone. Thus, we derive the *profile* log-likelihood function by substituting the MLEs back into (Equation 4.2):

$$\ell_{profile}(\theta;\hat{\beta},\hat{\sigma}^2,Z) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\hat{\sigma}^2) - \frac{1}{2}\log|\Omega| - \frac{n}{2}. \quad (4.2)$$

The primary advantage to using the profile log-likelihood is that it reduces the dimensionality of the problem, which can decrease the amount of time required to find a numerical solution. This is especially important when there are a large number of models to compare.

The correlation matrix, Ω, can be computed using any of the autocorrelation functions we mentioned previously, but here we use the exponential autocorrelation function as an example.

$$C_1(d;\theta_1,\theta_2) = \begin{cases} 1 & \text{if } d = 0 \\ (1 - \theta_1)\exp(-d/\theta_2) & \text{if } d > 0 \end{cases} \quad (4.3)$$

where θ_1 is the *proportion* of nugget effect and θ_2 is the range parameter. The nugget is estimated by $\hat{\theta}_1\hat{\sigma}^2$ where θ_1 is restricted between zero and one. The approximate range over which sites are considered to be correlated is $3\theta_2$ (Cressie, 1993). The value d

represents the distance between any two sites relative to the distance measure, e.g., SLD or WAHD. The correlation matrix for the WAHD measure is generated by taking the Hadamard (element-wise) product of (Equation 4.3) and the spatial weights matrix (Ver Hoef et al., in press).

The MLE for θ is found by maximizing the profile log-likelihood (Equation 4.2) using a quasi-Newton method (Byrd et al., 1995), which is in turn used to compute the MLEs for the model parameters: β and σ^2. To promote numerical stability, we standardized the response and explanatory variables to have mean zero and unit variance. We also scaled the distances to fall between zero and one, but the estimates reported here have been converted to the original units.

We used the spatial Akaike Information Corrected Criterion (AICC) statistic (Hoeting et al., in press) to select the geostatistical model with the most support in the data from each model set. The spatial AICC is defined as

$$AICC = -2\ell_{profile}(\theta;\beta,\sigma^2,Z) + 2n\frac{p+k+1}{n-p-k-2}, \tag{4.4}$$

where n is the number of observations, $p-1$ is the number of covariates, and k is the number of autocorrelation parameters.

The model with the smallest AICC from each set was used to generate predictions using the Universal Kriging algorithm (Cressie, 1993). We used a leave-one-out cross-validation method to calculate the mean square prediction error (MSPE) for each model of the nine models (Equation 4.5). The MSPE is defined as

$$MSPE = \frac{\sum_{i=1}^{n_p}(Z_i - \hat{Z}_i)^2}{n_p}, \tag{4.5}$$

where Z_i is the observed value at site *i*, $\hat{Z}_i$ is the predicted value at site *i*, and n_p is the total number of predictions. Models with small MSPE are desirable. The models constructed using different distance measures have unique variance structures and therefore cannot be compared using the spatial AICC. The MSPE provided a way to compare models constructed using different distance measures and to determine which measure, if any, was more able to account for the variability in the response variable. We generated cross-validation intervals for the covariate parameters, which contain 95% of the 312 parameter values. In addition, we calculated the r^2 for the predictions and observations.

The seven interbasins surveyed in 1996 contained 3083 first, second, and third order non-tidal stream segments. We created 3083 prediction locations by calculating the location for the downstream node of each stream segment. We used the downstream node to ensure that the entire segment was located within the same watershed. This caused more than one prediction location to be positioned at stream confluences. However, the covariates for the prediction locations represented the watershed conditions of the individual segment rather than the confluence, which would include all of the segments that flow into that location.

We generated a distance matrix that contained the SLD between both observed and unobserved sites. We used the Functional Linkage of Watersheds and Streams (FLoWs) tools (Theobald et al., 2005) developed for ArcGIS version 9 (ESRI, 2002) to calculate the watershed covariates (WATER, EMERGWET, WOODYWET, FELPERC, and MINTEMP) for the 3083 prediction locations. We used the fitted covariances based on the Mariah autocovariance model and the Universal Kriging algorithm to generate predictions and their variances at the prediction locations (Cressie, 1993). The prediction values and variances were assigned back to each stream segment in a GIS to visualize the results.

RESULTS

Table 4.2 contains a summary of the distribution for each variable. DOC values ranged between 0.6 and 15.9 mg/l. There were 207 sites located in ER64, 16 in ER65, 11 in ER66, 39 in ER67, and 19 in ER69. The remaining 20 sites were located in ER63, which was not included in the model. The Maryland DNR set a threshold level of 8.0 mg/l for DOC and determined that values greater than the threshold were a possible indicator of environmental stress resulting from anthropogenic influences (MDDNR, 1999). Only eight MBSS sites collected in 1996 and located west of Chesapeake Bay had DOC values greater than 5 mg/l and only two were greater than 8.0 mg/l (Figure 4.2). Larger DOC values were found to the north and east of Chesapeake Bay (4.8 to 15.9 mg/l). However, two of these sites were rated as naturally acidic blackwater streams and their large DOC concentrations would not be considered an indicator of stress (Boward et al., 1999).

As we expected, the predictive ability of the geostatistical models based on SLD and WAHD differed. The SLD models explained more variance in DOC than models based on WAHD for every autocovariance model except the spherical model (Table 4.3).

However, with the exception of the SLD spherical model, the predictive ability of models within distance measure was comparable.

The SLD geostatistical models based on the exponential, Mariah, and rational quadratic autocovariance models had the lowest MSPE values, produced the most accurate predictions, and were essentially equal in their predictive ability (Table 4.3). The rational quadratic model included six covariates: WATER, EMERGWET, WOODYWET, FELPERC, MINTEMP, and ER67. The exponential and Mariah models were similar, but excluded the ER67 covariate. Yet, the models produced dissimilar ranges of spatial autocorrelation. Sites are considered uncorrelated at a distance greater than $3\theta_2$ (Cressie, 1993). Therefore, the effective range of spatial autocorrelation for the exponential, Mariah, and rational quadratic models were 60.24 km, 21.09 km, and 35.14 km, respectively. Despite this difference, the correlation coefficients for the model predictions indicated that the models produced nearly identical predicted values ($r^2 \geq$ 0.990). A comparison of the model composition, MSPE values, and r^2 values suggested that there were no distinct differences between the predictions produced by the exponential, Mariah, and rational quadratic SLD models. We limited additional model exploration to the SLD Mariah model because it had that lowest MSPE value and the largest r^2 value.

The SLD Mariah model included five covariates. DOC had a positive relationship with WATER, EMERGWET, WOODYWET, and MINTEMP (Table 4.4). In contrast, FELPERC had a slight negative correlation with DOC. The cross-validation intervals for the regression coefficients were narrow (Table 4.4). The model described 72% of the variance in DOC. However, there was one DOC value (observed value = 15.9 mg/l) that had an unusually large effect on the correlation coefficient (Figure 4.4) and the model only explained 66% of the variance in DOC when that observation was removed.

The square prediction error (SPE) values for the individual model predictions produced by the SLD Mariah model were between 0 and 18.706, but were generally low. Eighty-nine percent of the values were less than 1.5 and 71% of the values were less than 0.5. However, there appeared to be an east-west trend associated with the spatial location of the SPE values (Figure 4.5). The low SPE values in the western portion of Maryland indicated that the model fit the data well in this area. In contrast, larger SPE values were found in the central and eastern portions of Maryland. We examined the 35 SPE values greater than 1.5 and found that the model produced conservative estimates. It underestimated large DOC values 29 times and overestimated lower DOC values only 6 times. These errors occurred at sites where the covariate values were similar to neighboring sites, but the observed value was considerably different from nearby values. We examined other watershed characteristics that were not included in the model, such as %FOREST, %URBAN, mean slope, and watershed area, to determine whether sites with large SPE values had unique characteristics that differed from conditions at other sites collected in 1996. However, the covariate distributions taken from the sites with large SPE values were similar to the overall statistical distribution of those covariates and to the watershed covariates at nearby sites.

The parameter estimates for the SLD Mariah model used to make the predictions at unobserved segments are listed in Table 4.5. The prediction segments represented 5973 km of streams in Maryland (Figure 4.6). The prediction values ranged from 0.76 to 40.44 mg/l and the prediction variances were between 0.05 and 2.6 (Table 4.6). In Maryland, DOC values less than 5 mg/l are considered low and values greater than 8 mg/l are high (MDDNR, 1999). The model predicted that 90.2% of streams had DOC values less than 5 mg/l, 6.7% were between 5 and 8 mg/l, and 3.1% of streams produced values greater than 8 mg/l (Figure 4.6). There were 18 prediction values that were greater than 15.9 mg/l and these segments also possessed the largest prediction variances (1.1 to 2.6). Although the prediction values exceeded the largest observed value in the 1996 data, they may be somewhat reasonable since the largest DOC value found in the complete MBSS dataset was 32.9 mg/l. Sixteen of the prediction values greater than 15.9 mg/l were located in watersheds with WATER, EMERGWET, or WOODYWET values that were substantially larger than those represented in the observed data. The other two segments with large prediction values resulted from artifacts in the streams data. The two stream segments drain directly into large reservoirs and a portion of the reservoir was erroneously included in the watershed, which caused the WATER covariate to be artificially high.

DISCUSSION

Our initial model selection procedure narrowed the field of covariates to five attributes that represented watershed conditions and five that represented Omernik's ecoregions. However, the only ecoregion covariate included in the final geostatistical models was ER67, which suggested that the variance in DOC that was previously explained by Omernik's ecoregion could also be explained using the covariances between neighboring sites. The initial model selection method was non-spatial and we questioned whether other watershed covariates would have been selected if spatial autocorrelation was accounted for during the model selection process. We attempted to address this question by replacing the ecoregion covariates with three watershed covariates that were part of the MBSS dataset: percent urban, percent conifer forest type, percent mixed forest type and two watershed covariates that we calculated in a GIS: mean slope and mean annual precipitation (Table 4.1). We generated another set of geostatistical models using the same methodology and the new set of covariates. Substituting watershed covariates for ecoregion covariates did not improve the accuracy of the model predictions. Although watershed slope and forest cover are correlated to DOC concentration in other studies (Eckhardt and Moore, 1990; Canham et al., 2004; Prusha and Clements, 2004), they do not appear to strongly influence DOC concentration in Maryland at this scale of analysis.

The positive relationships between stream DOC and percent water and wetland area that the SLD Mariah model demonstrated (Table 4.4) have been well documented in the ecological literature (Eckhardt and Moore, 1990; Wetzel, 1992; Kortelainen, 1993). Wetlands and water are closely related because wetlands are generally located at the land-water interface. Primary productivity is high and the decomposition of macrophytes is a significant source of nutrients (Wetzel, 1992). Wetlands also act as depositional areas for organic matter from terrestrial sources, such as soil, groundwater, and dead terrestrial plant material. Erosion within wetlands is rare because of low topographic gradients. As a result, the organic matter pool is large and its decomposition provides a continuous supply of nutrients to the stream (Wetzel, 1992).

Interestingly, there was a strong positive relationship between DOC and MINTEMP (Table 4.4). Hejzlar and others (2003) also reported a positive relationship between chemical oxygen demand, which they used as a surrogate for DOC, and increasing monthly mean air temperature. They attributed the relationship between temperature and DOC to the combined effect of increased precipitation and temperature on soil moisture or diminished periods of winter freezing. There is very little correlation between precipitation covariates and MINTEMP in our dataset ($r^2 < 0.0007$) and precipitation does not appear to be directly correlated to DOC ($r^2 < 0.0002$). Therefore, we believe that the positive relationship between DOC and MINTEMP may be due to diminished periods of winter freezing and the resultant acceleration of ecological processes associated with DOC cycling. These processes might include the timing of snowmelt, seasonal runoff volumes, terrestrial and aquatic primary production, and microbial decomposition.

We observed a weak negative relationship between percent felsic rock type in the watershed (FELPERC) and DOC (Table 4.4). Felsic rock is an igneous rock type that is rich in feldspar and silica (American Geological Institute, 1987) (Appendix IV). Previous studies have shown that DOC in soil water is adsorped when organic ligands form metal-organic complexes at the mineral surface of felsic rocks (Amrhein and Suarez, 1988; Bennett, 1991). DOC is converted from a dissolved phase to a solid phase during this process and we believe that this slightly reduced the concentration of DOC that moved from the soil to the stream.

The SLD measure consistently produced geostatistical models that described a greater amount of variance in DOC than the WAHD measure. Stream DOC concentration is significantly influenced by the terrestrial production of organic matter (Wetzel, 1992). In the absence of human modifications, landscape characteristics, such as vegetative composition or soil type, tend to be similar in neighboring watersheds. In addition, wetlands are generally located in watersheds with gently sloping topography and poorly drained soils. SLD may represent variance in DOC related to organic matter production more accurately than WAHD because spatial patterns associated with soil type, vegetative composition, and topography are not constrained to watershed boundaries.

The WAHD model performance may have also been hindered by a lack of neighboring sites. We examined the spatial neighborhood for the SLD measure using a 21.09 km range of spatial autocorrelation and found that the mean number of neighbors was 36.88. Every site had at least five neighbors and 279 sites had more than 15 neighbors. However, the WAHD spatial neighborhood is restricted to flow-connected sites, which dramatically reduced the number of neighboring sites. We did not impose a range of spatial autocorrelation on the WAHD spatial neighborhood and instead examined all flow-connected sites. The mean number of neighbors per site using the WAHD was only 1.11. There were 170 sites that did not have neighbors and only 87 sites had more than one neighbor. When a spatial neighborhood is deficient or absent for a specific site, the geostatistical model is essentially non-spatial. It has the ability to explain the broad-scale mean in the data, but does not provide additional predictive ability at that site. This is a common feature for geostatistical models. The associated standard error for prediction sites with many observed neighbors is small compared to sites that have few or no neighbors. Thus, the WAHD model had the ability to explain the broad-scale mean in the data, but did not provide additional predictive ability at sites with few or no neighbors.

The cross-validation intervals for the SLD Mariah model regression coefficients were all narrow, which indicated that the relationships between the covariates and DOC were consistent throughout Maryland (Table 4.4). There were few extreme regression coefficient values, but we examined the largest and smallest values and determined that the majority were not produced by common sites and that the sites were not clustered in space. This suggested that a local-scale factor, such as a point source of organic waste, affected stream DOC and was not explained by our model. The narrow cross-validation intervals for the regression coefficients and the lack of extreme outliers indicated that the spatial location of the sites was not as important as the watershed characteristics.

Spatial patterns were evident in the distribution of stream DOC predictions and prediction variances throughout Maryland (Figures 4.5 and 4.6). The majority of streams west of Chesapeake Bay produced low DOC predictions with low prediction variances. However, there was a small cluster of elevated DOC predictions in stream segments located within Baltimore. The watershed conditions at these sites were not the source of elevated DOC predictions. Instead, relatively large observed values at nearby sites, which likely result from organic pollution discharged into Baltimore waterways, increased the prediction values. Stream segments to the north and east of Chesapeake Bay had larger DOC predictions and larger prediction variances. Some areas of coastal Maryland are characterized by blackwater streams, which are naturally acidic and have elevated levels of stream DOC (Boward et al., 1999). We believe that the majority of the elevated stream predictions east of Chesapeake Bay are the result of natural environmental processes. It is not surprising that the variances associated with these predictions were also large since there were only two observed sites classified as blackwater in the MBSS dataset. Consequently, the unique ecological conditions that produce blackwater streams were not well represented in the observed dataset and the model was unable to account for this natural variability.

Overall, the geostatistical model that we developed described more variance in stream DOC than previous models based solely on coarse-scale landscape data. The model fit the data better in western and central Maryland compared to the eastern coastal regions (Figure 4.5) for a number of reasons. The regression equation was fit to the mean in the data and nearly all of the observed sites had low DOC values. In contrast to western and central Maryland, eastern Maryland contained blackwater sites that were not well represented in the observed data. In addition, there appears to be less variation in stream DOC in western and central Maryland and consequently neighboring sites tend to be similar. The separation distances for survey sites are shorter to the west of Chesapeake Bay compared to the east (Figure 4.2). Large separation distances result in a weaker covariance between observed and predicted sites and less confidence in the predictions. Given the statistical and spatial distribution of the observed data, it is not surprising that the model was able to predict DOC values more accurately in western and central Maryland. It may be necessary to develop a separate geostatistical model, which is fit to data collected in the coastal region to provide more accurate DOC predictions in eastern Maryland. Nonetheless, the ecological processes that influence stream DOC to the west of Chesapeake Bay appear to be similar and can be described using a single geostatistical model.

Although the geostatistical model fit the data well, it was developed to provide a general estimate of stream DOC throughout Maryland and there are clearly influential factors that were not adequately represented. For instance, the model was unable to account for abrupt differences in DOC values between neighboring sites when the sites had similar watershed conditions. These differences may result from local factors, such as point sources of organic inputs, which were not detected at our scale of analysis. It is also possible that elevated DOC values result from non-point sources of pollution and that our lumped watershed attributes were too general to capture the information. For example, a wetland area 100 meters upstream should have a larger impact on site DOC than a wetland five kilometers upstream. Lumped watershed attributes are non-spatial and any differences that result from the spatial location of the landuse within the watershed are not represented. In addition, it is challenging to represent ecological processes, such as the flow path that water takes from the terrestrial landscape to the stream, using coarse-scale lumped watershed attributes.

We acknowledge that there is a tradeoff between building a cost-efficient regional model and model accuracy. It may not be appropriate to use our model to identify point sources of organic pollution, but it can be used to provide an estimate of regional stream DOC values. Moreover, our model was conservative, meaning that it tended to under predict DOC values. This gives us confidence that a large predicted value with a low prediction variance actually represents an elevated stream DOC concentration.

The Maryland DNR analyzed the MBSS data and generated a statistically valid estimate of DOC levels in stream miles (Boward et al., 1999). However, their analysis did not indicate where stream segments with elevated DOC concentrations were located. We successfully generated predictions for every stream segment in the state and evaluated the predictions using threshold values developed by the Maryland DNR (Figure 4.6). Although we chose to model DOC, which is not currently regulated in Maryland, the same methodology could be applied to regulated constituents, such as pH or nitrate (Peterson et al., in review[2]). The Technical and Regulatory Services Administration within the Maryland Department of the Environment is currently working on a project to modify the USGS National Hydrography Dataset (NHD) (USGS, 2004) to include watershed impairments and stream-use designations by NHD segment (F. Siano, personal communication). The addition of these attributes can be used with the methodology that we presented and will provide a straightforward means of categorizing stream segment predictions into potentially impaired or unimpaired status. We believe that this methodology is an improvement over probability-based inferences because every stream segment in a large area is remotely surveyed in a cost-efficient manner to provide a regional estimate of stream health.

CONCLUSIONS

The geostatistical models that we developed generated more accurate DOC predictions than previous models, but did not appear to fit the data equally well throughout the study area. This raises the question whether it is appropriate to use a single geostatistical model to predict stream DOC throughout Maryland. Blackwater conditions were underrepresented in the observed data and the accuracy of model predictions might be improved if additional survey sites were located in these ecosystems. On the other hand, ecological conditions in eastern and western Maryland are dissimilar and it may be impossible to use a single model to accurately predict DOC. If the latter were true, it would be possible to develop more than one model to obtain accurate predictions of DOC throughout Maryland.

We believe that our methodology has clear advantages related to regional water quality monitoring of regulated water quality constituents, such as nitrate or pH, and that it could be used by states, territories, and tribes to comply with the Clean Water Act more easily. Existing data which has been collected using probability-based random survey designs can be used to develop geostatistical models based on SLD. Therefore, it is not necessary to collect additional field data to generate the preliminary geostatistical model. In addition, inferences about regional stream condition can be generated and this methodology can be used to locate potentially impaired stream segments in a rapid and cost-efficient manner. Each stream segment receives a prediction value with a measure of uncertainty, which allows future field efforts to be concentrated in areas with large amounts of uncertainty or a greater potential for water quality impairment. This ensures that supplementary survey sites are located in areas where additional information about water quality conditions would be most valuable. In addition, resources can be conserved and made available for the TMDL calculation for a specific segment. The model results can also be displayed visually, which allows professionals to communicate results to a wide variety of audiences.

LITERATURE CITED

American Geological Institute (1987) *Glossary of geology, 3rd Edition.* Thomson-Shore, Inc., Alexandria, VA, USA.

Amrhein C. and Suarez, D.L. (1988) The use of a surface complexation model to describe the kinetics of ligand-promoted dissolution of anorthite. *Geochimica et Cosmochimica Acta,* 52, 2785-2793.

Bailey T.C. and Gatrell A.C. (1995) *Interactive spatial data analysis.* Pearson Education Limited, Essex, England.

Bennett P.C. (1991) Quartz dissolution in organic-rich aqueous systems. *Geochimica et Cosmochimica Acta*, 55, 1781-97.

Boward D., Kayzak P., Stranko S., Hurd M., and Prochaska A. (1999) *From the mountains to the sea: The state of Maryland's freshwater streams*. EPA/903/R-99/023, Maryland Department of Natural Resources, Monitoring and Non-tidal Assessment Division, Annapolis, MD, USA.

Byrd R.H., Lu P., Nocedal J.l., Zhu C., and Siam J. (1995) A limited memory algorithm for bound constrained optimization. *Scientific Computing,* 16, 1190-1208.

Canham C.D., Pace M.L., Papaik M.J., Primack A.G.B., Roy K.M., Maranger R.J., Curran, R.P. and Spada D.M. (2004) A spatially explicit watershed-scale analysis of dissolved organic carbon in Adirondack lakes. *Ecological Applications*, 14, 839-54.

Copeland, C. (2002) *Water quality: Implementing the Clean Water Act*. IB89102, Congressional Research Service, Resources, Science, and Industry Division.

Creed I.F., Sanford S.E., Beall F.D., Molot L.A., and Dillon P.J. (2003) Cryptic wetlands: Integrating hidden wetlands in regression models of the export of dissolved organic carbon from forested landscapes. *Hydrological Processes,* 17, 3629-48.

Cressie N. (1993) *Statistics for spatial data, Revised Edition*. John Wiley and Sons, New York, NY, USA.

Currie W.S. and Aber J.D. (1997) Model leaching as a decomposition process in humid montane forests. *Ecology*, 78, 1844-60.

Driscoll C.T., Blette V., Yan C., Schofield C.L., Munson R., and Holsapple J. (1995) The role of dissolved organic carbon in the chemistry and bioavailability of mercury in remote Adirondack lakes. *Water, Air, and Soil Pollution*, 80, 499-508.

Earth System Science Center (ESSC) (1998) *Soil information for environmental modeling and ecosystem management*. ESSC, Pennsylvania State University, Web page, Accessed November 5, 2004, http://www.essc.psu.edu/soil_info/ index.cgi? soil_data&index.html.

Eckhardt B.W. and Moore T.R. (1990) Controls on dissolved organic carbon concentrations in streams, Southern Quebec. *Canadian Journal of Fisheries and Aquatic Sciences*, 47, 1537-1544.

Environmental Systems Research Institute (ESRI) (2002) *ArcGIS version 8.3*. Redlands, CA.

Furnival G. and Wilson R. (1974) Regression by leaps and bounds. *Technometrics,* 16, 499-511.

Gardner B. and Sullivan P.J. (2003) Spatial and temporal stream temperature prediction: Modeling nonstationary temporal covariance structures. *Water Resources Research*, 40, Art. No. W01102.

Gergel S.E., Turner M.G., and Kratz T.K. (1999) Dissolved organic carbon as an indicator of the scale of watershed influence on lakes and rivers. *Ecological Applications*, 9, 1377-90.

Haneberg W. C. (2005) The ins and outs of airborne lidar: An introduction for practicing engineering geologists. *Association of Engineering Geologists News* 48, 16-19.

Hejzlar J., Dubrovsky M., Buchtele J., and Ruzicka M. (2003) The apparent and potential effects of climate change on the inferred concentration of dissolved organic matter in a temperate stream (the Malse River, South Bohemia). *The Science of the Total Environment*, 310, 143-52.

Helsel D.R. and Hirsch R.M. (1992) *Statistical methods in water resources.* Elsevier Science Publishing Co., New York, NY, USA.

Herlihy A.T., Stoddard J.L., and Johnson C.B. (1998) The relationship between stream chemistry and watershed land cover data in the Mid-Atlantic region, US. *Water, Air, and Soil Pollution*, 105, 377-86.

Herlihy A.T., Larsen D.P., Paulsen S.G., Urquhart N.S., and Rosenbaum B.J. (2000) Designing a spatially balanced, randomized site selection process for regional stream surveys: The EMAP Mid-Atlantic Pilot Study. *Environmental Monitoring and Assessment*, 63, 95-113.

Hoeting J.A., Davis R.A., and Merton A.A. (in press) Model selection for geostatistical models. *Ecological Applications.*

Houle D., Carignan R., Lachance M., and Dupont J. (1995) Dissolved organic carbon and sulfur in Southwestern Quebec lakes: Relationships with catchment and lake properties. *Limnology and Oceanography*, 40, 710-717.

Ihaka R, and Gentleman R. (1996) R: A language for data analysis and graphics. *Journal of the Computational and Graphical Statistics*, 5, 299-314.

Kellum B. (2003) Analysis and modeling of acid neutralizing capacity in the Mid-Atlantic Highlands area. MS Thesis, Colorado State University, Fort Collins, CO, USA, 69 p.

Kiffney P.M., Clements W.H., and Cady T.A. (1997) Influence of ultraviolet radiation on the colonization dynamics of a Rocky Mountain stream benthic community. *Journal of the North American Benthological Society*, 16, 520-530.

Kortelainen P. (1993) Content of total organic carbon in Finnish lakes and its relationship to catchment characteristics. *Canadian Journal of Fisheries and Aquatic Sciences,* 50, 1477-83.

Maryland Department of Natural Resources (MDDNR) (1999) *State of the streams: 1995-1997 Maryland Biological Stream Survey results.* Maryland Department of Natural Resources, Annapolis, MD, USA.

Mercurio G., Chaillou J.C., and Roth N.E. (1999) *Guide to using 1995-1997 Maryland Biological Stream Survey data*, Maryland Department of Natural Resources, Annapolis, MD, USA.

Mulholland P.J. (2003) Large-scale patterns in dissolved organic carbon concentration, flux, and sources. in: S.E.G. Findlay, and R.L. Sinsabough (eds.), *Aquatic Ecosystems: Interactivity of Dissolved Organic Matter,* Academic Press, San Diego, CA, USA, pp. 139-157.

Multi-Resolution Land Characteristics Consortium (MRLC) (2003) U.S. Federal Region III land cover data set. Web page, Accessed June 17, 2005, http://www.epa.gov/mrlc/.

Neff J.C. and Asner G.P. (2001) Dissolved organic carbon in terrestrial ecosystems: Synthesis and a model. *Ecosystems*, 4, 29-48.

Olea R.A. (1991) *Geostatistical glossary and multilingual dictionary.* Oxford University Press, New York, NY, USA.

Olsen A.R. and Ivanovich M. (1993) *EMAP monitoring strategy and sampling design.* Office of Research and Development, Environmental Research Laboratory Corvallis (OR): U.S. Environmental Protection Agency, Corvallis, OR, USA. Video.

Omernik J.M. (1987) Ecoregions of the conterminous United States. *Annals of the Association of American Geographers*, 77,118-25.

Ouyang Y. (2003) Simulating dynamic load of naturally occurring TOC from watershed into a river. *Water Research,* 37, 823-32.

Peterson E.E., Theobald D.M., and Ver Hoef J.M. (in review[1]) Support for geostatistical modeling on stream networks: Developing valid covariance matrices based on hydrologic distance and stream flow. *Freshwater Biology.*

Peterson E.E., Merton A.A., Theobald D.M., and Urquhart N.S. (in review[2]) Patterns of spatial autocorrelation in stream water chemistry. *Environmental Monitoring and Assessment.*

Prusha B.A. and Clements W.H. (2004) Landscape attributes, dissolved organic C, and metal bioaccumulation in aquatic macroinvertebrates (Arkansas River Basin, Colorado). *Journal of the North American Benthological Society,* 23, 327-39.

Qualls R.G. and Haines B.L. (1992) Biodegradability of dissolved organic matter in forest throughfall, soil solution, and stream water. *Soil Science Society of American Journal*, 56, 578-86.

Rasmussen J.B., Godbout L., and Schallenberg M. (1989) The humic content of lake water and its relationship to watershed and lake morphometry. *Limnology and Oceanography*, 34, 1336-43.

Spatial Climate Analysis Service (SCAS) (1996) *Parameter-elevation Regressions on Independent Slopes Model (PRISM)*. SCAS, Oregon State University, Web page, Accessed October 23, 2004, http://www.ocs.orst.edu/prism/.

Strahler A.N. (1957) Quantitative analysis of watershed geomorphology. *Transaction of the American Geophysical Union*, 21, 913-20.

Sullivan T.J., Driscoll C.T., Gherini S.A., Munson R.K., Cooks R.B., Charles D.F., and Yatsko C.P. (1989) Influence of aqueous aluminum and organic acids on measurement of acid neutralizing capactiy in surface waters. *Nature*, 338, 408-410.

Theobald D.M., Norman J., Peterson E., and Ferraz S. (2005) Functional linkage of watersheds and streams (FLoWs): Network-based ArcGIS tools to analyze freshwater ecosystems. in: *Proceedings of the ESRI User Conference 2005*, July 25 - 29, 2005, San Diego, CA, USA.

U.S. Environmental Protection Agency (USEPA) (1987) Determination of dissolved organic carbon and dissolved inorganic carbon. in: *Handbook of Methods for Acid Deposition Studies: Laboratory Analysis for Surface Water Chemistry*. EPA 600/4-87/026, U.S. Environmental Protection Agency, Washington, DC, USA.

U.S. Environmental Protection Agency (USEPA) (2001) *Survey designs for sampling surface water condition in the West.* EPA A620/R-01/004c, U.S. Environmental Protection Agency, Office of Research and Development, Washington, DC.

U.S. Environmental Protection Agency (USEPA) (2005) *Level III Ecoregions.* Web page, Accessed August 13, 2003, http://www.epa.gov/wed/pages/ecoregions/ level_iii.htm .

U.S. Geological Survey (USGS) (2003) *National Elevation Dataset.* Web page, Accessed January 4, 2005, Available at http://ned.usgs.gov/.

U.S. Geological Survey (USGS) (2004) *National Hydrography Dataset.* Web page, Accessed January 27, 2005, Available at http://nhd.usgs.gov/ .

U.S. Geological Survey (USGS) (2005) *Land Use Land Cover (LULC).* Web page, Accessed May 25, 2005, Available at http://edc.usgs.gov/products /landcover/lulc.html.

Ver Hoef J.M., Peterson E.E., and Theobald D.M. (in press) Some new spatial statistical models for stream networks', *Environmental and Ecological Statistics.*

Vogelmann J.E., Howard S.M., Yang L., Larson C.R., Wylie B.K., and Van Driel N. (2001) Completion of the 1990's National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. *Photogrammetric Engineering and Remote Sensing*, 67, 650-652.

Wetzel R.G. (1992) Gradient-dominated ecosystems: Sources and regulatory functions of dissolved organic matter in freshwater ecosystems. *Hydrobiologia*, 229, 181-98.

Williamson C.E., Stemberger R.S., Morris D.P., Frost T.M., and Paulsen S.G. (1996) Ultraviolet radiation in North American lakes: Attenuation estimates from DOC measurements and implications for plankton communities. *Limnology and Oceanography*, 41, 1024-34.

Yuan L.L. (2004) Using spatial interpolation to estimate stressor levels in unsampled streams. *Environmental Monitoring and Assessment*, 94, 23-38.

Table 4.1. Potential waterhshed covariates. Potential covariates include lumped watershed attributes provided in the Maryland Biological Stream Survey (MBSS) dataset and lumped watershed attributes that we calculated in a GIS. Potential covariates that demonstrated collinearity with other covariates are not included in this table. The scale and source of the data are included.

Covariate	Description	Spatial Resolution	Source
AREA	Catchment area (ha)	30 meter	USGS, 2003
URBAN	% Urban	30 meter	MRLC Consortium, 2003
BARREN	% Barren landcover	30 meter	MRLC Consortium, 2003
WATER	% Open water	30 meter	MRLC Consortium, 2003
CONIFER	% Conifer or evergreen forest type	30 meter	MRLC Consortium, 2003
DECIDFOR	% Deciduous forest type	30 meter	MRLC Consortium, 2003
MIXEDFOR	% Mixed forest type	30 meter	MRLC Consortium, 2003
EMERGWET	% Emergent herbacious wetlands	30 meter	MRLC Consortium, 2003
WOODYWET	% Woody or shrubby wetlands	30 meter	MRLC Consortium, 2003
COALMINE	% Coalmine land use	30 meter	MRLC Consortium, 2003
EASTING	Easting - Albers Equal Area Conic	1 foot	Projected from MBSS
NORTHING	Northing - Albers Equal Area Conic	1 foot	Projected from MBSS
ER63-ER69	Omernik's Level III Ecoregion	1:7,500,000	USEPA, 2005

Table 4.1. continued

Covariate	Description	Spatial Resolution	Source
MEANELEV	Mean elevation in the watershed	30 meter	USGS, 2003
SLOPE	Mean slope in the watershed	30 meter	USGS, 2003
ARGPERC	% Argillaceous rock type in watershed	1:250,000	A. Herlihy, pers. comm.
CARPERC	% Carbonic rock type in watershed	1:250,000	A. Herlihy, pers. comm.
FELPERC	% Felsic rock type in watershed	1:250,000	A. Herlihy, pers. comm.
MAFPERC	% Mafic rock type in watershed	1:250,000	A. Herlihy, pers. comm.
SILPERC	% Siliceous rock type in watershed	1:250,000	A. Herlihy, pers. comm.
MEANK	Mean soil erodability factor in watershed (adjusted for rock fragments)	1 kilometer	ESSC, 1998
MAXTEMP	Mean annual maximum temperature (°C)	4 kilometer	SCAS, 1996
MINTEMP	Mean minimum temperature for	4 kilometer	SCAS. 1996
PRECIP	Mean precipitation for January - April (mm)	4 kilometer	SCAS, 1996
ANPRECIP	Mean annual precipitation	4 kilometer	SCAS, 1996

Table 4.2. Summary statistics for dissolved organic carbon (DOC) and model covariates.

Variable	Min	1st Quartile	Median	Mean	3rd Quartile	Max	σ2
DOC (mg/l)	0.6	1.2	1.7	1.9	2.7	15.9	1.8
WATER (%)	0	0	0.16	0.25	0.28	4.64	0.44
EMERGWET (%)	0	0	0.13	0.26	0.35	4.85	0.44
WOODYWET (%)	0	0	0.27	1.24	1.15	22.01	3.28
FELPERC (%)	0	0	0.31	26.81	55.26	100	36.14
MINTEMP (°C)	-5.88	-3.06	-2.39	-2.49	-1.4	0.03	1.47

Table 4.3. Model results. Mean square prediction error (MSPE) and r^2 for the "best" model within the straight-line distance (SLD) and weighted asymmetric hydrologic distance (WAHD) model sets.

Autocovariance Model	Distance Measure	MSPE	r^2
Exponential	SLD	0.9394	0.7190
	WAHD	1.2337	0.6368
Spherical	SLD	1.3391	0.6029
	WAHD	1.2187	0.6428
Mariah	SLD	0.9311	0.7221
	WAHD	1.2326	0.6346
Hole Effect	SLD	1.0136	0.6983
Linear with Sill	WAHD	1.2141	0.6388
Rational Quadratic	SLD	0.9447	0.7177

Table 4.4. Cross-validation intervals for straight-line distance Mariah regression coefficients. Regression coefficients represent the change in $\log_{10}$ dissolved organic carbon (mg/l) per unit of X.

Statistic	WATER (%)	EMERGWET (%)	WOODYWET (%)	FELPERC (%)	MINTEMP (°C)
Minimum	0.0469	0.0306	0.0156	-0.0006	0.0616
Maximum	0.0537	0.0425	0.0187	-0.0004	0.071
Mean	0.0501	0.0344	0.0176	-0.0005	0.0655
Standard Dev	0.0007	0.0009	0.0002	0.00005	0.0007
95% Lower Limit	0.0485	0.0322	0.017	-0.0006	0.0643
95% Upper Limit	0.0522	0.0366	0.0179	-0.0005	0.0669

Table 4.5. Unstandarized parameter estimates for the straight-line distance Mariah model that was used to make predictions at unobserved stream segments. The regression coefficients represent the change in $\log_{10}$ dissolved organic carbon (mg/l) per unit of X. The nugget value is the proportion nugget effect.

Nugget	Sill	Range	Intercept	WATER	EMERGWET	WOODYWET	FELPERC	MINTEMP
0.15	0.28	7.02	0.2800	0.0503	0.0366	0.0173	-0.0005	0.0662

Table 4.6. Summary statistics for dissolved organic carbon predictions and prediction variances produced by the straight-line distance Mariah model.

Variable	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Predictions (mg/l)	0.8	1.5	1.9	2.7	3.0	40.4
Prediction Variance $(mg/l)^2$	0.049	0.095	0.122	0.171	0.193	2.597

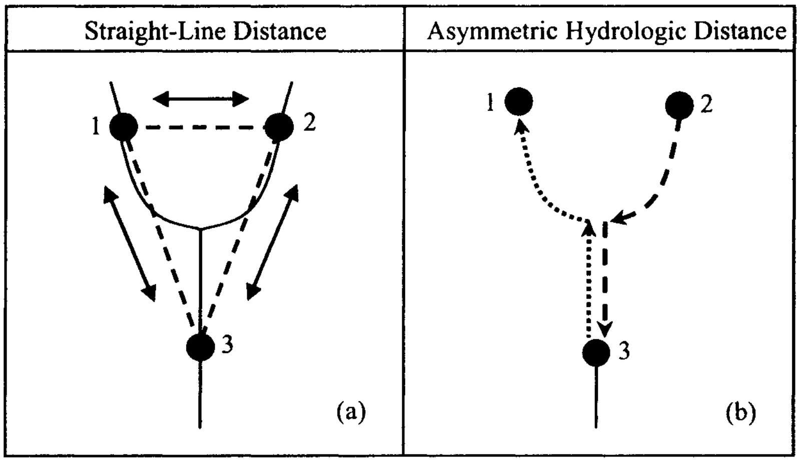


Figure 4.1. Straight-line distance and asymmetric hydrologic distance. The stream network is represented by a solid line while distance measurements are represented with dashed lines. Sites 1, 2, and 3 are all neighbors to one another when straight-line distance (a) is used. Sites 1 and 2 are neighbors to site 3, but not to each other when downstream asymmetric hydrologic distance (b) is used.

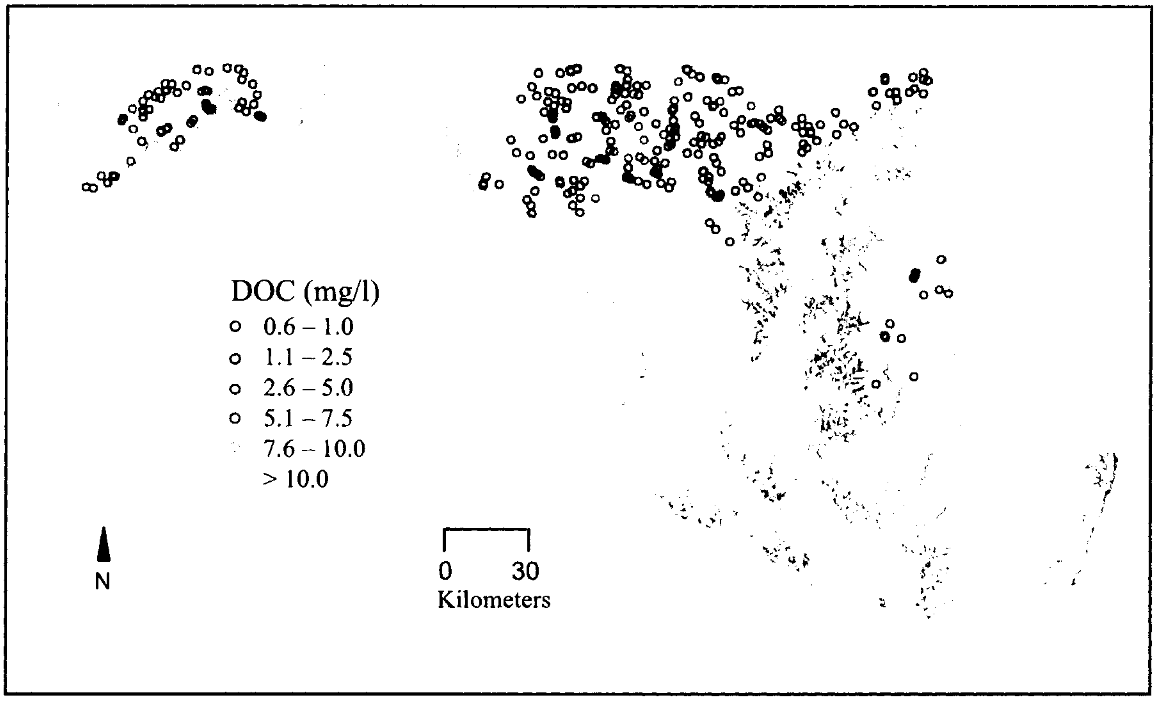
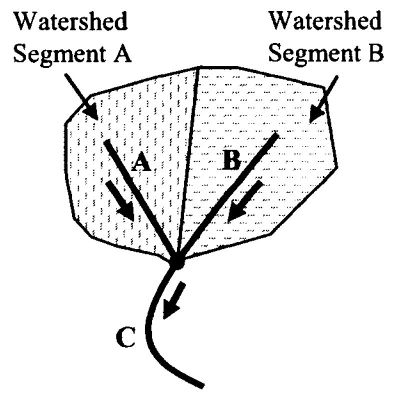


Figure 4.2. The Maryland Biological Stream Survey sites and their dissolved organic carbon (DOC) values in mg/l collected by the Maryland Department of Natural Resources in 1996.



$$\text{Segment PI of A} = \frac{\text{Watershed Area A}}{\text{Watershed Area B}}$$

Figure 4.3. The segment proportional influence (PI) represents a segment's PI on the segment directly downstream. It is calculated by dividing the segment's cumulative watershed area by the total incoming area at its downstream node.

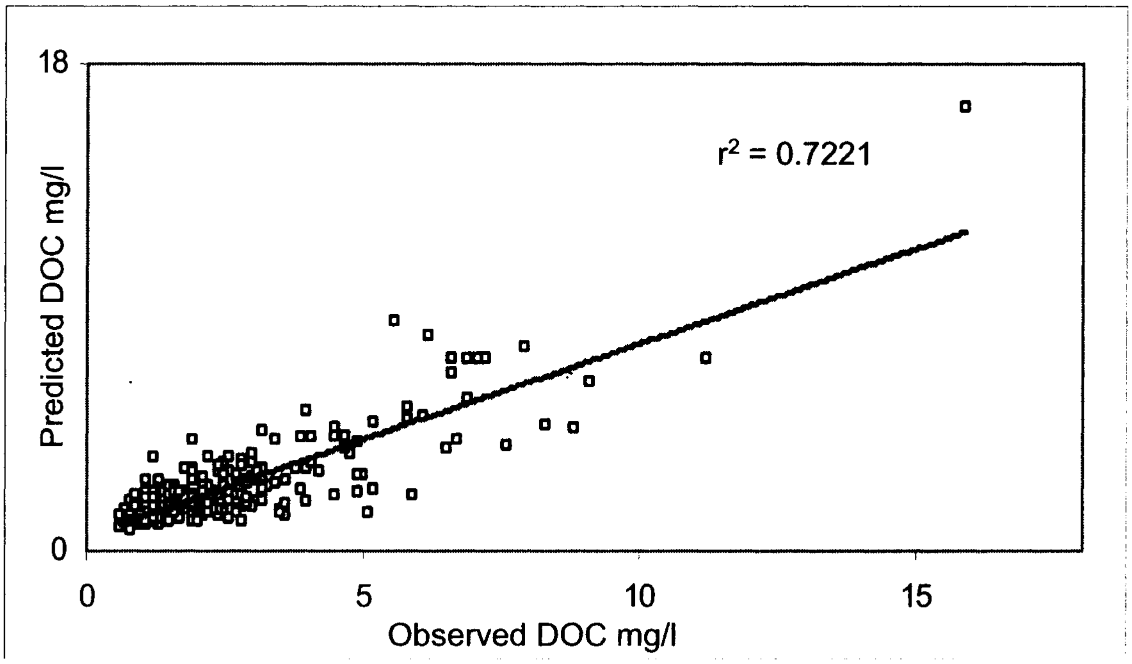


Figure 4.4. r^2 for the observed versus predicted values of dissolved organic carbon (DOC) using the straight-line distance Mariah model.

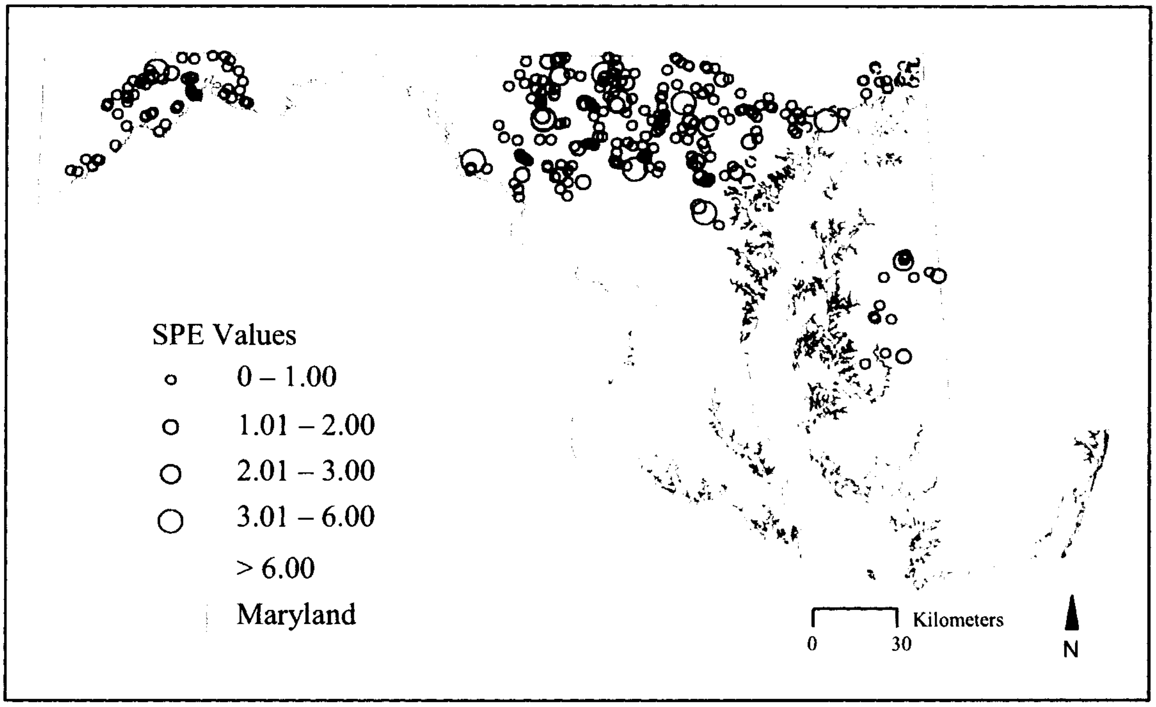
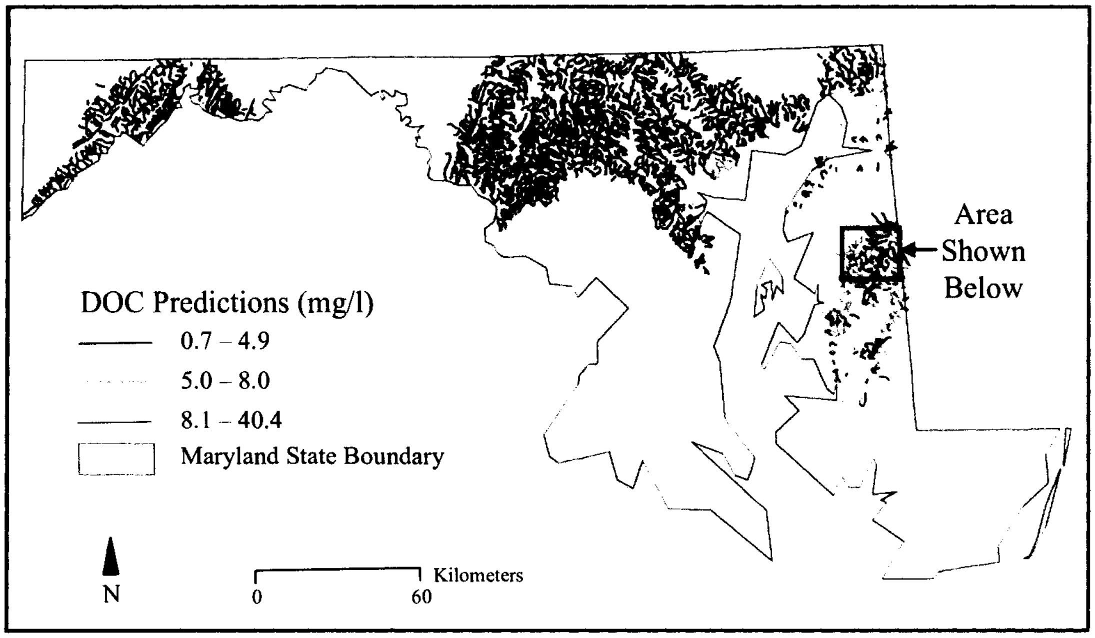


Figure 4.5. Square prediction error (SPE) values for the straight-line distance Mariah model. SPE values are low in western Maryland, but increase in central and eastern Maryland near Chesapeake Bay.



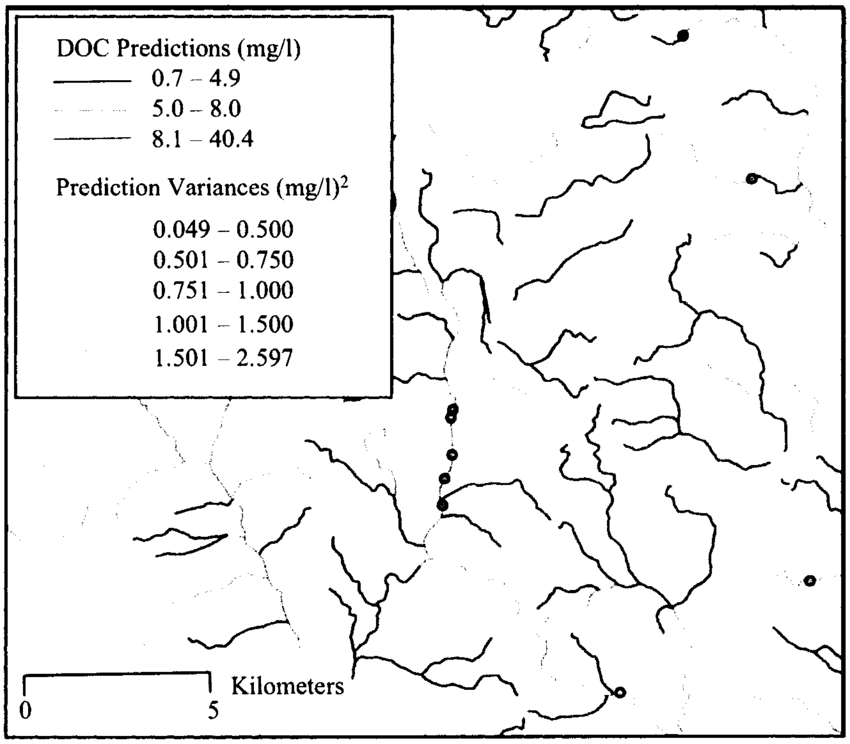


Figure 4.6. Map of the dissolved organic carbon (DOC) predictions and prediction variances for 3083 first, second, and third order non-tidal streams in seven interbasins throughout Maryland sampled in 1996.

CONCLUSIONS

GEOSTATISTICAL MODELING AND IMPLICATIONS FOR WATER QUALITY MONITORING

Spatial autocorrelation in stream water chemistry

My dissertation research clearly demonstrated that spatial autocorrelation exists in stream chemistry data at a relatively coarse scale and that chemical response variables exhibit dissimilar ranges and patterns of spatial autocorrelation. Based on the results of my study, straight-line distance appears to be more suitable for regional geostatistical modeling of conductivity, acid neutralizing capacity, nitrate, sulfate, temperature, dissolved organic carbon, dissolved oxygen, and temperature in Maryland than symmetric hydrologic and weighted asymmetric hydrologic distance when the data are collected using a probability-based random survey design. It may be that coarse-scale patterns reflect the effects of influential ecological processes at the regional scale, which are better represented using straight-line distance. However, it is also possible that there were too few neighboring sites to make full use of the weighted asymmetric hydrologic distance measure. Therefore, further research is needed to assess the capacity of hydrologic distance measures to explain additional variability in stream water chemistry at all scales. This research might include study areas with different spatial extents, such as sub-regions or watersheds. In addition, the spatial configuration of survey sites could be altered to determine whether new survey designs would provide more information about spatial relationships within a stream network.

Geostatistical models improved the accuracy of model predictions for every chemical response variable. However, the accuracy of the geostatistical model predictions varied, which is likely due to the influence of ecological processes acting at different filter scales. For example, conductivity and acid neutralizing capacity are strongly influenced by the weathering of parent material in the watershed and the model predictions were quite accurate. Conversely, the landscape filter scale appears to have a weak influence on temperature and dissolved oxygen, which are strongly influenced by fine filter scales, such as water depth, that were not included in my analysis. These results show that coarse-scale landscape variables can be used to accurately predict chemical characteristics in stream networks that are strongly influenced by broad-scale ecological processes.

Scale of analysis

Streams are clearly influenced by a hierarchy of filter scales, which are not limited to the landscape, segment, and microhabitat scales that I discussed here. The environment is complex and categories are imposed to make environmental analysis manageable. Simplification is necessary, but it inevitably introduces bias since some information must be excluded from the analysis. It is difficult to determine which scale will provide the most information about processes affecting stream water chemistry because different patterns of spatial autocorrelation will likely emerge depending on the scale of observation. In my case, the categories that I selected were based on the availability of GIS data and the processing time required for analysis. In addition, the CWA demands that results are reported at the segment scale. Therefore, it seemed judicious to tailor the analysis and reporting methodology to this scale.

I believe that the segment scale is also an ecologically appropriate scale of analysis. The purpose of the CWA is to maintain the chemical, physical, and biological integrity of surface waters, which suggests that legislators intended to protect overall stream condition. The segment scale is suitable for monitoring because it acts as a functional linkage between the landscape or watershed and the microhabitat filter scales. Information at the segment scale may help managers to make inferences about conditions at the microhabitat scale, allow them to report on the conditions at the segment scale, and to easily extrapolate to gain an estimate of regional water quality conditions.

Geostatistical modeling and regional water quality monitoring

This study demonstrates that geostatistical modeling can improve regional water quality monitoring and I believe that states and tribes can use it to comply with the requirements of the Clean Water Act (CWA) more efficiently. Probability-based random survey designs are currently used to generate an estimate of water quality attainment in stream miles and to meet the requirements of section 305(b) of the CWA. However, probability-based methods do not incorporate space into the water quality inventory. The methodology that I employed can be used to develop a regional estimate of stream condition based on individual segment attainment, but also provides information about the spatial distribution of water quality conditions. In addition, data collected to meet the requirements of section 305(b) can also be used to identify potentially impaired stream segments, which is required in section 303(d) of the CWA.

I fit a geostatistical model to dissolved organic carbon data collected throughout the state of Maryland, which is the reporting unit specified in the CWA. Like most geostatistical modeling attempts, the regional geostatistical model that I developed appears to have violated the stationarity assumption, which requires the mean in the data to be similar across a study area. This violation may have been averted if additional survey sites were added in the eastern portion of Maryland. However, adding survey sites is costly and prevents rapid water quality assessment. Clearly, the reporting units specified by the CWA may be too ecologically diverse to be described using a single geostatistical model. Instead, it may be necessary to subdivide the area to ensure stationarity within the different parts. One option would be to partition the region based upon ecological similarities prior to analysis. However, it is desirable to subdivide the area as little as possible because regional monitoring is most efficient when models are developed for large areas. Therefore, I recommend subdividing a study area following the initial analysis. The preliminary model that I presented shows that areas with a poor model fit also have relatively large prediction variances, which are easy to identify in a GIS. Post analysis subdivision enables managers to find a comfortable balance between the number of models and model accuracy.

Agencies must be conservative about making changes to their monitoring methods because they invest substantial amounts of time and money developing, supporting, and collecting data. Consequently, I wanted this methodology to compliment rather than replace existing methodologies. Existing data collected using probability-based random survey designs can be used to develop geostatistical models based on SLD. Therefore, it is not necessary to collect additional field data to generate the preliminary geostatistical model. In addition, inferences about stream condition can be generated and this methodology can be used to find the geographic location of potentially impaired stream segments in a rapid and cost-efficient manner. GIS also provides a way to visualize the spatial patterns in water quality throughout a region. This allows future survey efforts to be focused in areas where additional information will provide the most benefit. Large prediction variances may indicate that specific ecological conditions are underrepresented in the observed data and that supplemental survey sites are required. Other areas with small prediction variances and large prediction values could be targeted for future field surveys to determine whether water quality standards have truly been violated.

FUTURE RESEARCH

Functional distance measures

The geostatistical models that I developed were based on ecologically representative and statistically reliable modeling techniques that are applicable in stream networks. I believe that future ecological and statistical collaborations should focus on the creation of other ecologically valid distance measures for stream networks. The distance measures that I used were based on physical distances, but other functional distances could incorporate physical characteristics such as flow velocity, stream gradient, or structures such as dams or weirs. Network connectivity could also go beyond simple flow connectivity and include chemical, physical, and biological barriers, such as pH, waterfalls, and predators. These measures would better reflect the energy an organism expends to move from one location to another and would make the potential movement of organisms and material more realistic.

Survey designs for stream networks

Hydrologic distances did not provide additional information about stream chemistry conditions in this study. I believe that this is due in part to the survey design used to collect the data. Probability-based surveys do not ensure that an adequate number of neighbors are selected for an asymmetric hydrologic distance measure. Therefore, new survey designs must be developed to thoroughly examine the ability of functional distance measures to explain variability associated with stream water chemistry. These survey designs will also be needed to test the ability of new functional distance measures because they will likely include additional restrictions on site connectivity. These new survey designs need not include more survey sites. Instead, the configuration of sites must be altered so that both coarse and fine scales of variability within the stream network are captured, while maintaining a random component in the design. For example, a set of survey sites could be randomly selected using a probability-based random survey design. Then, additional satellite sites located on nearby tributaries could be either randomly or non-randomly selected and surveyed to ensure that finer scale hydrologic relationships are captured. Consequently, interdisciplinary research is needed to create survey designs for stream networks that provide the most information at multiple scales with the least amount of financial and physical effort. The models that I developed may be useful because the configuration of neighboring sites can be examined at prediction locations with low mean square prediction errors. These explorations may give insight into which configuration is most favorable.

ADDITIONAL EXPERIENCES, CHALLENGES, AND THOUGHTS

Working in an interdisciplinary environment has been rewarding and I am thankful for the opportunity. The experience has taught me that there are many challenges, but more opportunities, in cross-disciplinary research. This chapter describes some of my experiences, challenges, and thoughts.

TECHNICAL CHALLENGES

Many of the statistical tools described in traditional statistics courses are affected by spatial autocorrelation in the data. When the data contains positive spatial autocorrelation, a single data value contains less information than an independent data value. This may increase the Type 1 errors, which represent the probability that the null hypothesis will be rejected when it is true. Consequently, I was wary of traditional statistical tests such as the t-test, the Wilcoxon Rank Sum test, or confidence intervals, when the data were spatially correlated and I avoided them whenever possible. For example, I calculated cross-validation intervals for the regression coefficients rather than confidence intervals. The models produced in the cross-validation procedure provided a range of regression coefficients that included the affects of spatial autocorrelation. I used these calculate an interval that included 95% of the values. Obviously, it may not always be possible to avoid traditional statistical tests. The key is to recognize that spatial autocorrelation affects these tests and may significantly change the values they produce.

Another challenge that I discovered resulted when I moved data between software packages. I used Microsoft Excel, ArcGIS version 8.3, S-Plus, and R statistical software and each package made changes to the data. For example, blank cells were used to represent 'no data' values in Excel, but were changed to zeros when the file was imported to ArcGIS. This was a challenge since zeros can be true water chemistry values. Importing files that contain text into S-Plus and R can also cause the data fields to shift irregularly if there are spaces in the text. For example, one row may have 12 columns and the next only 11. These issues are not uncommon, but can be especially difficult and time consuming to solve when they are unexpected.

To promote numerical stability, I standardized the response and explanatory variables to have mean zero and unit variance and scaled the distances to fall between zero and one. I calculated the mean square prediction error (MSPE) for each model, but discovered that the model MSPE changed relative to other model MSPE values if they were based on standardized or unstandarized response values. This occurred because the standardized values reduced the contribution of large differences in the predicted and observed values to the MSPE. I was most interested in the true response value on the ground and consequently chose to use the unstandardized response units.

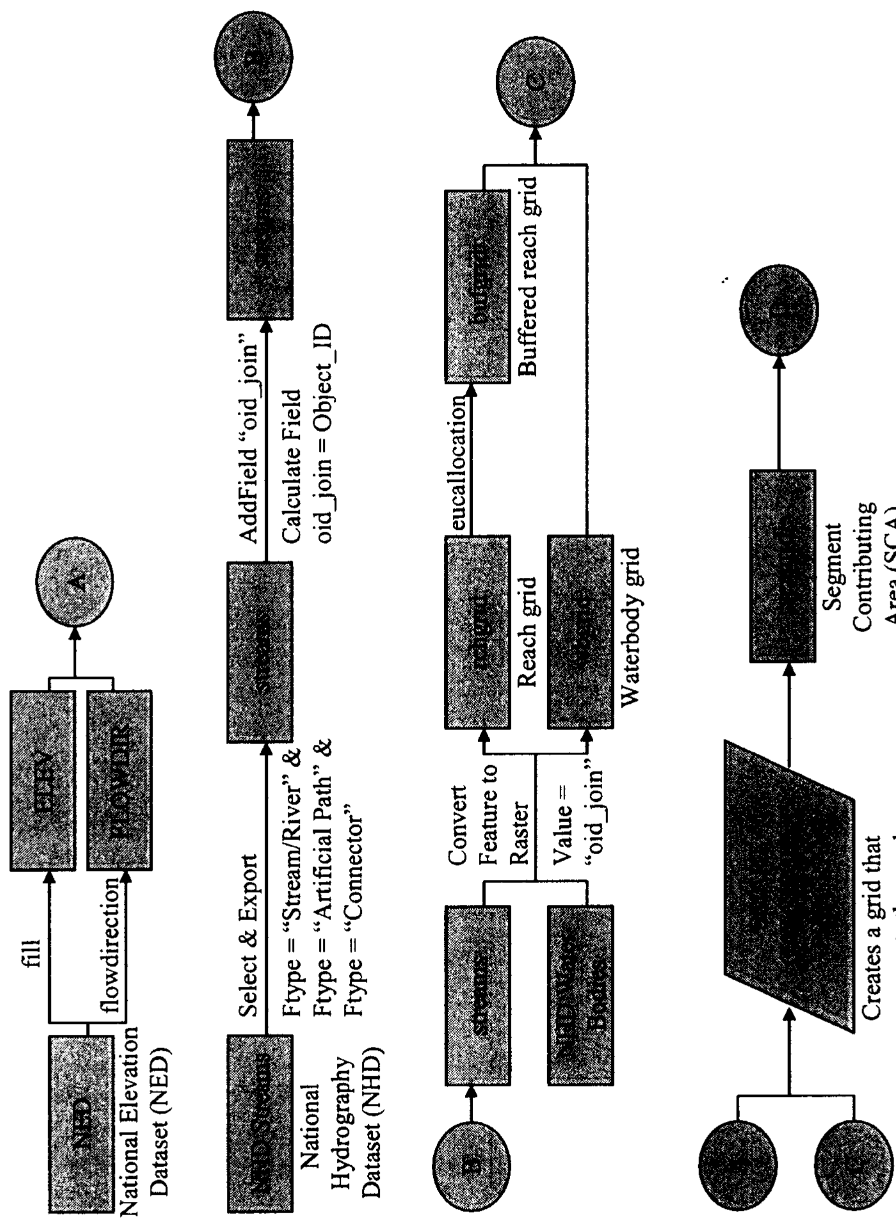
INTERDISCIPLINARY COMMUNICATION

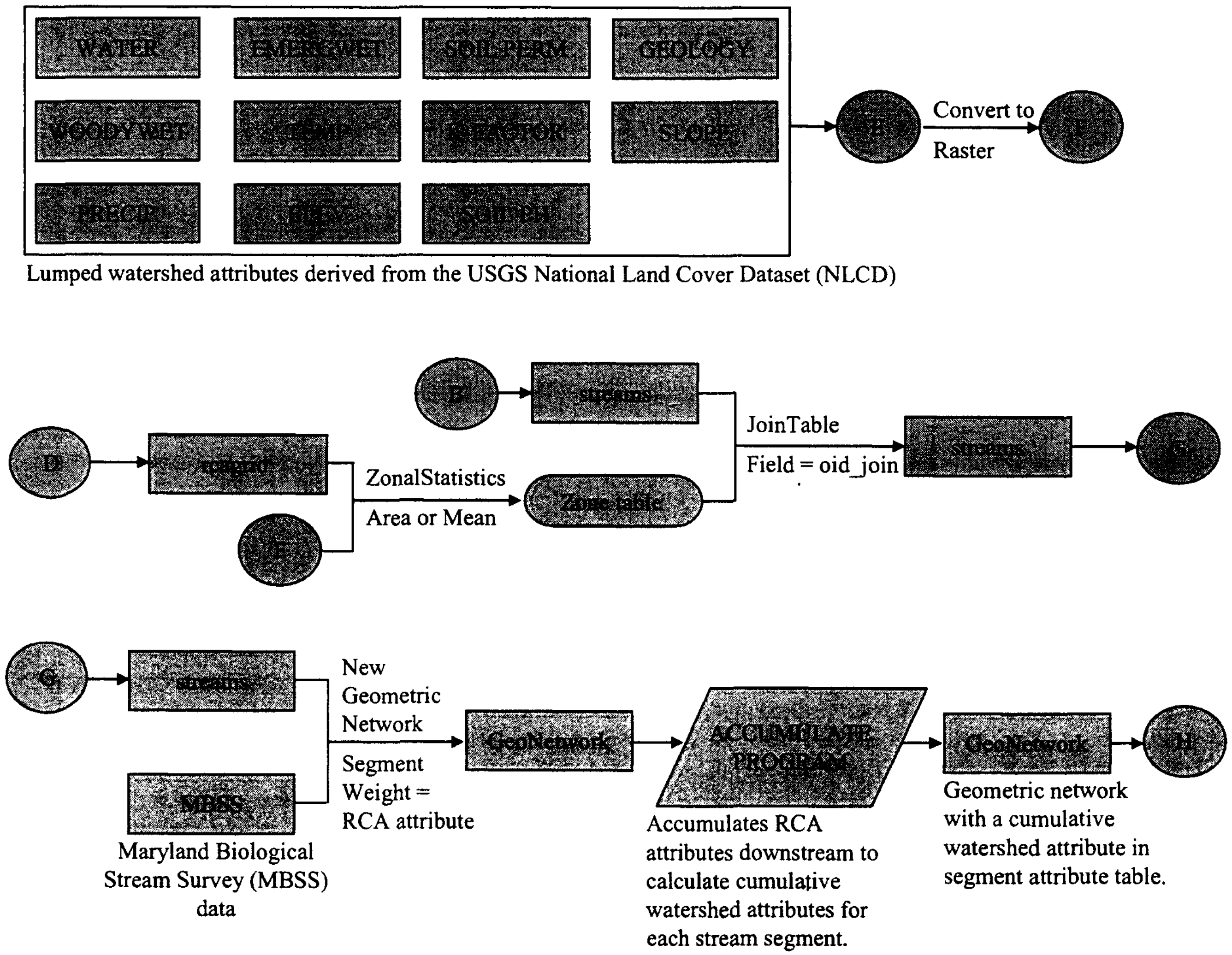
Communicating between disciplines can be difficult because common terminology often has different meanings. For example, a person familiar with geographical information systems would define a lattice as a surface of equally spaced points with a common origin. A statistician would define a lattice as a regular set of square cells made up of rows and columns. These semantic situations occur frequently and are confusing since neither person is typically aware of the difference in terminology. Discovering and working through these situations requires a conscience effort and some patience.

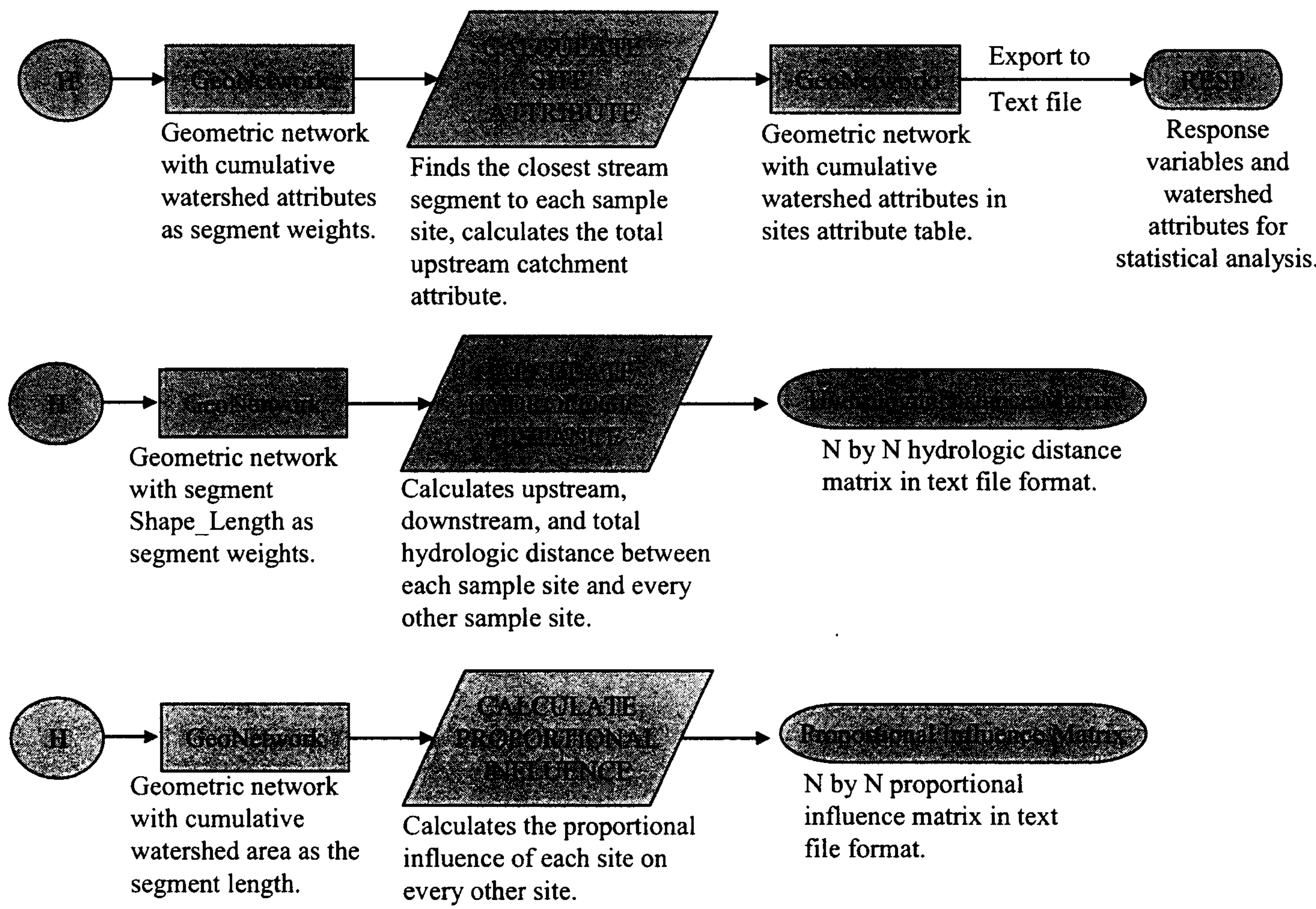
I find interdisciplinary research stimulating because it contributes to the knowledge base in more than one discipline. Scientists to come together from different backgrounds and share their unique perspectives, which often leads to new and exciting research. However, it also requires researchers to relinquish control in some ways because no one can be an expert in every discipline. There may be a large component of the project that each person does not understand. Therefore, I think it is critical to the success of interdisciplinary projects to have respect for the other disciplines, to recognize their value, and to have confidence in the other researcher's abilities.

Interdisciplinary projects have become more common in recent years, but working within a project does not ensure that people will work together. In the past, academia has promoted intellectual ownership and there have not been incentives to share ideas, data, or to work together. Although that philosophy is discouraged in interdisciplinary projects, the attitude seems to be ingrained in scientists. I think that researchers must do more than write an interdisciplinary proposal to overcome these traditions. One solution is to be officially invested in each others projects. This may simply mean that professors from other disciplines are members of a student's committee. It could also be achieved by agreeing to work together and explicitly discussing shared benefits, such as publications. This way, every person involved is invested in the project. They share in the benefits and do not feel that their contribution goes unrewarded.

APPENDIX I: FLOWCHART OF METHODOLOGY FOR SPATIAL ANALYSIS




Convert to
Raster
Lumped watershed attributes derived from the USGS National Land Cover Dataset (NLCD)
ZonalStatistics
Area or Mean
JoinTable
Field = oid_join
New
Geometric
Network
Segment
Weight =
RCA attribute
Maryland Biological
Stream Survey (MBSS)
data
Accumulates RCA
attributes downstream to
calculate cumulative
watershed attributes for
each stream segment.
Geometric network
with a cumulative
watershed attribute in
segment attribute table.


GeoNetwork
CALCULATE
SITE
ATTRIBUTE
Export to
Text file
GeoNetwork
CALCULATE
PROPORTIONAL
INFLUENCE
Geometric network with cumulative watershed attributes as segment weights.
Finds the closest stream segment to each sample site, calculates the total upstream catchment attribute.
Geometric network with cumulative watershed attributes in sites attribute table.
Response variables and watershed attributes for statistical analysis.
Geometric network with segment Shape_Length as segment weights.
Calculates upstream, downstream, and total hydrologic distance between each sample site and every other sample site.
N by N hydrologic distance matrix in text file format.
Geometric network with cumulative watershed area as the segment length.
Calculates the proportional influence of each site on every other site.
N by N proportional influence matrix in text file format.

APPENDIX II: GIS CODE

SNAP POINTS TO LINES PROGRAM

```
'basSnapPointsToLines.bas
'created by Erin Peterson
'program is complete
'last updated 3/4/04


'PURPOSE: The purpose of this program is to snap points to lines. It selects all of the
'points contained in the first layer of the TOC. It starts with a small snap tolerance
'designated by the user and iterates through the points. If the points are within the snap
'tolerance, they are snapped to the line and a "Y" is placed in the "snapped" field of the
'point layer. When the program has iterated through all of the points, it reselects all of the
'points that still have not been snapped. This continues until the stop snapping distance
'(also designated by the user is reached. The user can then examine all of the points that
'were not snapped to lines, delete the ones that are errors, and re-run the program with a
'larger stop snapping distance so that all of the points will be snapped to the lines.
'
'INPUT:  (1) Points feature layer. Add two new fields to the attribute table.
'          Name: snapped
'          Type: Text
'
'          Name: snapdist
'          Type: Long Integer
'
'          Initialize the snapped field to N
'
'         (2) Line feature layer.
'         Additional preprocessing instructions are contained in the
'          SnapPointsToLine_preprocessing.doc
'
'OUTPUT:  The points in the feature layer will be snapped to the lines if they fall within
'the designated snap tolerance. The points attribute table will have the two new fields
'populated. The snapped field represent whether the point's location was changed (Y =
'yes, N= no). The snapdist field shows the snap tolerance value when the point was
'moved.
'
'
'
```

```
'*****************************************************************
'constants must be designated by user
'*****************************************************************

Option Explicit

Const dblSnapTolerance As Double = 5  'snapping tolerance in map units to be with the
'program
Const lngAddSnapTolerance As Long = 5 'amount to add to the snap tolerance at each
'iteration
Const lngStopSnapping As Long = 100   'snap tolerance in map units to end with
Const blnRerunProgram As Boolean = False  're-run status. False means it is the first
'time the program has been run on these points. True means that it has already been run.

'*****************************************************************

Private pApp As IApplication
Private pMxDoc As IMxDocument
Private pMap As IMap
Private pEditor As IEditor
Private pEditLayers As IEditLayers
Private pFeatLines As IFeatureLayer                'line layer
Private pFeatPoints As IFeatureLayer               'points layer
Private pPointTable As ITable                      'attribute table for points layer
Private lngSnappedStateFieldPos As Long            'field position of "snapped"
Private lngSnapDistFieldPos As Long                'field position of "snapdist"
Private pFeatSel As IFeatureSelection              'selection of points
Private pSnapEnv As ISnapEnvironment
Private pFeatSnap As IFeatureSnapAgent
Private lngSnapDistance As Long                    'distance that point was snapped


Private Sub SnapPointsToLine()

Dim pQueryFilter As IQueryFilter
Dim pQueryFilter2 As IQueryFilter
Set pApp = Application
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pFeatLines = pMap.Layer(1)
Set pFeatPoints = pMap.Layer(0)
Set pPointTable = pFeatPoints.FeatureClass
lngSnapDistance = dblSnapTolerance
lngSnappedStateFieldPos = pPointTable.FindField("snapped")
lngSnapDistFieldPos = pPointTable.FindField("snapdist")
```

```
Call GetEditorReference

Set pEditLayers = pEditor

pEditLayers.SetCurrentLayer pFeatPoints, 0

pEditor.StartOperation

'Select all the points in the point layer
Set pFeatSel = pFeatPoints
Set pQueryFilter = New QueryFilter

If blnRerunProgram = True Then
pQueryFilter.WhereClause = "SNAPPED = 'N'"
End If

pFeatSel.SelectFeatures pQueryFilter, esriSelectionResultNew, False

Do Until lngSnapDistance = lngStopSnapping

Dim pEnum As IEnumFeature, pFeat As IFeature, lCount As Long, pNetFeat As
INetworkFeature
Dim pGeom As IGeometry, pExtent As IEnvelope, pNewGeom As IGeometry
Dim dblOID As Double
Dim pRow As IRow
Set pExtent = New Envelope
Set pEnum = pEditor.EditSelection
Set pFeat = pEnum.Next
lCount = 0

'Set up the Snap Environment
Set pSnapEnv = pEditor

'clear all existing snap agents
pSnapEnv.ClearSnapAgents

'create a new feature snap agent
Set pFeatSnap = New FeatureSnap
Set pFeatSnap.FeatureClass = pFeatLines.FeatureClass
pFeatSnap.HitType = esriGeometryPartBoundary
pSnapEnv.SnapTolerance = lngSnapDistance
pSnapEnv.SnapToleranceUnits = esriSnapToleranceMapUnits
pSnapEnv.AddSnapAgent pFeatSnap


'Make sure we have a snapping environment
```

```
If pSnapEnv.SnapAgentCount = 0 Then
MsgBox "You need to turn on at least one snapping agent."
Exit Sub
End If

'Check to make sure you are editing
If pEditor.EditState <> esriStateEditing Then Exit Sub

'Check to make sure something is selected
If pEditor.SelectionCount = 0 Then Exit Sub

While Not pFeat Is Nothing
Set pGeom = pFeat.Shape
Set pNewGeom = SnapFeature(pGeom)

'change the location of the point
If Not pNewGeom Is Nothing Then
pExtent.Union pFeat.Extent
Set pFeat.Shape = pNewGeom
pFeat.Store
pExtent.Union pFeat.Extent
lCount = lCount + 1

'record the results in the point attribute table
dblOID = pFeat.OID
Set pRow = pPointTable.GetRow(dblOID)
pRow.Value(lngSnappedStateFieldPos) = "Y"
pRow.Value(lngSnapDistFieldPos) = lngSnapDistance
pRow.Store

End If

Set pFeat = pEnum.Next
Wend

'add 5 map units to the snap tolerance
lngSnapDistance = lngSnapDistance + lngAddSnapTolerance

'Re-select all the points in the point layer that have not been snapped
Set pFeatSel = pFeatPoints
Set pQueryFilter2 = New QueryFilter
pQueryFilter2.WhereClause = "SNAPPED = 'N'"
pFeatSel.SelectFeatures pQueryFilter2, esriSelectionResultNew, False

Loop
```

```
pEditor.StopOperation " Snapping Points "
pEditor.StopEditing True
Set pEditor = Nothing

End Sub

Private Function SnapFeature(pGeom As IGeometry) As IGeometry
'make a copy of the original point and attempt to snap it to the line. If it moves,
'return the new x,y coordinates from the function.

Dim pPointColl As IPointCollection, pPoint As IPoint, pTempGeom As IGeometry
Dim lLoop As Long, bFlag As Boolean
Dim pSnapEnv As ISnapEnvironment, pClone As IClone, dOrigX As Double, dOrigY
As Double

'Clone the geometry
Set pClone = pGeom
Set pTempGeom = pClone.Clone

'Attempt to snap the feature
Set pSnapEnv = pEditor

If pGeom.GeometryType = esriGeometryPoint Then

Set pPoint = pTempGeom

'save the original x,y coordinates
dOrigX = pPoint.x
dOrigY = pPoint.Y
pSnapEnv.SnapPoint pPoint

'Check to see if the location of the point changed
If Not dOrigX = pPoint.x Or Not dOrigY = pPoint.Y Then
Set SnapFeature = pTempGeom
End If
End If

Exit Function
End Function

Public Sub GetEditorReference()
'creates an editor reference needed to start an editing session
Dim pID As esriCore.UID

Set pID = New esriCore.UID
pID = "esriCore.Editor"
```

```
Set pEditor = Application.FindExtensionByCLSID(pID)
StartEditing pEditor

End Sub

Public Sub StartEditing(pEditor As IEditor)
'starts an editing session

Dim pDataset As IDataset
Dim pDatasetWorkspace As IWorkspace

If Not pEditor.EditState = esriStateNotEditing Then Exit Sub
Set pDataset = pFeatPoints
Set pDatasetWorkspace = pDataset.Workspace
pEditor.StartEditing pDataset.Workspace
End Sub
```

CALCULATE PROPORTIONAL INFLUENCE PROGRAM

```
Attribute VB_Name = "calcProp"
'basCalcProp.bas program
'Created by Erin Peterson
'Program is complete   Last updated: 2/17/04

'PURPOSE: The purpose of this program is to calculate the proportion of influence that
'each segment has on downstream neighbors. This is accomplished by selecting each
'junction in the network, querying the adjacent edges, selecting those edges that flow into
'thejunction, and summing their cumulative upstream catchment area to get the total
'upstream catchment area at each junction. The proportion of influence for each segment
'is calculated by dividing the cumulative upstream area of each incoming segment by the
'total upstream area at each junction. This value is reported in a new field in the edges
'attribute table.

'INPUT: Geometric network containing sample sites. The sample sites should not fall
'along the edges, rather they should be used to split the edges so that they fall at the ends.
'The network should have cumulative upstream catchment area as a weight value.

'OUTPUT: A geometric network with a new field containing the proportion of
'downstream influence that each segment has on downstream neighbors. These values
'will fall between 0 and 1, the sum of proportions for segments flowing into each junction
'will always = 1.

'PREPROCESSING: The basSplitLineWithPoint.bas program should be used to split the
'edges with the the sample points. Then the preprocessing steps for the
```

```
'basSmartCatchment.bas program should be completed (these can be found in the file
'smartcatchment_preprocessing.doc) and then the SmartCatchment.bas program should
'be run. The preprocessing steps for the CatchArea.bas program should be completed
'(contained in the CatchArea_preprocessing.doc file)and the CatchArea.bas script should
'be run. The geometric network should be rebuilt and the cumulative catchment area
'should be included as a network weight. At this point, the network should be ready as
'input into the basCalcProp.bas program.

'******ALL CONSTANTS MUST BE DEFINED BY THE USER

Private Const strTableName As String = "streams_3"  'name of streams feature class
Private Const strmName As String = "streams_3"      'name of streams feature class
                                                     'included in the geometric network
Private Const fDatasetName As String = "fdataset" 'name of feature dataset that contains
                                                     'geometric network
Private Const geodatabasePathName As String = "pathname here"
                    'pathname to the geodatabase that contains the geometric network
Private Const strTableWorkspace As String = "pathname here" 'workspace where all
                                                     'tables will be saved
Private Const strPropFieldName As String = "prop_infl" 'name of new field in network
                                                     'table for cumulative upstream HCA area
Private Const lngSaveEdits As Long = 500    'number of edits completed before the
                                             'program saves them

'*********************************************************************

Option Explicit

Private pApp As IApplication
Private pMxDoc As IMxDocument
Private pMap As IMap
Private pActiveView As IActiveView
Private pWorkspaceFactory As IWorkspaceFactory
Private pWorkspace As IWorkspace
Private pFeatureWorkspace As IFeatureWorkspace
Private PFDataset As IFeatureDataset
Private pNetworkWorkspace As INetworkWorkspace
Private pFeatureClassContainer As IFeatureClassContainer
Private pUID As New esriCore.UID  'ID for UtilityNetworkAnalysisExt
Private pTraceSolver As INetworkAnalysisExt
Private pEditor As IEditor
Private pLine As IFeatureClass        'stream feature class
Private pJuncSel As IEnumNetEID   'enumeration of all the junctions in the network
Private pEdgeSel As IEnumNetEID  'enumeration of all the edges in the network
Private pGeometricNetwork As IGeometricNetwork  'geometric network loaded into the
                                                 'map frame
```

```
Private pNetwork As INetwork
Private pEdgeTable As ITable          'attribute table for edges in geometric network
Private pSortTable As ITable          'dbf table created to sort edges
Private pAdjacentEdges() As Long      'an array of adjacent edges returned by the
IForwardStar
Private vWtVals() As Variant          'an array of adjacent edge weights returned by the
IForwardStar
Private bRev() As Boolean             'an array containing the direction of the adjacent
                                      'edges returned by the IForwardStar
Private lngCumCatchArea As Long       'variable containing the cumulative upstream
                                      'catchment area for a single edge
Private pNextJuncEID As Long          'next downstream junction EID
Private lngDownstreamEdgeOID As Long          'the downstream edge OID
Private pJuncEID As Long              'variable containing the EID for a junction in the
                                      'network
Private lngPropFieldPos As Long       'field position of the cumulative catchment area
(CumArea_sqM) in the pEdgeTable
Private dblCumCatchArea As Double           'cumulative catchment area
Private pJuncCollection As New Collection  'collection of edges that have a Null value
                                           'for CumArea_sqM
Private lngSaveCount As Long


Sub Main()

Set pApp = New AppRef
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pActiveView = pMxDoc.FocusMap

pApp.Caption = "Setting up workspace."

'New AccessWorkspaceFactory is IWorkspaceFactory for personal geodatabase feature
'classes
Set pWorkspaceFactory = New AccessWorkspaceFactory
Set pWorkspace = pWorkspaceFactory.OpenFromFile(geodatabasePathName, 0)
Set pFeatureWorkspace = pWorkspace  'QI for the IFeatureWorkspace
Set PFDataset = pFeatureWorkspace.OpenFeatureDataset(fDatasetName)
Set pNetworkWorkspace = PFDataset.Workspace

'need feature class container to access feature classes within a feature dataset
Set pFeatureClassContainer = PFDataset
Set pLine = pFeatureClassContainer.ClassByName(strmName)
Set pEdgeTable = OpenTable(strTableName)
Dim pWorkspaceProperties As IWorkspaceProperties
```

```
Dim pWkspProp As IWorkspaceProperty
Set pWorkspaceProperties = pWorkspace

pUID = "esricore.UtilityNetworkAnalysisExt"
Set pTraceSolver = pApp.FindExtensionByCLSID(pUID)

'program will stop if there isn't a geometric network loaded
Set pGeometricNetwork = pTraceSolver.CurrentNetwork
Set pNetwork = pGeometricNetwork.Network

lngSaveCount = 0

'add a new field to the Edges attribute table that will hold the proportional influence
Call AddField

'create an enumeration of all junctions found in the network.
Set pJuncSel = GetCurrentEIDs(esriETJunction)

'Get the editor reference and set pEditor
Call GetEditorReference

Dim z As Long
z = 0
pJuncSel.Reset

'select each junction in the network and find the adjacent edges
For z = 0 To pJuncSel.Count - 1
' Select each junction and then find adjacent edges
pJuncEID = pJuncSel.Next
Call FindUpstreamInfluence(pJuncEID)
Next

pEditor.StopEditing True
End Sub


Private Function OpenTable(TableName As String) As ITable
'function opens a table and returns it as an ITable using the table name as an input

On Error GoTo Errorhandler
Dim pDatasetname As IDatasetName
Dim pName As IName
Dim pWorkspaceName As IWorkspaceName
Set pWorkspaceName = New WorkspaceName
pWorkspaceName.WorkspaceFactoryProgID = "esricore.AccessWorkspaceFactory"
pWorkspaceName.PathName = geodatabasePathName
```

```
'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenTable = pName.Open
Exit Function 'exit to avoid error handler

Errorhandler:
Set OpenTable = Nothing
End Function



Private Function GetCurrentEIDs(eElementType As esriElementType) As IEnumNetEID
'return an enumeration of network EID's for a given element type

pApp.Caption = "GetCurrentEIDs"

Dim pNetwork As INetwork   'holds a reference to the current network
Dim pEnumNetEID As IEnumNetEID   'stores a list of network EID's

Set pNetwork = pGeometricNetwork.Network
Set pEnumNetEID = pNetwork.CreateNetBrowser(eElementType)

'return the enumeration
Set GetCurrentEIDs = pEnumNetEID
End Function


Private Sub FindUpstreamInfluence(pJuncEID As Long)

Dim lngAdjEdgeCount As Long
Dim adjEdges() As Long
Dim blnRevOr() As Boolean
Dim adjWeights() As Variant

'Set up the IForwardStar Object
Dim pNetWeightNothing As INetWeight
Dim pNetArea As INetWeight
Dim pNetSchema As INetSchema
Dim pForwardStar As IForwardStar
Dim pNetElems As INetElements
Dim adjEdgesCount As Long
```

```
Dim NextEdgesCount As Long
Dim adjEdgeEID As Long, AdjEdgeWeight As Variant, bDirection As Boolean
Dim edgeElement As Variant
Dim dblTotalArea As Double
Dim q As Integer
Dim edgeWeight As Double
Dim lngEdgeProp As Double
Dim elementEID, elementClassID As Long, elementOID As Long, elementSub As Long

Set pNetSchema = pNetwork
Set pNetArea = pNetSchema.WeightByName("TotalArea")
Set pNetWeightNothing = Nothing

'In this case, you are only interested in edge weights on your network, so
'you can set the other weights using the NetWeightNothing object (pNetWeight)
Set pForwardStar = pNetwork.CreateForwardStar(True, pNetWeightNothing, pNetArea,
pNetArea, pNetWeightNothing)

pForwardStar.FindAdjacent 0, pJuncEID, adjEdgesCount

'erase the old arrays
Erase adjEdges()
Erase blnRevOr()
Erase adjWeights()

'reallocate memory for the new arrays
ReDim adjEdges(adjEdgesCount) As Long
ReDim blnRevOr(adjEdgesCount) As Boolean
ReDim adjWeights(adjEdgesCount) As Variant

pForwardStar.QueryAdjacentEdges adjEdgesCount, adjEdges(0), blnRevOr(0),
adjWeights(0)

dblTotalArea = 0
q = 0

'look at each adjacent edge, determine whether it flows into the junction, and if so add it's
' upstream cumulative catchment area
For Each edgeElement In adjEdges
'if it is an upstream edge from junction
If blnRevOr(q) = False Then
edgeWeight = adjWeights(q)
dblTotalArea = dblTotalArea + edgeWeight
End If

q = q + 1
```

```
Next

'use the total upstream catchment area calculated in the previous step to calculate the
'proportion of downstream influence for each segment that flows into the junction
If dblTotalArea > 0 Then
q = 0

For Each edgeElement In adjEdges
elementEID = edgeElement

'calculate proportion of catchment area
If blnRevOr(q) = False Then
edgeWeight = adjWeights(q)
lngEdgeProp = edgeWeight / dblTotalArea

If elementEID > 0 Then
Set pNetElems = pTraceSolver.CurrentNetwork.Network
pNetElems.QueryIDs elementEID, esriETEdge, elementClassID, elementOID,
elementSub
'report the results in the edge attribute table

Call RecordResults(elementOID, lngEdgeProp)
End If
End If

q = q + 1
Next

End If
End Sub



Private Sub AddField()
'this subroutine checks the edges feature class in the geometric network to see if there
'is a field for the proportional downstream influence (strPropFieldName) for each
'segment. If so, it deletes it and adds a new field. If not, it just adds the new field.

pApp.Caption = "Create a new field"
Dim pAddField As IField
Dim pFieldEdit As IFieldEdit

'create new field
Set pAddField = New Field
Set pFieldEdit = pAddField
```

```
With pFieldEdit
.AliasName = strPropFieldName
.Type = esriFieldTypeDouble
.Length = 10
.Name = strPropFieldName
.Editable = True
.IsNullable = True
End With

Set pAddField = pFieldEdit
Set pFieldEdit = Nothing

If pEdgeTable.FindField(strPropFieldName) = -1 Then
pEdgeTable.AddField pAddField
Else

Dim pFields As IFields
Dim pDeleteField As IField

Set pFields = pEdgeTable.Fields
lngPropFieldPos = pFields.FindField(strPropFieldName)
Set pDeleteField = pFields.Field(lngPropFieldPos)

pEdgeTable.DeleteField pDeleteField
pEdgeTable.AddField pAddField
End If

lngPropFieldPos = pFields.FindField(strPropFieldName)
Exit Sub

Set pEdgeTable = OpenTable(strTableName)
End Sub


Private Sub RecordResults(intEdgeOID As Long, lngEdgeProp As Double)
'this function records the cumulative catchment attribute in the Edges attribute table

pApp.Caption = "RecordResults"
Dim pRecordRow As IRow

If intEdgeOID > 0 Then
Set pRecordRow = pEdgeTable.GetRow(intEdgeOID)

'record proportional influence value
pEditor.StartOperation
```

```
'add this value to the proportional influence attribute field
pRecordRow.Value(lngPropFieldPos) = lngEdgeProp
pRecordRow.Store

'add one to the save count
lngSaveCount = lngSaveCount + 1
End If

pEditor.StopOperation ("RecordResults")

'if a specified number of edits have been completed, save them
If lngSaveCount = lngSaveEdits Then
pEditor.StopEditing True
StartEditing pEdito
lngSaveCount = 0
End If

'release memory
Set pRecordRow = Nothing
End Sub


Public Sub GetEditorReference()
'creates an editor reference needed to start an editing session
Dim pID As esriCore.UID

Set pID = New esriCore.UID
pID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pID)
StartEditing pEditor
End Sub


Public Sub StartEditing(pEditor As IEditor)
'starts an editing session

Dim pDataset As IDataset
Dim pDatasetWorkspace As IWorkspace

If Not pEditor.EditState = esriStateNotEditing Then Exit Sub

Set pDataset = pLine
Set pDatasetWorkspace = pDataset.Workspace
pEditor.StartEditing pDataset.Workspace
End Sub
```

```
Private Sub Reset()
'releases memory of objects

pEditor.StopEditing True
Set pEditor = Nothing
End Sub
```

ACCUMULATE CATCHMENT ATTRIBUTE PROGRAM

```
'AccumCatchAtt.bas program
'Created by Erin Peterson
'Program is complete   Last updated: 11/15/04
'it works in most network cases, but needs to programmatically address network issues
'such as circular flow. Also, sample sites must lie at the ends of edges because it does not
'interpolate catchment attribute values along the edges.
'
'Purpose: The purpose of this program is to calculate the cumulative upstream
'catchment attribute for each stream segment. There are 3 parts to this
'program: locating 1st order streams and finding adjacent edges, finding adjacent edges
'when there were multiple downstream edges, and locating edges that have not been
'recorded and calculating a cumulative attribute for them. Part 1 starts with a 1st order
'stream and records it's cumulative catchment attribute in the edges attribute table. The
'next downstream edge is selected and the program checks to see if all of the upstream
'edges have had their cumulative upstream catchment attribute calculated and recorded. If
'so, the downstream edge the next downstream edge is selected. When there is more than
'one downstream edge, the program can only move to one downstream edge. The other
'downstream edge EID(s) are saved in a collection of downstream edges. This process of
'recording cumulative attributes and moving downstream continues until the cumulative
'catchment attribute for one of the upstream edges has not been recorded. At this point,
'the program moves on to another 1st order stream. When the program has found every
'first order stream and moved downstream as far as possible, it moves on to Part 2. The
'collection of downstream edge EIDs was created in part one and now the program
'attempts to record the cumulative attributes for each of them. If possible, it will move
'downstream from these edges recording cumultive attributes. In Part 3, a collection
'EdgeEID values that do not have a cumulative catchment attribute recorded
'(MissedEdges) is created. These "missed" edges are selected again and the cumulative
'attribute is calculated and recorded in the edges attribute table. This is a recursive
'process and is not complete until each "missed" edge has a cumulative upstream
'catchment attribute recorded in the edges attribute table.
'
'Input:   A geometric network that contains the local hydrologic contributing area
'(HCA)attribute for each edge (or stream segment) as a weight in the geometric network.
'The weight attribute cannot contain NULL values.
'
'*********************************************************************
```

```
' Go to the CreateMissedEdgeCollection and change the [fieldname] to match the
' strCumAttFieldName in the constants section
'*************************************************************************
'
'
'Output:  A new field in the geometric network (located in the edges attribute table) that
'contains 'the cumulative upstream catchment attribute for each edge (or stream
'segment).
'
'PREPROCESSING: A geometric network that contains the sample sites and streams and
'includes local upstream catchment attribute as a weight. The instructions for pre-
'processing attribute contained in the AccumCatchAtt_preprocessing.doc file. This
'program was designed to use the output from the Smart_catchment.bas program.

'******ALL CONSTANTS MUST BE DEFINED BY THE USER******

Private Const strTableName As String = "streams"  'name of streams feature class
                                                   'included in geometric network
Private Const strmName As String = "streams"       'name of streams feature class
                                                   'included in geometric network
Private Const fDatasetName As String = "fdataset"  'name of feature dataset that contains
                                                   'geometric network
Private Const geodatabasePathName As String = "pathname here"
                  'pathname to the geodatabase that contains the geometric network
Private Const strTableWorkspace As String = "pathname here" 'workspace where all
                                                            'tables will be saved
Private Const strCumAttFieldName As String = "CumSil_m2" 'name of new field in
                        'network table for cumulative upstream HCA attribute
Private Const strWeightName As String = "SilHCA" 'name of weight to accumulate
                                                 'CASE SENSITIVE!!
Private Const blnFirstProgramRun As Boolean = True  'True = first time you have run
                  'the program; False= program has someCumulative attribute
                  'values in the edges attribute table, but some are still empty
Private Const lngSaveEdits As Long = 500  'number of edits completed before the
                                          'program saves them

'*************************************************************************
Option Explicit

Private pApp As IApplication
Private pMxDoc As IMxDocument
Private pMap As IMap
Private pActiveView As IActiveView
Private pWorkspaceFactory As IWorkspaceFactory
Private pWorkspace As IWorkspace
Private pFeatureWorkspace As IFeatureWorkspace
```

```
Private pFDataset As IFeatureDataset
Private pNetworkWorkspace As INetworkWorkspace
Private pFeatureClassContainer As IFeatureClassContainer
Private pUID As New esriCore.UID          'ID for UtilityNetworkAnalysisExt
Private pTraceSolver As INetworkAnalysisExt
Private pEditor As IEditor
Private pLine As IFeatureClass        'stream feature class
Private pJuncSel As IEnumNetEID  'enumeration of all the junctions in the network
Private pEdgeSel As IEnumNetEID  'enumeration of all the edges in the network
Private pGeometricNetwork As IGeometricNetwork  'geometric network loaded into the
                                                 'map frame
Private pNetwork As INetwork
Private pEdgeTable As ITable          'attribute table for edges in geometric network
Private pSortTable As ITable          'dbf table created to sort edges
Private pAdjacentEdges() As Long   'an array of adjacent edges returned by the
IForwardStar
Private vWtVals() As Variant          'an array of adjacent edge weights returned by the
                                      'IForwardStar
Private bRev() As Boolean             'an array containing the direction of the adjacent
                                      'edges returned by the IForwardStar
Private lngFirstOrderEdges() As Long
Private vFirstOrderWtVals() As Variant
Private bFirstOrderRev() As Boolean
Private lngCumCatchAtt As Long    'variable containing the cumulative upstream
catchment attribute for a single edge
Private pNextJuncEID As Long       'next downstream junction EID
Private lngDownstreamEdgeEID As Long   'the downstream edge OID
Private pJuncEID As Long      'variable containing the EID for a junction in the network
Private lngCumAttFieldPos As Long         'field position of the cumulative catchment
                                      'attribute (strCumAttFieldName) in the pEdgeTable
Private dblCumCatchAtt As Double          'cumulative catchment attribute
Private MissedEdges As New Collection     'collection of edges that have a Null value
                                          'for strCumAttFieldName
Private downstreamEIDs As New Collection
Private pCursor As ICursor
Private pNetElems As INetElements
Private lngSaveCount As Long, lngEdgeEID As Long

Sub Main()

Set pApp = New AppRef
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pActiveView = pMxDoc.FocusMap
Set downstreamEIDs = Nothing
pApp.Caption = "Setting up workspace."
```

```
Set pWorkspaceFactory = New AccessWorkspaceFactory
Set pWorkspace = pWorkspaceFactory.OpenFromFile(geodatabasePathName, 0)
Set pFeatureWorkspace = pWorkspace 'QI for the IFeatureWorkspace
Set pFDataset = pFeatureWorkspace.OpenFeatureDataset(fDatasetName)
Set pNetworkWorkspace = pFDataset.Workspace

'need feature class container to access feature classes within a feature dataset
Set pFeatureClassContainer = pFDataset
Set pLine = pFeatureClassContainer.ClassByName(strmName)
Set pEdgeTable = OpenTable(strTableName)
pUID = "esricore.UtilityNetworkAnalysisExt"
Set pTraceSolver = pApp.FindExtensionByCLSID(pUID)
Set pNetElems = pTraceSolver.CurrentNetwork.Network

'program will stop if there isn't a geometric network loaded
Set pGeometricNetwork = pTraceSolver.CurrentNetwork
Set pNetwork = pGeometricNetwork.Network
lngSaveCount = 0

If blnFirstProgramRun = True Then
Call AddField
End If

'check to make sure that a cumulative attribute field was added in the AddField
'Subroutine
lngCumAttFieldPos = pEdgeTable.FindField(strCumAttFieldName)
If lngCumAttFieldPos < 0 Then
MsgBox CStr(pLine.AliasName & " does not contain a cumulative attribute field named,
" & strCumAttFieldName)
Exit Sub
End If

'set flow direction of network
Call SetFlowDirection

'create an enumeration of Junction EIDs
Set pJuncSel = GetCurrentEIDs(esriETJunction)

'Set up the IForwardStar Object
Dim pNetAttribute As INetWeight    'this represents HCA attribute for each segment
Dim pNetWeightNothing As INetWeight
Dim pNetSchema As INetSchema
Dim pForwardStar As IForwardStar
Dim pForwardStar2 As IForwardStar
Dim adjEdgesCount As Long, i As Double
Dim NextEdgesCount As Long, intArrayPos As Integer
```

```
Dim juncElement As Variant, intOutEdgeCount As Integer
Dim FirstOrderEID As Long, FirstOrderWeight As Double, FirstOrderDirection As
Boolean
Dim blnMissed As Boolean
Dim elementEID As Long, elementOID As Long, elementClassID As Long, elementSub
As Long
Set pNetSchema = pNetwork
Set pNetAttribute = pNetSchema.WeightByName(strWeightName)
Set pNetWeightNothing = Nothing
blnMissed = False

'Get the editor reference and set pEditor
Call GetEditorReference

'**********************************************************************
'Part 1: Locate 1st order streams and find adjacent edges
'**********************************************************************
pApp.Caption = "1st order streams"

'In this case, since you do not have any junction or turn weights on your network
'you can use the NetWeightNothing object (pNetWeight) for these parameters.
Set pForwardStar = pNetwork.CreateForwardStar(True, pNetWeightNothing,
pNetAttribute,
pNetAttribute, pNetWeightNothing)
pJuncSel.Reset

'look at all of the junctions in the network and find their adjacent edges
For i = 0 To pJuncSel.Count - 1
pJuncEID = pJuncSel.Next
pForwardStar.FindAdjacent 0, pJuncEID, adjEdgesCount

If adjEdgesCount > 0 Then

'the QueryAdjacentEdges function returns arrays - so reallocate memory using
' adjacent edges count returned by FindAdjacent
Erase lngFirstOrderEdges
Erase vFirstOrderWtVals
Erase bFirstOrderRev

ReDim lngFirstOrderEdges(adjEdgesCount - 1) As Long
ReDim vFirstOrderWtVals(adjEdgesCount - 1) As Variant
ReDim bFirstOrderRev(adjEdgesCount - 1) As Boolean

'query for edges adjacent to ToJunc
pForwardStar.QueryAdjacentEdges adjEdgesCount, lngFirstOrderEdges(0),
bFirstOrderRev(0), vFirstOrderWtVals(0)
```

```
'reset the out flowing edges count
intOutEdgeCount = 0

For Each juncElement In bFirstOrderRev
FirstOrderDirection = juncElement

'if this edge flows out of the junction....
If Not FirstOrderDirection = False Then
'add one to the out flowing edges count
intOutEdgeCount = intOutEdgeCount + 1
End If
Next

'If this is a first order stream
If intOutEdgeCount = adjEdgesCount Then
intArrayPos = 0
For Each juncElement In lngFirstOrderEdges
dblCumCatchAtt = vFirstOrderWtVals(intArrayPos)
FirstOrderEID = juncElement

'substitute zero for NULL values
If IsNull(dblCumCatchAtt) Then
dblCumCatchAtt = 0
End If

If FirstOrderEID > 0 Then
'find the OID for the edge so that it can be recorded
pNetElems.QueryIDs FirstOrderEID, esriETEdge, elementClassID, elementOID,
elementSub
Call RecordResults(elementOID, dblCumCatchAtt)

'Find the next adjacent junction and pass it to the FindAdjacentEdges function
Dim ToJuncEID As Long, ToJuncWeight As Variant
ToJuncWeight = 0
Set pForwardStar2 = Nothing
Set pForwardStar2 = pNetwork.CreateForwardStar(True, pNetWeightNothing,
pNetAttribute, pNetAttribute, pNetWeightNothing)
pForwardStar.QueryAdjacentJunction 0, ToJuncEID, ToJuncWeight

'pass the ToJuncEID to FindAdjacentEdges
pNextJuncEID = FindAdjacentEdges(ToJuncEID, blnMissed)

'If the cumulative catchment attribute was recorded for the previous edge and if 'there is
another 'downstream edge, keep moving downstream and calculating/recording
cumulative attribute
While Not pNextJuncEID = -9999
```

```
pNextJuncEID = FindAdjacentEdges(pNextJuncEID, blnMissed)
Wend
End If
intArrayPos = intArrayPos + 1
Next
End If
End If
Next i

'***********************************************************************
'Part 2: Find adjacent edges when there were multiple downstream edges
'***********************************************************************

Dim DSElement As Variant
Dim lngEID As Long
Dim c As Double, g As Double, Index As Double

DSEdge: c = downstreamEIDs.Count
pApp.Caption = "Extra D.S. Edges: " & c
Index = 1

For Each DSElement In downstreamEIDs
lngEID = DSElement
While Not lngEID = -9999
lngEID = FindAdjacentEdges(lngEID, blnMissed)
pApp.Caption = "Extra D.S. Edges: " & downstreamEIDs.Count
If downstreamEIDs.Count > 0 Then
downstreamEIDs.Remove 1
End If
Index = Index + 1
Wend
Next

If downstreamEIDs.Count > 0 Then GoTo DSEdge:

'***********************************************************************
'Part 3: locate edges that have not been recorded and calculate a cumulative attribute for
them
'***********************************************************************
'this function creates a new table of EdgeEID values and their cum catch attributes
'it sorts them according to EdgeEID and creates a collection of EdgeEIDs that do not
'have a cumulative catchment attribute recorded (MissedEdges)

Call CreateMissedEdgeCollection

pApp.Caption = "Filling in missing edges " & MissedEdges.Count
```

```
'fill in cumulative catchment attribute for missed edges
Dim e As Double
Dim lngMissedOID As Long
Dim pNetTopology As INetTopology
Dim lngMissedJuncEID As Long
Dim lngFromEID As Long, lngToEID As Long
Dim lngUpdatedCount As Double, lastMissedCount As Double
Dim MissedElement As Variant, pMissedRow As IRow
Dim lngFCID As Long, lngSubID As Long

Set pNetTopology = pNetwork
lngFCID = pLine.FeatureClassID
lngSubID = -9999

'look through the collection of missed edges
Line1: For Each MissedElement In MissedEdges
blnMissed = True
lastMissedCount = MissedEdges.Count
lngMissedOID = MissedElement
lngEdgeEID = pNetElems.GetEID(lngFCID, lngMissedOID, lngSubID, esriETEdge)

Set pMissedRow = pEdgeTable.GetRow(lngMissedOID)
If IsNull(pMissedRow.Value(lngCumAttFieldPos)) Then

'Get the junctions that are adjacent to the missed edge the from and to node variables are
'switched below because the NHD data is digitized against flow
pNetTopology.GetFromToJunctionEIDs lngEdgeEID, lngToEID, lngFromEID
pNextJuncEID = FindAdjacentEdges(lngFromEID, blnMissed)

'fill in the cum. catchment attribute values for the missed edges
While Not pNextJuncEID = -9999
pNextJuncEID = FindAdjacentEdges(pNextJuncEID, blnMissed)
Wend
End If

pApp.Caption = "Filling in missing edges " & MissedEdges.Count
Next

Set MissedEdges = Nothing
Call CreateMissedEdgeCollection

lngUpdatedCount = MissedEdges.Count

'If the cumulative attribute was not calculated for any of the missed edges or all the edges
'have been recorded exit the program. Otherwise go back and attempt to fill in more
'missed edges
```

```
If lngUpdatedCount = lastMissedCount Then GoTo Line2
If lngUpdatedCount = 0 Then GoTo Line2 Else GoTo Line1

pEditor.StopEditing True
Line2: Call Reset
pApp.Caption = "Program is complete!"
End Sub


Private Sub SetFlowDirection()
'This subroutine sets the flow direction in the geometric network against the digitized
'direction.

Dim pUtilityNetwork As IUtilityNetwork    'necessary to get or set flow direction
Dim lngEdgeEIDCount As Long         'stores the number of edge EID's in the network
Dim lngEdgeEID As Long              'stores the current edge EID
Dim i As Double                     'loop iterator

pApp.Caption = "SetFlowDirection"

'get a reference to the current network
Set pUtilityNetwork = pGeometricNetwork.Network

'Get the editor reference and set pEditor
Call GetEditorReference

'create an edit operation enabling an undo for this operation
pEditor.StartOperation

'get a list of the EID's for edges in the network
Set pEdgeSel = GetCurrentEIDs(esriETEdge)

'set the flow direction for each edge in the network reset the enumeration
pEdgeSel.Reset
'determine the number of edges in the enumeration
lngEdgeEIDCount = pEdgeSel.Count
For i = 0 To lngEdgeEIDCount - 1
'get the next edge EID
lngEdgeEID = pEdgeSel.Next
'set the flow direction for this edge
pUtilityNetwork.SetFlowDirection lngEdgeEID, esriFDAgainstFlow
Next i

'stop the edit operation and specify the name of this edit operation
pEditor.StopOperation "Set Flow Direction"
pEditor.StopEditing True
```

```
End Sub


Public Sub GetEditorReference()
'an editor reference is needed to start an editing session
Dim pID As esriCore.UID
Set pID = New esriCore.UID
pID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pID)
StartEditing pEditor
End Sub


Public Sub StartEditing(pEditor As IEditor)
Dim pDataset As IDataset
Dim pDatasetWorkspace As IWorkspace

If Not pEditor.EditState = esriStateNotEditing Then Exit Sub
Set pDataset = pLine
Set pDatasetWorkspace = pDataset.Workspace
pEditor.StartEditing pDataset.Workspace
End Sub


Private Function GetCurrentEIDs(eElementType As esriElementType) As IEnumNetEID
'return an enumeration of network EID's for a given element type

'pApp.Caption = "GetCurrentEIDs"

Dim pNetwork As INetwork   'holds a reference to the current network
Dim pEnumNetEID As IEnumNetEID   'stores a list of network EID's
Set pNetwork = pGeometricNetwork.Network
Set pEnumNetEID = pNetwork.CreateNetBrowser(eElementType)

'return the enumeration
Set GetCurrentEIDs = pEnumNetEID
End Function


Private Sub RecordResults(intEdgeOID As Long, lngCatchAtt As Double)
'this function records the cumulative catchment attribute in the Edges attribute table

Dim pRecordRow As IRow

If intEdgeOID > 0 Then
Set pRecordRow = pEdgeTable.GetRow(intEdgeOID)
```

```
'if the program has been run before, then check the cumulative attribute field before
'recording in it
If blnFirstProgramRun = False Then

'check to see if the cumulative attribute has already been recorded
If IsNull(pRecordRow.Value(lngCumAttFieldPos)) Then
pApp.Caption = "RecordResults"

'create an edit operation enabling an undo for this operation
pEditor.StartOperation

'add this value to the cumulative upstream catchment attribute field
pRecordRow.Value(lngCumAttFieldPos) = lngCatchAtt
pRecordRow.Store

'add one to the count
lngSaveCount = lngSaveCount + 1
End If

'if this is the first time the program has been run, go ahead and record cum. attribute value
'w/out checking the field

Else
pEditor.StartOperation

'add this value to the cumulative upstream catchment attribute field
pRecordRow.Value(lngCumAttFieldPos) = lngCatchAtt
pRecordRow.Store

'add one to the save count
lngSaveCount = lngSaveCount + 1
End If
End If

pEditor.StopOperation ("RecordResults")

'if a specified number of edits have been completed, save them
If lngSaveCount = lngSaveEdits Then
pEditor.StopEditing True
StartEditing pEditor
lngSaveCount = 0
End If

'release memory
Set pRecordRow = Nothing
```

```
End Sub


Private Sub Reset()

pApp.Caption = "Flush memory"
'release the memory

Set pWorkspaceFactory = Nothing
Set pWorkspace = Nothing
Set pFeatureWorkspace = Nothing
Set pFDataset = Nothing
Set pNetworkWorkspace = Nothing
Set pFeatureClassContainer = Nothing
Set pGeometricNetwork = Nothing
Set pMxDoc = Nothing
Set pMap = Nothing
Set pActiveView = Nothing
Set pTraceSolver = Nothing
pEditor.StopEditing True
Set pEditor = Nothing
Set pLine = Nothing
Set pJuncSel = Nothing
Set pEdgeSel = Nothing
Set pGeometricNetwork = Nothing
Set pNetwork = Nothing
Set pEdgeTable = Nothing
Set pSortTable = Nothing
Set MissedEdges = Nothing
Set downstreamEIDs = Nothing

lngCumCatchAtt = 0          'variable containing the cumulative upstream catchment
                            'attribute for a single edge
pNextJuncEID = 0            'next downstream junction EID
lngDownstreamEdgeEID = 0 'the downstream edge OID
pJuncEID = 0                'variable containing the EID for a junction in the network
lngCumAttFieldPos = 0
dblCumCatchAtt = 0

End Sub


Private Function FindAdjacentEdges(ToJuncEID As Long, blnMissed As Boolean) As
Long

'find downstream junction
```

```
Dim pNetAttribute As INetWeight            'Catchment attribute
Dim pNetWeightNothing As INetWeight
Dim pNetSchema As INetSchema
Dim pPassJuncEID As Long
Dim elementOID As Long, elementClassID As Long, elementSub As Long
Dim dsEID As Long

Set pNetSchema = pNetwork
Set pNetAttribute = pNetSchema.WeightByName(strWeightName)
Set pNetWeightNothing = Nothing

Dim pForwardStar3 As IForwardStar
Dim pForwardStar4 As IForwardStar
Dim MultipleDSEdgeEIDs() As Long
Dim LocalAttArray() As Variant

'find the adjacent edges from To junction
Dim adjEdgesCount As Long
Set pForwardStar3 = pNetwork.CreateForwardStar(True, pNetWeightNothing,
pNetAttribute,
pNetAttribute, pNetWeightNothing)
pForwardStar3.FindAdjacent 0, ToJuncEID, adjEdgesCount

'the QueryAdjacentEdges function returns arrays - so reallocate memory using
'adjacent edges count returned by FindAdjacent
Erase pAdjacentEdges
Erase vWtVals
Erase bRev

ReDim pAdjacentEdges(adjEdgesCount - 1) As Long
ReDim vWtVals(adjEdgesCount - 1) As Variant
ReDim bRev(adjEdgesCount - 1) As Boolean

'query for edges adjacent to ToJunc
pForwardStar3.QueryAdjacentEdges adjEdgesCount, pAdjacentEdges(0), bRev(0),
vWtVals(0)

Dim lngUpstreamEdgeEID As Long
Dim dblUpstreamEdgeWeight As Double
Dim dblDownstreamEdgeWeight As Double
Dim j As Double
Dim foundCount As Double, downstreamEdgeCount As Integer
Dim pRow As IRow

dblCumCatchAtt = 0
j = 0
```

```
foundCount = 0
downstreamEdgeCount = 0

lngDownstreamEdgeEID = 0
dblDownstreamEdgeWeight = 0

'look at each of the edges which are adjacent to the junction
For j = 0 To adjEdgesCount - 1
lngUpstreamEdgeEID = 0
dblUpstreamEdgeWeight = 0

'If the edge is Upstream then......
If bRev(j) = 0 Then
lngUpstreamEdgeEID = pAdjacentEdges(j)

If lngUpstreamEdgeEID > 0 Then
pNetElems.QueryIDs lngUpstreamEdgeEID, esriETEdge, elementClassID, elementOID,
elementSub

Set pRow = pEdgeTable.GetRow(elementOID)

'check to see if the cumulative HCA attribute has been recorded
If Not IsNull(pRow.Value(lngCumAttFieldPos)) Then

'add this value to the cumulative upstream catchment attribute
dblCumCatchAtt = dblCumCatchAtt + pRow.Value(lngCumAttFieldPos)
foundCount = foundCount + 1

'if cum HCA attribute hasn't been recorded, exit function
pPassJuncEID = -9999
FindAdjacentEdges = pPassJuncEID
Exit Function
End If
End If

Set pRow = Nothing

'the edge is downstream
Else
'If filling in a missed edges - only record the missed edge
If blnMissed = True Then
downstreamEdgeCount = downstreamEdgeCount + 1
foundCount = foundCount + 1

If pAdjacentEdges(j) = lngEdgeEID Then
lngDownstreamEdgeEID = pAdjacentEdges(j)
```

```
dblDownstreamEdgeWeight = vWtVals(j)
End If

Else
downstreamEdgeCount = downstreamEdgeCount + 1
If downstreamEdgeCount > 1 Then

'create an array of network element EIDs for downstream edges
ReDim MultipleDSEdgeEIDs(adjEdgesCount - 1) As Long
ReDim LocalAttArray(adjEdgesCount - 1) As Variant

MultipleDSEdgeEIDs(0) = lngDownstreamEdgeEID
MultipleDSEdgeEIDs(downstreamEdgeCount - 1) = pAdjacentEdges(j)
LocalAttArray(0) = dblDownstreamEdgeWeight
LocalAttArray(downstreamEdgeCount - 1) = vWtVals(j)
foundCount = foundCount + 1

Else
'Else the edge is downstream.......
lngDownstreamEdgeEID = pAdjacentEdges(j)
dblDownstreamEdgeWeight = vWtVals(j)
foundCount = foundCount + 1
End If
End If
End If
Next j

Set pRow = Nothing

'pPassJuncEID is the junction id # that is returned by this function (-9999 = nodata)
pPassJuncEID = -9999

Dim MissedDSEdgeCount As Integer
MissedDSEdgeCount = downstreamEdgeCount

If blnMissed = True Then downstreamEdgeCount = 1

'If all of the upstream edges have cum. attribute reported and the downstream edge is
valid then
If foundCount = adjEdgesCount And lngDownstreamEdgeEID > 0 Then

'If there is more than one downstream edge
If downstreamEdgeCount > 1 Then
Dim MDSEdgeElement As Variant
Dim MultCumCatchAtt As Double
Dim ArrayIndex As Long
```

```
ArrayIndex = 0
MultCumCatchAtt = 0

For Each MDSEdgeElement In MultipleDSEdgeEIDs
lngDownstreamEdgeEID = MDSEdgeElement
dblDownstreamEdgeWeight = LocalAttArray(ArrayIndex)
MultCumCatchAtt = (dblCumCatchAtt / downstreamEdgeCount) +
dblDownstreamEdgeWeight

If lngDownstreamEdgeEID > 0 Then
pNetElems.QueryIDs lngDownstreamEdgeEID, esriETEdge, elementClassID,
elementOID, elementSub

Call RecordResults(elementOID, MultCumCatchAtt)
pApp.Caption = "FindAdjacentEdges"

If ArrayIndex = 0 Then
pPassJuncEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
Else
dsEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
downstreamEIDs.Add dsEID
End If

ArrayIndex = ArrayIndex + 1
End If
Next
Else

If blnMissed = True Then
If lngDownstreamEdgeEID > 0 Then
dblCumCatchAtt = (dblCumCatchAtt / MissedDSEdgeCount) +
dblDownstreamEdgeWeight
pNetElems.QueryIDs lngDownstreamEdgeEID, esriETEdge, elementClassID,
elementOID, elementSub

Call RecordResults(elementOID, dblCumCatchAtt)
pPassJuncEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
End If
Else

dblCumCatchAtt = dblCumCatchAtt + dblDownstreamEdgeWeight
If lngDownstreamEdgeEID > 0 Then
pNetElems.QueryIDs lngDownstreamEdgeEID, esriETEdge, elementClassID,
elementOID, elementSub
Call RecordResults(elementOID, dblCumCatchAtt)
pPassJuncEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
```

```
End If
End If
End If
End If

Set pRow = Nothing
FindAdjacentEdges = pPassJuncEID
End Function


Private Sub AddField()
'this subroutine checks the edges feature class in the geometric network to see if there
'is a field for the cumulative HCA attribute (const strCumAttFieldName). If so, it deletes
it and  'adds a new field. If not, it just adds the new field

pApp.Caption = "Create a new field"

Dim pAddField As IField
Dim pFieldEdit As IFieldEdit

'create new field
Set pAddField = New Field
Set pFieldEdit = pAddField

With pFieldEdit
.AliasName = strCumAttFieldName
.Type = esriFieldTypeDouble
.Length = 30
.Name = strCumAttFieldName
.Editable = True
.IsNullable = True
End With

Set pAddField = pFieldEdit
Set pFieldEdit = Nothing

If pEdgeTable.FindField(strCumAttFieldName) = -1 Then
pEdgeTable.AddField pAddField
Else
Dim pFields As IFields
Dim pDeleteField As IField
Set pFields = pEdgeTable.Fields
lngCumAttFieldPos = pFields.FindField(strCumAttFieldName)
Set pDeleteField = pFields.Field(lngCumAttFieldPos)
pEdgeTable.DeleteField pDeleteField
pEdgeTable.AddField pAddField
```

```
End If
Exit Sub
End Sub


Private Function OpenTable(TableName As String) As ITable
'Private Function OpenTable(strTableWorkspace As String, TableName As String) As
ITable
On Error GoTo Errorhandler
Dim pDatasetname As IDatasetName
Dim pName As IName

Dim pWorkspaceName As IWorkspaceName
Set pWorkspaceName = New WorkspaceName
pWorkspaceName.WorkspaceFactoryProgID = "esricore.AccessWorkspaceFactory"
pWorkspaceName.PathName = geodatabasePathName
pApp.Caption = "Open Table"

'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenTable = pName.Open

Exit Function 'exit to avoid error handler
Errorhandler:
Set OpenTable = Nothing
End Function


Private Sub UpdateCollection(lngDownstreamEID As Long)
Dim e As Double
Dim lngMissedOID As Long
Dim mElement As Variant
e = 1
pApp.Caption = "Updating missed edge collection"

For Each mElement In MissedEdges
lngMissedOID = mElement
If lngMissedOID = lngDownstreamEID Then
MissedEdges.Remove e
Exit For
End If
```

```
e = e + 1
Next

pApp.Caption = "done updating me" & MissedEdges.Count
End Sub


Private Function LocateDownstreamJunc(lngEdgeEID As Long) As Long
'uses a junction EID that is passed into the function to find the next downstream junction.
'It returns the downstream junction EID

pApp.Caption = "locate downstream junction"

Dim pNetTopology As INetTopology
Dim lngFromEID As Long, lngToEID As Long
Set pNetTopology = pNetwork

'Get the To Junction for the edge- the from and to node variables are switched below
'because the NHD data is digitized against flow
pNetTopology.GetFromToJunctionEIDs lngEdgeEID, lngToEID, lngFromEID
LocateDownstreamJunc = lngToEID
End Function


Private Sub CreateMissedEdgeCollection()
'creates a collection of edges that do not have their cumulative catchment attribute
'recorded
pApp.Caption = "creating m.e. collection"
Dim pFields As IFields
Dim pField As IField
Dim OIDField As Long
Dim sOIDField As String

pApp.Caption = "Create missed edge collection"

'set up the parameters for the ITableSort
Set pFields = pEdgeTable.Fields
Set pField = pFields.Field(0)
sOIDField = pField.Name
OIDField = pEdgeTable.FindField(pField.Name)
Dim pRow As IRow
Dim pQueryFilter As IQueryFilter
Set pQueryFilter = New QueryFilter
'***********************************************************************
'the [fieldname] must be changed to match the strCumAttFieldName in the constants
'section
```

```
'*******************************************************************
pQueryFilter.WhereClause = "[CumSil_m2]IS NULL"

'create the ITableSort
Dim pTableSort As ITableSort
Set pTableSort = New TableSort
With pTableSort
.Fields = sOIDField
.Ascending(sOIDField) = True
Set .QueryFilter = pQueryFilter
Set .Table = pEdgeTable
End With

pTableSort.Sort Nothing

'use the pTableSort to create a cursor containing the missed edges
Dim pCursor As ICursor
Dim OID As Long
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow

Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow

'use the pCursor populate the missed edges collection
Do While Not pRow Is Nothing
OID = pRow.Value(0)
'add this EdgeOID to a collection
MissedEdges.Add OID
Set pRow = pCursor.NextRow
Loop
pApp.Caption = "done w/ m.e. collection " & MissedEdges.Count
End Sub
```

ACCUMULATE MEAN CATCHMENT ATTRIBUTE PROGRAM

```
'AccumMeanCatchAtt.bas program
'Created by Erin Peterson
'Program is complete   Last updated: 12/31/04
'
'Purpose: The purpose of this program is to calculate a MEAN cumulative upstream
catchment
'attribute for each stream segment. The means are WEIGHTED by cumulative catchment
'area. There are 3 parts to this program: locating 1st order streams and finding adjacent
```

```
'edges, finding adjacent edges when there were multiple downstream edges, and locating
'edges that have not been recorded and calculating a cumulative mean attribute for them.
'Part 1 starts with a 1st order stream and records it's cumulative mean attribute in the
'edges attribute table. The next downstream edge is selected and the program checks to
'see if all of the upstream edges have had their cumulative upstream mean attribute
'calculated and recorded. If so, the downstream edge cumulative mean attribute is
'calculated/recorded and the next downstream edge is selected. When there is more than
'one downstream edge, the program can only move to one downstream edge. The other
'downstream edge EID(s) are saved in a collection of downstream edges. This process of
'recording cumulative mean attributes and moving downstream continues until the
'cumulative mean attribute for one of the upstream edges has not been recorded. At this
'point, the program moves on to another 1st order stream. When the program has found
'every first order stream and moved downstream as far as possible, it moves on to Part 2.
'The collection of downstream edge EIDs was created in part one and now the program
'attempts to record the cumulative mean attributes for each of them. If possible, it will
'move downstream from these edges recording cumultive attributes. In Part 3, a
'collection EdgeEID values that do not have a cumulative mean attribute recorded
'(MissedEdges) is created. These "missed" edges are selected again and the cumulative
'attribute is calculated and recorded in the edges attribute table. This is a recursive
'process and is not complete until each "missed" edge has a cumulative upstream
'catchment attribute recorded in the edges attribute table.
'
'Input:  A geometric network that contains the cumulative area
'attribute for each edge (or stream segment) as a weight in the geometric network. Null
'values should be stored as -9999
'************************************************************************
' Go to the CreateMissedEdgeCollection and change the [fieldname] to match the
' strCumMeanFieldName in the constants section
'************************************************************************
'
'
'Output:  A new field in the geometric network (located in the edges attribute table) that
' contains the cumulative upstream catchment attribute for each edge (or stream segment).
'
'
'PREPROCESSING: A geometric network that contains the sample sites and streams and
'includes cumulative upstream catchment attribute as a weight. The instructions for pre-
'processing attribute contained in the AccumCatchAtt_preprocessing.doc file. This
'program was designed to use the output from the Smart_catchment.bas program.

'******ALL CONSTANTS MUST BE DEFINED BY THE USER************

Private Const strTableName As String = "strm"     'name of streams feature class
                                                  'included in geometric network
Private Const strmName As String = "strm"         'name of streams feature class
                                                  'included in geometric network
```

```
Private Const fDatasetName As String = "testmean" 'name of feature dataset that contains
                                                  'geometric network
Private Const geodatabasePathName As String = "pathname  here"
                 'pathname to the geodatabase that contains the geometric network
Private Const strTableWorkspace As String = "pathname here"
                                      'workspace where all tables will be saved
Private Const strRCAMeanFieldName As String = "slope" 'name of RCA field in
                          'network table for cumulative upstream HCA attribute
Private Const strRCAreaFieldName As String = "HCA_Area"
Private Const strCumMeanFieldName As String = "CumSlope"
Private Const strWeightName As String = "CumArea_m2"    'name of weight to
                                              'accumulate CASE SENSITIVE!!
Private Const blnFirstProgramRun As Boolean = True 'True = first time you have run the
                  'program; False= program has some cumulative attribute values in
                  'the edges attribute table, but 'some are still empty
Private Const lngSaveEdits As Long = 3          'number of edits completed before
                                              'the program saves them

'**********************************************************************
Option Explicit

Private pApp As IApplication
Private pMxDoc As IMxDocument
Private pMap As IMap
Private pActiveView As IActiveView
Private pWorkspaceFactory As IWorkspaceFactory
Private pWorkspace As IWorkspace
Private pFeatureWorkspace As IFeatureWorkspace
Private pFDataset As IFeatureDataset
Private pNetworkWorkspace As INetworkWorkspace
Private pFeatureClassContainer As IFeatureClassContainer
Private pUID As New esriCore.UID  'ID for UtilityNetworkAnalysisExt
Private pTraceSolver As INetworkAnalysisExt
Private pEditor As IEditor
Private pLine As IFeatureClass  'stream feature class
Private pJuncSel As IEnumNetEID 'enumeration of all the junctions in the network
Private pEdgeSel As IEnumNetEID 'enumeration of all the edges in the network
Private pGeometricNetwork As IGeometricNetwork  'geometric network loaded into the
                                              'map frame
Private pNetwork As INetwork
Private pEdgeTable As ITable          'attribute table for edges in geometric network
Private pSortTable As ITable          'dbf table created to sort edges
Private pAdjacentEdges() As Long      'an array of adjacent edges returned by the
                                      'IForwardStar
Private vWtVals() As Variant          'an array of adjacent edge weights returned by the
                                      'IForwardStar
```

```
Private bRev() As Boolean           'an array containing the direction of the adjacent
                                    'edges returned by the IForwardStar
Private lngFirstOrderEdges() As Long
Private vFirstOrderWtVals() As Variant
Private bFirstOrderRev() As Boolean
Private lngCumCatchAtt As Long      'variable containing the cumulative upstream
                                    'catchment attribute for a single edge
Private pNextJuncEID As Long        'next downstream junction EID
Private lngDownstreamEdgeEID As Long   'the downstream edge OID
Private pJuncEID As Long   'variable containing the EID for a junction in the network
Private lngCumMeanFieldPos As Long       'field position of the cumulative catchment
                                    'attribute (strCumAttFieldName) in the pEdgeTable
Private lngRCAMeanFieldPos As Long       'field position of RCA mean attribute
Private lngRCAreaFieldPos As Long        'variable containing the RCA area field
Private dblCumCatchAtt As Double         'cumulative catchment attribute
Private MissedEdges As New Collection    'collection of edges that have a Null value
                                         'for strCumAttFieldName
Private downstreamEIDs As New Collection
Private pCursor As ICursor
Private pNetElems As INetElements
Private lngSaveCount As Long, lngEdgeEID As Long

Sub Main()

Set pApp = New AppRef
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pActiveView = pMxDoc.FocusMap
Set downstreamEIDs = Nothing

pApp.Caption = "Setting up workspace."

'New AccessWorkspaceFactory is IWorkspaceFactory for personal geodatabase feature
classes
Set pWorkspaceFactory = New AccessWorkspaceFactory
Set pWorkspace = pWorkspaceFactory.OpenFromFile(geodatabasePathName, 0)
Set pFeatureWorkspace = pWorkspace        'QI for the IFeatureWorkspace
Set pFDataset = pFeatureWorkspace.OpenFeatureDataset(fDatasetName)
Set pNetworkWorkspace = pFDataset.Workspace

'need feature class container to access feature classes within a feature dataset
Set pFeatureClassContainer = pFDataset
Set pLine = pFeatureClassContainer.ClassByName(strmName)
Set pEdgeTable = OpenTable(strTableName)
pUID = "esricore.UtilityNetworkAnalysisExt"
Set pTraceSolver = pApp.FindExtensionByCLSID(pUID)
```

```
Set pNetElems = pTraceSolver.CurrentNetwork.Network

'program will stop if there isn't a geometric network loaded
Set pGeometricNetwork = pTraceSolver.CurrentNetwork
Set pNetwork = pGeometricNetwork.Network

lngSaveCount = 0

If blnFirstProgramRun = True Then
Call AddField
End If

pApp.Caption = "Check fields."
'check to make sure that a new cumulative mean attribute field was added in the AddField
'Subroutine
lngCumMeanFieldPos = pEdgeTable.FindField(strCumMeanFieldName)

If lngCumMeanFieldPos < 0 Then
MsgBox CStr(pLine.AliasName & " does not contain a cumulative mean field named, "
& strCumMeanFieldName)
Exit Sub
End If

'check to make sure that a local mean attribute field exists in the streams attribute table
lngRCAMeanFieldPos = pEdgeTable.FindField(strRCAMeanFieldName)
If lngRCAMeanFieldPos < 0 Then
MsgBox CStr(pLine.AliasName & " does not contain a RCA mean attribute field named,
" & strRCAMeanFieldName)
Exit Sub
End If

'check to make sure that a RCA area attribute field exists in the streams attribute table
lngRCAreaFieldPos = pEdgeTable.FindField(strRCAreaFieldName)
If lngRCAreaFieldPos < 0 Then
MsgBox CStr(pLine.AliasName & " does not contain a RCA attribute field named, " &
strRCAreaFieldName)
Exit Sub
End If

pApp.Caption = "Flow direction"
'set flow direction of network
Call SetFlowDirection

'create an enumeration of Junction EIDs
Set pJuncSel = GetCurrentEIDs(esriETJunction)
```

```
'Set up the IForwardStar Object
Dim pNetAttribute As INetWeight 'this represents HCA attribute for each segment
Dim pNetWeightNothing As INetWeight
Dim pNetSchema As INetSchema
Dim pForwardStar As IForwardStar
Dim pForwardStar2 As IForwardStar
Dim adjEdgesCount As Long, i As Double
Dim NextEdgesCount As Long, intArrayPos As Integer
Dim juncElement As Variant, intOutEdgeCount As Integer
Dim FirstOrderEID As Long, FirstOrderWeight As Double, FirstOrderDirection As
Boolean
Dim blnMissed As Boolean
Dim elementEID As Long, elementOID As Long, elementClassID As Long, elementSub
As Long
Dim pFirstOrderRow As IRow

Set pNetSchema = pNetwork
Set pNetAttribute = pNetSchema.WeightByName(strWeightName)
Set pNetWeightNothing = Nothing
blnMissed = False

'Get the editor reference and set pEditor
Call GetEditorReference

'************************************************************************
'Part 1: Locate 1st order streams and find adjacent edges
'************************************************************************
pApp.Caption = "1st order streams"

'In this case, since you do not have any junction or turn weights on your network
'you can use the NetWeightNothing object (pNetWeight) for these parameters.
Set pForwardStar = pNetwork.CreateForwardStar(True, pNetWeightNothing,
pNetAttribute, pNetAttribute, pNetWeightNothing)

pJuncSel.Reset

'look at all of the junctions in the network and find their adjacent edges
For i = 0 To pJuncSel.Count - 1
pJuncEID = pJuncSel.Next
pForwardStar.FindAdjacent 0, pJuncEID, adjEdgesCount

If adjEdgesCount > 0 Then
'the QueryAdjacentEdges function returns arrays - so reallocate memory using
'adjacent edges count returned by FindAdjacent
Erase lngFirstOrderEdges
Erase vFirstOrderWtVals
```

```
Erase bFirstOrderRev

ReDim lngFirstOrderEdges(adjEdgesCount - 1) As Long
ReDim vFirstOrderWtVals(adjEdgesCount - 1) As Variant
ReDim bFirstOrderRev(adjEdgesCount - 1) As Boolean

'query for edges adjacent to ToJunc
pForwardStar.QueryAdjacentEdges adjEdgesCount, lngFirstOrderEdges(0),
bFirstOrderRev(0), vFirstOrderWtVals(0)

'reset the out flowing edges count
intOutEdgeCount = 0

For Each juncElement In bFirstOrderRev
FirstOrderDirection = juncElement

'if this edge flows out of the junction....
If Not FirstOrderDirection = False Then add one to the out flowing edges count
intOutEdgeCount = intOutEdgeCount + 1
End If
Next

'If this is a first order stream
If intOutEdgeCount = adjEdgesCount Then
intArrayPos = 0

For Each juncElement In lngFirstOrderEdges
'find the OID for the edge
FirstOrderEID = juncElement
pNetElems.QueryIDs FirstOrderEID, esriETEdge, elementClassID, elementOID,
elementSub

Set pFirstOrderRow = pEdgeTable.GetRow(elementOID)
dblCumCatchAtt = pFirstOrderRow.Value(lngRCAMeanFieldPos)

'substitute -9999 for NULL values
If IsNull(dblCumCatchAtt) Then
dblCumCatchAtt = -9999
End If

If FirstOrderEID > 0 Then
'pNetElems.QueryIDs FirstOrderEID, esriETEdge, elementClassID, elementOID,
elementSub
Call RecordResults(elementOID, dblCumCatchAtt)

'Find the next adjacent junction and pass it to the  FindAdjacentEdges function
```

```
ToJuncEID As Long, ToJuncWeight As Variant
ToJuncWeight = 0

Set pForwardStar2 = Nothing
Set pForwardStar2 = pNetwork.CreateForwardStar(True, pNetWeightNothing,
pNetAttribute, pNetAttribute, pNetWeightNothing)
pForwardStar.QueryAdjacentJunction 0, ToJuncEID, ToJuncWeight

'pass the ToJuncEID to FindAdjacentEdges
pNextJuncEID = FindAdjacentEdges(ToJuncEID, blnMissed)

'If the cumulative catchment attribute was recorded for the previous edge and there is
another 'downstream edge, keep moving downstream and calculating/recording
cumulative 'attribute
While Not pNextJuncEID = -9999
pNextJuncEID = FindAdjacentEdges(pNextJuncEID, blnMissed)
Wend
End If

intArrayPos = intArrayPos + 1
Next
End If
End If
Next i

'**********************************************************************
'Part 2: Find adjacent edges when there were multiple downstream edges
'**********************************************************************
Dim DSElement As Variant
Dim lngEID As Long
Dim c As Double, g As Double, Index As Double

DSEdge: c = downstreamEIDs.Count
pApp.Caption = "Extra D.S. Edges: " & c
Index = 1

For Each DSElement In downstreamEIDs
lngEID = DSElement
While Not lngEID = -9999
lngEID = FindAdjacentEdges(lngEID, blnMissed)
pApp.Caption = "Extra D.S. Edges: " & downstreamEIDs.Count
If downstreamEIDs.Count > 0 Then
downstreamEIDs.Remove 1
End If
Index = Index + 1
Wend
```

```
Next

If downstreamEIDs.Count > 0 Then GoTo DSEdge:
'*************************************************************************
'Part 3: locate edges that have not been recorded and calculate a cumulative attribute for
'them
'*************************************************************************
'this function creates a new table of EdgeEID values and their cum catch attributes
'it sorts them according to EdgeEID and creates a collection of EdgeEIDs that do not
'have a cumulative catchment attribute recorded (MissedEdges)

Call CreateMissedEdgeCollection

pApp.Caption = "Filling in missing edges " & MissedEdges.Count
'fill in cumulative catchment attribute for missed edges
Dim e As Double
Dim lngMissedOID As Long
Dim pNetTopology As INetTopology
Dim lngMissedJuncEID As Long
Dim lngFromEID As Long, lngToEID As Long
Dim lngUpdatedCount As Double, lastMissedCount As Double
Dim MissedElement As Variant, pMissedRow As IRow
Dim lngFCID As Long, lngSubID As Long

Set pNetTopology = pNetwork
lngFCID = pLine.FeatureClassID
lngSubID = -9999

'look through the collection of missed edges
Line1: For Each MissedElement In MissedEdges
blnMissed = True
lastMissedCount = MissedEdges.Count
lngMissedOID = MissedElement

lngEdgeEID = pNetElems.GetEID(lngFCID, lngMissedOID, lngSubID, esriETEdge)

Set pMissedRow = pEdgeTable.GetRow(lngMissedOID)
If IsNull(pMissedRow.Value(lngCumMeanFieldPos)) Then

'Get the junctions that are adjacent to the missed edge the from and to node variables are
'switched below because the NHD data is digitized against flow
pNetTopology.GetFromToJunctionEIDs lngEdgeEID, lngToEID, lngFromEID
pNextJuncEID = FindAdjacentEdges(lngFromEID, blnMissed)

'fill in the cum. catchment attribute values for the missed edges
While Not pNextJuncEID = -9999
```

```
pNextJuncEID = FindAdjacentEdges(pNextJuncEID, blnMissed)
Wend
End If
pApp.Caption = "Filling in missing edges " & MissedEdges.Count
Next

Set MissedEdges = Nothing
Call CreateMissedEdgeCollection

lngUpdatedCount = MissedEdges.Count

'If the cumulative attribute was not calculated for any of the missed edges or all the edges
' have been recorded exit the program. Otherwise go back and attempt to fill in more
' missed edges
If lngUpdatedCount = lastMissedCount Then GoTo Line2
If lngUpdatedCount = 0 Then GoTo Line2 Else GoTo Line1
pEditor.StopEditing True
Line2: Call Reset
pApp.Caption = "Program is complete!"
End Sub


Private Sub SetFlowDirection()
'This subroutine sets the flow direction in the geometric network against the digitized
'direction.

Dim pUtilityNetwork As IUtilityNetwork  'necessary to get or set flow direction
Dim lngEdgeEIDCount As Long  'stores the number of edge EID's in the network
Dim lngEdgeEID As Long   'stores the current edge EID
Dim i As Double  'loop iterator
pApp.Caption = "SetFlowDirection"

'get a reference to the current network
Set pUtilityNetwork = pGeometricNetwork.Network

'Get the editor reference and set pEditor
Call GetEditorReference

'create an edit operation enabling an undo for this operation
pEditor.StartOperation

'get a list of the EID's for edges in the network
Set pEdgeSel = GetCurrentEIDs(esriETEdge)
'set the flow direction for each edge in the network
'reset the enumeration
pEdgeSel.Reset
```

```
'determine the number of edges in the enumeration
lngEdgeEIDCount = pEdgeSel.Count
For i = 0 To lngEdgeEIDCount - 1
'get the next edge EID
lngEdgeEID = pEdgeSel.Next
'set the flow direction for this edge
pUtilityNetwork.SetFlowDirection lngEdgeEID, esriFDAgainstFlow
Next i

'stop the edit operation and specify the name of this edit operation
pEditor.StopOperation "Set Flow Direction"
pEditor.StopEditing True
End Sub


Public Sub GetEditorReference()
'an editor reference is needed to start an editing session
Dim pID As esriCore.UID

Set pID = New esriCore.UID
pID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pID)
StartEditing pEditor
End Sub


Public Sub StartEditing(pEditor As IEditor)
Dim pDataset As IDataset
Dim pDatasetWorkspace As IWorkspace

If Not pEditor.EditState = esriStateNotEditing Then Exit Sub
Set pDataset = pLine
Set pDatasetWorkspace = pDataset.Workspace
pEditor.StartEditing pDataset.Workspace
End Sub


Private Function GetCurrentEIDs(eElementType As esriElementType) As IEnumNetEID
'return an enumeration of network EID's for a given element type
pApp.Caption = "GetCurrentEIDs"

Dim pNetwork As INetwork   'holds a reference to the current network
Dim pEnumNetEID As IEnumNetEID   'stores a list of network EID's
Set pNetwork = pGeometricNetwork.Network
Set pEnumNetEID = pNetwork.CreateNetBrowser(eElementType)
```

```
'return the enumeration
Set GetCurrentEIDs = pEnumNetEID
End Function


Private Sub RecordResults(intEdgeOID As Long, lngCatchAtt As Double)
'this function records the cumulative catchment attribute in the Edges attribute table

Dim pRecordRow As IRow

If intEdgeOID > 0 Then
Set pRecordRow = pEdgeTable.GetRow(intEdgeOID)

'if the program has been run before, then check the cumulative attribute field before
'recording in it
If blnFirstProgramRun = False Then
'check to see if the cumulative attribute has already been recorded
If IsNull(pRecordRow.Value(lngCumMeanFieldPos)) Then
pApp.Caption = "RecordResults"
'create an edit operation enabling an undo for this operation
pEditor.StartOperation

'add this value to the cumulative upstream catchment attribute field
pRecordRow.Value(lngCumMeanFieldPos) = lngCatchAtt
pRecordRow.Store

'add one to the count
lngSaveCount = lngSaveCount + 1
End If

'if this is the first time the program has been run, go ahead and record cum. attribute
'value w/out checking the field
Else
pEditor.StartOperation
'add this value to the cumulative upstream catchment attribute field
pRecordRow.Value(lngCumMeanFieldPos) = lngCatchAtt
pRecordRow.Store

'add one to the save count
lngSaveCount = lngSaveCount + 1
End If
End If

pEditor.StopOperation ("RecordResults")

'if a specified number of edits have been completed, save them
```

```
If lngSaveCount = lngSaveEdits Then
pEditor.StopEditing True
StartEditing pEditor
lngSaveCount = 0
End If

'release memory
Set pRecordRow = Nothing
End Sub


Private Sub Reset()
pApp.Caption = "Flush memory"
'release the memory
Set pWorkspaceFactory = Nothing
Set pWorkspace = Nothing
Set pFeatureWorkspace = Nothing
Set pFDataset = Nothing
Set pNetworkWorkspace = Nothing
Set pFeatureClassContainer = Nothing
Set pGeometricNetwork = Nothing
Set pMxDoc = Nothing
Set pMap = Nothing
Set pActiveView = Nothing
Set pTraceSolver = Nothing
pEditor.StopEditing True
Set pEditor = Nothing
Set pLine = Nothing
Set pJuncSel = Nothing
Set pEdgeSel = Nothing
Set pGeometricNetwork = Nothing
Set pNetwork = Nothing
Set pEdgeTable = Nothing
Set pSortTable = Nothing
Set MissedEdges = Nothing
Set downstreamEIDs = Nothing

lngCumCatchAtt = 0          'variable containing the cumulative upstream catchment
                            'attribute for a single edge
pNextJuncEID = 0            'next downstream junction EID
lngDownstreamEdgeEID = 0 'the downstream edge OID
pJuncEID = 0                'variable containing the EID for a junction in the network
lngCumMeanFieldPos = 0
dblCumCatchAtt = 0

End Sub
```

```
Private Function FindAdjacentEdges(ToJuncEID As Long, blnMissed As Boolean) As
Long
'pApp.Caption = "Finding Adjacent Edges"

'find downstream junction
Dim pNetAttribute As INetWeight    'Catchment attribute
Dim pNetWeightNothing As INetWeight
Dim pNetSchema As INetSchema
Dim pPassJuncEID As Long
Dim elementOID As Long, elementClassID As Long, elementSub As Long
Dim dsEID As Long, dblDSEdgeRCArea As Double, USCumArea As Double,
dblEdgeRCAMean As Double


Dim pForwardStar3 As IForwardStar
Dim pForwardStar4 As IForwardStar
Dim MultipleDSEdgeEIDs() As Long
Dim LocalAttArrayR() As Variant

Dim dblDSMean() As Double
Dim dblUSMean() As Double
Dim dblDSArea() As Double
Dim dblUSArea() As Double

Set pNetSchema = pNetwork
Set pNetAttribute = pNetSchema.WeightByName(strWeightName)
Set pNetWeightNothing = Nothing


'find the adjacent edges from To junction
Dim adjEdgesCount As Long
Set pForwardStar3 = pNetwork.CreateForwardStar(True, pNetWeightNothing,
pNetAttribute, pNetAttribute, pNetWeightNothing)
pForwardStar3.FindAdjacent 0, ToJuncEID, adjEdgesCount

'the QueryAdjacentEdges function returns arrays - so reallocate memory using
'adjacent edges count returned by FindAdjacent
Erase pAdjacentEdges
Erase vWtVals
Erase bRev

ReDim pAdjacentEdges(adjEdgesCount - 1) As Long
ReDim vWtVals(adjEdgesCount - 1) As Variant
ReDim bRev(adjEdgesCount - 1) As Boolean
```

```
'query for edges adjacent to ToJunc
pForwardStar3.QueryAdjacentEdges adjEdgesCount, pAdjacentEdges(0), bRev(0),
vWtVals(0)

Dim lngUpstreamEdgeEID As Long
Dim dblUpstreamEdgeWeight As Double
Dim dblDownstreamEdgeWeight As Double
Dim j As Double
Dim foundCount As Double, downstreamEdgeCount As Integer
Dim pRow As IRow
Dim DSfoundcount As Double, USfoundcount As Double

dblCumCatchAtt = 0
j = 0
foundCount = 0
DSfoundcount = 0
USfoundcount = 0
downstreamEdgeCount = 0

lngDownstreamEdgeEID = 0
dblDownstreamEdgeWeight = 0

'Reallocate memory for mean attribute array and cumulative area array
Erase dblDSMean
Erase dblUSMean
Erase dblDSArea
Erase dblUSArea

ReDim dblDSMean(adjEdgesCount - 1) As Double
ReDim dblUSMean(adjEdgesCount - 1) As Double
ReDim dblDSArea(adjEdgesCount - 1) As Double
ReDim dblUSArea(adjEdgesCount - 1) As Double

' look at each of the edges which are adjacent to the junction
For j = 0 To adjEdgesCount - 1
lngUpstreamEdgeEID = 0
dblUpstreamEdgeWeight = 0

'***************If the edge is UPSTREAM If bRev(j) = 0 Then
lngUpstreamEdgeEID = pAdjacentEdges(j)

If lngUpstreamEdgeEID > 0 Then
pNetElems.QueryIDs lngUpstreamEdgeEID, esriETEdge, elementClassID, elementOID,
elementSub
```

```
Set pRow = pEdgeTable.GetRow(elementOID)

'check to see if the mean cumulative HCA attribute has been recorded
If Not IsNull(pRow.Value(lngCumMeanFieldPos)) Then
'add values to the upstream area and mean arrays
dblUSArea(USfoundcount) = vWtVals(j)
dblUSMean(USfoundcount) = pRow.Value(lngCumMeanFieldPos)
'dblCumCatchAtt = dblCumCatchAtt + pRow.Value(lngCumMeanFieldPos)
foundCount = foundCount + 1
USfoundcount = USfoundcount + 1

Else
'if cum Mean attribute hasn't been recorded, exit function
pPassJuncEID = -9999
FindAdjacentEdges = pPassJuncEID
Exit Function
End If
End If

Set pRow = Nothing

'***************IF THE EDGE IS DOWNSTREAM *****************
Else

'***************MISSED EDGES *********************
'If filling in a missed edges - only record the missed edge
If blnMissed = True Then
'Add one to the counts
downstreamEdgeCount = downstreamEdgeCount + 1
foundCount = foundCount + 1
DSfoundcount = DSfoundcount + 1

'If the edge matches the missed edge EID
If pAdjacentEdges(j) = lngEdgeEID Then
lngDownstreamEdgeEID = pAdjacentEdges(j)
'save cumulative catchment area
dblDownstreamEdgeWeight = vWtVals(j)
End If
'***************NOT MISSED***********
Else
downstreamEdgeCount = downstreamEdgeCount + 1
DSfoundcount = DSfoundcount + 1

'******MORE THAN ONE DOWNSTREAM EDGE - NOT MISSED********
If downstreamEdgeCount > 1 Then
If downstreamEdgeCount < 3 Then
```

```
'create an array of network element EIDs for downstream edges
ReDim MultipleDSEdgeEIDs(adjEdgesCount - 1) As Long
ReDim LocalAttArrayR(adjEdgesCount - 1) As Variant

MultipleDSEdgeEIDs(0) = lngDownstreamEdgeEID
LocalAttArrayR(0) = dblDownstreamEdgeWeight

MultipleDSEdgeEIDs(downstreamEdgeCount - 1) = pAdjacentEdges(j)
LocalAttArrayR(downstreamEdgeCount - 1) = vWtVals(j)
Else

MultipleDSEdgeEIDs(downstreamEdgeCount - 1) = pAdjacentEdges(j)
LocalAttArrayR(downstreamEdgeCount - 1) = vWtVals(j)

End If

MultipleDSEdgeEIDs(0) = lngDownstreamEdgeEID
LocalAttArrayR(0) = dblDownstreamEdgeWeight
MultipleDSEdgeEIDs(downstreamEdgeCount - 1) = pAdjacentEdges(j)
LocalAttArrayR(downstreamEdgeCount - 1) = vWtVals(j)
foundCount = foundCount + 1

'***************ONLY ONE DOWNSTREAM EDGE - NOT MISSED********
Else
'Else the edge is downstream.......
lngDownstreamEdgeEID = pAdjacentEdges(j)
dblDownstreamEdgeWeight = vWtVals(j)
foundCount = foundCount + 1
End If
End If
End If
Next j

Set pRow = Nothing

'*****REMOVE UPSTREAM NO DATA VALUES FROM THE CALCULATIONS
Dim USEdges As Variant
Dim p As Integer
Dim Mean As Double
p = 0

For Each USEdges In dblUSMean
Mean = USEdges
If Mean = -9999 Then
dblDownstreamEdgeWeight = dblDownstreamEdgeWeight - dblUSArea(p)
adjEdgesCount = adjEdgesCount - 1
```

```
foundCount = foundCount - 1
USfoundcount = USfoundcount - 1
End If
p = p + 1
Next

'*****CALCULATE CUMULATIVE MEAN FOR EDGE & RECORD****
'pPassJuncEID is the junction id # that is returned by this function (-9999 = nodata)
pPassJuncEID = -9999
'reset the upstream cumulative area variable
USCumArea = 0
Dim WMean As Double
Dim MissedDSEdgeCount As Integer
MissedDSEdgeCount = downstreamEdgeCount

If blnMissed = True Then downstreamEdgeCount = 1
'If all of the upstream edges have cum. attribute reported and the downstream edge is
'valid then
If foundCount = adjEdgesCount And lngDownstreamEdgeEID > 0 Then

'*********GREATER THAN ONE DOWNSTREAM EDGE****
If downstreamEdgeCount > 1 Then
Dim MDSEdgeElement As Variant
Dim MultCumCatchAtt As Double
Dim EdgeEID As Variant
Dim ArrayIndex As Long
ArrayIndex = 0
MultCumCatchAtt = 0

For Each MDSEdgeElement In MultipleDSEdgeEIDs
lngDownstreamEdgeEID = MDSEdgeElement

If lngDownstreamEdgeEID > 0 Then
dblDownstreamEdgeWeight = LocalAttArrayR(ArrayIndex)

'Get downstream RCA area
pNetElems.QueryIDs lngDownstreamEdgeEID, esriETEdge,elementClassID,
elementOID, elementSub
Set pRow = pEdgeTable.GetRow(elementOID)
dblDSEdgeRCArea = pRow.Value(lngRCAreaFieldPos)
dblEdgeRCAMean = pRow.Value(lngRCAMeanFieldPos)

If dblEdgeRCAMean = -9999 Then
dblDownstreamEdgeWeight = dblDownstreamEdgeWeight - dblDSEdgeRCArea
End If
```

```
Dim DSE As Variant
Dim a As Integer
a = 0
WMean = 0

For a = 0 To a = USfoundcount
USCumArea = dblUSArea(a)
If Not dblUSMean(a) = -9999 Then
If Not dblDownstreamEdgeWeight = 0 Then
WMean = WMean + (((USCumArea / MissedDSEdgeCount) /
dblDownstreamEdgeWeight) * dblUSMean(a))
End If
End If

a = a + 1
Next

If dblEdgeRCAMean = -9999 Then
dblCumCatchAtt = WMean
Else
dblCumCatchAtt = WMean + ((dblDSEdgeRCArea / dblDownstreamEdgeWeight) *
dblEdgeRCAMean)
End If

Call RecordResults(elementOID, dblCumCatchAtt)
pApp.Caption = "FindAdjacentEdges"

If ArrayIndex = 0 Then
pPassJuncEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
Else
dsEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
downstreamEIDs.Add dsEID
End If

ArrayIndex = ArrayIndex + 1
End If
Next

'******************ONLY ONE DOWNSTREAM EDGE*******
Else

'*************ONE DOWNSTREAM EDGE - MISSED
If blnMissed = True Then
If lngDownstreamEdgeEID > 0 Then
'Get downstream RCA area
```

```
pNetElems.QueryIDs lngDownstreamEdgeEID, esriETEdge, elementClassID,
elementOID, elementSub
Set pRow = pEdgeTable.GetRow(elementOID)
dblDSEdgeRCArea = pRow.Value(lngRCAreaFieldPos)
dblEdgeRCAMean = pRow.Value(lngRCAMeanFieldPos)

'If no data for RCA mean, then subtract RCA Area from CumArea
If dblEdgeRCAMean = -9999 Then
dblDownstreamEdgeWeight = dblDownstreamEdgeWeight - dblDSEdgeRCArea
End If

Dim DSEM As Variant
Dim b As Integer
b = 0
WMean = 0

For Each DSEM In dblUSArea
USCumArea = DSEM

If Not dblUSMean(b) = -9999 Then
WMean = WMean + (((USCumArea / MissedDSEdgeCount) /
dblDownstreamEdgeWeight) * dblUSMean(b))
End If
End If
b = b + 1
Next

If dblEdgeRCAMean = -9999 Then
dblCumCatchAtt = WMean
Else
dblCumCatchAtt = WMean + ((dblDSEdgeRCArea / dblDownstreamEdgeWeight) *
dblEdgeRCAMean)
End If

Call RecordResults(elementOID, dblCumCatchAtt)
pPassJuncEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
End If

'*************ONE DOWNSTREAM EDGE AND NOT MISSED***************
Else
'Get downstream RCA area
pNetElems.QueryIDs lngDownstreamEdgeEID, esriETEdge, elementClassID,
elementOID, elementSub
Set pRow = pEdgeTable.GetRow(elementOID)
dblDSEdgeRCArea = pRow.Value(lngRCAreaFieldPos)
dblEdgeRCAMean = pRow.Value(lngRCAMeanFieldPos)
```

```
If lngDownstreamEdgeEID > 0 Then

'If no data for RCA mean, then subtract RCA Area from CumArea
If dblEdgeRCAMean = -9999 Then
dblDownstreamEdgeWeight = dblDownstreamEdgeWeight - dblDSEdgeRCArea
End If

Dim DSENM As Variant
Dim c As Integer
c = 0
WMean = 0

For Each DSENM In dblUSArea
USCumArea = DSENM

If Not dblUSMean(c) = -9999 Then
If Not dblDownstreamEdgeWeight = 0 Then
WMean = WMean + ((USCumArea / dblDownstreamEdgeWeight) * dblUSMean(c))
End If
End If
c = c + 1
Next

If dblEdgeRCAMean = -9999 Then
dblCumCatchAtt = WMean
Else
dblCumCatchAtt = WMean + ((dblDSEdgeRCArea / dblDownstreamEdgeWeight) *
dblEdgeRCAMean)
End If

pNetElems.QueryIDs lngDownstreamEdgeEID, esriETEdge, elementClassID,
elementOID, elementSub
Call RecordResults(elementOID, dblCumCatchAtt)
pPassJuncEID = LocateDownstreamJunc(lngDownstreamEdgeEID)
End If
End If
End If
End If

Set pRow = Nothing
FindAdjacentEdges = pPassJuncEID
End Function

Private Sub AddField()
'this subroutine checks the edges feature class in the geometric network to see if there
```

```
'is a field for the cumulative HCA attribute (const strCumAttFieldName). If so, it deletes
it and adds a new
'field. If not, it just adds the new field

pApp.Caption = "Create a new field"
Dim pAddField As IField
Dim pFieldEdit As IFieldEdit

'create new field
Set pAddField = New Field
Set pFieldEdit = pAddField

With pFieldEdit
.AliasName = strCumMeanFieldName
.Type = esriFieldTypeDouble
.Length = 30
.Name = strCumMeanFieldName
.Editable = True
.IsNullable = True
End With

Set pAddField = pFieldEdit
Set pFieldEdit = Nothing

If pEdgeTable.FindField(strCumMeanFieldName) = -1 Then
pEdgeTable.AddField pAddField
Else
Dim pFields As IFields
Dim pDeleteField As IField
Set pFields = pEdgeTable.Fields
lngCumMeanFieldPos = pFields.FindField(strCumMeanFieldName)
Set pDeleteField = pFields.Field(lngCumMeanFieldPos)
pEdgeTable.DeleteField pDeleteField
pEdgeTable.AddField pAddField
End If
Exit Sub
End Sub


Private Function OpenTable(TableName As String) As ITable
'Private Function OpenTable(strTableWorkspace As String, TableName As String) As
ITable
On Error GoTo Errorhandler
Dim pDatasetname As IDatasetName
Dim pName As IName
Dim pWorkspaceName As IWorkspaceName
```

```
Set pWorkspaceName = New WorkspaceName
pWorkspaceName.WorkspaceFactoryProgID = "esricore.AccessWorkspaceFactory"
pWorkspaceName.PathName = geodatabasePathName

'pApp.Caption = "Open Table"

'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenTable = pName.Open

Exit Function 'exit to avoid error handler
Errorhandler:
Set OpenTable = Nothing
End Function


Private Sub UpdateCollection(lngDownstreamEID As Long)
Dim e As Double
Dim lngMissedOID As Long
Dim mElement As Variant
e = 1
pApp.Caption = "Updating missed edge collection"

For Each mElement In MissedEdges
lngMissedOID = mElement
If lngMissedOID = lngDownstreamEID Then
MissedEdges.Remove e
Exit For
End If
e = e + 1
Next

pApp.Caption = "done updating me" & MissedEdges.Count
End Sub


Private Function LocateDownstreamJunc(lngEdgeEID As Long) As Long
'uses a junction EID that is passed into the function to find the next downstream junction.
'it returns the downstream junction EID
Dim pNetTopology As INetTopology
Dim lngFromEID As Long, lngToEID As Long
```

```
Set pNetTopology = pNetwork

'Get the To Junction for the edge- the from and to node variables are switched below
'because the NHD data is digitized against flow
pNetTopology.GetFromToJunctionEIDs lngEdgeEID, lngToEID, lngFromEID

LocateDownstreamJunc = lngToEID
End Function


Private Sub CreateMissedEdgeCollection()
'creates a collection of edges that do not have their cumulative catchment attribute
'recorded

pApp.Caption = "creating m.e. collection"
Dim pFields As IFields
Dim pField As IField
Dim OIDField As Long
Dim sOIDField As String

'set up the parameters for the ITableSort
Set pFields = pEdgeTable.Fields

Set pField = pFields.Field(0)
sOIDField = pField.Name
OIDField = pEdgeTable.FindField(pField.Name)

Dim pRow As IRow
Dim pQueryFilter As IQueryFilter
Set pQueryFilter = New QueryFilter
'*********************************************************************
'the [fieldname] must be changed to match the strCumAttFieldName in the constants
'section
'*********************************************************************
pQueryFilter.WhereClause = "[CumElev_m]IS NULL"

'create the ITableSort
Dim pTableSort As ITableSort
Set pTableSort = New TableSort
With pTableSort
.Fields = sOIDField
.Ascending(sOIDField) = True
Set .QueryFilter = pQueryFilter
Set .Table = pEdgeTable
End With
```

```
pTableSort.Sort Nothing

'use the pTableSort to create a cursor containing the missed edges
Dim pCursor As ICursor
Dim OID As Long
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow
Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow

'use the pCursor populate the missed edges collection
Do While Not pRow Is Nothing
OID = pRow.Value(0)
'add this EdgeOID to a collection
MissedEdges.Add OID
Set pRow = pCursor.NextRow
Loop
pApp.Caption = "done w/ m.e. collection " & MissedEdges.Count
End Sub
```

CALCULATE WATERSHED ATTRIBUTES

```
'bas calcWatershedAtt.bas
'Created by Erin Peterson
'Program is complete   Last updated: 4/7/04

'PURPOSE: The purpose of this program is to record the watershed attribute for each
'sample point. This is accomplished by selecting each sample point in the network,
'querying the adjacent edges, selecting those edges that flow into the junction, and
'summing their cumulative upstream attribute to get the total upstream attribute for each
'site. This value is reported in a new field in the sample sites attribute table.

'INPUT: Geometric network containing sample sites. The sample sites should not fall
'along the edges, rather they should be used to split the edges so that they fall at the ends.
'The network should have ONE cumulative upstream attribute as a weight value. This
'program should be run after the AccumCatchArea.bas program, which will assign
'cumulative watershed attributes to each stream segment.

'OUTPUT: A geometric network with a new field in the sample sites feature class that
'contains the total upstream watershed attribute.

'PREPROCESSING: The basSplitLineWithPoint.bas program should be used to split the
'edges with the the sample points. Then the preprocessing steps for the
'basSmartCatchment.bas program should be completed (these can be found in the file
```

```
'smartcatchment_preprocessing.doc) and then the SmartCatchment.bas program should
'be run. The preprocessing steps for the AccumCatchArea.bas program should be
'completed (contained in the AccumCatchArea_preprocessing.doc file) and the
'AccumCatchArea.bas script should be run. The geometric network should be rebuilt and
'the cumulative catchment attribute should be included as a network weight. At this
'point, the network should be ready as input into the basWatershedAtt.bas program. This
'program is designed to operate on a geometric network that only has ONE weight
'assigned to it.


'******ALL CONSTANTS MUST BE DEFINED BY THE USER
Private Const strEdgeTableName As String = "rch"
Private Const strSiteTableName As String = "sites_all"
Private Const strSortName As String = "sort"
Private Const strmName As String = "rch"   'name of streams feature class
                                            'included in geometric network
Private Const siteName As String = "sites_all"
Private Const fDatasetName As String = "accumfd"  'name of feature dataset that
                                            'contains geometric network
Private Const geodatabasePathName As String = "c:\erin\jay\accum\accum2.mdb"
            'pathname to the geodatabase that contains the geometric network
Private Const strTableWorkspace As String = "pathname here" 'workspace where all
                                            'tables will be saved
Private Const strCatchAttFN As String = "CumArea_m2"   'name of new field in network
                                            'table for cumulative upstream HCA area
Private Const strWeightName As String = "Total_Area" 'name that was given to the
                                            'weight in the geometric network
'**********************************************************************
Option Explicit
Private pApp As IApplication
Private pMxDoc As IMxDocument
Private pMap As IMap
Private pActiveView As IActiveView
Private pWorkspaceFactory As IWorkspaceFactory
Private pWorkspace As IWorkspace
Private pFeatureWorkspace As IFeatureWorkspace
Private PFDataset As IFeatureDataset
Private pNetworkWorkspace As INetworkWorkspace
Private pFeatureClassContainer As IFeatureClassContainer
Private pUID As New esriCore.UID          'ID for UtilityNetworkAnalysisExt
Private pTraceSolver As INetworkAnalysisExt
Private pEditor As IEditor
Private pLine As IFeatureClass            'stream feature class
Private pPoint As IFeatureClass           'sample points feature class
Private pJuncSel As IEnumNetEID 'enumeration of all the junctions in the network
Private pEdgeSel As IEnumNetEID 'enumeration of all the edges in the network
```

```
Private pGeometricNetwork As IGeometricNetwork  'geometric network loaded into the
                                                'map frame
Private pNetwork As INetwork
Private pEdgeTable As ITable        'attribute table for edges in geometric network
Private pSiteTable As ITable        'attribute table for sample sites in geometric
                                    'network
Private pSortTable As ITable        'dbf table created to sort edges
Private pAdjacentEdges() As Long    'an array of adjacent edges returned by the
                                    'IForwardStar
Private vWtVals() As Variant        'an array of adjacent edge weights returned by the
                                    'IForwardStar
Private bRev() As Boolean           'an array containing the direction of the adjacent
                                    'edges returned by the IForwardStar
Private lngCumCatchArea As Long     'variable containing the cumulative upstream
                                    'catchment area for a single edge
Private pNextJuncEID As Long        'next downstream junction EID
Private lngDownstreamEdgeOID As Long    'the downstream edge OID
Private pJuncEID As Long 'variable containing the EID for a junction in the network
Private lngCatchAreaFPos As Long    'field position of the cumulative catchment area
(CumArea_sqM) in the pEdgeTable
Private dblCumCatchArea As Double       'cumulative catchment area
Private pJuncCollection As New Collection  'collection of edges that have a Null value
                                           'for CumArea_sqM
Private pNetworkElem As NetworkElement

Sub Main()
Set pApp = New AppRef
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pActiveView = pMxDoc.FocusMap

pApp.Caption = "Setting up workspace."

'New AccessWorkspaceFactory is IWorkspaceFactory for personal geodatabase feature
classes
Set pWorkspaceFactory = New AccessWorkspaceFactory
Set pWorkspace = pWorkspaceFactory.OpenFromFile(geodatabasePathName, 0)
Set pFeatureWorkspace = pWorkspace  'QI for the IFeatureWorkspace
Set PFDataset = pFeatureWorkspace.OpenFeatureDataset(fDatasetName)
Set pNetworkWorkspace = PFDataset.Workspace

'need feature class container to access feature classes within a feature dataset
Set pFeatureClassContainer = PFDataset
Set pLine = pFeatureClassContainer.ClassByName(strmName)
Set pPoint = pFeatureClassContainer.ClassByName(siteName)
Set pEdgeTable = OpenTable(strEdgeTableName)
```

```
Set pSiteTable = OpenTable(strSiteTableName)

Dim pWorkspaceProperties As IWorkspaceProperties
Dim pWkspProp As IWorkspaceProperty
Set pWorkspaceProperties = pWorkspace
Set pWkspProp = pWorkspaceProperties.Property(esriWorkspaceTablePropertyGroup,
esriTablePropCanDeleteField)
'MsgBox " can delete table field? " & pWkspProp.PropertyValue

pUID = "esricore.UtilityNetworkAnalysisExt"
Set pTraceSolver = pApp.FindExtensionByCLSID(pUID)

'program will stop if there isn't a geometric network loaded
Set pGeometricNetwork = pTraceSolver.CurrentNetwork
Set pNetwork = pGeometricNetwork.Network

'add a new field to the Edges attribute table that will hold the proportional influence
Call AddField

'create an enumeration of all junctions found in the network.
Set pJuncSel = GetCurrentEIDs(esriETJunction)
Set pJuncCollection = CreateJunctionCollection(pJuncSel)
Dim z As Integer
z = 0

'select each junction in the network and find the adjacent edges
For z = 1 To pJuncCollection.Count
' Select each junction and then find adjacent edges
Set pNetworkElem = pJuncCollection.Item(z)
pJuncEID = pNetworkElem.ElementID

If pJuncEID > 0 Then
Call FindUpstreamInfluence(pJuncEID)
End If
Next
End Sub

Private Function OpenTable(TableName As String) As ITable
'function opens a table and returns it as an ITable using the table name as an input

On Error GoTo Errorhandler
Dim pDatasetname As IDatasetName
Dim pName As IName
Dim pWorkspaceName As IWorkspaceName
Set pWorkspaceName = New WorkspaceName
pWorkspaceName.WorkspaceFactoryProgID = "esricore.AccessWorkspaceFactory"
```

```
pWorkspaceName.PathName = geodatabasePathName

'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenTable = pName.Open

Exit Function 'exit to avoid error handler
Errorhandler:
Set OpenTable = Nothing
End Function


Private Function GetCurrentEIDs(eElementType As esriElementType) As IEnumNetEID
'return an enumeration of network EID's for a given element type

pApp.Caption = "GetCurrentEIDs"

Dim pNetwork As INetwork   'holds a reference to the current network
Dim pEnumNetEID As IEnumNetEID   'stores a list of network EID's
Set pNetwork = pGeometricNetwork.Network
Set pEnumNetEID = pNetwork.CreateNetBrowser(eElementType)

'return the enumeration
Set GetCurrentEIDs = pEnumNetEID
End Function


Private Sub FindUpstreamInfluence(pJuncEID As Long)
Dim lngAdjEdgeCount As Long
Dim adjEdges() As Long
Dim blnRevOr() As Boolean
Dim adjWeights() As Variant

'Set up the IForwardStar Object
Dim pNetWeightNothing As INetWeight
Dim pNetArea As INetWeight
Dim pNetSchema As INetSchema
Dim pForwardStar As IForwardStar
Dim pNetElems As INetElements
Dim adjEdgesCount As Long
Dim NextEdgesCount As Long
```

```
Dim adjEdgeEID As Long, AdjEdgeWeight As Variant, bDirection As Boolean
Dim edgeElement As Variant
Dim dblTotalArea As Double
Dim q As Integer
Dim edgeWeight As Double
Dim lngEdgeProp As Double
Dim elementEID, elementClassID As Long, elementOID As Long, elementSub As Long

Set pNetSchema = pNetwork
Set pNetArea = pNetSchema.WeightByName(strWeightName)
Set pNetWeightNothing = Nothing

'In this case, you are only interested in edge weights on your network, so
'you can set the other weights using the NetWeightNothing object (pNetWeight)
Set pForwardStar = pNetwork.CreateForwardStar(True, pNetWeightNothing, pNetArea,
pNetArea, pNetWeightNothing)

pForwardStar.FindAdjacent 0, pJuncEID, adjEdgesCount

'erase the old arrays
Erase adjEdges()
Erase blnRevOr()
Erase adjWeights()
'reallocate memory for the new arrays
ReDim adjEdges(adjEdgesCount) As Long
ReDim blnRevOr(adjEdgesCount) As Boolean
ReDim adjWeights(adjEdgesCount) As Variant

pForwardStar.QueryAdjacentEdges adjEdgesCount, adjEdges(0), blnRevOr(0),
adjWeights(0)

dblTotalArea = 0
q = 0

'look at each adjacent edge, determine whether it flows into the junction, and if so add it's
' upstream cumulative catchment area
For Each edgeElement In adjEdges
'if it is an upstream edge from junction
If blnRevOr(q) = False Then
edgeWeight = adjWeights(q)
dblTotalArea = dblTotalArea + edgeWeight
End If

q = q + 1
Next
```

```
'***********************************************************************
Set pNetElems = pTraceSolver.CurrentNetwork.Network
pNetElems.QueryIDs pJuncEID, esriETJunction, elementClassID, elementOID,
elementSub
'report the results in the edge attribute table
Call RecordResults(elementOID, dblTotalArea)
'***********************************************************************
End Sub


Private Sub AddField()
'this subroutine checks the edges feature class in the geometric network to see if there
'is a field for the proportional downstream influence (strCatchAttFN) for each segment.
'If so, it deletes it and adds a new field. If not, it just adds the new field

pApp.Caption = "Create a new field"

Dim pAddField As IField
Dim pFieldEdit As IFieldEdit

'create new field
Set pAddField = New Field
Set pFieldEdit = pAddField

With pFieldEdit
.AliasName = strCatchAttFN
.Type = esriFieldTypeDouble
.Length = 20
.Name = strCatchAttFN
.Editable = True
.IsNullable = True
End With

Set pAddField = pFieldEdit
Set pFieldEdit = Nothing

If pSiteTable.FindField(strCatchAttFN) = -1 Then
pSiteTable.AddField pAddField
Else
Dim pFields As IFields
Dim pDeleteField As IField
Set pFields = pSiteTable.Fields
lngCatchAreaFPos = pFields.FindField(strCatchAttFN)
Set pDeleteField = pFields.Field(lngCatchAreaFPos)
'MsgBox " prop field position : " & lngCatchAreaFPos
'MsgBox "pDelete field " & pDeleteField.Name
```

```
pSiteTable.DeleteField pDeleteField
pSiteTable.AddField pAddField
End If
lngCatchAreaFPos = pFields.FindField(strCatchAttFN)
Exit Sub
Set pSiteTable = OpenTable(strSiteTableName)
End Sub


Private Sub RecordResults(intJuncOID As Long, dblTotalArea As Double)
'this function records the proportional influence of each edge in the edges attribute table

pApp.Caption = "RecordResults"
'Get the editor reference and set pEditor
Call GetEditorReference
'create an edit operation enabling an undo for this operation
pEditor.StartOperation

' create cursor containing all records in the edges table
Dim pRecordRow As IRow

If intJuncOID > 0 Then
Set pRecordRow = pSiteTable.GetRow(intJuncOID)
'Debug.Print pRow.Value(2)
pApp.Caption = "Looking for record results row"
'add this area to the cumulative upstream catchment area field position
pRecordRow.Value(lngCatchAreaFPos) = dblTotalArea
pRecordRow.Store
End If

pEditor.StopOperation ("RecordResults")
pEditor.StopEditing True
pApp.Caption = "Done looking for record results row"
Set pRecordRow = Nothing
End Sub


Public Sub GetEditorReference()
'creates an editor reference needed to start an editing session
Dim pID As esriCore.UID

Set pID = New esriCore.UID
pID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pID)
StartEditing pEditor
End Sub
```

```
Public Sub StartEditing(pEditor As IEditor)
'starts an editing session

Dim pDataset As IDataset
Dim pDatasetWorkspace As IWorkspace

If Not pEditor.EditState = esriStateNotEditing Then Exit Sub
Set pDataset = pPoint
Set pDatasetWorkspace = pDataset.Workspace
pEditor.StartEditing pDataset.Workspace
End Sub

Private Sub Reset()
'releases memory of objects

pEditor.StopEditing True
Set pEditor = Nothing
End Sub


Private Function CreateJunctionCollection (pSelection As IEnumNetEID) As Collection
'creates a collection of network elements that are sample sites
Dim lSiteFCID As Long
Dim pNetElems As INetElements
Dim pFeature As IFeature
Dim pNetworkElem As NetworkElement
'Dim isSite As Boolean
Dim bAddSite As Boolean
Dim pSelCollection As New Collection
Dim elementEID As Long, elementOID As Long, elementClassID As Long, elementSub
As Long

pApp.Caption = "CreateJunctionCollection"

'Get the feature class IDs for the site layer
lSiteFCID = pPoint.FeatureClassID

Set pNetElems = pTraceSolver.CurrentNetwork.Network
pSelection.Reset

'Convert each selected site EID to a NetworkElement object and add to a collection
elementEID = pSelection.Next
bAddSite = False
```

```
While elementEID <> 0
pNetElems.QueryIDs elementEID, esriETJunction, elementClassID, elementOID,
elementSub

'check to see if this is a site feature
If elementClassID = lSiteFCID Then
Set pFeature =
pFeatureClassContainer.ClassByID(elementClassID).GetFeature(elementOID)
bAddSite = True
Else
bAddSite = False
End If

'Add this site to the collection of sites to trace to and from
If bAddSite Then
Set pNetworkElem = New NetworkElement
pNetworkElem.ElementID = elementEID
pNetworkElem.FeatureClassID = elementClassID
pNetworkElem.ObjectID = elementOID
pNetworkElem.SubID = elementSub
pSelCollection.Add pNetworkElem, CStr(elementOID)
End If

bAddSite = False
elementEID = pSelection.Next
Wend

Set CreateJunctionCollection = pSelCollection
End Function
```

ESTIMATE MEAN WATERSHED ATTRIBUTE

```
'bas EstMeanWatershedAtt.bas
'Created by Erin Peterson
'Program is incomplete   Last updated: 11/11/04

'PURPOSE: The purpose of this program is to record the watershed attribute for each
'sample point. This is accomplished by selecting each sample point in the network,
'finding the closest edge and estimating the upstream attribute value using the percent
'distance along the edge, the local attribute value, and the cumulative upstream attribute
'value. If the point happens to fall at a confluence then the program queries the adjacent
'edges, selecting those edges that flow into the junction, and summing their cumulative
'upstream attribute to get the total upstream attribute for each site. This value is reported
'in a new field in the sample sites attribute table.
```

```
'INPUT: Geometric network containing sample sites. It does not matter if the sample sites
'fall directly on the edges.The network should have ONE cumulative upstream attribute
'as a weight value. This program should be run after the AccumCatchArea.bas program,
'which will assign cumulative watershed attributes to each stream segment.

'OUTPUT: A geometric network with a new field in the sample sites feature class that
'contains the total upstream watershed attribute.

'PREPROCESSING: The preprocessing steps for the AccumCatchArea.bas program
'should be completed (contained in the AccumCatchArea_preprocessing.doc file)and the
'AccumCatchArea.bas script should be run. The geometric network should be rebuilt and
'the cumulative catchment attribute should be included as a network weight. At this
'point, the network should be ready as input into the basWatershedAtt.bas program. This
'program is designed to operate on a geometric network that only has ONE weight
'assigned to it.


'******ALL CONSTANTS MUST BE DEFINED BY THE USER
Private Const strEdgeTableName As String = "streams"
Private Const strSiteTableName As String = "sites"
Private Const strSortName As String = "sort"
Private Const strmName As String = "streams"        'name of streams feature class
                                                    'included in geometric network
Private Const siteName As String = "sites"
Private Const fDatasetName As String = "fdataset4"  'name of feature dataset that
                                                    'contains geometric network
Private Const geodatabasePathName As String = "pathname here"
            'pathname to the geodatabase that contains the geometric network
Private Const strTableWorkspace As String = "pathname here" 'workspace where all
                                                            'tables will be saved
Private Const strMeanAttFN As String = "CumMaxT"       'name of new field in site
                                               'table for cumulative upstream HCA area
Private Const strWeightName As String = "AnMaxT" 'name that was given to the weight
                                                 'in the geometric network
Private Const sMeanAttName As String = "CumMaxT"
Private Const dblSnapTol As Double = 50

'************************************************************************
Option Explicit
Private pApp As IApplication
Private pMxDoc As IMxDocument
Private pMap As IMap
Private pActiveView As IActiveView
Private pWorkspaceFactory As IWorkspaceFactory
Private pWorkspace As IWorkspace
```

```
Private pFeatureWorkspace As IFeatureWorkspace
Private pFDataset As IFeatureDataset
Private pNetworkWorkspace As INetworkWorkspace
Private pFeatureClassContainer As IFeatureClassContainer
Private pUID As New esriCore.UID                'ID for UtilityNetworkAnalysisExt
Private pTraceSolver As INetworkAnalysisExt
Private pEditor As IEditor
Private pLine As IFeatureClass                  'stream feature class
Private pPoint As IFeatureClass                 'sample points feature class
Private pJuncSel As IEnumNetEID  'enumeration of all the junctions in the network
Private pEdgeSel As IEnumNetEID  'enumeration of all the edges in the network
Private pGeometricNetwork As IGeometricNetwork  'geometric network loaded into the
                                                'map frame
Private pNetwork As INetwork
Private pEdgeTable As ITable        'attribute table for edges in geometric network
Private pSiteTable As ITable 'attribute table for sample sites in geometric network
Private pSortTable As ITable 'dbf table created to sort edges
Private pAdjacentEdges() As Long   'an array of adjacent edges returned by the
                                   'IForwardStar
Private vWtVals() As Variant 'an array of adjacent edge weights returned by the
IForwardStar
Private bRev() As Boolean          'an array containing the direction of the adjacent
                                   'edges returned by the IForwardStar
Private pNextJuncEID As Long       'next downstream junction EID
Private lngDownstreamEdgeOID As Long   'the downstream edge OID
Private pJuncEID As Long           'variable containing the EID for a junction in the
                                   'network
Private lngCatchAreaFPos As Long  'field position of the cumulative catchment area
                                  '(CumArea_sqM) in the pEdgeTable
Private pJuncCollection As New Collection  'collection of edges that have a Null value
                                           'for CumArea_sqM
Private pNetworkElem As NetworkElement
Private lMeanAttFieldPos As Long, lLocalAttFieldPos As Long


Sub Main()

Set pApp = New AppRef
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pActiveView = pMxDoc.FocusMap

pApp.Caption = "Setting up workspace."
'New AccessWorkspaceFactory is IWorkspaceFactory for personal geodatabase feature
'classes
Set pWorkspaceFactory = New AccessWorkspaceFactory
```

```
Set pWorkspace = pWorkspaceFactory.OpenFromFile(geodatabasePathName, 0)
Set pFeatureWorkspace = pWorkspace  'QI for the IFeatureWorkspace
Set pFDataset = pFeatureWorkspace.OpenFeatureDataset(fDatasetName)
Set pNetworkWorkspace = pFDataset.Workspace

'need feature class container to access feature classes within a feature dataset
Set pFeatureClassContainer = pFDataset
Set pLine = pFeatureClassContainer.ClassByName(strmName)
Set pPoint = pFeatureClassContainer.ClassByName(siteName)
Set pEdgeTable = OpenTable(strEdgeTableName)
Set pSiteTable = OpenTable(strSiteTableName)

Dim pWorkspaceProperties As IWorkspaceProperties
Dim pWkspProp As IWorkspaceProperty
Set pWorkspaceProperties = pWorkspace
Set pWkspProp = pWorkspaceProperties.Property(esriWorkspaceTablePropertyGroup,
esriTablePropCanDeleteField)

pUID = "esricore.UtilityNetworkAnalysisExt"
Set pTraceSolver = pApp.FindExtensionByCLSID(pUID)

'program will stop if there isn't a geometric network loaded
Set pGeometricNetwork = pTraceSolver.CurrentNetwork
Set pNetwork = pGeometricNetwork.Network

'Determine whether CumAttribute field exists in streams layer and store position
lMeanAttFieldPos = pLine.FindField(sMeanAttName)
If lMeanAttFieldPos < 0 Then
MsgBox CStr(pLine.AliasName & " does not contain a Cumulative Attribute field.")
Exit Sub
End If

'add a new field to the Edges attribute table that will hold the proportional influence
Call AddField

'create an enumeration of all junctions found in the network.
Set pJuncSel = GetCurrentEIDs(esriETJunction)
Set pJuncCollection = CreateJunctionCollection(pJuncSel)

Dim z As Integer
Dim lOID As Long
z = 0

'select each junction in the network and find the adjacent edges
For z = 1 To pJuncCollection.Count
Set pNetworkElem = pJuncCollection.Item(z)
```

```
pJuncEID = pNetworkElem.ElementID

If pJuncEID > 0 Then
lOID = pNetworkElem.ObjectID
Call FindUpstreamInfluence(pJuncEID, lOID)
End If
Next
End Sub


Private Function OpenTable(TableName As String) As ITable
'function opens a table and returns it as an ITable using the table name as an input

On Error GoTo Errorhandler
Dim pDatasetname As IDatasetName
Dim pName As IName
Dim pWorkspaceName As IWorkspaceName
Set pWorkspaceName = New WorkspaceName
pWorkspaceName.WorkspaceFactoryProgID = "esricore.AccessWorkspaceFactory"
pWorkspaceName.PathName = geodatabasePathName

'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenTable = pName.Open

Exit Function 'exit to avoid error handler
Errorhandler:
Set OpenTable = Nothing
End Function



Private Function GetCurrentEIDs(eElementType As esriElementType) As IEnumNetEID
'return an enumeration of network EID's for a given element type

pApp.Caption = "GetCurrentEIDs"
Dim pNetwork As INetwork   'holds a reference to the current network
Dim pEnumNetEID As IEnumNetEID   'stores a list of network EID's

Set pNetwork = pGeometricNetwork.Network
Set pEnumNetEID = pNetwork.CreateNetBrowser(eElementType)
```

```
'return the enumeration
Set GetCurrentEIDs = pEnumNetEID
End Function


Private Sub FindUpstreamInfluence(pJuncEID As Long, lOID As Long)
Dim lngAdjEdgeCount As Long
Dim adjEdges() As Long
Dim blnRevOr() As Boolean
Dim adjWeights() As Variant

'Set up the IForwardStar Object
Dim pNetWeightNothing As INetWeight
Dim pNetArea As INetWeight
Dim pNetSchema As INetSchema
Dim pForwardStar As IForwardStar
Dim pNetElems As INetElements
Dim adjEdgesCount As Long
Dim NextEdgesCount As Long
Dim adjEdgeEID As Long, AdjEdgeWeight As Variant, bDirection As Boolean
Dim edgeElement As Variant
Dim dblTotalArea As Double, dblLocalArea As Double
Dim q As Integer
Dim edgeWeight As Double
Dim lngEdgeProp As Double
Dim elementEID As Long, elementClassID As Long, elementOID As Long, elementSub
As Long
Dim lEdgeEID As Long, lEdgeClassID As Long, lEdgeOID As Long, lEdgeSub As Long
Dim pPointToEID As IPointToEID 'used to find the closest edge to the site
Dim pNewLoc As IPoint 'New location of site
Dim dpercent As Double 'percent of distance along the nearest edge (From node --> To
node).
Dim spercent As Single 'percent distance along the edge stored as a single
Dim pFeature As esriCore.IFeature
Dim pSite As IPoint 'IPoint representing site
Dim pEdgeElem As NetworkElement 'Closest edge
Dim pEdgeRow As IRow
Dim calcCount As Integer, lEdgeID As Long

Set pNetSchema = pNetwork
Set pNetArea = pNetSchema.WeightByName(strWeightName)
Set pNetWeightNothing = Nothing

'*************************************************************************
'In this case, you are only interested in edge weights on your network, so
```

```
'you can set the other weights using the NetWeightNothing object (pNetWeight)
Set pForwardStar = pNetwork.CreateForwardStar(True, pNetWeightNothing, pNetArea,
pNetArea, pNetWeightNothing)

pForwardStar.FindAdjacent 0, pJuncEID, adjEdgesCount

'erase the old arrays
Erase adjEdges()
Erase blnRevOr()
Erase adjWeights()
'reallocate memory for the new arrays
ReDim adjEdges(adjEdgesCount) As Long
ReDim blnRevOr(adjEdgesCount) As Boolean
ReDim adjWeights(adjEdgesCount) As Variant

pForwardStar.QueryAdjacentEdges adjEdgesCount, adjEdges(0), blnRevOr(0),
adjWeights(0)

'****************if the point falls at the end of one or more edges
If adjEdgesCount > 0 Then
dblTotalArea = 0
q = 0
calcCount = 0
lEdgeID = 0

'look at each adjacent edge, determine whether it flows into the junction, and if so add it's
' upstream cumulative catchment area
For Each edgeElement In adjEdges
lEdgeID = edgeElement

If lEdgeID <> 0 Then
'if it is an upstream edge from junction
If blnRevOr(q) = False Then
edgeWeight = adjWeights(q)
dblTotalArea = dblTotalArea + edgeWeight
calcCount = calcCount + 1
End If

q = q + 1
End If
Next

'**********If the point does not fall on an edge
Else
calcCount = 0
```

```
Set pNetElems = pTraceSolver.CurrentNetwork.Network
pNetElems.QueryIDs pJuncEID, esriETJunction, elementClassID, elementOID,
elementSub
'report the results in the edge attribute table

'get the site using it's OID
Set pFeature = pPoint.GetFeature(lOID)
Set pSite = pFeature.Shape

'Find closest edge to the site
Set pPointToEID = New PointToEID

With pPointToEID
Set .GeometricNetwork = pGeometricNetwork
Set .SourceMap = pMap
.SnapTolerance = dblSnapTol  ' set a snap tolerance in map units in the constants section
.GetNearestEdge pSite, lEdgeEID, pNewLoc, dpercent
End With

If lEdgeEID <> 0 Then
'convert the edge EID to a feature class ID, OID, and sub ID
pNetElems.QueryIDs lEdgeEID, esriETEdge, lEdgeClassID, lEdgeOID, lEdgeSub

Set pEdgeRow = pEdgeTable.GetRow(lEdgeOID)
dblTotalArea = pEdgeRow.Value(lMeanAttFieldPos)
calcCount = calcCount + 1
End If
End If

'calculate the mean of upstream edges
If calcCount <> 0 Then
dblTotalArea = dblTotalArea / calcCount
Call RecordResults(lOID, dblTotalArea)
End If
End Sub



Private Sub AddField()
'this subroutine checks the edges feature class in the geometric network to see if there
'is a field for the proportional downstream influence (strMeanAttFN) for each segment.
'If so, it deletes it and adds a new field. If not, it just adds the new field

Dim pAddField As IField
Dim pFieldEdit As IFieldEdit
```

```
'create new field
Set pAddField = New Field
Set pFieldEdit = pAddField

With pFieldEdit
.AliasName = strMeanAttFN
.Type = esriFieldTypeDouble
.Length = 20
.Name = strMeanAttFN
.Editable = True
.IsNullable = True
End With

Set pAddField = pFieldEdit
Set pFieldEdit = Nothing

If pSiteTable.FindField(strMeanAttFN) = -1 Then
pSiteTable.AddField pAddField
Else

Dim pFields As IFields
Dim pDeleteField As IField

Set pFields = pSiteTable.Fields
lngCatchAreaFPos = pFields.FindField(strMeanAttFN)
Set pDeleteField = pFields.Field(lngCatchAreaFPos)

pSiteTable.DeleteField pDeleteField
pSiteTable.AddField pAddField
End If

lngCatchAreaFPos = pFields.FindField(strMeanAttFN)
Exit Sub
Set pSiteTable = OpenTable(strSiteTableName)
End Sub



Private Sub RecordResults(intJuncOID As Long, dblTotalArea As Double)
'this function records the proportional influence of each edge in the edges attribute table

'Get the editor reference and set pEditor
Call GetEditorReference
'create an edit operation enabling an undo for this operation
pEditor.StartOperation
```

```
'create cursor containing all records in the edges table
Dim pRecordRow As IRow

 If intJuncOID > 0 Then
Set pRecordRow = pSiteTable.GetRow(intJuncOID)
'Debug.Print pRow.Value(2)
pApp.Caption = "Looking for record results row"

'add this area to the cumulative upstream catchment area field position
pRecordRow.Value(lngCatchAreaFPos) = dblTotalArea
pRecordRow.Store
End If

pEditor.StopOperation ("RecordResults")
pEditor.StopEditing True
Set pRecordRow = Nothing
End Sub


Public Sub GetEditorReference()
'creates an editor reference needed to start an editing session
Dim pID As esriCore.UID

Set pID = New esriCore.UID
pID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pID)
StartEditing pEditor
End Sub


Public Sub StartEditing(pEditor As IEditor)
'starts an editing session
Dim pDataset As IDataset
Dim pDatasetWorkspace As IWorkspace

If Not pEditor.EditState = esriStateNotEditing Then Exit Sub
Set pDataset = pPoint
Set pDatasetWorkspace = pDataset.Workspace
pEditor.StartEditing pDataset.Workspace
End Sub



Private Sub Reset()
'releases memory of objects
pEditor.StopEditing True
```

```
Set pEditor = Nothing
End Sub


Private Function CreateJunctionCollection(pSelection As IEnumNetEID) As Collection
'creates a collection of network elements that are sample sites
Dim lSiteFCID As Long
Dim pNetElems As INetElements
Dim pFeature As IFeature
Dim pNetworkElem As NetworkElement
Dim bAddSite As Boolean
Dim pSelCollection As New Collection
Dim elementEID As Long, elementOID As Long, elementClassID As Long, elementSub
As Long
pApp.Caption = "CreateJunctionCollection"

'Get the feature class IDs for the site layer
lSiteFCID = pPoint.FeatureClassID

Set pNetElems = pTraceSolver.CurrentNetwork.Network

pSelection.Reset

'Convert each selected site EID to a NetworkElement object and add to a collection
elementEID = pSelection.Next
bAddSite = False
While elementEID <> 0
pNetElems.QueryIDs elementEID, esriETJunction, elementClassID, elementOID,
elementSub

'check to see if this is a site feature
If elementClassID = lSiteFCID Then
Set pFeature =
pFeatureClassContainer.ClassByID(elementClassID).GetFeature(elementOID)
bAddSite = True
Else
bAddSite = False
End If

'Add this site to the collection of sites to trace to and from
If bAddSite Then
Set pNetworkElem = New NetworkElement
pNetworkElem.ElementID = elementEID
pNetworkElem.FeatureClassID = elementClassID
pNetworkElem.ObjectID = elementOID
pNetworkElem.SubID = elementSub
```

```
pSelCollection.Add pNetworkElem, CStr(elementOID)
End If

bAddSite = False
elementEID = pSelection.Next
Wend
Set CreateJunctionCollection = pSelCollection
End Function
```

OUTPUT MATRIX

```
'outputMatrix.bas file
'file is complete last updated 2/23/04
'Created by Erin Peterson

'PURPOSE: The purpose of this program is to output an NxN matrix of values for sample
'sites. The values represent relationships between each site and every other site. For
'example, proportional influence of one site on another or hydrologic distances between
'sites. The output file is saved as a comma delimited text file.

'INPUT: A feature class containing the sample sites. This can be contained in the
'geometric network or can exist as a stand alone feature class in a feature dataset. The
'input is a dbf file created by either the calcProp.bas program or the calcHydroDist.bas
'program.

'OUTPUT: An NxN matrix in comma delimited text format containing values that
'represent a relationship between each sample site and every other sample site. Rows are
'To OID and columns are From OID.

'PREPROCESSING: None.

'******ALL CONSTANTS MUST BE DEFINED BY THE USER

Private Const strTableWorkspace As String = "pathname here" 'the workspace where
                                                            'input table resides
Private Const strTableName As String = "all_weight"         'input table name
Private Const siteName As String = "sites"                  'name of sites feature class
Private Const geodatabasePathName As String = "pathname here" 'geodatabase pathname
Private Const fDatasetName As String = "fdataset"           'name of feature dataset
Private Const sTextFileName = "pathname here" 'location where text files will be saved
                                'do not save text files in same directory as .dbf files
'*****************************************************************************
Private pApp As IApplication
Private pMxDoc As IMxDocument
```

```
Private pMap As IMap
Private pActiveView As IActiveView
Private pWorkspaceFactory As IWorkspaceFactory
Private pTextWkspFactory As IWorkspaceFactory
Private pWorkspace As IWorkspace
Private pTextWksp As IWorkspace
Private pFDataset As IFeatureDataset
Private pFeatureWorkspace As IFeatureWorkspace
Private pFeatureClassContainer As IFeatureClassContainer

Option Explicit

Private Function WriteMatrix()
Dim pTable As ITable
Dim pRow As IRow
Dim pEnumLayers As IEnumLayer
Dim pSites As IFeatureClass
Dim i As Integer
Dim pCursor As ICursor
Dim lFileID As Long
Dim lngToCounter As Long
Dim lngFromCounter As Long
Dim lngFeatCount As Long
Dim sText As String
Dim dblDistance() As Double

Set pApp = New AppRef
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pActiveView = pMxDoc.FocusMap
Set pWorkspaceFactory = New AccessWorkspaceFactory
Set pWorkspace = pWorkspaceFactory.OpenFromFile(geodatabasePathName, 0)
Set pFeatureWorkspace = pWorkspace
Set pFDataset = pFeatureWorkspace.OpenFeatureDataset(fDatasetName)
Set pTextWkspFactory = New TextFileWorkspaceFactory
Set pFeatureClassContainer = pFDataset
Set pSites = pFeatureClassContainer.ClassByName(siteName) 'sites feature class

'number of sites
lngFeatCount = pSites.FeatureCount(Nothing)

'if the text file already exists, delete it
If Len(Dir(sTextFileName)) > 0 Then Kill sTextFileName

'Create and open the text file
lFileID = FreeFile
```

```
Open sTextFileName For Output As #lFileID

'open the .dbf export table
Set pTable = OpenTable(strTableWorkspace, strTableName)

'if it isn't returned, then exit program
If pTable Is Nothing Then
Exit Function
End If

'populate the cursor with the rows in the export table
Set pCursor = pTable.Search(Nothing, False)
Set pRow = pCursor.NextRow
lngToCounter = 1

'do this for each possible ToNode
Do Until lngToCounter = lngFeatCount + 1
lngFromCounter = 1

If Not pRow Is Nothing Then
'clear the distance array when you start with a new ToNode
Erase dblDistance()
ReDim dblDistance(lngFeatCount)

'if the row is not part of the header
If pRow.Value(3) > 0 Then

'do this for each possible FromNode
Do Until lngFromCounter = lngFeatCount + 1

If Not pRow Is Nothing Then
If pRow.Value(1) = lngFromCounter And pRow.Value(2) = lngToCounter Then
'write the value to the distance array
dblDistance(lngFromCounter - 1) = pRow.Value(3)

Set pRow = pCursor.NextRow
lngFromCounter = lngFromCounter + 1
Else
'if FromNode doesn't flow into ToNode, record a 0 for distance
dblDistance(lngFromCounter - 1) = 0
lngFromCounter = lngFromCounter + 1
End If
Else
Exit Do
End If
Loop
```

```
Dim z As Integer
z = 1
For z = 1 To lngFeatCount
Write #lFileID, dblDistance(z - 1);
Next
Write #lFileID,

'add one to the ToCounter
lngToCounter = lngToCounter + 1
'If pRow.Value(3) = 0, it is part of the header. Go on to next row
Else
Set pRow = pCursor.NextRow
End If

'If you have reached the end of the table...
Else
Exit Do
End If
Loop

'if the end of the table has been reached before data has been recorded for all of the
'ToNodes, fill in the remaining ToNode rows with zeros in the text file
FinishTable: If lngToCounter < lngFeatCount + 1 Then
Erase dblDistance()
ReDim dblDistance(lngFeatCount)

z = 1
For z = 1 To lngFeatCount
Write #lFileID, dblDistance(z - 1);
Next

Write #lFileID,
lngToCounter = lngToCounter + 1
GoTo FinishTable
End If

'close the text file and release memory
Close #lFileID
Set pWorkspaceFactory = Nothing
End Function


Private Function OpenTable(strTableWorkspace As String, TableName As String) As
ITable
'open a table found in a shapefile workspace
```

```
On Error GoTo Errorhandler
Dim pShapefileWorkspaceName As IWorkspaceName
Dim pDatasetname As IDatasetName
Dim pName As IName

'Create the workspace name object
Set pShapefileWorkspaceName = New WorkspaceName
pShapefileWorkspaceName.PathName = strTableWorkspace
pShapefileWorkspaceName.WorkspaceFactoryProgID =
"esricore.shapefileworkspacefactory.1"

'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pShapefileWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenTable = pName.Open
Exit Function 'exit to avoid error handler
Errorhandler:
Set OpenTable = Nothing
End Function
```

CALCULATE HYDROLOGIC DISTANCE PROGRAM

```
'created by Erin Peterson
'Last updated 5/3/04
'this program is complete

' This purpose of this program is to calculate the direction (upstream vs. downstream) and
'hydrologic distance between sample sites. Only the downstream distance is recorded in
'the distance table, but the upstream and total hydrologic distances are also stored in the
'table. For instance, the downstream hydrologic distance from site A to site B is recorded
'in the table. The upstream hydrologic distance from site A to site B is equal the
'downstream hydrologic distance from site B to site A. The total hydrologic distance
'between sites A and B is equal to the downstream hydrologic distance from site A to Site
'B + the downstream hydrologic distance from site B to site A. The downstream
'distances are stored in a dbf file, which is converted to an NxN matrix of downstream
'distances and output as a comma delimited text file.

'INPUT: Geometric network with sample points included as network junctions. Create
'two length weights in the geometric network: downstream length and upstream length.
```

```
'Set the downstream length equal to the shape_length attribute. Create the upstream
'length weight, but do not assign an attribute.This allows the program to assign a length
'of zero to upstream segments.
'
'OUTPUT: Two output formats: 1) A .dbf file containing downstream hydrologic
'distances between sample sites. Inserts a header that contains the To feature in field 1
'and the row that it begins in field 2. The third field will contain a zero to indicate that
'this is the header. Header ends with To feature: -9. 2) An NxN matrix of downstream
'distances between every sample site and every other sample site.

'PRE-PROCESSING STEPS: Create a geometric network with sample points included as
'network junctions. Create two length weights in the geometric network: downstream
'length and upstream length. Set the downstream length equal to the shape_length
'attribute. Create the upstream length weight, but do not assign an attribute. This allows
'the program to assign a length of zero to upstream segments.

'****** NetworkElement.cls File must be included for this program to run

'******ALL CONSTANTS MUST BE DEFINED BY THE USER

Private Const sSiteIDFieldName As String = "Id"    'Name of attribute field in sites
                                        'feature class indicating whether it is a site
Private Const sLength As String = "Shape_Length"   'the attribute field name that
                                        'represents the downstream length value
Private Const sDSWeightName As String = "DSLength" 'Name given to the downstream
                                'weight when the geometric network was created
Private Const sUSWeightName As String = "USLength" 'Name given to the upstream
                                'weight when the geometric network was created
Private Const strTmpDistTableName As String = "tmpDistTable"  'name of the
                        'temporary distance table (dbf) the program will create
Private Const strDistName As String = "distance" ' Name of the output distance (.dbf)
                                        'table that the program produces
Private Const strmName As String = "streams_2"    'name of streams feature class
                                        'included in geometric network
Private Const siteName As String = "sites_2"       'name of sample sites feature class
                        'included in geometric network these are the points the
                        'program will measure from
Private Const fDatasetName As String = "fdataset" 'name of feature dataset that contains
                                        'geometric network
Private Const geodatabasePathName As String = "pathname here"
                'pathname to the geodatabase that contains the geometric network
Private Const strTableWorkspace As String = "pathname here" 'workspace where all
                                        'tables will be saved
Private Const dblSnapTol As Double = 0.5   'snap tolerance for snapping sample sites to
                                        'edges
```

```
Private Const sTextFileName = "pathname here"     'location where text files will be
                    'saved **do not save text files in same directory as .dbf files

Option Explicit

Private pApp As IApplication
Private pMxDoc As IMxDocument
Private pMap As IMap
Private pActiveView As IActiveView
Private pWorkspaceFactory As IWorkspaceFactory
Private pTextWkspFactory As IWorkspaceFactory
Private pFeatureWorkspace As IFeatureWorkspace
Private pWorkspace As IWorkspace
Private pTextWksp As IWorkspace
Private pFeatureClassContainer As IFeatureClassContainer
Private pFDataset As IFeatureDataset
Private pTraceSolver As INetworkAnalysisExt
Private pEditor As IEditor
Private pLine As IFeatureClass                'stream feature class
Private pPoint As IFeatureClass               'sample sites feature class
Private pEdgeTable As ITable                  'streams attribute table
Private pJuncSel As IEnumNetEID               'enumeration of junctions - used many times
                                              'for different junction sets
Private pEdgeSel As IEnumNetEID 'enumeration of edges selected in the trace function
Private pGeometricNetwork As IGeometricNetwork        'geometric network
Private lSiteIDFieldPos As Long                       'site ID field position in table
Private pJuncCollection As Collection          'collection of all junctions in the network
Private lSegmentLengthFieldPos As Long        'length field position in streams feature class
                                              'table


Sub Main()

Dim pNetworkWorkspace As INetworkWorkspace
Dim pUID As New esriCore.UID
Dim pFlags As INetworkAnalysisExtFlags
Set pApp = New AppRef
Set pMxDoc = pApp.Document
Set pMap = pMxDoc.FocusMap
Set pActiveView = pMxDoc.FocusMap
Set pWorkspaceFactory = New AccessWorkspaceFactory
Set pWorkspace = pWorkspaceFactory.OpenFromFile(geodatabasePathName, 0)
Set pFeatureWorkspace = pWorkspace  'QI for the IFeatureWorkspace
Set pFDataset = pFeatureWorkspace.OpenFeatureDataset(fDatasetName)
Set pNetworkWorkspace = pFDataset.Workspace
```

```
'need feature class container to access feature classes within a feature dataset
Set pFeatureClassContainer = pFDataset
Set pLine = pFeatureClassContainer.ClassByName(strmName) 'streams feature class
Set pPoint = pFeatureClassContainer.ClassByName(siteName)  'sites feature class
Set pEdgeTable = OpenGeodatabaseTable(strmName) 'streams attribute table

pUID = "esricore.UtilityNetworkAnalysisExt"
Set pTraceSolver = pApp.FindExtensionByCLSID(pUID)

'program will stop if there isn't a geometric network loaded in the map document
Set pGeometricNetwork = pTraceSolver.CurrentNetwork
Set pFlags = pTraceSolver

'check to see if the sites are in the network
If Not SitesAreInNetwork(pPoint.FeatureClassID) Then
MsgBox "Affimative: " & pPoint.AliasName & " is not in the specified network.",
vbCritical, "Isolation Trace"
Exit Sub
End If

'Determine whether the id field exists in site layer and store position - you will need
'this to create a collection of sites
lSiteIDFieldPos = pPoint.FindField(sSiteIDFieldName)
If lSiteIDFieldPos < 0 Then
MsgBox CStr(pPoint.AliasName & " does not contain an ID field named, " &
sSiteIDFieldName)
Exit Sub
End If

'Determine whether length field exists in streams layer and store position
lSegmentLengthFieldPos = pLine.FindField(sLength)
If lSegmentLengthFieldPos < 0 Then
MsgBox CStr(pLine.AliasName & " does not contain a Length field.")
Exit Sub
End If

'return an enumeration of network EID's for ALL network junctions
Set pJuncSel = GetCurrentEIDs(esriETJunction)

'Sort though the enumeration of network EID's and create a collection of sites
Set pJuncCollection = CreateJunctionCollection(pJuncSel)

'create a temporary table that will be used to store the downstream hydrologic distances
'as the traces are performed
Call CreateTable(strTmpDistTableName)
```

```
'set flow direction of network against the digitized direction
Call SetFlowDirection

'Perform the trace. Downstream distances are stored in the temporary distance table
Call PerformTrace(pFlags)

'sort the temporary table and output the final distance dbf table
Call SortTable(strTmpDistTableName, strDistName)

'delete the temporary table
Call DeleteTable(strTmpDistTableName)

'create an NxN matrix of downstream hydrologic distances in comma delimited text
format
Call WriteMatrix
pApp.Caption = "The Program is complete!"
End Sub


Private Function SitesAreInNetwork(lFeatureClassID As Long) As Boolean
'given a FeatureClassID, this function tests to see if the corresponding feature class
'is contained in the current geometric network

Dim pGeometricNetwork As IGeometricNetwork
Set pGeometricNetwork = pTraceSolver.CurrentNetwork

pApp.Caption = "Check if sites are in network."

'Return false if this is just the orphan junction class
If lFeatureClassID = pGeometricNetwork.OrphanJunctionFeatureClass.FeatureClassID
Then
Exit Function
End If

'Get an enumeration of the simple junction feature classes in the network
Dim pEnumFeatureClass As IEnumFeatureClass
Set pEnumFeatureClass = pGeometricNetwork.ClassesByType(esriFTSimpleJunction)

Dim pFeatureClass As IFeatureClass
SitesAreInNetwork = False
pEnumFeatureClass.Reset
Set pFeatureClass = pEnumFeatureClass.Next
Do Until pFeatureClass Is Nothing
If pFeatureClass.FeatureClassID = lFeatureClassID Then
SitesAreInNetwork = True
Exit Function
```

```
End If
Set pFeatureClass = pEnumFeatureClass.Next
Loop
Exit Function
End Function


Private Function GetCurrentEIDs(eElementType As esriElementType) As IEnumNetEID
'return an enumeration of network EID's for a given element type
pApp.Caption = "GetCurrentEIDs"

Dim pNetwork As INetwork  'holds a reference to the current network
Dim pEnumNetEID As IEnumNetEID  'stores a list of network EID's
Set pNetwork = pGeometricNetwork.Network
Set pEnumNetEID = pNetwork.CreateNetBrowser(eElementType)

'return the enumeration
Set GetCurrentEIDs = pEnumNetEID
End Function

Private Function CreateJunctionCollection(pSelection As IEnumNetEID) As Collection
'sorts through an enumeration of network junctions and creates a collection of sample
'sites the NetworkElement.cls file is necessary to run this function

Dim lSiteFCID As Long  'feature class ID for sites feature class
Dim pNetElems As INetElements 'used to convert between EIDs and OIDs
Dim pFeature As IFeature
Dim pNetworkElem As NetworkElement
Dim bAddSite As Boolean    'boolean indicating whether to add the element to the site
                           'collection
Dim pSelCollection As New Collection 'collection of sites
Dim elementEID As Long, elementOID As Long, elementClassID As Long, elementSub
As Long

'Get the feature class IDs for the site layer
lSiteFCID = pPoint.FeatureClassID
Set pNetElems = pTraceSolver.CurrentNetwork.Network
pSelection.Reset

'Convert each selected site EID to a NetworkElement object and add to the site collection
elementEID = pSelection.Next
bAddSite = False
While elementEID <> 0
pNetElems.QueryIDs elementEID, esriETJunction, elementClassID, elementOID,
elementSub
```

```
'check to see if this is a site feature
If elementClassID = lSiteFCID Then
Set pFeature =
pFeatureClassContainer.ClassByID(elementClassID).GetFeature(elementOID)

'If site ID field exists...
If lSiteIDFieldPos > -1 Then
bAddSite = True
End If
'If site ID field does not exist
ElseIf lSiteIDFieldPos < 0 Then
bAddSite = False
End If

'Add this site to the collection of sites to trace to and from
If bAddSite Then
Set pNetworkElem = New NetworkElement
pNetworkElem.ElementID = elementEID
pNetworkElem.FeatureClassID = elementClassID
pNetworkElem.ObjectID = elementOID
pNetworkElem.SubID = elementSub
pSelCollection.Add pNetworkElem, CStr(elementOID)
End If

bAddSite = False
elementEID = pSelection.Next
Wend

Set CreateJunctionCollection = pSelCollection
End Function


Private Sub PerformTrace(pFlags As INetworkAnalysisExtFlags)
' perform trace from and to each of the sites, which are stored as elements in the
' pJuncCollection. The NetworkElement.cls file is needed to run this subroutine.

Dim iLoop1 As Integer       'iLoop represents index number of 1st site in collection
Dim iLoop2 As Integer       'iLoop2 represents index # of 2nd site in collection
Dim pPathTrace As ITraceFlowSolver        'path between sample sites
Dim pNetElems As INetElements             'used to convert between EIDs and OIDs of
                                          'network elements
Dim pEnumNetEIDBuilder As IEnumNetEIDBuilder
Dim pPointToEID1 As IPointToEID, pPointToEID2 As IPointToEID 'used to find the
                                          'closest edge to the site
Dim lEdgeEID1 As Long, lEdgeEID2 As Long      'Edge EIDs for from site
                                          '(lEdgeEID1) and the to site (lEdgeEID2)
```

```
Dim pNewLoc1 As IPoint, pNewLoc2 As IPoint      'New location of from site (1) and to
                                                'site (2)
Dim dpercent1 As Double, dpercent2 As Double    'percent of distance along the nearest
                                                'edge (From node --> To node).
Dim percent1 As Single, percent2 As Single      'percent distance along the edge
                                                'stored as a single
Dim pFeature1 As esriCore.IFeature, pFeature2 As esriCore.IFeature
Dim elementOID1 As Long, elementClassID1 As Long, elementSub1 As Long 'From
edge OID, FCID, and SubID
Dim elementOID2 As Long, elementClassID2 As Long, elementSub2 As Long 'To edge
OID, FCID, and SubID
Dim lOID1 As Long, lOID2 As Long 'OID for from site (1) and to site (2)
Dim dblTraceLength As Double 'downstream distance of path between sites
Dim pSite1 As IPoint, pSite2 As IPoint 'IPoint representing from site and to site
Dim pNetworkElem1 As NetworkElement, pNetworkElem2 As NetworkElement 'from
site and to site
Dim pEdgeElem1 As NetworkElement, pEdgeElem2 As NetworkElement 'from edge and
to edge
Dim dblSegmentLength As Double   'edge length - used with dpercent variable to
                                 'calculate the hydrologic distance when a site lies
                                 'midway along an edge
Dim pRow As IRow
Dim blnDownstream As Boolean     'boolean indicating whether the path can be found
                                 'downstream from the sample site
Set pNetElems = pTraceSolver.CurrentNetwork.Network

iLoop1 = 1                       'from site

'trace to each of the sites (pJuncCollection = all the sites)
For iLoop1 = 1 To pJuncCollection.count

'get the OID for the site
Set pNetworkElem1 = pJuncCollection.Item(iLoop1)
lOID1 = pNetworkElem1.ObjectID

'get the site using it's OID
Set pFeature1 = pPoint.GetFeature(lOID1)
Set pSite1 = pFeature1.Shape

'Find closest edge to the site and the percent distance along the edge
Set pPointToEID1 = New PointToEID
With pPointToEID1
Set .GeometricNetwork = pGeometricNetwork
Set .SourceMap = pMap
.snapTolerance = dblSnapTol   'set a snap tolerance in map units in the constants section
.GetNearestEdge pSite1, lEdgeEID1, pNewLoc1, dpercent1
```

```
End With

'convert the edge EID to a feature class ID, OID, and sub ID
pNetElems.QueryIDs lEdgeEID1, esriETEdge, elementClassID1, elementOID1,
elementSub1

'create edge network element
Set pEdgeElem1 = New NetworkElement
pEdgeElem1.ElementID = lEdgeEID1
pEdgeElem1.FeatureClassID = elementClassID1
pEdgeElem1.ObjectID = elementOID1
pEdgeElem1.SubID = elementSub1

iLoop2 = 1 'to site

'Begin looking for valid paths between the from site and a to site
For iLoop2 = 1 To pJuncCollection.count

'reset the trace length
dblTraceLength = 0

'get the OID for the to site
Set pNetworkElem2 = pJuncCollection.Item(iLoop2)
lOID2 = pNetworkElem2.ObjectID

'make sure that to and from site are not the same
If lOID1 = lOID2 Then GoTo MoveToNextSite

'get the site using it's OID
Set pFeature2 = pPoint.GetFeature(lOID2)
Set pSite2 = pFeature2.Shape

'Find closest edge to the site and the percent distance along the edge
Set pPointToEID2 = New PointToEID
With pPointToEID2
Set .GeometricNetwork = pGeometricNetwork
Set .SourceMap = pMap
.snapTolerance = dblSnapTol   ' set a snap tolerance in map units
.GetNearestEdge pSite2, lEdgeEID2, pNewLoc2, dpercent2
End With

'convert the edge EID to a feature class ID, OID, and sub ID
pNetElems.QueryIDs lEdgeEID2, esriETEdge, elementClassID2, elementOID2,
elementSub2

'create edge network element
```

```
Set pEdgeElem2 = New NetworkElement
pEdgeElem2.ElementID = lEdgeEID2
pEdgeElem2.FeatureClassID = elementClassID2
pEdgeElem2.ObjectID = elementOID2
pEdgeElem2.SubID = elementSub2

'store the percent distance along the edge as a single
percent1 = dpercent1
percent2 = dpercent2

'set flags at the two sites and get ready for the trace
Set pPathTrace = ModifiedSetup(pFlags, pEdgeElem1, pEdgeElem2, percent1, percent2)
pApp.Caption = "PerformTrace"

'Make sure the solver came back from the ModifiedSetup
If pPathTrace Is Nothing Then
Exit Sub
End If

'allocate memory for the costs array returned by the find path
Dim pCosts() As Variant
Erase pCosts()
ReDim pCosts(2)

'perform the trace - shortest distance between the two sites
pPathTrace.FindPath esriFMConnected, esriSPObjFnMinSum, pJuncSel, pEdgeSel, 1,
pCosts(0)

'set the length of the trace
dblTraceLength = pCosts(0)

'If the two sites are located on the same edge
If elementOID1 = elementOID2 Then
dblTraceLength = dblTraceLength * Abs(dpercent1 - dpercent2)
GoTo RecordResults

' if the sites are located on different edges, use the percentages to calculate
' the length of beginning and end edges and to recalculate the total length
Else

'check to see whether the path moves upstream or downstream from the 1st site
blnDownstream = IsPathDownstream(lEdgeEID1)
'if the path moves downstream from the 1st site
If blnDownstream = True Then
'get the total length of the edge that the site lies on
Set pRow = pEdgeTable.GetRow(elementOID1)
```

```
dblSegmentLength = pRow.Value(lSegmentLengthFieldPos)

'calculate the percentage of the edge that is upstream from the site
'and subtract it from the total distance between sites
dblTraceLength = dblTraceLength - (dblSegmentLength * (1 - dpercent1))

Else
' no calculations are necessary if it moves upstream from the first site
' because we don't record these lengths
End If

'check to see whether the path moves towards the 2nd site from the upstream or
'or downstream direction
blnDownstream = IsPathDownstream(lEdgeEID2)

'if the path is found downstream from the 2nd site
If blnDownstream = True Then
'no calculations are necessary if it moves upstream the second site
'because we don't record these lengths

'if the path is found upstream from the 2nd site (moving downstream towards the site)
Else
'get the total length of the edge that the site lies on
Set pRow = pEdgeTable.GetRow(elementOID2)
dblSegmentLength = pRow.Value(lSegmentLengthFieldPos)

'calculate the percentage of the edge that is downstream from the site
'and subtract it from the total distance between sites
dblTraceLength = dblTraceLength - (dblSegmentLength * dpercent2)
End If
End If

'record the downstream distance
RecordResults:   Call RecordResults(lOID1, lOID2, dblTraceLength)

'reset the trace
Set pPathTrace = Nothing
MoveToNextSite:  Next iLoop2
Next iLoop1
End Sub


Private Function ModifiedSetup(pFlags As INetworkAnalysisExtFlags, pEdgeElem1 As
NetworkElement, pEdgeElem2 As NetworkElement, percent1 As Single, percent2 As
Single) As ITraceFlowSolver
'Routine for setting the Network up for tracing between two sites
```

```
'clear all existing network flags
Set pFlags = pTraceSolver
pFlags.ClearFlags

Dim pNetwork As INetwork
Dim pUtilityNetwork As IUtilityNetwork
Set pNetwork = pGeometricNetwork.Network
Set pUtilityNetwork = pNetwork

' Establish the trace flow solver
Dim pTraceFlowSolver As ITraceFlowSolver
Dim pNetSolver As INetSolver
Dim pNetSolverWeights As INetSolverWeights
Dim pNetSchema As INetSchema
Dim pNetDSLength As INetWeight, pNetUSLength As INetWeight
Set pTraceFlowSolver = New TraceFlowSolver
Set pNetSolver = pTraceFlowSolver
Set pNetSolver.SourceNetwork = pNetwork
Set pNetSolverWeights = pTraceFlowSolver

Dim pEdgeFlags(1) As IEdgeFlag
Dim pEdgeFlag1 As IEdgeFlag, pEdgeFlag2 As IEdgeFlag
Dim pNetFlag(1) As INetFlag
'Set the first edge flag for path finding
Set pNetFlag(0) = New EdgeFlag
Set pEdgeFlag1 = pNetFlag(0)

pEdgeFlag1.Position = percent1
pNetFlag(0).UserClassID = pEdgeElem1.FeatureClassID
pNetFlag(0).UserID = pEdgeElem1.ObjectID
pNetFlag(0).UserSubID = pEdgeElem1.SubID
pNetFlag(0).Label = "Origin"

'Set the second edge flag for path finding
Set pNetFlag(1) = New EdgeFlag
Set pEdgeFlag2 = pNetFlag(1)
pEdgeFlag2.Position = percent2
pNetFlag(1).UserClassID = pEdgeElem2.FeatureClassID
pNetFlag(1).UserID = pEdgeElem2.ObjectID
pNetFlag(1).UserSubID = pEdgeElem2.SubID
pNetFlag(1).Label = "Origin"
'create the array of edge flags
Set pEdgeFlags(0) = pNetFlag(0)
Set pEdgeFlags(1) = pNetFlag(1)
```

```
'Add the TraceFlowSolver object
pTraceFlowSolver.PutEdgeOrigins 2, pEdgeFlags(0) 'pass # of junction origins
                'and then pass the name of the first junction origin set upstream and
                'downstream lengths as the weights (cost fields) to solve on
Set pNetSchema = pNetwork
Set pNetDSLength = pNetSchema.WeightByName(sDSWeightName)
Set pNetUSLength = pNetSchema.WeightByName(sUSWeightName)
Set pNetSolverWeights = pTraceFlowSolver          'the upstream and downstream
                                        'weights are set up backwards because the NHD data is
                                        'digitized against flow
Set pNetSolverWeights.ToFromEdgeWeight = pNetDSLength

'Pass back the trace solver.
Set ModifiedSetup = pTraceFlowSolver
Exit Function
End Function


Public Function CreateTable(sName As String) As ITable
'create a new table - if the strType is neigh then create a neighborhood table, if strType
'is dist, then create a distance table

Dim pWorkspaceFactory2 As IWorkspaceFactory
Dim pWorkspace2 As IWorkspace
Dim pFeatureWorkspace As IFeatureWorkspace
Dim pDataset As IDataset
Dim pNetworkWorkspace As INetworkWorkspace
Dim pED As IEnumDataset 'enumeration of datasets in workspace
Dim pOutDSName As IDatasetName 'output dataset name
pApp.Caption = "CreateTable"

'New ShapefileWorkspaceFactory is IWorkspaceFactory for shapefiles
Set pWorkspaceFactory2 = New ShapefileWorkspaceFactory
Set pWorkspace2 = pWorkspaceFactory2.OpenFromFile(strTableWorkspace, 0)
Set pFeatureWorkspace = pWorkspace2  'QI for the IFeatureWorkspace

'create the fields
Dim pFeatClass As IFeatureClass
Dim pFieldEdit As IFieldEdit
Dim pFieldsEdit As IFieldsEdit
Dim pFields As IFields
Dim pField As IField
'Create new Fields collection
Set pFields = New Fields
Set pFieldsEdit = pFields
pFieldsEdit.FieldCount = 4
```

```
'Create OID Field
Set pField = New Field
Set pFieldEdit = pField
With pFieldEdit
.AliasName = "FID"
.Name = "OBJECTID"
.Type = esriFieldTypeOID
End With
Set pFieldsEdit.Field(0) = pField

'Create From Site Field
Set pField = New Field
Set pFieldEdit = pField
With pFieldEdit
.AliasName = "FromOID"
.Name = "FromOID"
.Type = esriFieldTypeInteger
End With
Set pFieldsEdit.Field(1) = pField

'Create To Site Field
Set pField = New Field
Set pFieldEdit = pField
With pFieldEdit
.AliasName = "ToOID"
.Name = "ToOID"
.Type = esriFieldTypeInteger
End With
Set pFieldsEdit.Field(2) = pField

'Create Distance Field
Set pField = New Field
Set pFieldEdit = pField
With pFieldEdit
.AliasName = "Distance"
.Name = "Distance"
.Type = esriFieldTypeDouble
End With
Set pFieldsEdit.Field(3) = pField

'Check if .dbf table already exists: if yes, delete it.
Set pOutDSName = New TableName
pOutDSName.Name = sName
Set pWorkspace2 = pWorkspaceFactory2.OpenFromFile(strTableWorkspace, 0)
Set pED = pWorkspace2.Datasets(esriDTTable)
```

```
Set pDataset = pED.Next

Do Until pDataset Is Nothing
If pDataset.Name = pOutDSName.Name Then
pDataset.Delete
Exit Do
End If
Set pDataset = pED.Next
Loop

Set CreateTable = pFeatureWorkspace.CreateTable(pOutDSName.Name, pFields,
Nothing, Nothing, "")
End Function



Private Function OpenTable(strTableWorkspace As String, TableName As String) As
ITable
'open a table found in a shapefile workspace

On Error GoTo Errorhandler
Dim pShapefileWorkspaceName As IWorkspaceName
Dim pDatasetname As IDatasetName 'table to be opened
Dim pName As IName 'name of table to be opened

'Create the workspace name object
Set pShapefileWorkspaceName = New WorkspaceName
pShapefileWorkspaceName.PathName = strTableWorkspace
pShapefileWorkspaceName.WorkspaceFactoryProgID =
"esricore.shapefileworkspacefactory.1"

'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pShapefileWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenTable = pName.Open

Exit Function 'exit to avoid error handler
Errorhandler:
Set OpenTable = Nothing
End Function
```

```
Private Sub RecordResults(SiteOID1 As Long, SiteOID2 As Long, dblTraceLength As
Double)
'records the site OID for each sample site and the downstream distance between sites

Dim pTable As ITable 'temporary distance table
Dim pNewRow As IRow

'open the temporary distance table
Set pTable = OpenTable(strTableWorkspace, strTmpDistTableName)

'record the values in a new row
Set pNewRow = pTable.CreateRow
pNewRow.Value(1) = SiteOID1
pNewRow.Value(2) = SiteOID2
pNewRow.Value(3) = dblTraceLength
pNewRow.Store
End Sub


Private Sub SortTable(strSortTable As String, strNewTable As String, strType As String)
'Sorts the temporary distance table according to the To site and then the From site. Inserts
a 'header that contains the To feature in field 1 and the row that it begins in field 2. The
third field 'will contain a zero to indicate that this is the header. Header ends with To
feature: -9

Dim pSortTable As ITable 'name of table to be sorted
Dim pNewTable As ITable 'new table to create
Dim pFields As IFields 'collection of columns in the table
Dim pField As IField 'a specific column in the table
Dim sFromField As String, sToField As String 'name of from and to fields
Dim vOldToValue As Variant, vOldFromValue As Variant 'store the last to and from
values when sorting the table
Dim xCounter As Long
Dim xTotal As Long 'number of features
pApp.Caption = "SortingTable"

'open the table and populate the fields collection
Set pSortTable = OpenTable(strTableWorkspace, strSortTable)
Set pFields = pSortTable.Fields

If (pFields.FieldCount > 2) Then
Set pField = pFields.Field(1)
sFromField = pField.Name
Set pField = pFields.Field(2)
sToField = pField.Name
```

```
Else
Err
End If

Dim pRow As IRow, j As Long
Dim pQueryFilter As IQueryFilter
Set pQueryFilter = New QueryFilter

'sort the table according to ToFeatureOID and then FromFeatureOID
Dim pTableSort As ITableSort
Set pTableSort = New TableSort
With pTableSort
.Fields = sToField & "," & sFromField
.Ascending(sFromField) = True
.Ascending(sToField) = True
Set .QueryFilter = pQueryFilter
Set .Table = pSortTable
End With

pTableSort.Sort Nothing

'Check to see if distance table already exists: if yes, delete it.
Dim pWorkspaceFactory As IWorkspaceFactory
Dim pOutDSName As IDatasetName
Dim pWorkspace As IWorkspace
Dim pED As IEnumDataset 'enumeration of datasets in the workspace
Dim pDataset As IDataset

Set pWorkspaceFactory = New ShapefileWorkspaceFactory
Set pOutDSName = New TableName
pOutDSName.Name = strNewTable
Set pWorkspace = pWorkspaceFactory.OpenFromFile(strTableWorkspace, 0)
Set pED = pWorkspace.Datasets(esriDTTable)
Set pDataset = pED.Next

Do Until pDataset Is Nothing
If pDataset.Name = pOutDSName.Name Then
pDataset.Delete
Exit Do
End If
Set pDataset = pED.Next
Loop
'create a new table
Set pNewTable = CreateTable(strNewTable)

'create a cursor that will be used to sort the rows
```

```
Dim pCursor As ICursor
Dim pNewRow As IRow
Dim i As Integer
Dim FromFeat As Integer
Dim ToFeat As Integer
Dim TotalDist As Double
Set pCursor = pTableSort.Rows

'create a new table with sorted values and insert a header
Dim pFields3 As IFields
Set pFields3 = pSortTable.Fields

'loop through the table to create the header information
'find the number of features and check for duplicates

'get the total number of sites
xTotal = pPoint.FeatureCount(pQueryFilter)

'put the header information in the table
Set pCursor = pTableSort.Rows

Set pRow = pCursor.NextRow
xCounter = 2
xTotal = xTotal + xCounter

'initialize old value
vOldToValue = pRow.Value(2)
'create the header information and record it in the new table
Do While Not pRow Is Nothing

'if it is not a duplicate
If (pRow.Value(2) <> vOldToValue) Then
Set pNewRow = pNewTable.CreateRow
pNewRow.Value(1) = vOldToValue
pNewRow.Value(2) = CVar(xTotal)
pNewRow.Store
xTotal = xTotal + xCounter
xCounter = 1
'if it is a duplicate

Else
xCounter = xCounter + 1
End If

'reset the old to and from values
vOldToValue = pRow.Value(2)
```

```
vOldFromValue = pRow.Value(1)
Set pRow = pCursor.NextRow
Loop

'record the last site in the header information
Set pNewRow = pNewTable.CreateRow
pNewRow.Value(1) = vOldToValue
pNewRow.Value(2) = CVar(xTotal)
pNewRow.Store

'put end of file line in the header
Set pNewRow = pNewTable.CreateRow
pNewRow.Value(1) = CVar(-9)
pNewRow.Value(2) = CVar(xTotal + xCounter)
pNewRow.Store

Set pCursor = pTableSort.Rows
Set pRow = pCursor.NextRow

'record downstream hydrologic distance data in new table
Do While Not pRow Is Nothing
Set pNewRow = pNewTable.CreateRow
i = 1

'loop through and record each field in each new row
Do Until i = pFields.FieldCount
If Not IsEmpty(pRow.Value(i)) Then
If pFields.Field(i).Editable Then
pNewRow.Value(i) = pRow.Value(i)
Else
MsgBox "The table is not editable!"
Exit Sub
End If
End If
i = i + 1
Loop
pNewRow.Store
Set pRow = pCursor.NextRow
Loop
End Sub


Private Sub SetFlowDirection()
'set the flow direction in the network against digitization because NHD route.rch file is
'digitized against flow
```

```
Dim pUtilityNetwork As IUtilityNetwork
Dim lngEdgeEIDCount As Long   'stores the number of edge EID's in the network
Dim lngEdgeEID As Long   'stores the current edge EID
Dim i As Long

pApp.Caption = "SetFlowDirection"

'get a reference to the current network
Set pUtilityNetwork = pGeometricNetwork.Network

'Get the editor reference and set pEditor
Call GetEditorReference

'create an edit operation enabling an undo for this operation
pEditor.StartOperation

'get a list of the EID's for edges in the network
Set pEdgeSel = GetCurrentEIDs(esriETEdge)

'set the flow direction for each edge in the network
'reset the enumeration
pEdgeSel.Reset
'determine the number of edges in the enumeration
lngEdgeEIDCount = pEdgeSel.count

For i = 0 To lngEdgeEIDCount - 1
'get the next edge EID
lngEdgeEID = pEdgeSel.Next
'set the flow direction for this edge
pUtilityNetwork.SetFlowDirection lngEdgeEID, esriFDAgainstFlow
Next i

'stop the edit operation and specify the name of this edit operation
pEditor.StopOperation "Set Flow Direction"
pEditor.StopEditing True
End Sub


Public Sub GetEditorReference()
'get a reference to the editor. Needed to start an editing session in StartEditing subroutine

Dim pID As esriCore.UID
Set pID = New esriCore.UID
pID = "esriCore.Editor"
Set pEditor = Application.FindExtensionByCLSID(pID)
StartEditing pEditor
```

```
End Sub


Public Sub StartEditing(pEditor As IEditor)
'start an editing session
Dim pDataset As IDataset
Dim pDatasetWorkspace As IWorkspace

'if there is already an editing session in progress, exit
If Not pEditor.EditState = esriStateNotEditing Then Exit Sub

Set pDataset = pLine
Set pDatasetWorkspace = pDataset.Workspace
pEditor.StartEditing pDataset.Workspace
End Sub


Private Function OpenGeodatabaseTable(TableName As String) As ITable
'Opens a table in a geodatabase and returns it as an ITable using the table name as an
input

On Error GoTo Errorhandler
Dim pDatasetname As IDatasetName
Dim pName As IName

Dim pWorkspaceName As IWorkspaceName
Set pWorkspaceName = New WorkspaceName
pWorkspaceName.WorkspaceFactoryProgID = "esricore.AccessWorkspaceFactory"
pWorkspaceName.PathName = geodatabasePathName

'create the table name object
Set pDatasetname = New TableName
pDatasetname.Name = TableName
Set pDatasetname.WorkspaceName = pWorkspaceName

'Open the table
Set pName = pDatasetname
Set OpenGeodatabaseTable = pName.Open

Exit Function 'exit to avoid error handler
Errorhandler:
Set OpenGeodatabaseTable = Nothing
End Function


Private Sub DeleteTable(sTableName As String)
```

```
'Delete a table if it already exists

Dim pWorkspaceFactory2 As IWorkspaceFactory
Dim pWorkspace2 As IWorkspace
Dim pFeatureWorkspace As IFeatureWorkspace
Dim pDataset As IDataset
Dim pNetworkWorkspace As INetworkWorkspace
Dim pED As IEnumDataset
Dim pOutDSName As IDatasetName
pApp.Caption = "Deleting temporary table"

Set pWorkspaceFactory2 = New ShapefileWorkspaceFactory
Set pWorkspace2 = pWorkspaceFactory2.OpenFromFile(strTableWorkspace, 0)
Set pFeatureWorkspace = pWorkspace2  'QI for the IFeatureWorkspace

'Check if .dbf table already exists: if yes, delete it.
Set pOutDSName = New TableName
pOutDSName.Name = sTableName
Set pWorkspace2 = pWorkspaceFactory2.OpenFromFile(strTableWorkspace, 0)
Set pED = pWorkspace2.Datasets(esriDTTable)
Set pDataset = pED.Next

Do Until pDataset Is Nothing
If pDataset.Name = pOutDSName.Name Then
pDataset.Delete
Exit Do
End If
Set pDataset = pED.Next
Loop
End Sub



Private Function IsPathDownstream(lngEdgeEID As Long) As Boolean
' Returns a boolean value, which indicates whether the path path is found in downstream
' direction from site

Dim pNetwork As INetwork
Dim pNetTopology As INetTopology
Dim lngToEID As Long, lngFromEID As Long 'EIDs of To and From junctions for a
specific edge
Dim lngAdjEdgesCount As Long 'number of adjacent edges
Dim adjEdges() As Long 'array of EID values for adjacent edges
Dim RevOr() As Boolean 'array of digitization direction of adjacent edges
Dim lngPathEID As Long 'edge EID that was returned by the trace function and is part of
the path
```

```
Dim lngAdjEdgeEID As Long 'adjacent edge EID
Dim blnDownstream As Boolean 'indicates whether the path approaches from the
downstream direction
Dim i As Integer, j As Integer
Dim elementClassID1 As Long, elementOID1 As Long, elementSub1 As Long

Set pNetwork = pGeometricNetwork.Network
Set pNetTopology = pNetwork
blnDownstream = False

'get the From junction EID - we are interested in this because we want the downstream
junction
'and NHD data is digitized against flow
pNetTopology.GetFromToJunctionEIDs lngEdgeEID, lngFromEID, lngToEID

'get the number of edges adjacent to from junction
lngAdjEdgesCount = pNetTopology.GetAdjacentEdgeCount(lngFromEID)

're-allocate memory for the arrays returned by the GetAdjacentEdges function
Erase adjEdges()
Erase RevOr()
ReDim adjEdges(lngAdjEdgesCount - 1) As Long
ReDim RevOr(lngAdjEdgesCount - 1) As Boolean

'get an enumeration of edge EIDs adjacent to from junction and their digitized direction
pNetTopology.GetAdjacentEdges lngFromEID, lngAdjEdgesCount, adjEdges(0),
RevOr(0)

'compare adjacent EIDs to FindPath EIDs to determine if path is found in downstream
direction
pEdgeSel.Reset
i = 1
For i = 1 To pEdgeSel.count
lngPathEID = pEdgeSel.Next
j = 0
For j = 0 To lngAdjEdgesCount - 1
lngAdjEdgeEID = adjEdges(j)
'If the downstream edge is part of the path
If lngPathEID = lngAdjEdgeEID And lngPathEID <> lngEdgeEID Then
'path is found in downstream direction from site
blnDownstream = True
GoTo PassBackValue
End If
Next j
Next i
PassBackValue: IsPathDownstream = blnDownstream
```

```
End Function


Private Function WriteMatrix()
'Create an NxN matrix in comma delimited text format.
Dim pTable As ITable
Dim pRow As IRow
Dim pEnumLayers As IEnumLayer
Dim i As Integer
Dim pCursor As ICursor
Dim lFileID As Long
Dim lngToCounter As Long
Dim lngFromCounter As Long
Dim lngFeatCount As Long 'number of sites
Dim sText As String
Dim dblDistance() As Double
Set pTextWkspFactory = New TextFileWorkspaceFactory

'number of sites
lngFeatCount = pPoint.FeatureCount(Nothing)

'if the text file already exists, delete it
If Len(Dir(sTextFileName)) > 0 Then Kill sTextFileName

'Create and open the text file
lFileID = FreeFile
Open sTextFileName For Output As #lFileID

'open the .dbf export table
Set pTable = OpenTable(strTableWorkspace, strDistName)

'if the .dbf export table isn't returned, then exit program
If pTable Is Nothing Then
Exit Function
End If

'populate the cursor with the rows in the export table
Set pCursor = pTable.Search(Nothing, False)
Set pRow = pCursor.NextRow
lngToCounter = 1

'do this for each possible To Site
Do Until lngToCounter = lngFeatCount + 1
lngFromCounter = 1

If Not pRow Is Nothing Then
```

```
'clear the distance array when you start with a new To Site
Erase dblDistance()
ReDim dblDistance(lngFeatCount)

'if the row is not part of the header
If pRow.Value(3) > 0 Then

'do this for each possible From Site
Do Until lngFromCounter = lngFeatCount + 1

If Not pRow Is Nothing Then
If pRow.Value(1) = lngFromCounter And pRow.Value(2) = lngToCounter Then

'write the value to the distance array
dblDistance(lngFromCounter - 1) = pRow.Value(3)

Set pRow = pCursor.NextRow
lngFromCounter = lngFromCounter + 1

Else
'if From site is not connected to To site, record a 0 for distance
dblDistance(lngFromCounter - 1) = 0
lngFromCounter = lngFromCounter + 1
End If

Else
Exit Do
End If
Loop

Dim z As Integer
z = 1
For z = 1 To lngFeatCount
Write #lFileID, dblDistance(z - 1);
Next
Write #lFileID,

'add one to the ToCounter
lngToCounter = lngToCounter + 1

'If pRow.Value(3) = 0, it is part of the header. Go on to next row
Else
Set pRow = pCursor.NextRow
End If

'If you have reached the end of the table...
```

```
Else
Exit Do
End If
Loop

'if the end of the table has been reached before data has been recorded for all of the
'To sites, fill in the remaining To site rows with zeros in the text file
FinishTable: If lngToCounter < lngFeatCount + 1 Then
Erase dblDistance()
ReDim dblDistance(lngFeatCount)

z = 1
For z = 1 To lngFeatCount
Write #lFileID, dblDistance(z - 1);
Next

Write #lFileID,
lngToCounter = lngToCounter + 1
GoTo FinishTable
End If

'close the text file and release memory
Close #lFileID
Set pWorkspaceFactory = Nothing
End Function
```

CREATE THE NETWORK ELEMENT OBJECT

```
VERSION 1.0 CLASS
BEGIN
MultiUse = -1 'True
END
Attribute VB_Name = "NetworkElement"
Attribute VB_GlobalNameSpace = False
Attribute VB_Creatable = False
Attribute VB_PredeclaredId = False
Attribute VB_Exposed = True
'clsNetworkElement.cls
' Downloaded from ArcObjects Help website 10/03 (http://arcscripts.esri.com/)

'this program should be loaded when the basHydroDist.bas, basHydroNeighborhood.bas,
'and the basHydroDistNeigh.bas programs are used. The purpose of this class module is
'to create a user defined object called NetworkElement. This object has 4 properties:
'ElementID, FeatureClassID, ObjectID, and the SubID which must be assigned.
```

```
Option Explicit
Private mEID As Long
Private mFCID As Long
Private mOID As Long
Private mSubID As Long

Public Property Let ElementID(ByVal vData As Long)
mEID = vData
End Property

Public Property Get ElementID() As Long
ElementID = mEID
End Property

Public Property Let FeatureClassID(ByVal vData As Long)
mFCID = vData
End Property

Public Property Get FeatureClassID() As Long
FeatureClassID = mFCID
End Property

Public Property Let ObjectID(ByVal vData As Long)
mOID = vData
End Property

Public Property Get ObjectID() As Long
ObjectID = mOID
End Property

Public Property Let SubID(ByVal vData As Long)
mSubID = vData
End Property

Public Property Get SubID() As Long
SubID = mSubID
End Property
```

APPENDIX III: STATISTICS CODE

MODEL SELECTION USING SPATIAL AICC

```
# Main function for model selection using spatial AICC and hydrologic distance.
# Code written by Andrew Merton.

source("source code text file here")

# Read data
Z <- standardize(resp)                                    # Response vector
X <- as.matrix(cbind(1,apply(pred,2,standardize)))        # Covariate matrix
np <- 5                                                   # No. of covariates
colnames(X) <- c("cst","X1","X2","X3","X4","X5")          # Covariate labels
maxdist<-max(stream.distance.data)
D<-stream.distance.data/maxdist                           # Standardized distances

D <- pmax(D,t(D))                                         # Force symmetry of D
A <- pmax(A,t(A))                                         # Force symmetry of A
diag(A) <- 1                                              # Correct flow matrix
n <- length(Z)                                            # No. of records

# Run Spatial AICC
system.time({
output <- NULL
for (i in 1:32) {                          # Controls the number of models to loop
                                           #through
print(paste("Model",i))
ind <- predictors(i)
tmp <- stream.fit(Z,X[,ind],D,A,ma.fcn="exp")
beta.tmp <- rep(NA,np+1)
beta.tmp[ind] <- tmp$beta
output <- rbind(output,c(i,tmp$aicc,tmp$likelihood,tmp$theta, beta.tmp,tmp$sigma2))
}

dimnames(output) <-
list(NULL,c("ExpModel","aicc","likelihood","th0","th1","th2","b0","b1","b2","b3","b4",
"b5","sigma2"))
})
```

```
# Output data to file
write.table(output,file="text file name here",append = FALSE, sep = ",",
row.names=FALSE, col.names=TRUE)
q()
y
```

R FUNCTIONS FOR SPATIAL AICC MODEL SELECTION

```
# R functions for spatial AICC Model Selection using in-stream distance and directional
# flow. Created a new covariance structure that coincides with that presented in VerHoef
# and others (2005). Code written by Andrew Merton
# ------------------------------------------------------------------------ #
# Function STANDARDIZE: Standardizes a vector such the scaled mean is zero
# and the variance is one.
standardize <- function(x) {
x <- as.vector(x)
x <- (x-mean(x))/sqrt(var(x))
return(x)
}

# ------------------------------------------------------------------------ #
# Function FLOW.ACF: Generates the variance-covariance matrix taking into account
the flow    # connectivity.
flow.acf <- function(theta,D,A,fcn="exp") {
th0 <- theta[1]
th1 <- theta[2]
th2 <- theta[3]
h  <- D/th2
if (fcn=="exp")   {Rho <- exp(-h)}
if (fcn=="gau")   {Rho<-exp(-h^2)}
if (fcn=="lin")   {Rho<-(1-h)*(h<1)}
if (fcn=="sph")   {Rho<-(1-1.5*h+0.5*h^3)*(h<1)}
if (fcn=="hole")  {Rho<-1*(h==0);Rho[Rho!=1]<-sin(h[h>0])/h[h>0]}
if (fcn=="quad")  {Rho<-1/(1+h^2)}
if (fcn=="mariah") {Rho<-1*(h==0);Rho[Rho!=1]<-log(h[h>0]+1)/h[h>0]}

Rho <- th1*Rho
diag(Rho) <- th0+th1
return(A*Rho)
}

# ------------------------------------------------------------------------ #
# Function PREDICTORS - Determines which X variables to include in the model.
predictors <- function(i,nmax=1024) {
```

```
params <- NULL
if (i<=nmax) {
resid  <- i-1
while (resid>0) {
params <- c(params,floor(log(resid)/log(2)))
resid  <- (i-1)-sum(2^(params))
}}
return(c(1,rev(params+2)))
}

# -------------------------------------------------------------------------- #
# Function PROFILE.LIKELIHOOD - Evaluates the log-profile likelihood of theta given
# the theta,  the distance matrix, the covariates, and the observed data.
# Note: Maximize(log.profile.likelihood) = Minimize(-log.profile.likelihood).

profile.likelihood <- function(theta,Z,X,D,A,fcn="exp") {

# Compute the correlation matrix.
n <- length(Z)
Rho <- flow.acf(theta,D,A,fcn=fcn)

# Invert the correlation matrix.
L <- t(chol(Rho))
L.inv <- forwardsolve(L,diag(nrow=n))
Rho.inv <- t(L.inv)%*%L.inv

# Compute the MLE for beta.
beta <- solve(t(X)%*%Rho.inv%*%X)%*%t(X)%*%Rho.inv%*%Z

# Compute the MLE for sigma.
resid <- Z-X%*%beta
sigma2 <- t(resid)%*%Rho.inv%*%resid/n

# Evaluate -log-profile likelihood.
log.L.det <- sum(log(diag(L)))
prof  <- log.L.det+n*log(sigma2)/2+n/2+n*log(2*pi)/2
return(prof)
}

# -------------------------------------------------------------------------- #
## Function STREAM.FIT: Fits the theta parameter to a stream network.
#  Z          - Response variable
#  X          - Matrix of covariates.  The leading column must be a column of ones.
#  D          - Stream distance matrix.
#  A          - Flow connectivity matrix.
#  ma.fcn     - Defines the function to be used in the autocorrelation function.
```

```
#               Defaults to "exp" which is the exponential ACF. Other options
#               include "gau" (Gaussian), "lin" (linear with Sill),
#               "sph" (spherical),"hole" (hole effect), "quad" (Rational
#               quadratic), and "mariah" (moving average reciprocal 1 add h).

# The function returns the MLEs for theta, beta, and sigma2 as well as the
# value of the likelihood function and the corresponding corrected AIC (AICC).

stream.fit <- function(Z,X,D,A,ma.fcn="exp") {

# Set optimization parameters:
theta  <- c(0.1,0.300,0.03)                  # Initial theta guesses.
llimits <- c(0.001,0.001,min(D[D!=0]))       # Upper and lower optimization limits.
ulimits <- c(0.999,0.999,max(D))

# Perform optimization.
profile.max <- optim(par=theta,fn=profile.likelihood,method="L-BFGS-
B",lower=llimits,upper=ulimits, Z=Z,X=X,D=D,A=A,fcn=ma.fcn)

ifelse (is.vector(X),p<-0,p<-length(X[1,])-1)      # Number of covariates.
k <- 3                                        # Number of autocovariance parameters.
theta <- profile.max$par
aicc <- 2*profile.max$value+2*n*(p+k+1)/(n-p-k-2)
Rho <- flow.acf(theta,D,A,fcn=ma.fcn)
L <- t(chol(Rho))
L.inv <- forwardsolve(L,diag(nrow=n))
Rho.inv <- t(L.inv)%*%L.inv
beta.mle <- solve(t(X)%*%Rho.inv%*%X)%*%t(X)%*%Rho.inv%*%Z
sigma2.mle <- as.numeric(t(Z-X%*%beta.mle)%*%Rho.inv%*%(Z-X%*%beta.mle)/n)
return(list(theta=theta,likelihood=profile.max$value,aicc=aicc,
beta=beta.mle,sigma2=sigma2.mle,ma.fcn=ma.fcn))
}
```

MAKE PREDICTIONS AT UNOBSERVED SITES

```
# main program for making predictions at unobserved sites
source("source code for prediction functions")

# Read data...
Z <- standardize(resp)                                   # Response
X <- as.matrix(cbind(1,apply(pred,2,standardize)))       # Covariates
np <- 5                                                  # No. of covariates
colnames(X) <- c("cst","X1","X2","X3","X4","X5")         # Covariate labels
```

```
n <- length(Z)                                  # No. of records
A<-pmax(A,t(A))                                 # Force symmetry of A
D<-pmax(D,t(D))                                 # Force symmetry of D
diag(A)<-1                                      # Correct flow matrix
theta0<-theta                                   # Set initial theta guesses
imodel<-32                                      # "best" model selected
                                                # during model selection

#Separate fit data from test data
# Create test data
testdata<-sample(1:n,100)

# Use a file with test data row numbers
# testdata <- scan("/home/nrel/eposton/asymmetric/anc/anc_td.txt", what = numeric(0),
sep = "\t")

# vector with 1's indicating locations with observed data, 0's indicating prediction
# locations
sampled.vector <- matrix(data = 1, nrow = n, ncol = 1)
sampled.vector[testdata]<- 0

# Refit model and make predictions
system.time({
pred.output <- NULL

print(paste("Model",imodel))
ind <- predictors(imodel)
tmp <- stream.fit2(theta0,Z,X[,ind],D,A,obs=sampled.vector,ma.fcn="exp")
beta.tmp <- rep(NA,np+1)
beta.tmp[ind] <- tmp$beta
pred.output <- rbind(pred.output,c(imodel,tmp$aicc,tmp$likelihood,tmp$theta,
beta.tmp,tmp$sigma2,))

predictions<-tmp$preds

dimnames(pred.output) <-
list(NULL,c("ExpModel","aicc","likelihood","th0","th1","th2","b0","b1","b2","b3","b4",
"b5","sigma2"))
})

# Output model parameters to file
write.table(pred.output,file="/home/nrel/eposton/asymmetric/anc/predoutput.csv",
append = FALSE, sep = ",",row.names=FALSE,col.names=TRUE)

# Output predictions to file
```

```
write.table(predictions,file="/home/nrel/eposton/asymmetric/anc/predictions.txt",append
= FALSE, sep =",",row.names=FALSE,col.names=FALSE)

q()
y
```

R FUNCTIONS FOR MAKE PREDICTIONS AT UNOBSERVED SITES

```
# R functions for Make Predictions at Unobserved Sites using in-stream distance and
# directional flow. Created a new covariance structure that coincides with that presented
# in VerHoef and others (2005). Code written by Andrew Merton

# -------------------------------------------------------------------------- #
# Function STANDARDIZE: Standardizes a vector such the the scaled mean is zero
# and the variance is one.
standardize <- function(x) {
x <- as.vector(x)
x <- (x-mean(x))/sqrt(var(x))
return(x)
}
# -------------------------------------------------------------------------- #
# Function FLOW.ACF: Generates the variance-covariance matrix taking into account
# the flow connectivity.
flow.acf <- function(theta,D,A,fcn="exp") {
th0 <- theta[1]
th1 <- theta[2]
th2 <- theta[3]
h  <- D/th2
if (fcn=="exp")   {Rho <- exp(-h)}
if (fcn=="gau")   {Rho<-exp(-h^2)}
if (fcn=="lin")   {Rho<-(1-h)*(h<1)}
if (fcn=="sph")   {Rho<-(1-1.5*h+0.5*h^3)*(h<1)}
if (fcn=="hole")  {Rho<-1*(h==0);Rho[Rho!=1]<-sin(h[h>0])/h[h>0]}
if (fcn=="quad")  {Rho<-1/(1+h^2)}
if (fcn=="mariah") {Rho<-1*(h==0);Rho[Rho!=1]<-log(h[h>0]+1)/h[h>0]}

Rho <- th1*Rho
diag(Rho) <- th0+th1
return(A*Rho)
}
# -------------------------------------------------------------------------- #
# Function PREDICTORS - Determines which X variables to include in the model.
predictors <- function(i,nmax=1024) {
params <- NULL
```

```
if (i<=nmax) {
resid <- i-1
while (resid>0) {
params <- c(params,floor(log(resid)/log(2)))
resid <- (i-1)-sum(2^(params))
}}
return(c(1,rev(params+2)))
}
# ------------------------------------------------------------------------ #
# Function PROFILE.LIKELIHOOD - Evaluates the log-profile likelihood of theta
# given the theta, the distance matrix, the covariates, and the observed data.

# Note: Maximize(log.profile.likelihood) = Minimize(-log.profile.likelihood).
profile.likelihood <- function(theta,Z,X,D,A,fcn="exp") {

# Compute the correlation matrix.
n <- length(Z)
Rho <- flow.acf(theta,D,A,fcn=fcn)

# Invert the correlation matrix.
L <- t(chol(Rho))
L.inv <- forwardsolve(L,diag(nrow=n))
Rho.inv <- t(L.inv)%*%L.inv

# Compute the MLE for beta.
beta <- solve(t(X)%*%Rho.inv%*%X)%*%t(X)%*%Rho.inv%*%Z

# Compute the MLE for sigma.
resid <- Z-X%*%beta
sigma2 <- t(resid)%*%Rho.inv%*%resid/n

# Evaluate -log-profile likelihood.
log.L.det <- sum(log(diag(L)))
prof <- log.L.det+n*log(sigma2)/2+n/2+n*log(2*pi)/2
return(prof)
}
# ------------------------------------------------------------------------ #
## Function STREAM.FIT2: Fits the theta parameter to a stream network.
# theta0      - Initial guess for the theta vector.
# Z           - Response variable.
# X           - Matrix of covariates.  The leading column must be a column of ones.
# D           - Stream distance matrix.
# A           - Flow connectivity matrix.
# obs         - Defines which records correspond to "observations" (1) and which
#               records are to be used for prediction (0).  Assumed to be a null
#               vector, i.e., treat all records as observations.
```

```
# ma.fcn      - Defines the function to be used in the autocorrelation function.
#               Defaults to "exp" which is the exponential ACF. Other options
#               include "gau" (Gaussian), "lin" (linear with Sill),
#               "sph" (spherical),"hole" (hole effect), "quad" (Rational
#               quadratic), and "mariah" (moving average reciprocal 1 add h).

# The function returns the MLEs for theta, beta, and sigma2 as well as the
# value of the likelihood function and the corresponding corrected AIC (AICC).

stream.fit2 <- function(theta0,Z,X,D,A,obs=NULL,ma.fcn="exp") {
n <- length(Z)
print(paste("inside stream.fit2"))

# Check to see if any predictions are to be made. If so, be sure to extract
# the proper rows and/or columns from each vector.matrix for optimization.
ifelse (is.null(obs),sample.vector<-matrix(1,nrow=n),sample.vector<-obs)
nn <- sum(sample.vector) # No. of observations.
np <- n-nn          # No. of prediction sites.
                    # Determine which records to use for estimation.
indices <- matrix((1:n)[sample.vector==1],nrow=nn)
ZZ <- Z[indices]
XX <- X[indices,]
DD <- D[indices,indices]
AA <- A[indices,indices]

# Set optimization parameters:
llimits <- c(0.001,0.001,min(D[D!=0])) # Optimization limits.
ulimits <- c(0.999,0.999,max(D))

# Perform optimization.
profile.max <- optim(par=theta0,fn=profile.likelihood, method="L-BFGS-
B",lower=llimits,upper=ulimits, Z=ZZ,X=XX,D=DD,A=AA,fcn=ma.fcn)
print(paste(profile.max$par))
print(paste("finished optim"))

ifelse (is.vector(X),p<-0,p<-length(X[1,])-1)  # Number of covariates.
k <- 3                          # Number of autocovariance parameters.
theta <- profile.max$par
aicc <- 2*profile.max$value+2*nn*(p+k+1)/(nn-p-k-2)
Rho <- flow.acf(theta,DD,AA,fcn=ma.fcn)
L <- t(chol(Rho))
L.inv <- forwardsolve(L,diag(nrow=nn))
Rho.inv <- t(L.inv)%*%L.inv

beta.mle  <- solve(t(XX)%*%Rho.inv%*%XX)%*%t(XX)%*%Rho.inv%*%ZZ
```

```
sigma2.mle <- as.numeric(t(ZZ-XX%*%beta.mle)%*%Rho.inv%*%(ZZ-
XX%*%beta.mle)/nn)
print(paste("finished beta mles"))

# Perform predictions (if any).
preds <- NULL
if (np>0) {
preds <- make.predictions(theta,Z,X,D,A,indices,ma.fcn="exp")
}

print(paste("finished predictions"))

output <- list(theta=theta,likelihood=profile.max$value,aicc=aicc,
beta=beta.mle,sigma2=sigma2.mle,ma.fcn=ma.fcn,preds=preds)
}

# -------------------------------------------------------------------------- #
# Function MAKE.PREDICTIONS - Computes the mean square prediction error
# and the percent coverage using the MLEs for theta and the "true" values at
# the prediction locations.
make.predictions <- function(theta,Z,X,D,A,indices,ma.fcn="exp") {
nn <- length(indices)
np <- length(Z)-nn

X1 <- X[ indices,]
X2 <- X[-indices,]

# Compute the weighted correlation matrix.
Rho <- flow.acf(theta,D,A,fcn=ma.fcn)

# Invert the correlation matrix for the observed data only, Rho11.
L <- t(chol(Rho[indices,indices]))
L.inv <- forwardsolve(L,diag(nrow=nn))
Rho11.inv <- t(L.inv)%*%L.inv

# Compute the MLEs for beta and sigma2.
beta  <- solve(t(X1)%*%Rho11.inv%*%X1)%*%t(X1)%*%Rho11.inv%*%Z[indices]
resid <- Z[indices]-X1%*%beta
sigma2 <- as.numeric(t(resid)%*%Rho11.inv%*%resid/n)

# Estimate the expected value and variance at the prediction locations.
Z.hat <- X2%*%beta+Rho[-indices,indices]%*%Rho11.inv%*%resid
Sigma <- sigma2*(Rho[-indices,-indices]-Rho[-
indices,indices]%*%Rho11.inv%*%Rho[indices,-indices])
resid <- X2-Rho[-indices,indices]%*%Rho11.inv%*%X1
Sigma <- Sigma+resid%*%solve(t(X1)%*%Rho11.inv%*%X1)%*%t(resid)
```

```
# Compute the mean square error and the proportion of (individual) confidence
# intervals that contain the "true" values at the prediction locations.
p <- length(beta)
k <- length(theta)
mspe <- t(Z[-indices]-Z.hat)%*%(Z[-indices]-Z.hat)/np
t.crit <- qt(0.975,nn-(p+1+k))
bound <- t.crit*sqrt(diag(Sigma))
pc <- length(split(1:np,bound-abs(Z[-indices]-Z.hat)>=0)$T)/np
return(list(z.hat=Z.hat,mspe=mspe,pc=pc))
}
```

APPENDIX IV: CLASSIFICATION OF PRIMARY LITHOLOGIES INTO FIVE CLASSES OF GEOLOGICAL SENSITIVITY

Class	Primary Lithology
Siliceous	arenite
	chert
	conglomerate
	conglomerate (sandstone)
	metasandstone
	orthoquartzite
	quartzite
	sandstone
Felsic	alaskite
	augen gneiss
	biotite gneiss
	conglomerate (arkose)
	dacite
	felsic gneiss
	felsic metavolcanic rock
	felsic volcanic rock
	gneiss
	granite
	granitic gneiss
	granodiorite
	granulite
	mica schist
	migmatite
	mylonite
	orthogneiss
	quartz diorite
	quartz monzonite
	rhyolite
	sedimentary breccia (arkose)
	syenite
Carbonic	dolomite (dolostone)
	dolostone (dolomite)
	limestone
	marble

APPENDIX IV: continued.

Class	Primary Lithology
Mafic	amphibole schist
	amphibolite
	anorthosite
	basalt
	diabase
	diorite
	dunite
	gabbro
	greenstone
	mafic gneiss
	mafic metavolcanic rock
	meta-basalt
	metavolcanic rock
	norite
	peridotite
	schist (actinolite)
	ultramafic intrusive rock
Argillaceous	black shale
	claystone
	conglomerate (graywacke)
	conglomerate (mudstone)
	conglomerate (shale)
	graywacke
	meta-argillite
	metasedimentary rock (graywacke)
	metasedimentary rock (meta-argillite)
	metasedimentary rock (mica schist)
	metasedimentary rock (phyllite)
	metasedimentary rock
	mudstone
	phyllite
	schist
	sedimentary breccia
	sedimentary breccia (mudstone clasts)
	shale
	siltstone
	slate