Bayesian Methods to Characterize Uncertainty

in Predictive Modeling of the Effect of Urbanization on Aquatic Ecosystems

by

Roxolana Oresta Kashuba

Department of Environment
Duke University

Date:_____

Approved:

_____
Kenneth H. Reckhow, Supervisor

_____
Song S. Qian

_____
Thomas F. Cuffney

_____
Gerard McMahon

_____
Emily S. Bernhardt

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in
the Department of Environment in the Graduate School
of Duke University

2010

ABSTRACT

Bayesian Methods to Characterize Uncertainty

in Predictive Modeling of the Effect of Urbanization on Aquatic Ecosystems

by

Roxolana Oresta Kashuba

Department of Environment
Duke University

Date:_____

Approved:

_____
Kenneth H. Reckhow, Supervisor

_____
Song S. Qian

_____
Thomas F. Cuffney

_____
Gerard McMahon

_____
Emily S. Bernhardt

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor of Philosophy in the Department of
Environment in the Graduate School
of Duke University

2010

Abstract

Urbanization causes myriad changes in watershed processes, ultimately disrupting the structure and function of stream ecosystems. Urban development introduces contaminants (human waste, pesticides, industrial chemicals). Impervious surfaces and artificial drainage systems speed the delivery of contaminants to streams, while bypassing soil filtration and local riparian processes that can mitigate the impacts of these contaminants, and disrupting the timing and volume of hydrologic patterns. Aquatic habitats where biota live are degraded by sedimentation, channel incision, floodplain disconnection, substrate alteration and elimination of reach diversity. These compounding changes ultimately lead to alteration of invertebrate community structure and function. Because the effects of urbanization on stream ecosystems are complex, multilayered, and interacting, modeling these effects presents many unique challenges, including: addressing and quantifying processes at multiple scales, representing major interrelated simultaneously acting dynamics at the system level, incorporating uncertainty resulting from imperfect knowledge, imperfect data, and environmental variability, and integrating multiple sources of available information about the system into the modeling construct. These challenges can be addressed by using a Bayesian modeling approach. Specifically, the use of multilevel hierarchical models and Bayesian network models allows the modeler to harness the hierarchical nature of the U.S.

Geological Survey (USGS) Effect of Urbanization on Stream Ecosystems (EUSE) dataset to predict invertebrate response at both basin and regional levels, concisely represent and parameterize this system of complicated cause and effect relationships and uncertainties, calculate the full probabilistic function of all variables efficiently as the product of more manageable conditional probabilities, and includes both expert knowledge and data. Utilizing this Bayesian framework, this dissertation develops a series of statistically rigorous and ecologically interpretable models predicting the effect of urbanization on invertebrates, as well as a unique, systematic methodology that creates an informed expert prior and then updates this prior with available data using conjugate Dirichlet-multinomial distribution forms. The resulting models elucidate differences between regional responses to urbanization (particularly due to background agriculture and precipitation) and address the influences of multiple urban induced stressors acting simultaneously from a new system-level perspective. These Bayesian modeling approaches quantify previously unexplained regional differences in biotic response to urbanization, capture multiple interacting environmental and ecological processes affected by urbanization, and ultimately link urbanization effects on stream biota to a management context such that these models describe and quantify how changes in drivers lead to changes in regulatory endpoint (the Biological Condition Gradient; BCG).

Contents

List of Tables

List of Figures

Acknowledgements

First and foremost, I would like to thank my fantastic doctoral committee, particularly its chair, Ken Reckhow, without whose insight and guidance this research could not have been accomplished. Song Qian freely shared his amazing statistical knowledge and modeling abilities whensoever I asked, and encouraged and inspired me continuously. Tom Cuffney's extensive expert ecological knowledge, open-mindedness to Bayesian ideas, and unending feedback and discussions were indispensable. Jerry McMahon managed the entire administrative structure under which this project prospered, helped me better frame and organize my ideas, enabled and encouraged analysis progress, and expertly navigated essential government processes. Emily Bernhardt lured me away from my computer by giving me the opportunity to engage in actual water quality sampling, and ensured that I couched my theoretical models in the framework of environmental reality. This combination of outstanding influences enriched my doctoral research substantially and I am beyond grateful for my committee's massive contribution of time and effort.

This research was funded by the U. S. Geological Survey and I could not imagine participating in any better, more beneficial working relationship with a funding agency. USGS covered full salary, frequent travel to conferences and meetings, publishing expenses, and even gave me office space and constant access to all their resources. More

importantly, the open exchange of ecological and statistical information created a vibrant learning environment with many useful new ideas and research products!

I would also like to extend sincerest appreciation to all our collaborators in the Maine Department of Environmental Protection, U.S. Environmental Protection Agency (USEPA), Tetra Tech, Inc., U.S. Geological Survey (USGS), Anne Arundal County, Maryland, and the Center for Watershed Protection, including Susan Davies, Dave Courtemanch, Tom Danielson, Susan Jackson, Jeroen Gerritsen, Jim Coles, Tom Cuffney, Peter Weiskel, Marilee Horn, Karen Beaulieu, David Armstrong, Chris Waldron, Faith Fitzpatrick, Marie Peppler, Barbara Scudder, Kernell Ries, Amanda Bell, Jason May, Larry Brown, Ron Bowen, Janis Markusic, Christopher Victoria, Hala Flores, Paul Sturm, Bill Stack, Joe MacDonald, and Karen Cappiella. These gifted scientists and planners shared their environmental, biological, and urban planning expertise with open minds, and helped me learn how to present and discuss modeling information with a variety of audiences.

I would like to thank my colleagues in the Nicholas School for many fruitful discussions and information exchange, especially in the current and former Reckhow lab: Melissa Kenney, Drew Gronewold, Ibrahim Alameddine, YoonKyung Cha, Boknam Lee, Eric Money, Kris Voss, Yun Jian, and Farnaz Nojavan, and my current and former office mates: Jason Jackson, Madeleine Baker-Goering, Soumya Hassan, Jianwei Li, Daniela Miteva, Dave Bell, Kevin Fritze, and Shana Starobin. Also, Meg Stephens was a

godsend in helping me wrestle with all manner of Nicholas School related administrative questions. And Ram Oren and Gabi Katul were instrumental in resolving major programmatic incompatibilities.

My family and friends have been an important source of support and enthusiasm. Мамцю і Татку, I can't even begin to describe everything you have done for me, how you've encouraged me, and how much your pride in me means. Адю і Ромчику, thanks for trying not to fall asleep when I discuss my environmental statistics passions but, more importantly, for believing in me and for being there for me. Aimee, Julie, Tania, Claudia, Patrick, Andrew, Emily, Hannah, Foos, Brian, Annie, Jordan, Beia, and YBB, you all made this experience much more pleasant than it would have otherwise been. And, finally, I'd like to thank my incredibly patient and supportive boyfriend, Michael Arribage, for spending the last four years talking me off many ledges, listening to countless practice presentations and offering feedback, and reminding me that I can do this especially when I was sure there was no way.

1. Introduction

1.1. Urbanization processes and their impacts

Urbanization of landscapes causes myriad compounding changes in watershed processes, ultimately leading to disruption of the structure and function of stream ecosystems (Walsh et al. 2005). Increasing impervious surface alters water movement in stream basins by limiting infiltration capacity and, therefore, escalating surface runoff (Klein 1979, Finkenbine et al. 2000). The resulting changes to the timing and volume of water delivery tend to decrease base flow and increase stream flashiness (Poff et al. 1997, Konrad and Booth 2002, McMahon et al. 2003). Superimposed on the effects of impervious surface are human population driven patterns of water use. These artificial withdrawals and returns further distort stream flow patterns (Weiskel et al. 2007). In addition to hydrological impacts, both impervious surface and human water use affect stream water quality (Beck 2005). That is, urban development creates large new contaminant sources (human waste, pesticides, industrial chemicals) (Hall and Anderson 1988, Pitt et al. 1995, Van Metre et al. 2000, Mahler et al. 2005, Gilliom et al. 2006) and speeds the delivery of those contaminants to streams through surface runoff and artificial drainage while bypassing the ability of the land to mitigate these contaminants through watershed-scale soil filtration and local riparian processes (Groffman et al. 2003). Destruction of riparian buffers via deforestation also leads to

declines in canopy cover, changes in energy inputs, and increases in water temperatures (Waite and Carpenter 2000, Jacobson 2001, Sprague et al. 2006). The combination of this buffer vegetation removal and hydrologic disruption further deteriorates aquatic habitats through sedimentation (Wolman and Schick 1967, Trimble 1997), channel incision (Booth 1990), floodplain disconnection (Haase 2003), substrate alteration (Sponseller et al. 2001, Roy et al. 2003) and elimination of reach diversity (Winterbourn and Townsend 1991). Collectively, these and other interrelated changes in the physical and chemical environment have been associated with degraded invertebrate assemblages in many urban areas (Jones and Clark 1987, Kennen 1999, Yoder et al. 1999, Ourso and Frenzel 2003). At a local scale, many different processes affect the functioning of an urban stream ecosystem, such as: hydrology (flashiness, connectedness, sewage systems and stormwater movement), soil, animal interactions, and pollution (Bernhardt et al. 2008).

1.2. Invertebrate community metrics

Aquatic benthic macroinvertebrates are organisms without backbones that can be seen without magnification and live at least part of their life cycles on the stream bottom. Such organisms include some members of phylums Annelida (worms and leeches) and Mollusca (mussels and snails), but most macroinvertebrates are of the phylum Arthropoda (crustaceans such as crayfish and sowbugs but, especially, insects such as

mayflies, stoneflies, caddisflies, water penny beetles, riffle beetles, dragonflies, black flies, and midge flies)(USEPA 1997).

The composition of benthic macroinvertebrate communities, both in terms of the types and amounts present, can be an excellent biological indicator of stream health. This is because different types of stream invertebrates require different physical and chemical conditions to prosper and, hence, invertebrate community structure as a whole responds to the combination of changes in these physical and chemical conditions (Hellawell 1986). Physical factors that affect invertebrates include stream temperature, stream channel shape and structure, stream bed stability, stream bottom substrate composition and diversity, substrate embeddedness, magnitude and frequency of hydrologic peaks, shear stress, dissolved oxygen concentrations, suspended solids, light, food availability, and riparian zone shading and filtration (Huryn et al. 2002, Kennen and Ayers 2002, Vølstad et al. 2003, Roy et al. 2003a). Influential chemical factors include concentrations of nutrients, organics, and toxins such as pesticides, metals, and industrial chemicals (Klein 1979, Jones and Clark 1987, Horner et al. 1997, Beasley and Kneale 2002). These factors vary regionally and locally (Vinson and Hawkins 1998). Additionally, changing land cover patterns have the potential to affect any number of these elements of stream quality.

Invertebrates are subject to the effects of these environmental stream conditions throughout their life cycles. Therefore, sampling invertebrate community structure provides a more integrated, functional estimate of stream quality than an instantaneous snapshot of any of these physical or chemical measurements individually (Barbour et al. 1999). Other benefits of using benthic macroinvertebrate biomonitoring to characterize stream health include their ubiquitousness in existing in many different regions and aquatic habitats (Lenat et al. 1980) allowing for their wide use in biomonitoring and enabling comparison between sites, their limited mobility compared to fish or other larger species to better spatially link an invertebrate sample to its habitat of residence (Abel 1989), their longer lifecycles compared to algae or other smaller species to allow for longer scale temporal assessment of stream conditions (Lammert and Allan 1999), and the relative ease of their collection and identification (Hellawell 1986).

As there are hundreds of different species of benthic macroinvertebrates which can be found in a stream, invertebrate metrics have been developed to summarize information about community structure. Typically, these metrics report either taxa count (richness), individual count (abundance), or percent composition (relative abundance) of specific types of invertebrates grouped in different, meaningful ways. The most common grouping criteria are based on taxa classification, pollution

sensitivity, functional feeding patterns, or habitat and behavioral characteristics. Each of these groups can be reported in richness, abundance or relative abundance.

Description of invertebrate community structure based on taxa classification can be done at the order, family, genus, or species level, depending on the lowest taxonomic level of classification available for a particular dataset. Common reported invertebrate taxa include order Coleoptera (beetles), order Diptera (flies), family Chironomidae (midges), class Gastropoda (snails), class Oligochaeta (worms), and many more in different permutations (eg., non-midge Diptera, non-insect, etc.). Each taxa is known to be tolerant of a particular set of conditions, so what types of taxa groups are present in what amounts are indicative of what conditions exist in that stream (Margalef 1958). The most used invertebrate metric grouping indicative of pollution sensitivity is EPT, orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies). If less EPT taxa are present, stream conditions are more likely to be impaired. Invertebrates grouped by functional feeding patterns include predators, omnivores, collectors (filter collectors or gatherer collectors), scrapers/grazers, shredders, and piercers. Information on the balance of different invertebrate feeding strategies provides an estimate of the stability of food web dynamics in a sampled ecosystem (Karr et al. 1986). Invertebrates grouped by habitat and behavioral characteristics include clingers, climbers, sprawlers, burrowers and swimmers. Knowing about the distribution of

invertebrates sorted by how they function informs about stream habitat conditions (Merritt et al. 1996).

Additionally, biologists and ecologists have derived for use as invertebrate metrics indices representing community characteristics such as tolerance, diversity or similarity. Tolerance indices attempt to assess invertebrate ability to withstand stressors. One common general tolerance index is EPA's 0 to 10 scale tolerance values (USEPA 1997, Barbour et al. 1999), and other more specific tolerance indices include the Hilsenhoff Biotic Index (HBI) (Hilsenhoff 1987, 1988) or The Florida Index (Ross and Jones 1979) for tolerance of organic pollution, and Biotic Condition Index (Winget and Mangum 1979) for tolerance of sedimentation. Diversity indices use both abundance and richness to estimate how large of a variety of different kinds of invertebrate taxa live in a particular stream reach. Some commonly used diversity indices include Shannon's index (Shannon 1948), Simpson's index (Simpson 1949), and Margalef's index (Margalef 1958). The goal of similarity indices, on the other hand, is to compare the communities of invertebrates at different sites to determine how compositionally similar they are. Whereas tolerance and diversity based indices assign scores to a single site at a time, independent of other sites, similarity indices are meant to quantify similarities between sites and a score for a site is interpreted in context of other sites. Common similarity indices include Jaccard coefficient (Jaccard 1912), Bray-Curtis dissimilarity (Bray and

Curtis 1957), and Pinkham-Pearson Community Similarity index (Pinkham and Pearson 1976).

Another method of comparing information found in a set of samples is ordination analysis. Multivariate ordination techniques can be used to condense correlation information found in invertebrate measures into hypothetical latent variables called axes. These axes attempt to explain variation in the dataset and can be attributed to variation causing factors such as urbanization. Several such ordination techniques include principle components analysis (PCA) (Hottelling 1933), principle coordinates analysis (PCoA) (Gower 1966), and nonmetric multidimensional scaling (NMDS) (Clarke and Gorley 2001).

1.3. Biological Condition Gradient

Although the amount and composition of benthic macroinvertebrate taxa found at a stream sample site can tell much about the biological health of that ecosystem (Lenat and Barbour 1994, Barbour et al. 1999, USEPA 2005), the interpretation of these invertebrate community metrics is very specific to the region in which they are measured and to the detailed sampling protocol used. This is makes it challenging to compare invertebrate community health across regions and sampling programs.

The Biological Condition Gradient (BCG) systematically defines six levels of ecosystem health using ten characteristics of ecosystem structure and function that

respond to increasing stress (Davies and Jackson 2006). BCG levels, or Tiers, help to standardize interpretation and communication of the condition of aquatic biota (collected with differing sampling methods and in diverse ecological settings), facilitate detection of incremental changes in ecological condition, and enable linking of management actions to meaningful environmental outcomes.

However, in order to apply this BCG conceptual framework to management scenarios, it must be quantitatively parameterized in terms of its links to regional invertebrate community metrics and, ultimately, in terms of its relationships to manageable environmental system stressors. (See Chapter 4 for additional discussion of BCG and its use as a probabilistic model endpoint.)

1.4. Effects of Urbanization on Stream Ecosystems (EUSE) Dataset

Data from a study of the Effects of Urbanization on Stream Ecosystems (EUSE) are used to characterize and understand the myriad cascading effects of urbanization. This investigation, undertaken by the United States Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Program, measured approximately 1400 variables in each of 262 watersheds. The goal of the study was to characterize the physical, chemical, and biological condition of streams on an urban gradient (Table 1). Approximately 30 streams in watersheds of differing levels of urbanization were sampled in each of 9 metropolitan areas. By substituting space for time, the EUSE study

was designed to detect the multitude of environmental and ecological stream changes occurring with increasing urbanization.

Table 1: Effect of Urbanization on Stream Ecosystems variable type summary.

General category	Specific category	Number of metrics in dataset
Urban intensity	Census	69 metrics
	Landcover	98 metrics
	Infrastructure	6 metrics
Physical	Hydrology	70 metrics
	Habitat	89 metrics
	Water temperature	33 metrics
	Climate	26 metrics
	Soils	25 metrics
	Topography	12 metrics
Chemical	Nutrients	52 metrics
	Pesticides	96 metrics
	SPMD[a] Chemistry	29 metrics
Biological	Invertebrate response	194 metrics
	Algal response	414 metrics
	Fish response	196 metrics

The EUSE studies are based on a common study design (McMahon and Cuffney 2000, Tate et al. 2005) and consistent measures of urban intensity (Cuffney and Falcone 2008) and sample-collection and processing methods (Fitzpatrick 1998, Moulton 2002). Urban intensity was measured using land cover and census data to report variables such as percent urban land cover in the watershed basin, percent impervious cover in the watershed basin, population, housing density, and road density. Physical stream characteristics measured include hydrological metrics describing water flow, habitat

metrics describing channel shape, light and litter input, soil size and texture, air and water temperature, precipitation, topography and many other measurements. Chemical measurements include various nitrogen and phosphorus nutrient indicators, a suite of pesticide concentrations, and many organic compound measurements using semipermeable membrane devices. Biological variables include community metrics describing fish, invertebrates and algae. EUSE studies have already been used to describe the effects of urbanization on fish (Brown et al. 2009), benthic macroinvertebrates (Cuffney et al. 2009a), algae (Coles et al. 2009), habitat (Faith A. Fitzpatrick, U.S. Geological Survey, written commun., 2009), and water chemistry (Sprague et al. 2007).

1.5. Dissertation Objectives

The aim of this dissertation research is to develop and apply new modeling methods to assess the relationship between urban stressors and aquatic assemblage responses in streams of the United States with implications for regulation and restoration. This will be accomplished by developing models which connect urbanization to biological response and, eventually, biological condition interpretation at both regional and local scales while simultaneously characterizing uncertainty and incorporating expert knowledge. Three major gaps in basin-level, single-predictor regression urbanization modeling addressed by this dissertation include: (1)

understanding and quantifying differences between regional responses of biota to urbanization, (2) incorporating influences of multiple stressors acting simultaneously, and (3) linking the effects of urbanization to a management endpoint.

Specifically, these gaps will be addressed by: (Chapter 2) using multilevel hierarchical modeling to connect urbanization to ecosystem response at two, nested scales across nine metropolitan regions of the conterminous United States, (Chapter 3) developing Bayesian network construction methodology to describe and quantify the interrelated system of urbanization impacts on aquatic invertebrate community metrics in streams of the Southeast United States, and (Chapter 4) applying that Bayesian network methodology to interpret invertebrate response in terms of probability of attaining ecosystem quality goals as defined by Biological Condition Gradient (BCG) tier in the Northeast United States. Each chapter either already is or will be published separately in peer-reviewed scientific investigation reports and journal articles.

Dissertation structure for each chapter follows Fisher's (1922) general framework of a modeling process: model formulation, followed by parameter estimation, concluding with model evaluation and interpretation. For both multilevel models and Bayesian network models, deciding how to best represent relationships between concepts (ultimately, variables) to address the question of interest was key. This first step of the modeling process requires an assessment of scientific knowledge about the

system of interest and the translation of that knowledge into mathematical form, being careful to balance complexity with parameterizeability. Both types of model developments required extensive consultation with subject matter experts to determine how to best characterize the desired aspects of urbanization effects. Then, model parameter estimation is typically the easiest step, once model formulation is established, since it only requires simple mathematical calculations given model form and data. Once multilevel and Bayesian network model structure were created, it was elementary to use R and Hugin, respectively, to calculate two levels of regression coefficients and conditional probability tables. Finally, evaluating the completed model and interpreting the results is crucial in order to be able to learn from models and continue the iterative process of scientific knowledge generation. Multilevel hierarchical model evaluation involved comparison of different invertebrate response patterns as a result of different regional factors and assessment of goodness of fit through DIC (Deviance Information Criteria). Bayesian network model evaluation was more challenging due to the multidimensional nature of the modeling framework, and a large contribution of this dissertation is the development and assessment of Bayesian network evaluation techniques, including prior weight method comparison, predictive and diagnostic probability consideration, mutual information, sensitivity analysis, and conditional probability table row averaging and comparison.

This work is significant in the field of environmental modeling because it will result in the development of effective new methods of analysis for extensive stream ecosystem datasets. Using these powerful models, we can understand relationships in urban streams at multiple tiers and point to systemic solutions for minimizing the effect of urbanization. Additionally, by utilizing a probabilistic framework this work will integrate component uncertainties and their interactions into a thorough evaluation of the processes and relationships of the entire system.

This research is original in that I develop a new analysis methodology with which to reinterpret a comprehensive existing dataset (United States Geological Survey's Effects of Urbanization on Stream Ecosystems; USGS EUSE). It is clear that the process of urbanization affecting aquatic ecosystems is complex and consists of many interacting factors on multiple levels. Previous analysis of stream ecosystem data attempts to find empirical relationships between pairs of compartments, without incorporating additional influential variables or known ecological information about the system. In contrast, not only do these research models link urbanization to ecological responses but they do so through linking together a suite of all major system components into an interactive, parameterized network. In this way, we can now model the effect of a change in one variable on every other component in the entire system, enabling multidimensional data interpretation in a way not previously possible. This network

can eventually function as a decision analysis tool for ecosystem managers by also including a link to the probability of attaining a meaningful, defined water quality condition or standard (BCG tier).

Major contributions to the literature include: (1) the development and use of a statistically more rigorous and ecologically more interpretable multilevel hierarchical modeling methodology, (2) creation of a systematic method of prior Bayesian network model elicitation from expert knowledge, (3) updating of an expert prior Bayesian network model with data (whereas Bayesian network development typically stops at prior model construction), (4) developing model evaluation procedures to compare a Bayesian network prior to data to posterior, (5) exploring problem areas of a Bayesian network model approach in environmental problem solving, (6) parameterization of the Biological Condition Gradient (BCG) relative to urban stressors, and (7) introducing a Bayesian network framework for urbanization management modeling.

2. Multilevel Hierarchical Modeling of Benthic Macroinvertebrate Responses to Urbanization in Nine Metropolitan Regions across the Conterminous United States

Note: the text from this Chapter appears in the peer-reviewed U.S. Geological Survey Scientific Investigations Report 2009-5243 (Kashuba et al., 2010). This Report was co-authored by YoonKyung Cha, Ibrahim Alameddine, Boknam Lee, and Thomas Cuffney. It was completed initially as part of the work from the course "Topics in Environmental and Ecological Statistics" led by Dr. Song S. Qian of the Nicholas School of the Environment at Duke University through a cooperative agreement between the USGS and Duke University. Dr. Qian was instrumental in teaching of multilevel hierarchical modeling philosophy and providing statistical consult throughout the course of this analysis. The authors thank Dr. Kenneth H. Reckhow of the Nicholas School of the Environment at Duke University for project conception and funding acquisition. Dr. Gerard McMahon of the U.S. Geological Survey expertly managed and guided the continuation of this work to meet USGS study goals. Dr. Andrew Gelman of Columbia Unviersity and Jason May of the U.S. Geological Survey provided insightful technical reviews resulting in the improvement of this Report.

2.1. Introduction

Stream ecosystems are increasingly affected by urban development associated with human population growth (Booth and Jackson 1997, Paul and Meyer 2001, Walsh et al. 2001, Tate et al. 2005, Walsh et al. 2005). Deforestation destroys riparian buffer zones

and leads to declines in canopy cover, changes in energy inputs, and increases in water temperatures (Waite and Carpenter 2000, Jacobson 2001, Sprague et al. 2006). Residential and industrial development introduces human waste, pesticides, and industrial chemicals into the water and sediment (Hall and Anderson 1988, Pitt et al. 1995, Van Metre et al. 2000, Mahler et al. 2005, Gilliom et al. 2006). Impervious surfaces reduce rainfall infiltration, increase surface runoff, and alter the frequency and magnitude of peak and base flows (Klein 1979, Poff et al. 1997, USEPA 1997, Finkenbine et al. 2000, Konrad and Booth 2002, McMahon et al. 2003, Roy et al. 2005b). Altering the hydrology changes channel morphology, degrades aquatic habitats (Winterbourn and Townsend 1991), and increases sedimentation rates (Wolman and Schick 1967, Trimble 1997, Sponseller et al. 2001, Roy et al. 2003b). Changes in land cover, hydrology, and impervious surface also affect stream temperature (Sinokrot and Stefan 1993, LeBlanc et al. 1997, Paul and Meyer 2001). Collectively, these and other changes in the physical and chemical environment have been associated with degraded invertebrate assemblages in many urban areas (Klein 1979, Jones and Clark 1987, Schueler and Galli 1992, Lenat and Crawford 1994, Yoder and Rankin 1996, Horner et al. 1997, Kemp and Spotila 1997, Kennen 1999, Yoder et al. 1999, Beasley and Kneale 2002, Huryn et al. 2002, Kennen and Ayers 2002, Morley and Karr 2002, Morse et al. 2003, Ourso and Frenzel 2003, Vølstad et al. 2003, Roy et al. 2003a, Fitzpatrick et al. 2004, Brown and Vivas 2005).

In 1999, the U.S. Geological Survey (USGS) initiated a series of urban stream studies (Effects of Urbanization on Stream Ecosystems, EUSE) as part of the National Water- Quality Assessment (NAWQA) Program. The EUSE studies are based on a common study design (McMahon and Cuffney 2000, Coles et al. 2004, Cuffney et al. 2005, Tate et al. 2005) and consistent measures of urban intensity (Cuffney and Falcone 2008) and sample-collection and processing methods (Fitzpatrick et al. 1998, Moulton et al. 2002). Nine major metropolitan regions—Boston, MA (BOS); Raleigh, NC (RAL); Atlanta, GA (ATL); Birmingham, AL (BIR); Milwaukee–Green Bay, WI (MGB); Denver, CO (DEN); Dallas–Fort Worth, TX (DFW); Salt Lake City, UT (SLC); and Portland, OR (POR) (Figure 1)—were chosen to represent the effects of urbanization in regions of the country that differ in potential natural vegetation, air temperature, precipitation, basin relief, elevation, and basin slope (Table 2).

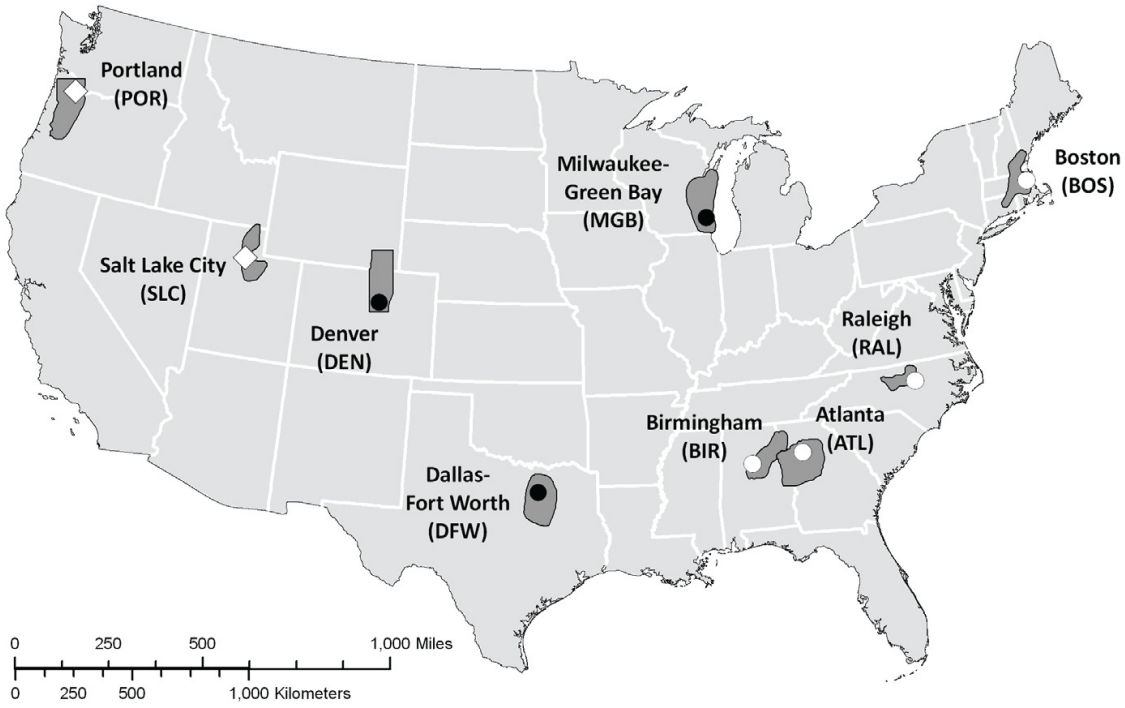

Figure 1: Locations of the nine metropolitan regions for which benthic macroinvertebrate responses to urbanization were modeled. [White circles indicate East regions; black circles indicate Central regions (high antecedent agriculture); white diamonds indicate West regions]

Table 2: Major environmental characteristics of the nine metropolitan regions. [Antecedent agricultural land cover combines row crop and grazing lands at candidate basins with MA-NUII ≤ 10. MA-NUII is the metropolitan area national urban intensity index.]

Metropolitan region	Antecedent agricultural land cover (%)	Mean annual air temperature (°C)	Mean annual precipitation (cm)	Number of candidate basins	Number of basins in EUSE study
Atlanta, GA (ATL)	17.4	16.3	133.5	116	30
Birmingham, AL (BIR)	15.0	16.0	146.8	854	28
Boston, MA (BOS)	10.3	8.7	123.2	76	30
Denver, CO (DEN)	88.0	9.2	43.0	204	28
Dallas-Fort Worth, TX (DFW)	81.7	18.3	104.2	166	28
Milwaukee-Green Bay, WI (MGB)	79.4	7.6	85.5	56	30
Portland, OR (POR)	16.9	10.8	152.8	148	28
Raleigh, NC (RAL)	24.4	14.9	119.2	871	30
Salt Lake City, UT (SLC)	12.2	9.7	68.0	9	30

Each metropolitan region represents a geographically extensive region that includes numerous cities and towns in addition to the one for which it is named. EUSE studies have been used to describe the effects of urbanization on fish (Brown et al. 2009), benthic macroinvertebrates (Cuffney et al. 2009b), algae (Coles et al. 2009), habitat (Faith A.

Fitzpatrick, U.S. Geological Survey, written commun., 2009), and water chemistry (Sprague et al. 2007).

2.1.1. Purpose and scope

The purpose of this Chapter is to develop and document an innovative multilevel hierarchical modeling framework that can be used to describe the response of benthic macroinvertebrates to urbanization (percentage of basin area in developed land, URB) and climatic factors (precipitation and air temperature) within and across the nine metropolitan regions that are included in the EUSE studies. The scope of benthic invertebrate response is limited to four assemblage metrics: nonmetric multidimensional scaling first axis ordination basin scores (NMDS1), total taxa richness (RICH), combined richness of Ephemeroptera, Plecoptera, and Trichoptera (EPTRICH), and richness-weighted mean tolerance of taxa at a basin (RICHTOL). The derivation of the models developed in this study involves:

1. Description and analysis of the selected response variables along with the considered predictor variables ,
2. Description of the methodology used to develop and assess the multilevel hierarchical models linking invertebrate assemblage responses to URB, precipitation, and air temperature,

3. Development of multilevel hierarchical models that combine both local (basin) and regional variables within the model structure to predict the response of invertebrate assemblages to increased urbanization,
4. Assessment of interactions between urbanization, climate, and the condition of stream benthic invertebrate assemblages,
5. Comparison of the results generated by the different models and the metrics used to characterize invertebrate responses, and
6. Identification of the main constraints associated with the modeling methodology, selected computational execution techniques, and model formulation.

The multilevel hierarchical models described in this Chapter link invertebrate responses to urbanization and important climate parameters (precipitation and air temperature) within each metropolitan region while simultaneously explaining differences in the rates at which invertebrate assemblages respond to urbanization across the nine metropolitan regions. These models, in combination with the four invertebrate response metrics, are used to identify appropriate predictor variables and invertebrate metrics for describing changes in the condition of the invertebrate assemblages. Demonstration of the utility of the multilevel hierarchical models will serve as a template for modeling the response of other biological indicators (fish and algal assemblages) to increased urbanization.

This Chapter does not incorporate a thorough analysis of all possible region-level influences. The selection of invertebrate response and environmental variables was guided by previous analysis (Cuffney et al. 2009b). Of all EUSE data collected, it is possible that there are other region-level variables (such as chemical concentrations, elevations, and many more) that account for differences between regions in intercept and slope. The focus of this project is the development of analytical statistical methodology and does not incorporate a thorough scientific analysis to determine all possible factors at which scales affect species composition, abundance, and richness in stream ecosystems. The ecological community effects predicted from urbanization indicators are broad and hard to generalize. Because there are many different species of macroinvertebrates, all potentially affected by different influences, finding predictor variables that apply to the entire aquatic assemblage is challenging. The goal of this research effort is to introduce a new way of looking at urbanization effects modeling data, to explain why this modeling approach makes sense in this context, to show exactly how to fit these types of models (including providing R code in Appendix A), and to demonstrate how to interpret the results ecologically. Ultimately, the multilevel hierarchical models can be used to describe and quantify ecologically relevant relations and can become a useful tool for ecological data analysis and interpretation.

2.2. Methods

The nine urban studies were conducted using a common study design (McMahon and Cuffney 2000, Coles et al. 2004, Cuffney et al. 2005, Tate et al. 2005) that used nationally available Geographic Information System (GIS) variables (Falcone et al. 2007) to define a population of candidate basins (typically basins drained by second to third order streams) from which 28–30 basins were selected to represent a gradient of urbanization within each region. Local and national GIS variables that represented the natural environmental setting (for example, ecoregion, climate, elevation, stream size) were used to minimize the effects of environmental variability by dividing candidate basins into groups with relatively homogenous environmental features. Urban intensity was defined for each candidate basin by combining housing density, percentage of basin area in developed land cover, and road density into an index (metropolitan area national urban intensity index, MA-NUII) scaled to range from 0 (little or no urban) to 100 (maximum urban) within each metropolitan region (Cuffney and Falcone 2008). Once groups of basins with relatively homogeneous environmental features were defined, 28–30 basins were selected to represent the gradient of urbanization in each metropolitan region. This spatially distributed sampling network was intended to represent changes in urbanization through time (that is, substitute space for time).

Conditions in each basin were verified by field reconnaissance. If conditions in a basin deviated substantially from what was expected, based on the GIS data (for

example, substantial new development that was not present on the GIS coverage), or if a sampling reach (150-meter [m] stream section at the outflow of the basin) was disturbed by local-scale effects (for example, channelization, road or building construction), an alternate basin from the same group or a group with similar natural environmental characteristics was selected to represent the same level of urban intensity. The BOS, BIR, and SLC metropolitan regions were studied during 1999–2000; ATL, DEN, and RAL were studied during 2002–2003; and DFW, MGB, and POR were studied during 2003–2004. Details of the study designs are discussed in Coles and others (2004), Cuffney and others (2005), and Tate and others (2005).

The SLC design differed from the other metropolitan regions in that many of the basins were nested, one within another. This modification was necessary because of the small number of streams that are present in SLC and was only possible because urban development in the SLC basins has progressed upstream over time, which ensures that urban intensity increases downstream. The SLC landscape characterizations were restricted to the portions of the basins that were located in the Central Basin and Range ecoregion (Omernik 1987). Portions in the Wasatch and Uinta Mountains ecoregion were excluded because no urban development has occurred in this area, and the biology and geomorphology of the streams in this ecoregion are different from the Central Basin and Range ecoregion. Large reservoirs in DEN constituted major discontinuities that effectively isolated the upper and lower portions of many of the candidate basins.

Consequently, landscape characterizations in DEN were restricted to the portions of the basins that were below the major reservoirs.

2.2.1. Data collection

All data used in this modeling effort were collected as part of the EUSE studies. Of all measured EUSE variables, this Chapter specifically includes analysis of only benthic macroinvertebrate data, land-cover data, and climate data.

2.2.1.1. Benthic Macroinvertebrate Data

The NAWQA Program sampling protocols were used to collect benthic macroinvertebrates over a 1–4 week period during summer low base flows (Cuffney and others [1993] was used for BOS, BIR, and SLC; Moulton and others [2002] was used for ATL, DEN, DFW, MGB, POR, and RAL). Five quantitative richest targeted habitat (RTH) samples were collected from five riffles in each sampling reach using a Slack Sampler (1.25 square meters [m2] total area sampled) except in ATL, DFW, and one SLC basin (Kays Creek at Layton, UT) where woody snags were sampled (1.4 m2 mean snag area sampled) because riffles were not available. Samples were preserved in 10-percent buffered formalin and sent to the USGS National Water Quality Laboratory in Denver, CO, for taxa identification and enumeration (Moulton et al. 2000). The USGS Invertebrate Data Analysis System (IDAS; Cuffney 2003) was used to resolve taxonomic ambiguities and calculate assemblage metrics and diversity measures. Ambiguous taxa (Cuffney 2003) were resolved independently for each metropolitan region by

distributing ambiguous parents among children (DPAC-Ck) for quantitative samples and deleting ambiguous parents (RPKC-C) for qualitative samples. These options have been shown to be suitable for analyzing responses along urban gradients (Cuffney et al. 2007). Invertebrate attribute data (tolerances and functional groups) were optimized for four regions of the country—mid-Atlantic (BOS), southeast (ATL, BIR, RAL, DFW), midwest (MGB), and northwest (DEN, SLC, POR) on the basis of the attributes compiled by Cuffney (2003). Quantitative richest targeted habitat data (RTH) were converted to densities (number per square meter) prior to resolving ambiguous taxa and calculating assemblage metrics.

2.2.1.2. Land-Cover Data

Land-cover data for ATL, BOS, BIR, RAL, and SLC were based on the National Land Cover Data 2001 (NLCD01) dataset (USGS 2005). Land-cover data for POR were derived (Falcone et al. 2007) by using the NLCD01 class structure to process data from the National Oceanic and Atmospheric Administration (NOAA) Coastal Change Analysis Program (NOAA 2005). NOAA land-cover classes were recoded to match the NLCD01 classes. Land-cover data for DEN, DFW, and MGB were derived using identical methods and protocols as the NLCD01 program (Falcone and Pearson 2006). The 16 NLCD01 land-cover classes (USGS 2005) were aggregated into eight Anderson Level I classes. For example, "deciduous forest," "evergreen forest," and "mixed forest"

were aggregated into "forest" (Anderson et al. 1976) because the broader Level I classes were deemed to be more reliable than the Level II classes (Falcone et al. 2007).

2.2.1.3. Climate Data

Mean monthly precipitation (in centimeters) and air temperature (in degrees Celsius) were derived for each of the candidate drainage basins on the basis of 1-kilometer (km) resolution (Daymet 2005) model data. These data represented 18-year (1980–1997) temperature and precipitation means obtained from terrain-adjusted daily climatological observations (Falcone et al. 2007). Region-level mean annual temperature and mean annual precipitation were obtained by averaging the annual temperature and annual precipitation for the candidate basins in each metropolitan region.

2.2.2. Data summary

The dataset for the EUSE studies includes variables that characterize the biological, chemical, and physical conditions of 261 basins located in nine metropolitan regions. In this section, the variables that were analyzed are discussed, including six urbanization indicators that delineate census and infrastructure, four assemblage metrics of benthic macroinvertebrates, six land-cover variables, two climate parameters (ambient air temperature and precipitation), along with two fragmentation and one basin size variables. These variables were grouped into four main categories: (1) urbanization measures, (2) macroinvertebrate response variables, (3) land-cover types, and (4) scale-dependent variables. In addition to summarizing the characteristics of the

variables of primary concern, this section spans preliminary analyses, such as simple regressions of these variables, to offer a rationale for the methodology selected and variables incorporated in further analysis using multilevel hierarchical regression models.

2.2.2.1. Urbanization Measures

The process of urbanization is a complex, multidimensional, and dynamic process that is difficult to quantitatively define. As such, it is often hard to identify a suitable indicator that is capable of adequately characterizing the degree of urban development in a region. Several studies have used impervious surface area to represent urban gradients and their effects on stream biota. The degree of imperviousness was shown to affect the stream ecosystem by altering the hydrology and geomorphology of the stream. More frequent and larger floods, increased peak flows (and reduced base flow), and an acceleration of bed and bank erosion were observed to occur with increases in impervious surfaces (Klein 1979, Schueler 1994, Finkenbine et al. 2000, Walsh et al. 2001, Morse et al. 2003). However, Karr and Chu (2000) showed that imperviousness was not capable of explaining other crucial influences of urbanization, such as loss of the riparian cover.

Recent studies (Morley and Karr 2002, Alberti et al. 2007, Horwitz et al. 2008) have used the percentage of urban area in a basin to determine urbanization effects on stream ecology, while others have tried to use population density (Jones and Clark

1987), building density, and paved road density (Bolstad and Swank 1997) to describe the effects of urbanization on the condition of stream biota. Given the heterogeneity of urbanization processes and the diversity of background conditions at each site, finding a single urbanization surrogate that clearly correlates with effects on aquatic systems is challenging. The EUSE studies addressed this challenge by combining the percentage of developed land in the basin with road density and housing unit density to develop indices that describe urban intensity (Cuffney and Falcone 2008). These indices were scaled to represent urbanization in each metropolitan region (metropolitan area national urban intensity index, MA-NUII) as well as nationally (national urban intensity index, NUII). The NUII scaling compensates for differences in the rates at which the urbanization variables change among metropolitan regions as a function of population density. That is, the NUII accounts for the effects of basin and regional scales on the measurement of urban intensity, whereas the MA-NUII does not.

In this study, six potential urbanization measures were analyzed: national urban intensity index (NUII), percentage of urban area (URB), densities of housing units (HU), population density (POP), road density (RD), and percentage of impervious surface area (IS). The NUII represents the degree of urbanization across all nine metropolitan regions ranging from non-urban 0 to fully urban 100. Figure 2 presents box plots of the six urbanization measures that were analyzed for each of the nine regions. The figure shows that SLC exhibited strongest urban intensity, while DFW showed the lowest.

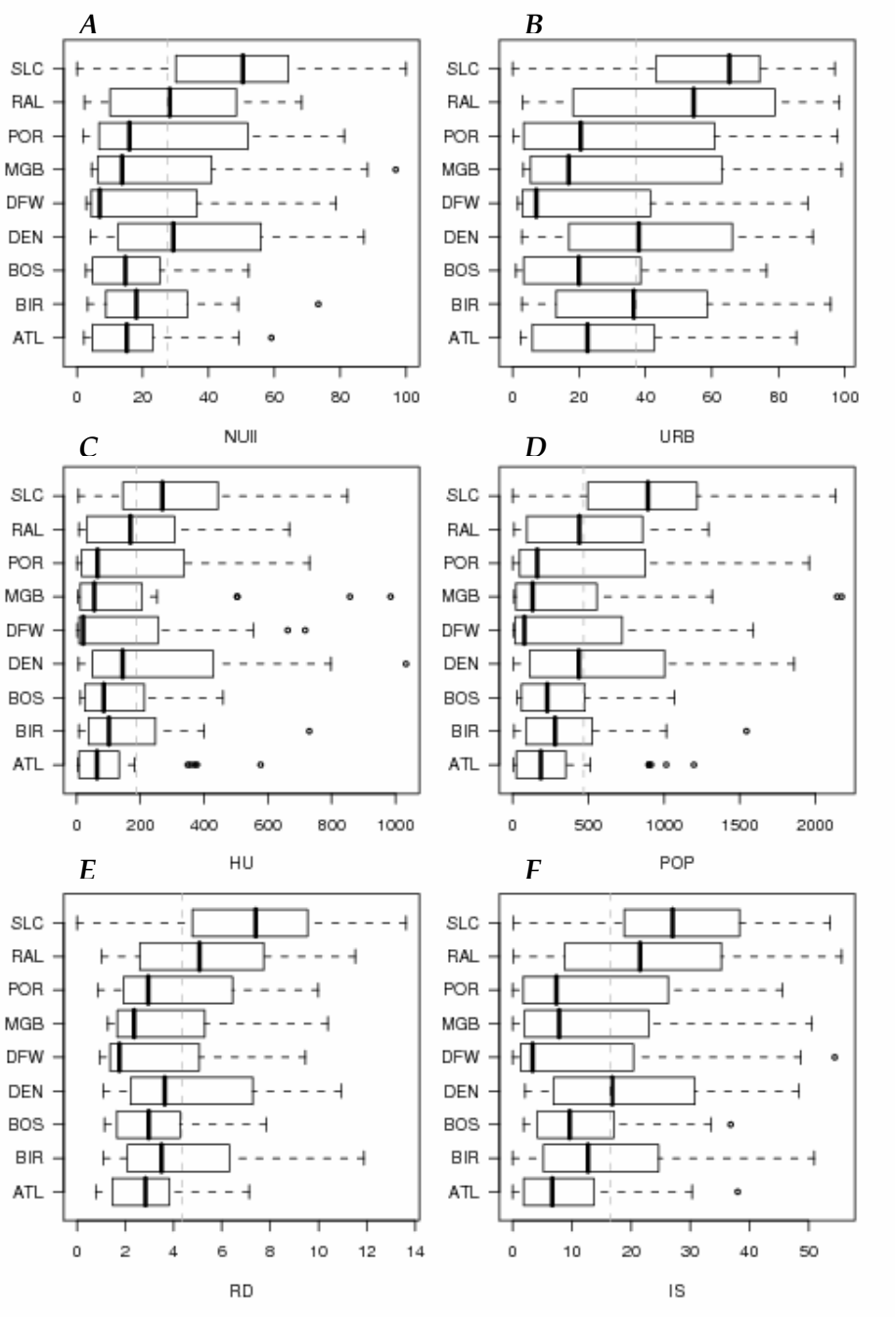

Figure 2: Box plots of urbanization measures: (A) NUII, national urban intensity index; (B) URB, percent urban area; (C) HU, housing unit density (units/km²); (D) POP, population density (people/km²); (E) RD, road density (km of road/km²); (F) IS, percent impervious surface area in the drainage basin. A vertical dashed line represents the overall mean value across the nine regions.

The six urbanization measures are highly correlated (Table 3; Figure 3), so a single measure was used to describe the level of urbanization. Percentage of urban area (Figure 2*B*) was selected as the candidate urbanization surrogate because it has the broadest coverage along the urban intensity gradient, and it is easy to monitor and simple to portray to urban planners and decisionmakers. Even though the NUII integrates the other five measures of urbanization, either directly (URB, HU, RD) or indirectly through high correlation with URB (r = 0.93 for URB and IS; r = 0.95 for POP and URB), NUII was not used to represent urban intensity because the scaling of this index already incorporates the multilevel (basin and region) effects that are the objective of multilevel hierarchical regression.

Table 3: Spearman rank correlation coefficients for urbanization measures: NUII, national urban intensity index; URB, percent urban area in the drainage basin; HU, density of housing units; POP, population density; RD, road density; IS, mean percent impervious surface in the drainage basin.

	NUII	URB	HU	POP	RD	IS
NUII	1.00					
URB	0.97	1.00				
HU	0.97	0.94	1.00			
POP	0.97	0.94	0.99	1.00		
RD	0.97	0.94	0.94	0.95	1.00	
IS	0.97	0.96	0.96	0.96	0.95	1.00

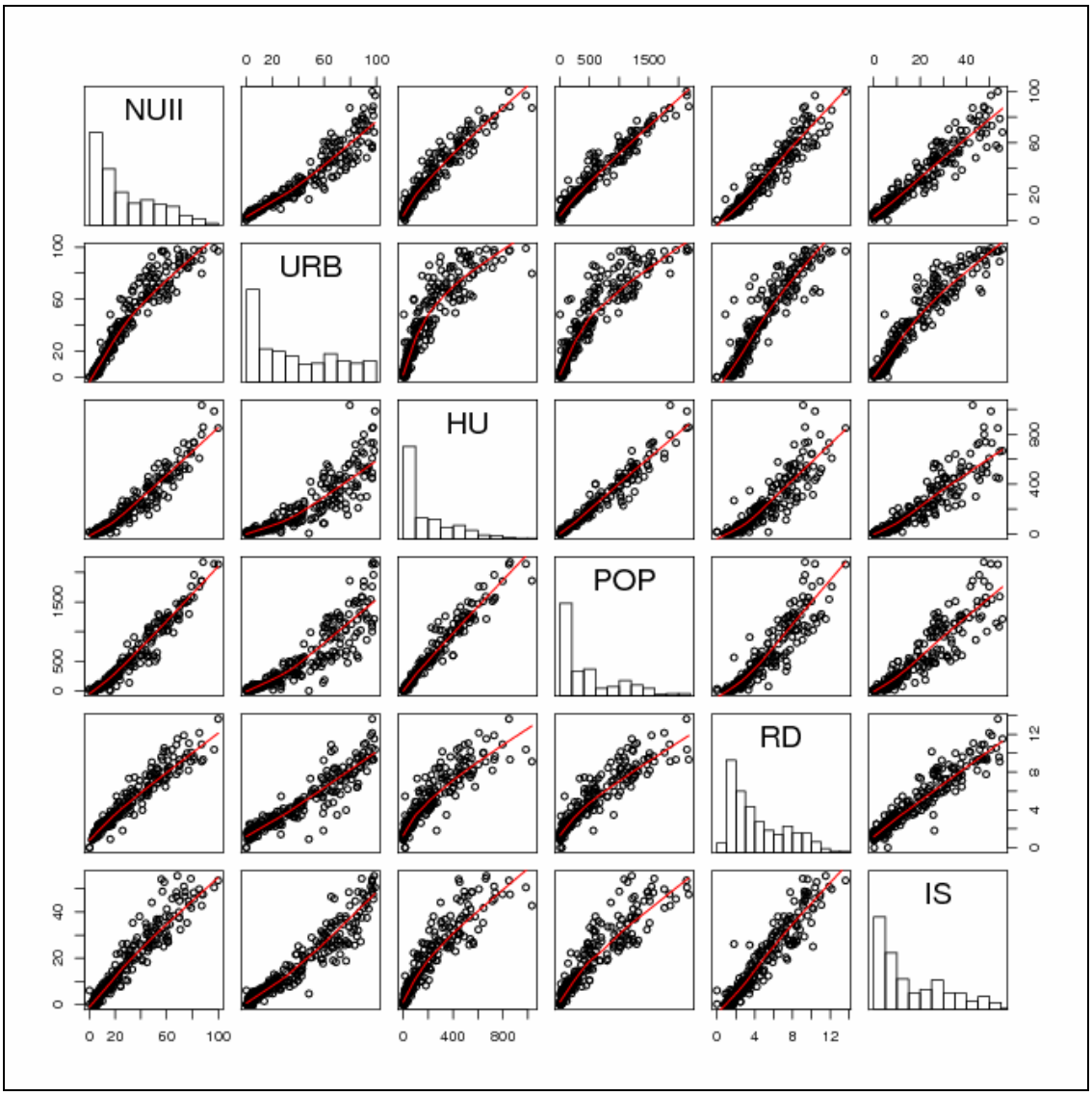

Figure 3: Histograms and scatterplots of urbanization measures: NUII, national urban intensity index; URB, percent urban area; HU, housing unit density (units/km2); POP, population density (people/km2); RD, road density (km of road/km2); IS, percent impervious surface area in the drainage basin.

2.2.2.2. Macroinvertebrate Response Variables

In this Chapter, benthic macroinvertebrate community metrics were used to represent the response of stream biota to urban stressors. Macroinvertebrates are

omnipresent in streams and show wider variety as compared to fish. Macroinvertebrates also tend to better integrate conditions over time due to their relative immobility (Lammert and Allan 1999). Several metrics describing macroinvertebrate assemblages were analyzed in comparison with a full range of human disturbances represented by urban gradients across the nine defined regions and their corresponding basins.

Previous analyses have indicated that richness metrics are more reliable indicators of urbanization than abundance metrics (Cuffney et al. 2009b). As such, two richness metrics, total taxa richness (RICH) and EPT taxa richness (EPTRICH), are included as response variables in the current study. Both RICH and EPTRICH are commonly used macroinvertebrate parameters (Wallace et al. 1996). RICH measures the number of taxa of macroinvertebrates found in a sample, while EPTRICH measures the number of taxa in the orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) found in a sample (Barbour et al. 1999). The number of taxa, both RICH and EPTRICH, tends to decrease as the condition of the aquatic system degrades (Barbour et al. 1999). Although both metrics tend to decrease as perturbation proceeds, EPTRICH has been known to be more useful than RICH in terms of evaluating water quality, as EPT insect orders are not only indicative of stream disturbances, but also easy to identify and apply (Wallace et al. 1996). Although all mayfly, stonefly, and caddisfly species are found in streams under various conditions, they are likely to be most abundant in clean waters often with high levels of dissolved oxygen. Thus,

increasing EPTRICH is indicative of increasing diversity of intolerant macroinvertebrate species, which suggests healthier aquatic environments. In a previous study, Roy et al. (2003b) examined 30 basins in the Etowah River basin in Georgia. They found that RICH ranged between 21 and 62 with a mean value of 43, while EPTRICH ranged between 3 and 31 with a mean value of 16. The current EUSE study showed comparable results with regard to taxa richness, whereby RICH ranged between 17 and 60 with a mean value of 32 (Figure 4*B*), and EPTRICH ranged between 0 and 26 with a mean value of 8 (Figure 4C). Results from this study also showed that richness, particularly EPTRICH, was strongly affected by urbanization.

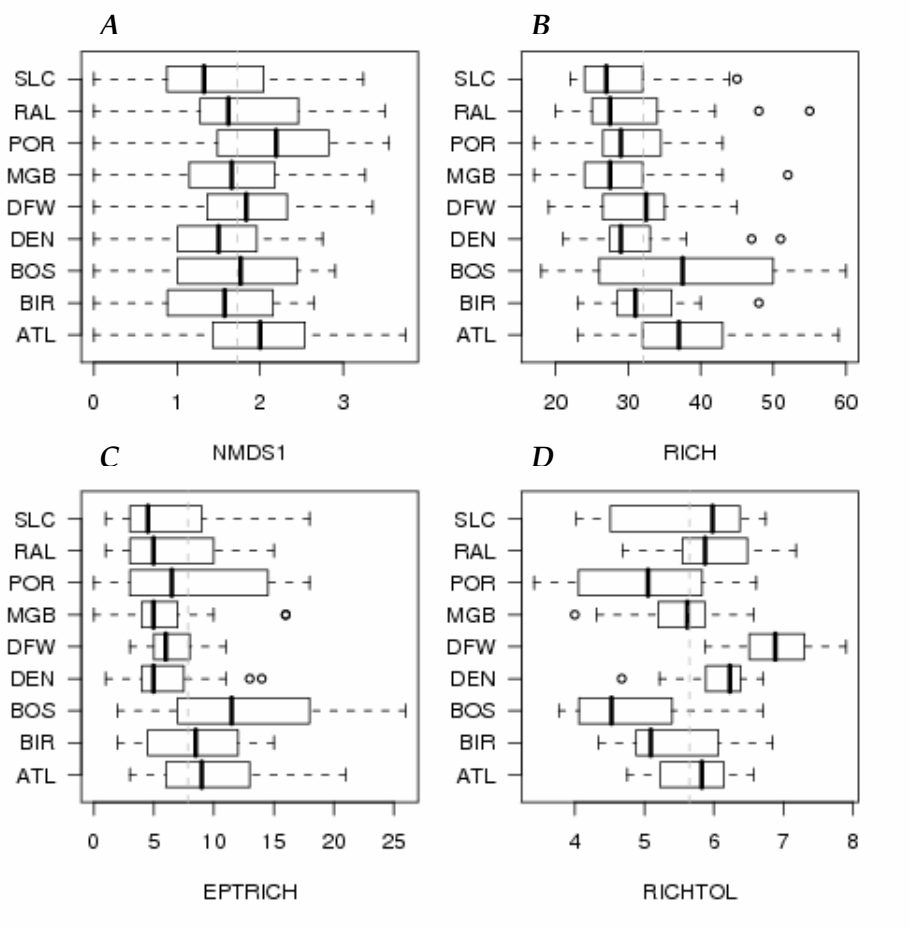

Figure 4: Box plots of macroinvertebrate response variables: (A) NMDS1, first axis adjusted nonmetric multidimensional scaling site score; (B) RICH, total taxa richness; (C) EPTRICH, EPT taxa richness; (D) RICHTOL, richness-weighted tolerance. A vertical dashed line represents the overall mean value across the nine regions.

In addition to the richness measures, a multivariate and a tolerance measure were examined as response variables. First axis ordination sample scores (NMDS1), a multivariate measure, were derived using nonmetric multidimensional scaling (Clarke and Gorley 2001) of quantitative richest targeted habitat (RTH) invertebrate samples. Separate ordination analyses were conducted for each metropolitan region using fourth-

root transformed abundance data and Bray-Curtis similarities. First axis site scores (NMDS1) were used to represent responses to urbanization as this axis was most closely associated with changes in urban intensity. NMDS is a data reduction technique that locates sites along axes that represent latent variables. That is, the process of ordination condenses information found in measures of species abundance into hypothetical variables (ordination axes) that utilize the correlation structure within the dataset to convey the same amount of information using fewer variables. The positions of sites along the axis are proportional to the similarity among the assemblages with sites with similar assemblages located close to one another and sites with dissimilar assemblages located far apart. While the relative positions of sites along the axes are important, the actual site scores are relatively arbitrary, that is, how they relate (increase or decrease) to an explanatory variable (URB) is immaterial unless the response can be tied to a known biological response, such as EPTRICH (decreases with increasing urbanization). Consequently, the NMDS1 scores were adjusted by either subtracting each value from the maximum score, if the scores decreased with decreasing EPTRICH, or subtracting the minimum score from each value, if the scores increased with decreasing EPTRICH (Cuffney et al. 2005). This rescaling produced adjusted NMDS1 scores that decreased as EPTRICH decreased and ranged from a maximum value at minimum urban intensity to zero at maximum urban intensity without affecting the relations among sites or the range of scores in each region. The resulting first axis adjusted ordination score

(NMDS1) can be interpreted as a measure of "ecological distance" in species composition between sample basins. Basins with similar macroinvertebrate populations have similar NMDS1 values while basins with different species have dissimilar NMDS1 values. High NMDS1 values correlate with high EPT taxa richness and vice versa. Ordination plots (axes 1 and 2) were also examined for outliers that might obscure the structure of the data. One such outlier (station 41365910437001, Bear Creek above Little Bear Creek near Phillips, WY) was removed from the Denver dataset. This basin was large (459 km^2) with very little developed land (< 2% of basin area) that was not comparable to other basins. Subsequent ordination analyses and modeling efforts excluded this site. NMDS1 ranged between 0 and 3.8 (Figure 4*A*), with higher ordination scores representing more dissimilar aquatic assemblages across the sample basins.

Furthermore, a tolerance measure, richness-based tolerance (RICHTOL), was selected as a macroinvertebrate response variable (Barbour et al. 1999, Cuffney 2003, NCDENR 2006). RICHTOL reflects the richness-weighted mean tolerance of taxa found at a sampled basin. The calculated RICHTOL metric for each basin is derived based on the tolerance values that were reported by the U.S. Environmental Protection Agency (USEPA 1997, Barbour et al. 1999) and North Carolina Department of Environment and Natural Resources (NCDENR 2006). Invertebrate tolerances are scaled from 0 to 10 (Figure 4*D*) with the most intolerant species receiving a score of 0 and the most tolerant assigned a score of 10. RICHTOL behaves differently from the other response variables

(NMDS1, EPTRICH, and RICH) in response to increasing disturbance. While NMDS1, EPTRICH, and RICH are expected to decrease with increased urbanization, RICHTOL is expected to increase (Figure 5; Table 4) as fragile species are lost in favor of more hardy macroinvertebrates that are able to persist in disturbed environments.

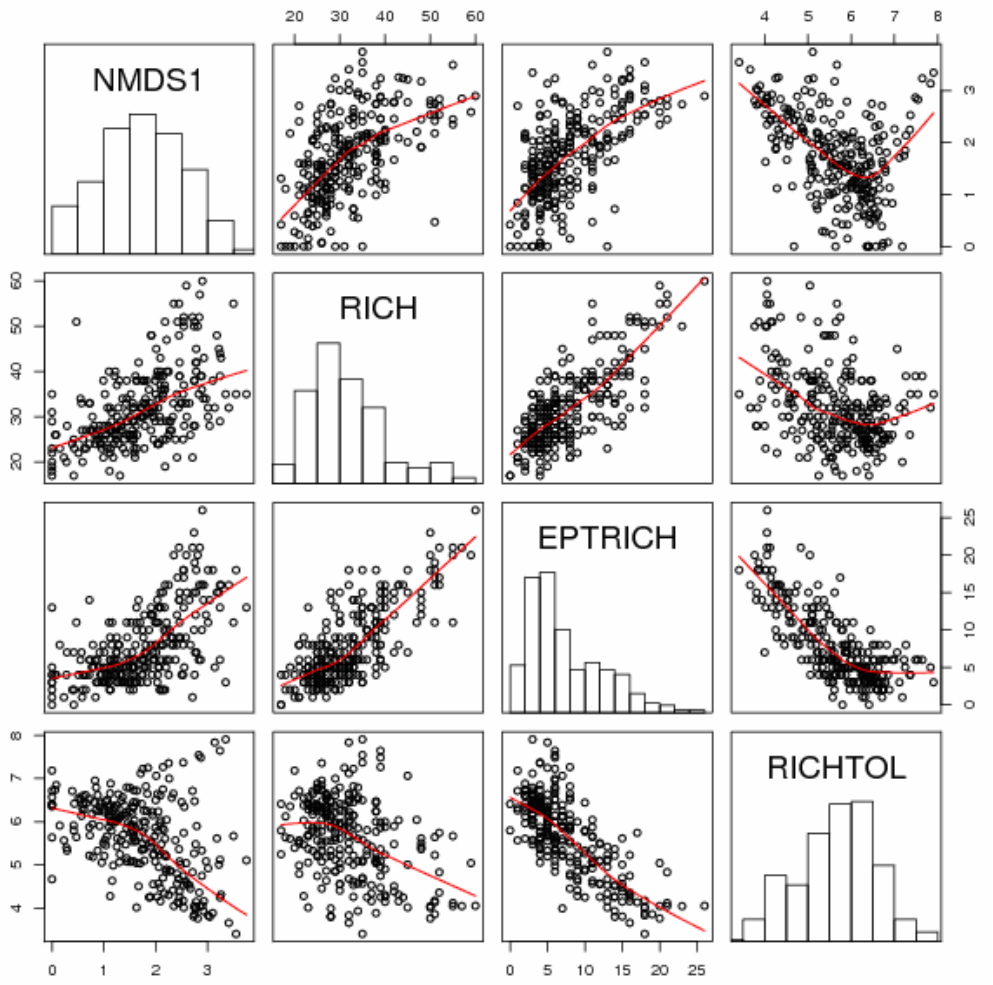

Figure 5: Histograms and scatterplots of macroinvertebrate response variables: NMDS1, first axis adjusted nonmetric multidimensional scaling site sore; RICH, total taxa richness; EPTRICH, EPT taxa richness; RICHTOL, richness-weighted tolerance.

Table 4: Spearman rank correlation coefficients for macroinvertebrate response variables: NMDS1, first axis scaled ordination score; RICH, total richness; EPTRICH, EPT taxa richness; RICHTOL, richness-weighted tolerance.

	NMDS1	RICH	EPTRICH	RICHTOL
NMDS1	1.00			
RICH	0.56	1.00		
EPTRICH	0.67	0.69	1.00	
RICHTOL	-0.54	-0.36	-0.73	1.00

EPTRICH shows strong correlations (r = 0.67 – 0.73 in absolute values) with the other response variables when all nine regions are combined (Table 4), which is expected because EPTRICH shares certain attributes with the other variables: EPTRICH is a subset of RICH, it was used as a reference to calibrate NMDS1 values, and both EPTRICH and RICHTOL represent tolerance differences among groups of macroinvertebrates. In contrast, RICH, NMDS1, and RICHTOL exhibit moderate correlations (0.36 – 0.56 in absolute values) with each other (Table 4). Because of regional differences affecting invertebrate response, correlations between invertebrate metrics are even greater within regions.

2.2.2.3. Land-Cover Types

Land-cover data from the National Land Cover Dataset 2001 (NLCD01) were reclassified into six main types (Falcone and Pearson 2006): urban (NLCD2), agriculture (NLCD7 and NLCD8), forest (NLCD4 and NLCD5), water (NLCD1), wetlands (NLCD9), and barren (NLCD3). Each land-cover type is expressed as a percentage of the total basin area. The NLCD01 reclassification was conducted by redefining the agricultural

areas as those that include croplands, pastures, and grasslands, while the forest class was reclassified to include both forests and shrub lands. The focus in this study is on analyzing the effects that urbanization and agricultural land cover have on the stream macroinvertebrate communities. The decision to include agriculture alongside urbanization as a predictor of stream ecosystem health stems from previous studies that have shown that agricultural practices destabilize stream banks, affect flow regimes, increase temperature, and impair ambient water quality (Lenat and Crawford 1994, Richards and Host 1994, Roth et al. 1996, Wichert and Rapport 1998, Walser and Bart 1999, Wang et al. 2000, Booth et al. 2002). Current agricultural land cover is greater in the Midwest (DEN, DFW, and MGB) than in the East and West (ATL, BIR, BOS, POR, RAL, and STL; Figure 6). Forest land cover is not used in the current analysis as it is highly

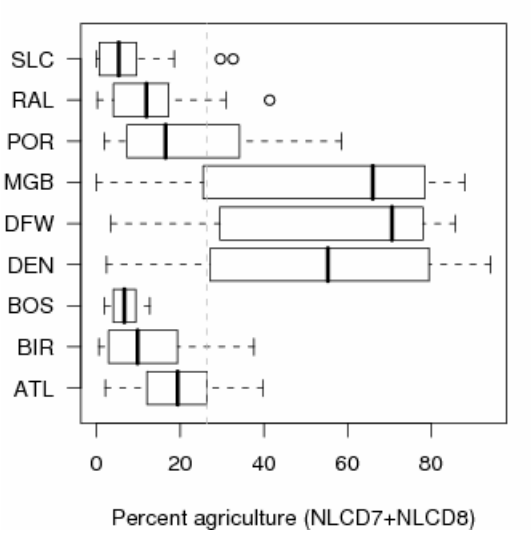

Figure 6: Box plot of the percentage of current agriculture (NLCD7 and NLCD8) in each region. A vertical dashed line represents the overall mean value across the nine regions.

correlated (negatively) with the sum of urban and agricultural land coverage, as expected both from the logic of land-use patterns and from the use of percentage data, which sums to 100 for each basin.

Diverse types of land cover, including cropland, pastures, and forests, have been developed into urban landscapes (McDonnell and Pickett 1990, Booth and Jackson 1997). A number of findings indicate decisive contribution of past land-use activity to the health of terrestrial or aquatic ecosystems (Moscrip and Montgomery 1997, Foster et al. 2003). Specifically, Harding et al. (1998) found cumulative degradation of aquatic diversity caused by long-term past agricultural activities, irrespective of mitigation efforts.

To investigate the effects of previous land use, the degree of agricultural land present in each region prior to urbanization was estimated. This antecedent agriculture (AG) was determined by calculating the mean percentage of basin area in NLCD classes 7 (grasslands) and 8 (crop and pasture lands) for non-urbanized basins (MA-NUII ≤ 10) in each metropolitan region. Grasslands were included in the estimation of antecedent agriculture because these areas are used extensively for livestock grazing. AG was calculated from the complete population of candidate basins from which the 28–30 study basins were selected. This approach provided a more extensive characterization of antecedent agricultural condition though the AG values obtained this way were very similar to AG calculated from the study basins only. MA-NUII was used to define non-

urbanized basins because it is the index upon which the urbanization gradients are defined in each metropolitan region. Performing these calculations is a way to describe past agricultural activity in these regions in the absence of readily available past records of land cover prior to urbanization. This approach follows the work of Fitzpatrick et al. (2004) who also used spatial variability to substitute for temporal changes. In this approach, mean agricultural land-cover data in non-urbanized basins at each region are extracted and used as a surrogate for representing conditions prior to the onset of urbanization. As a result of this calculation (Table 5; Figure 7), the nine study regions

Table 5: Antecedent agriculture (AG) of each region for non-urbanized basins. [AG, mean percentage of basin area in NLCD classes 7 (grasslands) and 8 (crop and pasture lands) (MA-NUII ≤ 10) in each metropolitan region. UCL, upper 95% confidence limit; LCL, lower 95% confidence limit; N, number of sample basins]

	UCL	LCL	Mean	N
ATL	19.5	15.4	17.4	116
BIR	16.0	14.1	15.0	854
BOS	11.7	9.0	10.3	76
DEN	90.9	85.1	88.0	204
DFW	82.8	80.5	81.7	166
MGB	81.4	77.2	79.3	56
POR	18.2	15.5	16.9	148
RAL	25.3	23.5	24.4	871
SLC	19.8	4.6	12.2	9

can be divided into two main categories: regions with agriculture-dominated antecedent land cover (DEN, DFW, and MGB) and regions with non-agriculture-dominant antecedent land cover (mainly forested lands; ATL, BIR, BOS, POR, RAL, and STL). These AG categories mirror patterns of current agricultural land cover (Figure 6).

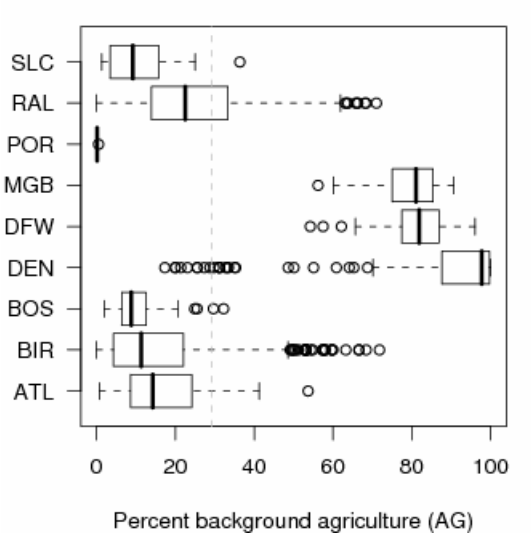

Figure 7: Estimates of the percentage of antecedent agriculture (AG) in each region. [AG is NLCD7 and NLCD8 for basins with MA-NUII ≤ 10 (N=9 to 871. A vertical dashed line represents the overall mean value across the nine regions.]

2.2.2.4. Scale-Dependent Variables

It is well known that physical processes operating at different spatial scales result in varied responses of stream ecosystems. Nonetheless, most previous studies address neither the scaling issues nor the multilevel hierarchical nature of the problem at hand. Moreover, previous studies circumvent the issue of scale by separating the effects at the region and basin levels from the effects at the local, riparian buffer level (Hunsaker and Levine 1995, Lammert and Allan 1999, Morley and Karr 2002, Strayer et al. 2003, Cuffney et al. 2005, Alberti et al. 2007). In the current study, an attempt was made to incorporate basin-level predictors with spatially larger scale (region level) variables in order to present a more thorough multiscale explanation of macroinvertebrate responses to urbanization and agricultural disturbances.

2.2.2.4.1. Basin Scale

In addition to the percentage of urban land cover at the basin level, several other basin-level variables were examined as plausible candidates for improving understanding of the response of stream ecosystems to the effects of increased urbanization at the basin level. Basin size has been used in several previous studies on the subject, whereby the size of a basin has been shown to affect the stream response by changing the physical environment of the stream (Johnson et al. 1995, Strayer et al. 2003). Nevertheless, basin size has been shown to be a poor predictor of macroinvertebrate assemblage variability (Morley and Karr 2002). In the current study, basin size is not an important parameter to include in the models given that the design of the EUSE study controlled for basin size by selecting basins in the nine regions that have relatively homogeneous environmental settings, including stream size (Figure 8).

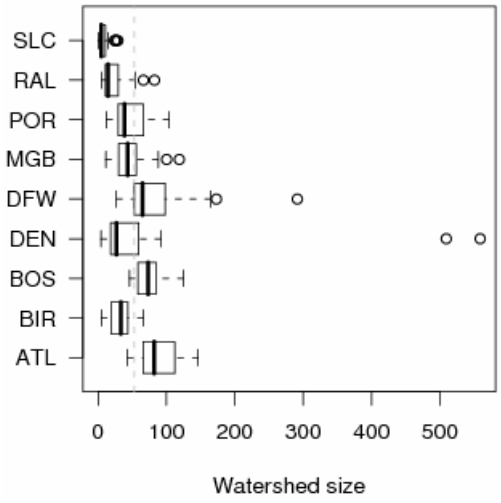

Figure 8: Box plot of drainage basin area (km²) in each region. [A vertical dashed line represents the overall mean value across the nine regions.]

Additional basin-scale variables that were considered for this study include land-cover fragmentation variables, such as patch density, largest patch index, and mean patch area. The land-cover fragmentation variables that showed strong correlations with macroinvertebrate responses also were strongly correlated with other basin-scale measures of land cover, such as percentage of basin area in developed land. Therefore, land-cover fragmentation variables are not included in the models.

2.2.2.4.2. Regional Scale

Three regional scale parameters are considered as suitable surrogates for regional scale processes that may affect the response of macroinvertebrate communities to increased urbanization. These parameters include ambient air temperature, annual precipitation, and percentage of antecedent agricultural land cover (as discussed in a previous section). Both mean annual precipitation (Figure 9*A*) and mean annual ambient

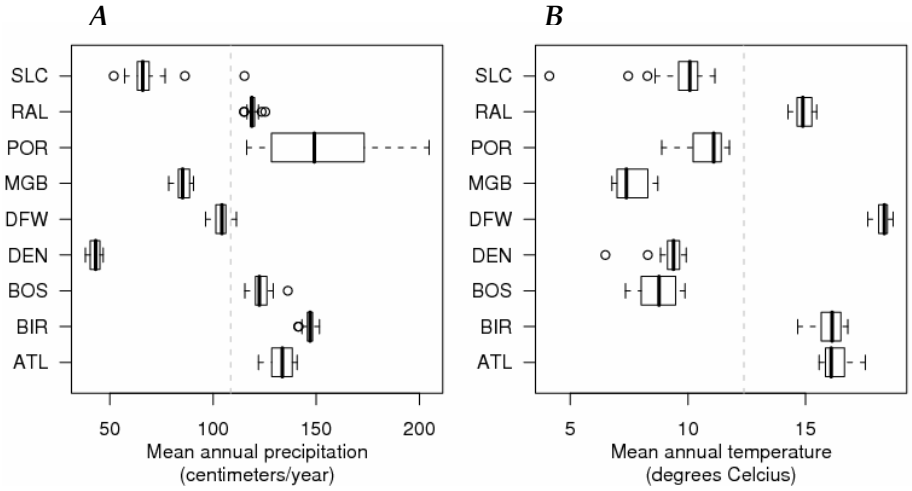

Figure 9: Boxplots of (A) mean anual precipitation and (B) mean annual ambient temperature for the period from 1980 to 1997 in each region. [A vertical dashed line represents the overall mean value across the nine regions.]

air temperature (Figure 9*B*) differ among the nine regions as the within-region variances are smaller than the between-region variances. As expected, annual precipitations are higher in coastal regions (ATL, BIR, BOS, POR, and RAL), whereas annual temperatures are higher in southern regions (ATL, BIR, DFW, and RAL). The variations in precipitation and temperature between regions may define and govern the patterns and structure of stream ecosystems at the regional scale by affecting the macroinvertebrates communities (different communities prefer different ambient conditions) along with the riparian and basin vegetation (forest as opposed to grass and shrub lands), energy inputs, channel shading, water temperatures, water chemistry, and hydrology. Specifically, temperature is known to have a role in determining large-scale distributions of invertebrates independent of urbanization (Sweeney and Vannote 1978), and the movement of precipitation through the ecosystem is known to be affected by urbanization (Walsh et al. 2005). However, at the regional scale, an annual mean temperature metric averaged over annual temporal variation may result in a less accurate description of regional climate differences relative to a cumulative precipitation metric.

2.2.3. Technique: Multilevel Hierarchical Models

A multilevel hierarchical model is a statistical modeling approach that allows the simultaneous analysis of hierarchically structured data, for example, data collected at multiple spatial scales, such as basin-level land cover and region-level precipitation.

Higher-level variables can be derived independently or by aggregating properties of lower-level variables (for example, averaging air temperatures at basins within a region to obtain region-level mean temperature). Unlike commonly used linear models, multilevel hierarchical models provide a natural framework in which to examine relations between response variables and explanatory variables that have a hierarchical arrangement. In other words, multilevel hierarchical models can be used to identify the effects of group-level variables on individual level outcomes. Multilevel hierarchical models are more statistically efficient than conventional regression models, which must use dummy variables and interactions terms in multiple linear or generalized linear regression models in order to represent multi-tiered effects.

Without multilevel hierarchical model capabilities, regression parameters for multiple groups of similar data can be calculated assuming all groups are different (unpooled) or all groups are the same (completely pooled). Group-level parameters in multilevel hierarchical models, however, are calculated by partially pooling estimates from only the group of interest (unpooled—separate analyses of each region) with estimates from the entire dataset across all groups (completely pooled—analyses for all regions combined). Both component estimates are weighted by group sample size and variation within and between groups (regions). For instance, the multilevel hierarchical mean estimate for a given group (region) *j* is estimated as follows (Gelman and Hill 2006):

$$\hat{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \quad (1)$$

where

$\hat{\alpha}_j^{multilevel}$ is the partially pooled, multilevel hierarchical mean estimate for group (region) *j*,

$\bar{y}_j$ is the unpooled mean from group (region) *j*,

$\bar{y}_{all}$ is the completely pooled mean for all groups (regions),

n_j is the sample size of group (region) *j*,

σ_y^2 is the within-group (region) variance, and

σ_α^2 is the between-group (region) variance.

In this way, partially pooled mean estimates experience shrinkage toward the completely pooled mean. That is, relative to unpooled means, partially pooled mean estimates are weighted toward the overall completely pooled mean. Level of shrinkage depends on sample size relative to within- and between-group variance. The multilevel hierarchical mean estimate $\hat{\alpha}_j^{multilevel}$ is close to the complete pooling mean (that is, greater shrinkage) when the region-level variation (σ_y^2) and sample size (n_j) are small, whereas the estimate is close to a no-pooling mean (that is, less shrinkage) when between-region variance (σ_α^2) and sample size (n_j) are large. This means that, unlike

classical regression models, multilevel hierarchical regression models can estimate model parameters even when sample sizes in some groups are very low. This is accomplished by borrowing strength from completely pooled estimates to reduce high variance in unpooled estimates from small sized groups. When there is high certainty in a group-level estimate, however, due to a large sample and large differences between groups, multilevel model estimates do not vary greatly from unpooled group estimates.

The multilevel hierarchical model framework can be illustrated by assuming that 30 basin-level samples of an ecological response (ECO) and urban condition (URB) are measured in each of three regions (A, B, and C). The three sets of basin-level data can each be modeled such that ECO response is a linear function of URB predictor within each region (Figure 10*A*–*C*). However, instead of modeling the three regions separately as three independent regressions, a multilevel hierarchical model can be constructed by constraining model parameters to be part of a higher tier region-level distribution of parameters. Conceptually, the three intercepts for Region A, B, and C regressions are not independent but belong to a distribution of intercepts (Figure 10*D*). Similarly, the three slopes for Region A, B, and C regressions belong to a distribution of slopes (Figure 10*E*). Statistically, intercept and slope variability is represented by a joint bivariate normal distribution that accounts for the correlation between intercept and slope. Figure 10 shows only the univariate distribution of intercept at mean slope (Figure 10*D*) and the

univariate distribution of slope at mean intercept (Figure 10*E*). If there are no region-level variables available to explain differences between regions, then the distribution

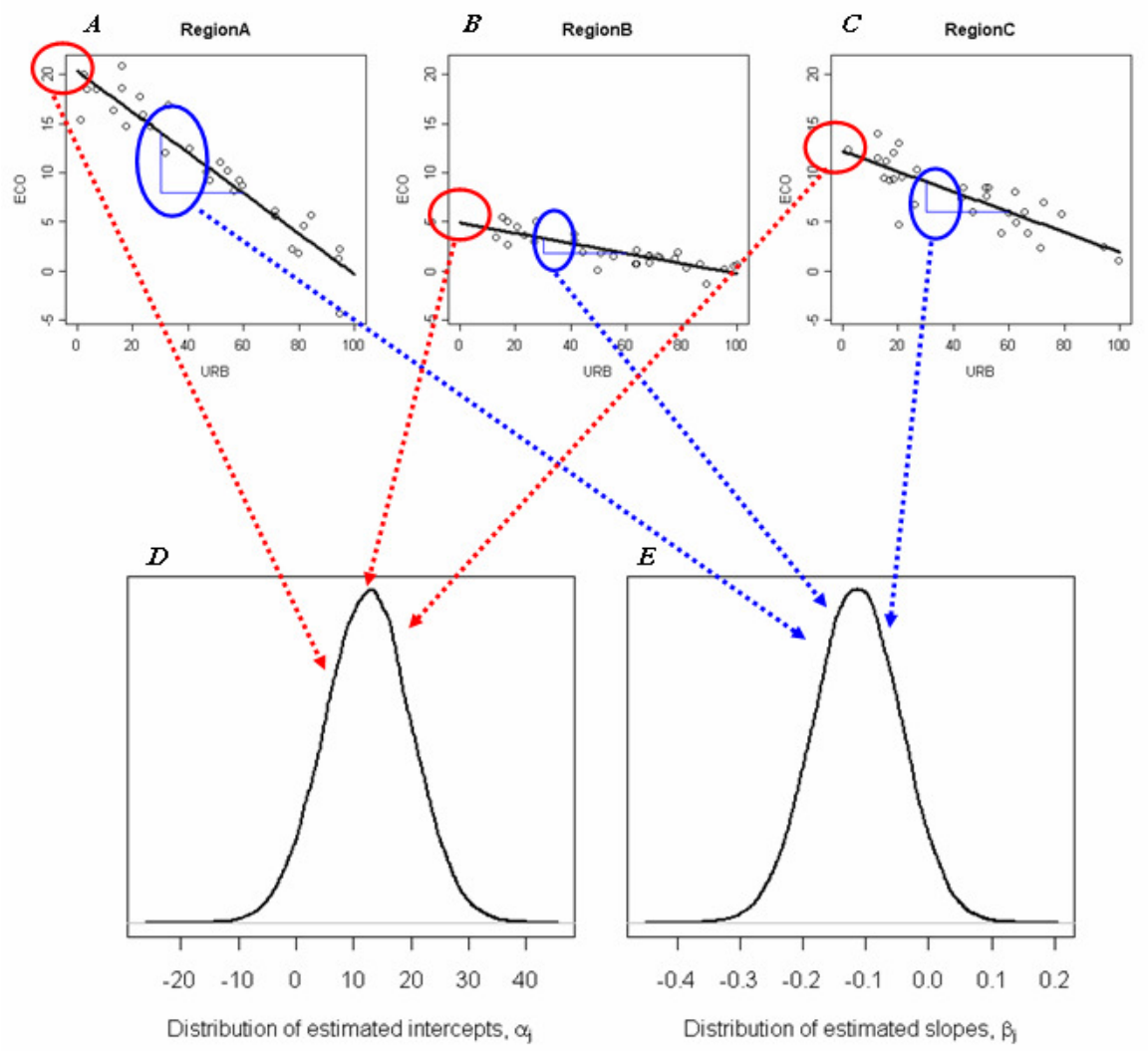

Figure 10: Multilevel hierarchical modeling framework without group-level predictors. Within each region (A, B, C), ecological response (ECO) is modeled as a linear function of urban land cover (URB) as shown in Equation 5 where ECO is y_{ij} and URB is x_{ij}. Across regions, basin-level intercepts (red) and slopes (blue) for regions A, B, and C are modeled jointly as a bivariate normal distribution with a constant mean vector as defined by eq. 6 and depicted marginally for (D) intercept and (E) slope.

of intercepts and slopes would be centered on a constant mean pair. This model structure can be represented statistically as follows. At the basin level, the following three distributions describe the top three graphs (Figure 10*A*–*C*):

$$y_{ij} \sim N(\alpha_{j=A} + \beta_{j=A} x_{ij}, \ \sigma_y^2), \ \textit{for } i=1,\ldots\ldots, \textit{30 samples} \quad (2)$$

$$y_{ij} \sim N(\alpha_{j=B} + \beta_{j=B} x_{ij}, \ \sigma_y^2), \ \textit{for } i=1,\ldots\ldots, \textit{30 samples} \quad (3)$$

$$y_{ij} \sim N(\alpha_{j=C} + \beta_{j=C} x_{ij}, \ \sigma_y^2), \ \textit{for } i=1,\ldots\ldots, \textit{30 samples} \quad (4)$$

where

i is the basin index,

j is the region index,

x_{ij} is the urban predictor variable, URB (basin level),

y_{ij} is the ecological response variable, ECO (basin level),

$\alpha_{j=A}$ is the estimated intercept for the Region A regression (region level),

$\alpha_{j=B}$ is the estimated intercept for the Region B regression (region level),

$\alpha_{j=C}$ is the estimated intercept for the Region C regression (region level),

$\beta_{j=A}$ is the estimated slope for the Region A regression (region level),

$\beta_{j=B}$ is the estimated slope for the Region B regression (region level),

$\beta_{j=C}$ is the estimated slope for the Region C regression (region level), and

σ_y^2 is the within-region variance in ecological response.

This also can be written in a compressed form as:

$$y_{ij} \sim N(\alpha_j + \beta_j x_{ij}, \sigma_y^2), \text{ for } i=1,\ldots\ldots, 30 \text{ samples and } j=A,B,C \text{ regions} \quad (5)$$

where the index *j* is used to represent the three different regional slopes and intercepts.

At the region level, the following distribution describes the bottom two graphs (Figure 10*D*, *E*):

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right), \text{ for } j= A, B, C \text{ regions} \quad (6)$$

where

α_j is the estimated intercept for the Region *j* regression (region level);

β_j is the estimated slope for the Region *j* regression (region level);

μ_α is the mean of region-level intercepts,

μ_β is the mean of region-level slopes,

σ_α^2 is between-region variance in intercept;

σ_β^2 is between-region variance in slope; and

ρ is correlation of model coefficients α_j and β_j.

This model indicates that the three regional intercepts and three regional slopes are random draws from a bivariate normal distribution of intercepts and slopes where intercepts are centered on μ_α and slopes are centered on μ_β.

In addition to just being able to represent different groups as related at a higher tier, a greater strength of the multilevel model structure is the ability to actually explain differences between these groups by incorporating a group-level predictor variable. If such region-level (group-level) information is available, intercepts and slopes can be modeled as a function of a region-level variable (Figure 11*D*, *E*). In this case, the means of the joint distribution are not constant as discussed above but change depending on the value of the region-level variable. Ideally, higher region-level intercepts can be attributed to higher values of region-level variables (such as Region A) and low intercepts to low values of region-level variables (such as Region B). This would imply a relation between the region-level variable and ecological response at zero urbanization (intercept); for example, baseline ecological response could be higher for regions with more precipitation. Intercepts and slopes can each be modeled with a different region-level variable or variables within the same model structure.

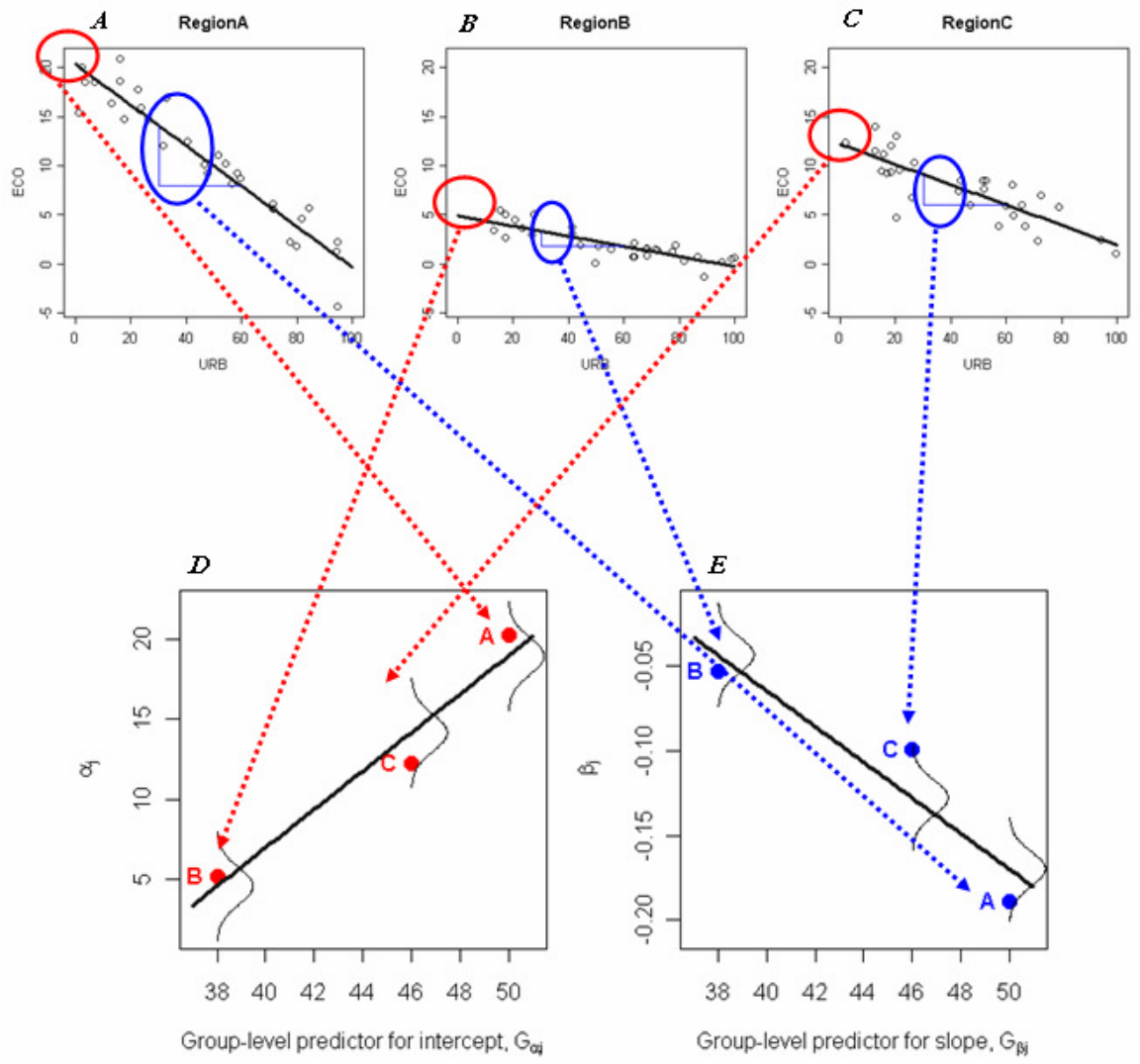

Figure 11: Multilevel hierarchical modeling framework with group-level predictors. Within each region (A, B, C), ecological response (ECO) is modeled as a linear function of urban land cover (URB) as shown in Equation 5 where ECO is y_{ij} and URB is x_{ij}. Across regions, intercepts (red) and slopes (blue) for regions A, B, and C are modeled jointly as a function of one or more group-level predictor(s) at the regional level as defined by eq. 7 and depicted marginally for (D) intercept and (E) slope.

The basin-level model structure for a hierarchical multilevel model incorporating region-level variables is statistically identical to the basin-level model structure shown

above for a model without region-level variables. At the region level, however, the following distribution describes the bottom two graphs in Figure 11*D* and *E* instead:

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_{\alpha 0} + \gamma_{\alpha 1} G_{\alpha j} \\ \gamma_{\beta 0} + \gamma_{\beta 1} G_{\beta j} \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho \sigma_\alpha \sigma_\beta \\ \rho \sigma_\alpha \sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right), \text{ for } j = A, B, C \text{ regions} \quad (7)$$

where

α_j is the estimated intercept for the Region *j* regression (region level),

β_j is the estimated slope for the Region *j* regression (region level),

$G_{\alpha j}$ is the group/region predictor variable for the intercept (region level),

$G_{\beta j}$ is the group/region predictor variable for the slope (region level),

$\gamma_{\alpha 0}, \gamma_{\alpha 1}, \gamma_{\beta 0}$, and $\gamma_{\beta 1}$ are region-level intercepts and slopes describing the relation between a region-level predictor and α_j and β_j,

σ_α^2 is between-region variance in intercept;

σ_β^2 is between-region variance in slope; and

ρ is correlation of model coefficients α_j and β_j.

2.2.3.1. Model Structure

The multilevel hierarchical models used to analyze the effects of urbanization on stream benthic macroinvertebrate assemblages for the nine EUSE metropolitan regions incorporate predictor variables at two levels, basin (URB land cover) and region (antecedent agricultural land cover, AG; mean annual precipitation, PRECIP; and mean

annual air temperature, TEMP). The combination of one basin-level and three region-level predictor variables produced eight models to evaluate (Table 6). Separate models were constructed for the four different macroinvertebrate response indicators (NMDS1, RICH, EPTRICH, and RICHTOL). Two different multilevel hierarchical model forms, linear and Poisson, are used to characterize the different response variables. Linear multilevel hierarchical models are constructed for the continuous variables NMDS1 and RICHTOL, whereas Poisson generalized linear multilevel hierarchical models are developed for RICH and EPTRICH because those variables were collected as count variables and follow the discrete Poisson distribution.

Table 6: Delineation of variables included in the eight multilevel hierarchical models per response variable. Model structure templates are presented in the Model Structure section of the Chapter. URB, percentage of urban area in the drainage basin; PRECIP, mean cumulative annual precipitation; TEMP, mean annual ambient air temperature; AG, mean percentage of basin area in row crop agriculture and grazing lands for non-urbanized basins (MA-NUII < 10).

Model number	Model structure	Basin-level predictor	Region-level predictor(s)	
			Basin intercept predictor(s)	Basin slope predictor(s)
Model 1	Template 1	URB	none	none
Model 2	Template 2	URB	PRECIP	PRECIP
Model 3	Template 2	URB	TEMP	TEMP
Model 4	Template 2	URB	TEMP	PRECIP
Model 5	Template 2	URB	Continuous AG	Continuous AG
Model 6	Template 3	URB	PRECIP and Categorical AG	PRECIP and Categorical AG
Model 7	Template 3	URB	TEMP and Categorical AG	TEMP and Categorical AG
Model 8	Template 3	URB	TEMP and Categorical AG	PRECIP and Categorical AG

The linear multilevel hierarchal models were developed using the following three template structures.

Template 1. Model 1 (no region-level predictor):

$$y_{ij} \sim N\left(\alpha_j + \beta_j x_{ij}, \sigma_y^2\right) \tag{8}$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right) \tag{9}$$

where

i=1,……, ~30 basins within each region,

j=1,……, 9 regions,

y_{ij} is basin-level ecological response of benthic macroinvertebrates (for NMDS1 or RICHTOL),

x_{ij} is basin-level urban land-cover percentage (URB),

α_j is region-level intercept (the estimated invertebrate response at URB=0),

β_j is region-level slope (the estimated change in invertebrate response per unit change in URB),

σ_y^2 is within-region variance in invertebrate response,

μ_α. and μ_β are means of region-level intercepts and slopes, respectively,

σ_α^2 is between-region variance in intercept;

σ_β^2 is between-region variance in slope; and

ρ is correlation of model coefficients α_j and β_j.

The intercept (α_j) and slope (β_j) are allowed to vary by region in order to account for the difference of urbanization effect among regions. Instead of region-level information on regional differences, intercept (α_j) and slope (β_j) are modeled as a joint bivariate normal distribution centered on a vector of constant means ($\mu_{\alpha.}$ and μ_β). The error term σ_y^2 represents variation within regions, including the measurement error and natural variation in invertebrate response metrics at a basin, and variation between basins beyond what is explained by the urbanization indicator (URB). The errors σ_α^2 and σ_β^2 represent unexplained variation between regions. The term ρ accounts for the correlation between α_j and β_j, as typically increasing intercept is associated with decreasing slope and vice versa.

Template 2. Models 2–5 (one continuous variable region-level predictor):

$$y_{ij} \sim N\left(\alpha_j + \beta_j x_{ij}, \sigma_y^2\right) \tag{10}$$

$$\begin{pmatrix}\alpha_j \\ \beta_j\end{pmatrix} \sim N\left(\begin{pmatrix}\gamma_{\alpha 0} + \gamma_{\alpha 1} P_{\alpha j} \\ \gamma_{\beta 0} + \gamma_{\beta 1} P_{\beta j}\end{pmatrix}, \begin{pmatrix}\sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2\end{pmatrix}\right) \tag{11}$$

where

i=1,……, ~30 basins within each region,

j=1,……, 9 regions,

y_{ij} is basin-level ecological response of benthic macroinvertebrates (for NMDS1 or RICHTOL),

x_{ij} is basin-level urban land cover percentage (URB),

α_j is region-level intercept (the estimated invertebrate response at URB=0),

β_j is region-level slope (the estimated change in invertebrate response per unit change in URB),

σ_y^2 is within-region variance in invertebrate response,

$P_{\alpha j}$ and $P_{\beta j}$ are region-level conditions (such as PRECIP, TEMP, or AG as a percentage) predicting intercept and slope, respectively,

$\gamma_{\alpha 0}, \gamma_{\alpha 1}, \gamma_{\beta 0}$, and $\gamma_{\beta 1}$ are hyperparameter intercepts and slopes describing the relation between a region-level predictor and α_j and β_j,

σ_α^2 is between-region variance in intercept;

σ_β^2 is between-region variance in slope; and

ρ is correlation of model coefficients α_j and β_j.

The intercept (α_j) and slope (β_j) are still allowed to vary by region, however, in this second template structure; intercept (α_j) and slope (β_j), are each modeled as a linear function of a region-level predictor. Instead of centering on a constant vector of means as in Model 1, in Models 2–5 intercept (α_j) and slope (β_j) are modeled as a joint

bivariate normal distribution centered on a different mean vector for each region. These means are predicted from the values of continuous variables $P_{\alpha j}$ and $P_{\beta j}$ in each region. In Model 2, $P_{\alpha j} = P_{\beta j}$ = PRECIP, in Model 3, $P_{\alpha j} = P_{\beta j}$ = TEMP, in Model 4, $P_{\alpha j}$ = TEMP and $P_{\beta j}$ = PRECIP, and in Model 5, $P_{\alpha j} = P_{\beta j}$ = AG (continuous form). The error term σ_y^2 still represents variation within regions. The errors σ_α^2 and σ_β^2 in Models 2–5 represent variation between regions, beyond what is explained by the region-level predictors. The term ρ again accounts for the correlation between α_j and β_j.

AG percentages for the nine regions take on values either <30% or >70%. Therefore, AG can also be represented categorically as either low or high. This categorical AG variable can be incorporated into the model structure as follows:

Template 3. Models 6–8 (two region-level predictors: one continuous variable and one categorical variable):

$$y_{ij} \sim N\left(\alpha_j + \beta_j x_{ij}, \sigma_y^2\right) \tag{12}$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} = \begin{pmatrix} \gamma_{\alpha 0} + \gamma_{\alpha 1} P_{\alpha j} \\ \gamma_{\beta 0} + \gamma_{\beta 1} P_{\beta j} \end{pmatrix} + \begin{pmatrix} \delta_{\alpha j} \\ \delta_{\beta j} \end{pmatrix} + \begin{pmatrix} \delta_{\alpha k} \\ \delta_{\beta k} \end{pmatrix} \tag{13}$$

$$\begin{pmatrix} \delta_{\alpha j} \\ \delta_{\beta j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha j}^2 & \rho_j \sigma_{\alpha j} \sigma_{\beta j} \\ \rho_j \sigma_{\alpha j} \sigma_{\beta j} & \sigma_{\beta j}^2 \end{pmatrix}\right) \tag{14}$$

$$\begin{pmatrix} \delta_{\alpha k} \\ \delta_{\beta k} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha k}^2 & \rho_k \sigma_{\alpha k} \sigma_{\beta k} \\ \rho_k \sigma_{\alpha k} \sigma_{\beta k} & \sigma_{\beta k}^2 \end{pmatrix}\right) \tag{15}$$

where

i=1,……, ~30 basins within each region,

j=1,……, 9 regions,

k=0 for low levels of AG and k=1 for high levels of AG,

y_{ij} is basin-level ecological response of benthic macroinvertebrates (for NMDS1 or RICHTOL),

x_{ij} is basin-level urban land cover percentage (URB),

α_j is region-level intercept (the estimated invertebrate response at URB=0),

β_j is region-level slope (the estimated change in invertebrate response per unit change in URB),

σ_y^2 is within-region variance in invertebrate response,

$P_{\alpha j}$ and $P_{\beta j}$ are region-level conditions (such as PRECIP or TEMP) predicting intercept and slope, respectively,

$\gamma_{\alpha 0}, \gamma_{\alpha 1}, \gamma_{\beta 0}$, and $\gamma_{\beta 1}$ are hyperparameter intercepts and slopes describing the relation between a region-level predictor and α_j and β_j,

$\delta_{\alpha j}$ and $\delta_{\beta j}$ are effects of region on intercept and slope, respectively,

$\delta_{\alpha k}$ and $\delta_{\beta k}$ are effects of AG group on intercept and slope, respectively,

$\sigma_{\alpha j}^2$ is between-region variance of the regional effect on intercept,

$\sigma_{\beta j}^2$ is between-region variance of the regional effect on slope,

ρ_j is correlation of model coefficients $\delta_{\alpha j}$ and $\delta_{\beta j}$,

$\sigma^2_{\alpha k}$ is between-AG group variance of the AG effect on intercept,

$\sigma^2_{\beta k}$ is between-AG group variance of the AG effect on slope, and

ρ_k is correlation of model coefficients $\delta_{\alpha k}$ and $\delta_{\beta k}$.

In this third template, intercept (α_j) and slope (β_j) vary by region and by AG category (low or high). The intercept (α_j) and slope (β_j) are again modeled as a joint bivariate normal distribution centered on a different mean vector for each region-AG group combination. Intercept and slope are predicted from the sum of a linear regression with continuous variables $P_{\alpha j}$ and $P_{\beta j}$, a region effect term and an AG effect term. In Model 6, $P_{\alpha j} = P_{\beta j}$ = PRECIP, in Model 7, $P_{\alpha j} = P_{\beta j}$ = TEMP, and in Model 8, $P_{\alpha j}$ = TEMP and $P_{\beta j}$ = PRECIP. The region-varying components of the intercept and slope, $\delta_{\alpha j}$ and $\delta_{\beta j}$, respectively, are modeled as a bivariate normal distribution with the prior centered on zero, and with $\sigma^2_{\alpha j}$ and $\sigma^2_{\beta j}$ accounting for between-region variance of the regional effect on intercept and slope, respectively. The AG-varying components of the intercept and slope, $\delta_{\alpha k}$ and $\delta_{\beta k}$, respectively, are also modeled as a bivariate normal distribution with the prior centered on zero, and with $\sigma^2_{\alpha k}$ and $\sigma^2_{\beta k}$ accounting for between-AG group variance of the AG effect on intercept and slope, respectively. The error term σ^2_y still represents variation within regions. Finally, the terms ρ_j and ρ_k

account for the correlation between regional effects on slope and intercept and AG effects on slope and intercept, respectively.

In this template, the influence of AG is modeled as a grouping factor. But because there are only two AG groups, it may not be the most appropriate model structure to evaluate using the lmer fitting function (Gelman and Hill 2006). Model structure may be computationally appropriate as well as more intuitively interpretable if the influence of AG in modeled as a numeric predictor instead of as a grouping factor. This alternate model structure could incorporate both continuous and categorical regional predictors in the following manner (representing the effect of AG, in R output terminology, as a "fixed" effect rather than a "random" effect as in the structure above):

$$y_{ij} \sim N\left(\alpha_j + \beta_j x_{ij}, \sigma_y^2\right) \tag{16}$$

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_{\alpha 0} + \gamma_{\alpha 1} P_{\alpha j} + \delta_{\alpha AG} AG \\ \gamma_{\beta 0} + \gamma_{\beta 1} P_{\beta j} + \delta_{\beta AG} AG \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\ \sigma_\alpha \sigma_\beta \\ \rho\ \sigma_\alpha \sigma_\beta & \sigma_\beta^2 \end{pmatrix}\right) \tag{17}$$

where

i=1,……, ~30 basins within each region,

j=1,……, 9 regions,

y_{ij} is basin-level ecological response of benthic macroinvertebrates (for NMDS1 or RICHTOL),

x_{ij} is basin-level urban land cover percentage (URB),

α_j is region-level intercept (the estimated invertebrate response at URB=0),

β_j is region-level slope (the estimated change in invertebrate response per unit change in URB),

σ_y^2 is within-region variance in invertebrate response,

$P_{\alpha j}$ and $P_{\beta j}$ are region-level conditions (such as PRECIP or TEMP) predicting intercept and slope, respectively,

$\gamma_{\alpha 0}, \gamma_{\alpha 1}, \gamma_{\beta 0}$, and $\gamma_{\beta 1}$ are hyperparameter intercepts and slopes describing the relation between a region-level predictor and α_j and β_j,

$\delta_{\alpha AG}$ and $\delta_{\beta AG}$ are effects of AG predictor on hyperparameter intercept and slope, respectively,

AG is 0 for high AG values and 1 for low AG values,

σ_α^2 is between-region-and-AG-group variance in intercepts around their mean, $\gamma_{\alpha 0} + \gamma_{\alpha 1} P_{\alpha j} + \delta_{\alpha AG}$,

σ_β^2 is between-region-and-AG-group variance in slopes around their mean, $\gamma_{\beta 0} + \gamma_{\beta 1} P_{\beta j} + \delta_{\beta AG}$, and

ρ is correlation of model coefficients α_j and β_j.

However, this reparameterization shows no significant difference on model results and conclusions, likely because of the large difference between AG groups, a difference which is explained by the model regardless of specific parameterization. Therefore, the first model Template 3 parameterization is utilized in this Chapter.

These three template structures are modified to accommodate the count response variables of RICH and EPTRICH by fitting a basin-level Poisson-distributed generalized linear model with a log link function in place of the basin-level normally-distributed linear model used for NMDS1 and RICHTOL. The normal distribution for y_{ij} in each of the three templates is replaced by the following equations:

$$y_{ij} \sim Poisson(\lambda_{ij}) \tag{18}$$

$$\log_e(\lambda_{ij}) = \alpha_j + \beta_j x_{ij} \tag{19}$$

where

i=1,……, ~30 basins within each region,

j=1,……, 9 regions,

y_{ij} is basin-level ecological response of benthic macroinvertebrates (for RICH or EPTRICH),

x_{ij} is basin-level urban land cover percentage (URB),

α_j is region-level intercept (the estimated invertebrate response at URB=0),

β_j is region-level slope (the estimated change in invertebrate response per unit change in URB), and

λ_{ij} is the Poisson distribution parameter specific to a basin within a region.

Thus, the ecological count response variable y_{ij} is modeled as Poisson distributed around parameter, λ_{ij} the logarithm of which is a linear function of x_{ij}. This

amended model structure adjusts the skewness of the count data distributions as well as prevents negative predicted values. There is no variance parameter for the Poisson basin-level models because Poisson distribution variance equals the distribution mean and, therefore, changes as the mean changes (as opposed to normal distribution variance which is modeled as a constant σ_y^2 at the basin level for NMDS1 and RICHTOL response variables). The regression coefficient α_j now represents the log of the mean ecological response at URB=0, and the regression coefficient β_j now represents the expected change in the log of the mean ecological response per unit change in the predictor x_{ij}. The region-level tier of the Poisson multilevel hierarchical models remain the same as the region-level tier of the linear multilevel hierarchical models.

These multilevel hierarchical models, thus, provide a systematic structure to fit the individual measurements and to account for variation among the nine regions. Based on the above model frameworks, the multilevel hierarchical models are developed in eight steps by adding environmental predictors at the individual and region levels. All models are compared to a model using no region-level predictors (Model 1). PRECIP and TEMP are used as continuous region-level predictors alone (Models 2 and 3) or in combination (Model 4). Antecedent agriculture (AG) is used as a continuous region-level predictor alone (Model 5) or as a categorical variable in addition to a PRECIP or TEMP predictor to further divide regions into groups of high and low

antecedent agriculture land cover (Models 6, 7, and 8). Interaction between group-level variables is not included in the models.

2.2.3.2. Model Fitting

Linear and Poisson multilevel hierarchical models are fit using the functions lmer (linear mixed-effect model) and glmer (generalized linear mixed-effect model), respectively, from the lme4 package in R statistical program, version 2.7.1, released on June 23, 2008 (R 2008). The lmer and glmer functions use the sparse matrix and Laplace approximation methods to estimate the crossed random effects, which are reflected in the model parameters. Model coefficient estimates are calculated by fitting all coefficients simultaneously to basin and region level data and reporting the maximum likelihood estimate of each coefficient. The lmer and glmer functions are expected to be robust and resilient in dealing with various types of model structure and data because of the advanced algorithms incorporated in these functions.

Hierarchical model output is created and translated into coefficients as described below, using Model 2 as an example. An lmer model output object is created using the following code in R:

```
> M2 <- lmer(ECO ~ URB + PRECIP + URB:PRECIP + (1+URB|REGION)
```

The desired basin-level response variable (ECO) is modeled as a function of the basin-level predictor (URB), the region-level predictor (PRECIP), and the interaction between URB and PRECIP. The final term of the model allows basin-level intercept (as indicated

in the code by '1') and slope (as indicated in the code by 'URB') to change between regions (REGION). This model code will produce output coefficients for each regional regression equation in the following format:

```
ECO = c0 + c1*URB + c2*PRECIP + c3*URB*PRECIP
```
(20)

These coefficients from R output can be rearranged as follows to show the desired basin-level linear regression predicting ECO from URB, including the effect of precipitation:

```
ECO = (c0 + c2*PRECIP) + (c1 + c3*PRECIP)*URB
```
(21)

Then, Model 2 intercept,

α_j = c0 + c2*PRECIP (22)

And Model 2 slope with URB,

β_j = c1 + c3*PRECIP (23)

Intercepts and slopes are calculated this way for each region. This calculation procedure applies to intercept and slope coefficients for Models 2–5. Model 1 has no region-level predictor term and, therefore, coefficients are simply c0 for intercept and c1 for slope.

Models 6–8 add a second categorical predictor. Using Model 6 as an example, an lmer model output object is created using the following code in R:

```
> M6 <- lmer(ECO ~ URB + PRECIP + URB:PRECIP
+(1+URB|REGION)+(1+URB|AG)
```

The desired basin-level response variable (ECO) is modeled as a function of the basin-level predictor (URB), the region-level continuous predictor (PRECIP), and the

interaction between URB and PRECIP. The (1+URB|REGION) term of the model allows basin-level intercept (1) and slope (URB) to change between regions (REG) and the (1+URB|AG) term allows basin-level intercept (1) and slope (URB) to change between the two antecedent agriculture groups. This model code will produce output coefficients for each regional regression equation in the following format:

```
ECO = c0 + c1*URB + c2*PRECIP + c3*URB*PRECIP
```
(24)

where coefficients are changing not only by region but also by AG group.

Rearranging again into:

```
ECO = (c0 + c2*PRECIP) + (c1 + c3*PRECIP)*URB
```
(25)

this time c0 is composed of c0_mean + c0_REG +c0_AG, and c1 is composed of c1_mean + c1_REG +c1_AG. The R syntax calls c0_mean a "fixed effect" and c0_REG and c0_AG "random effects," although the terms fixed and random are relative. Then, Model 6 intercept,

α_j = c0_mean + c0_REG +c0_AG + c2*PRECIP (26)

And Model 2 slope with URB,

β_j = c1_mean + c1_REG +c1_AG + c3*PRECIP (27)

Intercepts and slopes are calculated this way for each region. This calculation procedure applies to intercept and slope coefficients for Models 6–8.

If Models 6–8 are parameterized with the influence of AG as a numeric predictor instead of as a grouping factor, R code would look like the following:

```
> M6.2 <- lmer(ECO ~ URB + PRECIP + AG + URB:PRECIP + URB:AG
+(1+URB|REGION)
```

where the desired basin-level response variable (ECO) is modeled directly as a function of the basin-level predictor (URB), the region-level continuous predictor (PRECIP), the interaction between URB and PRECIP, and the region-level categorical predictor (AG), and the interaction between URB and AG. Here, the (1+URB|REGION) term of the model allows basin-level intercept (1) and slope (URB) to change between regions (REG), but this time the intercept and slope estimates incorporate both PRECIP and AG effects directly. R output now reports coefficients in the format:

```
ECO = c0 + c1*URB + c2*PRECIP + c3*URB*PRECIP +
c4*AG + c5*URB*AG                                    (28)
```

where AG effect coefficients are now reported directly in output as "fixed effects." Rearranging into regional slope-intercept form,

```
ECO = (c0 + c2*PRECIP + c4*AG) + (c1 + c3*PRECIP +c5*AG)*URB   (29)
```

Model 6.2 intercept can then be calculated directly from "fixed effects" by

$$\alpha_j = c0 + c2*PRECIP + c4*AG \quad (30)$$

And Model 6.2 slope with URB,

$$\beta_j = c1 + c3*PRECIP + c5*AG \quad (31)$$

Again, constructing models in R in this way (M6.2) results in nearly identical conclusions to models constructed as M6 code above and, therefore, M6 format analysis is reported for all Models 6–8 in this Chapter.

Model results are summarized graphically by plotting each set of nine α_j estimates and nine β_j estimates against their respective region-level predictor variable values. Average trend lines (Equations 11 and 13) are drawn using $\gamma_{\alpha 0}$ and $\gamma_{\alpha 1}$ coefficients as the region-level intercept and slope, respectively, predicting α_j estimates from region-level predictor (for example, PRECIP), and $\gamma_{\beta 0}$ and $\gamma_{\beta 1}$ coefficients as the region-level intercept and slope, respectively, predicting β_j estimates from region-level predictor. Error bars representing one standard deviation are drawn on each α_j and β_j estimate (see Appendix A for detailed R code).

A Deviance Information Criterion (DIC) is used for evaluating and comparing the eight multilevel hierarchical models for each response variable. This criterion has been developed specifically for comparing hierarchical multilevel model results. It quantitatively evaluates a model by balancing better fitness to the data with less complexity of the model structure. Lower DIC values indicate better model fit. DIC cannot be compared between response variables, as it is a relative and not absolute measure (Spiegelhalter et al. 2002).

2.2.3.3. Model Limitations

This preliminary analysis uses the R command lmer (and its generalized linear corollary, glmer) to calculate multilevel hierarchical model results. To conserve computational power, this command reports only the peaks of posterior distributions of

the coefficients and does not use prior distributions. Even though the updated version of R (R–2.7.1; released on June 23, 2008) improves the performance of hierarchical multilevel regression models by offering more robust and flexible algorithms, several drawbacks were discovered to using the lmer and glmer functions to approximate a fully Bayesian analysis.

For example, the lmer and glmer functions occasionally produce errors due to difficult matrix inversions and then do not converge on parameter estimates. One possible cause of such errors is that certain constraints on estimated variance-covariance matrix—which is positive-definite and symmetric—might be violated. This type of error can possibly be resolved by centering the predictor or by increasing the number of groups in the grouping factor. However, these solutions cannot be applied in the current study because the centering transformation restricts interpretability of the regression plots along the predictor-axis. Also, increasing the number of groups in the categorical agricultural predictor is not an option because AG is distributed only at high and low values. More importantly, regardless of whether the error is resolved or not, the reliability of resultant coefficients as well as variance components cannot be confirmed because the running algorithms of glmer (and lmer) are not accessible at the R user level. Thus, introducing Bayesian inference using WinBugs appears to be necessary in further studies in order to more explicitly comprehend the factors of uncertainty in model parameters. Using the lmer and glmer functions in primary analysis enabled quicker

running of simple introductory models to choose initial responding variables and easily plot results.

In order to work around the problem of convergence, decimal fractions for URB (fraction urban land cover) are used instead of percentages during model fitting, and all other variables are used in their natural scale. The only remaining questionable results involve RICH Models 6 and 7, where, when parameterized in this manner, variance between high and low AG groups is calculated to be zero. However, non-zero variance calculation can be coerced by using URB as a percentage, not a fraction, and standardizing precipitation as PRECIP divided by two standard deviations of PRECIP. Neither method gives overt errors in R, but it is not clear which method produces better estimates. Philosophically, it is unacceptable for model convergence on estimates of intercept, slope, and variance parameters to depend on variable scaling. These different results are likely due to variations in numerical approximation algorithms causing the reporting of different results for different variable scales. This is a potential problem for any estimates not calculated analytically. In the current study, the method using fractional URB to be consistent with the parameterization of the rest of the models is shown. A potential parameterization to try to solve this problem may involve fitting models with AG as a direct categorical regression variable instead of a group divider as it is currently modeled.

Alternately, the convergence problems may be exacerbated by the effect of the small number of groups involved in this analysis (nine for precipitation, temperature, and percentage of antecedent agriculture and two for categorical antecedent agriculture) on the numerical estimation algorithm. For small groups, the lmer function does not estimate variance well and, hence, is unable to converge on posterior peaks of coefficients. An algorithm with more flexible computations is required to ensure accurate calculation of model parameters. In addition to causing computational problems, small group sample size statistically leads to low confidence and high uncertainty in region-level predictors.

Finally, as mentioned above, lmer does not allow the incorporation of prior information into the model specification; therefore, models solved with lmer are not fully Bayesian. Solutions calculated by lmer essentially are identical to those calculated by using a Markov Chain Monte Carlo (MCMC) method using non-informative prior distributions. For the purposes of this Chapter, prior information is not yet available for parameters that are being estimated; however, once the information becomes available, a different analysis method will need to be used in order to incorporate prior information. Relative to fully Bayesian approaches, it is not clear exactly how lmer is calculating model estimates. Distributions and parameters are not defined explicitly nor is it evident how model format accounts for all variables and interrelations. For all these reasons, in subsequent post-preliminary analyses, multilevel hierarchical analysis on the EUSE

dataset should be performed using WInBUGS software, a program designed to implement MCMC simulation analysis to the solution of Bayesian systems and problems.

2.3. Predicting and Understanding Effects of Urbanization

Prior to conducting multilevel hierarchal modeling, land cover data was explored to understand trends associated with agricultural land use as the impacts of this use may confound or compound impacts of urbanization. Subsequently, multilevel hierarchal model results are presented and interpreted.

2.3.1. Preliminary Land Cover Analysis

In this study, the nine regions can be divided into currently agriculture-dominant (DEN, DFW, and MGB) and non-agriculture-dominant (ATL, BIR, BOS, POR, RAL, and STL) groups (Figure 6). These two groups divided by current agricultural practices match those established for previous AG practices (Figure 7). Generally, urbanization appears to have greater negative effects in the non-agriculture-dominant group as evidenced by significant, negative slopes for unpooled regressions of ecological response across URB only for ATL, BIR, BOS, POR, RAL, and STL (Table 7).

Table 7: Summary of unpooled regressions between URB and invertebrate response variables based on 243 equivalent degrees of freedom. For NMDS1 and RICHTOL, simple linear regression was used, and for RICH and EPTRICH, simple Poisson regression was used; * indicates values significant at P<0.05; ** indicates values significant at P<0.001. NMDS1, first axis adjusted nonmetric multidimensional scaling site score; RICH, total taxa richness; EPTRICH, combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders; RICHTOL, richness-weighted mean tolerance of taxa at a basin.

Response	Region	Intercept	Slope	Response	Region	Intercept	Slope
NMDS1	ATL	2.814**	-0.030**	**RICH**	ATL	3.792**	-0.006**
	BIR	2.418**	-0.023**		BIR	3.578**	-0.003*
	BOS	2.557**	-0.036**		BOS	3.938**	-0.015**
	DEN	0.822**	0.005		DEN	3.504**	-0.002
	DFW	2.120**	-0.010*		DFW	3.500**	-0.002
	MGB	1.886**	-0.007*		MGB	3.452**	-0.003*
	POR	2.685**	-0.018**		POR	3.473**	-0.002*
	RAL	2.908**	-0.021**		RAL	3.644**	-0.005**
	SLC	2.495**	-0.018**		SLC	3.712**	-0.006**
EPTRICH	ATL	2.753**	-0.019**	**RICHTOL**	ATL	5.221**	0.018**
	BIR	2.654**	-0.015**		BIR	4.555**	0.021**
	BOS	3.044**	-0.028**		BOS	4.048**	0.031**
	DEN	1.803**	0.000		DEN	6.033**	0.002
	DFW	1.924**	-0.003		DFW	6.937**	-0.001
	MGB	1.992**	-0.006*		MGB	5.360**	0.004
	POR	2.548**	-0.017**		POR	4.367**	0.020**
	RAL	2.489**	-0.016**		RAL	5.144**	0.016**
	SLC	2.834**	-0.018**		SLC	3.931**	0.028**

Despite the known assumption that agricultural activities degrade stream ecosystems, unpooled linear regression slopes did not show a clear pattern of decrease in ecological response with increasing agricultural land cover (Table 8). In fact, NMDS1, RICH, and EPTRICH tend to increase (have positive slope) and RICHTOL tends to decrease (have negative slope) with increasing agricultural land cover in the majority of regions,

although interpreting regression coefficients across agricultural land cover may not be appropriate because of the limited range of agricultural land cover in many of the regions (Figure 6).

Table 8: Summary of regressions between percentage of agriculture land-cover (NLDC7 and NLDC8) and invertebrate response variables based on 243 equivalent degrees of freedom. For NMDS1 and RICHTOL, simple linear regression was used, and for RICH and EPTRICH, simple Poisson regression was used; * indicates values significant at P<0.05; ** indicates values significant at P<0.001. NMDS1, first axis adjusted nonmetric multidimensional scaling site score; RICH, total taxa richness; EPTRICH, combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders; RICHTOL, richness-weighted mean tolerance of taxa at a basin.

Response	Region	Intercept	Slope	Response	Region	Intercept	Slope
NMDS1	ATL	1.020**	0.049**	**RICH**	ATL	3.352**	0.014**
	BIR	0.850**	0.054**		BIR	3.391**	0.006
	BOS	0.444	0.185**		BOS	3.200**	0.060**
	DEN	1.340**	-0.006		DEN	3.354**	0.001
	DFW	1.240**	0.011*		DFW	3.383**	0.001
	MGB	1.402**	0.005		MGB	3.198**	0.002*
	POR	2.309**	-0.010		POR	3.461**	-0.002
	RAL	1.113**	0.056**		RAL	3.205**	0.014**
	SLC	1.359**	0.012		SLC	3.345**	0.004
EPTRICH	ATL	1.672**	0.031**	**RICHTOL**	ATL	6.334**	-0.033*
	BIR	1.708**	0.031**		BIR	5.988**	-0.049**
	BOS	1.759**	0.106**		BOS	5.852**	-0.158**
	DEN	1.795**	0.000		DEN	6.142**	-0.001
	DFW	1.705**	0.002		DFW	6.785**	0.002
	MGB	1.539**	0.004		MGB	5.512**	-0.001
	POR	2.434**	-0.015**		POR	4.715**	0.013
	RAL	1.230**	0.039**		RAL	6.472**	-0.041**
	SLC	1.908**	-0.004		SLC	5.713**	-0.013

In the non-agriculture-dominant group, these results can be attributed to the distribution of the percentage of forest associated with the corresponding percentage of

agriculture (Figure 12); that is, in ATL, BIR, BOS, and RAL, responses are likely to increase with increasing percentage of agriculture (Table 8), not because agricultural

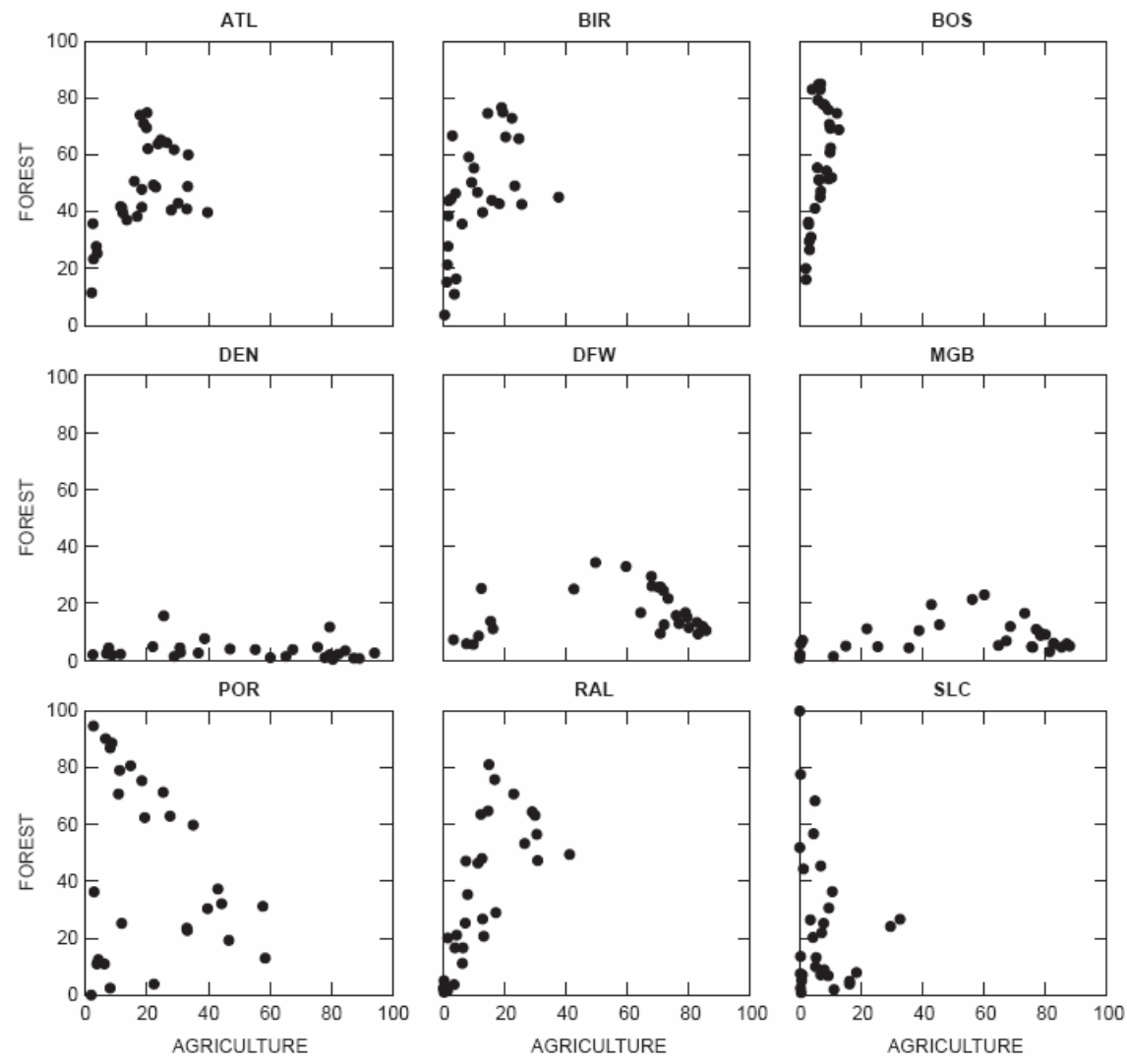

Figure 12: Scatterplots of the percentage of agriculture (NLCD7 and NLCD8) compared to the percentage of forest (NLCD4 and NLCD5) for each region.

practices yield positive effects on invertebrate communities, but more convincingly because percentage of agriculture is positively related to percentage of forest (Figure 12, ATL, BIR, BOS, and RAL graphs), whereas it is negatively associated with urbanization

(Figure 13, ATL, BIR, BOS, and RAL graphs). Conversely, in POR and SLC, invertebrate communities show mostly insignificant responses to increasing percentage of

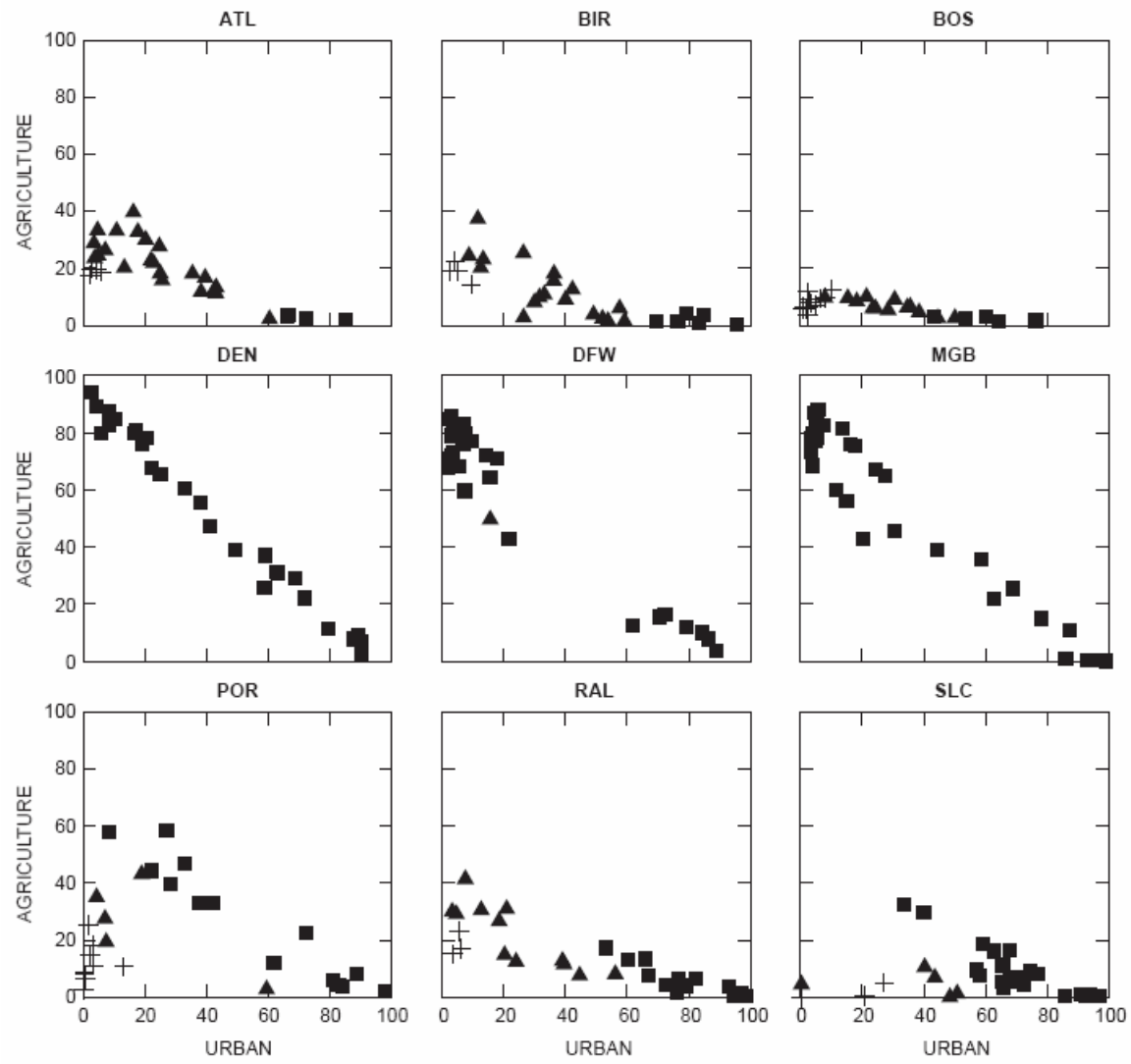

Figure 13: Scatterplots of the percentage of urban (NLCD2) compared to the percentage of agriculture (NLCD7 and NLCD8) for each region. [The filled squares represent the basins where percent forest (NLCD4 and NLCD5) is less than or equal to 33,the triangles represent the basins where percent forest (NLCD4 and NLCD5) is between 33 and 67, and the crosses represent the basins where percent forest (NLCD4 and NLCD5) is greater than or equal to 67]

agriculture (Table 8), because there is more variation in the relations between percentage of agriculture and forest and urban land cover (Figure 12 and Figure 13, POR and SLC graphs).

In contrast to the non-agriculture-dominant group, the agriculture-dominant group generally involves small percentages of forest (Figure 12); thus, agricultural land cover is nearly the inverse of urban land cover (Figure 13). The noticeable common feature in this agriculture-dominant group is that the assemblage metrics already reach low levels (or high levels in the case of RICHTOL) at the low range of both the urban and agricultural gradients (see intercepts in Table 7 and Table 8); accordingly, the rate of change in the responses along the gradients are relatively small or not detectable (see slopes in Table 7 and Table 8). However, in this group, low percentages of agriculture (or high urbanization) correspond consistently to the low levels of response, while high percentages of agriculture (or low urbanization) correspond to more variability in the responses (Figure 14), agreeing with results of previous studies (Wang et al. 2000).

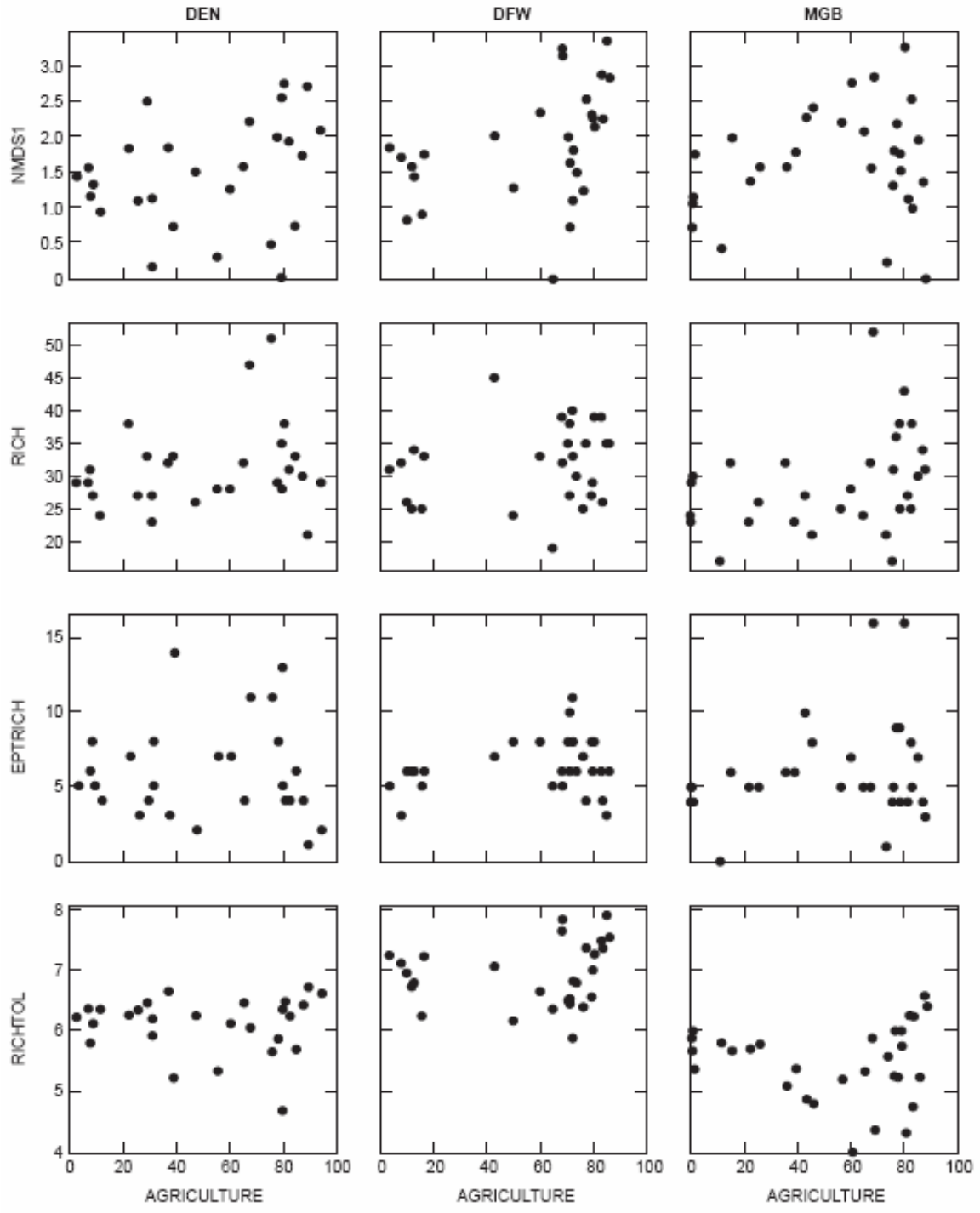

Figure 14: Scatterplots of the percentage of agriculture (NLCD7 and NLCD8) compared to macroinvertebrate response variables (NMDS1, RICH, EPTRICH, and RICHTOL) for agriculture-dominant regions.

2.3.2. Model Summary (Model Results)

Eight different multilevel hierarchical models of the form described in the *Methods* section, with varying region-level intercept and slope predictors, are fit to NMDS1, RICH, EPTRICH, and RICHTOL data from nine different regions. Deviance information criteria (DIC) are calculated (Table 9) as a measure of goodness of fit across the eight models for each response variable. At the basin level, NMDS1 and RICHTOL

Table 9: Deviance information criteria (DIC) for the eight different multilevel hierarchical models per response variable. [NMDS1, first axis adjusted nonmetric multidimensional scaling site score; RICH, total taxa richness; EPTRICH, combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders; RICHTOL, richness-weighted mean tolerance of taxa at a basin. See Table 5 for model definitions]

	Model1	Model2	Model3	Model4	Model5	Model6	Model7	Model8
NMDS1	499.5	472.8	489.9	484.0	462.2	471.4	479.5	474.4
RICH	344.8	344.3	339.9	337.7	339.9	344.3	339.9	336.5
EPTRICH	356.9	351.8	356.8	353.7	334.9	345.7	347.1	345.6
RICHTOL	443.4	425.1	425.3	412.7	402.1	417.8	410.4	404.1

models predict a continuous response from a continuous URB predictor using a linear regression, and RICH and EPTRICH models predict a discrete response from a continuous URB predictor using a generalized linear regression (Poisson data distribution with log link function). These regressions are defined by a set of nine region-level intercepts (α_j) and nine slopes (β_j), as well as a measure of average within-region variance (σ_y^2) for the linear regressions.

The region-level slopes and intercepts themselves belong to a higher tier distribution either centered on a constant pair of means ($\mu_{\alpha.}$ and μ_{β}, Model 1), predicted from a linear regression with a region-level variable (Models 2–5), or predicted from linear regressions with two region-level variables (Models 6–8). Each higher tier regression is defined by a hyperparameter intercept predicting region-level intercepts ($\gamma_{\alpha 0}$), a hyperparameter slope predicting region-level intercepts ($\gamma_{\alpha 1}$), between-region variance in intercept (σ_{α}^2), a hyperparameter intercept predicting region-level slopes ($\gamma_{\beta 0}$), a hyperparameter slope predicting region-level slopes ($\gamma_{\beta 1}$), and between-region variance in slope (σ_{β}^2). For the two region-level variable models, an additional term (δ) is added to hyperparameter intercept estimates to account for high (k=1) or low (k=0) antecedent agriculture. (See *Model Structure* section for model templates with coefficients.) Region-level annual precipitation and temperature are used as continuous variables (Models 2–4 and 6–8), and region-level antecedent agricultural percentage is used as both a continuous (Model 5) and categorical variable (Models 6–8). When fit in R using the lmer or glmer commands, all models converge on peak values of posterior distributions of coefficients, calculated numerically using maximum likelihood estimation. Model coefficients are presented in Tables 9–20. Model graphs are presented in Figures 15–46.

2.3.2.1. Model 1: No Region-Level Predictor

In Model 1, each response variable is modeled hierarchically from basin-level URB without any region-level predictors, allowing the intercept and slope to vary by region around constant means. In order to provide a benchmark for comparison, non-multilevel hierarchical linear regression models are also calculated. Completely pooled models are constructed by fitting a linear regression or generalized linear regression to data from all nine regions simultaneously while unpooled models are constructed by fitting a linear regression or generalized linear regression to each region independently. The results from these simple completely pooled and separate unpooled models are compared with Model l to understand how a hierarchical multilevel model bridges the gap between information obtained from these two modeling extremes.

NMDS1, RICH, and EPTRICH generally decrease with increasing urbanization while RICHTOL increases with increasing urbanization. These trends support the notion that the expansion of urban developed area co-occurred with detrimental effects on macroinvertebrate assemblages. While complete pooling of all nine regions summarize the aggregate pattern of decline with urbanization for each response, unpooled models show that original conditions and response rates differ between regions.

For NMDS1, the negative trend of decreasing NMDS1 values with increasing urban land cover in complete pooling of all nine regions (Figure 15, dashed line) shows

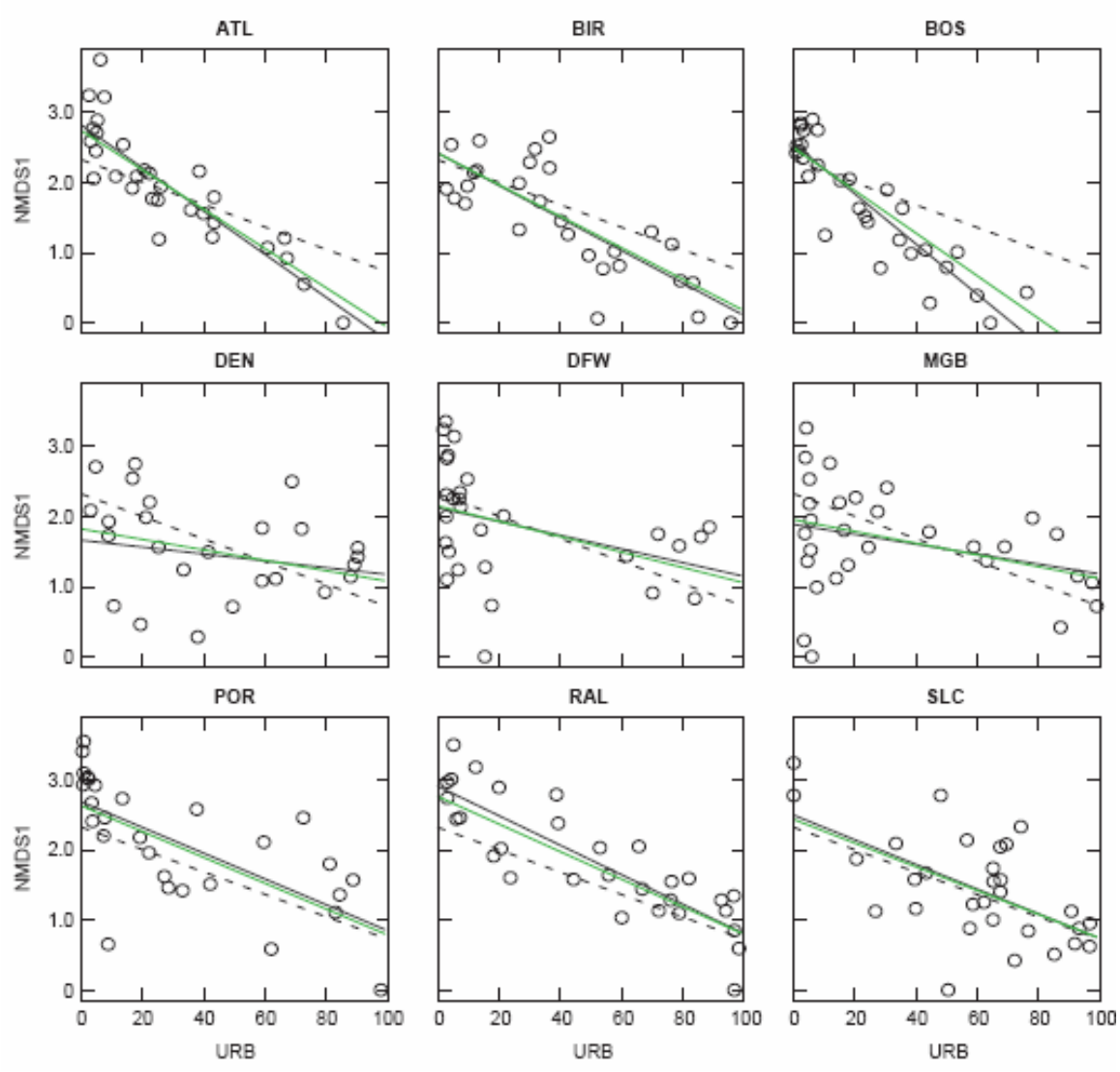

Figure 15: Scatterplots of NMDS1 compared to URB for each region with complete pooling, no pooling, and Model 1: Noregion-level predictors (Equations 8 and 9, Template 1). [Dashed line, complete pooling; black line, no pooling; green line, Model 1 multilevel hierarchical partial pooling. NMDS1, nonmetric multidimensional scaling first axis ordination basin scores; URB, percentage of basin area in developed land]

substantial scatter (Table 10, σ_y^2 = 0.445). When each region is modeled separately

(Figure 15, black line), the unpooled regression models for POR, RAL, and SLC are

similar to the completely pooled model. The unpooled models for ATL, BIR, and BOS have steeper slopes than the completely pooled, and the unpooled slopes of DFW, MGB, and DEN are statistically not different from zero. That is, the urbanization effects in POR, RAL, and SLC are close to the nine-region average, the effects in ATL, BIR, and BOS are above average (NMDS1 responds more to increasing urbanization than

Table 10: Intercept (α) and slope (β) coefficient estimates, representing background condition prior to urbanizationand rate of change with urbanization, respectively, and variance coefficient estimates for invertebrate response NMDS1 Model 1, unpooled model, and completely pooled model. [N/A, not applicable because there are no α or β distributions for the unpooled and completely pooled models. Model 1 has only higher tier constant mean predictors and no group-level predictors, unpooled model addresses each region separately, and completely pooled model combines all regions]

	Model 1 (Partially pooled)		**Unpooled**			**Completely pooled**	
	α_j	β_j	$\alpha_{unpooled}$	$\beta_{unpooled}$	σ_y^2	α	β
ATL	2.73	-2.80	2.81	-3.07	0.159	2.32	-1.60
BIR	2.42	-2.24	2.42	-2.31	0.238		
BOS	2.50	-3.03	2.56	-3.57	0.144		
DEN	1.83	-0.74	1.67	-0.49	0.590		
DFW	2.15	-1.09	2.12	-0.97	0.575		
MGB	1.95	-0.84	1.89	-0.71	0.539		
POR	2.63	-1.84	2.69	-1.84	0.442		
RAL	2.76	-1.97	2.91	-2.10	0.193		
SLC	2.44	-1.69	2.50	-1.76	0.384		
σ_y^2	0.360		Separate values for each *j*; see above			0.445	
μ_α	2.378		N/A			N/A	
μ_β	-1.805		N/A			N/A	
σ_α^2	0.139		N/A			N/A	
σ_β^2	0.794		N/A			N/A	

average), and the effects in DFW, MGB, and DEN are below average (NMDS1 responds less to increasing urbanization than average). Additionally, baseline NMDS1 values at no urbanization (intercepts) for unpooled models in BIR, BOS, DFW, and SLC are similar to completely pooled un-urbanized baseline. However, ATL, POR, and RAL appear to have higher NMDS1 values at no urbanization than overall average, while DEN and MGB have lower values at no urbanization than average.

Negative relations between RICH and percentage of urban land area in all nine regions (Figure 16; Table 11) show between and within region variability, as well. Urbanization effects also vary among the nine regions with different intercepts and slope rates. The unpooled rate of change in total richness with urbanization appears near average (completely pooled) for BIR, RAL, and SLC; below average in DEN, DFW, MGB, and POR; and above average in ATL. BOS, however, stands out as showing a substantially faster degrading of RICH relative to urbanization compared to other regions.

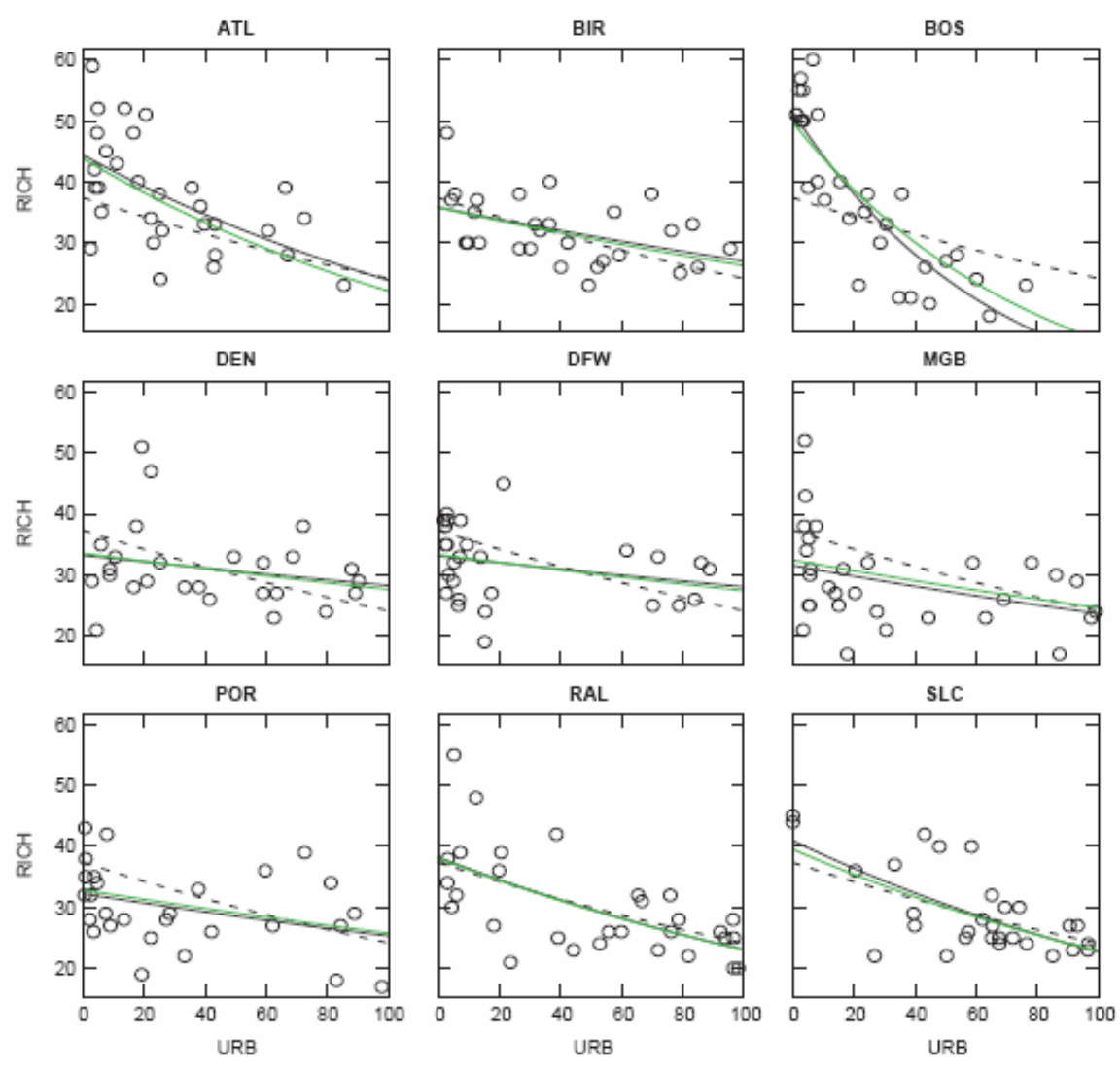

Figure 16: Scatterplots of RICH compared to URB for each region with complete pooling, no pooling, and Model 1: No region-level predictors (Equations 8 and 9, Template 1). [Dashed line, complete pooling; black line, no pooling; green line, Model 1 multilevel hierarchical partial pooling. RICH, total taxa richness; URB, percentage of basin area in developed land]

Table 11: Intercept (α) and slope (β) coefficient estimates, representing background condition prior to urbanization andrate of change with urbanization, respectively, and variance coefficient estimates for invertebrate response RICH Model 1, unpooled model, and completely pooled model. [N/A, not applicable because data variance is equal to the mean parameter (λ) in Poisson models and, therefore, changes with changing λ_j, and there are no α or β distributions for the unpooled and completely pooled models. Model 1 has only higher tier constant mean predictors and no group-level predictors, unpooled model addresses each region separately, and completely pooled model combines all regions]

	Model 1 (Partially pooled)		Unpooled		Completely pooled	
	α_j	β_j	$\alpha_{unpooled}$	$\beta_{unpooled}$	α	β
ATL	3.78	-0.68	3.79	-0.62	3.62	-0.44
BIR	3.58	-0.31	3.58	-0.28		
BOS	3.91	-1.26	3.94	-1.51		
DEN	3.51	-0.19	3.50	-0.16		
DFW	3.51	-0.19	3.50	-0.17		
MGB	3.48	-0.27	3.45	-0.29		
POR	3.49	-0.24	3.47	-0.24		
RAL	3.64	-0.50	3.64	-0.50		
SLC	3.68	-0.55	3.71	-0.59		
σ_y^2	N/A		N/A		N/A	
μ_α	3.618		N/A		N/A	
μ_β	-0.467		N/A		N/A	
σ_α^2	0.022		N/A		N/A	
σ_β^2	0.116		N/A		N/A	

Declining EPTRICH response to urbanization appears to have slightly less variability than RICH (Figure 17; Table 12) but, again, no-pooling models reveal the varying response of EPTRICH among the regions to increasing urbanization. While the intercepts and slopes of BIR, POR, and RAL approximately overlay the overall averages, ATL, BOS, and SLC show higher intercepts and steeper slopes than overall averages.

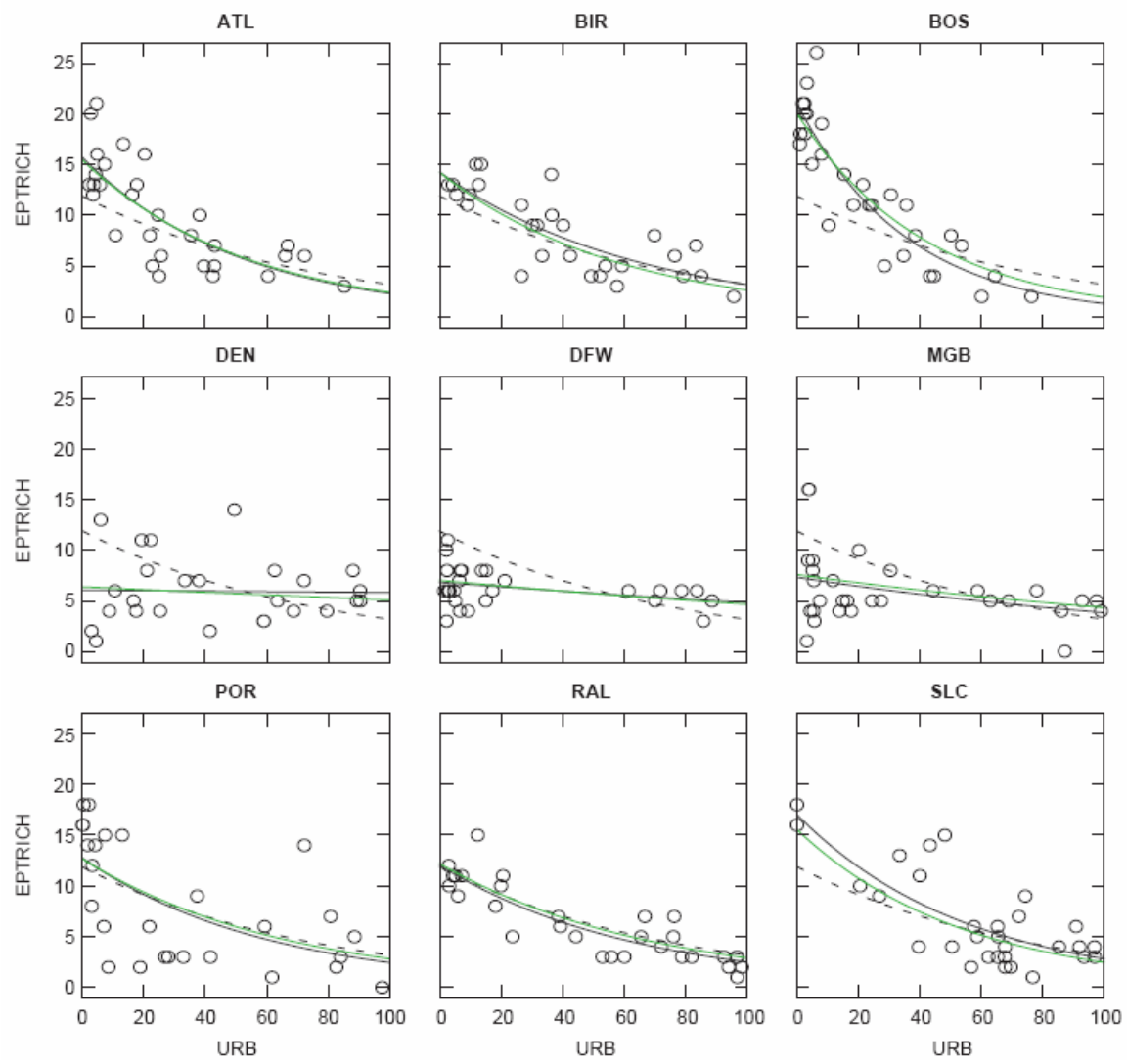

Figure 17: Scatterplots of EPTRICH compared to URB for each region with complete pooling, no pooling, and Model 1: No region-level predictors (Equations 8 and 9, Template 1). [Dashed line, complete pooling; black line, no pooling; green line, Model 1 multilevel hierarchical partial pooling. EPTRICH, combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders; URB, percentage of basin area in developed land]

Conversely, DEN, DFW, and MGB show lower intercepts and less steep slopes than overall averages, meaning that EPTRICH in these regions might be already reduced

prior to urbanization, and thus, the metric in these regions does not sensitively respond to increasing urbanization as in the other regions.

Table 12: Intercept (α) and slope (β) coefficient estimates, representing background condition prior to urbanization and rate of change with urbanization, respectively, and variance coefficient estimates for invertebrate response EPTRICH Model 1, unpooled model, and completely pooled model. [N/A, not applicable because data variance is equal to the mean parameter (λ) in Poisson models and, therefore, changes with changing λ_j, and there are no α or β distributions for the unpooled and completely pooled models. Model 1 has only higher tier constant mean predictors and no group-level predictors, unpooled model addresses each region separately, and completely pooled model combines all regions]

	Model 1 (Partially pooled)		**Unpooled**		**Completely pooled**	
	α_j	β_j	$\alpha_{unpooled}$	$\beta_{unpooled}$	α	β
ATL	2.74	-1.86	2.75	-1.92	2.48	-1.32
BIR	2.65	-1.69	2.65	-1.50		
BOS	2.99	-2.35	3.04	-2.78		
DEN	1.86	-0.22	1.80	-0.04		
DFW	1.95	-0.41	1.92	-0.34		
MGB	2.03	-0.56	1.99	-0.65		
POR	2.54	-1.50	2.55	-1.65		
RAL	2.50	-1.44	2.49	-1.57		
SLC	2.74	-1.84	2.80	-1.80		
σ_y^2	N/A		N/A		N/A	
μ_α	2.442		N/A		N/A	
μ_β	-1.313		N/A		N/A	
σ_α^2	0.154		N/A		N/A	
σ_β^2	0.524		N/A		N/A	

Finally, the complete pooling of RICHTOL across all nine regions (Figure 18, dashed line) shows a pattern opposite the other three metrics. While total richness, EPT richness, and ordination scores scaled to EPT richness decrease with urbanization,

richness-weighted tolerance generally increases with increasing urban land cover. Overall, this pattern shows considerable scatter (Table 13, $\sigma_y^2 = 0.704$), greater than that for NMDS1. Modeled individually, POR, MGB, and SLC have the greatest within-region variability in RICHTOL response to urban land cover (Table 13, $\sigma_y^2 = 0.555, 0.389, 0.354$, respectively). The unpooled regressions (Figure 18, black line) match the slope of the completely pooled model fairly well in BIR, ATL, RAL, and POR but exhibit steeper slopes than the slope of the completely pooled model for BOS and SLC and flatter slopes than those calculated in the completely pooled model for MGB, DFW, and DEN. This means that the RICHTOL in BOS and SLC responds more drastically to increasing urbanization than the overall average, while in MGB, DFW, and DEN the rate at which RICHTOL increases is less affected by increases in urbanization as compared to the overall average.

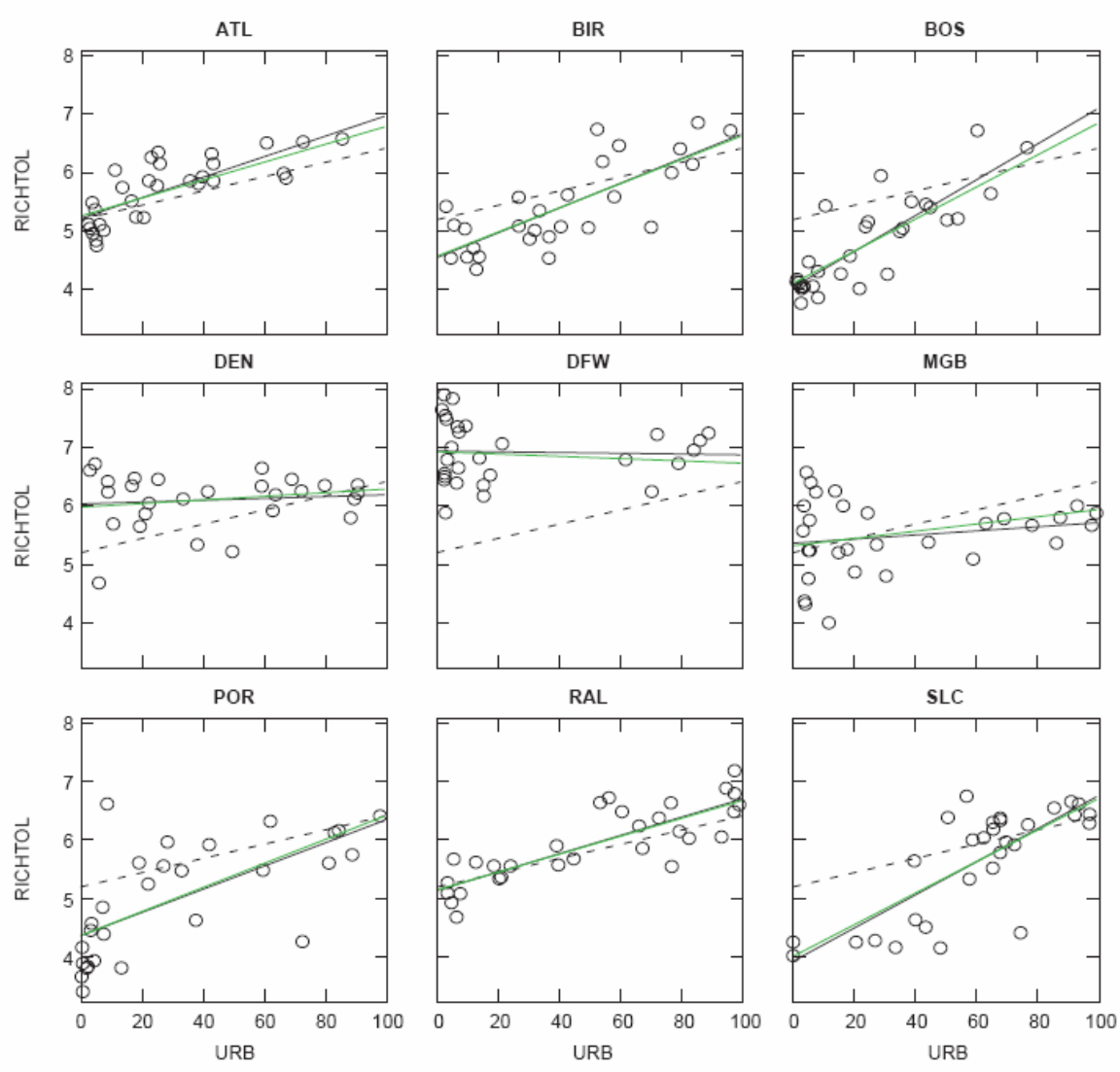

Figure 18: Scatterplots of RICHTOL compared to URB for each region with complete pooling, no pooling, and Model 1: No region-level predictors (Equations 8 and 9, Template 1). [Dashed line, complete pooling; black line, no pooling; green line, Model 1 multilevel hierarchical partial pooling. RICHTOL, richness-weighted mean tolerance of taxa at a basin; URB, percentage of basin area in developed land]

Table 13: Intercept (α) and slope (β) coefficient estimates, representing background condition prior to urbanization and rate of change with urbanization, respectively, and variance coefficient estimates for invertebrate response RICHTOL Model 1, unpooled model, and completely pooled model. [N/A, not applicable because there are no α or β distributions for the unpooled and completely pooled models. Model 1 has only higher tier constant mean predictors nd no group-level predictors, unpooled model addresses each region separately, and completely pooled model combines all regions]

	Model 1 (Partially pooled)		Unpooled			Completely pooled	
	α_j	β_j	$\alpha_{unpooled}$	$\beta_{unpooled}$	σ_y^2	α	β
ATL	5.26	1.60	5.22	1.75	0.122	5.20	1.22
BIR	4.58	2.13	4.55	2.11	0.231		
BOS	4.12	2.75	4.05	3.05	0.187		
DEN	5.98	0.21	6.03	0.16	0.226		
DFW	6.92	-0.16	6.94	-0.06	0.282		
MGB	5.32	0.59	5.36	0.35	0.389		
POR	4.37	2.12	4.37	1.98	0.555		
RAL	5.15	1.55	5.14	1.57	0.131		
SLC	4.01	2.60	3.93	2.83	0.354		
σ_y^2	0.275		Separate values for each *j*; see above			0.839	
μ_α	5.079		N/A			N/A	
μ_β	1.482		N/A			N/A	
σ_α^2	0.916		N/A			N/A	
σ_β^2	1.169		N/A			N/A	

Multilevel hierarchical Model 1 combines data from an individual region with the overall trend from the nine regions to calculate partially pooled estimates of intercept (α_j) and slope (β_j) for each region, as well as estimates of within-region variance in invertebrate response (σ_y^2), between-region variance in intercept (σ_α^2), and

between-region variance in slope (σ_{β}^{2}) for each of the four ecological response models. As in the simple linear regressions, the multilevel hierarchical model intercept represents the estimated value of the ecological metric at zero urbanization, and the slope represents the rate of change of the ecological metric with urbanization. However, the difference is that multilevel hierarchical model estimates are calculated taking into account both the regional data and the overall data across all regions. These partially pooled α_j and β_j values are, therefore, shrunk toward the overall mean. For example, the estimated Model 1 NMDS1 α_j value for ATL (2.73, Table 10) is a partial pooling of intercept information from the unpooled NMDS1 ATL model (2.81, Table 10) and intercept information from the overall completely pooled NMDS1 model (2.32, Table 10).

By assigning multilevel hierarchical structure to the data, individual region estimates borrow strength from the entire dataset and decrease the biasing influence of group sample size. In this analysis, all of the nine regions have approximately equal sample sizes (28–30). Also, between-region variance is generally greater than within-region variance, meaning that the pattern of ecological response to urbanization varies substantially from region to region for each response variable. Therefore, unpooled regressions have greater weight than the completely pooled regression on partially pooled estimates. This is why, although the partially pooled multilevel hierarchical models (Figure 15-Figure 18, green lines, for NMDS1, RICH, EPTRICH, and RICHTOL, respectively) technically lie between their respective completely pooled and separate

unpooled regressions for each region (Figure 15-Figure 18, dashed and black lines, for NMDS1, RICH, EPTRICH, and RICHTOL, respectively), they do not appear to differ significantly from results obtained by separate unpooled regressions. This trend holds for Models 2–8 and, therefore, the visual model results at the basin level are not depicted for the remaining models. Since Model 1 imposes a higher tier distribution structure on the data by assuming that slopes and intercepts for the nine regions are drawn from the same distribution, residual variance decreases for the linear multilevel hierarchical models (σ_y^2 = 0.360 and 0.275, for NMDS1 and RICHTOL, Table 10 and Table 13, respectively) relative to the completely pooled models (σ_y^2 = 0.445 and 0.704, for NMDS1 and RICHTOL, Table 10 and Table 13, respectively) which has no higher tier structure requirement. Poisson model variance is not held constant by definition and, therefore, the same comparison cannot be made for RICH and EPTRICH.

Because it does not incorporate region-level data, Model 1 assumes that slopes and intercepts for regions are exchangeable and are drawn from the same distribution centered on a constant, μ. When region-level variables are available, however, differences in slope and intercept between regions are no longer random. These differences can now be explained by differences in region-level variable values, as shown below. In this way, multilevel hierarchical analysis incorporates data measured at different scales into the same model and enables multilevel hierarchical models to harness greater interpretive power than simple single-tier regression.

2.3.2.2. Model 2: PRECIP Region-Level Predictor

Unlike Model 1, Models 2–8 assume that differences between regions in response to urbanization are not random but mediated by physical environmental factors. Model 2 uses the differences in annual precipitation between the nine regions to account for differences in regional intercept and slope. Partially pooled estimates of intercept (α_j) and slope (β_j) for each region, within-region variance in invertebrate response (σ_y^2), between-region variance in intercept (σ_α^2), and between-region variance in slope (σ_β^2) are again calculated for each of the four ecological response models. In addition, Model 2 estimates region-level intercepts and slopes ($\gamma_{\alpha 0}, \gamma_{\alpha 1}, \gamma_{\beta 0}$, and $\gamma_{\beta 1}$) which describe the relation between the regional precipitation predictor and α_j and β_j. This means that in Model 2, σ_α^2 and σ_β^2 now describe variation of estimated intercepts and slopes around a higher tier regression line, instead of variation around a constant mean as in Model 1. Incorporating a regional, explanatory variable into intercept and slope estimation results in better model fit for all four response variables, as measured by lower DIC for Model 2 (and all subsequent Models) relative to Model 1 (Table 9).

When precipitation is included as a region-level predictor, NMDS1 (Figure 19; Table 14), RICH (Figure 20; Table 15), and EPTRICH (Figure 21; Table 16) intercepts increase with increasing precipitation and their slopes become more negative with increasing precipitation. Conversely, RICHTOL intercepts decrease with increasing

precipitation and slopes become more steeply positive with increasing precipitation (Figure 22; Table 17). This finding means that precipitation affects both baseline pre-urbanization conditions and rate of change in ecological condition with urbanization.

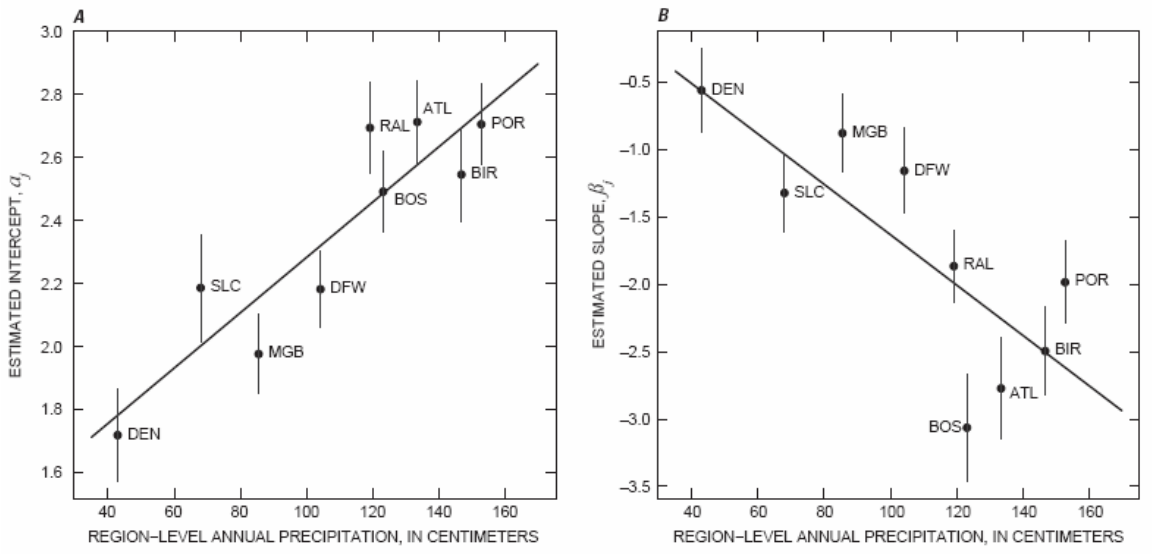

Figure 19: NMDS1 multilevel hierarchical Model 2: Region-level precipitation predictor for (A) intercept and (B) slope. Within each region, NMDS1 (first axis adjusted nonmetric multidimensional scaling site score) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

Table 14: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response NMDS1 Models 2–5. [See Table 5 for model definitions]

	Model 2		Model 3		Model 4		Model 5	
	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
ATL	2.71	-2.77	2.76	-2.86	2.73	-2.77	2.60	-2.44
BIR	2.55	-2.49	2.44	-2.30	2.43	-2.29	2.59	-2.60
BOS	2.49	-3.06	2.49	-3.09	2.44	-3.02	2.50	-3.07
DEN	1.72	-0.56	1.80	-0.70	1.80	-0.69	1.83	-0.75
DFW	2.18	-1.16	2.18	-1.10	2.19	-1.11	1.98	-0.68
MGB	1.98	-0.88	1.92	-0.80	1.94	-0.85	1.96	-0.85
POR	2.71	-1.98	2.61	-1.81	2.64	-1.91	2.75	-2.08
RAL	2.69	-1.86	2.78	-1.98	2.81	-2.01	2.69	-1.83
SLC	2.19	-1.32	2.39	-1.632	2.42	-1.65	2.78	-2.24
σ_y^2	0.360		0.359		0.359		0.359	
$\gamma_{\alpha 0}$	1.407		1.901		1.948		2.818	
$\gamma_{\alpha 1}$	0.009		0.0383		0.035		-0.011	
$\gamma_{\beta 0}$	0.240		-1.199		-0.836		-2.784	
$\gamma_{\beta 1}$	-0.019		-0.049		-0.009		0.024	
σ_α^2	0.046		0.139		0.134		0.016	
σ_β^2	0.436		0.922		0.551		0.108	

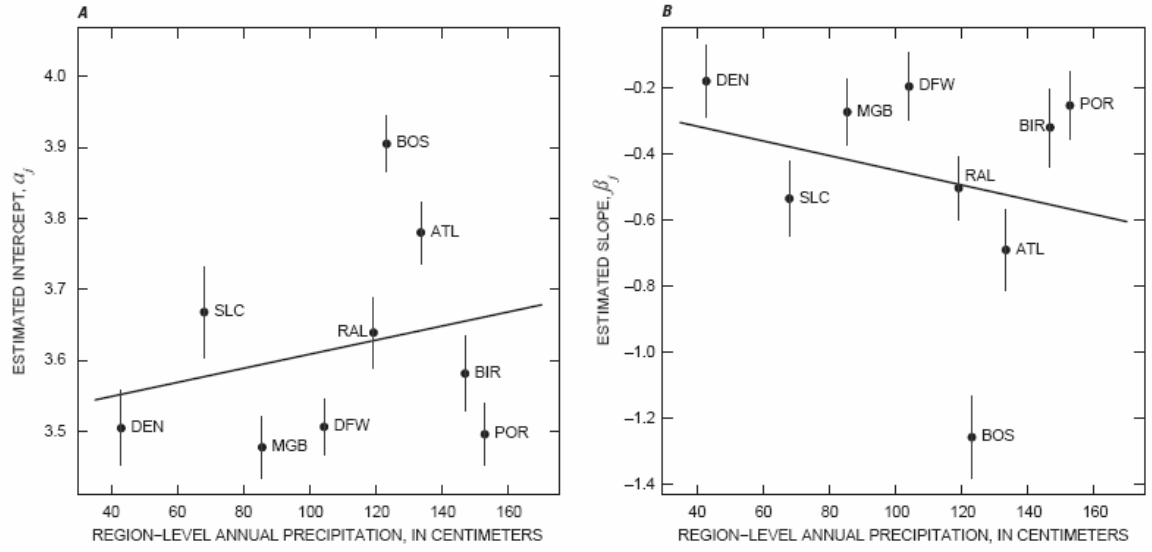

Figure 20: RICH multilevel hierarchical Model 2: Region-level precipitation predictor for (A) intercept and (B) slope. Within each region, RICH (total taxa richness) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

Table 15: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response RICH Models 2–5. [N/A, not applicable because data variance is equal to the mean parameter (λ) in Poisson models and, therefore, changes with changing λ_j. See Table 5 for model definitions]

	Model 2		Model 3		Model 4		Model 5	
	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
ATL	3.78	-0.69	3.78	-0.68	3.79	-0.75	3.78	-0.68
BIR	3.58	-0.32	3.58	-0.29	3.58	-0.34	3.60	-0.34
BOS	3.91	-1.26	3.90	-1.25	3.90	-1.19	3.90	-1.25
DEN	3.50	-0.18	3.49	-0.21	3.50	-0.19	3.49	-0.15
DFW	3.51	-0.19	3.52	-0.13	3.53	-0.09	3.49	-0.18
MGB	3.48	-0.27	3.48	-0.29	3.47	-0.26	3.47	-0.26
POR	3.50	-0.25	3.49	-0.24	3.48	-0.29	3.51	-0.27
RAL	3.64	-0.50	3.65	-0.50	3.66	-0.50	3.65	-0.51
SLC	3.67	-0.53	3.64	-0.52	3.64	-0.53	3.70	-0.58
σ_y^2	N/A		N/A		N/A		N/A	
$\gamma_{\alpha 0}$	3.510		3.602		3.442		3.742	
$\gamma_{\alpha 1}$	0.001		0.001		0.014		-0.003	
$\gamma_{\beta 0}$	-0.226		-0.723		-0.273		-0.718	
$\gamma_{\beta 1}$	-0.002		0.021		-0.002		0.006	
σ_α^2	0.020		0.021		0.023		0.012	
σ_β^2	0.111		0.111		0.104		0.075	

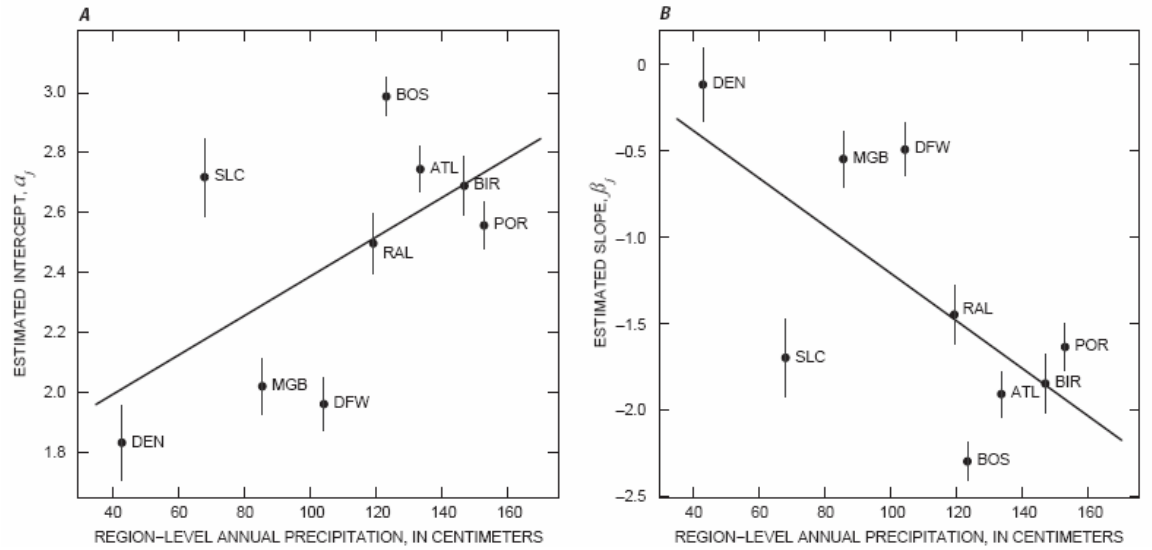

Figure 21: EPTRICH multilevel hierarchical Model 2: Region-level precipitation predictor for (A) intercept and (B) slope. Within each region, EPTRICH (combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

Table 16: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response EPTRICH Models 2–5. [N/A, not applicable because data variance is equal to the mean parameter (λ) in Poisson models and, therefore, changes with changing λ_j. See Table 5 for model definitions]

	Model 2		Model 3		Model 4		Model 5	
	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
ATL	2.75	-1.91	2.74	-1.84	2.74	-1.83	2.73	-1.84
BIR	2.69	-1.85	2.65	-1.67	2.68	-1.80	2.70	-1.75
BOS	2.99	-2.30	3.00	-2.37	3.00	-2.40	2.94	-2.28
DEN	1.83	-0.12	1.85	-0.23	1.86	-0.17	1.80	-0.15
DFW	1.96	-0.50	1.95	-0.38	1.95	-0.37	1.91	-0.36
MGB	2.02	-0.55	2.03	-0.58	2.01	-0.64	1.97	-0.48
POR	2.56	-1.64	2.54	-1.51	2.55	-1.73	2.63	-1.60
RAL	2.50	-1.45	2.50	-1.43	2.49	-1.40	2.59	-1.55
SLC	2.72	-1.70	2.74	-1.84	2.78	-1.79	2.75	-1.86
σ_y^2	N/A		N/A		N/A		N/A	
$\gamma_{\alpha 0}$	1.729		2.444		2.269		2.926	
$\gamma_{\alpha 1}$	0.007		0.000		0.014		-0.012	
$\gamma_{\beta 0}$	0.165		-1.373		-0.886		-2.183	
$\gamma_{\beta 1}$	-0.014		0.005		-0.004		0.023	
σ_α^2	0.105		0.154		0.162		0.008	
σ_β^2	0.307		0.530		0.442		0.041	

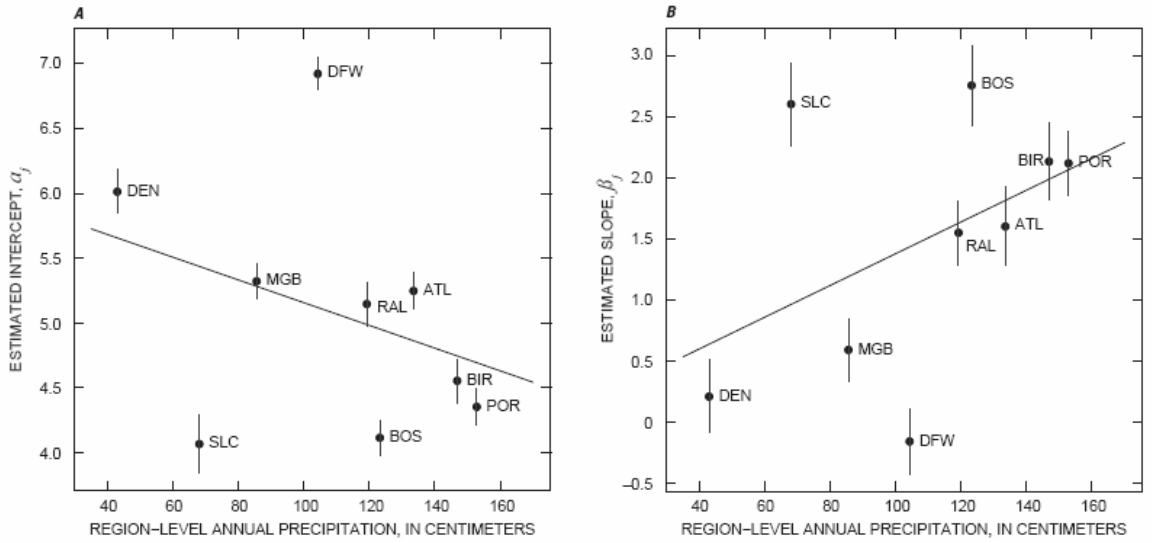

Figure 22: RICHTOL multilevel hierarchical Model 2: Region-level precipitation predictor for (A) intercept and (B) slope. Within each region, RICHTOL (richness-weighted mean tolerance of taxa at a basin) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

Table 17: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response RICHTOL Models 2–5. [See Table 5 for model definitions]

	Model 2		Model 3		Model 4		Model 5	
	α_j	β_j	α_j	β_j	α_j	β_j	α_j	β_j
ATL	5.25	1.60	5.25	1.63	5.26	1.55	5.17	1.87
BIR	4.55	2.13	4.53	2.32	4.52	2.41	4.53	2.19
BOS	4.11	2.75	4.16	2.41	4.21	2.05	4.19	2.43
DEN	6.01	0.21	6.01	0.14	5.99	0.21	6.10	0.00
DFW	6.92	-0.16	6.93	-0.19	6.92	-0.22	6.94	-0.22
MGB	5.32	0.59	5.32	0.62	5.32	0.67	5.35	0.50
POR	4.35	2.12	4.36	2.12	4.38	1.93	4.31	2.25
RAL	5.15	1.55	5.15	1.58	5.15	1.60	5.04	1.78
SLC	4.06	2.60	4.03	2.62	4.00	2.71	4.19	2.39
σ_y^2	0.274		0.275		0.278		0.272	
$\gamma_{\alpha 0}$	6.032		3.786		3.937		4.171	
$\gamma_{\alpha 1}$	-0.009		0.104		0.092		0.024	
$\gamma_{\beta 0}$	0.078		1.989		1.962		2.656	
$\gamma_{\beta 1}$	0.013		-0.042		-0.005		-0.031	
σ_α^2	0.924		0.850		0.745		0.275	
σ_β^2	1.088		1.263		1.245		0.049	

Because NMDS1 scores are scaled such that they decrease from a maximum value at minimum urban intensity to zero at maximum urban intensity, the intercept provides a measure of the approximate extent to which the assemblages at minimum and maximum urbanization (URB) differ. Regions with little precipitation have lower intercepts than do regions with higher precipitation (Figure 19*A*). Consequently, the amount of change that can occur over the urbanization gradient is

lower in regions with lower precipitation. This finding is also reflected in the slopes (Figure 19*B*), which are higher in regions with higher precipitation and lower in regions with lower precipitation. That is, given a minimum possible low NMDS1 value of zero, higher NMDS1 starting values have more parameter space in which to fall. This correlation between intercept and slope is addressed with the ρ model coefficient term.

RICH and EPTRICH metrics are absolute, not scaled, measures of total and EPT taxa richness. Higher precipitation is also associated with greater total richness and greater EPT richness at zero urbanization, although both relations look considerably more scattered than NMDS1 Model 2. This greater initial richness results in more negative rates of change of richness with urbanization at high precipitation, and lower initial richness at low precipitation results in the opposite (less negative rates of change of richness with urbanization). RICHTOL metrics are richness weighted and defined on a scale of 0 to 10. However, RICHTOL values in the EUSE dataset range from only 4 to 7, showing that typical macroinvertebrate communities were neither completely tolerant nor completely intolerant. Following expected response direction of the previous three variables, higher precipitation was associated with lower richness-weighted tolerance and steeper increase in tolerance with urbanization. Similar to NMDS1, RICHTOL intercepts and slopes also show inverse correlation, lending ecological credence to observed patterns.

Therefore, in addition to response shapes derived from statistical consequences, Model 2 is scientifically interpretable. In general, macroinvertebrate condition metrics appear to be worse in dry regions with little precipitation and better in wet regions with greater precipitation prior to urbanization. This can be interpreted to mean that wet regions are likely associated with greater indicators of healthier stream macroinvertebrate communities than dry regions in the absence of urban land cover. Precipitation also appears to speed up the decline of macroinvertebrate communities with urbanization.

2.3.2.3. Model 3: TEMP Region-Level Predictor

Model 3 replaces annual precipitation with annual temperature as a region-level intercept and slope predictor. Model format and coefficient interpretations are identical to Model 2, exchanging TEMP for PRECIP. In contrast to models including precipitation, only NMDS1 and RICHTOL appear to respond to temperature, while RICH and EPTRICH, for the most part, do not.

For NMDS1, Model 3 goodness of model fit evaluated visually and quantitatively (DIC = 489.9, Table 9) is less than the fit of Model 2 (DIC = 472.8, Table 9). Nonetheless, there is still a fairly strong increase in intercept with increasing temperature (Figure 23*A*; Table 14) and increase in negative slope steepness with increasing temperature (Figure 23*B*; Table 14). Decline in NMDS1 with urbanization (in all regions except DEN) appears to happen more steeply relative to urbanization in

regions with warmer temperatures. The negative relation of slope with temperature is not very strong, and regions such as BOS, DEN, and DFW do not follow this pattern closely.

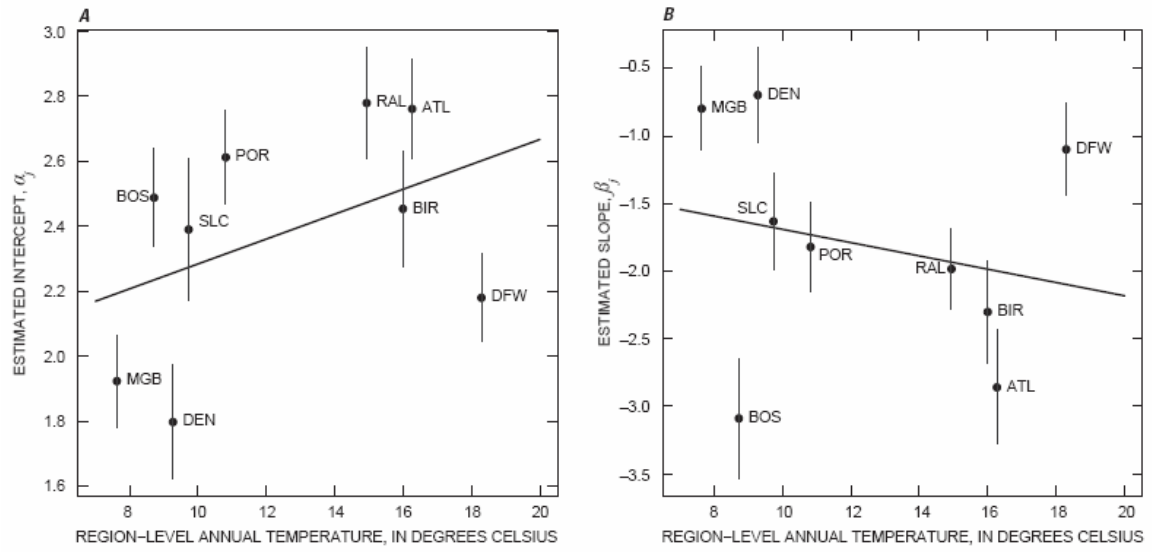

Figure 23: NMDS1 multilevel hierarchical Model 3: Region-level temperature predictor for (A) intercept and (B) slope. Within each region, NMDS1 (first axis adjusted nonmetric multidimensional scaling site score) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature as shown in Equation 11 (Template 2, region level).

Temperature does not appear to have any effect on RICH prior to urbanization (Figure 24*A*; Table 15). Total richness prior to urbanization appears to vary randomly across both colder regions (such as SLC and MGB) and warmer regions (such as ATL and DFW). Somewhat of a pattern with temperature is observed under urbanization conditions. There appears to be less decline in total richness per unit change of urbanization with warmer temperature (Figure 24*B*), which implies that warmer climate

is more optimal for high total macroinvertebrate richness. However, this response to temperature is driven largely by a BOS outlier, which has an estimated slope substantially more negative than the remaining regions and also the second coldest annual temperature. Without the BOS point, it appears that slope estimates vary randomly with temperature, as well. Therefore, it is surprising that DIC assigns better model fit to Model 3 (DIC = 339.9, Table 9) than to Model 2 (DIC = 344.3, Table 9) for RICH data. It is not clear why this occurs.

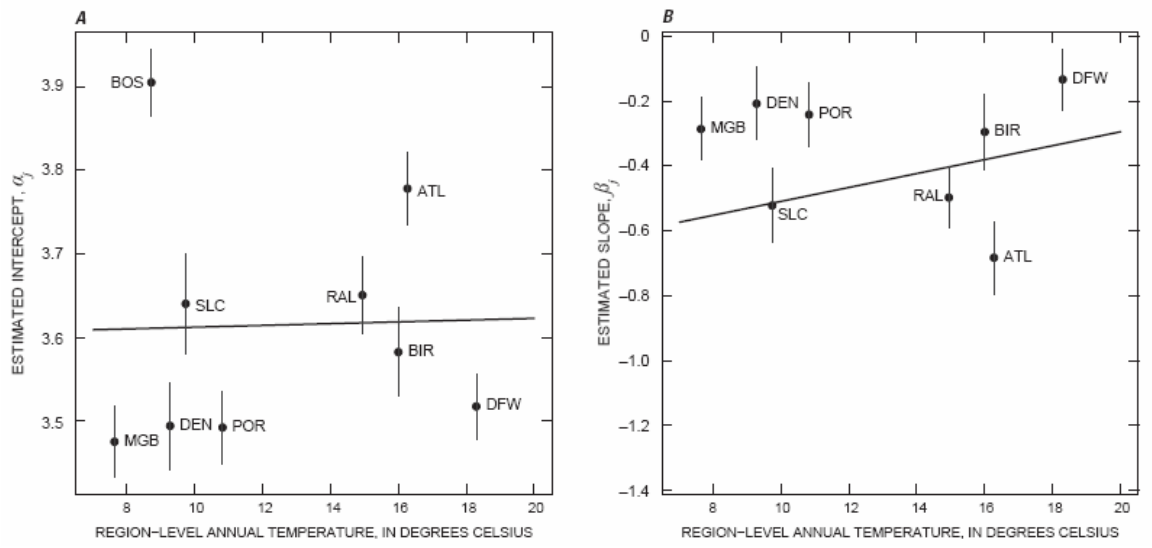

Figure 24: RICH multilevel hierarchical Model 3: Region-level temperature predictor for (A) intercept and (B) slope. Within each region, RICH (total taxa richness) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature as shown in Equation 11 (Template 2, region level).

For EPTRICH, adding a region-level predictor did not significantly reduce the DIC (356.8, Table 9), nor significantly improve the accuracy in estimating coefficients at

region level ($\sigma_\alpha^2 = 0.154$ and $\sigma_\beta^2 = 0.524$ for Model 1 [Table 12] relative to $\sigma_\alpha^2 = 0.105$ and $\sigma_\beta^2 = 0.307$ for Model 2 and $\sigma_\alpha^2 = 0.154$ and $\sigma_\beta^2 = 0.530$ for Model 3 [Table 16]). In fact, adding a region-level predictor actually slightly increased the unexplained region-level variance for Model 3 slope. One potential reason of this result could be the similar sample sizes (27 to 30) among regions, which minimizes the level of pooling effect toward the overall mean, thus also minimizing the degree of modifying σ_α^2 and σ_β^2, which is usually done through partial pooling. Another possibility could be that σ_α^2 and σ_β^2 increase because adding a region-level predictor actually unveils the hidden variation among regions by explaining correlation between the basin-level variable (URB) and region-level errors (σ_α^2 and σ_β^2) (Gelman and Hill 2006). Nonetheless, a region-level predictor still improves the model with regard to interpretability of region-level variations. Estimated intercepts and slopes do not show any significant tendency (of increasing or decreasing) along with changes in annual mean temperature (Figure 25). This result indicates that influential factors other than temperature likely determine the response of EPTRICH at the region-level. One of the potential factors might be the level of agricultural activities at each region because the regions that deviated from the others, by having lower intercepts and higher slopes, are the agriculture-dominant group (DEN, MGB, DFW; see *Preliminary Land Cover Analysis* section).

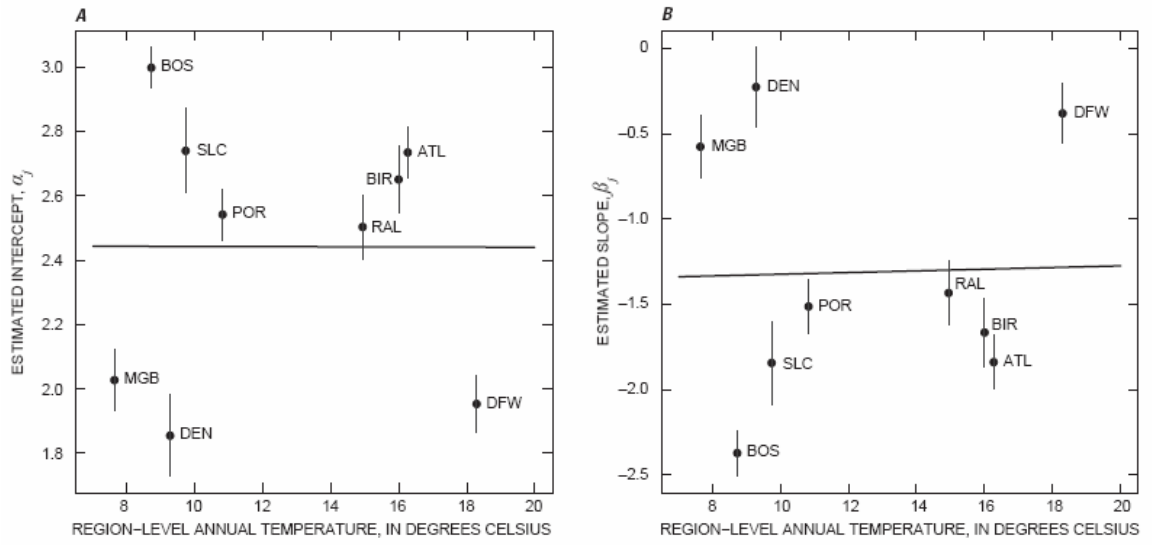

Figure 25: EPTRICH multilevel hierarchical Model 3: Region-level temperature predictor for (A) intercept and (B) slope. Within each region, EPTRICH (combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature as shown in Equation 11 (Template 2, region level).

For RICHTOL, Model 3 seems to fit (DIC = 425.3, Table 9) the data as well as Model 2 (DIC = 425.1, Table 9). Nonetheless, in Model 3 the intercept increases with increasing temperature (Figure 26*A*; Table 17), while the positive slope on urban development tends to decrease with increasing temperatures (Figure 26*B*; Table 17). Prior to urbanization, it appears that there are more tolerant species in warmer regions. And the positive relation between RICHTOL and urban development dampens slightly as regional mean annual ambient temperature increases. This could mean that warmer regions initially start with more hardy species and as such they have less opportunity to gain even more hardy species with increasing urbanization. This dampening of the slope

with increased temperature does not seem to be strong for regions such as MGB and DEN, which have flat slopes despite low mean temperatures. Again, it appears that high-agriculture regions (DEN, DFW, MGB) form a response group pattern visually separate from the remaining regions.

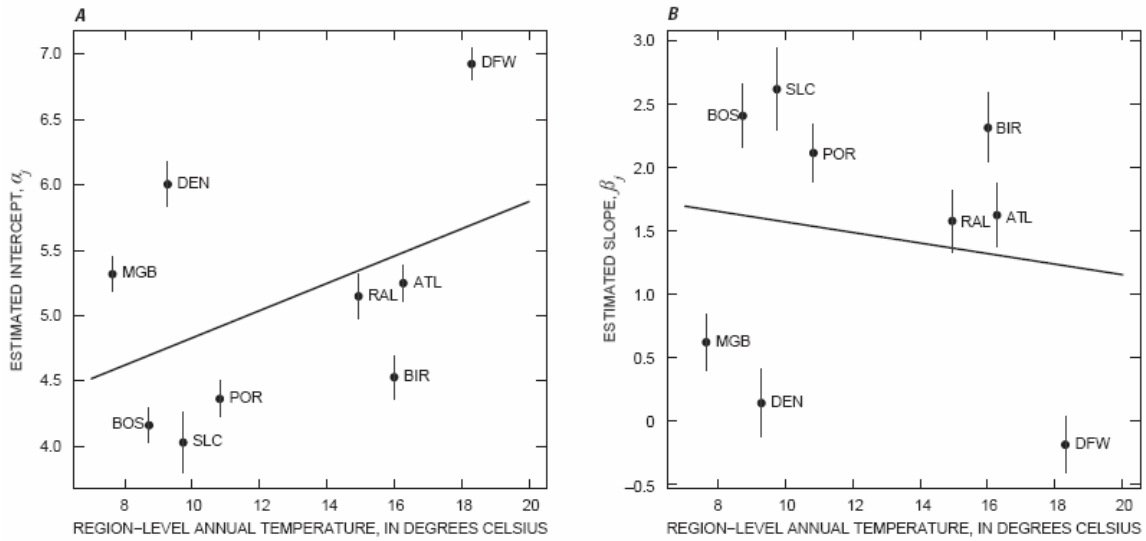

Figure 26: RICHTOL multilevel hierarchical Model 3: Region-level temperature predictor for (A) intercept and (B) slope. Within each region, RICHTOL (richness-weighted mean tolerance of taxa at a basin) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature as shown in Equation 11 (Template 2, region level).

2.3.2.4. Model 4: PRECIP and TEMP Region-Level Predictors

Model 4 combines the most scientifically plausible climate region-level predictors into one model where intercept varies with temperature, and slope varies with precipitation. The intent is to account for the possibility that temperature governs the baseline ecological response at pre-urbanization, while precipitation governs the rate of change in ecological response with urban development. In this model, partially pooled

estimates of basin-level intercept (α_j) are modeled at the region level as a linear function of TEMP using region-level intercept, $\gamma_{\alpha 0}$, and region-level slope, $\gamma_{\alpha 1}$. Meanwhile, partially pooled estimates of basin-level slope (β_j) are modeled as a linear function of PRECIP using region-level intercept, $\gamma_{\beta 0}$, and region-level slope, $\gamma_{\beta 1}$. Coefficients σ_α^2 and σ_β^2 then represent between-region variance in intercept as predicted by TEMP and between-region variance in slope as predicted by PRECIP, respectively.

For NMDS1, patterns of intercept increase with temperature (Figure 27*A*;

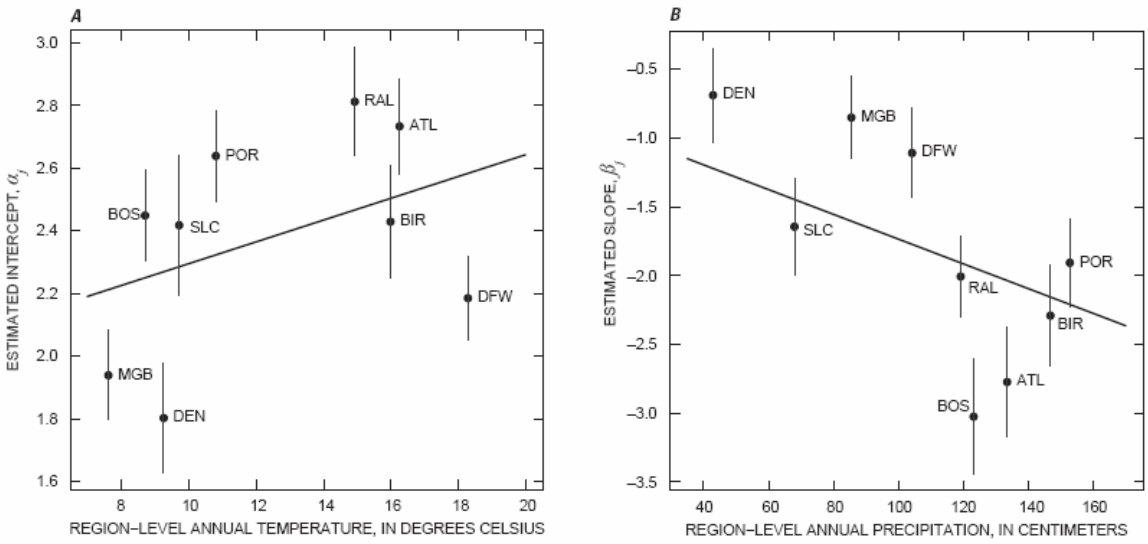

Figure 27: NMDS1 multilevel hierarchical Model 4: Region-level temperature predictor for (A) intercept and region-level precipitation predictor for (B) slope. Within each region, NMDS1 (first axis adjusted nonmetric multidimensional scaling site score) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

Table 14) and negative slope decrease with precipitation (Figure 27*B*; Table 14) do not change substantially from those in Models 3 and 2, respectively. Quantitatively, goodness of Model 4 fit (DIC = 484.0, Table 9) lies between that of Model 2 (DIC = 472.8, Table 9) and Model 3 (DIC = 489.9, Table 9). Scientifically, this model supports the assumption that temperature affects baseline macroinvertebrate species condition in regions with no urban land cover while precipitation affects the rate of change in species condition with increasing urbanization. However, the DIC value of this model also suggests that, although temperature is capable of explaining differences in NMDS1 intercepts, perhaps precipitation explains these differences better.

For RICH, annual precipitation appears to increase the steepness of negative slope whether intercept is modeled with temperature (Figure 28*B*; Table 15) or precipitation (Figure 20*B*). However, a relation between temperature and intercept is only evident when slope is modeled with precipitation (Figure 28*A*) but not when both slope and intercept are modeled with temperature (Figure 24*A*). This may mean that when variability in slope is better accounted for by precipitation, then temperature can explain variability in intercept. Otherwise, temperature is not able to explain differences in intercepts. Directionally, RICH Model 4 suggests that warmer temperature may be a driving force for greater richness of macroinvertebrates on un-urbanized land when higher precipitation is explaining spatially faster macroinvertebrate richness decline with urbanization.

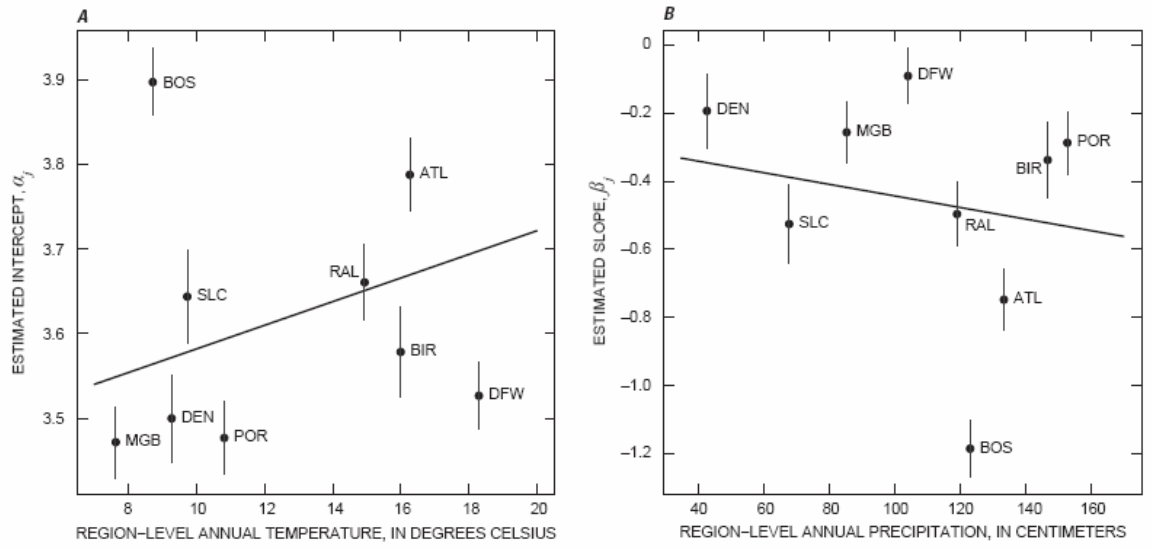

Figure 28: RICH multilevel hierarchical Model 4: Region-level temperature predictor for (A) intercept and region-level precipitation predictor for (B) slope. Within each region, RICH (total taxa richness) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 2, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

EPTRICH Model 4 shows a slight improvement in the capacity of temperature to explain regional differences in intercept (Figure 29*A*) compared to EPTRICH Model 3 (Figure 25*A*). Intercept increases minimally with temperature when slope is modeled with precipitation; however, between-region variation in intercept increases from Model 3 (σ_α^2 = 0.162 relative to 0.154, Table 16). In contrast, the ability of precipitation to predict differences in slopes is diminished when intercept is modeled with temperature (Figure 29*B*) instead of with precipitation as in Model 2 (Figure 21*B*). Nonetheless, slopes continue to become more negative with increasing precipitation although less so. Deviations of the agriculture-dominant group (DEN, DFW, MGB) from the others are

still distinguishable. Therefore, the effect of agriculture is expected to overrule either temperature or precipitation in terms of predicting the behavior of EPTRICH.

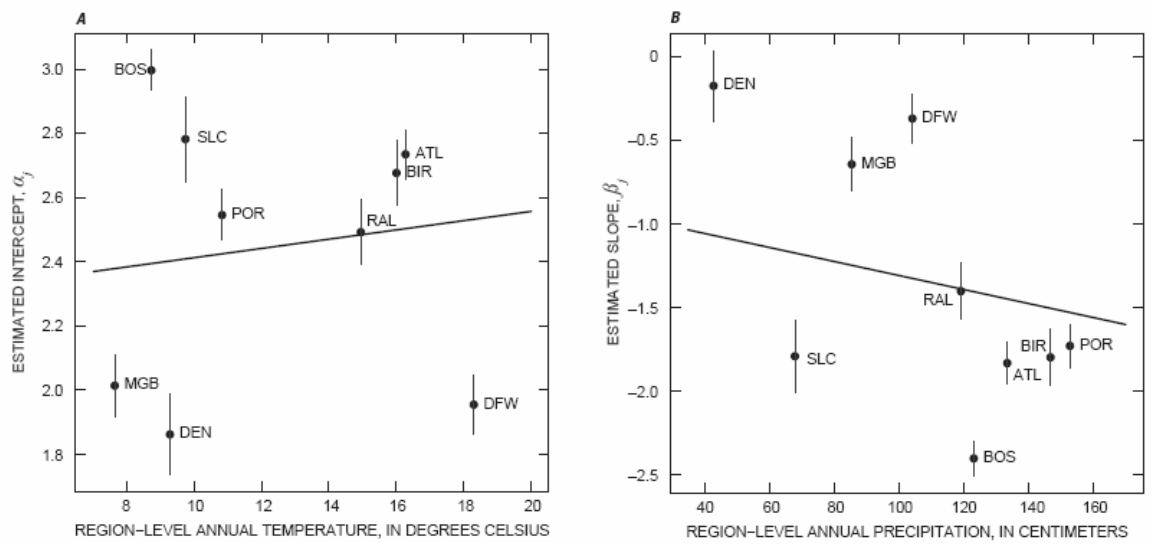

Figure 29: EPTRICH multilevel hierarchical Model 4: Region-level temperature predictor for (A) intercept and region-level precipitation predictor for (B) slope. Within each region, EPTRICH (combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 2, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

For RICHTOL, the overall pattern of intercept increasing with temperature in Model 4 (Figure 30*A*) does not change substantially from the pattern observed in Model 3 (Figure 26*A*). However, the pattern of increasing slopes with increasing precipitation seen in Model 2 (Figure 22*B*) is reversed in direction (Figure 30*B*). When intercept is modeled with temperature instead of precipitation, RICHTOL slopes decrease instead of increase with precipitation. However, greater between-region variance in slope is

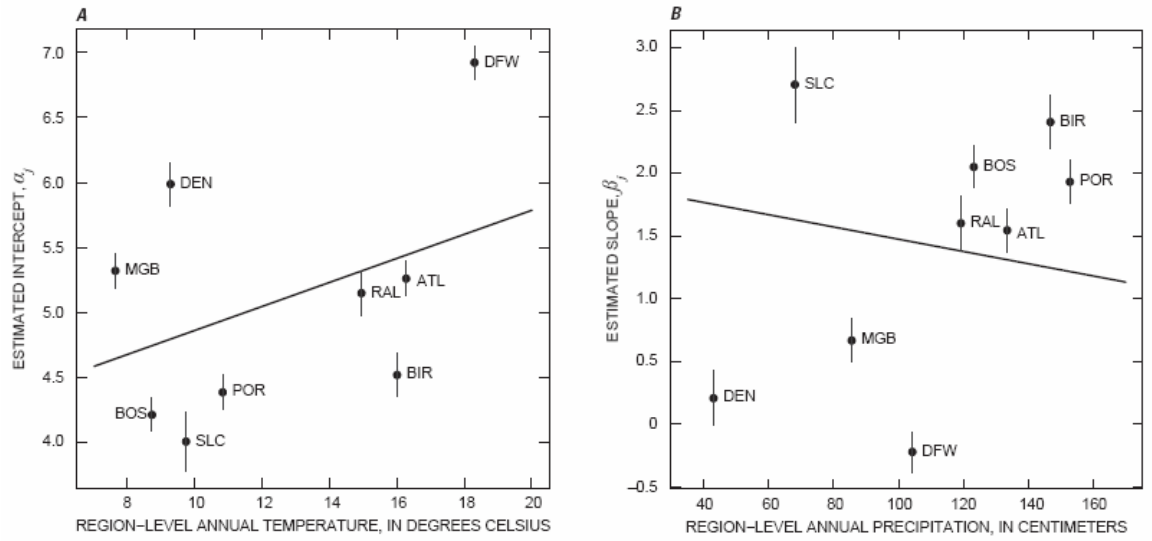

Figure 30: RICHTOL multilevel hierarchical Model 4: Region-level temperature predictor for (A) intercept and region-level precipitation predictor for (B) slope. Within each region, RICHTOL (richness-weighted mean tolerance of taxa at a basin) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and slopes (B) are modeled as a function of regional precipitation as shown in Equation 11 (Template 2, region level).

observed for Model 4 (σ_β^2 = 1.245, Table 17) than for Model 2 (σ_β^2 = 1.088 , Table 17). Decrease in BOS estimated slope from Model 2 to Model 4 appears to have influenced the reversal in relation of slopes with precipitation. Even though quantitatively the goodness of fit for Model 4 (DIC = 412.7, Table 9) is better than that of Models 2 and 3, the presence of high leverage points tends to skew the expected line. The lmer procedure does not allow for the ability to fix this problem. Scientifically, this model suggests that temperature is not well related to baseline macroinvertebrate species tolerance in regions with no urban land cover, while precipitation may either increase or decrease the rate of change in RICHTOL with increasing urbanization.

2.3.2.5. Model 5: AG (Continuous) Region-Level Predictor

Region-level antecedent agriculture and pasture land cover percentage are used to predict intercept and slope in Model 5. This region-level variable attempts to describe the amount of agricultural land use in drainage basins with low urban land use, as a surrogate for type of land cover in a region prior to urbanization of its drainage basins (in other words, antecedent agricultural land cover). Despite the use of a continuous variable format, antecedent agricultural land use in the nine regions is not distributed continuously but rather divides the regions into two major groups: one with low antecedent agriculture land cover (BOS, SLC, BIR, POR, ATL, RAL all <30 percent) and one with high antecedent agriculture land cover (DFW, MGB, DEN all >70 percent). In the previous models, DEN, DFW, and MGB intercepts and slopes often form a cluster separate from the six other regions. This pattern can now be accounted for by incorporating the antecedent agriculture variable. Coefficients are the same as in Models 2–4 and now represent slopes, intercepts, and variances as predicted by region-level antecedent agriculture.

Model 5 has the best quantitative fit of all tested models for all four ecological response variables, but this is likely due to the statistical implications of predicting a line with essentially two points. The distribution of antecedent agricultural land cover (AG) is also interesting from a national context. That is, how representative are these data in terms of the antecedent conditions from which cities are developing? Are there

examples of cities that are developing in areas of 40–70 percent agriculture, or are cities converting lands that are either primarily in forest or intensive agriculture? The NAWQA Program has data from three other metropolitan regions that were not part of the EUSE studies (Anchorage, AK; Chicago, IL; and Seattle, WA) that fall into the existing pattern: Anchorage, forest; Chicago, high agriculture; and Seattle, forest.

This division into two groups clearly shows that regions with low antecedent agriculture have high intercepts, and regions with high antecedent agriculture have low intercepts for NMDS1 (Figure 31*A*; Table 14), RICH (Figure 32*A*; Table 15), and

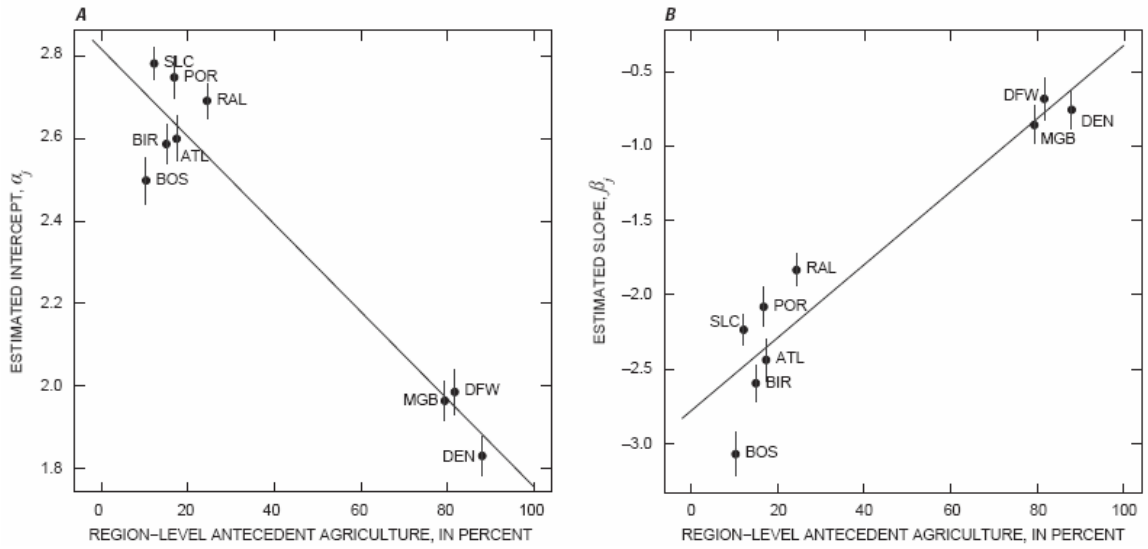

Figure 31: NMDS1 multilevel hierarchical Model 5: Region-level antecedent agriculture percent predictor for (A) intercept and (B) slope. Within each region, NMDS1 (first axis adjusted nonmetric multidimensional scaling site score) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of antecedent agriculture as shown in Equation 11 (Template 2, region level).

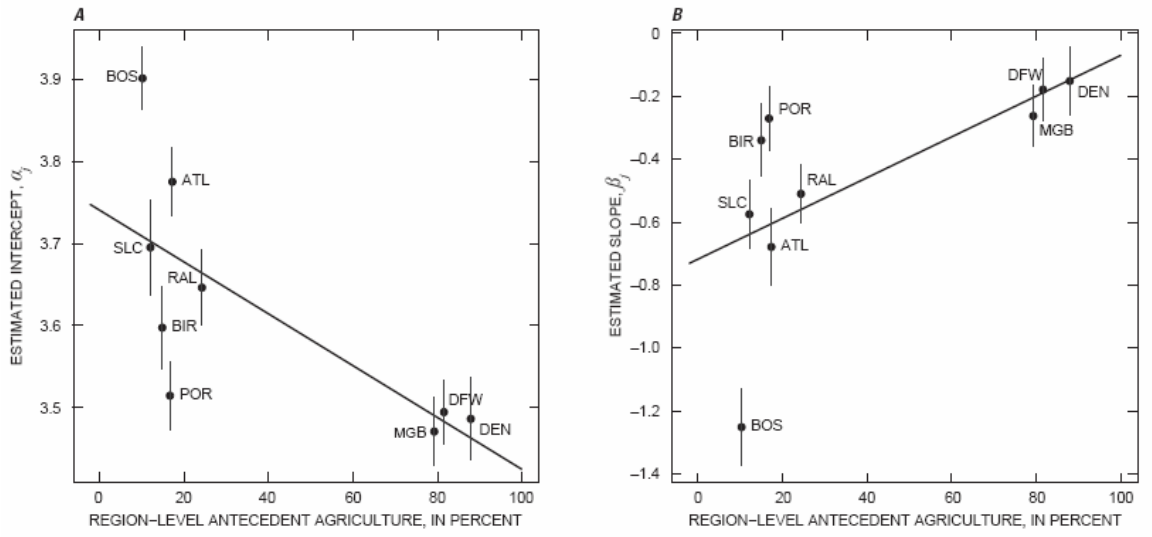

Figure 32: RICH multilevel hierarchical Model 5: Region-level antecedent agriculture percent predictor for (A) intercept and (B) slope. Within each region, RICH (total taxa richness) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of antecedent agriculture as shown in Equation 11 (Template 2, region level).

EPTRICH (Figure 33*A*; Table 16). RICHTOL, just as clearly, has low intercepts at low antecedent agriculture, and high intercepts for high antecedent agriculture (Figure 34*A*; Table 17). This means that, at zero urbanization, ecological communities living on land that was previously agricultural are more degraded than communities in areas that were not agricultural in the past. Agricultural activities, therefore, appear to degrade macroinvertebrate assemblages, even if there is no urbanization effect. High antecedent agriculture land use in DFW, MGB, and DEN leads to lower ordination scores, lower total and EPT richness, and higher tolerances in these regions prior to urbanization.

Regions with low antecedent agriculture also have steeper negative slopes than

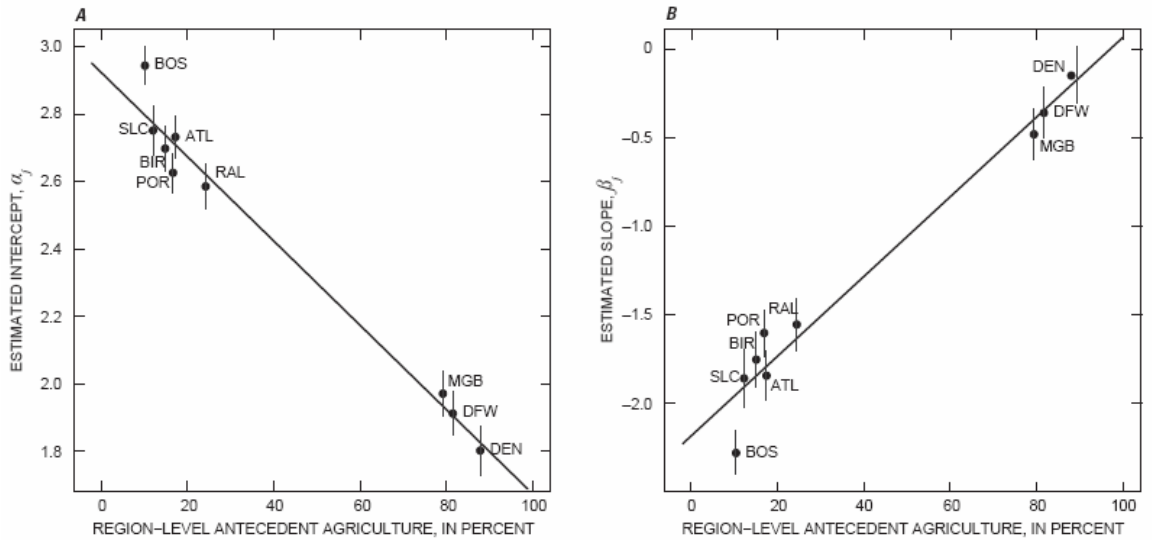

Figure 33: EPTRICH multilevel hierarchical Model 5: Region-level antecedent agriculture percent predictor for (A) intercept and (B) slope. Within each region, EPTRICH (combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders) is modeled as a log-linear function of URB (percent urban land cover) as shown in eqs. 18 and 19 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of antecedent agriculture as shown in eq. 11 (Template 2, region level).

regions with high antecedent agriculture for NMDS1 (Figure 31*B*), RICH (Figure 32*B*), and EPTRICH (Figure 33*B*) and steeper positive slopes for RICHTOL (Figure 34*B*). Rate of change of each of these four macroinvertebrate measures is closer to zero in regions with high antecedent agriculture. Since ecological communities in regions with a lot of previously converted agricultural land have already been disturbed, there is little further decline in macroinvertebrate response compared to regions with less disturbed pre-urbanization land cover. That is, the same amount of urbanization has a greater effect in regions with low antecedent agriculture than in regions with high antecedent agriculture. Model 5 has the best quantitative fit of all tested models for three of the four

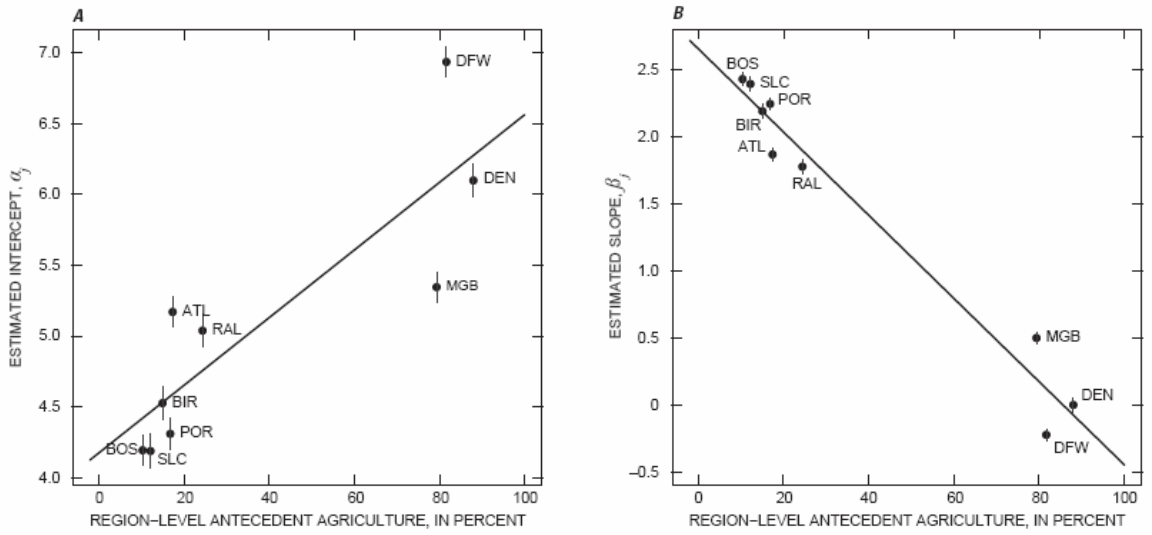

Figure 34: RICHTOL multilevel hierarchical Model 5: Region-level antecedent agriculture percent predictor for (A) intercept and (B) slope. Within each region, RICHTOL (richness-weighted mean tolerance of taxa at a basin) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 10 (Template 2, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of antecedent agriculture as shown in Equation 11 (Template 2, region level).

response variables (DIC = 462.2, 334.9, 402.1 for NMDS1, EPTRICH, and RICHTOL, respectively, Table 9), but this is likely due to the statistical implications of predicting a line with essentially two points. Regardless of the degree of fitness, this model strongly supports the idea speculated in the previous models that past agricultural practices determine the patterns of change in macroinvertebrate measures associated with urbanization. These model results imply that the effect of urbanization on stream ecosystems should not be analyzed in isolation.

2.3.2.6. Model 6: AG (Categorical) and PRECIP Region-Level Predictors

The continuous AG variable naturally divides the regions into two groups (Figure 35–Figure 38), which disrupts linear regression assumptions concerning coverage across the whole range of possible predictor variable values. Therefore, in

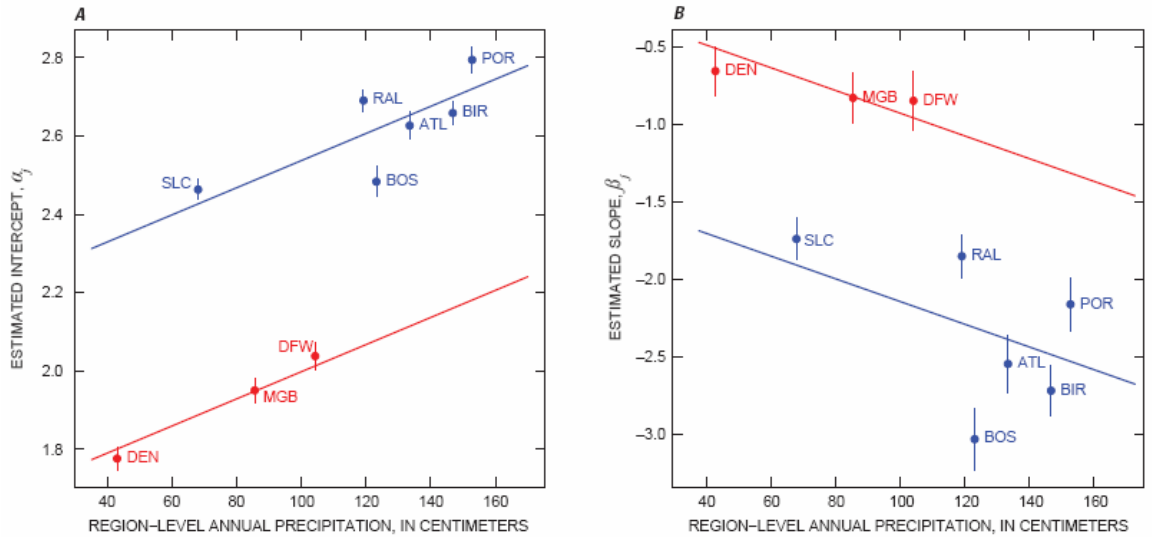

Figure 35: NMDS1 multilevel hierarchical Model 6: Region-level precipitation predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue: low AG) for (A) intercept and (B) slope. Within each region, NMDS1 (first axis adjusted nonmetric multidimensional scaling site score) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 12 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

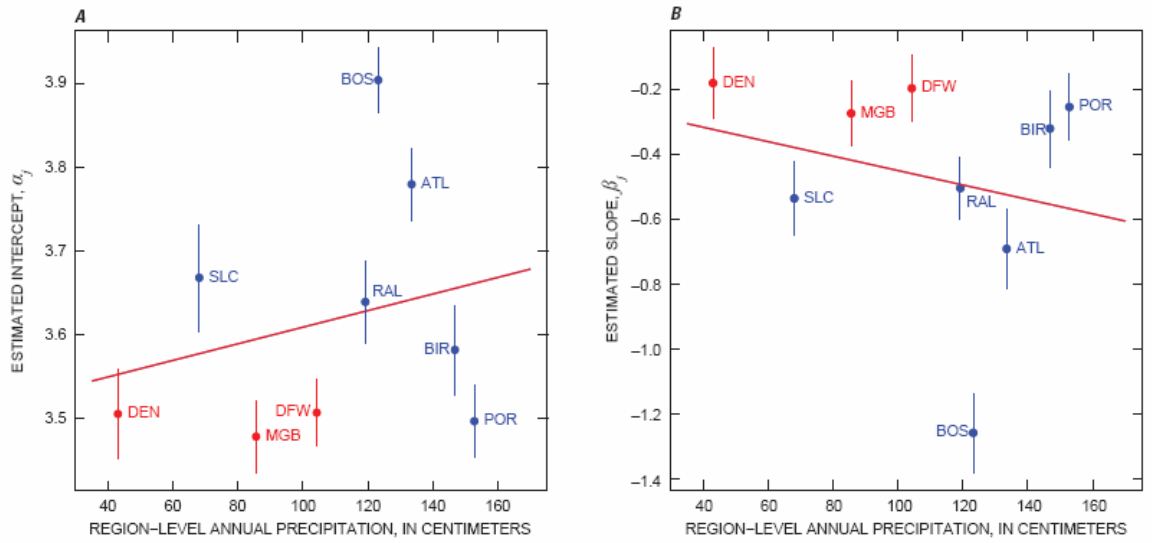

Figure 36: RICH multilevel hierarchical Model 6: Region-level precipitation predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue: low AG) for (A) intercept and (B) slope. Within each region, RICH (total taxa richness) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

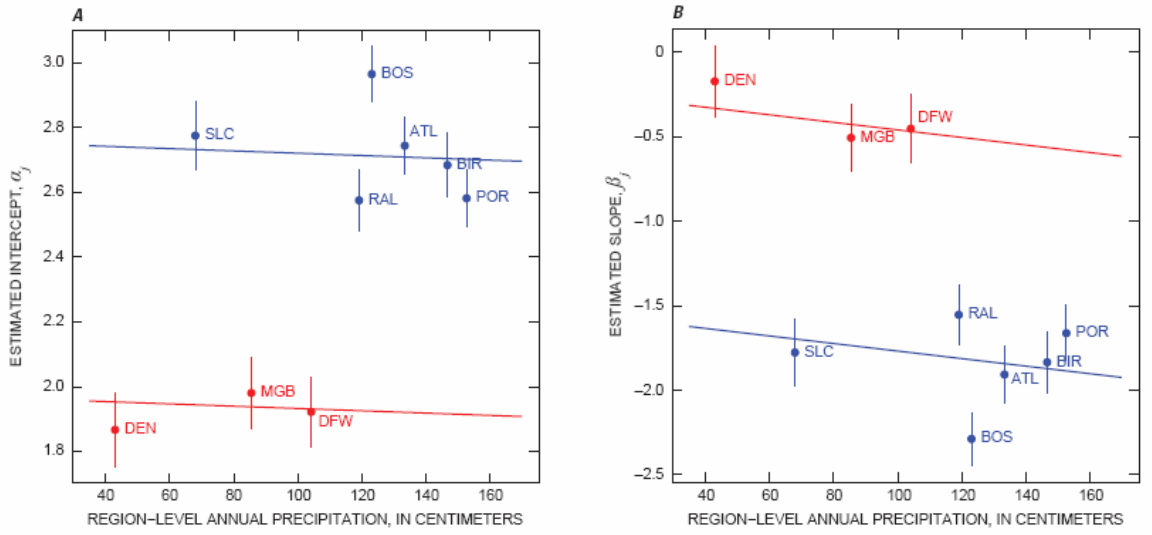

Figure 37: EPTRICH multilevel hierarchical Model 6: Region-level precipitation predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue: low AG) for (A) intercept and (B) slope. Within each region, EPTRICH

(combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

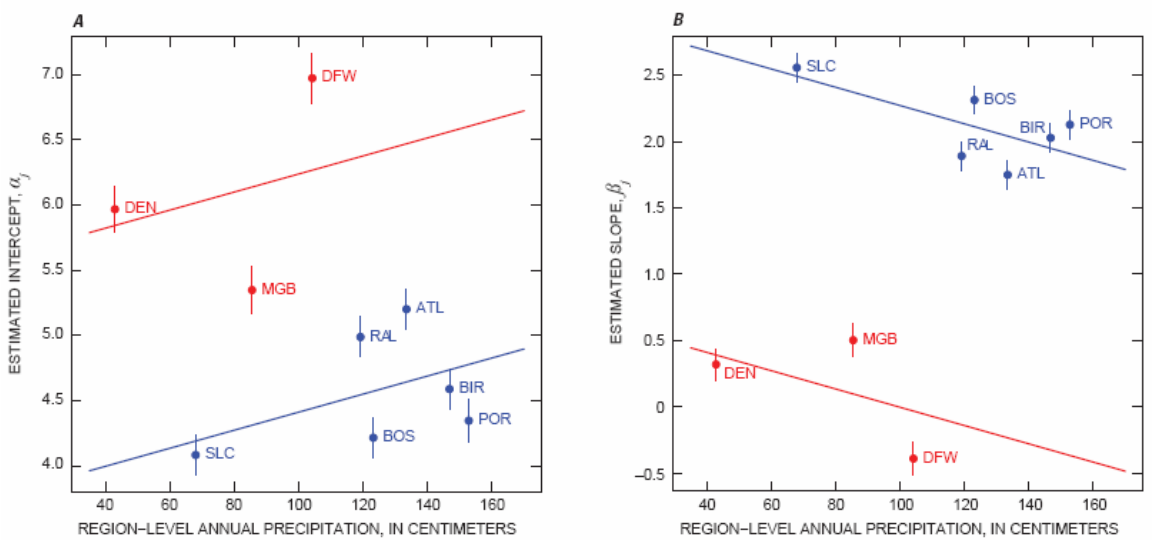

Figure 38: RICHTOL multilevel hierarchical Model 6: Region-level precipitation predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue: low AG) for (A) intercept and (B) slope. Within each region, RICHTOL (richness-weighted mean tolerance of taxa at a basin) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 12 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

Model 6, antecedent agriculture percentage is converted into a categorical high or low antecedent agriculture predictor. Region-level precipitation is also included as an additional region-level predictor of intercept and slope. Model 6, then, fits the ecological response data to a non-nested model, which incorporates two different levels of grouping: region and antecedent agricultural category. This model structure examines the dependence of ecological responses on percentage of urban land cover at the basin

level and annual mean precipitation at the region level, conditional on the levels of region and antecedent agriculture. The antecedent agriculture category is not allowed to interact with precipitation because region sample size (n = 9) is not large enough to support this estimation accurately. Coefficients α_j and β_j now represent the PRECIP component of the region-level intercept and slope, while α_k and β_k represent the AG component of the region-level intercept and slope. Together $\alpha_j + \alpha_k$ and $\beta_j + \beta_k$ represent the total of the region-level intercept and slope terms. Higher tier region-level regression parameters, $\gamma_{\alpha 0}, \gamma_{\alpha 1}, \gamma_{\beta 0}$, and $\gamma_{\beta 1}$, still describe only the relation between region-level predictors α_j and β_j and PRECIP. And variances are modeled using σ_y^2 for within-region variance in invertebrate response, σ_α^2 for between-region variance in intercept, σ_β^2 for between-region variance in slope, and σ_k^2 for between-category (AG) variance in slope and intercept.

For NMDS1 and EPTRICH, regions with low antecedent agriculture (Figure 35*A* and Figure 37*A*, blue lines; Table 18 and Table 19) continue to have estimated intercepts greater than regions with high antecedent agriculture (Figure 35*A* and Figure 37*A*, red lines) across all values of annual precipitation. Also, as shown previously, slope is steeper for regions with low antecedent agriculture (Figure 35*B* and Figure 37*B*, blue lines) than for regions with high antecedent agriculture (Figure 35*B* and Figure 37*B*, red lines). Low and high antecedent agriculture trend lines are parallel because Model 6

Table 18: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response NMDS1 Models 6–8. [See Table 5 for model definitions]

	Model 6		Model 7		Model 8	
	α_j	β_j	α_j	β_j	α_j	β_j
ATL	2.63	-2.54	2.70	-2.70	2.69	-2.66
BIR	2.66	-2.72	2.69	-2.74	2.68	-2.75
BOS	2.48	-3.03	2.44	-2.94	2.42	-2.92
DEN	1.78	-0.65	1.85	-0.79	1.83	-0.73
DFW	2.04	-0.84	2.10	-0.94	2.10	-0.93
MGB	1.95	-0.83	1.83	-0.62	1.84	-0.67
POR	2.79	-2.16	2.66	-1.90	2.69	-2.01
RAL	2.69	-1.85	2.77	-1.97	2.77	-1.96
SLC	2.46	-1.74	2.61	-1.97	2.59	-1.90
σ_y^2	0.358		0.356		0.356	
$\gamma_{\alpha 0} + \delta_{\alpha k}$	highAG (*k*=1): 1.652 lowAG (*k*=0): 2.191		highAG (*k*=1): 1.616 lowAG (*k*=0): 2.307		highAG (*k*=1): 1.609 lowAG (*k*=0): 2.298	
$\gamma_{\alpha 1}$	0.003		0.026		0.027	
$\gamma_{\beta 0} + \delta_{\beta k}$	highAG (*k*=1): -0.210 lowAG (*k*=0): -1.426		highAG (*k*=1): -0.497 lowAG (*k*=0): -2.057		highAG (*k*=1): -0.326 lowAG (*k*=0): -1.661	
$\gamma_{\beta 1}$	-0.007		-0.025		-0.006	
σ_α^2	Region (*j*): 0.006 AG (*k*): 0.157		Region (*j*): 0.004 AG (*k*): 0.245		Region (*j*): 0.006 AG (*k*): 0.244	
σ_β^2	Region (*j*): 0.169 AG (*k*): 0.797		Region (*j*): 0.187 AG (*k*): 1.248		Region (*j*): 0.162 AG (*k*): 0.918	

Table 19: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response EPTRICH Models 6–8. [N/A, not applicable because data variance is equal to the mean parameter (λ) in Poisson models and, therefore, changes with changing λ_j. See Table 5 for model definitions]

	Model 6		Model 7		Model 8	
	α_j	β_j	α_j	β_j	α_j	β_j
ATL	2.74	-1.91	2.72	-1.81	2.75	-1.89
BIR	2.68	-1.84	2.66	-1.69	2.69	-1.84
BOS	2.96	-2.29	2.97	-2.29	2.97	-2.33
DEN	1.87	-0.17	1.88	-0.27	1.85	-0.15
DFW	1.92	-0.45	1.90	-0.30	1.93	-0.43
MGB	1.98	-0.51	2.00	-0.52	1.98	-0.53
POR	2.58	-1.66	2.59	-1.59	2.58	-1.70
RAL	2.58	-1.55	2.57	-1.53	2.57	-1.53
SLC	2.77	-1.78	2.76	-1.88	2.76	-1.76
σ_y^2	N/A		N/A		N/A	
$\gamma_{\alpha 0} + \delta_{\alpha k}$	highAG (k=1): 1.963 lowAG (k=0): 2.756		highAG (k=1): 2.095 lowAG (k=0): 2.877		highAG (k=1): 1.883 lowAG (k=0): 2.651	
$\gamma_{\alpha 1}$	0.000		-0.014		0.005	
$\gamma_{\beta 0} + \delta_{\beta k}$	highAG (k=1): -0.228 lowAG (k=0): -1.546		highAG (k=1): -0.715 lowAG (k=0): -2.141		highAG (k=1): -0.137 lowAG (k=0): -1.408	
$\gamma_{\beta 1}$	-0.002		0.028		-0.003	
σ_α^2	Region (j): 0.002 AG (k): 0.162		Region (j): 0.019 AG (k): 0.158		Region (j): 0.025 AG (k): 0.154	
σ_β^2	Region (j): 0.077 AG (k): 0.450		Region (j): 0.067 AG (k): 0.527		Region (j): 0.085 AG (k): 0.421	

does not incorporate interaction between antecedent agriculture and precipitation. Similarly, RICHTOL intercepts continue to be lower for regions with low antecedent agriculture (Figure 38*A*, blue line) and higher for regions with high antecedent agriculture (Figure 38*A*, red line) across all values of annual precipitation. RICHTOL

slopes are higher for regions with low antecedent agriculture (Figure 38*B*, blue line) than for regions with high antecedent agriculture (Figure 38*B*, red line).

NMDS1 intercept still increases with increasing precipitation (Figure 35*A*) and negative slope becomes steeper with increasing precipitation (Figure 35*B*) in Model 6 compared to Model 2 (Figure 19); however, when categorical AG is introduced into the model, relations of EPTRICH and RICHTOL with precipitation change. Precipitation no longer has an increasing effect on EPTRICH intercept (Figure 21*A*) and a decreasing effect on EPTRICH negative slope (Figure 21*B*) when AG is included, instead little relation is shown between either intercept (Figure 37*A*) or slope (Figure 37*B*) and precipitation. Precipitation's decreasing effect on RICHTOL intercept (Figure 22*A*) and increasing effect on RICHTOL slope (Figure 22*B*) are actually reversed when AG is included (Figure 38). Model 6 shows that antecedent agriculture is an important categorical predictor of regional differences in ordination scores, indicator species richness and richness-weighted tolerance pre-urbanization condition, and rate of change with urbanization, in some cases even influencing the effect of precipitation.

For NMDS1, EPTRICH, and RICHTOL, despite having less favorable measures of fit (DIC = 471.4, 345.7, and 417.8, respectively, Table 9) than Model 5 (DIC = 462.2, 334.9, and, 402.1, respectively, Table 9), Model 6 visually appears to describe the data well. This is likely because Model 6 accounts for two major region-level influences on differences between basin-level models. Both antecedent agriculture levels and annual

precipitation affect how different regions respond to urbanization and account for region differences in baseline invertebrate assemblage conditions. As these region-level variables affect separate system elements (land use and climate), it is important to evaluate the influence of both simultaneously. This reasoning is quantitatively supported as inclusion of an agricultural predictor to the precipitation model improves model fit to the data for NMDS1, EPTRICH, and RICHTOL (DIC decreases from 472.8, 351.8, and 425.1, respectively, for Model 2 to 471.4, 345.7, and 417.8, respectively, for Model 6, Table 9).

RICH intercept and slope estimates, on the other hand, have zero variance between high and low AG groups (Figure 36; Table 20). This means RICH Model 6 calculates no differences in the number of taxa supported by regions with low antecedent agricultural land use and regions with high antecedent agricultural land use pre-urbanization or in rate of change in taxa with urbanization. Therefore, with no AG influence, the effect of precipitation mirrors the results of RICH Model 2 (Figure 20), and DIC does not differ between Model 6 and Model 2 (Table 9). Looking at RICH Model 5 relative to Model 5 for the other three response variables, low AG intercepts vary substantially across the range of all intercept estimates more for RICH (Figure 32*A*) than for NMDS1 (Figure 31*A*), EPTRICH (Figure 33*A*), and RICHTOL (Figure 34*A*). Also, with the exception of BOS, differences between high AG and low AG slopes are not great (Figure 32*B*). In RICH Model 6, these patterns are translated into no clear

Table 20: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response RICH Models 6–8. [N/A, not applicable because data variance is equal to the mean parameter (λ) in Poisson models and, therefore, changes with changing λ_j. See Table 5 for model definitions]

	Model 6		Model 7		Model 8	
	α_j	β_j	α_j	β_j	α_j	β_j
ATL	3.78	-0.69	3.78	-0.68	3.78	-0.75
BIR	3.58	-0.32	3.59	-0.31	3.59	-0.34
BOS	3.90	-1.26	3.90	-1.25	3.90	-1.24
DEN	3.50	-0.18	3.49	-0.19	3.48	-0.17
DFW	3.50	-0.19	3.51	-0.13	3.52	-0.12
MGB	3.47	-0.27	3.47	-0.28	3.46	-0.27
POR	3.50	-0.26	3.50	-0.25	3.49	-0.29
RAL	3.65	-0.51	3.66	-0.50	3.68	-0.51
SLC	3.69	-0.57	3.65	-0.53	3.67	-0.54
σ_y^2	N/A		N/A		N/A	
$\gamma_{\alpha 0} + \delta_{\alpha k}$	highAG (k=1): 3.510 lowAG (k=0): 3.510		highAG (k=1): 3.602 lowAG (k=0): 3.602		highAG (k=1): 3.353 lowAG (k=0): 3.501	
$\gamma_{\alpha 1}$	0.001		0.001		0.014	
$\gamma_{\beta 0} + \delta_{\beta k}$	highAG (k=1): -0.226 lowAG (k=0): -0.226		highAG (k=1): -0.723 lowAG (k=0): -0.723		highAG (k=1): -0.058 lowAG (k=0): -0.295	
$\gamma_{\beta 1}$	-0.002		0.021		-0.002	
σ_α^2	Region (j): 0.020 AG (k): 0.000		Region (j): 0.021 AG (k): 0.000		Region (j): 0.018 AG (k): 0.007	
σ_β^2	Region (j): 0.111 AG (k): 0.000		Region (j): 0.111 AG (k): 0.000		Region (j): 0.094 AG (k): 0.018	

differentiation between high and low AG estimates as distributions for the two groups overlap across intercept (Figure 36*A*) and slope (Figure 36*B*) values. Intercepts of high AG regions are low and slopes are flatter but so are intercepts and slopes of POR and

BIR, indicating that the influence of PRECIP on taxa richness may be greater than the influence of antecedent agriculture.

This lack of variation between high and low AG groups is calculated using URB as a fraction and unscaled PRECIP variables in the R lmer command. If variables are rescaled to URB as a percentage and PRECIP rescaled by dividing by two standard deviations of PRECIP, then regions in the low antecedent agricultural land use support more taxa at zero urbanization and have steeper negative slopes than do regions with high antecedent agricultural land use (ecologically similar to trends observed for NMDS1, EPTRICH, and RICHTOL), and response to PRECIP looks nearly flat (similar to EPTRICH Model 6). This change in results depending on variable scaling showcases one of the drawbacks of lmer, because it appears that rescaling has a non-negligible effect on numerical estimation results.

2.3.2.7. Model 7: AG (Categorical) and TEMP Region-Level Predictors

Model 7 substitutes the region-level temperature predictor for the region-level precipitation predictor, retaining the categorical antecedent agriculture predictor (Figure 39–Figure 42). Again, conditional on the two non-nested levels of region and antecedent agriculture, this model examines the dependence of ecological responses on percentage of urban land cover at the basin level and annual mean temperature at the region level without accounting for interaction between the two group-level predictors. Coefficients remain the same as Model 6, substituting TEMP for PRECIP.

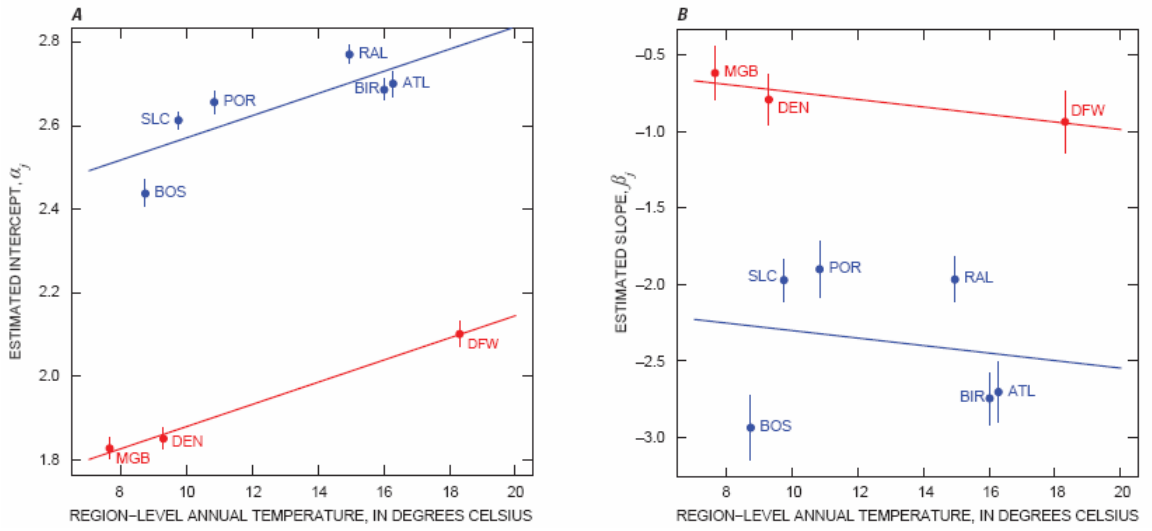

Figure 39: NMDS1 multilevel hierarchical Model 7: Region-level temperature predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue: low AG) for (A) intercept and (B) slope. Within each region, NMDS1 (first axis adjusted nonmetric multidimensional scaling site score) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 12 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

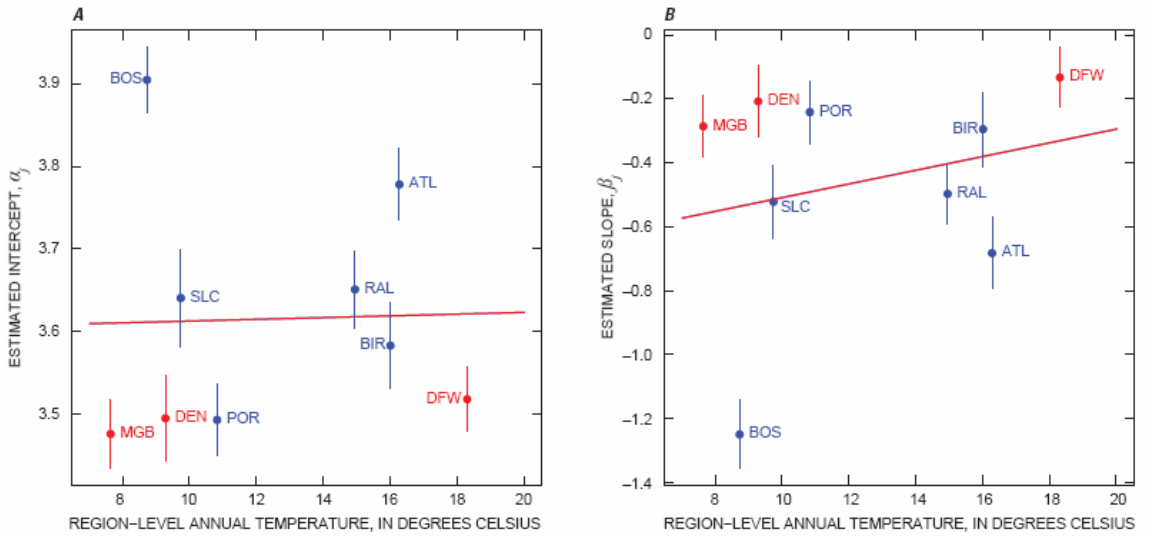

Figure 40: RICH multilevel hierarchical Model 7: Region-level temperature predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue:

low AG) for (A) intercept and (B) slope. Within each region, RICH (total taxa richness) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

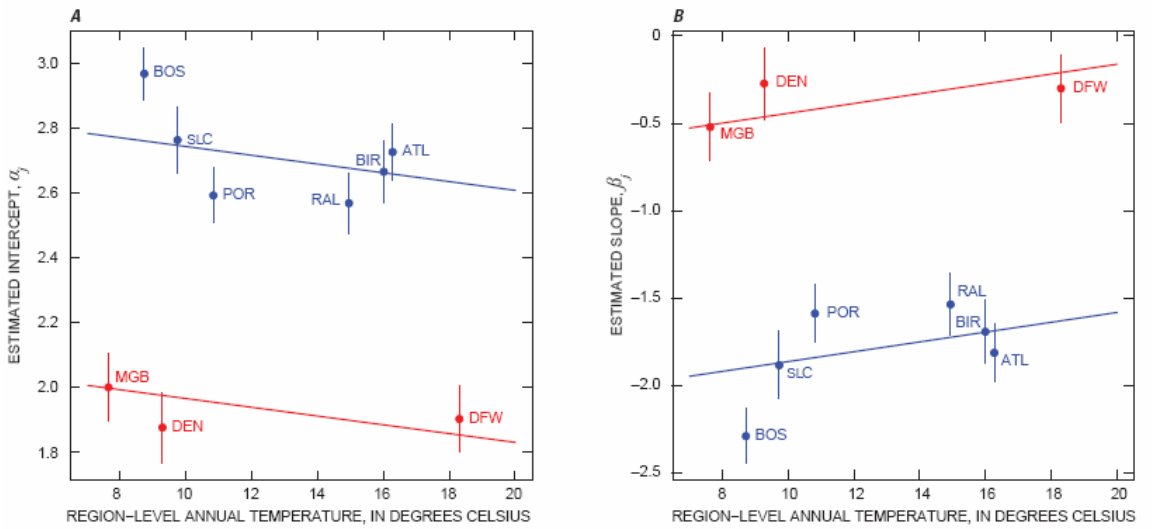

Figure 41: EPTRICH multilevel hierarchical Model 7: Region-level temperature predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue: low AG) for (A) intercept and (B) slope. Within each region, EPTRICH (combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

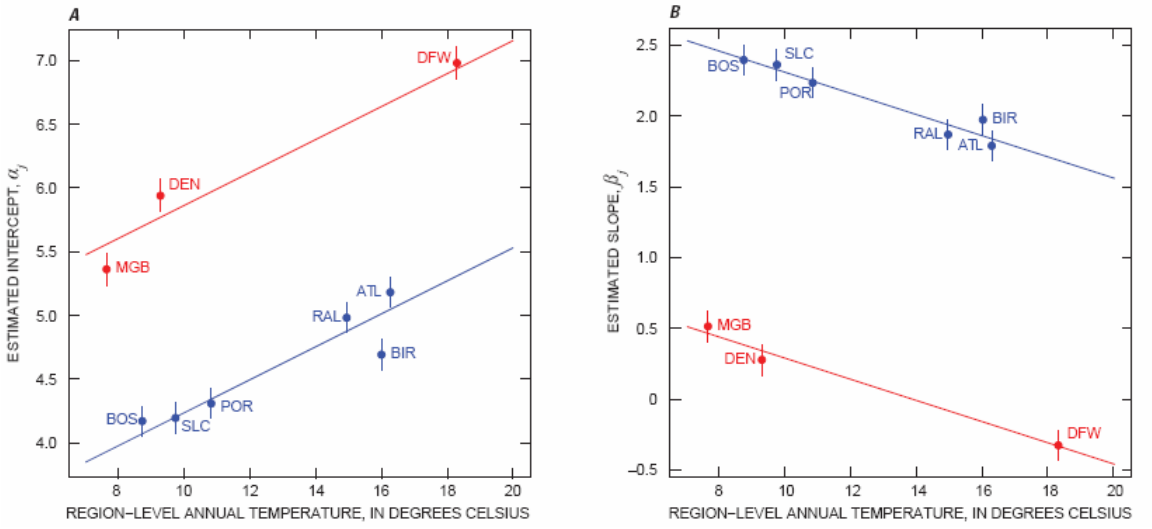

Figure 42: RICHTOL multilevel hierarchical Model 7: Region-level temperature predictor and categorical antecedent agriculture (AG) predictor (red: high AG; blue: low AG) for (A) intercept and (B) slope. Within each region, RICHTOL (richness-weighted mean tolerance of taxa at a basin) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 12 (Template 3, basin level). Across regions, intercepts (A) and slopes (B) are modeled as a function of regional temperature (in degrees Celsius) and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

Similar to Model 6, the effect of antecedent agriculture is evident and consistent for NMDS1, EPTRICH, and RICHTOL. Regions with low antecedent agriculture have higher intercepts (Figure 39*A* and Figure 41*A*, blue lines; Table 18 and Table 19) and steeper negative slopes (Figure 39*B* and Figure 41*B*, blue lines; Table 18 and Table 19) than regions with high antecedent agriculture (Figure 39 and Figure 41, red lines; Table 18 and Table 19) across all values of annual temperature for NMDS1 and EPTRICH, while, for RICHTOL, regions with low antecedent agriculture have lower intercepts (Figure 42*A*, blue lines; Table 21) and steeper positive slopes (Figure 42*B*, blue lines; Table 21) than regions with high antecedent agriculture (Figure 42, red lines; Table 21)

across all values of annual temperature. Again, the two antecedent agriculture trend lines are parallel because Model 7 does not incorporate interaction between antecedent agriculture and temperature.

Table 21: Regional intercept (α_j) and slope (β_j) coefficient estimates, representing regional background condition prior to urbanization and regional rate of change with urbanization, respectively, hyperparameter intercept and slope coefficient estimates, and variance coefficient estimates for invertebrate response RICHTOL Models 6–8. [See Table 5 for model definitions]

	Model 6		Model 7		Model 8	
	α_j	β_j	α_j	β_j	α_j	β_j
ATL	5.20	1.75	5.18	1.79	5.17	1.80
BIR	4.59	2.03	4.69	1.97	4.61	2.19
BOS	4.22	2.31	4.17	2.40	4.23	2.01
DEN	5.97	0.32	5.94	0.28	5.96	0.27
DFW	6.97	-0.39	6.98	-0.33	6.92	-0.14
MGB	5.35	0.50	5.36	0.51	5.41	0.34
POR	4.35	2.13	4.31	2.24	4.39	1.87
RAL	4.99	1.89	4.98	1.87	4.98	1.94
SLC	4.08	2.56	4.19	2.37	4.11	2.56
σ_y^2	0.275		0.275		0.276	
$\gamma_{\alpha 0} + \delta_{\alpha k}$	highAG (*k*=1): 5.547 lowAG (*k*=0): 3.721		highAG (*k*=1): 4.570 lowAG (*k*=0): 2.943		highAG (*k*=1): 4.881 lowAG (*k*=0): 3.276	
$\gamma_{\alpha 1}$	0.007		0.129		0.103	
$\gamma_{\beta 0} + \delta_{\beta k}$	highAG (*k*=1): 0.686 lowAG (*k*=0): 2.964		highAG (*k*=1): 1.036 lowAG (*k*=0): 3.057		highAG (*k*=1): 0.663 lowAG (*k*=0): 2.850	
$\gamma_{\beta 1}$	-0.007		-0.075		-0.006	
σ_α^2	Region (*j*): 0.306 AG (*k*): 1.713		Region (*j*): 0.044 AG (*k*): 1.339		Region (*j*): 0.063 AG (*k*): 1.306	
σ_β^2	Region (*j*): 0.089 AG (*k*): 2.667		Region (*j*): 0.015 AG (*k*): 2.065		Region (*j*): 0.049 AG (*k*): 2.424	

NMDS1 intercepts continue to increase with increasing temperature (Figure 39*A*), and slopes continue to steepen negatively with increasing temperature (Figure

39*B*). Model 7 quantitative fit (DIC = 479.5, Table 9) is not as good as Model 6 (DIC = 471.4, Table 9), similar to how Model 2 fit with precipitation (DIC = 472.8, Table 9) exceeded Model 3 fit with temperature (DIC = 489.9, Table 9) prior to inclusion of an agricultural predictor. Despite lower DIC for Model 6 than for Model 7, the addition of the agricultural predictor to the temperature model offers valuable interpretation improvements. In Model 3, the deviations of regions, which did not appear to show a linear pattern with intercept and outliers with slope, can now be explained with categorical antecedent agriculture. Before antecedent agriculture was introduced, DEN, MGB, and DFW had lower intercepts (Figure 23*A*) and less negative slopes (Figure 23*B*) than regression lines with temperature. Antecedent agricultural land use explains this pattern. Analysis by visual inspection is supported quantitatively. Accounting for agriculture improves the originally poorer fit of the temperature model (DIC decreases from 489.9 for Model 3 to 479.5 for Model 7, Table 9).

For EPTRICH, when the effect of low or high antecedent agriculture is additionally considered as a grouping factor, an effect of temperature is evident within each group (Figure 41) as compared to Model 3 (Figure 25) which showed no effect of temperature on all regions together. As with NMDS1, observed differences in intercepts and slopes relative to temperature are finally explained and quantified by antecedent agriculture. Within each AG group, intercept now decreases and slope flattens with increasing annual temperature. However, interpreting the regression results at the

region level becomes more difficult when the nine regions are divided into two groups because region-level regressions are performed with merely three to six data points each. As for NMDS1, model fit improves with inclusion of an antecedent agriculture predictor (DIC = 356.8 for Model 3 and 347.1 for Model 7, Table 9).

Unlike the direction of effect reversal caused by addition of AG to the RICHTOL precipitation model, RICHTOL intercepts continue to increase with increasing temperature (Figure 42*A*) and slopes continue to decrease in magnitude with increasing temperature (Figure 42*B*) as they did prior to the inclusion of AG. Quantitative fit (DIC = 410.4, Table 9) is better than all previous models except Model 5 (DIC = 402.1, Table 9). The addition of the agricultural predictor is also able to visually explain intercept and slope differences better than the model using a temperature predictor alone. DEN, MGB, and DFW had higher intercepts (Figure 26*A*) and lower slopes (Figure 26*B*) than the region-level temperature regression lines in Model 3. In Model 6, these differences are explained by differences in antecedent agriculture.

Again, RICH Model 7 calculates no differences between response patterns for regions with high as opposed to low antecedent agriculture (Figure 40; Table 20). Intercept does not appear to vary with temperature (Figure 40*A*), but slope flattens slightly with increasing temperature (Figure 40*B*). As the variance between antecedent agriculture levels is zero, this pattern is essentially mathematically identical to RICH Model 3 (Figure 24) with minor numerical approximation error differences in

coefficients. DIC between RICH Model 3 and Model 7 is identical (Table 9). Previous discussion for RICH Model 6 involving intercept and slope distribution overlap and lmer limitations applies for RICH Model 7, as well.

2.3.2.8. Model 8: AG (Categorical), PRECIP, TEMP Region-Level Predictors

Model 8 combines a region-level temperature predictor for the intercept with a region-level precipitation predictor for the slope and continues to include the categorical antecedent agriculture predictor for both without interaction. Coefficients and model structure are identical to Models 6 and 7 with intercept terms predicted by temperature and slope terms predicted by precipitation.

NMDS1 results (Figure 43; Table 18) duplicate those of Model 7 for intercept (Figure 39*A*) and Model 6 for slope (Figure 35*B*), with quantitative fit between those for Models 6 and 7 (DIC = 474.4 relative to DIC = 471.4 for Model 6 and DIC = 479.5 for Model 7, Table 9). This means intercept increased with temperature and negative slope became more negative with precipitation. Fit improved from that of Model 4 (DIC = 484.0 , Table 9) with the addition of categorical agriculture but was not better than the fit for Model 6, which used only region-level precipitation and categorical agriculture. Therefore, it appears that region-level precipitation describes baseline NMDS1 condition at zero urbanization more accurately than region-level temperature. This may be because an annual mean temperature measure averages over temporal variation resulting in more uncertainty and less descriptive power than a measure of cumulative

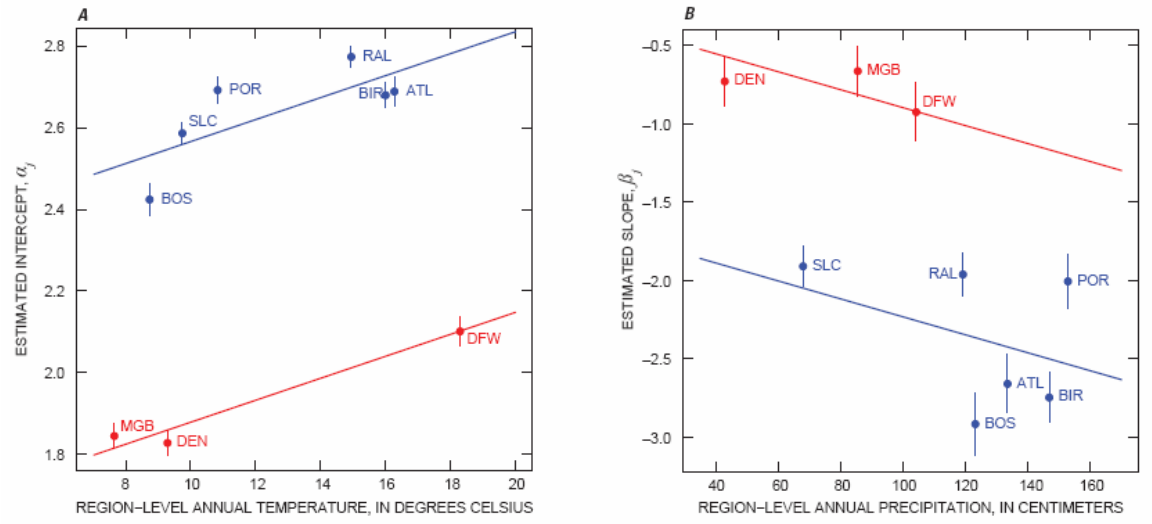

Figure 43: NMDS1 multilevel hierarchical Model 8: Region-level temperature and categorical antecedent agriculture (AG) predictors (red: high AG; blue: low AG) for (A) intercept and region-level precipitation and categorical antecedent agriculture predictors for (B) slope. Within each region, NMDS1 (first axis adjusted nonmetric multidimensional scaling site score) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 12 (Template 3, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and categorical antecedent agriculture and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

precipitation. High antecedent agriculture continued to correlate with lower intercepts and flatter slopes with a large predictive distinction between high and low antecedent agriculture groups.

For RICH, Model 8 was the only model that was able to differentiate between high and low AG in the presence of an additional continuous physical predictor variable. As in Model 5, high antecedent agriculture was associated with lower average intercepts (Figure 44*A*) and flatter average slopes (Figure 44*B*). However, visually, there did not appear to be a large distinction between the two AG groups, with much

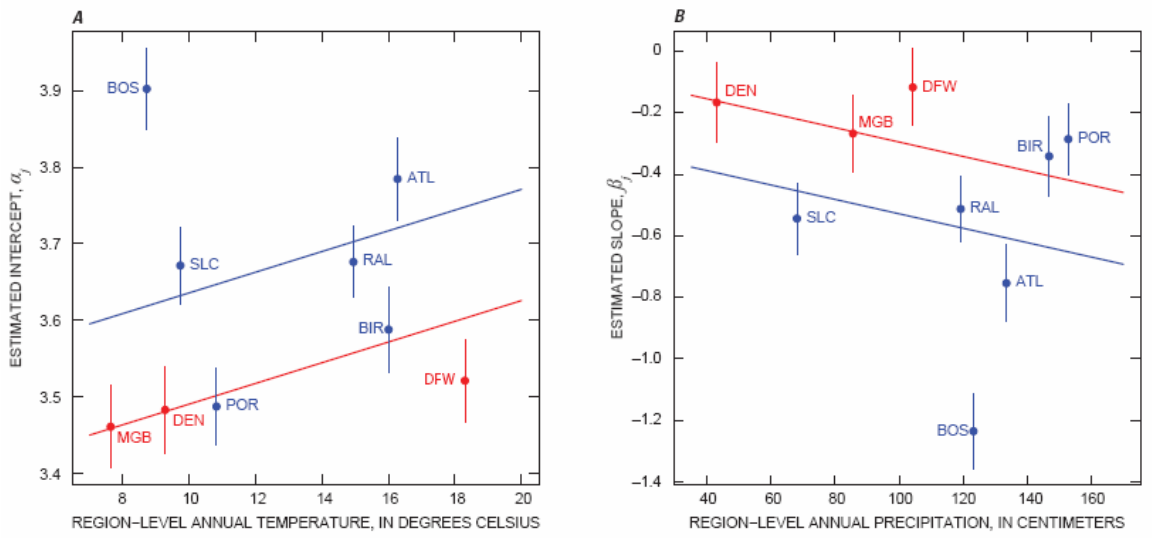

Figure 44: RICH multilevel hierarchical Model 8: Region-level temperature and categorical antecedent agriculture (AG) predictors (red: high AG; blue: low AG) for (A) intercept and region-level precipitation and categorical antecedent agriculture predictors for (B) slope. Within each region, RICH (total taxa richness) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 3, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and categorical antecedent agriculture and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

observed overlap and small between-agriculture-group variance (σ_α^2 = 0.007 and σ_β^2 = 0.018, Table 20). Similar patterns of intercept increase with temperature and negative slope decrease with precipitation were observed in Model 8 as in Model 4 (Figure 28), accompanied by slight DIC improvement with the addition of AG (from 337.7 for Model 4 to 336.5 for Model 8, Table 9).

For EPTRICH, when the effect of low or high antecedent agriculture is additionally considered as a grouping factor, the effects of both temperature and precipitation do not change from Model 4 (Figure 29). That is, within each group, the

intercept slightly increases with increasing annual temperature (Figure 45*A*) and the negative slope decreases further with increasing precipitation (Figure 45*B*). Additionally, the group of high antecedent agriculture (DEN, DFW, and MGB) has lower intercept and higher slope on average. As a result, this additional grouping improves the quantitative fitness of Model 8 (DIC = 345.6, Table 9) over EPTRICH Model 4 (DIC = 353.7, Table 9).

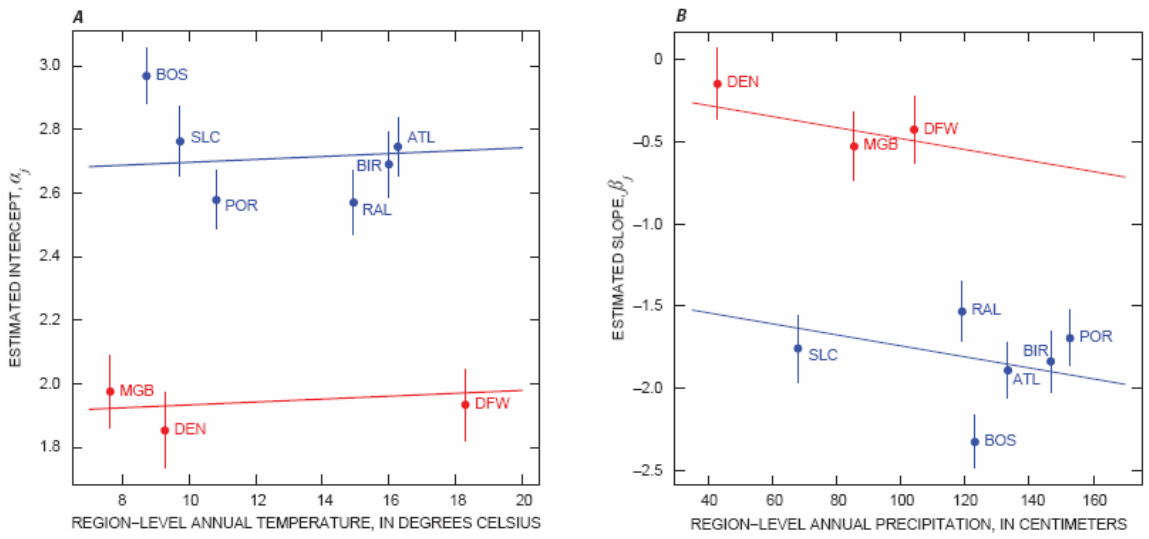

Figure 45: EPTRICH multilevel hierarchical Model 8: Region-level temperature and categorical antecedent agriculture (AG) predictors (red: high AG; blue: low AG) for (A) intercept and region-level precipitation and categorical antecedent agriculture predictors for (B) slope. Within each region, EPTRICH (combined richness of Ephemeroptera, Plecoptera, and Trichoptera orders) is modeled as a log-linear function of URB (percent urban land cover) as shown in Equations 18 and 19 (Template 3, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and categorical antecedent agriculture and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

Similar to NMDS1, RICHTOL results (Figure 46; Table 21) duplicate those of Model 7 for intercept and Model 6 for slope, with quantitative fit between those for

Models 6 and 7 (DIC = 404.1 relative to DIC = 417.8 for Model 6 and DIC = 410.4 for Model 7, Table 9). Direction of change in intercepts and slopes with temperature and precipitation was the same as in Model 4 (Figure 30), with intercept increasing with temperature (Figure 46*A*) and slope becoming less positive with precipitation (Figure 46*B*). Both high and low AG groups followed these trends with high AG regions tending to have higher tolerance and flatter slopes. As AG helped explain differences in intercepts and slopes, model fit improved over that of Model 4 (DIC = 412.7, Table 9).

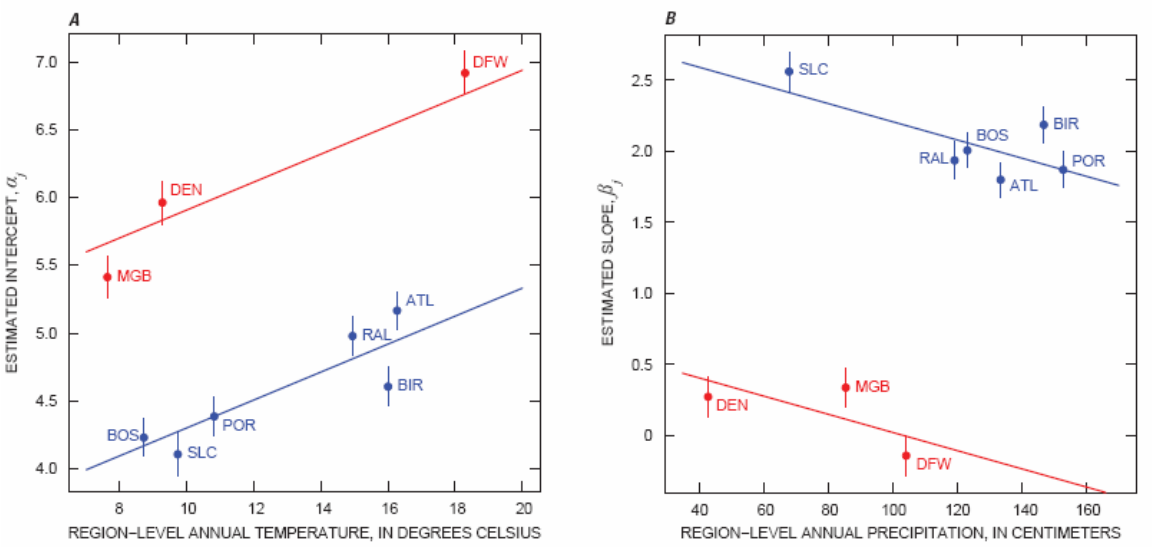

Figure 46: RICHTOL multilevel hierarchical Model 8: Region-level temperature and categorical antecedent agriculture (AG) predictors (red: high AG; blue: low AG) for (A) intercept and region-level precipitation and categorical antecedent agriculture predictors for (B) slope. Within each region, RICHTOL (richness-weighted mean tolerance of taxa at a basin) is modeled as a linear function of URB (percent urban land cover) as shown in Equation 12 (Template 3, basin level). Across regions, intercepts (A) are modeled as a function of regional temperature and categorical antecedent agriculture and slopes (B) are modeled as a function of regional precipitation and categorical antecedent agriculture as shown in Equations 13, 14, and 15 (Template 3, region level)

2.3.3. Model Interpretation

Precipitation, air temperature and antecedent agricultural were all found to affect the response of macroinvertebrates to urbanization in different ways.

2.3.3.1. Effect of Precipitation

NMDS1, RICH, and EPTRICH exhibit the same pattern of increasing intercept with precipitation. Logically, RICHTOL responds oppositely with decreasing intercept across precipitation. This means that, in general, ecological condition before urbanization was found to improve with precipitation. This makes sense ecologically because it shows that, at zero urbanization, regions with higher annual rainfall have greater baseline abundance and richness than regions with less rain (Cuffney et al. 2009a). That is, there is a greater amount and diversity of organisms in wet climates as opposed to dry climates. Since organisms require water to live, this result is well supported by biology. At zero urbanization, regions with an abundance of annual rain also have lower richness-weighted tolerance values than regions with less rain. This means, when there is an abundance of rain, less tolerant organisms can survive but when there is less rain, even at zero urbanization, organisms that are more tolerant (hardy) will thrive. In drier regions, baseline sites have more tolerant organisms without even considering the effect of urbanization. Additionally, precipitation is correlated with vegetation. In order for forests to exist, a certain amount of precipitation is needed. Regions with higher precipitation tend to be more heavily forested than regions with

lower precipitation, which tend to be dominated by grass- or shrublands. Forested regions are conducive to supporting a diverse assemblage of invertebrates because they moderate extremes in air and water temperatures, reduce extremes in streamflow, reduce erosion and sedimentation, support more diverse and stable habitats, incorporate fewer sources of chemical contamination, and provide more variable and sustainable food sources (Lenat and Crawford 1994, Richards and Host 1994, Roth et al. 1996, Rieman et al. 2001, Roy et al. 2005a, Horwitz et al. 2008). Therefore, higher precipitation almost always corresponded to higher adjusted ordination score, richness, and EPT richness and lower tolerance at zero urbanization.

The few exceptions to this pattern in intercepts occurred when categorical antecedent agriculture (AG) was added to the models, after which EPTRICH intercept no longer responded to precipitation, and RICHTOL intercept actually increased with increasing precipitation. Even though EPTRICH intercept increases with precipitation when modeled alone (Figure 21*A*), EPTRICH pre-urbanization condition is likely more affected by antecedent agriculture levels than by precipitation, thereby reducing the explanative capacity of precipitation once AG is included in the model (Figure 37*A*). Additionally, regions with high AG tend to have low precipitation, which may confound the two predictors. SLC, in contrast, has low AG and low precipitation. In fact, the unusually low intercept of this SLC outlier, combined with generally high model variability and low region sample size, may be causing the apparent reversal of the

RICHTOL intercept pattern with precipitation after the inclusion of AG (Figure 38*A*–Figure 42*A*).

NMDS1, RICH, and EPTRICH also exhibit the same pattern of decreasing (negative) slope with precipitation. RICHTOL again responds oppositely with increasing (positive) slope with precipitation (although only when intercept was also modeled with precipitation). This means that measures of ecological condition decline greater with increasing urbanization (have more negative slopes) in regions with more rain than in regions with less rain. That is, regions with an abundance of rainfall experience more drastic ecological disruption than dry regions for the same amount of urbanization. This result is likely because more rain ensures that the byproducts of urbanization (sediment, chemicals, pollutants) run off into the streams, therefore decreasing ecosystem quality more for the same amount of urbanization. Increased precipitation also exacerbates the effect of urbanization by way of changes in hydrology. When a region has an abundance of rain, urbanization can affect flow through less infiltration, less groundwater flow, faster transfer to stream channels through drains, higher and shorter duration high flows, lower and longer duration low flows, less available aquatic habitat, and numerous changes in channel shape and incision. In places where rainfall is less, the same amount of urbanization causes less ecosystem damage and, hence, has a (negative) slope of smaller absolute value with urbanization. Most slopes of change in tolerance with urbanization are positive (urbanization is related to

macroinvertebrate communities of higher tolerance) and increase as precipitation increases. This is for the same reason—in regions where there is an abundance of rain, tolerance increases faster with urbanization than in regions that have less rain.

When antecedent agriculture is included in the model (Figure 38*B*, Figure 46*B*) or when intercept is modeled with temperature instead of precipitation (Figure 30B), RICHTOL slope instead decreases with precipitation. This result could be due to the fact that the influence of antecedent agriculture is overriding the effect of precipitation. Even though AG is not explicitly included in the model with temperature-predicted intercept, it is clear that the regions with high AG form a separate group from the rest of the low AG regions, potentially obscuring a precipitation driver. A similar decreasing of precipitation influence can be seen for EPTRICH when the addition of an AG predictor flattens the relation between EPTRICH slope with urbanization and precipitation (Figure 37*B*) compared to the precipitation slope model without AG (Figure 21*B*).

2.3.3.2. Effect of Temperature

The relation of ecological condition with temperature is not clear from this modeling effort. NMDS1 and RICHTOL intercepts both increase with increasing temperature (although with significant variability) while RICH and EPTRICH intercepts do not vary systematically with temperature. When slope is predicted by precipitation alone or precipitation and antecedent agriculture, RICH and EPTRICH intercepts increase slightly with temperature (but, again, with variability). It appears that there

may be an optimal temperature range for some macroinvertebrate responses (regions with median average temperature values tend to have highest NMDS1 and EPTRICH intercepts) above and below which pre-urbanization ecological condition declines because it is either too hot or too cold to produce maximum community population health.

This final interpretation is supported by biology. Invertebrates are known to have thermal optima where their growth, survival, and fecundity are at their peak (Sweeney and Vannote 1978). However, there is too much variability in the temperature models to unequivocally decipher a pattern in these data. Contrary to original assumptions, model results show that precipitation more clearly and consistently explains differences in regional macroinvertebrate conditions prior to urbanization. Perhaps a linear relation is not the best model form to use to capture the effect of temperature. Using an annual temperature mean may capture seasonal temperature variation that could be driving ecological response differences, resulting in a time scale mismatch problem. Alternately, perhaps the perceived explanatory power of precipitation is spurred by its correlation with antecedent agriculture.

A consistent relation between temperature and rate of urbanization for the four ecological measures does not appear to exist. NMDS1 slope becomes more negative with increasing temperature, while RICH slopes becomes less negative and EPTRICH slope does not change with temperature. RICHTOL slope with urbanization becomes less

positive with temperature. This means warmer temperatures are increasing the effect of urbanization for NMDS1, decreasing the effect of urbanization for RICH and RICHTOL, and not affecting EPTRICH. These relations, again, are not clear and involve scatter. Qualitatively, it seems like precipitation also has a tighter relation with slope than temperature. It makes sense logically that precipitation would have more direct effects on urbanization than temperature. Urbanization in cold places is likely similar to urbanization in warm places, while urbanization in wet places is different from urbanization in dry places. When AG is included as a slope predictor with temperature for most measures (Figure 39*B*, Figure 41*B*, Figure 42*B*), it is clear that AG explains more than temperature does for all ecological responses except EPTRICH (Figure 40*B*).

2.3.3.3. Effect of Antecedent Agriculture

Ecological condition prior to urbanization is consistently poorer for regions with high antecedent agricultural activity than regions with low antecedent agriculture. NDMS1, RICH, and EPTRICH intercepts decrease with AG, and RICHTOL intercepts increase with AG when AG is modeled either continuously or categorically. This result is intuitive. As agricultural land was already disturbed from a natural state prior to urbanization, ecological communities living in past agricultural areas are more degraded than communities in past non-agricultural areas even at zero urbanization. If previous agricultural activity changed the macroinvertebrate community composition to include more pollutant-resistant species, ordination would be able to readily detect this shift in

assemblage structure and the shift would also likely be reflected in lower richness and EPT richness and higher richness-weighted tolerance.

When AG is modeled continuously, RICH intercepts have high variability for the low antecedent agriculture group (Figure 32*A*). This large range of intercept values for all regions having low AG likely contributes to the fact that both precipitation and temperature mask the effect of agriculture when RICH is modeled with both an environmental and AG influence (Figure 36*A* and Figure 40*A*). This masking does not occur for any other response variable. Also, low intercepts for DEN, DFW, and MGB match better with the precipitation pattern (all three high-agriculture regions happen to have lower precipitation values, so they are close to each other on the precipitation axis) than with the temperature pattern (MGB and DEN have low temperature and DFW has high temperature, so their similar, low intercepts are not near each other on the temperature axis). This distribution of precipitation values could possibly be making it look like precipitation is a better predictor than temperature, when really the important predictor is antecedent agriculture.

Similarly, rate of decline in ecological condition with urbanization is consistently greater for regions with low antecedent agriculture. That is, all macroinvertebrate measures show slopes closer to zero (less negative slopes for NMDS1, RICH, and EPTRICH, and less positive slopes for RICHTOL) for regions with high AG parameterized both continuously and categorically. Agricultural disturbance dampens

the effect of urban disturbance on macroinvertebrates. Since ecological communities in regions with a lot of previously converted agricultural land have already been disturbed, there is little further decline in macroinvertebrate response with urbanization compared to regions where urbanization starts with less disturbed land cover. This pattern is likely observed because the macroinvertebrates living in areas with a lot of antecedent agriculture have already been degraded by the effects of agriculture and are, therefore, less influenced by further human activity.

As with intercept, this pattern of slope change with AG does not appear to be as strong when ecological condition is measured using total taxa richness. Slopes for POR and BIR are estimated as low (close to zero) as average slopes for the high AG group (Figure 32*B*) and, therefore, a categorical AG predictor is not able to distinguish differences in overall slope trend between regions with high and low AG when precipitation and temperature predictors are included (Figure 36*B* and Figure 40*B*). Even when the slope dampening effect of high agriculture is observed when temperature predicts intercept and precipitation predicts slope (Figure 44*B*), slopes of POR and BIR continue to overlap with those of DEN, DFW, and MGB. Past agricultural activity may not affect total richness as much as it affects abundance-based ordination scores, EPTRICH and RICHTOL.

2.4. Conclusions

A multilevel hierarchical modeling approach is an appropriate means of describing the multiple tiers of EUSE data statistically and ecologically because it allows discernment of relations between responses and predictors at multiple scales. The connections between urbanization and ecosystem status at the basin level are clearer if modeled in a framework that accounts for region-level effects, which change the nature of the relation between URB and macroinvertebrate community response in different regions. In this way, the understanding of both the basin- and region-scale effects of urbanization on aquatic macroinvertebrate species abundance can be increased, and previously unexplained regional differences in response to urbanization can be quantified.

2.4.1. General Ecological Trends

Observed changes in intercepts and slopes with region-level predictors are ecologically interpretable. Intercepts represent macroinvertebrate condition at zero urbanization, in other words, before urbanization and its effects take place. In general, this baseline macroinvertebrate condition prior to urbanization is found to be better across all four metrics in locations with high rainfall and low previous agriculture. That is, regions with high cumulative annual precipitation and low antecedent agriculture exhibited invertebrate metric values generally considered indicators of healthy streams such as high diversity (RICH), high amounts of sensitive taxa (EPTRICH and NMDS1

scaled to EPTRICH), and low richness-weighted mean tolerance (RICHTOL). These same high rain and low antecedent agriculture regions were also associated with faster decline of macroinvertebrate communities with urbanization, as measured by steeper slopes of change in invertebrate metrics for the same amount of increasing urbanization.

Temperature alone has no consistent effect on macroinvertebrate pre-urban condition or on rate of response to urbanization. However, when modeled with antecedent agriculture, cold regions show better baseline conditions prior to urbanization but faster decline with urbanization when combined with low previous agriculture. These multilevel hierarchical models demonstrate that the effects of urbanization are influenced by both natural (rain, temperature) and human factors (antecedent agriculture). To effectively quantify the effects of these different factors, it is important to model their influences at the appropriate scales, as well as to incorporate the simultaneous influences of multiple factors.

2.4.2. Utility of Modeling Methodology

The multilevel hierarchical model structure is a natural framework for analyzing data with hierarchical arrangement. As part of the EUSE study, such data were collected by measuring variables across basins within a region as well as higher tier variables across regions. Applying this modeling methodology to sets of variables measured at different levels located in nested tiers creates better understanding of the relation between predictors and responses at multiple scales. As opposed to calculating basin-

level coefficients separately for each region, modeling these data hierarchically decreases model uncertainty by borrowing strength from the entire dataset (rather than just one region at a time) when calculating region specific model coefficients. Alternately, using a multilevel hierarchical model is more statistically efficient than attempting to account for regional differences non-hierarchically using dummy variables. Once multilevel hierarchical models are fit to the hierarchical EUSE dataset, they provide a means of making predictions about ecological effects for regions where there are region-level predictor measures but no basin-level measures. This type of inference is accomplished by predicting basin-level intercept and slope estimates for a new region from the higher tier inter-region coefficients and error terms. Model users are unable to make such predictions with non-multilevel individual region models, which require availability of data at the basin level in order to establish regression relations.

The use of multilevel hierarchical modeling also requires the application of the exchangeability assumption in two different ways, depending on the lack or contribution of specific group-level predictors. Without any information to distinguish regions at the group level, the hierarchical hyperparameter distribution becomes the equivalent of a noninformative prior distribution (Gelman and Hill 2006, Qian 2010). In this case, exchangeability is assumed between regional parameters in the sense that each is modeled as a random draw from the single higher tier bivariate normal distribution of possible intercepts and slopes centered on a constant mean. Without prior information

about differences between regions, it is assumed that differences result only from normal random variability and that regions are exchangeable. However, as soon as differences between regions can be explained by differences in a regional level variable, this model's prior hyperparameter distribution now has much more information in it and can now be described by the equivalent of an informed prior distribution. In this case, regions are no longer exchangeable in that differences between them are now meaningful and not random. Prior knowledge about how regional differences are expected to affect intercept and slope can be incorporated in the prior hyperparameter distribution. But because regional intercept and slope can not be predicted exactly, exchangeability now applies to hyperparameter distribution error. While the expected mean regional intercept and slope can be predicted from the value of a regional variable, the variability of those estimates around the hyperparameter regional regression prediction lines is random and exchangeable. Using these different kinds of exchangeability assumptions allows flexibility in the multilevel hierarchical modeling structure to represent as much prior information as is available in each modeling scenario, distinguishing between that which can be explained and that which is randomly exchangeable.

2.4.3. Measures of Model Fit

It is difficult to establish a quantitative measure of model fit for multilevel hierarchical models with varying parameters and forms. Deviance Information Criterion

(DIC) attempts to quantify the balance between predictive benefits of additional variables and drawbacks of overparameterization, but DIC values between different model structures are not comparable. Additionally, DIC does not capture the scientific validity or plausibility of a model, and certain statistical principles may influence the relative magnitude of DIC (Spiegelhalter et al. 2002). For example, of the eight models tested, the model with the smallest DIC for all response variables is Model 5, the model that uses a continuous measure of antecedent agriculture percentage. However, it is not clear that Model 5 offers the best fit to the data. The reason quantitative fit of Model 5 is evaluated so highly is because antecedent agriculture of the nine regions clusters around two groups so that, at the region level, linear regressions are essentially fit to two points, resulting in low estimates of variance. In terms of interpretability, Model 5 is not necessarily the most useful model because there is no sample information for agriculture values between the two groups, and it is unclear whether this tight linear relation exists across the entire range of possible antecedent agriculture percentages.

2.4.4. Variable Limitations

Despite producing reasonable models that link urbanization effects to macroinvertebrate response, the use of RICHTOL as a measure to represent the ecosystem condition has two main limitations. The first being that RICHTOL is not a truly continuous variable as it is defined to vary only between 0 and 10. In theory, this limitation can be fixed through adopting the logit transformation on the variable.

However, in this case, the logit transformation does not help since most values are centered in the middle of the range and as such the transformation complicates the interpretability of the response variable while slightly increasing the spread of the data. Possible limitations of RICH and EPTRICH include over-summarization of ecological condition and difficulty of communicating Poisson (count based) regression models and results.

Variables that condense and combine a large amount of information into a few measures (for example, NMDS1 or NUII) may not be the most appropriate to use, both in modeling patterns and in explaining or understanding what variables mean on their own and in relation to other variables. Adjusted ordination score (NMDS1) does not unambiguously represent macroinvertebrate assemblage data. Ordination is a multivariate method that is designed to summarize the relations among sites based on the similarity (or dissimilarity) among the assemblages. Ordination is a data reduction technique that reduces the dimensionality of the data to typically two or three dimensions while retaining as much of the structure in the original multidimensional data as possible. Conceptually, the dimensions of the ordination (ordination axes) can be associated with derived environmental gradients that capture the major axes of variability (that is, assemblage change) in the datasets. Sites are located along these axes such that sites with similar assemblages are located close together and dissimilar sites are located far apart. Consequently, the relative position of sites along the ordination

axes has ecological meaning in that it represents degree of ecological community difference. The ordination axis explaining this difference can then be attributed to an environmental change such as urbanization. The relative distances among sites can then be interpreted as responses to urbanization by plotting ordination scores against an explanatory variable such as urban intensity. Ordination analysis would appear to be an ideal method of analyzing responses along an urbanization gradient since these methods are designed to detect assemblage changes along potential explanatory gradients. However, there are a number of problems associated with ordination that make it difficult to use ordination results in multilevel analyses and to communicate ordination results to water-resource managers. Despite producing models with interpretable effects on macroinvertebrate abundance, it is not clear that NMDS1 is the best measure to represent ecosystem condition.

The calculation of ordination score obfuscates the details of actual ecological condition found in the original data. Complex matrix manipulation does not identify which basins are grouped in which ways. For basins with similar NMDS1 values, it is not clear which species these basins have in common nor is it clear exactly how similarities change for basins with dissimilar NMDS1 values. For example, what if two basins have different amounts of different species (dissimilar macroinvertebrate assemblages) but both basins represent a poor ecological condition? It is possible that the two basins would be far apart in ordination space yet they each represent a negative

effect of urbanization. Unadjusted ordination scores are directionless prior to chosen interpretation; they are only a measure of relative assemblage change. Ordination axes are only latent variables that account for variability in the original data. Counter to intuition, increasing or decreasing ordination score values have no absolute meaning until assigned by adjustment relative to another variable. But, how can an adjusted unidirectional continuous axis then be used to describe ecological response if there may be multiple but different ways to be ecologically unhealthy?

Ordination analysis also requires cumbersome recalculation of all values every time a basin is added to or removed from a dataset, making the measure impractical and inconsistent within regions. Removal of one data point should not have an effect on the response value of a different data point. While NMDS1 site scores are related to relative measures of assemblage similarity the actual units of measure are not clearly defined, consequently it is not known whether they follow a linear, logarithmic, or other kind of systematic scale. That is, there is no simple definition of what an NMDS1 value means. This ambiguity makes changes in NMDS1 difficult to communicate to both scientists and environmental managers.

Perhaps of greatest relevance to multilevel modeling is the question of whether ordination scores derived independently for different metropolitan regions are ecologically comparable among metropolitan regions. That is, does a change from 0.5 to 1.5 units on the ordination axis in BOS have the same meaning for POR? The consistency

of variable definition between regions is important for multilevel modeling because the logic of complete pooling relies on it. The interpretation reliability between regions is complicated by the rescaling required to place ordination in a consistent ecological context. For example, in this study, the ordinations were rescaled so that the first site axis scores decreased with increasing urbanization and ranged from a maximum value at minimum urbanization to zero at maximum urbanization. This rescaling eliminated negative numbers and preserved the range of the original ordination scores and the distances among sites. Rescaling the axis scores in this manner facilitates comparability among metropolitan regions by preserving the relative distances among sites and the differences in the range of ordination scores among metropolitan regions. However, an NMDS1 of 2 in BOS still does not have the exact same meaning in terms of macroinvertebrate assemblage status as an NMDS1 of 2 in POR. Hence comparing the two regions on the same scale of NMDS1 values is not entirely valid. This inconsistency presents theoretical objections to pooling NMDS1 values between regions as part of the hierarchical modeling calculations as the actual numbers mean different things in different regions.

The comparability of NMDS1 multilevel response models to models based on more straight forward assemblage response metrics (EPTRICH, RICHTOL) provides empirical evidence that the independently derived ordination scores captured regional scale differences in the change in invertebrate assemblages across the urbanization

gradient. While the ordinations provided ecologically meaningful results that incorporate data from the entire assemblage, the complexity of calculation, theoretical considerations, and the difficulty of explaining what ordination scores represent make this a less attractive response variable than metrics that emphasize only a portion of the assemblage (EPTRICH) or only one aspect of the assemblage (RICHTOL). Ordination is a valuable tool for analyzing assemblages, and it is only through the use of these more powerful tools that confidence can be gained in the presentation of the more simple metrics. But for the reasons listed above, directly measurable total richness, EPT richness, and richness-weighted tolerance may be better metrics to use to quantify stream ecosystem macroinvertebrate assemblages. Although NMDS has data summarizing advantages that metrics lack, the data analysis tools must fit the analysis methods, and it is not clear that ordination scores are an appropriate response variable to use in inter-region hierarchical modeling.

A similar problem with information condensation exists for the National Urban Intensity Index (NUII) predictor metric. The NUII predictor is an artificially scaled amalgamation of multiple covarying variables that have been scaled to account for regional differences in the rates at which the component variables (housing density, road density, percentage developed land) change across metropolitan regions. A benefit of combining covarying variables from different data sources is to prevent possible errors in one data source from biasing the whole index. Multimetric measures such as

NUII are commonly used in bioassessment for this reason. However, there is no statistical strength in combining multiple covarying variables because no new information is gained. Additionally, because NUII combines many different measures into one score, it is hard to define explicitly what that score represents. There is no definition, for example, for what a NUII of 30 means. Similar to NMDS1, NUII is calculated and calibrated per dataset so a NUII value cannot be measured on its own. Instead, a NUII value is dependent on the range of data in the dataset, and the entire set of NUII values has to be recalculated for every new data point. Therefore, NUII is not a clear, consistent, and unchanging measurement. Additionally, from a multilevel modeling standpoint, NUII already attempts to incorporate regional differences, though in a more primitive manner than multilevel modeling. Because of this, using NUII as a predictor in multilevel models may, in fact, confound regional and basin influences. It is statistically and intuitively more appropriate to directly account for regional differences at a higher tier in the multilevel models than to indirectly attempt to do so using NUII as a lower tier predictor. For these reasons, percentage of urban land cover was chosen over NUII as the basin-level predictor variable in this analysis.

Metrics should be simple, consistent, measurable, and usable. Metrics should be easy to obtain and easy for mangers to understand them in environmental models and decisions. Ultimately, neither NMDS1 nor NUII fulfill these criteria.

2.4.5. Future Directions

The next immediate model building step in this research is the implementation of WinBUGS fully Bayesian methodology to replace lmer convergence problems, incorporate prior information, and report distributions of model coefficients rather than just point estimates. Once methodology is upgraded, more specific and management orientated model drivers can be developed by incorporating other basin-level and region-level predictor variables, including alternate urbanization indicators resulting from the decomposition of NUII. This modeling approach can also be expanded to evaluate the effect of urbanization on algae and fish assemblages. Using region-level coefficients, ecological effects can be predicted in regions for which only region-level predictor measures exist and compared to future real world measurements (for example, datasets to be collected in Chicago, Anchorage, and Seattle). Development and implementation of quantitative model evaluation and verification criteria are essential in order to be able to judge the quality of results.

3. Combining Expert Knowledge with Data to Develop a Bayesian Network Describing the Impact of Urbanization on Aquatic Macroinvertebrates in the Southeast U.S.

3.1. Problem construct

The effects of urbanization on aquatic stream biota are complex, multilayered, and interacting, covering factors operating over multiple spatiotemporal scales. Modeling these effects presents many unique challenges. First and foremost, is the question of how to represent this complicated system in a manner that appropriately captures the major interacting dynamics as well as the uncertainty resulting from imperfect knowledge, imperfect data, or even irreducible environmental variability. Accounting for the scope of these relationships and all their associated uncertainty is currently mechanistically impossible. This detailed level of scientific understanding simply does not exist. Even if such a mechanistic model could be built, it would be impossible to verify due to its complexity. To represent known general associations between variables while acknowledging uncertainty in the specifics, a statistical approach must be utilized. However, depending on the number of variables and probabilistic connections between them, directly parameterizing such a complicated model statistically could be incredibly challenging in a different sense. A statistical representation requires developing a joint probability distribution which simultaneously characterizes the potential behavior of each variable as a function of all other variables. This task becomes exponentially more difficult with every additional variable, quickly

becoming intractably large for relatively few variables. The parameterization of this joint distribution remains a significant modeling challenge. Finally, in addition to independent datasets collected, we know a vast store of integrated expert knowledge exists about the processes of urbanization and their impacts on hydrology, habitat, water quality and biology. Ideally, both these sources of information (data and expert knowledge) should be included, appropriately weighted, in a final system model.

Previous modeling efforts aimed at understanding urbanization effects have not addressed any of these formidable challenges. Traditional analysis of stream ecosystem data attempts to find empirical relationships only between single pairs of environmental concepts at a time using simple regression techniques (Cuffney et al. 2009a), without incorporating the web of additional interconnected environmental variables, uncertainty characterization, or known ecological information about the system. In contrast, we transform ecosystem-urbanization modeling by employing a Bayesian network approach. This framework will allow us to concisely represent and parameterize the system of complicated cause and effect relationships and uncertainties, calculate the full joint distribution efficiently as the product of more manageable conditional probabilities, and incorporate both data and expert knowledge.

Relative to traditional frequentist modeling constructs, this approach is enticing to aquatic ecologists for its ability to model the entire set of major system components that are interacting at different levels, rather than just limited subsets of components at a

time. Using a Bayesian network to describe the effect of urbanization on invertebrate condition acknowledges the complexity of the environmental and ecological processes driving biological response and allows concurrent assessment of multiple driving factors.

3.1.1. Purpose and scope

The purpose of this Chapter is to present a methodology for developing a Bayesian network and report the results of a case study, using this methodology, to examine the effects of urbanization on aquatic invertebrates in the Southeast United States. The two main objectives are to advance both scientific understanding of the system of relationships describing environmental and ecological processes and methodological understanding of the benefits and drawbacks, successes and pitfalls of developing a Bayesian network model. The scope of this Bayesian network creation process involves using expert knowledge to develop prior distributions for relations between compartments in the model, updating with 85 data points specific to the Southeast United States, and calculation and evaluation of the resulting posterior model in itself and compared to expert knowledge and data inputs.

Multilevel hierarchical modeling analysis of all 270 streams in the 9 total metropolitan areas showed that baseline, pre-urbanization conditions and rates of response to urbanization differ among ecoregions (Kashuba et al. 2010). Therefore, in this seminal Bayesian network construction, we chose to focus efforts on model

development specific to the Southeast United States. We used EUSE data collected at 85 watersheds in the Atlanta, Birmingham, and Raleigh EUSE study areas, as well as the professional expertise of a USGS research ecologist who has worked in the Southeast United States for nearly 30 years.

3.2. Methods

The bulk of this research effort involves the development of a methodology process for use in urbanization system Bayesian network modeling. These methods can be applied to a variety of simple and complex questions concerned with the incorporation of expert knowledge, data, and uncertainty characterization.

3.2.1. Bayesian networks

Uncertainty can result from many sources. It may be a consequence of random or systematic measurement error which results in a discrepancy between a true value and collected datum. Inaccuracies in assumptions may increase uncertainty in cases where an assumed relationship between variables may not be exactly correct. A decision to measure a surrogate of a variable of interest may not capture all the true characteristics of the desired variable. In environmental field studies, there is often an inability to control all variables engendering uncertainties associated with observational data. The cost and time consumption of data collection may lead to a lack of empirical basis for making a true characterization. Due to these and other limitations or circumstances, we may have imperfect knowledge about how the system works, in

general. But even in the best scenarios of knowledge and measurement, simple random variability is associated with natural, environmental field systems at multiple scales. This type of uncertainty is, by definition, irreducible as it is a characteristic of populations (Morgan and Henrion 1990).

We know we are building models in the context of all this uncertainty. And to build the most useful characterization of the system of interest, we must account for it. To do so requires a probabilistic model framework where we describe variables with distributions and systems of related variables with joint distributions. But this gets very complicated very fast.

By the chain rule of probability, a joint probability can be specified as a product of conditional probabilities with each subsequently described variable conditioned on the remaining variables. For example, if you were interested in the joint distribution of variables A, B, C, and D, by the probability chain rule:

$$p(A,B,C,D) = p(D|C,B,A)p(C|B,A)p(B|A)p(A) \quad (32)$$

where

$p(A,B,C,D)$ is the joint distribution of variables A, B, C, and D,

$p(D|C,B,A)$ is the conditional distribution of variable D given variables C, B, and A,

$p(C|B,A)$ is the conditional distribution of variable C given variables B and A,

$p(B|A)$ is the conditional distribution of variable B given variable A, and

$p(A)$ is the marginal distribution of variable A.

As the number of variables increases, this specification becomes intractably difficult to define. A Bayesian network is essentially a compact way of storing all of the probabilistic information about a joint distribution of multiple related variables (Heckerman 1999, Spirtes et al. 2000, Neapolitan 2004, Taroni et al. 2006, Jensen and Nielsen 2007, Kjaerulff and Madsen 2008, Pourret et al. 2008). It takes the form of a directed acyclic graph (DAG), which is a set of items (or 'nodes') connected by ordered edges (or 'arrows') where no path following the arrows goes through the same node twice; i.e., there are no cycles in the model. In a Bayesian network DAG, nodes represent important concepts (variables) in the system of interest and arrows represent probabilistic relations between those concepts. By defining a system in this way, we greatly reduce the difficulty of parameterizing a model of high dimensionality without losing information in the joint distribution.

In contrast to using the non-graphical probability chain rule (Equation 32), a Bayesian network (or Bayes net) allows us to decompose this complicated joint distribution of all variables into the product of more easily specified conditional probabilities *between connected nodes only*. This specification greatly simplifies representation of the entire system of variables and allows for more intuitive parameterization only for directly related variables. Again, for a system with variables A, B, C, D, if you know that B is conditional only on A, C is conditional on only A, and D is conditional on only B and C, then, instead of using the chain rule as above, you can

describe the probabilistic relationship between A,B,C,D with a Bayes net where A points to B and C, B points to D, and C points to D (Figure 47). Then,

$$p(A,B,C,D) = p(D|C,B)\, p(C|A) p(B|A)\, p(A) \quad (33)$$

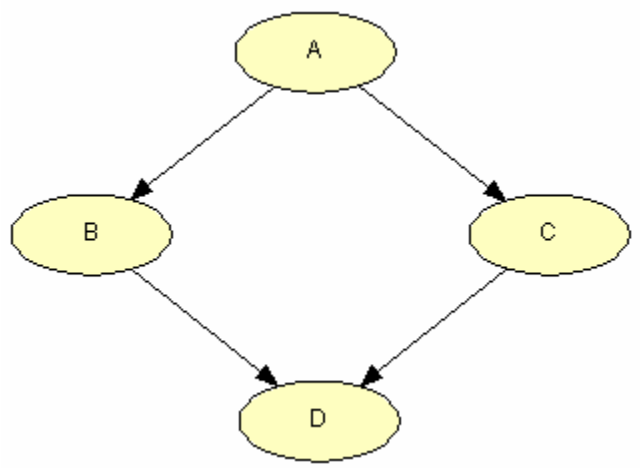

Figure 47: Simple explanatory Bayesian network.

Because of the conditional independencies defined by this Bayes net, now

$$p(D|C,B,A) = p(D|C,B), \text{ and} \quad (34)$$

$$p(C|B,A) = p(C|A) \quad (35)$$

This simplification makes the problem easier to solve, because it is easier to specify distributions for variables that are dependent on few other variables than for variables which are dependent on many other variables. This is especially true as the number of variables in a network increases. It is also intuitively easier to specify relationships between directly related variables than between indirectly related variables (such as D and A). Using the Bayesian network, the probability of D given C and B is now independent of A. For example, we could use a Bayesian network to describing the

relationships between urbanization, physical environmental properties, chemical environmental properties and biological ecological properties. Rather than attempt to derive a four variable joint probability function, we could instead recognize that physical properties and chemical properties are each conditional on urbanization, and that biological properties are conditional on both physical and chemical properties. This system of relationships can then be described by the Bayesian network introduced above where A is urbanization, B is physical environmental condition, C is chemical environmental condition, and D is biological ecologic condition (Figure 48). Now, instead of having to specify the indirect relationship between biological condition (D) and urbanization (A), we can specify the system in terms of more direct relationships.

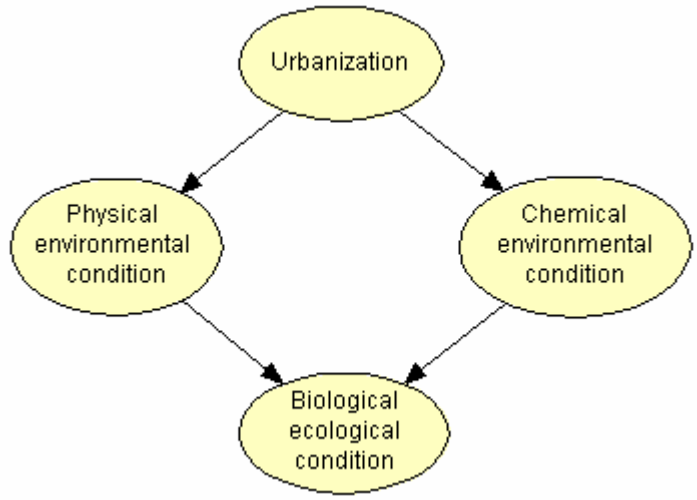

Figure 48: Simple explanatory Bayesian network example.

This modeling framework allows decomposition of a complicated system into sets of conditional probabilities describing a distribution for each separate child node (a node with arrows pointing into it) given values of its parent nodes (nodes with arrows

pointing from them). Child nodes are then themselves modeled as parent nodes of subsequent child nodes, in this way eventually linking the entire system of interest into one probabilistic network. Then, the multiplication of all the separately parameterized conditional probabilities results in the full joint distribution on all variables without having to calculate the joint distribution directly. The final network describes the interrelationships between all variables in the system simultaneously such that we can know the effect of changing one variable on all the remaining variables.

The concept of a Bayesian network derives from the principles of Bayes Theorem:

$$p(\theta \mid x) \propto p(\theta)\, p(x \mid \theta) \tag{36}$$

where

$p(\theta)$ is the prior probability distribution on θ, the parameter of interest,

$p(x \mid \theta)$ is the likelihood function characterizing x, the data, given θ, and

$p(\theta \mid x)$ is the posterior probability distribution on θ, which is a combination of the prior information and the information in the data collected.

This seminal theorem quantitatively describes the process of using new information ('data') to refine a previously held belief ('prior') resulting in an updated belief ('posterior'). There are two ways a Bayesian network can use Bayes Theorem to revise a prior belief based on additional information. The first is simple evidence propagation in a compiled network. This type of prior revision is also called predictive or diagnostic probability calculation (Jensen and Nielsen 2007) and is essentially a form

of conditional inference making given the already constructed and parameterized model. It uses a discrete form of Bayes theorem where θ represents a single probability value (eg., $p(\theta) = 0.25$) instead of a probability distribution (eg., $p(\theta) \sim$ Beta(1, 3)). Evidence propagation is similar, conceptually, to fitting a regression equation to a dataset resulting in parameter estimates (slope and intercept) that describe the relationship between x and y, and then using that model to predict y given evidence of a particular x value.

In the Bayesian network context, the probability of a certain node, A, being in a particular state, a, is considered a prior probability, i.e., $p(\theta) = p(A=a)$. Then, suppose evidence is gathered about the state of another node, D, in the network. D is known to be in state d. This evidence can be considered 'data' separate from the information available in the prior. Given the existing sets of linked conditional probabilities that make up the network, we can then calculate the posterior probability for node A being in state a given that node D is in state d, i.e., $p(A=a|D=d)$. This calculation is making inference on the marginal distribution of node A given a particular state of D using the static probabilistic relationships between the nodes (the Bayesian network encoded conditional probabilities), in the same way that one can make inference on the value of y given a particular state of x and a slope and an intercept. Using Bayes Theorem (and graph theory), we revise what we predict about a node based on hypothetical 'evidence' on another node. We do not, however, change the model itself during evidence

propagation. We simply use information in the model for prediction. The use of Bayes Theorem in this context is relegated to its definitions of basic statistical relationships between discrete conditional and marginal probabilities:

$$p(\theta \mid x) = \frac{p(\theta)\, p(x \mid \theta)}{p(x)} \tag{37}$$

In contrast, the continuous version of Bayes Theorem (Equation 36) can be invoked in a more philosophically Bayesian manner to update the model itself by updating the network's conditional probabilities. The linear regression corollary of this type of updating would be if additional data measuring both x and y were collected and updated slope and intercept parameter estimates were calculated using both the original data and the new data. In the Bayesian network situation, now the parameter θ refers not to a single marginal node probability, as it does in evidence propagation, but to a probability distribution describing a relationship between parent and child nodes. This updating of a prior conditional probability distribution with measured data is what is unequivocally accepted as 'true' Bayesian updating, as it is incorporating new information which changes the model itself instead of just using the given model for inference making as in evidence propagation. This Bayesian network updating is also referred to as adaptation (Hugin 2008).

In truly Bayesian updating, a probability distribution form is specified that represents prior knowledge about a conditional relationship between nodes. Then, data are collected about these same nodes to inform this conditional relationship. Using

Bayes Theorem (Equation 36), the prior probability distribution is multiplied by the likelihood function from the data (and normalized, which is not shown in Equation 36 for explanation simplification purposes). This calculation results in a posterior probability distribution which actually changes the Bayesian network conditional probabilities. Relative to evidence propagation, where we use what we already know about a relationship to make predictions, in Bayesian updating, we update what we know about a relationship between nodes based on measured data of that relationship.

In practice, the use of discrete Bayesian network software enables the quick and efficient calculation of both evidence propagation and Bayesian updating. Using the statistical principles described above, when hypothetical evidence is entered on any node, Bayesian network software automatically calculates inferred marginals for all nodes in the network. Similarly, when actual measured data is incorporated into the network, the entire series of linked conditional probability distributions are updated. These calculational efficiencies, in addition to intuitive and interactive graphical representation, are what contribute to the much greater user friendliness of a discrete Bayesian network model over continuous Bayesian modeling constructs.

If a Bayesian model is not constrained to use only discrete versions of data, both solutions and their visual representations are greatly complicated. In continuous form, Bayesian updated probability distributions often have no analytical solution and the use of Markov Chain Monte Carlo (MCMC) sampling is required to calculate an

approximate numerical solution (Lunn et al. 2000). This entire time-intensive and computer-processing heavy procedure must be repeated for every separate evidence propagation or conditional probability updating scenario. Also, for every scenario, there is no clear or instantaneous visual representation of the updated system of marginal or conditional distributions. (Representation might require four-dimensional continuous graph slices.) Even if a continuous visualization were created, it would not be directly manipulatable by the user as, again, it would require complete system re-sampling to incorporate any change.

Calculations are immensely simplified if a Bayesian network is parameterized in discrete form (dividing a variable into categories instead of continuous values). This is because, in discrete form, we can guarantee the use of parameter and data distributions that are conjugate and, therefore, simple analytical solutions are available. Conjugate distributions are defined as pairs of distribution forms where the prior and posterior parameter distributions are of the same functional form, given a particular data distribution (likelihood function). This construct requires selecting an appropriate combination of prior distribution and likelihood function such that when they are multiplied together per Bayes theorem (Equation 36), the product enables a posterior distribution to be expressed in the same form as the prior. When a conjugate distribution family is chosen and the form of the posterior coefficients are established from the conjugate prior and likelihood coefficients, it becomes very simple to calculate

posterior distribution coefficient values given prior distribution coefficient values and the data. This is not the case in many continuous Bayesian models where the posterior distribution often has a different functional form than the prior distribution. However, in a discrete Bayesian network, we can guarantee conjugacy by using the Dirichlet-multinomial conjugate distribution family.

If each node is divided into two or more mutually exclusive states, the probability that a sample will fall in each of those discrete, non-overlapping bins is logically described using a multinomial data distribution. The multinomial likelihood function is simply an extension of a binomial distribution to situations where there are more than two possible discrete states of a sample (e.g., low/medium/high rather than true/false). Unlike the binomial distribution, which is described by only one parameter, θ (since the probability of a sample falling in the other bins is 1- θ), the multinomial distribution is described by a number of parameters equal to one less than the number of bins. For example, if data could fall in one of four bins then the multinomial distribution would have to include parameters θ_1, θ_2, and θ_3 (since the probability that a sample will fall in the fourth bin is really just 1 - (θ_1+θ_2+θ_3). For interpretability, sometimes a parameter θ_4 is included in a four bin multinomial distribution specification, as it is easier to think symmetrically of four parameters for four bins.

The likelihood function for a four bin multinomial distribution is:

$$p(x_1, x_2, x_3, x_4 \mid \theta_1, \theta_2, \theta_3, \theta_4) \propto \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3} \theta_4^{x_4} \tag{38}$$

$$\sum_{i=1}^{4} x_i = n \tag{39}$$

where

$x_1 - x_4$ are the counts of samples belonging to each of four bins, respectively,

n is the sum of all counts (i.e., total sample size), and

$\theta_1 - \theta_4$ are the probabilities of a sample belonging to bins 1-4, respectively.

This four parameter likelihood function defines the likelihood of observing discrete data, x_1, x_2, x_3, and x_4 as a function of parameters θ_1, θ_2, θ_3, and θ_4. These data measure how many of n total samples fell in bins 1, 2, 3, and 4, respectively. For example, there could be $n = 10$ streams sampled of which $x_1 = 2$ have low conductivity (less than 92 μs; bin 1), $x_2 = 4$ have medium-low conductivity (between 92 and 129 μs; bin 2), $x_3 = 3$ have medium-high conductivity (between 129 and 281 μs; bin 3), and $x_4 = 1$ have high conductivity (greater than 281 μs; bin 4). Parameters θ_1, θ_2, θ_3, and θ_4 represent the probabilities that a random stream sample would fall into bins 1, 2, 3, or 4, respectively. Using the multinomial likelihood function, one could calculate the likelihood of seeing those data given different values of these parameters. Maximizing this likelihood function is often used as a parameter estimation technique in frequentist statistics. Maximum likelihood parameter values are calculated by setting the first derivative of the likelihood function to zero and solving for the parameter value that maximizes the likelihood of seeing this particular dataset. Often the logarithm of the

likelihood function is used to simplify calculations; this transformation does not affect the maximum likelihood since the logarithm function is monotonic. For a multinomial likelihood function in four parameters, the maximum likelihood estimate for the parameters, θ_1, θ_2, θ_3, and θ_4 can be calculated as follows:

$$\frac{\partial(\log(p(\theta_1,\theta_2,\theta_3,\theta_4 \mid x_1,x_2,x_3,x_4)))}{\partial\theta_i} = \frac{\partial(\log(\theta_1^{x_1}\theta_2^{x_2}\theta_3^{x_3}\theta_4^{x_4}))}{\partial\theta_i} = 0 \tag{40}$$

Rearranging the logarithmic exponents and substituting 1 - θ_1 - θ_2 - θ_3 for θ_4:

$$\frac{\partial(x_1\log\theta_1 + x_2\log\theta_2 + x_3\log\theta_3 + (n-x_1-x_2-x_3)\log(1-\theta_1-\theta_2-\theta_3)}{\partial\theta_i} = 0 \tag{41}$$

Taking the derivative (Equation 41) with respect to θ_1:

$$\frac{x_1}{\theta_1} + 0 + 0 + \frac{-(n-x_1-x_2-x_3)}{1-\theta_1-\theta_2-\theta_3} = 0 \tag{42}$$

Taking the derivative (Equation 41) with respect to θ_2:

$$0 + \frac{x_2}{\theta_2} + 0 + \frac{-(n-x_1-x_2-x_3)}{1-\theta_1-\theta_2-\theta_3} = 0 \tag{43}$$

Taking the derivative (Equation 41) with respect to θ_3:

$$0 + 0 + \frac{x_3}{\theta_3} + \frac{-(n-x_1-x_2-x_3)}{1-\theta_1-\theta_2-\theta_3} = 0 \tag{44}$$

Solving the system of three equations (Equations 42, 43, and 44) for three unknowns θ_1, θ_2, θ_3 results in the following maximum likelihood estimates (MLEs):

$$\hat{\theta}_1 = \frac{x_1}{n} \tag{45}$$

$$\hat{\theta}_2 = \frac{x_2}{n} \tag{46}$$

$$\hat{\theta}_3 = \frac{x_3}{n} \tag{47}$$

By definition,

$$\hat{\theta}_4 = 1 - \hat{\theta}_1 - \hat{\theta}_2 - \hat{\theta}_3 \tag{48}$$

and

$$x_4 = n - x_1 - x_2 - x_3 \tag{49}$$

Therefore,

$$\hat{\theta}_4 = 1 - \hat{\theta}_1 - \hat{\theta}_2 - \hat{\theta}_3 = 1 - \frac{x_1}{n} - \frac{x_2}{n} - \frac{x_3}{n} = \frac{n - x_1 - x_2 - x_3}{n} = \frac{x_4}{n} \tag{50}$$

In summary notation, for all multinomial distributions with *i* bins:

$$\hat{\theta}_i = \frac{x_i}{n} \tag{51}$$

where $\hat{\theta}_i$ is the maximum likelihood estimate of θ_i, the probability of a sample belonging to bin *i*.

This same maximum likelihood estimate can also be calculated in matrix form using a Lagrange multiplier. However, because, in the frequentist framework, data are considered random and parameters are considered fixed, inference from such a maximum likelihood analysis is not intuitive. Confidence intervals specify that the fixed but unknown parameter of interest lies within a particular reported range of values a

certain percent (most commonly, 95%) of the times a random sample is drawn from the true population.

In contrast, Bayesian analysis treats parameters themselves as random values and, therefore, Bayesian results are reported directly as distributions of the parameter of interest. Both the prior distribution and the posterior distribution describe how the parameters θ_1, θ_2, θ_3, and θ_4 are distributed. Before data are collected and analyzed, prior information about the probability of a sample falling into each of the four bins may be available from other studies or from expert knowledge. If this prior information is incorporated via a Dirichlet distribution, then the posterior distribution calculated by multiplying this prior by the multinomial data likelihood function will also be a Dirichlet distribution. Again, the prior and posterior have the same functional form because the Dirichlet parameter distribution and the multinomial likelihood function are conjugate.

The Dirichlet prior distribution form is:

$$p(\theta_1,\theta_2,\theta_3,\theta_4) \propto \theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\theta_3^{\alpha_3-1}\theta_4^{\alpha_4-1} \tag{52}$$

$$\sum_{i=1}^{4}\alpha_i = \alpha_0 \tag{53}$$

where

$\theta_1 - \theta_4$ are the probabilities of a sample belonging to bins 1-4, respectively,

$\alpha_1 - \alpha_4$ are coefficients describing the possible values of $\theta_1 - \theta_4$, respectively, and

α_0 is the sum of $\alpha_1 - \alpha_4$, also known as total prior weight or equivalent sample size.

The expected value for θ_i is:

$$E[\theta_i] = \frac{\alpha_i}{\alpha_0} \tag{54}$$

The expected variance for θ_i is:

$$\text{var}[\theta_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{{\alpha_0}^2(\alpha_0 + 1)} \tag{55}$$

where *i* can equal 1, 2, 3, or 4.

To calculate the posterior distribution, per Bayes Theorem (Equation 36), the Dirichlet prior distribution is multiplied by the multinomial likelihood function as follows:

$$p(\theta_1, \theta_2, \theta_3, \theta_4 \mid x_1, x_2, x_3, x_4) \propto p(\theta_1, \theta_2, \theta_3, \theta_4)\, p(x_1, x_2, x_3, x_4 \mid \theta_1, \theta_2, \theta_3, \theta_4) \propto \theta_1^{\alpha_1 - 1}\theta_2^{\alpha_2 - 1}\theta_3^{\alpha_3 - 1}\theta_4^{\alpha_4 - 1}\theta_1^{x_1}\theta_2^{x_2}\theta_3^{x_3}\theta_4^{x_4} \tag{56}$$

This functional form can then be rearranged into a Dirichlet posterior distribution with new coefficient values as follows:

$$p(\theta_1, \theta_2, \theta_3, \theta_4 \mid x_1, x_2, x_3, x_4) \propto \theta_1^{x_1 + \alpha_1 - 1}\theta_2^{x_2 + \alpha_2 - 1}\theta_3^{x_3 + \alpha_3 - 1}\theta_4^{x_4 + \alpha_4 - 1} \tag{57}$$

$$\sum_{i=1}^{4}(\alpha_i + x_i) = \alpha_0 + n \tag{58}$$

where

$\theta_1 - \theta_4$ are the probabilities of a sampled basin belonging to bins 1-4, respectively,

$\alpha_1 - \alpha_4$ are prior coefficients describing the possible values of $\theta_1 - \theta_4$, respectively, α_0 is the sum of $\alpha_1 - \alpha_4$, also known as total prior weight or equivalent sample size. $x_1 - x_4$ are the counts of sampled basins belonging to each of four bins, respectively, and n is the sum of all counts (i.e., total sample size).

The expected value for θ_i given the data is now:

$$E[\theta_i \mid x_i, n] = \frac{\alpha_i + x_i}{\alpha_0 + n} \tag{59}$$

This updated posterior is a compromise between information about θ_i from the prior $\left(\frac{\alpha_i}{\alpha_0}\right)$ and from the data $\left(\frac{x_i}{n}\right)$. The expected variance for θ_i given the data is now:

$$\text{var}[\theta_i \mid x_i, n] = \frac{(\alpha_i + x_i)(\alpha_0 + n - \alpha_i + x_i)}{(\alpha_0 + n)^2(\alpha_0 + n + 1)} \tag{60}$$

where *i* can equal 1, 2, 3, or 4.

Because of Dirichlet-multinomial conjugacy, Bayesian updating of a Dirichlet prior distribution with multinomial data results in a Dirichlet posterior distribution. Where the Dirichlet prior is described using coefficients α_1, α_2, α_3, and α_4, the Dirichlet posterior is of the same form as the prior, but instead described using coefficients $x_1+\alpha_1$, $x_2+\alpha_2$, $x_3+\alpha_3$, and $x_4+\alpha_4$. With such a simple, exact solution for the posterior distribution

available, the results of updating a multinomial-Dirichlet Bayesian network can be reported directly. There is no longer a need for cumbersome MCMC calculations.

The probabilistic results of a discrete form Bayesian network can then be easily represented by marginal bar graphs with one bar for every discrete state. The combination of this intuitive model representation and swift computation speed allows model users to change the marginal distribution of one or more nodes and instantaneously observe the effect on all other nodes' marginal distributions.

Ultimately, Bayesian networks describe a system of probability relationships between variables in terms of the likelihood of a variable showing a particular value. All solutions are in terms of probabilities; there are no deterministic equations relating variables. As such, Bayesian networks are not subject to the typical multivariate computational problems associated with finding a unique solution. While the updating process determines a unique set of conditional probability tables (i.e., there is no ambiguity in the calculation procedures), there are no unique predictions of node values. Solutions describe distributions of possible node values, not single values. Therefore, when we predict which values of certain drivers may lead to a possible outcome, we report those predictions in terms of the probability of a driver value falling in each of a set of categories, which is a type of solution that summarizes uncertainty.

Typically, when Bayesian networks have been developed to address environmental questions (Borsuk et al. 2004, McNay et al. 2006, Ticehurst et al. 2007),

modelers only develop a prior model and, therefore, are only capable of employing the discrete, basic use of Bayes Theorem (Equation 37) for marginal distribution prediction through evidence propagation. These models are not truly Bayesian in that they do not use Bayes Theorem to update the model parameters themselves, they simply make inference on a prior model. Additionally, this inference is made using a Bayesian network model version created through non-systematic combination of any available expert information, data, or submodels to parameterize the conditional probabilities. In contrast, we use separately elicited expert information to develop a stand-alone 'prior' Bayesian network and then update the conditional probability distributions with a set of 85 data points collected on each node to create a separate 'posterior' Bayesian network. This is unique in many ways. First of all, it fully applies the philosophical principles of a Bayesian modeling approach in that we update the entire set of prior knowledge based model parameters with new data instead of simply making predictions based on a prior model. This enables us to compare the current state of expert knowledge about the process of urbanization affecting stream biotic communities with measured data on the system. We can evaluate the posterior model to determine where in the network prior information or data dominate relationships. In very few projects have measurements and expert opinion been collected on all the variables of interest and their connections in an actual environmental context to make this type of analysis possible. Further, we develop a systematic, documented procedure for eliciting the prior exclusively from

expert knowledge. This procedure includes an evaluation of different methods of eliciting prior weight, a value needed to determine the weight to be placed on prior information relative to the data when calculating the posterior. Prior weights have not been elicited in the literature for discrete Bayesian networks. Finally, through this process of methods development, we uncover and address unique questions concerning how to discretize continuous variables (number of bins and bin endpoints), how many parent nodes should be elicited per child node, what to consider when selecting variables to be representative of model node concepts, and the impact of zero probability in the prior.

3.2.2. Prior development

The first step in constructing a Bayesian network is determining what, if any, prior information exists and should be incorporated into the model. In the case of urbanization processes and their ultimate effect on stream macroinvertebrates, there exists a large body of knowledge concerning many aspects of this phenomenon. This information is challenging to combine quantitatively as it has been measured according to many different kinds of study designs, collected in myriad different locations at a range of different scales both spatially and temporally. In fact, there is currently no accepted method for quantifying a Bayesian prior. Yet we know this prior information exists and ideally should inform our model. We also know that there exist expert scientists who have spent their careers collecting, analyzing, and interpreting these

varying datasets. This is the type of collective understanding that is already in use in model development, albeit implicitly. Choosing model structures and parameters is not merely 'objective,' as scientists often prefer to believe, but requires experts to amalgamate different kinds of available data to infer the most current hypothesis of process function. In a way that goes beyond simple data summary, experts provide an integrated estimate of system relationships and uncertainties based on all information synthesized from a career's worth of experience. The incorporation of this rich well of expert knowledge and judgment into the modeling process can be conducted explicitly through formal expert elicitation. Expert elicitation specifically refers to the translation of expert knowledge into a probabilistic framework (Morgan and Henrion 1990, Frey 1998, Winkler 2003, Reckhow et al. 2005, O'Hagan et al. 2006). The goal is to codify expert knowledge, appropriately recognizing the degree of uncertainty in that knowledge, and then use that knowledge directly in model development and parameterization.

The optimal way of informing a prior distribution in Bayesian analysis is still very much an active research question (Craig et al. 1998, Walls and Quigley 2001, Alfaro and Holder 2006, Choy et al. 2009). In the literature where expert opinion is used in the prior, the process of elicitation is often not explicit. Generally, the only available documentation for prior development only vaguely suggests that data and experts were consulted in the development process (Rieman et al. 2001, Dlamini 2009, Stewart-Koster

et al. 2009). Bromley et al. (2005), for example, state that the arrangement of nodes and arrows and variable bin endpoints are simply "users' choice". However, these selections are major decisions that greatly impact model results. Therefore, we develop and refine a unique, systematic methodology to create an informed expert prior. This comprehensive expert prior model development involves:

(1) formal elicitation of model structure (layout of nodes and arrows),

(2) variable and bin selection for the nodes,

(3) elicitation of conditional probability tables defined by the structure, and

(4) elicitation of prior weights to enable Bayesian updating.

3.2.2.1. Structure elicitation

Our expert elicited model structure needs to accurately capture how the expert, a seasoned USGS ecologist, thinks about the process of urbanization affecting stream biological condition. Ultimately, the elicited structure should be an adequate reflection of the expert's conceptual understanding of what affects what based on nearly 30 years of career experience with ecological aquatic systems in the Southeast United States. Model structure was elicited from scratch directly from the expert (as opposed to creating a literature review based model and asking for feedback) so as not to bias the expert's thinking process. Elicitation was initiated by conducting a 3-hour open ended expert interview systematically addressing the following questions:

- What things contribute to biologically degraded stream condition?

- How would you be able to tell how much invertebrate communities in a stream are degraded?
- What aspects of urbanization most affect aquatic invertebrate communities?
- Through what chain of events do these aspects affect aquatic invertebrate communities?

The goal was to elicit a directed causal narrative (Nadkarni and Shenoy 2004) describing step by step the process of urbanization affecting stream macroinvertebrate communities. During this interview, the expert identified ecosystem elements which directly affect macroinvertebrate community responses and worked backwards in a causal chain toward urban drivers resulting in a 16-page narrative transcript. Formal textual analysis was conducted on the interview transcript to code major concepts and causal relationships between concepts. Assuming coded concepts represented nodes and coded relationships represented arrows, an initial causal map of the expert's representation of the process was constructed (Figure 49). In this way, we derived a thorough set of cause-and-effect relationships describing the process of urban stressors impacting aquatic macroinvertebrate assemblage responses in streams of the Southeast United States. This model structure is specific to a particular set of ecoregions as different parts of country undergo different urbanization patterns (Kashuba et al. 2010) which the scope of this model was not designed to capture.

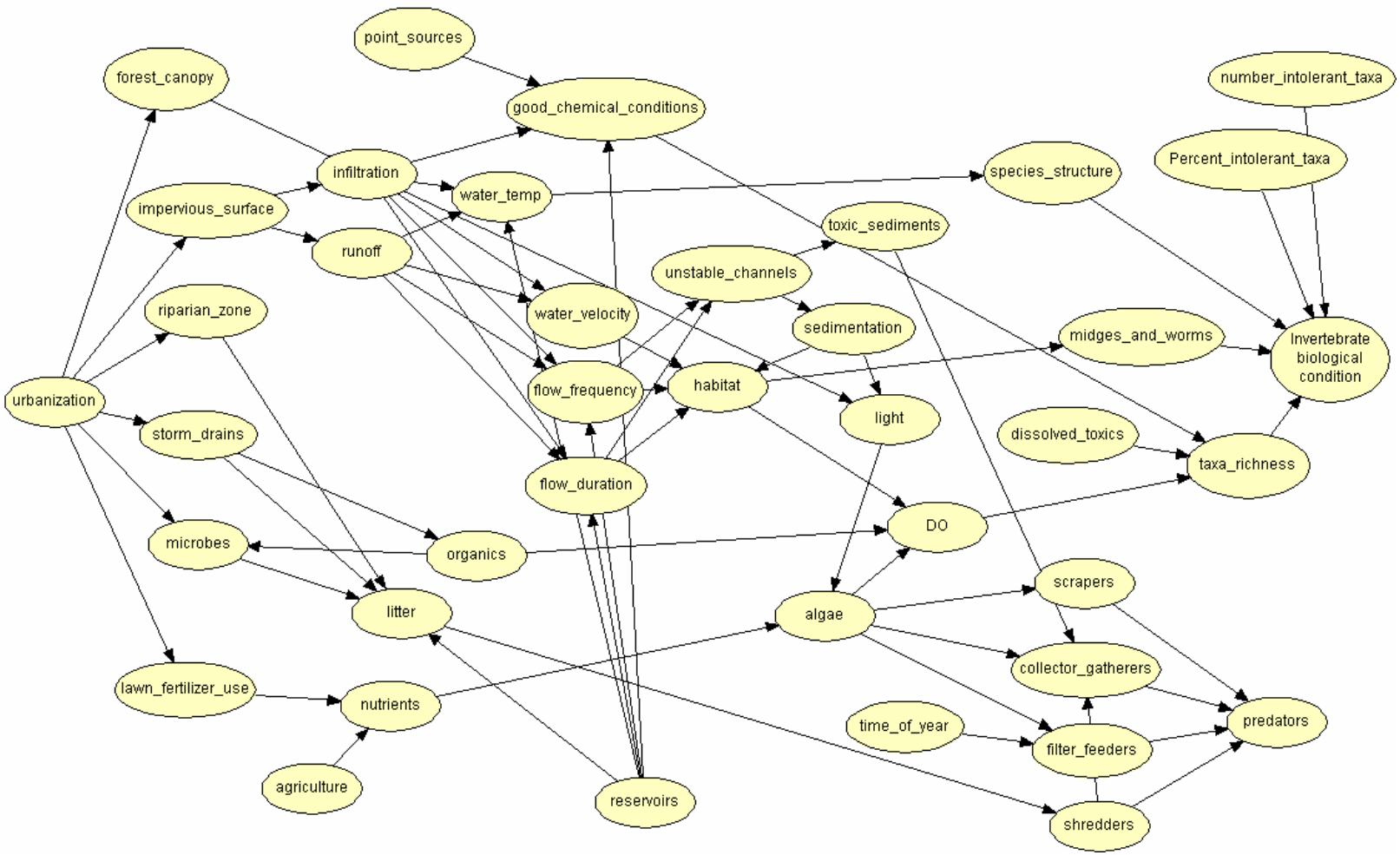

Figure 49: Initial representation of the process of urbanization affecting invertebrate biological condition based on textual analysis of a narrative expert elicitation interview.

In model structure development, it is important to eliminate reiteration of the same concepts, to clearly define causal pathways, and to balance model complexity with ability to parameterize in order to avoid overfitting. Additionally, in order for a Bayesian network to be statistically viable, it must be a directed, acyclic graph with no feedback loops. Following these guiding principles, the expert refined the model derived directly from causal narrative via logical reasoning such as: (1) in the Southeast, since the conversion of agricultural land to urban is such a minor component (Cuffney et al. 2009a), urban land cover and forest land cover are directly inversely related, so the model driver can really be just one node incorporating 'forest removal and conversion to urban', (2) the model should clearly track pathways of water as they are being affected by urbanization, and (3) the model endpoint was not well established in the initial structure; all the identified, overlapping invertebrate response metrics should be consolidated into one 'invertebrate condition' node. Using these and similar arguments as well as triaging most vital system components and relationships, the expert was able to pare down the initial structure (40 nodes and 62 arrows, Figure 49) to a more parsimonious second iteration structure (18 nodes and 36 arrows, Figure 50). This 'parsimonious' structure was, nonetheless, far more complex than previously-attempted one to four variable regression models.

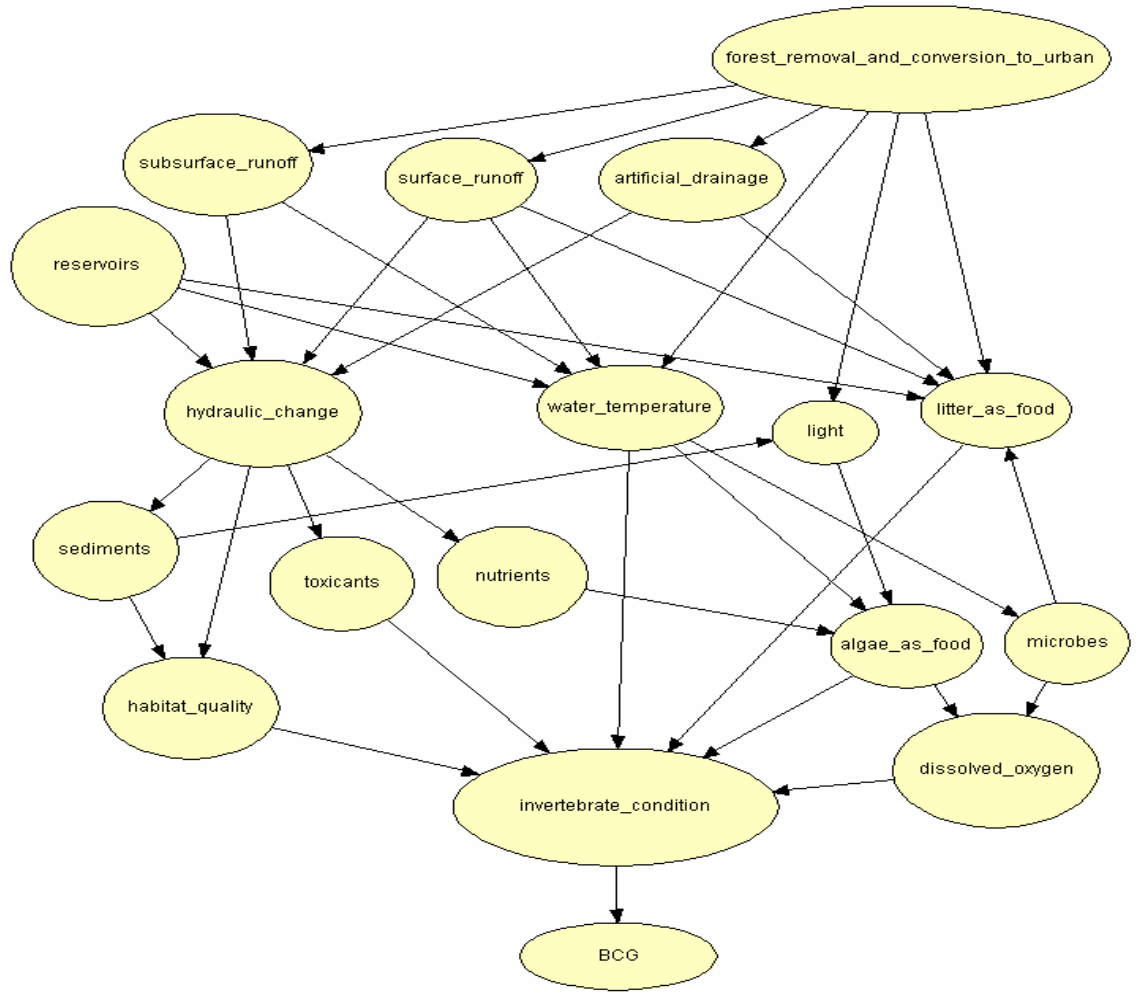

Figure 50: Southeast intermediate urbanization-to-macroinvertebrate-response model structure.

To augment model validity in moving from the second to final iteration model structure, additional expert judgment was solicited from a team of three state-level biologists familiar with invertebrate community structure metrics. This stage of model refinement aimed to limit the number of parent nodes a child node could have while maintaining thorough, defensible system characterization. In order to achieve expert parameterization of all conditional probability tables in the model, we needed to

consider the physiological limitations of human thought in analyzing multidimensional problems directly. Clearly, being able to reason in the seven dimensional conditional probabilities required by the second iteration model structure (where 'invertebrate condition' has six parents) are not within unaided human capacity. Research shows that people cannot think in terms of more than a few conditioning factors (Morgan and Henrion 1990). Therefore, we chose to stipulate that a child node should have no more than three parents. Given this limitation, experts were asked which three inputs determining macroinvertebrate community condition were the most important. Consensus quickly arrived at 'hydrology', 'habitat', and 'water quality'. Experts also considered the possibility of including other relevant concepts within the larger concepts of these three nodes such as incorporating sediments, light, and litter into the 'habitat' node. The realization that each of three 'flow generation' nodes in the second iteration model (Figure 50) pointed to the same set of child nodes ('hydraulic change', 'water temp', 'litter') led to the decision to combine these parent nodes into one general 'flow generation' node. Finally, the experts chose three invertebrate metrics that, in their joint opinion, best characterize invertebrate condition in the southeast United States.

Therefore, in a three-stage iterative process of expert feedback, the first causal map was pared down twice into a 9 node, 16 arrow model structure affirmed by expert consensus that represents all important processes and causal influences while also being

of sufficiently manageable size to allow for parameterization. This became the southeast model used for further Bayesian network methodology development (Figure 51).

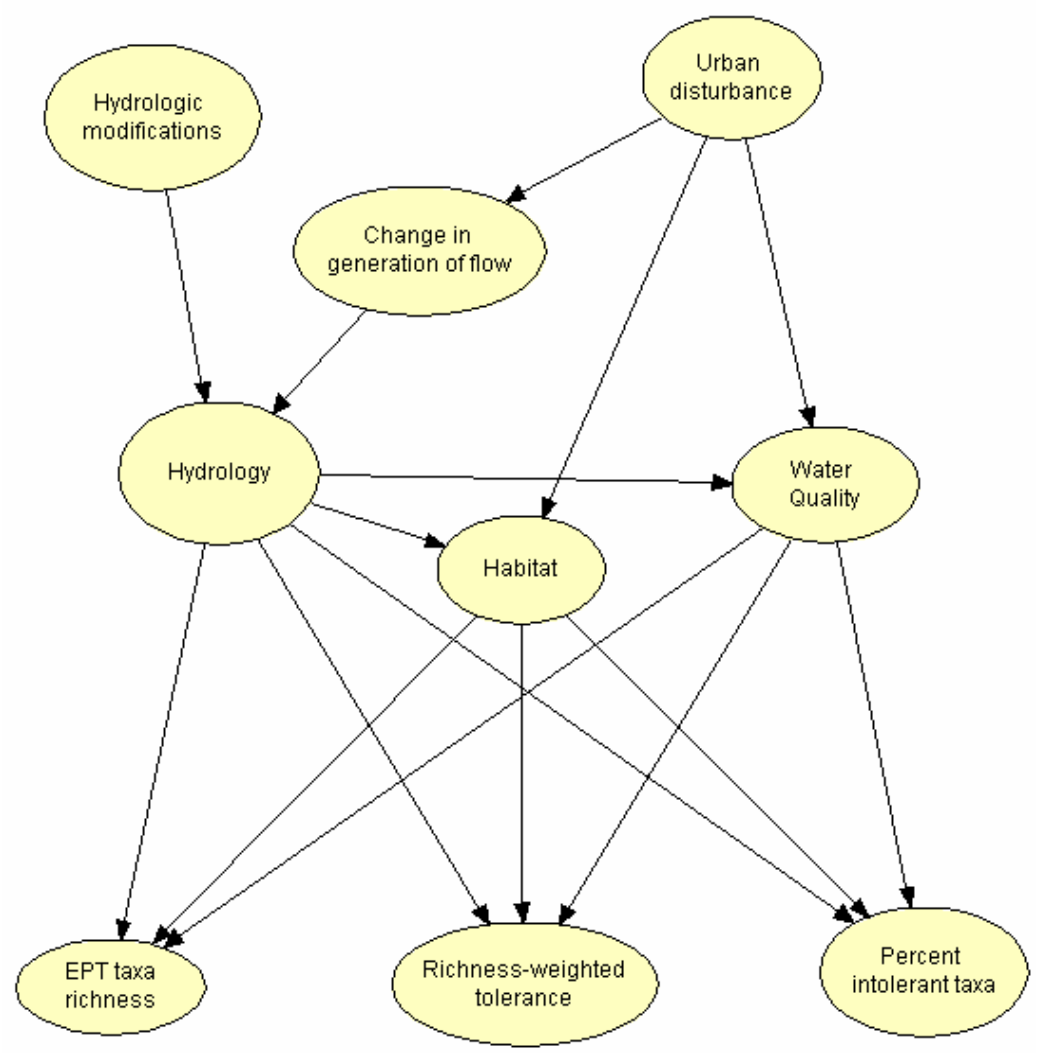

Figure 51: Southeast final urbanization-to-macroinvertebrate-response model structure

Although there are algorithms available to "learn" model structure from data in data mining literature, these algorithms were not used for the following reasons. First of all, depending on model size, data requirements for structure learning accuracy can be

extremely large. Even with abundant data, structure learning algorithms such as PC (Peter and Clark algorithm) (Spirtes et al. 2000) or NPC (Necessary Path Condition algorithm) (Steck 2001) are not incredibly robust, often "learning" incorrect or nonsensical models. In many cases, these algorithms cannot even re-create a known model structure from data sampled directly from that structure (de Campos and Castellano 2007). Additionally, algorithm-'learned' model structure is dependent on often arbitrary selections of parameters such as bin number or size (Alameddine et al. 2010). Structure learning is particularly inappropriate when applied to environmental model structure building due to the complexity of environmental systems and the variability of largely observational environmental data. As explained by Uusitalo (2007), "Structural learning, especially in the constraint-based form, has only limited use in environmental research, however. Often there are knowledge, theories, and hypotheses on the causal interactions between the variables, and it is then only reasonable to make use of them by building the model structures accordingly. Experience and simulations have shown that environmental interaction, which often includes a lot of variation and uncertainty, cannot be reliably estimated based on the available datasets. Theories about causal connections generally result in better models." This is why we choose to rely on expert elicitation rather than structure learning to build our model structure. In some cases, data-driven structure learning may be used for elicited model checking; however, with this study's limited sample size and known

structure learning inconsistencies in environmental contexts (Alameddine et al. 2010), we should not put too much emphasis on the structure "discovered" by data. Additionally, if available data were used to 'learn' the structure and build the prior model, there would be no remaining data to use for model updating. Using structure learning algorithms, even if facilitated by expert judgment as in the NPC algorithm, would eliminate the ability to systematically compare the state of expert knowledge to data to better quantify where there are gaps in measurement or understanding.

3.2.2.2. Variable selection

By necessity, a model is a simplified representation of reality. It cannot be too detailed because it will be difficult to parameterize and not useful for applied purposes (Reckhow 1999). As explained in the *Structure elicitation* section above, many nodes in the final structure consolidated multiple system components to achieve this required parsimony. After structure elicitation, the model exists as a graphical representation of qualitative concepts and their causal connections. To be able to convert such a causal map into a quantitative Bayesian network, a single, defined variable which is most representative of each of these larger concepts must be selected to represent each node. As expert elicited invertebrate metrics, 'EPT taxa richness', 'Richness-weighted tolerance', and 'Percent intolerant taxa' are already described by single variables, the larger concept nodes to parameterize include 'Urban disturbance', 'Hydrologic modifications', 'Change in generation of flow', , 'Hydrology', 'Habitat', and 'Water

Quality'. Starting with a list of system components encompassed by each node derived during model structure elicitation (Table 22), the USGS expert selected a single variable to represent each node (Table 23).

Table 22: System components included in larger concept final Southeast United States model nodes.

Node	Components
Urban disturbance	-impervious surface -housing unit density -percent developed land -fragmentation statistics -forest removal and conversion to urban -road density
Hydrologic modifications	-reservoirs -hydroelectric dam -stream simplification -sinuosity (reach scale and segment scale)
Change in generation of flow	-subsurface runoff -surface runoff -artificial drainage BMPs that would affect this node: -storm drains -retention ponds -rain gardens -buffers -swales
Hydrology	-flashiness
Habitat	-integrated habitat score -algal cover -shear stress -channel morphology (shape of channel) -substrate composition (sediment) -embeddedness -geomorphic channel characteristics -channel shading -light (PAR meter) -litter

Table 22 continued: System components included in larger concept final Southeast United States model nodes.

Node	Components
Water quality	-conductivity -pesticides/herbicides -nutrients -organics -dissolved oxygen (DO) -metals -temperature

Table 23: Variables chosen by experts to represent nodes in Southeast United States model. [NA is not applicable since Hydrologic modifications was discretized into only three bins since 40% of the samples had zero dam density.]

Node	Variable: units	Bins			
		1	2	3	4
Urban disturbance	Urban land cover: percent urban land cover in basin area	0-12%	>12-36%	>36-66%	>66%
Hydrologic modifications	Dam density: dams per 100 square kilometer basin area	0	>0-6	>6	NA
Change in generation of flow	Impervious surface: percent impervious surface	0-2%	>2-8%	>8-16%	>16%
Hydrology	Flashiness: rises greater than five times the median total rise	0-40	41-63	64-94	95+
Habitat	Channel ratio: mean channel width to depth ratio	0-25	>25-33	>33-44	>44
Water quality	Conductivity: at low base flow, µsiemens per centimeter at 25^{o}C	0-92	>92-129	>129-281	>281
EPT richness	EPT richness: number of EPT taxa	0-4	5-7	8-12	13+
Richness-weighted tolerance	Tolerance: average tolerance of taxa at a site on a scale of 0-1	<5	>5-5.6	>5.6-6.2	>6.2
Percent intolerant taxa	Percent intolerant taxa: percent of total richness composed of intolerant taxa	0-3%	>3-8%	>8-18%	>18%

Variable selection is inherently a subjective judgment, as many variables can be chosen from to represent a particular node. Therefore, we document how the expert relied upon career experience to make and justify this modeling decision. This documentation allows for transparent model development which facilitates honest and open model discussion and evaluation. The expert was to determine which type of measurement, including units, best represented the concept the node was meant to embody. Selections were to be made from among the ~1400 variables available in the USGS EUSE dataset (Giddings et al. 2009).

Percent urban land cover was chosen as a measurement of 'urban disturbance' as, in the expert's experience, it was closely correlated to other integrated measures of urban impact such as impervious surface and road and housing density in the Southeast (Cuffney and Falcone 2008), and was the most direct representation of urbanization at the watershed scale. Dam density was deemed representative of 'hydrologic modifications' that ultimately affect stream hydrology and change patterns of water movement across sampling locations. The 'change in generation of flow' node was meant to embody urban-induced changes to water movement across the landscape both under and above ground. Although these process disruptions are vital ways in which urbanization affects watersheds, these changes are very challenging to measure. Therefore, the available dataset did not include direct characterization of flow generation change in each watershed. As a surrogate, the expert chose watershed

impervious surface percentage to represent landscape flow changes because it is an easily available measure which relates to the amount of water which is no longer able to infiltrate in an urban watershed, therefore contributing to increased surface runoff. The expert acknowledges that this variable selection greatly oversimplifies the complicated process changes of interest, however, given data constraints, in the expert's professional opinion, it was the best possible compromise.

The concept of 'hydrology' was intended to account for the effect of urban-mediated extreme flows on stream invertebrates. This includes volume and timing changes in peak and base flow. A good synthesizing description of these patterns would be a measure of stream flashiness over time. Therefore, the expert chose to parameterize the 'hydrology' node with the number of rises greater than five times the median total rise during the period of record. This variable was selected over metrics calculating rises over three, seven and nine times the median rise because five times was known to correlate closely with urbanization. The expert acknowledged that number of rises over seven or nine times the median could also be used to characterize flashiness but these are more rare events and, therefore, less likely to occur in the sampling record. Ultimately, choosing to use rises greater than five times the median represented a compromise between flashiness extremeness and sample size with which the expert was comfortable. The 'habitat' node incorporates the most diverse assortment of elements that all contribute to the quality of immediate living conditions for invertebrates. The

expert originally considered using one of several integrated habitat scores but eventually conceded that these scores do not always consistently evaluate the full complexity of habitat quality nor are they well-linked to urbanization. It was also preferable to avoid the additional error structure complexities created by using a model within a model. Of all the individual habitat components of import, the export chose to use channel width to depth ratio as this metric is an integrated assessment of urban-related degradation of channel-scale properties including scour, incision, sedimentation, and substrate composition changes. Finally, the 'water quality' node refers to the ability of the stream water itself to support life in terms of what is dissolved in it and its temperature. Originally, the expert chose total pesticides detected as an appropriate aggregating measure of water toxicity, however, these data were unavailable for part of the dataset due to local area drought conditions during high flow sampling. After reconsideration, the expert eventually more confidently selected a measure of conductivity at low flow since it better integrated the influences of both natural and anthropogenic ion generation in the watershed. After original USGS expert variable elicitation, the chosen variables were corroborated by the additional team of state-level biologist experts.

3.2.2.3. Bin number and endpoint selection

As discussed in the *Bayesian networks* section above, the use of a discrete Bayesian network requires that each continuous variable be divided into a set of discrete categories. There is inherently a tradeoff involved in discretizing continuous variables

(Uusitalo 2007). Some of the discriminatory strength of continuous measurements may be lost and error may be created by assumptions associated with binning. The ability to easily define the functional form of a simple, known continuous relationship (e,g,. linear) is complicated by discretization which now requires a more complicated function to describe the same association with less accuracy. However, benefits gained include substantially increased Bayesian network calculation speeds, intuitiveness of graphical output, and parameterizablity of complex, multivariate relationships which would have been prohibitive to describe with unknown continuous functions. In this project, we greatly value the benefits of binning over the drawbacks.

Ideally, each created category should represent a group of values similar to each other and distinct from other categories. This is because, after discretization, each continuous number within a category will be treated identically as a member of that category. In practice, these clear distinctions are often difficult to achieve. There is no definitive process in the literature for choosing number of categories and endpoint values for each category, although these 'bins' may have significant effects on the resulting analysis. Uusitalo (2007) suggests that the more bins a variable is split into, the harder it will be to establish dependencies between model variables. Unfortunately, the model is less able to discriminate between states and evaluate changes to that variable when the variable has fewer possible states. Compromising between these opposing logic directions and also considering the ability of an expert to distinguish bins and data

availability for populating bins, we chose to split each continuous variable into four discrete categories. These categories represent low, medium-low, medium-high, and high values of each variable (with the exception of 'Hydrologic Modifications' discussed below) and can also be referred to as bin 1, 2, 3, and 4, respectively.

Bin endpoints were then elicited systematically. First, the expert was asked whether there were any thresholds of ecological or environmental significance which designate how a low, medium-low, medium-high, or high state of any of the chosen model variables would be defined. For variables to which no clear threshold applies, available data were split into quartiles to evaluate potential bin endpoints given equal frequency in each bin. Finally, starting with the a priori threshold and quartile bin definitions, bin endpoints were amended via expert opinion during conditional probability table elicitation. It was found that, when considering the effect of a parent node state on a child node state, the context of a causal question often clarified ecologically or environmentally significant thresholds. If the a priori bin assignment made sense in context, it was not changed.

In this southeast Bayesian network, initial bin endpoints were calculated using quartiles for all variables except dam density (used to parameterize the 'Hydrologic Modifications' node). The southeast dam density data showed that more than one fourth of the measurements were 0 dams/km^2. In fact, closer to one third of the dataset had no dams in the sampled watersheds. Since the same value cannot belong to more

than one bin, dam density was divided by frequency into three bins instead of four to accommodate the natural divisions of the data and maintain mutually exclusive bin definitions. During conditional probability elicitation, the expert then confirmed the eight quartile-calculated and one thirtile-calculated bin endpoint selections as acceptable ecologically without amendment, resulting in final model bin definitions for all nine variables (Table 23).

3.2.2.4. Conditional probability table elicitation

The heart of quantifying expert judgment and uncertainty in a Bayesian network lies in elicitation of conditional probability tables (CPTs). These tables define the relationships represented by the arrows between model nodes. It is the product of these conditional probabilities that enables the calculation of the joint distribution of all variables in the network. They are represented as discrete distributions across the states of a child node given a combination of possible states of its parent node(s). For example, consider again the simple Bayes net from the *Bayesian networks* section (Figure 47). Assume each node has two possible states: low and high. A conditional probability distribution for node D given nodes B and C may look like Table 24.

Table 24: Example conditional probability table (CPT).

		D	
B	**C**	**D = low**	**D = high**
B = low	C = low	0.90	0.10
B = low	C = high	0.75	0.25
B = high	C = low	0.20	0.80
B = high	C = high	0.01	0.99

The first two columns show the state of the parent nodes. In the first row, both B and C are in state 'low'. The last two columns show the distribution across the possible states of the child node given known, static values of the parent nodes. The first row shows that when B and C are low, D has a 90% chance of being low and a 10% chance of being high. All the uncertainty about a situation about how a low B and C would affect D is incorporated in the distribution of possible values of D. As discussed in the *Bayesian networks* section above, this uncertainty could be due to factors such as measurement error, imperfect knowledge or just random variability. The second row, then, shows the distribution of D when B is low and C is high (that is, D is 75% likely to be low and 25% likely to be high). Knowing how the distribution on D changes as its parent states change codifies the joint effect of B and C on D. In this way, conditional probability tables describe relationships between variables and uncertainty in those relationships. Every child node in a Bayesian network must be assigned such a series of discrete distributions conditional on its parent nodes. In our case, the prior set of conditional probability tables is assigned via expert elicitation.

There are several documented direct and indirect probability elicitation methods for binomial distributions (Winkler 1967, Morgan and Henrion 1990, Winkler 2003); however, the literature remains sparse on ideal elicitation methods for the multinomial distributions used in this model. This is because, in a binomial construct, there is only one probability, θ, to be elicited to describe the likelihood of an event occurring, as the

probability of the event not occurring is, by definition, fixed at 1-θ. In a multinomial scenario, it is much less clear how to assess an expert's desired distribution of probability weight among more than two categories. Therefore, we had to develop an adaptation of the direct fixed value elicitation method (Clemen 1991) to apply to multinomial elicitations, following the Stanford Research Institute Elicitation Protocol (Spetzler and von Holstein 1975) to attempt to minimize bias and error. The protocol specifies five steps to assist the expert in a rational, thoughtful evaluation of his or her expertise:

1. Motivating: Establish rapport

The first step in expert elicitation involves making sure that the expert has a comfortable understanding of the process. This includes explaining the nature of problem and analysis, giving the expert context on how his or her judgments will be used, discussing the general methodology of a probabilistic assessment, explaining heuristics expert can use, and identifying any potential motivational biases.

2. Structuring: Define uncertain quality

Once the expert is oriented as to the general what, how and why, the next step is to clearly define the specific questions about which the expert will be providing judgment. During this step, it is important to define variables of interest unambiguously and identify variable units as well as possible ranges of values. Variables can be

disaggregated into more elementary variables, if necessary, or combined into summary variables, as appropriate.

3. Conditioning: Think about all evidence

After the specific values to be elicited are chosen, the expert should then be prompted to think about all of his or her relevant expert knowledge concerning the variables and relationships of interest. This knowledge could include data, theoretical models, analogies with similar systems or other sources of understanding. The expert should be encouraged to think from different perspectives and draw on as much information as possible in order to overcome potential biases related to use of limited scope. For example, the elicitor can ask the expert to invent scenarios for extreme outcomes and ask the expert to explain how these different outcomes could occur.

4. Encoding: Quantify expert judgment

After proper preparation in previous steps, this step comprises the actual elicitation. Probabilistic information can be elicited according to many different proposed protocols and we chose the direct fixed value elicitation method as applied to multinomial questions. In this method, the elicitor fixes the variable value and elicits the probability directly. (In other methods, the elicitor can fix the probability and directly elicit the variable value or conduct an indirect reference lottery, among other things.) To apply this method to the elicitation of multiple probabilities, experts were shown a

combination of states of parent nodes and asked to directly provide probabilities for each fixed bin of the child node, using their expert judgment.

5. Verifying: Check the answer

Finally, after all desired probabilities are encoded, the elicitor should test the expert answers given to see if they correctly capture the expert's opinion. This can be done by rephrasing an expert's answer in another way to see if the expert still agrees with the assessment. If the expert does not confirm the answer, the elicitor may need to repeat conditioning and encoding steps.

Following this expert elicitation protocol, before eliciting conditional probability tables (CPTs), the expert was Motivated (step 1) via an explanation that the purpose of the exercise is to capture his or her expert knowledge about the process of urbanization affecting stream macroinvertebrate communities in the southeast US in a way that appropriately recognizes the degree of uncertainty due to randomness and imperfect knowledge. We explicitly discussed how the expert assessment will then be combined with data specific to the Southeast (see *Data* section above) to create a Bayesian network describing this urbanization process. The elicitation was then Structured (step 2) by reviewing which variables and variable units had been chosen to represent the nine nodes of the model structure developed through expert consensus. We clarified that the expert would be asked to provide judgment on seven separate conditional probability tables, one table for each child node that had parent nodes pointing to it. This judgment

would involve the expert's opinion of the causal effect of each combination of parent nodes states on a child node distribution. For example, the question of interest for the 'Hydrology' node is: how do urban-induced changes in generation of flow (measured by impervious surface percent) and hydrologic modifications (measured by dam density) affect stream flashiness (measured by number of rises greater than five times the median in a period of record)? The two marginal nodes with no parents require no conditional probability judgment; they were assigned uniform prior distributions, in accordance with their equal frequency binning, to limit bias in the model drivers. Next, the expert was Conditioned (step 3) via a review of the ecological reasoning used to choose the representative variables for each model node. Further, the expert was asked to discuss the available evidence on the connections between these variables, including dissenting views of which he or she was aware and extreme scenarios.

Encoding (step 4) was conducted in a series of sessions due to the large scope of information required (720 elicited probabilities total; Table 25). In all sessions, we used amended direct fixed value elicitation. The principle of this fixed value method is that the category values remain constant while the expert adjusts the probabilities of a random sample falling into each of the categories until the probabilities match his or her professional belief. In a binary scenario, this process amounts to providing just one probability as the second probability is automatically equal to 1 minus the first probability. In this multinomial context, the expert is instead asked to provide several

Table 25: Total number of conditional probabilities in final southeast model. For every combination of parent nodes states, the distribution across child node states is multinomial where the probabilities across all child node states must add to 1. Because of this bound, the final child node state probability is fixed. For example, if a child node has four bins, after three of the four probabilities are assigned, the final probability must be 1 minus the sum of the three assigned probabilities and is, therefore, not an independently required probability. This number of required child node state probabilities is multiplied by the number of parent node state combinations to calculate total conditional probabilities required per child node CPT.

Node	Number of child states	Parent nodes (number of parent states)	Number of parent state combinations	Total conditional probabilities required
Urban disturbance	4	None	0	0
Hydrologic modifications	3	None	0	0
Change in generation of flow	4	Urban disturbance (4)	4	3 * 4 = 12
Hydrology	4	Change in generation of flow (4) & Hydrologic modifications (3)	4 * 3 = 12	3 * 12 = 36
Habitat	4	Urban disturbance (4) & Hydrology (4)	4 * 4 = 16	3 * 16 = 48
Water quality	4	Urban disturbance (4) & Hydrology (4)	4 * 4 = 16	3 * 16 = 48
EPT richness	4	Hydrology (4), Habitat (4) & Water quality (4)	4 * 4 * 4 = 64	3 * 64 = 192
Richness-weighted tolerance	4	Hydrology (4), Habitat (4) & Water quality (4)	4 * 4 * 4 = 64	3 * 64 = 192
Percent intolerant taxa	4	Hydrology (4), Habitat (4) & Water quality (4)	4 * 4 * 4 = 64	3 * 64 = 192
			TOTAL	720

probabilities and gauge their distribution across four possibilities. In the binary scenario, while thinking about the probability of an event happening, the expert can

automatically assess the complementary probability of that event not happening. In contrast, asking for a probability for one of multiple categories at a time does not automatically invoke a thorough mental review of the likelihoods of all other possible categories. In fact, decision theory literature has shown that a person is most likely to overestimate his or her probability for one of multiple categories if asked for only one at a time because of bias associated with unpacking nested information (Tversky and Koehler 1994). This is why multinomial probabilities need to be elicited in the context of all the other options. Therefore, instead of asking for the probability of a specific event happening, as is done in typical binary direct fixed value elicitation, we instead phrased the elicitation question in a way that asked for a distribution of multiple events happening. We established a denominator of total events and then requested that the expert divide up the total into four possible categories. Per this amendment of the fixed value direct elicitation procedure, for every row in a CPT, the expert was asked to provide his or her probability for each state of a child node given a set of fixed state(s) for the parent node(s). The probabilities for each row must add to 1 and probabilities can be moved from bin to bin until the expert is satisfied with the distribution. This procedure was repeated for every combination of parent node states until the CPT was complete. Seven such CPTs of varying levels of complexity were elicited in three separate sessions to complete the model.

Taking the "Hydrology" node example again, CPT elicitation involves creating a table linking the "Hydrology" node with its parent nodes, "Change in generation of flow" and "Hydrologic modifications" (Figure 51). To do so, the expert is guided to consider his or her beliefs about the likelihood of streams having low, medium-low, medium-high, or high flashiness when impervious surface and dam density are each in a known, fixed state. The first row in the "Hydrology" node CPT is elicited by asking the expert: "Assume the watersheds of 100 streams in the southeast United States are known to have 0-2% impervious surface and 0 dams/100 km^2. Assume all other stream features are randomly distributed as if you took a random sample of low impervious surface, low dam density streams. How would you expect the flashiness of these 100 streams to be distributed? Specifically, how many of those 100 streams would you expect to have flashiness less than 40 rises greater than five times the median in a period of record (low flashiness), between 41 and 63 rises (medium-low flashiness), between 64 and 94 rises (medium-high flashiness) or more than 95 rises (high flashiness)?" The verbal cues are facilitated by visual representation of the question which includes an interactive count of how many stream remain to allocate (Table 26).

Table 26: Visual representation of conditional probability elicitation example.

Given:		*How would you expect 100 streams to be distributed?*			
Impervious surface	Dam density	Flashiness greater than 5 times the median			
0-2 %	0 dams/100 km^2	**0-40 rises**	**41-63 rises**	**64-94 rises**	**95+ rises**

The expert responded that in watersheds with low impervious surface and low dam density, he or she would expect the majority of streams to have low flashiness. Given all the uncertainty and unknowns, he or she estimated this to translate into about 88 streams of 100 likely having low flashiness. The expert then assessed that, of the remaining 12 low impervious surface/low dam density streams, 10 streams may have medium-low flashiness, 1 stream may have medium-high flashiness and 1 stream may have high flashiness. These expert frequencies easily normalized into probabilities which determine the first row of the prior CPT for the "Hydrology" node (Table 27). The rest of the rows are elicited by asking the expert to make the same assessment of the most likely flashiness distribution given (i.e., conditional on) each remaining possible combination of parent node states. This direct elicitation process captures how the expert expects flashiness to change in response to changes in impervious surface and dam density. The completed prior CPT (Table 27) shows that the expert believes that flashiness generally decreases with increasing dam density and increases with increasing impervious surface. The resulting probability distributions across flashiness also encompass the expert's professional estimate of the uncertainty in this relationship. In this way, we assess an expert's professional opinion in a way that is much more thorough and quantitative than the implicit, often unacknowledged use of expert judgment to inform non-Bayesian models.

Table 27: Prior conditional probability table and prior weight (α_0) for "Hydrology" node (flashiness bins in probability units and α_0 in equivalent data points). [Table rows selected for probability range prior weight elicitation are highlighted and prior expected values within those rows selected for hypothetical sample prior weight elicitation are shown in bold.]

Impervious surface	Dam density	Flashiness greater than 5 times the median				
		0-40 rises	41-63 rises	64-94 rises	95+ rises	α_0
0-2%	0 dams/100 km^2	0.88	0.10	0.01	0.01	6.45
>2-8%	0 dams/100 km^2	0.19	0.65	0.15	0.01	6.45
>8-16%	0 dams/100 km^2	0.05	0.15	0.60	**0.20**	6.45
>16%	0 dams/100 km^2	0.01	0.04	0.15	0.80	6.45
0-2%	>0-6 dams/100 km^2	0.93	0.05	0.01	0.01	6.45
>2-8%	>0-6 dams/100 km^2	0.40	0.54	0.05	0.01	6.45
>8-16%	>0-6 dams/100 km^2	0.05	0.40	0.50	0.05	6.45
>16%	>0-6 dams/100 km^2	0.05	0.10	0.30	0.55	6.45
0-2%	>6 dams/100 km^2	**0.93**	0.05	0.01	0.01	6.45
>2-8%	>6 dams/100 km^2	0.55	0.39	0.05	0.01	6.45
>8-16%	>6 dams/100 km^2	0.15	0.50	0.34	0.01	6.45
>16%	>6 dams/100 km^2	0.10	0.20	0.40	0.30	6.45

To facilitate assessment of the CPTs of the three most complex nodes, the invertebrate endpoint nodes, the expert chose to employ a belief submodel. These nodes are the most challenging to parameterize because they each have three parents which requires the expert to think in four dimensions about the effect of hydrology, habitat, and water quality on an invertebrate community measure. To simplify the overwhelmingness of this question and to help organize his or her thought process logically, the expert devised a ranking and weighting scheme to prioritize the 64 different combinations of parent states (i.e., four flashiness states times four channel states times four conductivity states). After the 64 states were sorted from best to worst

conditions, the expert was better able to assess how the distribution across each invert metric may change as driver states change. Parent combinations with the same weighted rank score received the same invertebrate distribution which represents the expert's assumption that is it possible to achieve a certain invertebrate metric value range from different combinations of drivers.

Particular care was taken to address zero probability 'certainty' as part of the elicitation procedure. In a Bayesian network, per Bayes theorem (Equation 36), when zero is entered in the prior, the probability for that bin can never be updated with data because anything multiplied by that zero still equals zero. This is logical because specifying zero in a probability question essentially signifies complete certainty that the outcome will never occur in the same way that specifying one indicates complete certainty that the outcome will always occur. And if the prior shows with total certainty that an outcome is impossible, then no amount of data should be able to overcome that complete certainty. In fact, in the Hugin software used in this analysis (Hugin 2008), a zero in the prior excludes the entire data row that has that combination of parent and child values from analysis. This is because if that combination of values is impossible with total certainty, then the software assumes there must be something wrong with the data if the data show the values to be possible. Therefore, to avoid discarding potentially useful data, it is important to make sure the expert differentiates between very, very unlikely outcomes and impossible ones. Experts are instructed to assign zero

probability to a bin (conditional on parent values) if they are certain that obtaining that outcome is impossible. Otherwise, if the outcome is possible but highly unlikely, experts should assign very low probability, instead.

After all the CPTs were elicited, the probabilities were Verified (step 5) by assembling the linked CPT structure of the Bayes net and determining if marginal Bayesian updating was consistent with expectation. That is, under different hypothetical network updating scenarios, the expert was asked if larger trends make sense. If the network was set to high urban land cover, did the expert agree with implications on drivers and endpoints both in direction and magnitude? Did the fully connected network correctly describe which parent variables were expected to be more important in affecting child distributions? Under many network incarnations, the expert confirmed separately elicited probabilities and, therefore, no changes were made to the seven conditional probability tables. These final prior CPTs for the entire Southeast Bayesian network (including uniform marginal priors for "Urban disturbance" and "Hydrologic modifications") are shown in Appendix B (Table 63-Table 70).

3.2.2.5. Prior weight elicitation

The final step in prior model creation is the elicitation of prior weights. While these weights do not change the prior conditional probability tables themselves, they provide information vital to the Bayesian updating of those CPTs. Per Bayes theorem (Equation 36), the updating process is to result in a weighted average of prior

information and data. Data weight is simply the number of data points. As a general statistical principle, larger sample sizes receive proportionately greater weight in analysis. The goal of the elicitation process, then, is to determine how much weight to place on the prior information relative to the available data. It is useful to think of prior weight as translating into an equivalent sample size related to degree of prior knowledge. If an expert reports a 0.60 probability, is this akin to 6 of 10 samples or 60 of 100 samples (or any other sample size)? While the elicited 0.60 probability does not change, the prior weight determines how many equivalent samples that 0.60 probability information is worth.

To determine this, experts must quantitatively specify their certainty about the provided conditional probability relationships in units of equivalent data points. However, as with model structure, variable, and CPT elicitation, there is no consensus in the literature as to the optimal way of obtaining this information. Therefore, we tested and compared three different methods of expert prior weight elicitation: (1) direct elicitation (Winkler 2009), (2) hypothetical sample (Winkler 2003, 2009), and (3) probability range (Cowell 1999, Jenson 2009).

3.2.2.5.1. Direct elicitation

In direct prior weight elicitation, the expert is asked to provide the prior weight directly, relative to a fixed number of data points to be collected (Winkler 2009). Using this method, the expert directly gauges how many equivalent data points he or she feels his or her prior is worth. To standardize this reasoning process, the expert is provided a

point of comparison in the form of a set measured sample size against which to judge his or her knowledge. The reported number of equivalent data points is to reflect the expert's confidence in his or her prior knowledge relative to a known amount of data measuring the same relationship.

The specification of prior weight is done in Hugin at the level of CPT row by defining the α_0 Dirichlet prior distribution parameter (Equation 53, where each CPT row is described by its own Dirichlet distribution). Using the direct elicitation method, it seems unlikely that an expert would be able to directly identify his or her level of certainty for every separate combination of parent states (i.e., CPT row) or that beliefs about certainty are even differentiable at such a fine scale. Direct intuition about the weight of prior information is much more reasonably elicited at the node level, which asks more broadly about an expert's certainty of his or her understanding of the relationship between a variable and its parent(s). Therefore, direct elicitation was focused at the wider child node scale.

To implement this direct elicitation method, we asked the expert to assume that 10 sites in the southeast were to be sampled for each of the nine variables in the Bayesian network model. Then, for each set of prior conditional probabilities that described a child node distribution given its parent(s), the expert was asked: "Relative to the amount of information you believe is available from these 10 data points, how much information does your prior contain in units of equivalent data points? In other words,

knowing that it will be combined with results from 10 measurements, how many data points do you think your prior represents?"

Each node can have a different number of average data points per CPT row, depending on how many different parent state combinations exist for that child node. If the 85 true samples must be divided into more parent state combinations (i.e., CPT rows), then the average number of samples in each CPT row is less. In this model, CPT row sample sizes range from ~21 data points/CPT row for the four parent states of the "Change in generation of flow" node (85/4) to ~1.3 data points/CPT row for the 64 parent state combinations of each of the invertebrate metrics nodes (i.e., the "EPT Richness", "Richness-weighted tolerance" and "Percent intolerant taxa" nodes) (85/64). (See Appendix B Table 71-Table 78 for total number of samples, n, for each individual CPT row in the model, from which node averages were calculated.) To account for these differences, we standardize elicitation for 10 data points and then convert the standardized value into actual value by accounting for the true average number of data points per child node CPT row. We do this by assuming that the elicited number of equivalent data points is proportional to the prior weight (α_0) normalized by the actual average number of data points available for a CPT row of a node:

(Elicited number of equivalent data points) / (10 reference samples) =

α_0 / (Actual average data points per child node CPT row) (61)

where

α_0 is the prior weight Dirichlet distribution parameter at the CPT row level.

Solving for CPT row prior weight, α_0:

α_0 = (Elicited number of equivalent data points * Actual average data points per child node CPT row) / 10 reference samples (62)

For example, if an expert judged his certainty about his "Hydrology" node prior to be worth 8 equivalent data points relative to 10 collected data points, and there were 7.08 average actual data points available for each of 12 "Hydrology" node CPT rows ($85/12 \approx 7$), then the actual (proportional) prior weight for each CPT row in the "Hydrology" node, α_0, would be 5.67 equivalent data points. (Units of equivalent data points do not have to be whole numbers per Dirichlet hyperparameter allowed values.) Using this proportionality assumption, the node level standardized-to-10-samples direct elicitation was converted into CPT row specific prior weights. This resulted in one final mean prior weight value per child node specific to the actual mean amount of data available for that node (Table 28). Standard deviation around the mean prior weight value represents within-node variability in n, total number of samples in CPT row. This variability is not a measure of variability resulting from using the direct elicitation method. Since standardized equivalent data point values were directly elicited only once per node, is not possible to calculate within-method variability for the direct elicitation method as is done for the other two elicitation methods.

Table 28: Direct prior weight elicitation results.

Child node (variable)	Directly elicited prior weight relative to 10 hypothetical data points	Mean ± standard deviation actual data points in CPT row	Proportional mean ± standard deviation prior weight (α_0) for all CPT rows in node
Change in generation of flow (Impervious surface)	10	21.25 ± 0.96	21.25 ± 0.96
Hydrology (Flashiness)	8	7.08 ± 2.84	5.67 ± 2.27
Habitat (Channel ratio)	5	5.31 ± 3.40	2.66 ± 1.70
Water Quality (Conductivity)	6	5.31 ± 3.40	3.19 ± 2.04
EPT taxa richness	7	1.33 ± 1.35	0.93 ± 0.94
Richness-weighted tolerance	6	1.33 ± 1.35	0.80 ± 0.81
Percent intolerant taxa	7	1.33 ± 1.35	0.93 ± 0.94

3.2.2.5.2. Hypothetical sample elicitation

Another possible method of prior weight elicitation involves showing the expert a hypothetical sample and asking how he would change his originally reported prior probability if he knew that this sample was actually collected (Winkler 2003, 2009). This elicitation method assumes that the expert is using Bayesian reasoning to appropriately weight his prior relative to the hypothetical data to devise his updated probability. If Bayesian reasoning was being used, we could infer the prior weight by observing how far the expert's updated probability is from his original prior probability. If the expert's reported updated probability is close to the prior, despite observing data that are

different from that prior, then the expert has high certainty in his prior assessment and a large prior weight should be assigned. Conversely, if the updated probability is closer to the data, this means the expert was not as confident in the prior since the hypothetical sample significantly affected his assessment. In this case, prior weight should be small.

These theoretically large and small prior weights can be quantified using the Dirichlet-multinomial version of Bayes theorem (Equation 56). The expert's provided conditional probability is considered the prior expected value for probability, $p(\theta_i)$. The hypothetical sample is considered the data (x_i, n) used to update the prior. And the expert's updated probability is considered the posterior expected value for probability, $p(\theta_i | x_i, n)$. Knowing the prior expected value, the data and the posterior expected value, the prior weight can be back-calculated using Dirichlet-multinomial Bayes Theorem. For example, the prior probability reported by the expert for impervious surface to be between 2 and 8% of watershed area (bin 2) given that urban land cover is between 12 and 36% of watershed area (bin 2) is 0.60. As such, the prior Dirichlet expected value for θ_2 in the medium-low urban land cover CPT row of the 'Change in generation of flow' child node is (from Equation 54):

$$E[\theta_2] = \frac{\alpha_2}{\alpha_0} = 0.60 \tag{63}$$

If the given hypothetical sample shows that 8 out of 10 streams in watersheds with 12-36% urban land cover have 2-8% impervious surface (i.e., belong to child node

bin 2), then $x_2 = 8$ and $n = 10$. If the expert reports that, if she saw these data, she would update her probability that a sample falls in 'Impervious surface' bin 2 to 0.70, then (from Equation 59):

$$E[\theta_2 \mid x_2, n] = \frac{\alpha_2 + x_2}{\alpha_0 + n} = 0.70 \tag{64}$$

Solving Equations 63 and 64 for α_0, the prior weight:

$$\alpha_0 = \frac{E[\theta_2 \mid x_2, n] * n - x_2}{E[\theta_2] - E[\theta_2 \mid x_2, n]} = \frac{0.70 * 10 - 8}{0.60 - 0.70} = 10 \tag{65}$$

In this example, the prior has the same weight as the data ($\alpha_0 = n = 10$) and, therefore, the expected value of the posterior (0.70) is exactly in between the expected value of prior (0.60) and maximum likelihood of the data (8/10 = 0.80, Equation 51). This is the actual result for the first hypothetical-sample-calculated prior weight of the 'Change in generation of flow' node (Table 29). If, however, the expert felt there was much uncertainty associated with her prior, then she would have relied greatly on the data in assessing how to update her probability. In this case, the reported posterior probability would have been closer to 0.80 (from data) than to 0.60 (from prior) and the calculated α_0 would be less than 10 (n, the weight of the data). For example, if the expert reported an updated probability of 0.73, the prior weight would equal 5 and the data (n = 10) would have twice as much weight as the prior in the calculation of the posterior. Conversely, if the expert was confident in her prior assessment and the data did not influence her very much, then her posterior probability would have been closer to 0.60

Table 29: Hypothetical sample prior weight elicitation results using two randomly selected rows from each conditional probability table and two hypothetical samples per row.

Child node (variable)	Parent node variable states (bin number)		Child variable state	Prior exp. value ($E[\theta_i]$)	Hypo-thetical data (x_i/n)	Elicited updated posterior ($E[\theta_i \mid x_i,n]$)	Calculated prior weight (α_0)
Change in generation of flow (Imp. surface)	Urban land cover						
	>12-36% (bin 2)		>2-8% (bin 2)	0.60	8/10	0.70	10
					5/10	0.55	10
	>66% (bin 4)		>8-16% (bin 3)	0.10	2/10	0.11	90
					7/10	0.20	50
	Mean ± within-method standard deviation						40 ± 38
Hydrology (Flashiness)	Imp. surface	Dam density					
	>8-16% (bin 3)	0 dams/100 km² (bin 1)	95+ rises (bin 4)	0.20	4/10	0.25	30
					1/10	0.15	10
	0-2% (bin 1)	>6 dams/100 km² (bin 3)	0-40 rises (bin 1)	0.93	5/10	0.90	133
					8/10	0.85	6.3
	Mean ± within-method standard deviation						45 ± 60
Habitat (Channel ratio)	Urban land cover	Flashiness					
	>66% (bin 4)	41-63 rises (bin 2)	0-25 (bin 1)	0.60	8/10	0.70	10
					2/10	0.55	70
	>36-66% (bin 3)	95+ rises (bin 4)	>33-44 (bin 3)	0.24	5/10	0.30	33
					2/10	0.20	0
	Mean ± within-method standard deviation						28 ± 31
Water quality (Conduc-tivity)	Urban land cover	Flashiness					
	0-12% (bin 1)	0-40 rises (bin 1)	0-92 µs (bin 1)	0.80	7/10	0.77	23
					9/10	0.83	23
	>66% (bin 4)	64-94 rises (bin 3)	>129-281 µs (bin 3)	0.35	3/10	0.32	6.7
					5/10	0.37	65
	Mean ± within-method standard deviation						30 ± 25

Table 29 continued: Hypothetical sample prior weight elicitation results using two randomly selected rows from each conditional probability table and two hypothetical samples per row.

Child node (variable)	Parent node variable states (bin number)			Child variable state	Prior exp. value ($E[\theta_i]$)	Hypothetical data (x_i/n)	Elicited updated posterior ($E[\theta_i \mid x_i,n]$)	Calculated prior weight (α_0)
EPT taxa richness	Flashiness	Chan. ratio	Conductivity					
	95+ rises (bin 4)	0-25 (bin 1)	>92-129 µs (bin 2)	>13 taxa (bin 4)	0.01	1/10	0.05	13
						3/10	0.05	63
	64-94 rises (bin 3)	0-25 (bin 1)	>281 µs (bin 4)	0-4 taxa (bin 1)	0.65	9/10	0.75	15
						4/10	0.60	40
	Mean ± within-method standard deviation							33 ± 24
Richness-weighted tolerance	Flashiness	Chan. ratio	Conductivity					
	0-40 rises (bin 1)	>44 (bin 4)	0-92 µs (bin 1)	<5 (bin 1)	0.88	10/10	0.90	50
						5/10	0.80	38
	64-94 rises (bin 3)	>44 (bin 4)	>281 µs (bin 4)	>5-5.6 (bin 2)	0.30	1/10	0.20	10
						4/10	0.35	10
	Mean ± within-method standard deviation							27 ± 20
Percent intolerant taxa	Flashiness	Chan. ratio	Conductivity					
	95+ rises (bin 4)	>33-44 (bin 3)	>281 µs (bin 4)	>8-18% (bin 3)	0.09	2/10	0.12	27
						3/10	0.15	25
	0-40 rises (bin 1)	>44 (bin 4)	>92-129 µs (bin 2)	>18% (bin 4)	0.70	6/10	0.65	10
						5/10	0.60	10
	Mean ± within-method standard deviation							18 ± 9.2

(from prior) even though the data showed a maximum likelihood estimate of 0.80. In this scenario, the expert may have reported an updated probability of 0.67, which would

result in a prior weight of 20, causing the prior information to have twice as much weight as the data (n=10).

In extreme situations, an expert may be so certain of his reported prior probability that, no matter what the hypothetical sample shows, he is not willing to change his prior at all. If the reported posterior is the same as the prior, no amount of data will be able to amend the prior. This result is, in practice, identical to the event of reporting a zero probability in the prior but occurs for a different mathematical reason. In this case, in order for a weighted average of data and prior to exactly equal the prior, the prior weight must equal infinity and, therefore, any real number weight of data will be irrelevant in comparison. As with zero conditional probability, the expert should be informed of the consequences of reporting an unchanged posterior during hypothetical sample elicitation. If the expert has total certainty in his prior, then an unchanged posterior reflects that. If, however, the expert has only near certain beliefs about a prior, he should report a slightly altered posterior that appropriately reflects his high but not absolute level of certainty.

Conversely, if an expert has no certainty in his prior to the point where he reports a posterior identical to the maximum likelihood of the data, this will result in a prior weight of zero. Opposite to the unchanging posterior situation, the only way for a weighted average of prior and data to equal exactly the data is for there to be no weight placed on the prior. However, if an expert truly has no prior knowledge about a

probability, he should assign a uniform distribution to that particular CPT row where every bin has an equal likelihood of occurring. If an expert assigns a non-uniform prior this means he believes he knows something about a conditional relationship, however little. Therefore, philosophically, a non-uniform prior should never be able to be overridden completely by data. If an expert provided a uniform prior, then placing zero prior weight on that uninformed prior is appropriate. However, if an expert provided a very uncertain but nonetheless not uniform prior, he should be instructed to assign the posterior in a manner which reflects the small amount of knowledge available in the prior and not a posterior equal to maximum likelihood of the data. If, upon reflection, the expert decides that he truly has no confidence in his prior knowledge relative to data then he should ask to amend his original non-uniform prior to a uniform prior.

This hypothetical sample method requires elicitation at the level of single reported prior probability expected value, $E[\theta_i]$. This model contains 960 such single expected values for which hypothetical sample prior weight could be elicited (since Dirichlet hyperparameter α_4 is not constrained by hyperparameters α_1, α_2 , and α_3 as multinomial parameter θ_4 is constrained by θ_1, θ_2 , and θ_3, resulting in only 720 θ_i parameters; Table 25). To evaluate the validity of this elicitation method in determining an expert's true prior weight, it would be prudent to use more than one hypothetical sample for every prior probability expected value and then check between samples for consistency. As such, it is impractical to elicit an updated posterior via multiple

applications of the hypothetical sample method for every one of the 960 prior expected values. Therefore, two prior probability expected values were randomly selected for each child node and updated posteriors were elicited for two random hypothetical samples per prior probability value, keeping a constant reference data weight (n=10). This process resulted in four separate prior weight calculations per node which were averaged to provide a single prior weight mean for each child node (Table 29).

As in the direct elicitation method, it was assumed that a single prior weight value can apply to all CPT rows in a child node. However, unlike direct elicitation, since hypothetical sample elicitation was performed multiple times for the same node, within-method variability can be estimated from the standard deviation of the four prior weight calculations (Table 29). This measure of variability can then be compared to the third prior weight elicitation method, probability range elicitation.

3.2.2.5.3. Probability range elicitation

A third possible prior weight elicitation method involves asking the expert to provide a range of probabilities that represent values that the expert believes a particular CPT parameter can have. If the expert is fairly certain about the value of a CPT probability, then, presumably, the expert does not believe that probability would take on very different values and reports a narrow range. This narrow range then translates into a high prior weight. Conversely, if an expert is not sure about a certain CPT probability, then he may believe that it could take on many possible values and reports a large range, consistent with a low prior weight. That is, the greater the range of possible

probabilities reported, the less certain the expert is about what the value of that probability should be, and the smaller the prior weight. Given a quantitative range as well as the reported probability, prior weight can be calculated from equations for the expected value and variance of a Dirichlet parameter distribution (Spiegelhalter et al. 1994, Cowell 1999, Jenson 2009). This is done by assuming that the reported probability represents the prior distribution mean (expected value) and the range represents an interval of one standard deviation in each direction around the original probability value provided (i.e., an interval two standard deviations total in length).

This interpretation of the elicited range was shown empirically (Spiegelhalter et al. 1994) in an experiment comparing 1000 probability judgments of medical experts against data (8000 observations). Reported ranges were assumed to represent an unknown number (x) of standard deviations around a reported probability mean. The correct value for x needed to accurately reflect the uncertainty of the expert knowledge. If x is too large, then most of the variability in a distribution occurs within the reported range, resulting in an overly precise prior. However, if x is too small, then the reported range only accounts for a small amount of the total variability and the prior is flatter than true uncertainty. Knowing that humans have difficulty in estimating true extremes (Morgan and Henrion 1990), the question is, how many standard deviations around the mean does a reported range represent? When the experts provided upper and lower bounds on their estimates, that range was found to correlate most closely with x=1

standard deviations from the mean in both directions. This was tested and confirmed using both Beta and Dirichlet distribution parameters (for binomial and multinomial data, respectively) and on alternate prior distribution shapes (such as log (θ/(1-θ)).

Eliciting a single probability and an uncertainty range of probabilities for a CPT parameter, θ_i, provides values for expected mean and variance. Prior weight can then be calculated via moment matching for the first two moments by finding a Dirichlet distribution that matches the reported mean and variance. Given $E[\theta_i]$ and $\text{var}[\theta_i]$ (Equations 54 and 55), one can solve for α_0.

Using the same example as in the *Hypothetical sample elicitation* section above, if the expert reported a prior probability of 0.60 for θ_2 in the medium-low urban land cover CPT row of the 'Change in generation of flow' child node, then the prior Dirichlet expected value (prior mean) for θ_2 is (from Equation 54):

$$\mu = E[\theta_2] = \frac{\alpha_2}{\alpha_0} = 0.60 \tag{66}$$

where the expected value of θ_2, $E[\theta_2]$, is assumed to represent the mean, μ, of the θ_2 Dirichlet prior distribution.

If the expert then reported a certainty range of 0.40 to 0.80, then

$$range = 2\sigma = 0.80 - 0.40 = 0.40 \tag{67}$$

And, therefore,

$$\sigma = (0.40/2) = 0.20 \tag{68}$$

From Dirichlet distribution properties (Equation 55),

$$\sigma^2 = \text{var}[\theta_2] = \frac{\alpha_2(\alpha_0 - \alpha_2)}{\alpha_0{}^2(\alpha_0 + 1)} = 0.20^2 \tag{69}$$

Solving Equations 66 and 69 for α_0, the prior weight:

$$\alpha_0 = \frac{\mu - \mu^2 - \sigma^2}{\sigma^2} = \frac{0.60 - 0.60^2 - 0.20^2}{0.20^2} = 5 \tag{70}$$

This prior weight of 5 is still reported in units of equivalent prior sample size. Using this probability range prior weight elicitation method, expert prior weight is half what was calculated using the hypothetical sample elicitation method. It is not clear why this occurs; prior elicitation method comparisons are sparse in the literature.

As discussed for the hypothetical sample elicitation method, probability ranges would have to be elicited 960 times in order to calculate a prior weight specific to each separate prior mean value. Again, this is impractical. Instead, probability ranges were elicited for each child node state in the same two randomly selected CPT rows as used for hypothetical sample elicitation so as to enable prior weight elicitation method comparison (Table 30). Probability range elicitation can be conducted considerably faster than hypothetical sample elicitation and, therefore, probability ranges were elicited for all four child node states in a selected CPT row rather than just the one used in hypothetical sample elicitation. This resulted in eight separate prior weight calculations per node. Although Cowell (1999) and Speigelhalter et al. (1994) suggest using the smallest calculated prior weight (i.e., minimum precision) if more than one

Table 30: Probability range prior weight elicitation results.

Child node (variable)	Parent node variable states (bin number)		Child variable state	Prior expected value (μ)	Elicited prior prob. range	Standard deviation from range (σ)	Calculated prior weight (α_0)
Change in generation of flow (Imp. surface)	Urban land cover						
	>12-36% (bin 2)		0-2%	0.09	0-0.20	0.10	7.19
			>2-8%	0.60	0.40-0.80	0.20	5.00
			>8-16%	0.30	0.20-0.40	0.10	20.00
			>16%	0.01	0-0.10	0.05	2.96
	>66% (bin 4)		0-2%	0.01	0-0.10	0.05	2.96
			>2-8%	0.09	0.05-0.20	0.08	13.56
			>8-16%	0.10	0.05-0.25	0.10	8.00
			>16%	0.80	0.60-0.95	0.18	4.22
	Mean ± within-method standard deviation						7.99 ± 5.97
Hydrology (Flashiness)	Imp. surface	Dam density					
	>8-16% (bin 3)	0 dams/100 km² (bin 1)	0-40 rises	0.05	0-0.15	0.08	7.44
			41-63 rises	0.15	0.05-0.30	0.13	7.16
			64-94 rises	0.60	0.40-0.80	0.20	5.00
			95+ rises	0.20	0.10-0.30	0.10	15.00
	0-2% (bin 1)	>6 dams/100 km² (bin 3)	0-40 rises	0.93	0.80-1	0.10	5.51
			41-63 rises	0.05	0.03-0.20	0.09	5.57
			64-94 rises	0.01	0-0.10	0.05	2.96
			95+ rises	0.01	0-0.10	0.05	2.96
	Mean ± within-method standard deviation						6.45 ± 3.83
Habitat (Channel ratio)	Urban land cover	Flashiness					
	>66% (bin 4)	41-63 rises (bin 2)	0-25	0.60	0.40-0.80	0.20	5.00
			>25-33	0.30	0.20-0.50	0.15	8.33
			>33-44	0.09	0.05-0.20	0.08	13.56
			>44	0.01	0-0.10	0.05	2.96
	>36-66% (bin 3)	95+ rises (bin 4)	0-25	0.25	0.10-0.35	0.13	11.00
			>25-33	0.50	0.30-0.70	0.20	5.25
			>33-44	0.24	0.10-0.40	0.15	7.11
			>44	0.01	0-0.08	0.04	5.19
	Mean ± within-method standard deviation						7.30 ± 3.52

Table 30 continued: Probability range prior weight elicitation results.

Child node (variable)	Parent node variable states (bin number)			Child variable state	Prior expected value (μ)	Elicited prior prob. range	Standard deviation from range (σ)	Calculated prior weight (α_0)
Water quality (Conductivity)	Urban land cover	Flashiness						
	0-12% (bin 1)	0-40 rises (bin 1)		0-92 μs	0.80	0.70-0.97	0.14	7.78
				>92-129 μs	0.10	0.05-0.35	0.15	3.00
				>129-281μs	0.05	0.01-0.15	0.07	8.69
				>281 μs	0.05	0.01-0.15	0.07	8.69
	>66% (bin 4)	64-94 rises (bin 3)		0-92 μs	0.20	0-0.25	0.13	9.24
				>92-129 μs	0.35	0.22-0.45	0.12	16.20
				>129-281μs	0.35	0.20-0.55	0.18	6.43
				>281 μs	0.10	0.01-0.25	0.12	5.25
	Mean ± within-method standard deviation							8.16 ± 3.87
EPT taxa richness	Flash-iness	Chan. ratio	Conduc-tivity					
	95+ rises (bin 4)	0-25 (bin 1)	>92-129 μs (bin 2)	0-4 taxa	0.55	0.45-0.75	0.15	10.00
				5-7 taxa	0.35	0.25-0.50	0.13	13.56
				8-12 taxa	0.09	0.01-0.15	0.07	15.71
				>13 taxa	0.01	0-0.10	0.05	2.96
	64-94 rises (bin 3)	0-25 (bin 1)	>281 μs (bin 4)	0-4 taxa	0.65	0.55-0.80	0.13	13.56
				5-7 taxa	0.30	0.25-0.55	0.15	8.33
				8-12 taxa	0.04	0.01-0.15	0.07	6.84
				>13 taxa	0.01	0-0.10	0.05	2.96
	Mean ± within-method standard deviation							9.24 ± 4.86
Richness-weighted tolerance	Flash-iness	Chan. ratio	Conduc-tivity					
	0-40 rises (bin 1)	>44 (bin 4)	0-92 μs (bin 1)	<5	0.88	0.75-0.95	0.10	9.56
				>5-5.6	0.10	0.05-0.25	0.10	8.00
				>5.6-6.2	0.01	0.01-0.15	0.07	1.02
				>6.2	0.01	0-0.10	0.05	2.96
	64-94 rises (bin 3)	>44 (bin 4)	>281 μs (bin 4)	<5	0.10	0-0.15	0.08	15.00
				>5-5.6	0.30	0.10-0.45	0.18	5.86
				>5.6-6.2	0.40	0.25-0.55	0.15	9.67
				>6.2	0.20	0.10-0.35	0.13	9.24
	Mean ± within-method standard deviation							7.66 ± 4.37

Table 30 continued: Probability range prior weight elicitation results.

Child node (variable)	Parent node variable states (bin number)			Child variable state	Prior expected value (μ)	Elicited prior prob. range	Standard deviation from range (σ)	Calculated prior weight (α_0)
Percent intolerant taxa	Flash-iness	Chan. ratio	Conduc-tivity					
	95+ rises (bin 4)	>33-44 (bin 3)	>281 μs (bin 4)	0-3%	0.55	0.45-0.75	0.15	10.00
				>3-8%	0.35	0.27-0.47	0.10	21.75
				>8-18%	0.09	0.01-0.18	0.09	10.34
				>18%	0.01	0-0.15	0.08	0.76
	0-40 rises (bin 1)	>44 (bin 4)	>92-129 μs (bin 2)	0-3%	0.01	0-0.1	0.05	2.96
				>3-8%	0.05	0.01-0.15	0.07	8.69
				>8-18%	0.24	0.10-0.40	0.15	7.11
				>18%	0.70	0.45-0.85	0.20	4.25
	Mean ± within-method standard deviation							8.23 ± 6.45

elicitation is performed per CPT row, neither study justifies this choice explicitly. Presumably, they implicitly acknowledge some degree of uncertainty in subjective prior weight elicitation and calculation, and, therefore, attempt to minimize the influence of prior weight. However, because there is no clear understanding of what is causing within-subject, within-method variability in prior weight response, there is no known reason why using the smallest prior weight would be any better or worse than using the largest prior weight. Therefore, we choose to average all eight prior weights calculated to provide a single prior weight mean for each child node. Averaging also maintains symmetry with hypothetical sample elicitation method treatment of prior weight. Within-method variability is calculated, as well, for comparison to hypothetical sample elicitation method.

Dirichlet distribution properties guarantee that the width of a range is evaluated relative to its mean value. For example, a standard deviation of 0.10 (i.e., range width of 0.20) around a mean of 0.09 translates to much greater relative uncertainty than a standard deviation of 0.10 around a mean of 0.30. Per two-moment-matching Dirichlet distribution calculations, this constant standard deviation but difference in mean translates into a lower prior weight of 7.19 for the former and a prior weight of 20 for the later (Table 30, first and third rows). These calculations quantify the intuitive sense that 0.09±0.10 provides a much less certain prediction of the value of θ_1 than 0.30±0.10 provides for θ_3. More specifically, per the shape of the Dirichlet distribution, ~86% of the probability density of θ_1 is within a 0.10 standard deviation of 0.09 while only ~67% of the probability density of θ_3 is within a 0.10 standard deviation for 0.30 (Figure 52).

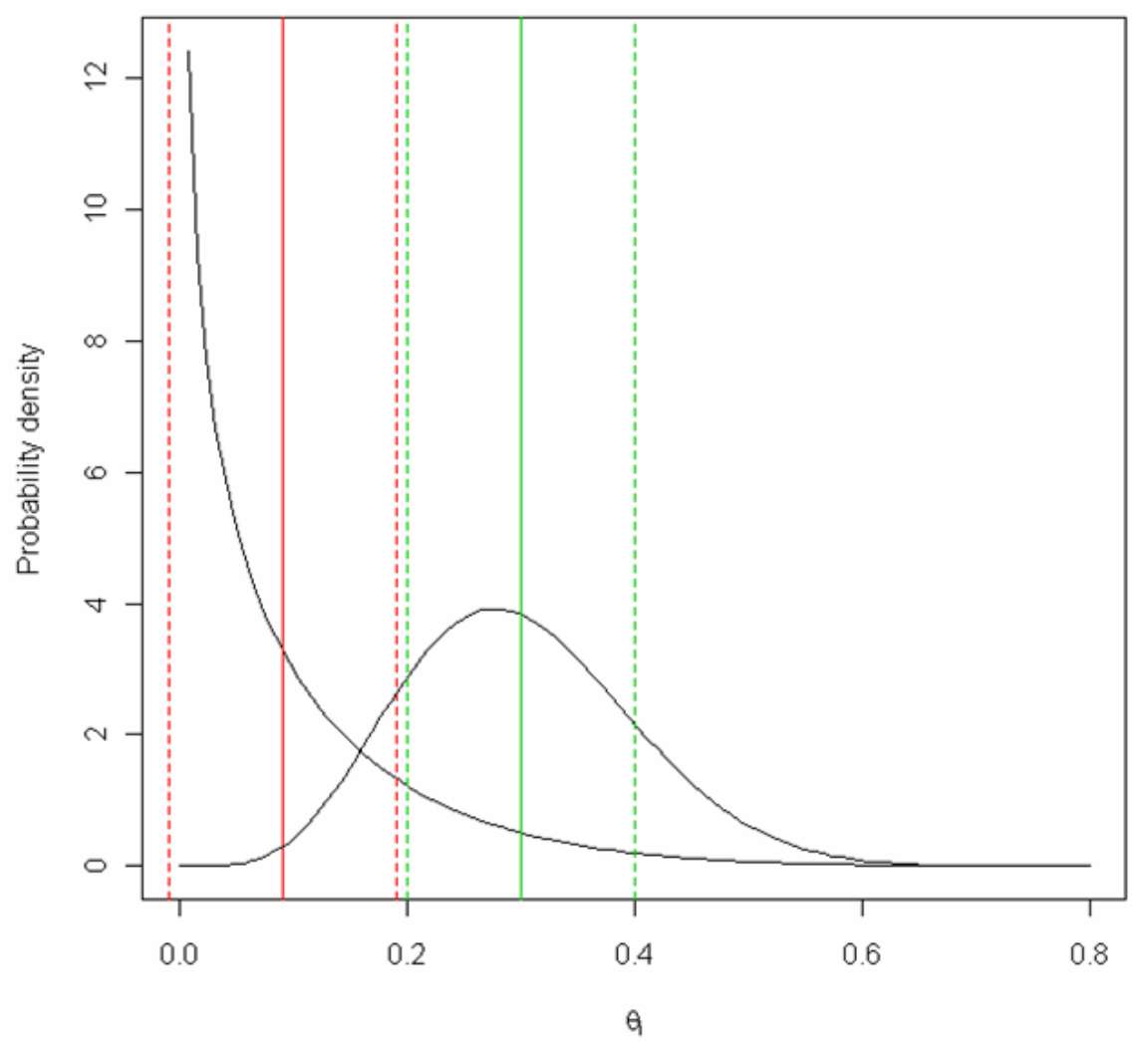

Figure 52: Different amounts of probability density covered by same standard deviation depending on where mean is for Dirichlet distribution. [Red is mean of θ_i = 0.09; green is mean of θ_i is 0.30; σ = 0.1; a 2σ interval around the mean in both cases shown by respectively colored dashed lines. Each mean is centered on its respective θ_i distribution shown in black.]

Since a 0.10 standard deviation defines level of uncertainty in a prior estimate for each parameter, the parameter distribution for which that 0.10 standard deviation occupies more probability weight should be associated with more uncertainty and, therefore, less prior weight. In fact, for the prior distribution on θ_1 to have a prior weight of 20, the

range-defined standard deviation would have to decrease from 0.10 to 0.063 to account for the more extreme location of the θ_1 mean. That is, to assign a prior weight of 20 equivalent samples, an expert would have to report a smaller range of 0.027-0.153 around a mean probability of 0.09 ($\sigma = 0.063$) relative to the larger range of 0.20-0.40 ($\sigma = 0.10$) around a mean of 0.30 that results in the same $\alpha_0 = 20$ prior weight.

The probability range method assumes an expert's natural human thinking mimics these mathematical principles of the extremes, where by instinct a smaller range would be assigned to extreme probabilities to show an equal amount of prior knowledge as would assigning a larger range for more middle probabilities. Spiegelhalter et al. (1994) empirically verified that this (or something that approximates it) did happen with their many elicitations but it is not clear whether this assumption about innate (or even subconscious) human reasoning processes is theoretically correct.

As with CPT elicitation and hypothetical sample elicitation, extreme situations specific to probability range elicitation require caution. If an expert is so certain of his reported mean probability that he does not provide a possible range (essentially reporting a range of 0.60-0.60 around a mean of 0.60), this would result in a prior weight of infinity and no amount of data could ever update that 0.60. If this is not what the expert intends, then he should provide a range whose small length accurately reflects his small uncertainty.

Prior weights elicited via the three methods described above produced different prior weight calculations for the same nodes (Table 31). In all three methods, a

Table 31: Summary of mean prior weight (α_0) and within-elicitation-method standard deviation using different elicitation methods.

Node		Direct elicitation	Hypothetical sample	Probability range
Change in generation of flow	mean α_0	21.25	40.00	7.99
	sd. of α_0		38.30	5.97
Hydrology	mean α_0	5.67	44.90	6.45
	sd. of α_0		59.87	3.83
Habitat	mean α_0	2.66	28.33	7.30
	sd. of α_0		31.09	3.52
Water Quality	mean α_0	3.19	29.58	8.16
	sd. of α_0		24.88	3.87
EPT taxa richness	mean α_0	0.93	32.50	9.24
	sd. of α_0		23.54	4.86
Richness-weighted tolerance	mean α_0	0.80	26.88	7.66
	sd. of α_0		20.14	4.37
Percent intolerant taxa	mean α_0	0.93	17.92	8.23
	sd. of α_0		9.17	6.45

node-averaged α_0 value was assigned to all CPT rows for a node. Therefore, within-node CPT row sample size variability was the same for all three methods (although reported only for direct elicitation, Table 28). This type of variability is derived only from different amounts of data points in each CPT row and is not an evaluation of the prior elicitation methods themselves. Within-method variability, however, could be calculated for hypothetical sample method and probability range method because these two methods each elicited more than one separate prior weight estimate for each node. No within-elicitation-method variability is available for direct elicitation as each node's

value was only elicited once. Between the two methods for which within-method variability could be calculated, hypothetical sample elicitation produced from 1.5 to 15.6 the amount of within-method variability as the probability range method. Per node standard deviation within prior weight elicitation method (s.d. 3.52-6.45 for range method; 9.17-59.87 for hypothetical sample method, Table 31) is greater than per node deviation between number of samples in each CPT row (s.d. 0.96-3.40,Table 28). This standard deviation size difference means that variations between CPT rows within a node are not affecting consistency of prior weight calculations nearly as much as the elicitation methods themselves.

It is not clear from the literature which prior elicitation method to choose in which kind of scenario, types of decision criteria to use, or even simple benefits and drawbacks of these three methods. In this project's comparison, the probability range elicitation method showed lower within-method variability and the most consistent prior weights among nodes. Additionally, between all three methods, the expert reported feeling most comfortable providing the type of information the probability range method required over the other two methods, and believed that reporting ranges expressed uncertainty more intuitively and accurately than reporting equivalent data points of prior worth or hypothetically updated posteriors. Therefore, we chose to use the prior weights elicited and calculated via the probability range method in further model development. Prior weight for parent nodes with no children (and hence no

elicited CPTs), 'Urban disturbance' and 'Hydrologic Modifications', were each assigned a prior weight of 1. Marginal distributions on these nodes were not expected to change with data updating since binning was based on data and no other information was included (i.e., they are not conditional on anything).

3.2.3. Posterior calculation

Once development of the prior model is complete and both conditional probability tables (CPTs) and prior weights are quantified, available data can be incorporated into the prior model using conjugate Dirichlet-multinomial updating (as described in the *Bayesian networks* section above) to create the posterior model. In order to combine prior beliefs with data, the data must be discretized and reformatted into tables which match the discrete arrangement of the prior CPTs. Hugin software (Hugin 2008) performs this discretization and prior updating automatically via the Adaptation subroutine, however, discussing this process is instructive. Each of the 85 available EUSE Southeast sampling sites has a measure of each of the nine final node variables. However, not all variables will be relevant for every node. For the two marginal parent nodes which are not themselves child nodes ('Urban disturbance' and 'Hydrologic modifications'), only the variable representing each marginal node is relevant. Upon updating, the 85 measures of urban land cover and the 85 measures of dam density will be discretized into their appropriate 4 and 3 node states, respectively, and number of total data points belonging to each bin will be summed. Data tables for these two nodes

show relatively uniform distribution of samples (for example, the 'Urban disturbance' node data table shows, of the 85 total samples, 21 samples have urban land cover between 0 and 12%, 22 between 12 and 36%, 21 between 36 and 66%, and 21 have urban land cover greater than 66%, Appendix B Table 71). This is not surprising since bin endpoints were determined from the data based on equal allotment to each bin by percentile.

Child node data tables are more complicated because data need to be discretized and divided by not only child state but also parent state combination. For each node's data table, all 85 available EUSE Southeast samples are essentially divided and categorized in a different way, depending on a specific child node's parent and child variables of interest. As with prior CPTs, child nodes with more states which are conditional on parent nodes with more parent state combinations result in more extensive data tables. The goal is to compare what the data say about a particular relationship between a child node and its parents to prior knowledge about the same relationship, parameterized in the same conditional way. For example, the 'Hydrology' child node data table (Table 32) has 12 rows, one for each of the possible state combinations of the 'Hydrology' node's two parents. The 85 available EUSE Southeast sampling sites are divided into 12 groups based on which discrete bin each sampling site's impervious surface and dam density values belong to. If a particular sampled stream was in a watershed that had 1% impervious surface and 0 dams/100 km^2, that

Table 32: Data table for "Hydrology" node (flashiness counts and *n* total number of samples per parent state combination).

Impervious surface	Dam density	Flashiness greater than 5 times the median				
		0-40 rises	41-63 rises	64-94 rises	95+ rises	*n*
0-2 %	0 dams/100 km^2	6	2	1	0	9
>2-8%	0 dams/100 km^2	4	2	1	0	7
>8-16%	0 dams/100 km^2	3	4	2	0	9
>16%	0 dams/100 km^2	3	1	2	4	10
0-2 %	>0-6 dams/100 km^2	2	7	1	0	10
>2-8%	>0-6 dams/100 km^2	0	1	4	1	6
>8-16%	>0-6 dams/100 km^2	1	0	0	4	5
>16%	>0-6 dams/100 km^2	0	0	0	2	2
0-2 %	>6 dams/100 km^2	1	1	0	0	2
>2-8%	>6 dams/100 km^2	2	2	5	0	9
>8-16%	>6 dams/100 km^2	0	0	3	4	7
>16%	>6 dams/100 km^2	0	1	2	6	9

stream would be assigned to the first row of the 'Hydrology' data table, and so on. After this first division of the 85 samples, we can calculate the *n* values for every row where *n* is the number of samples that have that row's combination of parent values. Each such row is considered its own multinomial distribution across possible child node states and, as such, this *n* represents the total sample size (the sum of all counts) in the multinomial likelihood function (Equation 38). The first row of the 'Hydrology' data table has *n*=9 total samples which have both impervious surface between 0 and 2% of watershed area and 0 dams/100 km^2.

All the *n* values in a data table should always add up to 85, the total number of data points. This is because we are working with a complete dataset where each of the 85 samples has a value for each of the 9 variables. In cases where data are missing,

several algorithms have been developed to estimate those missing values (such as the Expectation Maximization (EM) algorithm (Lauritzen 1995) or the Expectation Maximization-model selection (EM-MS) algorithm (Friedman 1997b)), however, this is outside of the scope of this project. In our case, all of the 85 data points are used to calculate a data table for each of the child nodes. Additionally, if desired, a completely different dataset could be used to calculate each separate data table. The only requirement for creating a child node's data table is the measurement of values for a child node's child and parent variables. Each separately created data table could be used to update only its appropriate prior CPT and nothing else. This is one of the benefits of Bayesian network decomposability, since it is often incredibly challenging to find a single dataset that measures all Bayesian network variables. More often, data from a single study may be available for some nodes but not others, and several different datasets must be acquired. Fortunately, the properties of Bayesian networks allow separate parameterization for each model child node. The EUSE dataset offers a rare opportunity to parameterize all of a network's data tables from one comprehensive dataset.

Once the data are split by child node parent state combination, within each parent combination, data points are further subdivided by child state. Again, using the first row of the 'Hydrology' node as an example, of the 9 sampling sites that have lowest impervious surface and lowest dam density, 6 have low flashiness, 2 have medium-low

flashiness, 1 has medium-high flashiness and none have high flashiness. These four counts (6, 2, 1, 0) represent the x_1, x_2, x_3, and x_4 values of the multinomial likelihood function (Equation 82) describing this table row. They are the number of samples that fall in each of the four child node bins. Data are discretized and divided this way for each data table row in each child node data table (Appendix B Table 71-Table 78). Having separate assessments of the distribution of a child node conditional on its parents, one from prior knowledge and one from discretized data, allows comparison of the current state of scientific thinking about ecological processes to field measurements of those processes as stipulated by the scientific method.

To directly compare the prior to data in probability units, data counts must be converted to multinomial maximum likelihood estimates (Equation 51). Again looking at the first row of the 'Hydrology' node data table, in maximum likelihood probability units, the data show that 67% of streams with low impervious surface and low dam density have low (>40 rises greater than five times the median) flashiness (Table 33). The prior derived from expert knowledge expects 88% of such low impervious surface and low dam density streams to have low flashiness. Probabilities of medium-low, medium-high, and high flashiness decrease from 22% to 11% to 0% in the data, while the expert believed them to be 10%, 1%, and 1%, respectively. Overall, data match expert opinion in trend direction, differing slightly in magnitude. Such comparisons can be drawn for any of the 242 rows in the full model. Hugin software also offers graphical

Table 33: Data table for "Hydrology" node (converted to maximum likelihood estimate units and n total number of samples per parent state combination).

Impervious surface	Dam density	Flashiness greater than 5 times the median				
		0-40 rises	41-63 rises	64-94 rises	95+ rises	*n*
0-2 %	0 dams/100 km^2	0.6667	0.2222	0.1111	0.0000	9
>2-8%	0 dams/100 km^2	0.5714	0.2857	0.1429	0.0000	7
>8-16%	0 dams/100 km^2	0.3333	0.4444	0.2222	0.0000	9
>16%	0 dams/100 km^2	0.3000	0.1000	0.2000	0.4000	10
0-2 %	>0-6 dams/100 km^2	0.2000	0.7000	0.1000	0.0000	10
>2-8%	>0-6 dams/100 km^2	0.0000	0.1667	0.6667	0.1667	6
>8-16%	>0-6 dams/100 km^2	0.2000	0.0000	0.0000	0.8000	5
>16%	>0-6 dams/100 km^2	0.0000	0.0000	0.0000	1.0000	2
0-2 %	>6 dams/100 km^2	0.5000	0.5000	0.0000	0.0000	2
>2-8%	>6 dams/100 km^2	0.2222	0.2222	0.5556	0.0000	9
>8-16%	>6 dams/100 km^2	0.0000	0.0000	0.4286	0.5714	7
>16%	>6 dams/100 km^2	0.0000	0.1111	0.2222	0.6667	9

displays of CPTs and maximum likelihood data tables (obtained by 'adapting' a blank prior with data) via probability color shading for visual comparison. Due to the size and complexity of discrete multidimensional model structure, Bayesian network CPT representation is still a nascent research field.

In order to incorporate all available knowledge about the system in the model, prior knowledge is then combined with data to create a posterior version of the parameterized Bayesian network. Each type of information is appropriately weighted according to its strength of evidence (either certainty of expert knowledge or sample size of data), and the posterior model comprehensively reports the best possible combination of evidence. Posterior calculation occurs through Dirichlet-multinomial updating at the level of table row, with a separate set of parameters for each row. Using the first row of

the 'Hydrology' node example, the posterior expected value for θ_1 given the data, x_1 and n, is (Equation 59):

$$E[\theta_1 \mid x_1, n] = \frac{\alpha_1 + x_1}{\alpha_0 + n} = \frac{5.68 + 6}{6.45 + 9} = 0.76 \tag{71}$$

where α_1 from the prior can be calculated by solving for α_1 knowing the expected value of the prior, $E[\theta_1]$ and the prior weight, α_0 (Equation 54):

$$E[\theta_i] = \frac{\alpha_i}{\alpha_0} = 0.88 = \frac{\alpha_1}{6.45} \tag{72}$$

The posterior expected value of θ_1, 0.76 (Table 34), is a weighted average of the prior expected value, 0.88, and the data maximum likelihood value, 0.67, weighted

Table 34: Posterior conditional probability table and posterior weights (α_0+n) for "Hydrology" node (flashiness bins in probability units and α0+n in equivalent data points).

Impervious surface	Dam density	Flashiness greater than 5 times the median				
		0-40 rises	41-63 rises	64-94 rises	95+ rises	α_0+n
0-2 %	0 dams/100 km²	0.7557	0.1712	0.0689	0.0042	15.45
>2-8%	0 dams/100 km²	0.3885	0.4604	0.1463	0.0048	13.45
>8-16%	0 dams/100 km²	0.2150	0.3215	0.3799	0.0835	15.45
>16%	0 dams/100 km²	0.1863	0.0765	0.1804	0.5568	16.45
0-2 %	>0-6 dams/100 km²	0.4862	0.4451	0.0647	0.0039	16.45
>2-8%	>0-6 dams/100 km²	0.2072	0.3601	0.3472	0.0855	12.45
>8-16%	>0-6 dams/100 km²	0.1155	0.2253	0.2817	0.3775	11.45
>16%	>0-6 dams/100 km²	0.0382	0.0763	0.2290	0.6565	8.45
0-2 %	>6 dams/100 km²	0.8282	0.1565	0.0076	0.0076	8.45
>2-8%	>6 dams/100 km²	0.3591	0.2923	0.3445	0.0042	15.45
>8-16%	>6 dams/100 km²	0.0719	0.2398	0.3861	0.3022	13.45
>16%	>6 dams/100 km²	0.0417	0.1482	0.2964	0.5136	15.45

slightly greater towards the data which have a greater sample size, 9, than the prior weight, 6.45. In full form, θ_1 actually has a whole distribution of values (not just a single expected value) which can be described via a full Dirichlet prior distribution, multinomial likelihood, and Dirichlet posterior distribution. However, for purposes of parsimony, Hugin interface reports only the single number expected values and maximum likelihood estimates instead of providing the entire distribution for every parameter. Such expected value posterior conditional probability tables are calculated for every child node in the model (Appendix B Table 79-Table 86).

Due to the adaptive nature of Bayes Theorem, if more data were to be collected about the nine variables of this Southeast stream urbanization model, the current posteriors would become the new priors and the model could be further informed. This flexible construct allows the model to 'learn' as more data are incorporated. Using this modeling framework, we can quantify the process of scientific hypothesis creation and verification in a transparent manner. This means clearly presenting which expert prior assumptions were included in which iteration of the model, instead of concealing implicit expert input and falsely calling model building 'objective'. Over time, as more data on the system become available, they will override incorrect scientific assumptions and fortify correct ones leading to a much more systematic implementation of the scientific process.

3.2.4. Model evaluation

Once the Bayesian network model is built, it is useful to evaluate its content and performance. Three major aspects of Bayesian network evaluation include assessment of (1) the strength of links between nodes, (2) the sensitivity of endpoints to changes in driver nodes, and (3) the comparison of prior information to data both in terms of node value magnitude and direction of relationship between child and parent nodes.

3.2.4.1. Mutual information

To better understand and interpret graphical modeling output, it may be prudent to evaluate the strength of the connections between the nodes of a Bayesian network (Boerlage 1992). One method of evaluating these linkage strengths is the calculation of mutual information (Nicholson and Jitnah 1998). Mutual information is essentially a measure of the dependence of two variables. This dependence can be quantified by comparing the joint distribution of the two variables with the product of their two marginal distributions. Per basic statistics,

$$p(X,Y) = p(X)p(Y) \tag{73}$$

if variables X and Y are independent. However, if X and Y are not independent, then their joint distribution and the product of their marginals will not be equal. Calculating the Kullback-Leibler divergence (Kullback 1968) between the joint distribution for X and Y and the product of each marginal distribution gives an estimate of the degree to which

X and Y are dependent. This application of Kullback-Leibler divergence is known as mutual information:

$$MI(X,Y) = \sum_{x,y} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{74}$$

If p(X,Y) and p(X)p(Y) do not diverge, then mutual information is zero and X and Y are independent. In this case, knowing X does not give any information on Y. The further p(X,Y) and p(X)p(Y) diverge, the more mutually dependent X and Y are and the greater the mutual information, maximizing when there is an exact functional form between X and Y. A measure of mutual information, then, can be used to evaluate which nodes make the most informative connections in a Bayesian network. Nodes with greater mutual information between them have a stronger link between their variables.

To calculate mutual information for a discrete Bayesian network built of conditional probabilities, we convert joint probabilities with the property:

$$p(X,Y) = p(X)p(Y \mid X) \tag{75}$$

This restatement in terms of conditional and marginal probabilities allows us to rewrite the mutual information formula as (Nicholson and Jitnah 1998):

$$MI(X,Y) = \sum_{x} p(x) \sum_{y} p(y \mid x) \log\left(\frac{p(y \mid x)}{p(y)} \right) \tag{76}$$

Since the Bayesian network encodes the model in terms of conditional and marginal probabilities of parent and child nodes, this formula can be used to calculate

the mutual information between parent node X and child node Y, both of which are represented by discrete variables with four bins, as follows:

$$MI(X,Y) = \sum_{i=1}^{4} p(x_i) \sum_{j=1}^{4} p(y_j \mid x_i) \log\left(\frac{p(y_j \mid x_i)}{p(y_j)} \right) \tag{77}$$

By summing permutations of the single probabilities that parameterize the Bayesian network over each discrete state of parent and child nodes, we can calculate the mutual information for that parent-child link. For example, to calculate the mutual information between the 'Change in generation of flow' node (Y) and the 'Urban disturbance' node (X), where Y is conditional on X in the posterior network:

$$\begin{aligned}
MI(X,Y) = {} & \\
& p(x_1)\left(\begin{array}{l} p(y_1 \mid x_1)\log\left(\frac{p(y_1 \mid x_1)}{p(y_1)} \right) + p(y_2 \mid x_1)\log\left(\frac{p(y_2 \mid x_1)}{p(y_2)} \right) + \\ p(y_3 \mid x_1)\log\left(\frac{p(y_3 \mid x_1)}{p(y_3)} \right) + p(y_4 \mid x_1)\log\left(\frac{p(y_4 \mid x_1)}{p(y_4)} \right) \end{array} \right) + \\
& p(x_2)\left(\begin{array}{l} p(y_1 \mid x_2)\log\left(\frac{p(y_1 \mid x_2)}{p(y_1)} \right) + p(y_2 \mid x_2)\log\left(\frac{p(y_2 \mid x_2)}{p(y_2)} \right) + \\ p(y_3 \mid x_2)\log\left(\frac{p(y_3 \mid x_1)}{p(y_3)} \right) + p(y_4 \mid x_2)\log\left(\frac{p(y_4 \mid x_2)}{p(y_4)} \right) \end{array} \right) + \\
& p(x_3)\left(\begin{array}{l} p(y_1 \mid x_3)\log\left(\frac{p(y_1 \mid x_3)}{p(y_1)} \right) + p(y_2 \mid x_3)\log\left(\frac{p(y_2 \mid x_3)}{p(y_2)} \right) + \\ p(y_3 \mid x_3)\log\left(\frac{p(y_3 \mid x_3)}{p(y_3)} \right) + p(y_4 \mid x_3)\log\left(\frac{p(y_4 \mid x_3)}{p(y_4)} \right) \end{array} \right) + \\
& p(x_4)\left(\begin{array}{l} p(y_1 \mid x_4)\log\left(\frac{p(y_1 \mid x_4)}{p(y_1)} \right) + p(y_2 \mid x_4)\log\left(\frac{p(y_2 \mid x_4)}{p(y_2)} \right) + \\ p(y_3 \mid x_4)\log\left(\frac{p(y_3 \mid x_4)}{p(y_3)} \right) + p(y_4 \mid x_4)\log\left(\frac{p(y_4 \mid x_4)}{p(y_4)} \right) \end{array} \right) = 0.63
\end{aligned} \tag{78}$$

Mutual information of different parents that have the same child node is additive (Nicholson and Jitnah 1998), so mutual information measures of all parent nodes of a single child node can simply be summed to assess the total mutual information between a child node and its parents (deCampos 2006). In this way, we can use mutual information in order to measure the degree of interaction between each variable and its parent variables in the network. Unfortunately, similar to the properties of r-squared, mutual information always increases with the number of parents. In models with many parents for each node, mutual information should be evaluated in context of a model size penalty term to avoid falsely inflating the amount of information contributed by additional parent nodes. Possible solutions to this problem include incorporation of a chi-squared punishment term (deCampos 2006) or evaluation of total weighted mutual information for sub-systems of connected nodes (Nicholson and Jitnah 1998). In the southeast Bayesian network model, by design, we limited parents to no more than three parents per child nodes. Therefore, we do not anticipate a large enough effect on mutual information from increasing number of parents to warrant correcting for it.

3.2.4.2. Sensitivity analysis

Another information strength inquiry of concern in Bayesian network evaluation is the assessment of the relative influence of model drivers on desired endpoints. This is important if the model is to be used in a management context, that is, to identify improvements to nodes that would result in the greatest proverbial 'bang for the buck'.

Typically, when this type of question is answered in traditional model evaluation contexts, each model 'input' is evaluated independently for degree of influence on output (Saltelli et al. 2000). While this approach is reasonable for controlled scenarios, for example engineering design models, in an environmental system input drivers do not act alone. Often they act in concert and it is difficult if not impossible to distinguish their effects, as one could distinguish independent vector components in a physics model. Therefore, it is not realistic to perform environmental model sensitivity analysis by holding all variables but the driver of interest constant and observing endpoint changes. It is much more desirable to take the response of the system into account, acknowledging that variables are related, when evaluating the effect of each driver.

A Bayesian network model can be used to assess such credible system-wide response. Additionally, Bayesian network sensitivity analysis incorporates all the uncertainties which are encoded by the network when predicting driver impacts on endpoints, and results are reported probabilistically for clearer characterization. Driver assessment sensitivity analysis was performed on the completed posterior Southeast Bayesian network in a relatively straightforward manner. Each driver was varied from 'worst' state associated with negatively impacted environmental conditions to 'best' state associated with natural conditions, and predicted changes in the three invertebrate metrics 'best' category were assessed. By varying each driver node in this manner, the

influence of each driver node on the three invertebrate metrics was determined both in terms of absolute endpoint likelihood and percent improvement.

3.2.4.3. Comparison of prior to posterior

Finally, the construct of this Bayesian network model affords a unique opportunity to quantitatively compare the information elicited from an expert to the information collected in the data via changes to the posterior with updating. Given the multidimensional nature of the model, particularly as the number of parent nodes per child node increases, this evaluation task is challenging to represent two-dimensionally. To attempt to synthesize the amount of information available in this model, standardized weighted averages are computed for every CPT row of each child node. This is done by assigning numbers 1-4 to each of the four bins of every child node and weighing each of the bins in a CPT row by its respective probability (for prior and posterior) or maximum likelihood estimate (for data). The resulting weighted averages can range from 1 to 4 between the two extremes of all probability weight assigned to the first or last bin, respectively. Each node is reported on this standardized scale for ease of interpretation, regardless of actual bin values in node specific units.

Each child node's standardized scores are then plotted across discrete gradients of changing parent node states for the prior, the data, and the posterior. From these plots, we observe differences in average child node score between what was expected by the expert and what the data measured, as well as difference in direction of change in

the child node score in response to changes in the parent node(s) states. Additionally, these graphs allows better visualization of how the posterior is itself a weighted average between the prior and the data, and highlight where the prior is weighted more than the data and vice versa.

3.3. Results

The final prior Bayesian network combined elicited model structure, variables, CPTs, and prior weights into one representation of node relationships and uncertainties (Figure 53). The model representation shows marginal discrete distributions for each of the nine nodes and arrows connecting the nodes as determined by the model structure. Conditional probability and prior weight information is embedded in the arrows pointing from parent to child nodes, and depicted marginal distributions are calculated from basic statistical properties relating conditional to marginal probabilities combined with principles of graph theory. Each node is divided into the expert-determined number of states (or bins) and each state is defined by the expert-determined bin endpoints. For each node, the first column of horizontal bars visually shows the probability of a Southeast sample falling in each node state so that the model user can see the distribution of samples between states of that node. Longer bars correspond to higher probabilities. The second column provides the actual numeric value for the probability of each state, and the third column defines the bin endpoints for each state.

Mean (μ) and variance (σ^2) values of each node's discrete marginal distribution are also provided in the Hugin model representation.

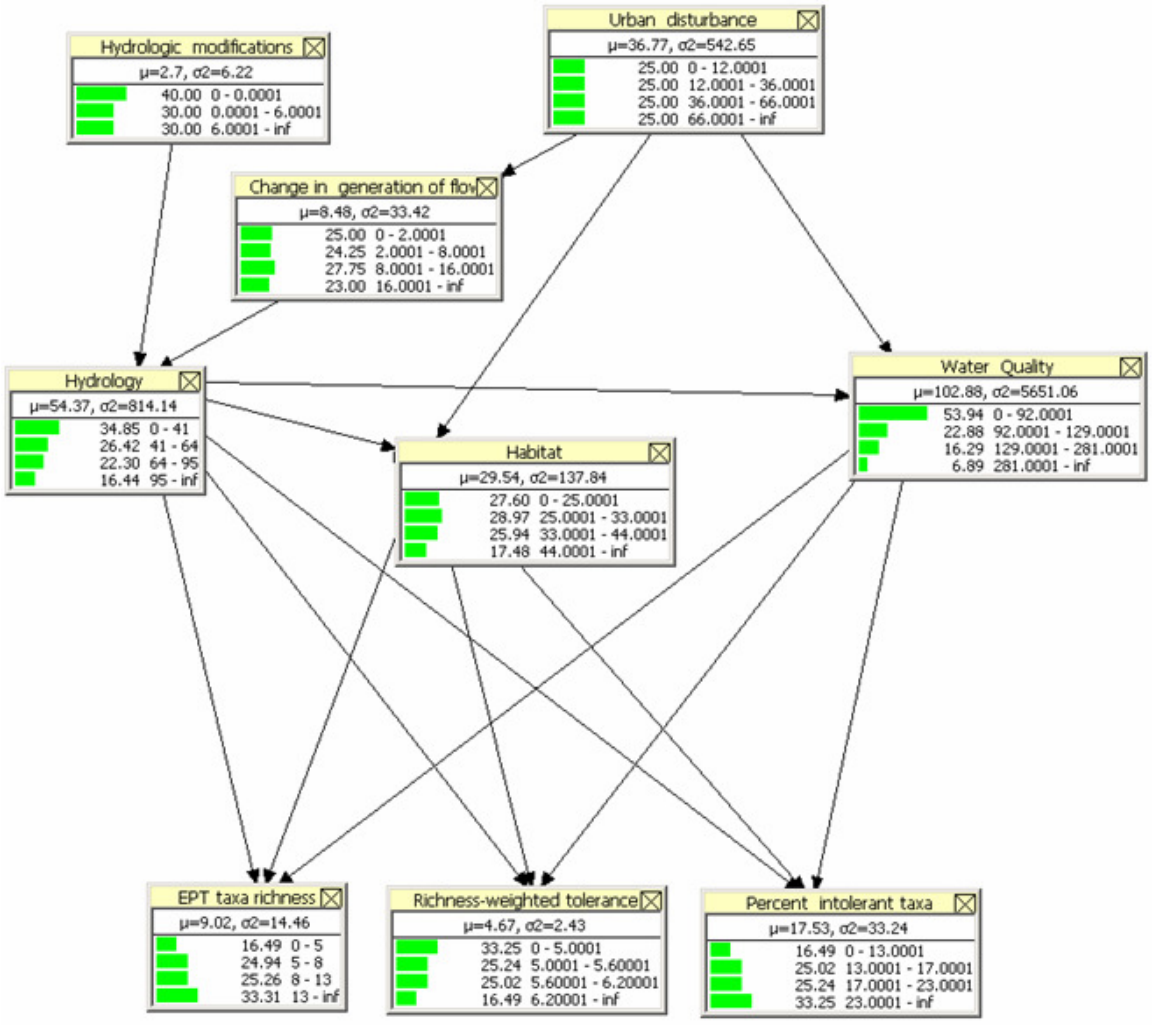

Figure 53: Prior Bayesian network marginal distributions for the southeast United States.

Typically, in environmental Bayesian network studies, the prior model marks the end of network development and analysis is performed on this prior model alone (Borsuk et al. 2004, McNay et al. 2006, Ticehurst et al. 2007). In this Bayesian network

creation, we went further than just developing a prior model by fully implementing Bayes theorem. To do so we used 85 data points to Bayesian update the expert-developed prior model. When looking at the data binned and tabulated (Table 32, Appendix B Table 71-Table 78), it is clear why expert-created priors are especially useful in Bayesian network construction. Even in this simplified network with only nine nodes, at least three situations arise where a child node with four bins is predicted from three parent nodes of four bins each, resulting in 64 possible parent state combinations (child nodes 'EPT richness', 'Richness-weighted tolerance', and 'Percent intolerant taxa' are three such nodes). The 85 data available points are inadequate to completely characterize this many CPT rows (e.g., for 'EPT richness' 19 of 64 CPT data rows are empty (n=0) because not even one observation exists to fill them, Appendix B Table 76). If the prior Bayesian network model was parameterized using only these data, in addition to not being fully Bayesian, the model would consist of several CPT rows containing no information. That is, the CPT rows for which no data is available would have flat distributions across all child node states (i.e., each child node state is described as equally likely). If future data were to be obtained, these distributions can be refined; but, in the current problem context, they would be uninformed.

An even more potentially damaging result of using only data to parameterize Bayesian network CPTs are the child node states assigned zero prior probability by the data. As mentioned in the *Conditional probability table elicitation* section, if a prior CPT

cell has a zero probability (i.e., a peaked distribution across child node states with one or more child node states at zero), that cell can never be updated with additional data (because a zero prior times any data always equals zero). To calculate accurate prior conditional probabilities from data alone, a larger sample size would be required. Therefore, prior information in combination with data is vital to establish relationships between nodes which cannot be established only from data. Very few Bayesian networks reported in the literature have taken the additional step of combining these different sources of information (expert knowledge and data) in a truly Bayesian manner.

Once expert prior information is updated with the data into a posterior Bayesian network (Figure 54), it can be interactively manipulated to see the effect of changes in one node on the rest of the system. This hypothetical evidence can be propagated in the forward (predictive) or backward (diagnostic) direction. We can also calculate various metrics which evaluate the contributions of nodes and connections (mutual information), determine which node has the greatest impact on improving desired outcomes (sensitivity analysis), or quantitatively compare information provided by the expert with the data (comparison of prior to posterior). Bayesian networks have not been well-evaluated in the literature and, therefore, we attempt to develop such missing effective evaluation procedures to determine model goodness.

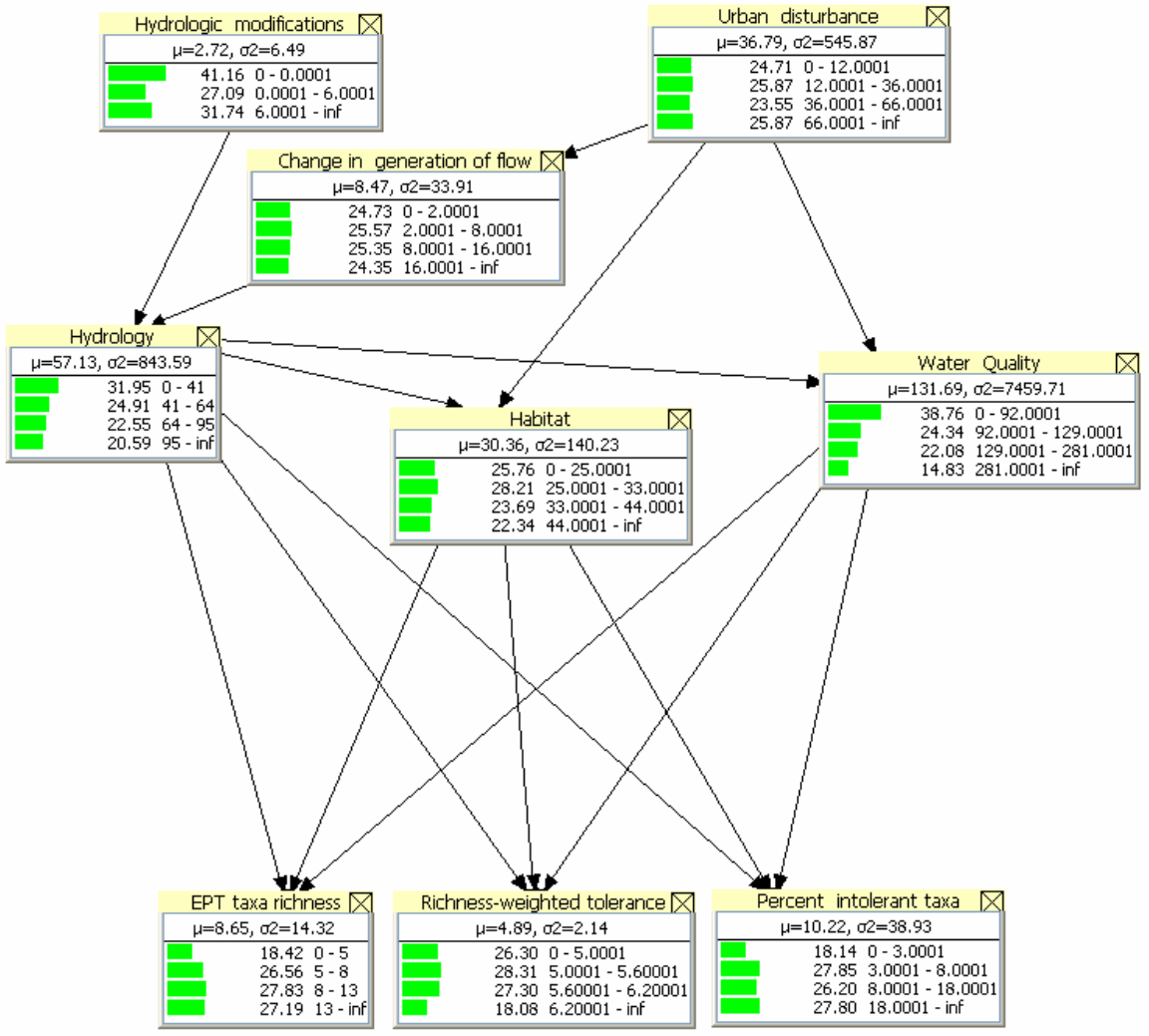

Figure 54: Posterior Bayesian network marginal distributions for the southeast United States.

3.3.1. Evidence propagation: predictive and diagnostic probabilities

Ultimately, a Bayesian network is just a model quantifying the probabilities of each node being in a particular state relative to the likely states of all other nodes. As described in the *Bayesian networks* section above, changing the marginal value of one or more nodes in a Bayesian network will revise the marginal probability distributions of

the remaining nodes in the network using Bayes law, conditional on model structure and new 'evidence' (i.e., the changed marginal value). Evidence propagation does not change conditional probability tables themselves, as Bayesian updating with new data does; it simply predicts changes in marginal probabilities given the relationships between the nodes defined by the conditional probability tables. These revised marginal probabilities then reflect probabilistically the influence of evidence on another part of the system, taking into account the model structure and all the information encoded in the model's probability tables. They are able to depict how the entire system is related and how it would respond to change. Through this evidence propagation, we can evaluate both predictive and diagnostic probabilities.

Predictively, we can ask: given a certain level of urbanization, what distribution of macroinvertebrate biological condition would be expected? In general, we expect higher urban land cover to be predictive of worse biological stream condition (i.e., distributions shifted towards higher likelihood of tolerant species and lower likelihood of sensitive species) and vice versa. Before entering evidence, the posterior network shows an average of 27.19% chance of EPT richness greater than 13 taxa across all streams in the southeast United States (Figure 54). Given evidence that urban disturbance is low (urban land cover <12% of basin area), we now have an revised belief of the state of EPT richness relative to having no specific evidence on urban disturbance (i.e., when the marginal belief of EPT richness is integrated across all possible values of

urban disturbance). This updated Bayesian network shows that high EPT richness probability increases to a 47.74% chance of EPT richness greater than 13 taxa (Figure 55). In fact, the whole discrete distribution of EPT richness values shifts to show greater likelihoods of streams having many sensitive species than when urban land cover is averaged over all possible values. Similarly, marginal distributions of other system nodes show that in conditions of low urbanization, it is more likely to find lower richness-weighted tolerance, higher percent intolerant taxa, lower flashiness, higher channel width to depth ratio, lower conductivity and lower impervious surface.

If, on the other hand, we have evidence that urban disturbance is high (urban land cover >66% of basin area), the updated belief about EPT richness shows only 10.81% chance of EPT richness greater than 13 taxa, and greater likelihoods of higher richness-weighted tolerance, lower percent intolerant taxa, higher flashiness, lower channel width to depth ratio, higher conductivity and higher impervious surface (Figure 56). Not only are trend directions established on an interactive system level, but the probabilistic magnitudes of those trends are quantified. These quantifications are comprehensive, including all uncertainties associated with expert opinion and data, making them incredibly useful in a decision management context. These types of relationships and uncertainties are intuitively experienced using the graphical interface of the Bayesian network.

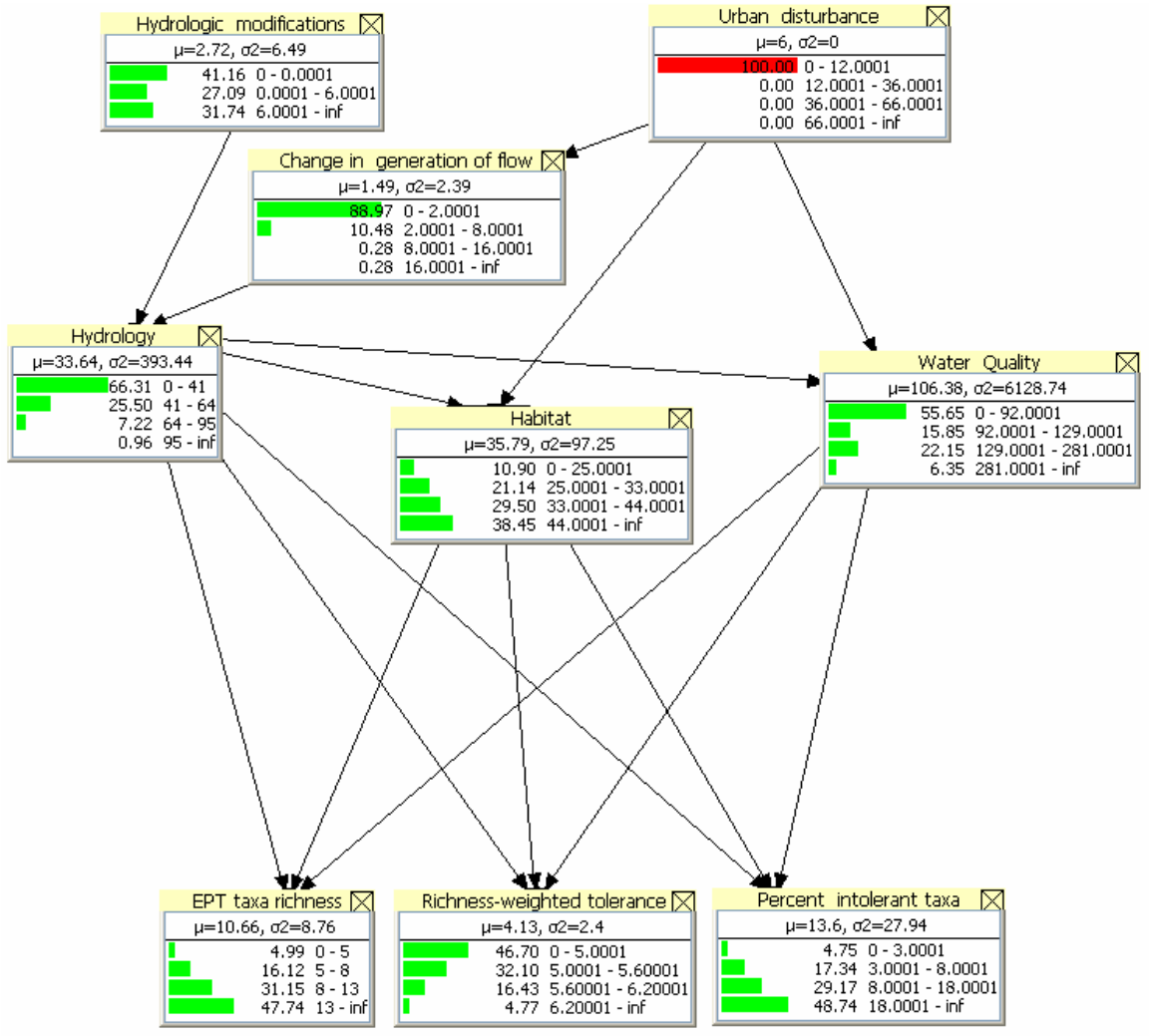

Figure 55: Posterior predictive marginal distributions for low urban disturbance in the southeast United States.

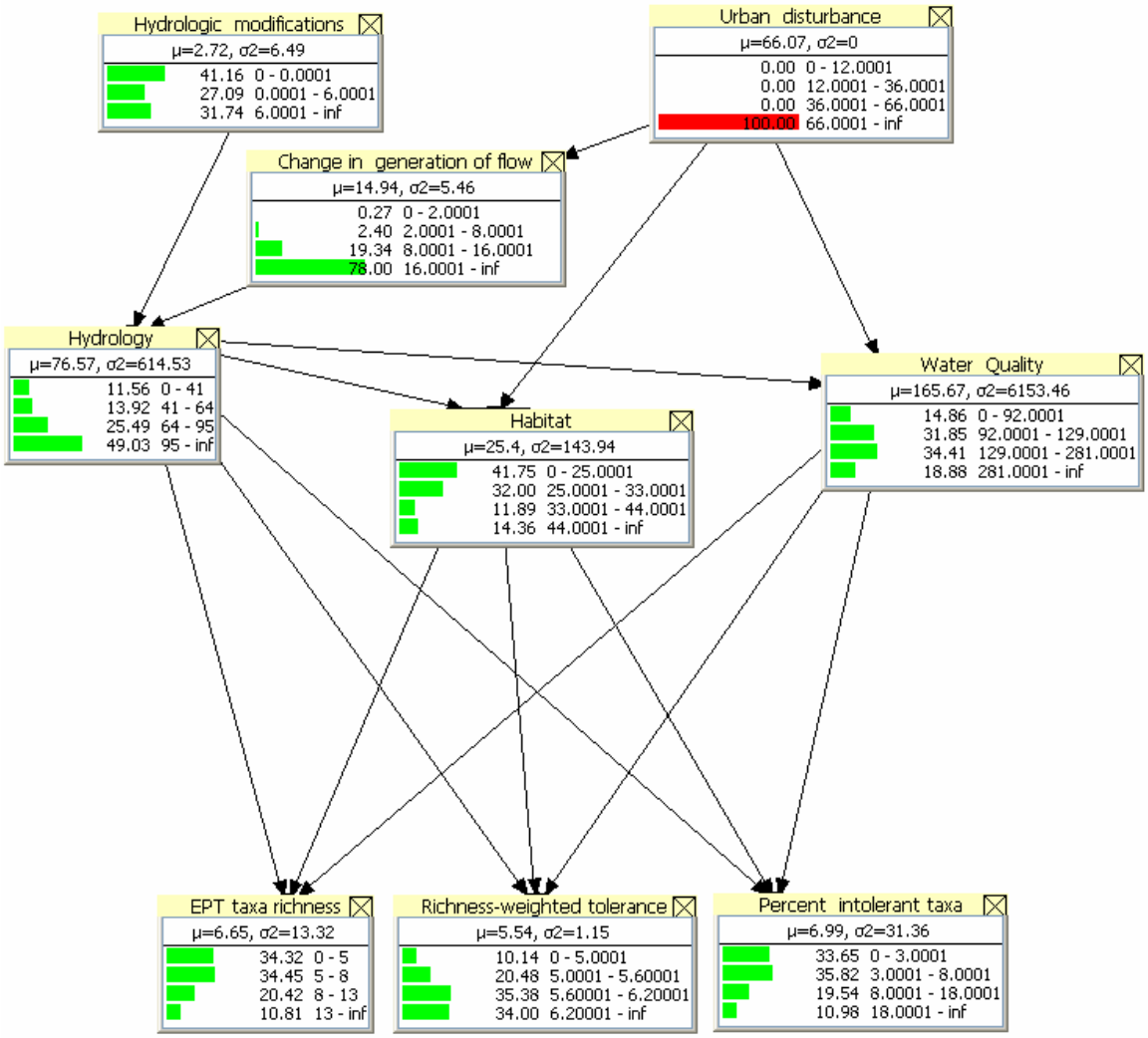

Figure 56: Posterior predictive marginal distributions for high urban disturbance in the southeast United States.

Diagnostically, we can ask: what does the level of urbanization have to be to achieve a certain biological condition? In this case, we have hypothetical evidence on an endpoint and are interested in the probabilities of drivers that would result in that endpoint. Essentially, we diagnose the probable values of parent nodes given the value of a child node. In this backwards-diagnosing direction, we expect good biological condition to be the result of a distribution of urban land cover with high probabilities of low urban land cover, and, conversely, bad biological condition to be the result of high probabilities of high urban land cover.

The posterior southeast Bayesian network supports these diagnoses, showing that EPT richness greater than 13 taxa would require a 43.39% likelihood of urban land cover less than 12% of the basin area (Figure 57) and EPT richness less than 5 taxa would be caused by a 48.21% likelihood of urban land cover greater than 66% of the basin area (Figure 58). Different uncertainties are taken into account to provide these diagnoses than to provide predictions because in diagnoses endpoint states are known conclusively while in prediction it is the drivers that are known, leading different parts of the model to be unknown in each situation.

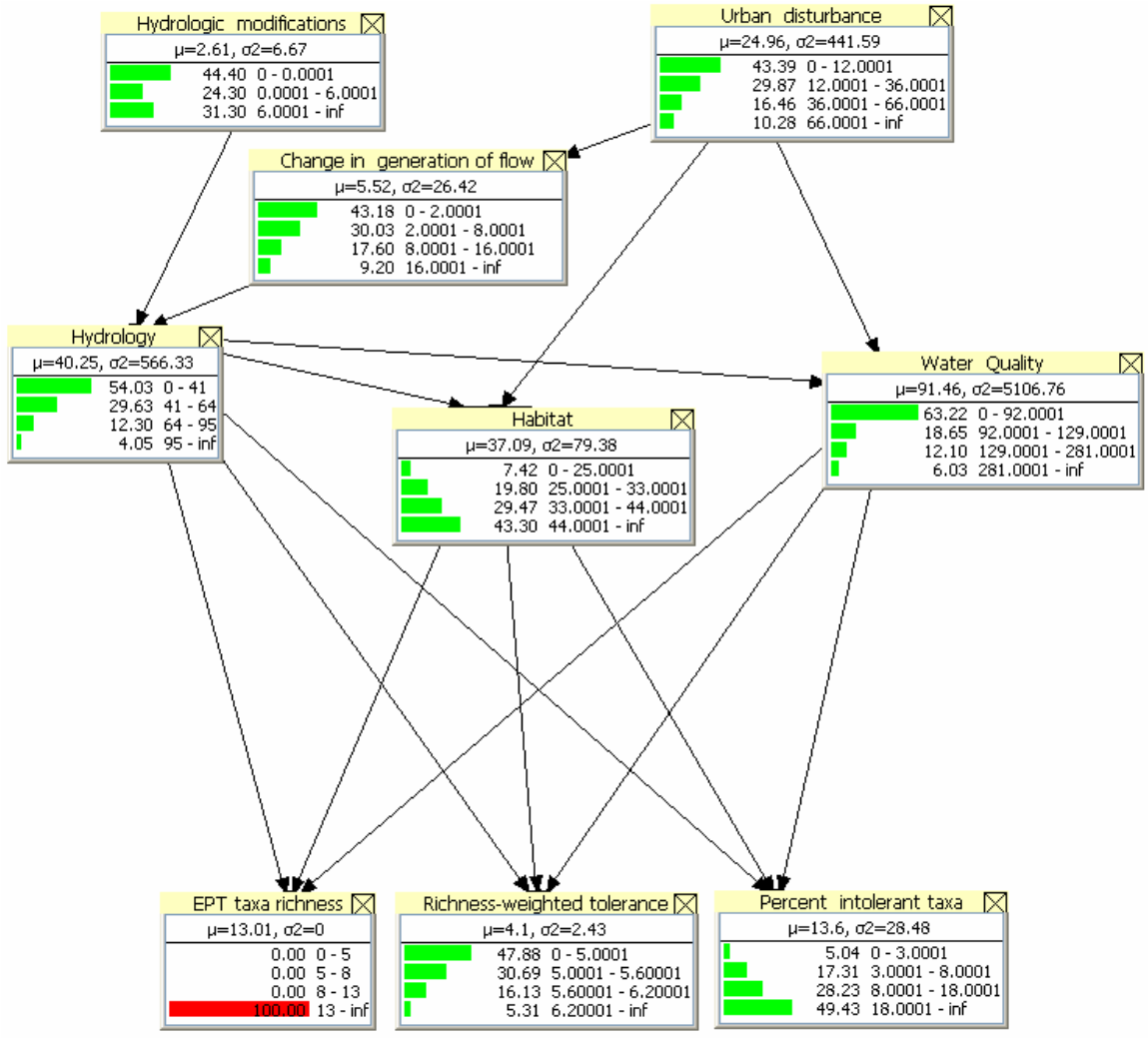

Figure 57: Posterior diagnostic marginal distributions for high EPT taxa richness in the southeast United States.

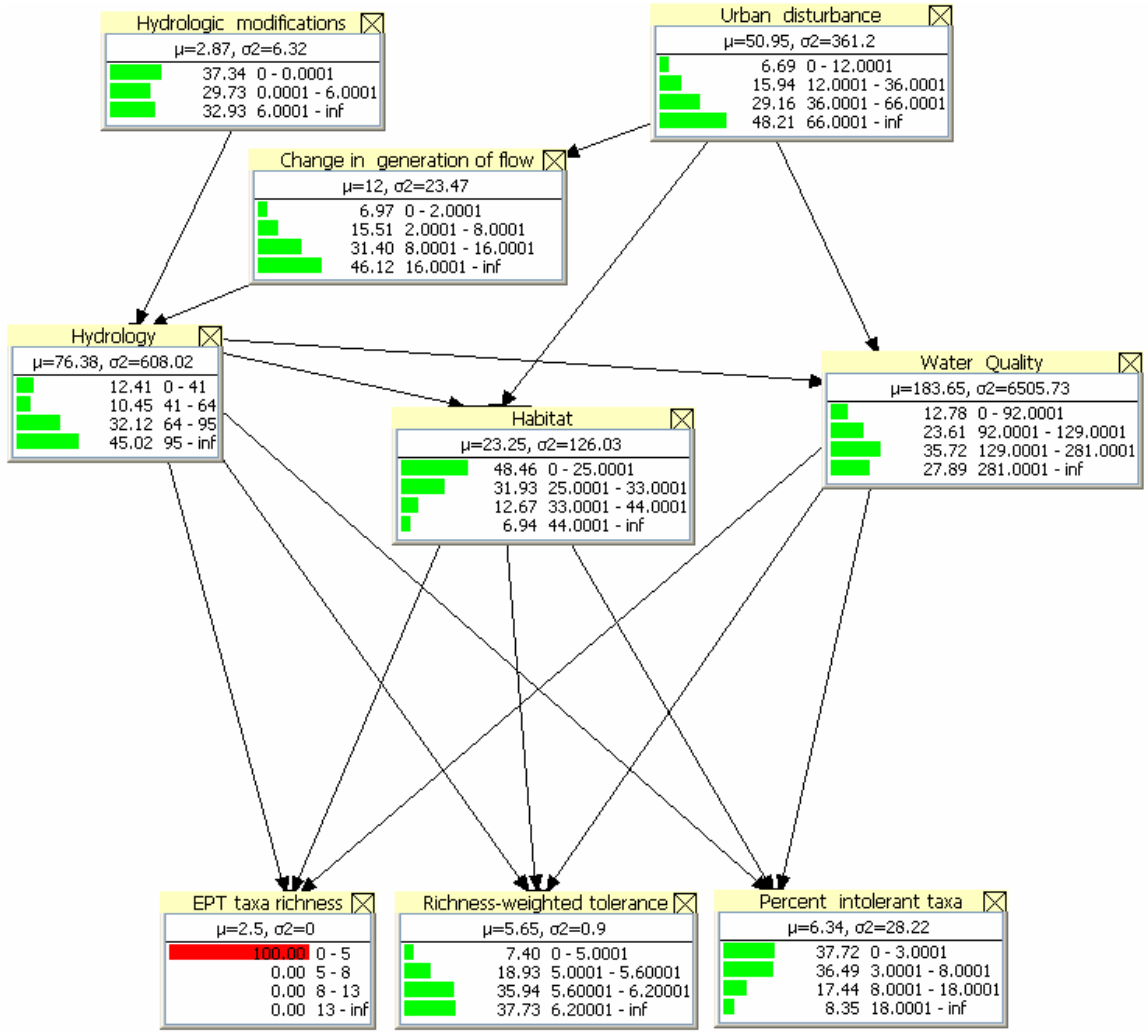

Figure 58: Posterior diagnostic marginal distributions for low EPT taxa richness in the southeast United States.

3.3.2. Mutual information: model link strengths

Mutual information was calculated for the prior model (Figure 59), data-only model (Figure 60), and posterior model (Figure 61). The greatest mutual information in the prior is between 'Urban disturbance' and 'Change in generation of flow' nodes (MI=0.63). This is not surprising since the expert noted that percent impervious surface

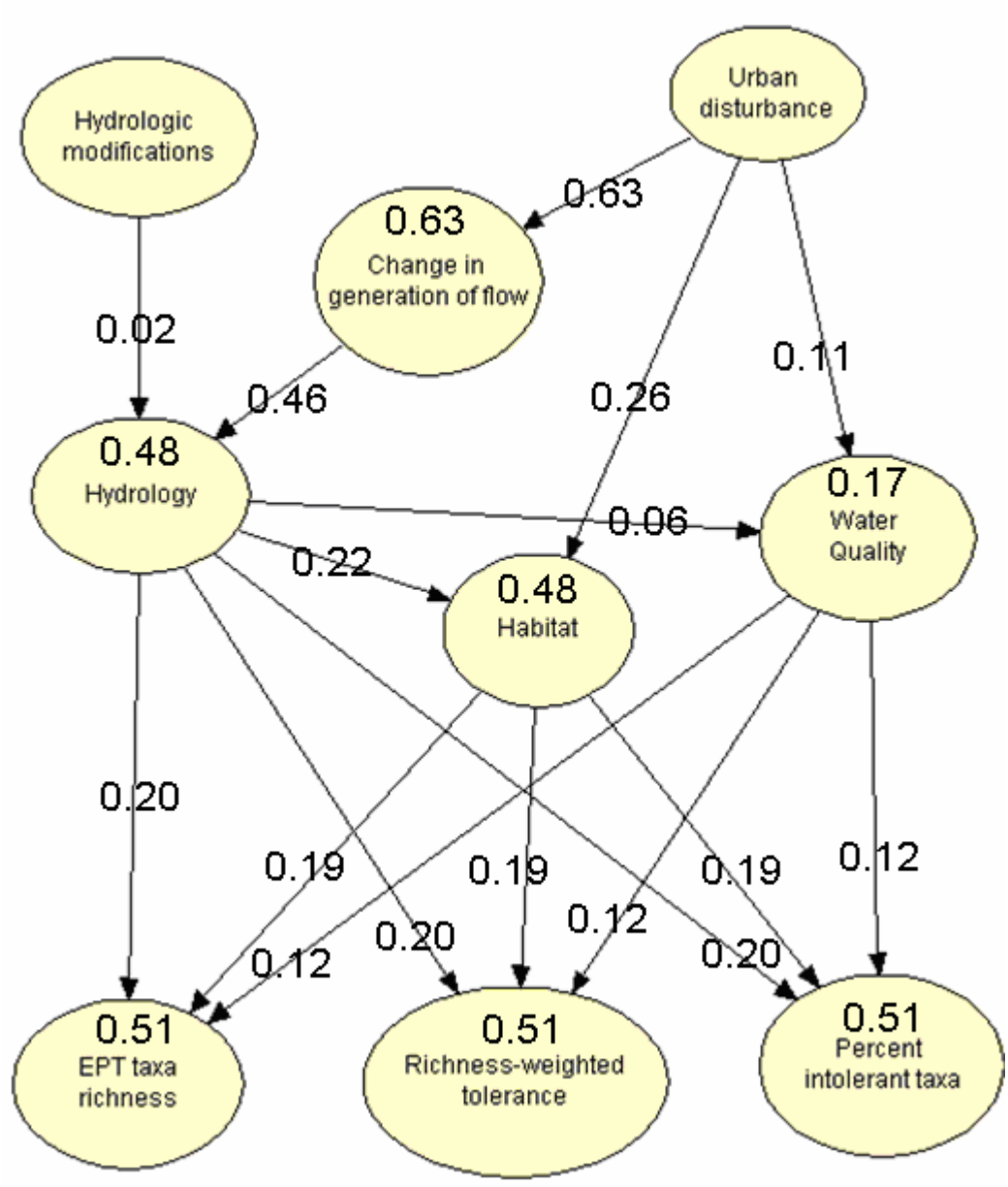

Figure 59: Mutual information for the prior southeast model. [Mutual information between two nodes is labeled directly on the arrow linking the two nodes; total mutual information for a child node given all its parents is labeled inside each child node.]

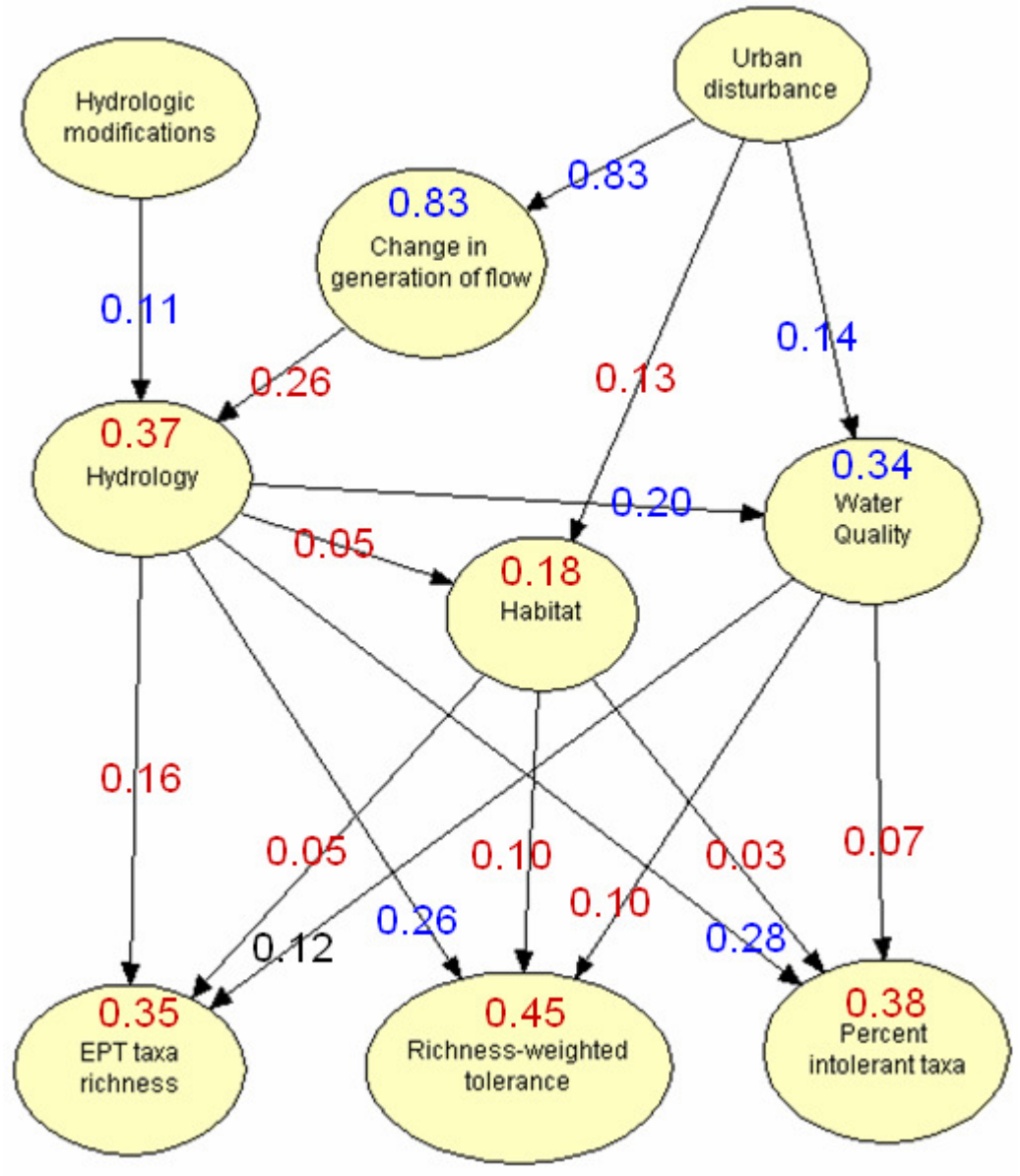

Figure 60: Mutual information for the data-only southeast model. [Mutual information between two nodes is labeled directly on the arrow linking the two nodes; total mutual information for a child node given all its parents is labeled inside each child node. Mutual information values for the data which are greater than prior mutual information are labeled in blue, values which are less than the prior are labeled in red, and values the same as the prior are labeled in black.]

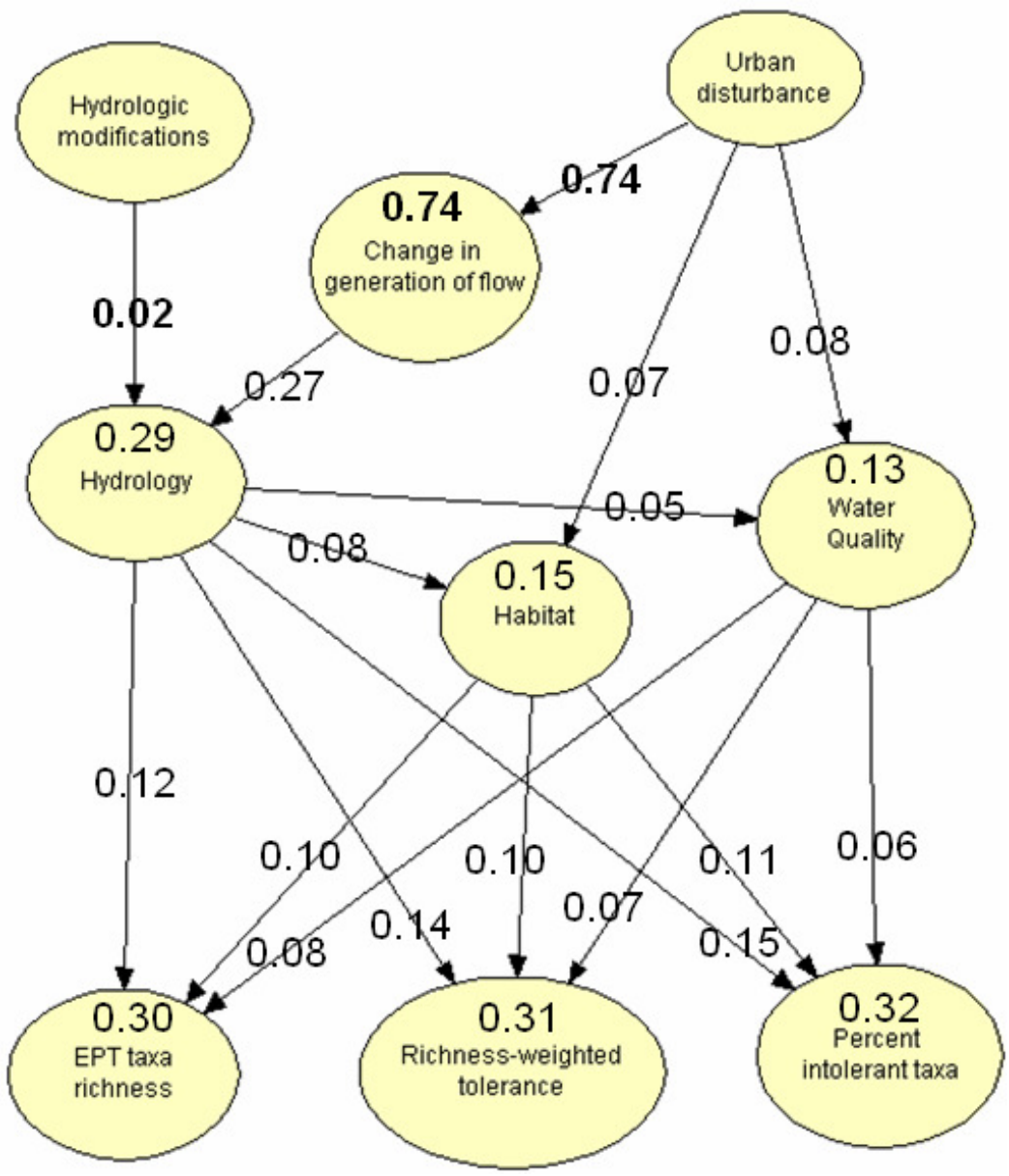

Figure 61: Mutual information for the posterior southeast model. [Mutual information between two nodes is labeled directly on the arrow linking the two nodes; total mutual information for a child node given all its parents is labeled inside each child node. Mutual information values for the posterior which are greater than or equal to prior mutual information are designated in bold.]

very closely approximates percent urban land cover in a watershed. Ideally, there should be a more direct measure of change in generation of flow, as impervious surface

is only an indirect approximation and is essentially a component of urban land cover which explains their close correlation. There was also relatively high mutual information between expert-designated important environmental stream drivers (hydrology, habitat, and water quality) and invertebrate community response (MI=0.51 for all three invertebrate metrics). This verifies that the expert considered these three environmental stream factors vital process components during both model structure elicitation and conditional probability elicitation (knowledge quantification). Hydrology and habitat have only slightly less mutual information with their parents (MI=0.48 for both). However, habitat seems to be relatively equally dependent on its two parents (hydrology and urban disturbance), hydrology is almost exclusively dependent on 'Change in generation of flow' and virtually independent of 'Hydrologic modifications'. This means that while the expert believed dam density to be important in theory, the results of the elicitation did not quantify it as such. Finally, water quality had the least mutual information shared with its parents (MI=0.17), and, particularly, less with hydrology than with urban disturbance. Perhaps this result is a manifestation of the expert's belief that one time water samples are a poor indicator of general water quality on a scale that affects invertebrates.

Comparing the prior to the data model (Figure 60, differences between the prior model and the data model are coded in red where the prior had greater mutual information and blue where the data had greater mutual information), the data also

showed highest mutual information between 'Urban disturbance' and 'Change in generation of flow' (MI=0.83) with magnitude even stronger than the prior. Mutual information for the invertebrate metrics and their parents remained important, albeit more weakly than established in the prior. Also, while, in the prior, the expert expected the three different metrics to be equally informed by environmental factors, the data show richness-weighted tolerance to have the greatest dependence on hydrology, habitat, and water quality (MI=0.45), followed by percent intolerant taxa (MI=0.38) and then EPT richness (MI=0.35). This invertebrate metric ranking is consistent with EUSE invertebrate regression strength (Cuffney et al. 2009a). Perhaps using an averaged tolerance index instead of abundance or taxa counts better integrates invertebrate response at a scale appropriate to the model. Alternately, perhaps the small range of possible tolerance values exaggerated dependencies with driver variables. Hydrology and habitat both have less mutual information with their parents in the data than in the prior, but water quality data actually have more dependence on its parents than in the prior. Perhaps the expert attempted to compensate for perceived sampling scale inaccuracies more than was warranted by the data. Or perhaps the data incorrectly appeared dependent on drivers by chance. Also, curiously, hydrologic modifications data show greater mutual information with hydrology than the expert expected.

Despite strong mutual information in the prior and data models, the posterior has lower mutual information between nodes for all but one connection (Figure 61).

Mutual information between 'Urban disturbance' and 'Change in generation of flow' (MI=0.74) increases from that reported for the prior, which is logical since the data showed even greater mutual information. While mutual information is still substantially greater than zero between the other nodes, they do not show as strong dependencies as in the prior or in the data. Perhaps, for these nodes with more parents and, hence, more complicated relationships with their parents, the functional forms of the dependencies between variables in the prior versus in the data were sufficiently different such that they dampened each other in the posterior. Posterior mutual information does not appear to be a weighted compromise between prior mutual information and sample mutual information, in contrast to Bayesian posterior distributions relative to prior distributions and data likelihoods.

3.3.3. Sensitivity analysis: model driver strength

Influences of driver nodes on invertebrate metric endpoints were assessed and compared by varying the state of each driver node from its 'worst' to its 'best' bin value. For example, when evidence was propagated that the 'Hydrology' node was in the high state (95+ rises greater than 5 times the median), high EPT richness (13+ EPT taxa) was 5.34% likely, low richness-weighted tolerance (<5) was 4.60% likely, and high percent intolerant taxa (18%+) was 4.64% likely (Figure 62). However, when evidence that the

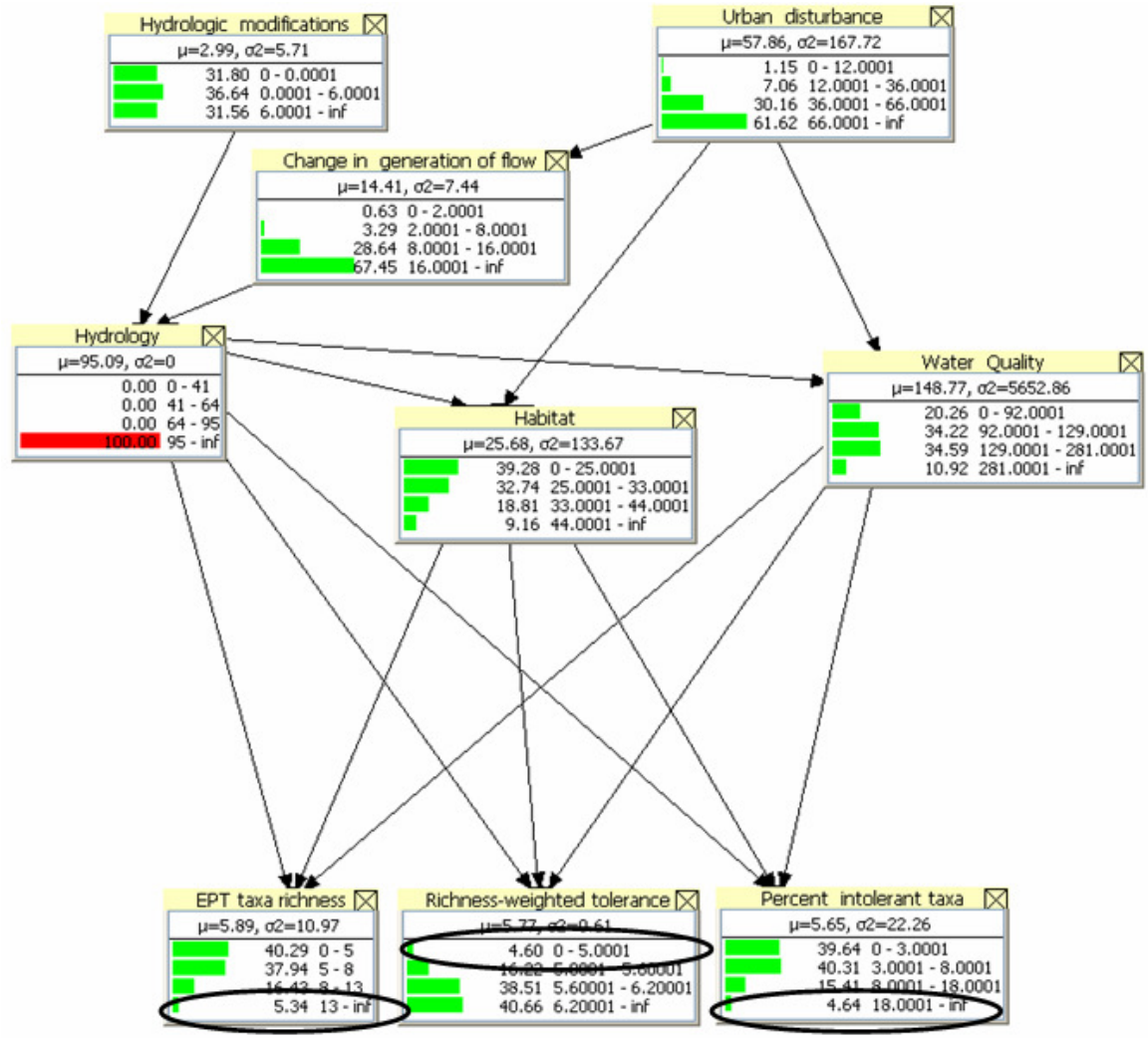

Figure 62: Posterior predictive marginal distributions for high flashiness as it affects invertebrate metrics in the southeast United States.

'Hydrology' node was in the low state (0-41 rises greater than 5 times the median) was propagated, likelihoods that invert metrics were in the best possible categories increased to 45.97%, 46.16%, and 49.34%, respectively (Figure 63). These represent changes of 40.63%, 41.56%, and 44.70% in likelihood of high EPT taxa richness, low richness-weighted tolerance, and high percent intolerant taxa, respectively.

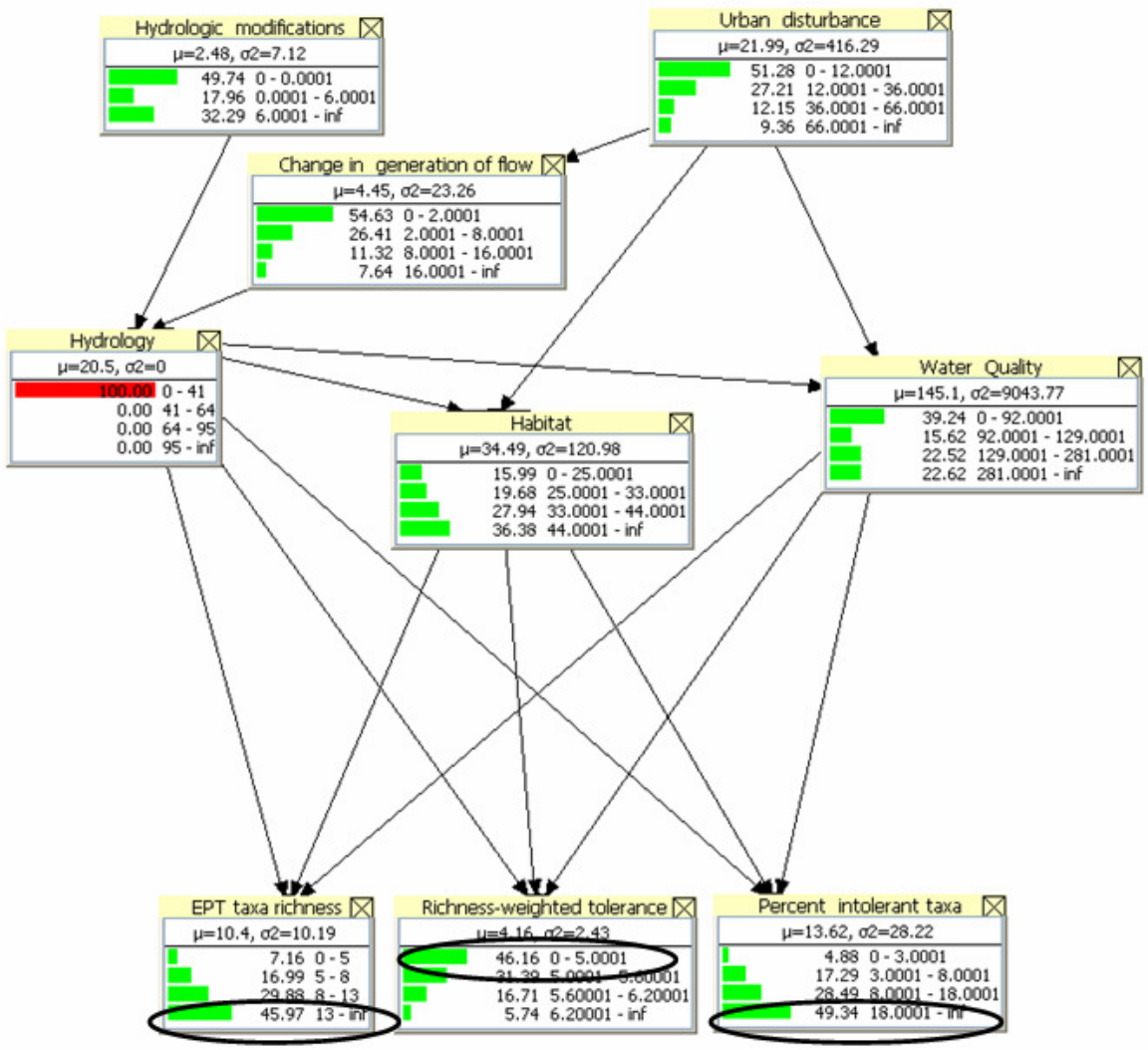

Figure 63: Posterior predictive marginal distributions for low flashiness as it affects invertebrate metrics in the southeast United States.

This sensitivity analysis was performed for each of the six non-endpoint driver nodes (Table 35). Visually (Figure 64), it is evident that all three invertebrate metrics responded very similarly to changes in driver conditions, differing at most about 8% in response to changes in water quality (conductivity). Urban land cover and impervious surface showed nearly identical influence on invertebrate metric improvement, both

Table 35: Possible 'best' state invertebrate metric improvements from full-scale improvements ('worst' to 'best') in driver nodes of posterior model.

Node variable	High EPT richness (13+ taxa)	Low richness-weighted tolerance (<5)	High percent intolerant taxa (>18%)
Urban land cover high (>66%) → low (0-12%)	10.81% → 47.47% Δ = 36.66%	10.14% → 46.70% Δ = 36.56%	10.98% → 48.74% Δ = 37.76%
Impervious surface high (>16%) → low (0-2%)	10.27% → 47.46% Δ = 37.19%	9.62% → 46.76% Δ = 37.14%	10.27% → 49.07% Δ = 38.80%
Dam density high (>6/km²) → low (0/km²)	26.81% → 29.33% Δ = 2.52%	25.84% → 28.65% Δ = 2.81%	27.11% → 30.51% Δ = 3.40%
Flashiness high (95+rises) → low (0-40rises)	5.34% → 45.97% Δ = 40.63%	4.60% → 46.16% Δ = 41.56%	4.46% → 49.34% Δ = 44.70%
Channel width:depth ratio low (0-25) → high (>44)	7.84% → 52.70% Δ = 44.86%	7.17% → 50.01% Δ = 42.84%	9.25% → 52.37% Δ = 43.12%
Conductivity high (>281 µs) → low (0-92 µs)	11.06% → 44.35% Δ = 33.29%	12.54% → 42.62% Δ = 30.08%	16.87% → 42.61% Δ = 25.74%

increasing the likelihood of observing 'best' bin invertebrate metrics to an average of 47-49% (which was 36-39% improvement from 'worst case' urban land cover and impervious surface scenarios). Changes to channel width to depth ratio (habitat) appear to have the greatest impact on invertebrate community improvement (up to ~53% chance of observing 'best' invertebrate metric values with 43-45% improvement from 'worst case' habitat), followed closely by flashiness (hydrology) improvements (up to ~49% chance of observing 'best' invertebrate metric values with 41-45% improvement from 'worst case' water quality). Dam density decreases have by far the least impact on macroinvertebrate condition (up to only about 31% chance of observing 'best' invertebrate metric values, improving only 2.5-3.4% from 'worst case' dam density).

This lack of significant influence on biological impacts of interest may be justification for removing the 'Hydrological modifications' node from model structure. Otherwise, the model appears fairly similarly responsive to changes in all other drivers.

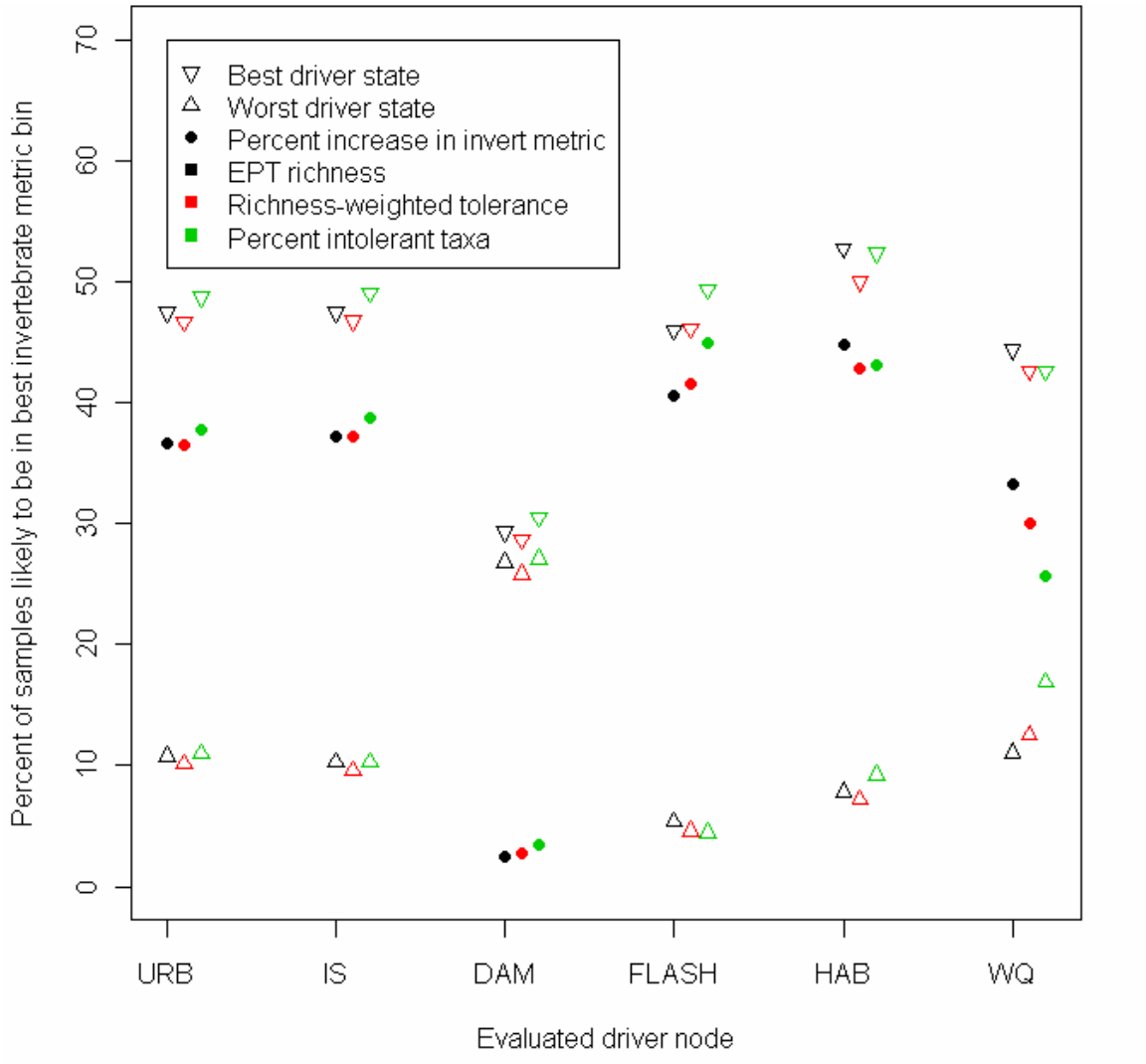

Figure 64: Sensitivity analysis of driver importance on increasing best invertebrate metric bin. [URB is urban land cover, IS is impervious surface, DAM is dam density, FLASH is hydrological flashiness, HAB is habitat channel width to depth ratio, and WQ is water quality conductivity.]

3.3.4. Comparing prior to posterior: information source and direction

Standardized weighted average scores ranging between 1 (all probability weight in 'low' bin 1) and 4 (all probability weight in 'high' bin 4) were calculated for each CPT row of each child node's prior CPT, maximum likelihood data table, and posterior CPT. To quantify how much posterior child node Dirichlet distributions differed from prior distributions, each CPT row's prior standardized score was subtracted from its respective posterior standardized score. This measure, which could range from -3 to 3, estimates how much data influenced prior elicited parameters during model updating, taking into account the prior weight assigned to the prior relative to the data sample size. The direction of change (positive or negative) identifies whether data increased or decreased the expert's expected child node average estimate. If the change between prior and posterior CPT row standardized score is positive, then data measured greater child node expected value (given that CPT row parent state combination) than the expert reported, and vice versa for a negative difference between prior and posterior. These magnitudes and directions of differences between priors, data, and posteriors, as they change with changing parent node states, are visually evident in the plots below.

Across all seven child nodes, posteriors only differed from priors a minimum of -0.63 to a maximum of 0.79. Given that these differences could have been as low as -3 and as high as 3, this much smaller actual range seems to indicate similarity between prior knowledge of and data measurements of the same phenomenon, relative to the

weight placed on each. Within a node, the magnitude of these differences between prior and posterior does not appear to be dependent on CPT row sample size and, hence, observing a posterior very different from a prior is not a function of only the amount of data available to alter the prior. Instead, sample content is likely more important than sample size in changing a prior, given a constant prior weight.

The prior, data, and posterior standardized scores of the 'Change in generation of flow' node (Figure 65) are easiest to represent visually since it is the only child node in this model with one parent ('Urban disturbance'). It is also the node with the smallest

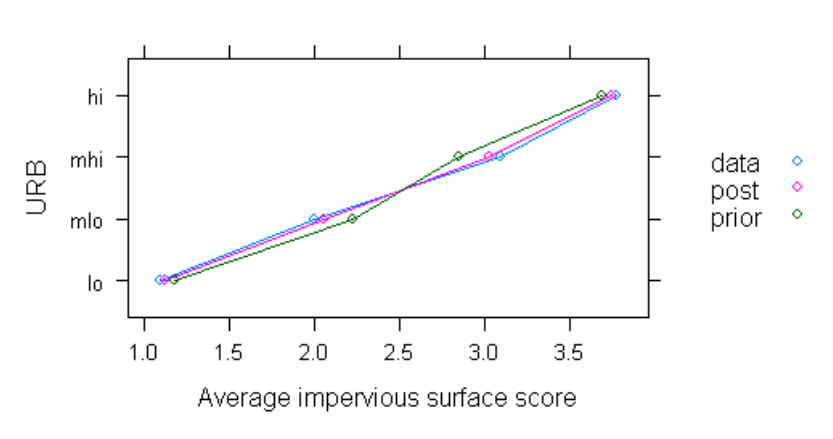

Figure 65: Average standardized CPT row change in generation of flow score for prior, data, and posterior compared across changing parent node states (URB is urban land cover of 'Urban disturbance' node).

differences between prior and data and the clearest response to change in parent node state. Prior and data standardized scores essentially overlap in describing increasing impervious surface in response to increasing urban land cover. The strength of this pattern confirms mutual information and sensitivity analysis findings.

Child nodes with two parents, such as 'Hydrology', 'Habitat', and 'Water quality' nodes, are depicted in two ways, by sorting first according to each parent, in order to better see patterns from different perspectives. 'Hydrology' node standardized score patterns (Figure 66) clearly show several important conclusions. First, as

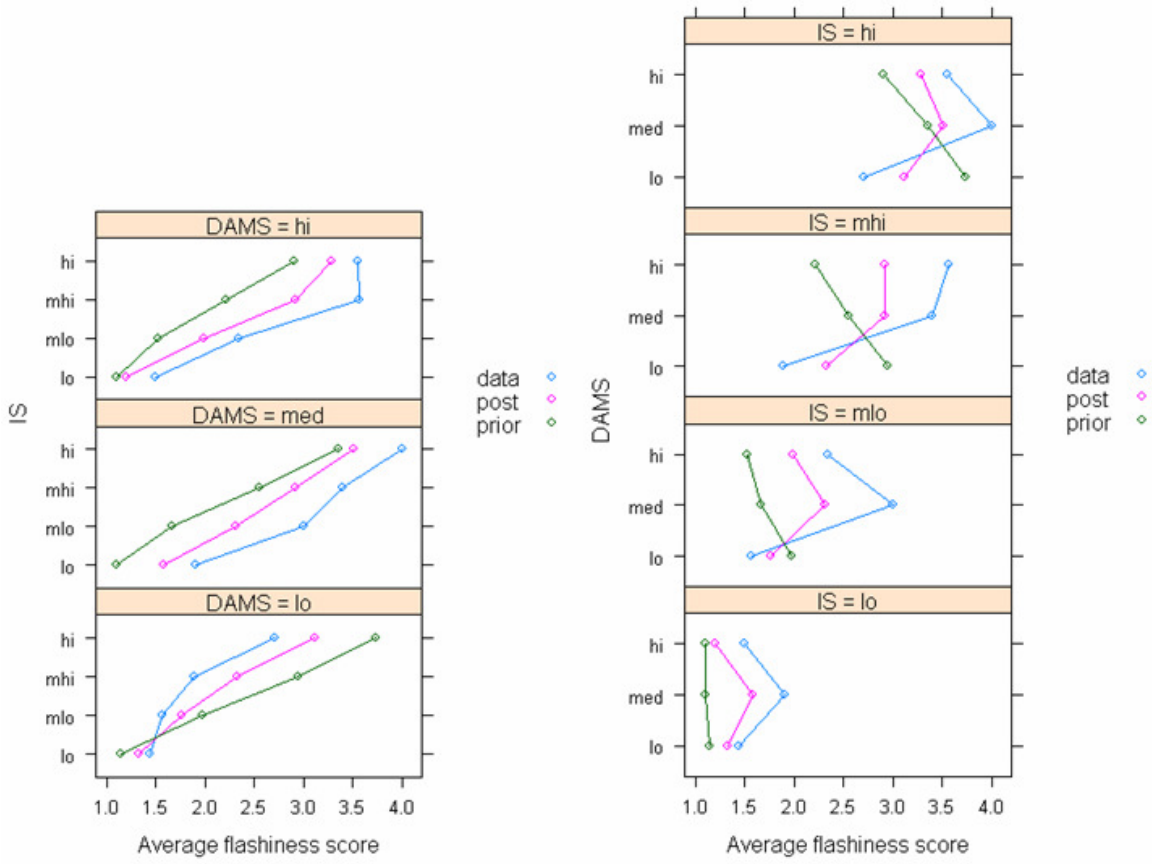

Figure 66: Average standardized CPT row hydrology score for prior, data, and posterior compared across changing parent node states (IS is impervious surface of 'Change in generation of flow' node and DAMS is dam density of 'Hydrologic modifications' node).

confirmed by mutual information (Figure 59-Figure 61) and sensitivity analysis (Figure 64), impervious surface is much more predictive of flashiness than dam density is.

When grouped by dam density, flashiness increases across an increasing impervious surface gradient for prior, data, and posterior of all dam density groups (as shown by parallel positively sloped lines, Figure 66, left graph), while flashiness scores remain in the same range regardless of changing dam density. This same conclusion is evident when grouped by impervious surface when average flashiness for prior, data, and posterior consistently increased as impervious surface state increases, with very little overlap between impervious surface categories.

In contrast, prior and data do not show similar responses of flashiness to dam density (Figure 66, left graph), as data (in blue) measured higher average flashiness than the prior expected (in green) for high and medium dam density (blue to the right of green) and lower average flashiness than the prior for low dam density (blue to the left of green). This contrast is even more evident when the responses are grouped by impervious surface (Figure 66, right graph), as prior (green) and data (blue) lines have very different shapes across increasing dam density for every impervious surface category, particularly in the transition from low to medium dam density. The expert expected flashiness to decrease as dam density increased, because typically dams mitigate sporadic water movement. However, data shows that flashiness actually increased from low to medium dam density, instead. Perhaps dam density is correlated with another factor that is causing the opposite of expected pattern. Alternately, these data were collected during a drought year which may affect typical expected patterns.

These plots also aptly demonstrate how, for each set of changing parent states, the posterior (pink line) is a weighted average of the prior (green line) and the data (blue line), since the pink line is always located in between the green and blue lines. The weight placed on the prior relative to the data during Bayesian updating is also evident from the specific location of the posterior relative to the prior and the data. For any CPT row (parent state combination), if the pink point is closer to the green point than to the blue point, this means the prior equivalent sample size has greater weight than the data sample size, and vice versa if the pink posterior point is closer to blue data point. This type of representation makes understanding of the influences on the posterior straightforward.

In contrast to the 'Hydrology' node, the 'Habitat' node did not appear to have a clear relationship between standardized average channel width:depth ratio score and its parents, 'Urban disturbance' (urban land cover) and 'Hydrology' (flashiness) (Figure 67). This result is consistent with conclusions drawn from mutual information analysis where the data and posterior showed greater mutual information between the 'Hydrology' node and its parents than between the 'Habitat' node and its parents (Figure 60 and Figure 61). While the expert expected channel width:depth ratio to decrease with increasing urban land cover and increasing flashiness (Figure 67, green lines), since urbanization is known to lead to channel incision, the data did not show this trend with changing parent node states (Figure 67, blue lines). The only parent state

combinations for which data scores somewhat agreed with prior scores were low and medium low urban land cover where data width:depth score increased across decreasing flashiness, as expected in the prior (Figure 67, right graph). Perhaps the remaining discrepancies are a consequence of variable choice— many more environmental concepts were originally desired to be represented by the 'Habitat' node than can actually be measured by a single variable. Flashiness, on the other hand, sums more efficiently the major influence of hydrology on the system.

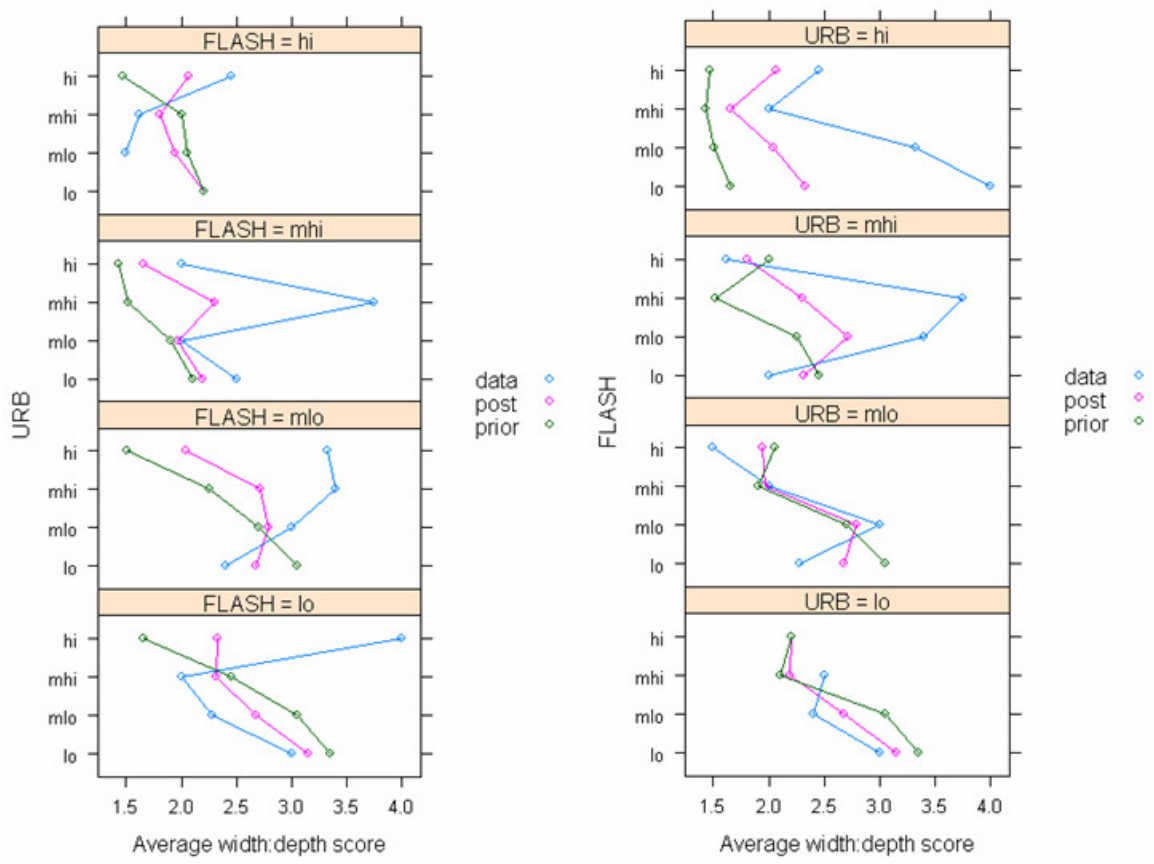

Figure 67: Average standardized CPT row habitat score for prior, data, and posterior compared across changing parent node states (FLASH is flashiness of 'Hydrology' node and URB is urban land cover of 'Urban disturbance' node).

For the 'Water quality' node, the expert expected conductivity to increase as urban land cover and flashiness increased, creating and delivering more contaminants to streams. Urban land cover was expected to have a greater impact than flashiness, as evidenced by the overlapping conductivity scores for all flashiness states (Figure 68, left graph) relative to the non-overlapping, increasing conductivity scores for increasing urban land cover state (Figure 68, right graph). For the most part, data reflected prior expectations for all parent node states except low flashiness. At low flashiness, data

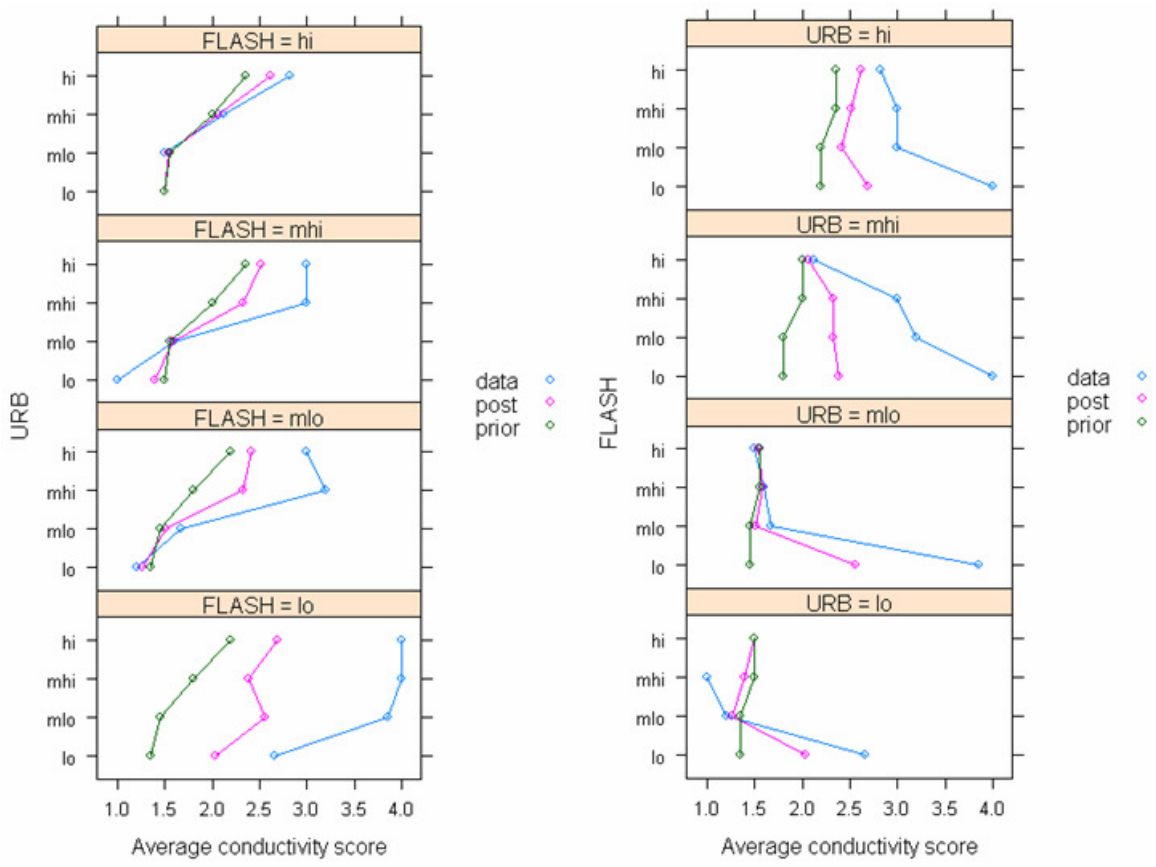

Figure 68: Average standardized CPT row water quality score for prior, data, and posterior compared across changing parent node states (FLASH is flashiness of 'Hydrology' node and URB is urban land cover of 'Urban disturbance' node).

show unexpected high conductivity both when grouped by flashiness (Figure 68, left graph, bottom) and by urban land cover (Figure 68, right graph). This is due to regional differences in background conductivity within the Southeast United States, where Birmingham, AL has naturally higher conductivity than Atlanta, GA and Raleigh-Durham, NC (Figure 69) likely due to regional geology differences. Atlanta and Raleigh

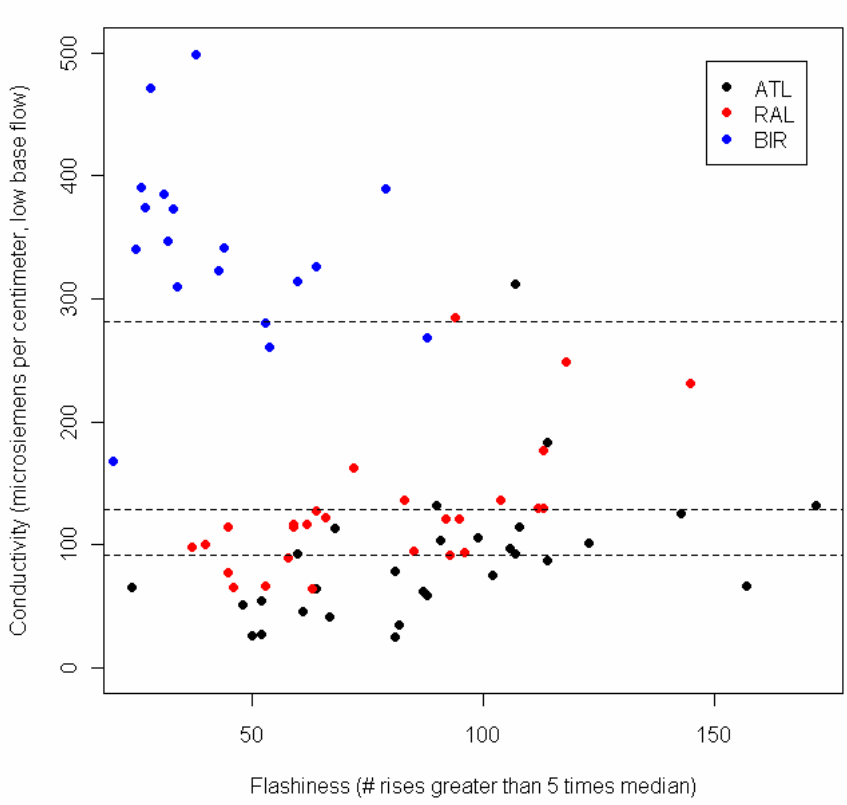

Figure 69: Southeast conductivity across flashiness by region (ATL is Alanta, GA, RAL is Raleigh-Durham, NC, and BIR is Birmingham, AL. Conductivity bin endpoints are shown with dashed horizontal lines.)

show the prior-expected pattern of increasing conductivity with increasing flashiness (as more pollutants are delivered to streams) but high conductivity background conditions seem to swamp this kind of pattern in Birmingham. This discovery supports the

Chapter 2 conclusion that urbanization affects different regions differently, and suggests that regions should be modeled separately prior to drawing conclusions across regions.

Nodes with three parents, 'EPT richness' (Figure 70), 'Richness-weighted

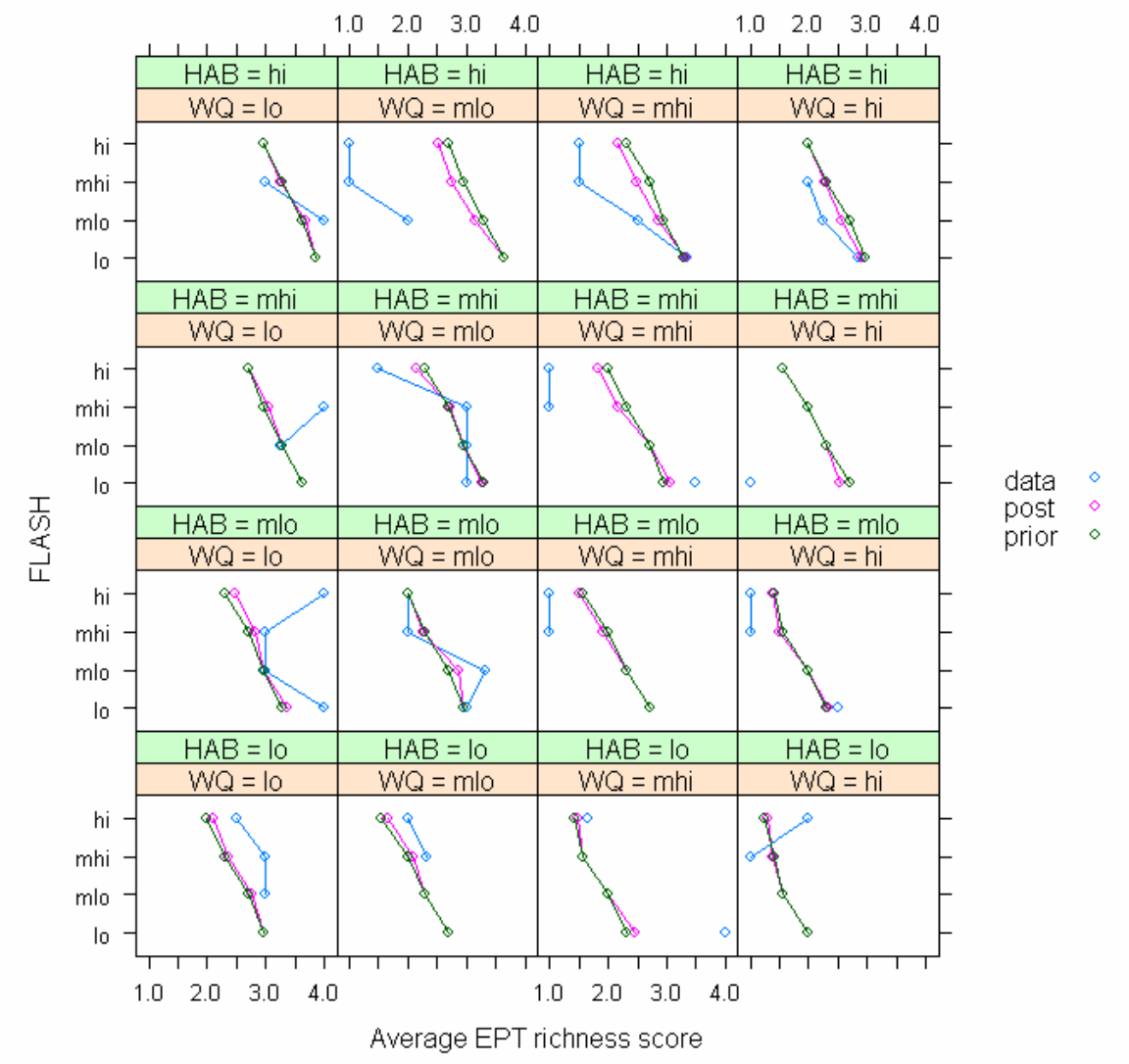

Figure 70: Average standardized CPT row EPT richness score for prior, data, and posterior compared across changing parent node states (FLASH is flashiness of 'Hydrology' node, HAB is width:depth channel ratio of 'Habitat' node, and WQ is conductivity of 'Water quality' node).

tolerance' (Figure 71), and 'Percent intolerant taxa' (Figure 72) are depicted in only one sorting configuration, as resorting in other parent state configurations did not improve interpretability. Due to small CPT row sample sizes, ranging from 0 to 6 (Table 76-

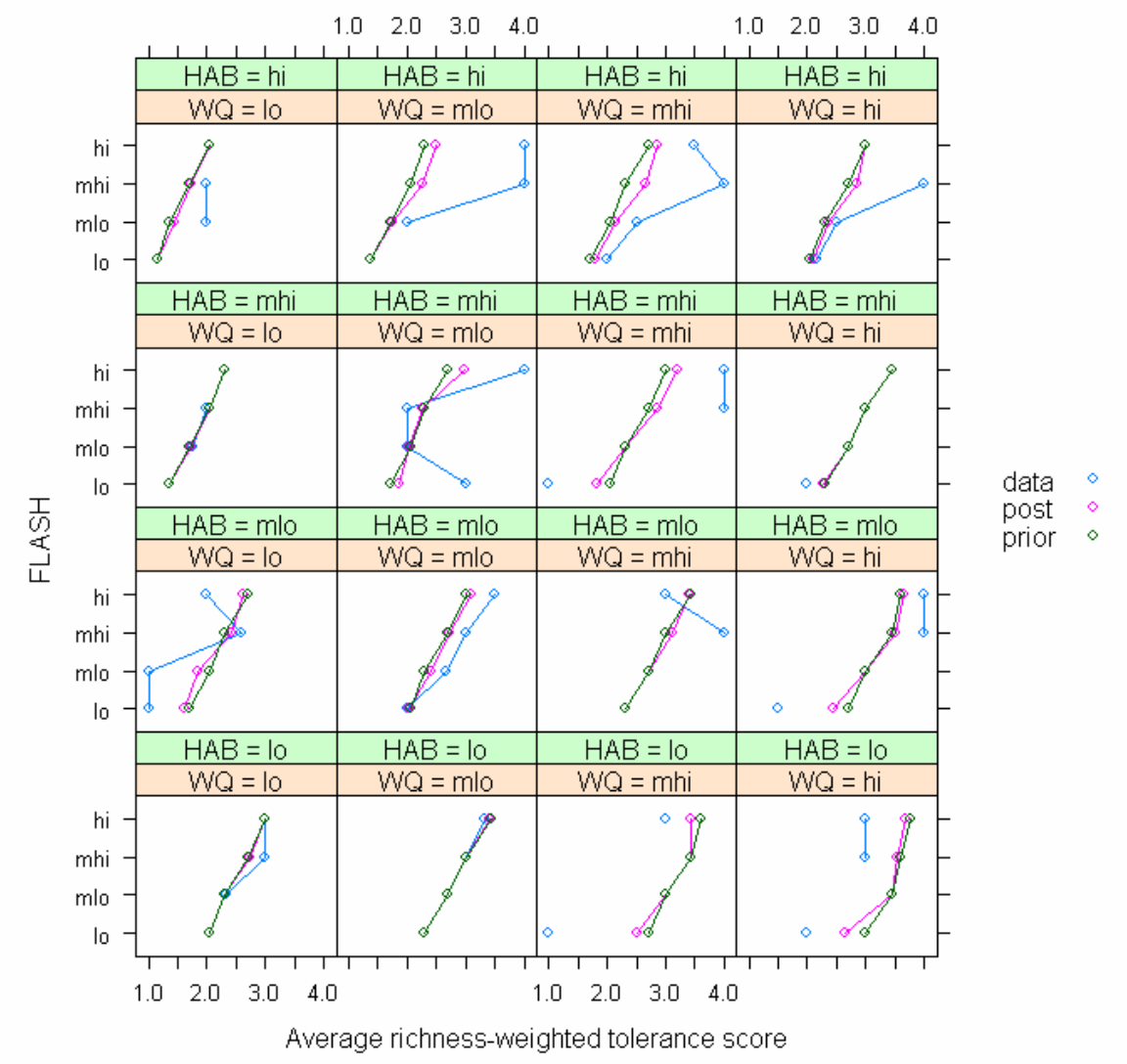

Figure 71: Average standardized CPT row Richness-weighted tolerance score for prior, data, and posterior compared across changing parent node states (FLASH is flashiness of 'Hydrology' node, HAB is width:depth channel ratio of 'Habitat' node, and WQ is conductivity of 'Water quality' node.)

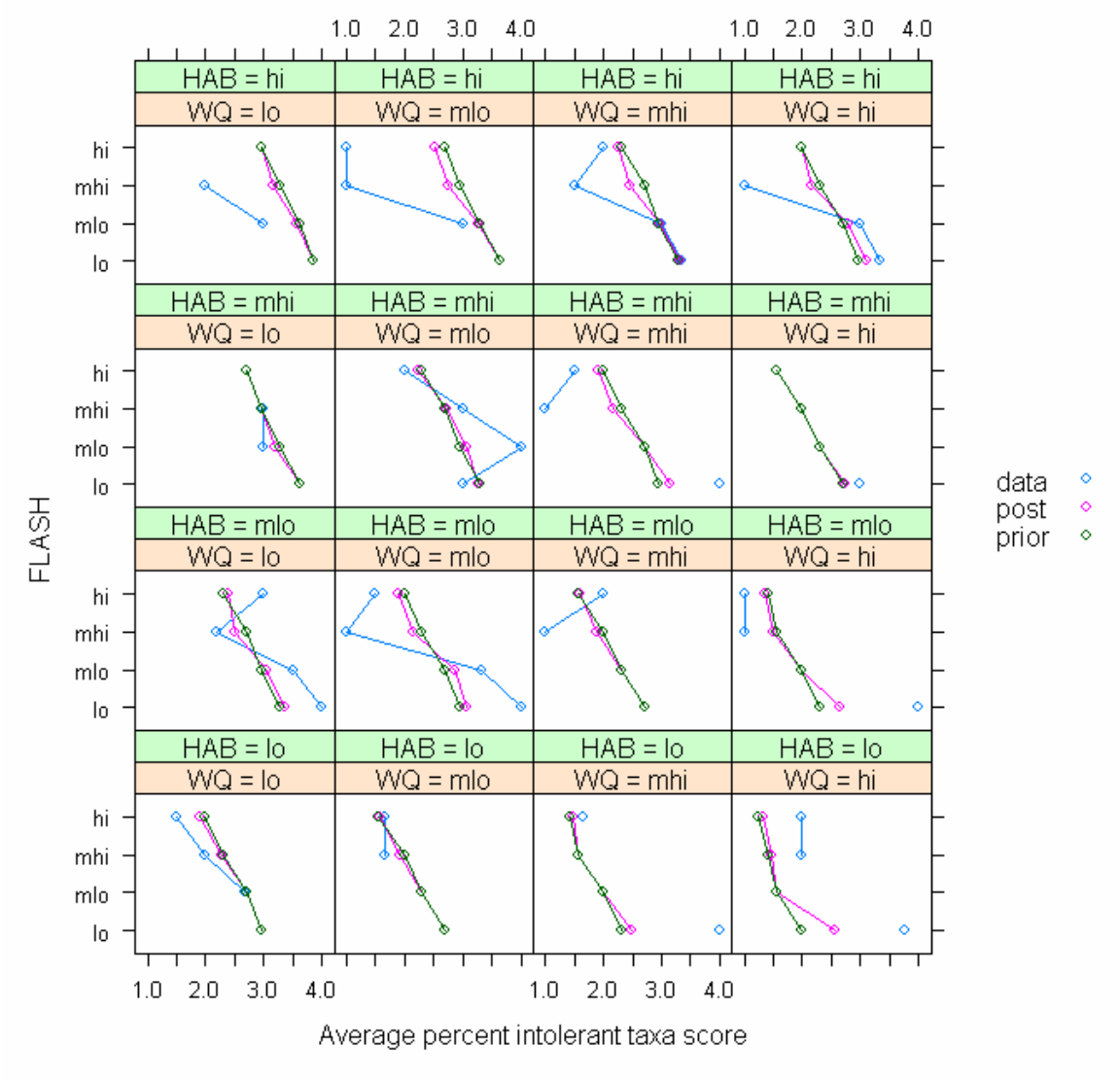

Figure 72: Average standardized CPT row Percent intolerant taxa score for prior, data, and posterior compared across changing parent node states (FLASH is flashiness of 'Hydrology' node, HAB is width:depth channel ratio of 'Habitat' node, and WQ is conductivity of 'Water quality' node.)

Table 78), relative to prior weights of 7.66-9.24 (Table 31) posteriors closely track priors for all three biologic endpoint nodes. Data patterns are difficult to discern (Figure 70-Figure 72) due to missing values for 19 of the 24 possible parent state combinations.

Additionally, data that is available tends toward average score extremes because of the small within-CPT-row sample sizes makes it more likely for a parent state combination to have the lowest or highest 'average' score (e.g., if the only data point is in the highest bin, the 'weighted' score is automatically 4.0).

Prior patterns (and, by default, posterior patterns), however, are more easily visualized. Within each of the 16 'Habitat'/'Water quality' state combinations, EPT richness decreases as flashiness increases (Figure 70). As EPT richness is a measure of generally sensitive taxa, this is an expected effect of urbanization-induced flashiness. Looking down the columns, EPT richness also increases as channel width:depth ratio increases (i.e., channel become less incised). And, looking across rows, EPT richness increases as conductivity decreases.

Richness-weighted tolerance priors (and posterior) have patterns in the opposite direction from EPT richness across flashiness, as expected, since tolerance is essentially measuring an opposite characteristic to EPT sensitivity. Therefore, richness-weighted tolerance (Figure 71) increases with increasing flashiness within each 'Habitat'/'Water quality' state combination, with decreasing width:depth ratio down the columns, and with increasing conductivity across the rows. Again, it is hard to draw conclusions from the richness-weighted tolerance data with so many missing values relative to the model construct.

Finally, the general direction of percent intolerant taxa changes relative to parent node state changes (Figure 72) mimics that of EPT richness. Data direction is somewhat consistent with the prior in many of the 'Habitat'/'Water quality' state combinations, although conspicuously different in a few (e.g., medium low width:depth ratio habitat state, intolerant taxa percent decreases as flashiness moves from high to medium high when the prior expected intolerant taxa to increase with decreasing flashiness). With more data, these types of patterns and consistencies can be evaluated in this manner more thoroughly in the future.

3.4. Discussion

This Bayesian network methodology development effort elucidated many benefits, drawbacks, learning opportunities, and research gaps in the use of this type of modeling to answer environmental questions. Overall, research goals of creating, parameterizing, and evaluating a model linking urbanization processes to stream biota responses were fulfilled. This final model represents a complicated network of cause and effect relationships and uncertainties, incorporates and appropriately weights both data and expert knowledge, and provides probabilistic predictions and diagnoses useful for system understanding and management.

The completed Southeast Bayesian network provides a level of understanding that supports previous modeling work predicting the effects of urbanization on aquatic invertebrates using the USGS EUSE dataset (Cuffney et al. 2009a, Cuffney et al. 2010,

Kashuba et al. 2010), but it does so in a more integrated and comprehensive framework that is able to model the entire system of effects rather than compare a few factors at a time. Also, this Bayesian network model actually has the ability to be directly manipulated by the user to explore probabilistically presented outcomes of various system scenarios. The potential for use of this type of model in environmental management is enormous.

3.4.1. Pros and cons of expert knowledge and data

Throughout the process of model building and quantifying, it was natural to compare, in practice, the kind of information available from expert knowledge versus data. Each is not without its own unique strengths and flaws. Expert opinion certainly allowed the model to incorporate advanced, integrated scientific thinking, a type of information data simply do not contain. This expert-driven model construction minimized the model's susceptibility to excessive influence from blind, possibly random, data correlation devoid of analytical reasoning. However, expert opinion is not immune from various kinds of innate human biases or subjectivity. In fact, there is a large field of applied psychology and decision analysis that studies the different ways human reasoning can be influenced when processing information and assessing uncertainty. Several consistent patterns of bias have been documented.

Availability bias occurs when assesors link their probabilities to the frequency with which they can recall past occurrences of an event (Tversky and Kahneman 1974).

Representativeness bias arises if an expert assumes a relationship seen in a small number of data points is representative of a whole population when this may not, in fact, be the case. This may be related to overconfidence bias where experts have difficulty in assessing the tails of a distribution (Wallsten and Budescu 1983) because they skew their judgments towards the mean. Anchoring and adjustment bias occurs when judgments are anchored at some starting value and adjusted outwards, usually insufficiently (Winkler 1967, Alpert and Raiffa 1982). Motivational bias may result where an expert is motivated to provide answers that support certain policy, results, or desired answer. Hindsight fallacy may ensue if the expert has seen the sample data, in which case the elicited opinion may have already been updated on the basis of the data (Morgan and Henrion 1990). Finally, conjunctive fallacy can occur when a higher probability is assigned to an event that is actually a subset of an event with a lower probability (Mullin 1986). These possible cognitive biases associated with expert knowledge can be confounded with linguistic uncertainty of node, variable, or bin definitions (Hosack et al. 2008). The ideal method for addressing these potential biases and translating the expert knowledge into quantified form is still under development.

We often would like to believe that data objectively transcend all these subjective influences of human-derived information. And real-world measurements of phenomena of interest are often able to do so. Data increase our confidence that hypothesized relationships exists and form the basis for understanding of all observable

phenomena. But data do not come without their own unique set of potential problems, either. First is the question of how close what you are measuring is to the actual concept of interest. Regularly, surrogates or indicators of true variables of concern are measured rather than the variables themselves due to cost, time, or logistics limitations. This final limitation is the case when stream substrate size or degree of incision is measured to estimate 'habitat quality' or when impervious surface is measured to approximate 'change in generation of flow'. The true concept of interest is often very difficult to measure directly, and validity of the surrogate measure depends on the correctness of assumptions about the relationship between the true value and the surrogate. A related data collection issue, in terms of cost and time limitations, is one of sample size and sampling frequency. When the budgeting and personnel allocations are less than sampling requirements, enough data cannot be collected to accurately depict true trends and patterns studied. Smaller than required sample sizes may also omit record of many possible scenarios of interest, particularly system extremes, making it difficult to understand the true range of the system.

In other situations, measuring the system actually alters it such that, inevitably, measurements no longer match the system. In environmental studies, specifically, the system of interest is not directly controllable as in a laboratory. Therefore, experimenters may not be able to hold all influences but the measured influence of interest constant, and results may be caused by something that is correlated to what you

are measuring but was itself unmeasured. Additionally, environmental systems are constantly changing such that it is not possible to take repeated samples of the same population and it is not clear to which questions the data collected apply. Data may contain measurement error, either random or systematic, or just be naturally variable. Finally, there may be a mismatch of scales between what is actually measured and what it is purported to represent. The clearest example of this situation, which was clarified during the Bayesian network building process, involves the water samples collected as part of the EUSE dataset. These one time samples, which were analyzed for a variety of water chemistry components, did not appear to correlate well with any urbanization indicators (Sprague et al. 2007). This is likely because a single sample is on a much smaller temporal scale than watershed-scale measures of impairment, which integrate expected influences of urbanization over annual scales. The single, instantaneous water sample is not representative of the water quality conditions of the watershed and would not be expected to correlate with measurements of larger temporal, or spatial, scale. Water chemistry is not measured often enough, in enough locations, to characterize water quality at the same scale as urban land cover characterizes urbanization or invertebrate community metrics characterize biological condition.

3.4.2. Model building questions and concerns

Several unchangeable properties of Bayesian networks have implications for environmental model building.

3.4.2.1. No feedback loops

Because of principles of graph theory and probability propagation, Baesyian networks are required to be in the form of a directed acyclic graph (DAG) and are, therefore, unable to capture the natural feedback loops of many environmental processes. This presents a problem of accuracy in system characterization in instances where a driver variable affects a response variable whose changes then alter the driver variable. For example, increased light penetration to a stream (perhaps caused by decreased riparian vegetation) may cause increased algal growth due to increased photosynthetic availability. However, if enough algae grow, the result is less light penetration to a stream because it is being blocked by the algae which grew in response to light. In a Bayesian framework, cause and effect can only be modeled unidirectionally. Feedback loops, therefore, have to be captured by having separate representations of the same variable through time (in a structure called a dynamic Bayesian network (Jensen and Nielsen 2007)), or redefining more specific aspects of a more general variable involved in a feedback loop (i.e., 'external light penetration' versus 'within' stream light penetration). Alternately, feedback loops can be eliminated by narrowing or broadening the scope of the model to either not include or average over the feedback loop in question. Depending on the circumstances, these solutions may or may not be appropriate to fulfill the modeling goals.

3.4.2.2. Determining ideal model scale

During model structure elicitation, we discovered it is vital to consider the spatial and temporal scale the model is to apply to and to ensure that variables and relationships between variables are represented at a matching, consistent scale. There is no single correct scale at which to use a Bayesian network; this is clearly a problem specific decision. However, if the modeling interest is to characterize large-scale, big-picture associations, then the data used should be appropriately aggregated or averaged to accurately represent this scale. In some sense, the fact that Bayesian network are assessments of the probabilities of mutual events occurring simultaneously may make them inappropriate for answering small-scale, process-based questions. If a detailed functional form is known to describe a particular cause and effect relationship, then discretizing variables and converting the continuous function into a conditional probability relationship may actually result in a loss of information. If relationships between small-scale system variables, particularly process based relationships, are known more exactly than in terms of discrete conditional probabilities, then perhaps they should be described deterministically or even using a continuous equation with error terms. Bayesian networks seem to be more appropriate for describing systems of a more probabilistic nature, where any combinations of events may occur with differing probabilities.

Additionally, when defining these events, it should be ensured that variables are measured and reported on similar scales. Otherwise, probabilities that they would be related are very low. In the original analysis of the EUSE macroinvertebrate dataset (Cuffney et al. 2009a), greater correlations were found directly between measures of urban disturbance and measures of invertebrate community response metrics than between invertebrate response and intermediate driver variables such as hydrology or water chemistry. Despite the well-documented fact that factors such as hydrology, habitat, and water quality affect invertebrates (Jones and Clark 1987, Kennen and Ayers 2002, Vølstad et al. 2003), the EUSE data may have sampled them on less integrated scales than aggregate urbanization and invertebrate measures and, therefore, found little relationship. Urbanization and invertebrate measures, however, may both be integrating influences on a larger watershed scale, and correlation findings may be mostly findings of scale match, not causal understanding. In fact, when structure learning was performed on the 85 data point Southeast EUSE dataset (Reckhow 2009), the strongest links 'learned' were those directly between urban land cover and invertebrate metrics. This outcome may have been simply the result of variable scale mismatch. It may not be reasonable to expect a reach-scale habitat measure or a grab-sample water quality measure to be representative of all the urban land cover in the whole watershed, given the diversity and variability in habitat between reaches and water quality between grab samples. In reality, given the state of urban land cover in

the watershed, there would likely be a different probability of seeing high or low conductivity. While this may be true for year-long average conductivity, it may not be true for an instantaneous measure of conductivity. Alternately, this finding may serve to highlight the difference between model use for prediction versus description purposes. While urban land cover may best predict invertebrate metric values due to variable scale match, this best prediction model offers little in terms of describing the system or offering management solutions for invertebrate condition mitigation.

Another aspect of scale, model structure scale, also affects the functioning of a Bayesian network model. If a model contains more nodes, more links between nodes (arrows) and more parents per node, it is more complicated to parameterize and (exponentially) more data are required to validate it. The Southeast model has only nine nodes, 16 arrows and, at most, three parents per node. However, even given this relatively low level of complexity, the model still has 967 parameters total in all its 7 conditional probability tables and 2 marginal probability tables (i.e., 242 separate Dirichlet distributions). (Each row of discrete probabilities in a conditional probability table is independent of each other row because there is no functional structure connecting them. They must each be parameterized independently.) Per child node CPT, this ranges from 16-256 parameters to estimate relative to 85 data points. Particularly for the most complicated invertebrate metric nodes, this results in many empty cells in their data tables (Appendix B Table 76-Table 78) because there are

roughly three times more parent/child state combinations probabilities than there are data to estimate them. Additionally, as the number of rows per CPT in the model increase, data weight per CPT decreases because there are fewer data to support each CPT row as the 85 samples are divided into greater and greater numbers of groups (parent/child state combinations).

This situation suggests many questions. Should modelers take data sample size into account when deciding the optimal network design? If so, how? Should experts take data sample size into account when providing prior probability weights, knowing that in data poor situations the posterior will be prior driven? How much data would be required to be reasonably confident in model maximum likelihood probabilities? What would be a reasonable ratio of number of parameters to data sample size? When data on a particular parent state combination are missing entirely and prior information is weak, what effect does putting a uniform distribution across all child node states for that parent state combination have on the model? Would this ever be appropriate to do in situations where that particular parent state combination is unlikely to occur? What criteria should be used to decide 'best' model structure scale?

In the southeast model, it is clear that expert elicitation is required to have any kind of estimate for the large data-empty model sections. Much more data than the 85 available measures would be required to even contemplate a data-only model without supplemental expert information. Data can be gathered from different sources to

parameterize different CPTs which may be a significant model improvement action considered in the future. Alternately, perhaps prior information can be strengthened with the use of multiple subject-matter experts providing knowledge about only CPTs which coincide directly with their expertise rather than using a single expert with general knowledge about the whole system. For now, it seems appropriate that the more complicated parts of the model have more posterior uncertainty associated with them in the form of smaller posterior weights from having less data weight during the combination of prior and data.

3.4.2.3. Discretization

A major unsolved concern with Bayesian network creation is the transformation of continuous variables into discrete ones. There is no clear answer to the question of how to best select number of bins and bin interval endpoints during Bayeian network parameterization. Factors considered include science-based cutoffs, equal data frequency in each bin, equal interval length in each bin, and moment matching with data. The benefits and drawbacks of these potential criteria have not been thoroughly studied, and there is no consensus on which type of method to use in which type of modeling situation. Unfortunately, implications of bin selection on model results and conclusions can be drastic. Once continuous variables are discretized, every possible value in a bin is treated the same way as every other value, significantly altering the type of information available from the data. In one study assessing structure learning

algorithms, totally different structures were learned from the same dataset using the same structure learning algorithm depending on how variable bins were assigned (Alameddine et al. 2010).

The number and size of bins, then, affects the dependencies that can be established in the model but it is not clear exactly how. In some cases, fewer bins lead to more certain descriptions of relationships (Myllymäki et al. 2002). This may be related to data sample size in that when more bins are created, more data are required to characterize those bins. For a constant sample size, fewer bins results in more samples per bin. Additionally, more bins lead to more uncertainty in experts' ability to discriminate between bins during prior elicitation. It is important for the future of Bayesian network modeling to better understand the effects of discretization on model functionality.

3.4.2.4. Inflexibility of model structure once defined

Conditional probability tables are constructed for a given model structure indicating which parent nodes affect a child node and for a defined number of bins and bin endpoint for all parent and child nodes involved. Once prior conditional probabilities are elicited, the model structure and bin definitions cannot change with the addition of data because then the prior information would not be defined in a consistent manner. In order to change nodes or arrows or bins, an entire set of prior probabilities must be re-elicited. So, while the Bayesian network construct allows for flexibility in the

original modeling decisions and in the ability to add unlimited future datasets, it is not flexible to changes in human-elicited prior specification. Since the elicitation of even simple conditional probability tables can be laborious and time consuming, this inflexibility places a major limitation on the incorporation of expert knowledge in modeling building. After completing model elicitation, if even minor changes to model structure want to be made, the entire time and energy intensive elicitation process must be repeated for all parent and child nodes affected by the change. To avoid this model structure limitation, perhaps after structure elicitation, model structure reasonableness should be assessed with data before expending effort on prior probability elicitation.

3.4.2.5. Ensuring conditional independence

A basic principle of Bayesian network calculation is the conditional independence of different parent and child node configurations. Yet rarely in Bayesian network model development is this principle effectively communicated to subject matter experts nor kept in consideration during the process of expert-based model building. In principle, if node A is a parent to both nodes B and C (Figure 47), then B and C are conditionally independent given A. This means that when A is unknown, then B and C are dependent because of their relationships to A, and information about B tells us something about the distribution of C. However, when A is observed, B and C become (conditionally) independent. Now, obtaining information about B will not garner any additional information about C over what is already known about C from A being

observed. Secondly, if B and C are both parents of D, and D is unobserved, then B and C are independent. This should be the case when we elicit the CPT of node D given nodes B and C. Only when D is observed are B and C dependent through their relationships to D. Finally, in this model structure (Figure 47), D is independent of A given B and C. If B and C are already known, learning A will give no additional information on D.

Without any observations on any nodes, multiple parent nodes are supposed to be independent of each other. If they are each independent influences on the child node then, only when the child node is observed can we expect to know something about one parent given information on another parent. A common example used to convey this point is one where an automated sprinkler (node B) and rain (node C) independently predict when grass may be wet (node D). Without knowing anything about the grass, knowing the sprinkler is automatically set to water the lawn at a particular time does not tell us anything about whether it is raining or not then. However, if the grass is wet, and we know the sprinkler was not on, then it is likely that it is raining. Similarly, in the southeast Bayesian network prior, before real of hypothetical evidence is available, each parent of a child node should be independent from all other parents.

In practice, it is not easy to ensure these properties hold during expert structure elicitation, which is focused on describing a larger-scale cause and effect relationship and is not intuitively concerned with conditional independencies. This may be the reason that some parent state combinations in the model are not likely to exist in reality. If parents were independent, then the state of one parent should not depend on the state of the other. However, situations in the southeast model arise where the state of one parent may be correlated with the state of another parent independent of the child node. For example, for both habitat and water quality child

nodes, no data exist for parent state combination low urban land cover (0-12%) and high flashiness (95+ rises) (Appendix B Table 74 and

Table 75). This may indicate that urban land cover and flashiness are not conceptually independent in the absence of data on habitat or water quality and should not both be parents of the habitat or water quality nodes. Or it just might mean that 85 samples are not enough to detect this particular parental combination. If large datasets were available, perhaps data could be checked for conditional independencies designated in the expert model as an iterative step in model development. If data were not available, perhaps conditional independence could be checked via hypothetical scenarios and thought experiments, although this solution is considerably less quantitative and less intuitive for experts. To build a technically correct Bayesian network, principles of conditional independence should be upheld during structure elicitation in a way experts can understand.

3.4.3. Prior weight evaluation

Another important consideration in Bayesian network building and updating is the assignment of prior weight. The three methods tested in this development project, direct elicitation (Winkler 2009), hypothetical sample (Winkler 2003, 2009), and probability range (Spiegelhalter et al. 1994, Cowell 1999, Jenson 2009) showed substantial variability both between and within methods, and the reason for this is not yet well understood. The best way to assess the strength of prior knowledge is still very

much an active research topic. The goal is to find a method that best captures expert certainty, but it is unclear which criteria should be used to decide that.

One major point of debate is the scale at which prior weight is most accurately elicited. Direct elicitation asked the expert to provide a prior weight describing strength of knowledge about an entire node while hypothetical sample and probability range methods asked for prior weight estimates at the level of individual cell (single child node state for one parent node state combination). The Hugin software, meanwhile, requires prior weight input at the level of CPT row. A major unanswered question is: how finely can humans differentiate their prior knowledge? Are individual cell elicitations meaningful? Are node scale elicitations too broad to capture within-node differences in certainty? Are prior weights elicited at one level applicable to narrower or broader scales? Another relevant consideration when choosing elicitation scale is desired time investment in prior elicitation, both in terms of whether a larger time investment will improve estimates substantially and how long a human being can continue to focus on providing accurate prior weights for a large problem construct.

Within-method accuracy was evaluated for hypothetical sample method and probability range method by calculating variability between prior weight estimates elicited for different CPT cells within the same node. Within-method variability was 1.4 to 15.6 times more variable for the hypothetical sample method than for the probability range method (Table 31), which was one of the reasons the probability range method

was selected for use in the model. Because direct elicitation was performed at the node level, there was no way to elicit more than one prior weight per node and, therefore, no way to assess within-method variability for a single node. Perhaps prior weight can be elicited directly for each CPT row but, again, can experts really distinguish between certainty of knowledge about a variable from certainty of knowledge about different states of a variable? As applied to all three prior weight elicitation methods, at which narrowness of scale do experts stop focusing on uncertainty about the specific nodes of interest and just report more general certainties about the extremes and the middle, no matter the variable?

This second major question asks about the effect of human bias in assessing extremes and whether it contributed to the documented differences between elicitation methods. Some studies have found that people report more confidence than warranted for probabilities of extremely low or extremely high values and less confidence for probabilities in the unknown middle (Wallsten and Budescu 1983, Morgan and Henrion 1990). If this human heuristic bias were to manifest itself in hypothetical sample elicitation and probability range elicitation, the expert would report smaller changes in hypothetical posteriors and tighter ranges for probability values closer to zero and 100. In fact, slightly greater prior weight does seem to have been elicited for extreme probabilities with the hypothetical sample method. However, this trend is not evident for prior weights elicited via probability range method (Figure 73). Perhaps another

benefit of the probability range method is minimization of extreme value overconfidence in the elicitation of prior weight.

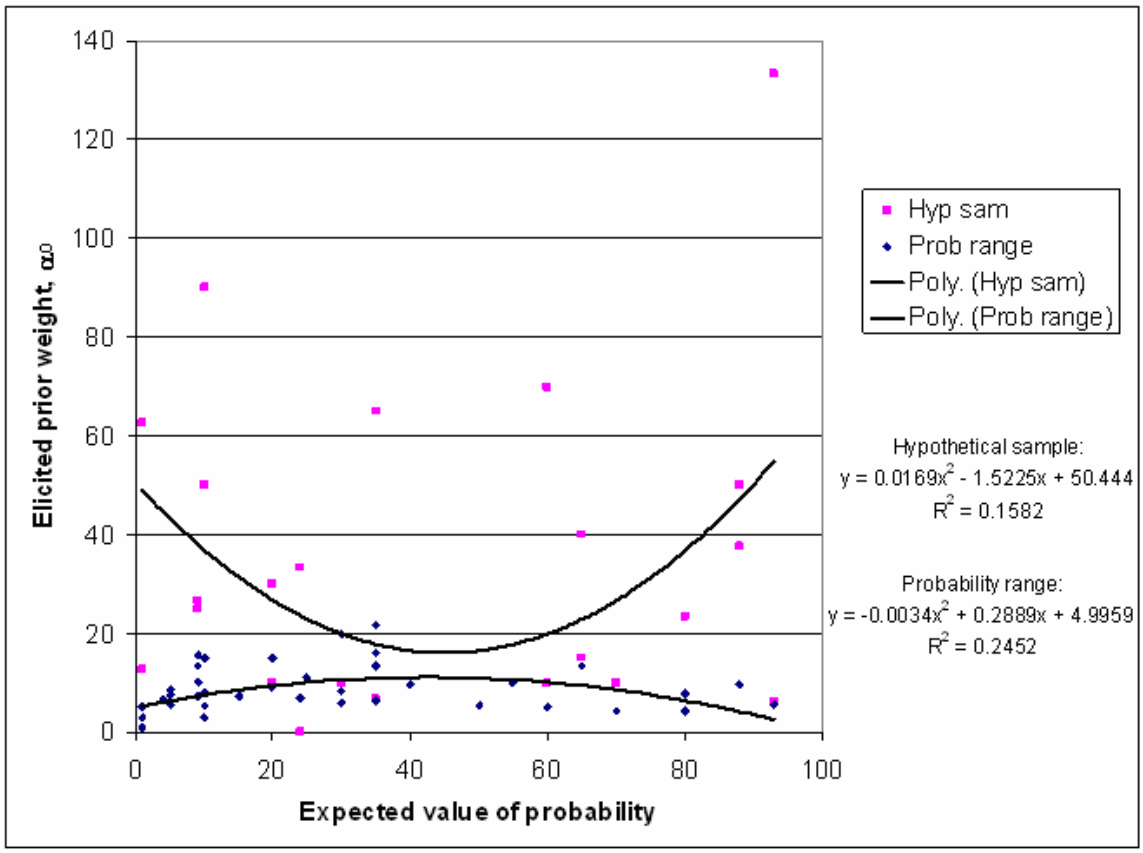

Figure 73: Assessment of possible expert overconfidence in extremes by comparing elicited prior weight across expected value of probability for hypothetical sample method ('Hyp sam', pink squares) and probability range method ('Prob range', navy diamonds) and eval

A third concern applicable to all three methods and, to some extent, to elicitation of conditional probabilities, as well, is the potentially biasing influence of experts thinking about probabilities in easily assessable units of 1 (whole numbers) or 5 or 10 when probabilities distribution assumptions are continuous. If experts are more likely, in reality, to report particular multiples of probabilities, is using continuous

distributions to describe these non-continuous probabilities introducing additional error? Should more analysis effort be expended to develop probability distributions that more closely reflect human thinking?

Using direct elicitation in reference to a constant number of data points led to a large range of prior weights for the seven model child nodes (from 21.25 for 'Change in generation of flow' to 0.80 for 'Richness-weighted tolerance', Table 31). Since prior weights were elicited proportionate to the amount of data available, this range was mostly due to the large differences in data sample size between nodes with as little as four parent states and, therefore, about 21 data points per CPT row ('Change in generation of flow') to as many as 64 parent state combinations and, hence, ~1.2 data points per CPT row (all three invertebrate metrics). It may be logical that prior weight decreases with increasing number of parents since this shows uncertainty increasing as the model becomes more complicated. However, should uncertainty increase so drastically? Is the expert really 23 times more certain of impervious surface knowledge than invertebrate metric knowledge? Relative to 10 reference data points, directly elicited prior weights only ranged between 5 and 10 equivalent data points (Table 28).

Final calculated prior weights, then, depended much more on data sample sizes than actual provided weights. Is this appropriate? Should prior weights be specific to how much data are available or independent of data? Posterior calculations completely depend on the prior weighting relative to the data weighting, so it seems reasonable to

elicit prior weights relative the actual data weighting to be used in posterior calculation, especially since data weight is so drastically different between child nodes. But does asking experts to compare their knowledge to information available from a reference number of data points provide accurate measures of prior knowledge weight? Without a reference, is a human mind capable of providing reliable estimates of numbers of equivalent data points that expert knowledge is worth? Even with a reference, how reliable are estimates? When asked, the expert reported being much less comfortable reporting prior weights directly than reporting hypothetical sample probability changes or probability ranges.

A fundamental question to ask of the hypothetical sample elicitation method is whether or not humans truly think in a Bayesian manner. If they do not update their prior beliefs with data in the same way that Bayes theorem would dictate, then the premise of the hypothetical sample method does not apply. For example, if an expert changes his probability assessment the same amount in response to different hypothetical samples, he is not thinking in a Bayesian manner. Bayes theorem would specify that, given a consistent prior weight (i.e., a consistent level of confidence in an assessment of a prior probability), a posterior probability should move further from the prior the further that the data are from the prior. That is, the prior weight specifies how much to weigh the prior probability relative to the data in the posterior, so the weighted average of the prior and data depends on the value of the data. Given a prior

probability of 0.10 and a prior weight of 5, data showing 2 of 10 samples updates the posterior to 0.17 while data showing 4 of 10 samples updates the posterior to 0.3 (Equation 59 where α_i is 0.5, based on $E[\theta_i]$=0.10 and α_0=5 in Equation 54, x_i is 2 and 4, respectively, and n is 10). In each case the prior weight specifies that the prior is worth half the weight of the data, and the posterior probability changes depending on what the data show.

However, it is not clear that humans think in this Bayesian manner, leading to mostly inconsistent prior weight calculations for two elicitations of the same prior weight using two different samples and the hypothetical sample elicitation method. For example, for the 'EPT taxa richness' child node first probability (flashiness bin 4, channel ratio bin 1, conductivity bin 2, and EPT taxa richness bin 4), the expert reported updating the prior probability of 0.01 to 0.05 regardless if data showed 1 or 3 of 10 samples to have high EPT taxa richness (Table 29). By definition, these results do not use a Bayesian thinking process. If Bayesian thinking were used, then the same prior weight should be used in calculation of the posterior in different data circumstances. But in this case, 0.05 is less than half of the way from 0.01 to 0.10 while 0.05 is less than one sixth of the way from 0.01 to 0.030, resulting in very different prior weights being calculated for the same prior probability depending on the hypothetical data (13 for x_i = 1 and 63 for x_i = 3; more prior weight when the posterior is closer to the prior than to the data). In fact, for all but 4 of 14 hypothetical sample elicitations, calculated prior weights

differed for the same prior probability, in many cases drastically so (the second 'Hydrology' probability for which prior weight was elicited shows prior weight of either 133 or 6.3, a range of 126.7 equivalent samples; Table 29). Not surprisingly, the expert reported the hypothetical sample method to be hard to conceptualize, especially in situations where hypothetical data were so far from the prior that they were not believable.

Although the probability range method appeared to be the most acceptable both in terms of expert comfort and variability and bias minimization, it was not without its own questionable aspects. First of all, how can we assure that the two randomly selected CPT rows used in elicitation are truly representative of the prior weight of the entire child node? The prior weight elicited may depend on which row is selected. Given the time and labor intenseness of the elicitation process, what is the best way to select rows to elicit that will provide a good characterization of 'true' prior weight? Even within a row, probability range elicited prior weights vary. How can this variation be minimized to consistently assess strength of prior knowledge?

As touched upon in the scaling discussion above, it may be invalid to assign the average of all elicited CPT rows to all possible CPT rows in the node. Perhaps some rows truly have more uncertainty associated with them than others. But, again, can these different uncertainties be distinguished? Particularly with the probability range elicitation method, is it not clear that experts can even distinguish uncertainties at

different nodes when asked to provide a range for a single probability. For such a narrow question, are experts really thinking about the subject matter or just reporting their general sense of uncertainty for the entire modeling context? Ranges, then, may depend more on values of mean (expected value) probability than on node, and however an expert responds, he or she would respond the same way for all nodes. How can we know if the fact that many of the reported probability range method prior weights are similar across nodes is due to true similarities in certainty of prior knowledge between nodes or too narrow of a focus on specific probability mean rather than uncertainty at node level?

3.4.4. Future improvements and research

In developing this southeast model describing the effects of urbanization and stream biota, it would be extremely useful to include explicit management nodes for the probabilistic evaluation of the effect of various management actions with uncertainty. As part of this model expansion, different groups of experts with career knowledge specific to each appropriate part of the model (such as urban planners, hydrologists, habitat scientists, and biologists) should be consulted for input on model structure as well as conditional probabilities.

Technical improvements include decreasing the number of bins per node since the southeast model expert reported a difficulty in distinguishing between the two middle (medium-low and medium-high) nodes. In addition to obtaining more precise

estimates of expert knowledge, this model amendment will likely result in tighter relationships between variables known to be associated with fewer bins (Myllymäki et al. 2002). Model flexibility can be improved by introducing latent variables to capture influences and responses of unmeasured or unmeasurable system components (such as flow generation). Incomplete datasets can also be supplemented by incorporating missing data approximation algorithms in posterior calculation.

Broader future research should be conducted on better visual representation of conditional probability tables, particularly as the number of dimensions in each table increases, to enable clearer understanding of discrete relationships between variables as they are changing in multiple dimensions. Expert knowledge elicitation methods should be refined to guarantee consistency and accuracy in both conditional probability elicitation and prior weight elicitation. Perhaps combining some form of relationship visualization to use during elicitation may achieve both goals.

The value of Bayesian network construction prior to the design of studies has greatly been realized over the course of this project. Incorporating expert knowledge systematically in the beginning to organize big picture system understanding can be immensely helpful in future work. This type of pre-assessment can help guide important study decisions, particularly those that define study site characteristics and select which variables to measure. Making these decisions in the greater framework of

system understanding can contribute to more efficient and productive study design and execution.

3.5. Conclusions

The main benefits of using a Bayesian network to model the process of urbanization affecting aquatic endpoints include:

(1) The ability to model this process as an interacting system instead of as separate, unrelated system components. This created network much more realistically captures the multiple, simultaneous effects of variables on all other variables in the system, and allows for interpretation of the modeling questions in terms of a manipulatable cause-and-effect network.

(2) The incorporation of both expert knowledge and data into the final model, which ensures that the model is informed with the greatest possible amount of available knowledge. Additionally, using the Bayesian modeling construct, we can designate the relative weight of incorporated expert knowledge and data according to how much each should contribute to final model.

(3) Explicit acknowledgment of many kinds of uncertainty inherent in environmental model development including random variability, biased and random measurement error, inaccuracies in assumptions, lack of empirical basis for making a characterization, or imperfect knowledge about how a system works. These

uncertainties are incorporated logically and comprehensibly, resulting in honest system characterization.

(4) Easy adaptability to the inclusion of more information on any variables or linkages in the network. Because of the use of a Bayesian construct, the current posterior probabilities can become the new prior probabilities to be updated later every time more data become available. New data will be weighed appropriately relative to amount of information already incorporated in the model.

(5) Flexible model construct does not require explicitly knowing the parametric form of all linked relationships. Conditional probability tables can be constructed empirically. This flexibility is especially useful in situations where we are not sure of exact parametric relationships because they are too complicated and multi-tiered, but we do understand simpler relational characteristics between variables such as which values increase when other values decrease and to what degree. Additionally, model structure is decomposable so that we can focus on and parameterize one set of relationships at a time, but can then seamlessly combine these sets of relationships into a model of an integrated, larger scope.

(6) Interactive end product which is easy for users to understand without necessarily being bogged down with the complex mathematics that the model contains. The final model can be used in management contexts to predict effects of possible actions and diagnose causes of desired results.

This developed methodology creates models which connect urbanization to biological response at multiple scales while simultaneously characterizing uncertainty and incorporating expert knowledge. This new way of prior development and data updating contributes to better understanding of the use of Bayesian networks in answering environmental and potentially many other types of complex questions.

4. Linking Urbanization to an Explicit Biological Standard in the Northeast United States Using a Bayesian Network Approach

4.1. Summary

Urbanization is known to adversely affect stream biology through a complicated system of impacts on vital environmental processes. The Biological Condition Gradient (BCG) is a method of systematically defining levels of ecosystem health using characteristics of ecosystem structure and function that respond to increasing stress. BCG levels, or Tiers, help to standardize interpretation and communication of the condition of aquatic biota (collected with differing sampling methods and in diverse ecological settings), facilitate detection of incremental changes in ecological condition, and enable linking of management actions to meaningful environmental outcomes. We translated the BCG conceptual framework into a Bayesian network of quantifiable nodes, relationships and probabilities to describe the effect of urbanization on macroinvertebrate biological condition in the Northeast U.S. This was done by integrating U.S. Geological Survey data with expert elicitation from experienced New England biologists to create a set of probabilistic linkages connecting urbanization metrics to interpretation of macroinvertebrate biological condition (via BCG Tiers). By incorporating not only information available from data but also expert knowledge, and uncertainty associated with both data and experts, these probabilistic linkages are able to thoroughly characterize the system of interest. Managers can interactively use this

parameterized Bayesian network to calculate the probability of attaining desired aquatic ecosystem BCG tiers assuming different levels of urban stress, environmental conditions and, ultimately, management options.

4.2. Introduction

4.2.1. Effect of urbanization on streams

Urbanization of a watershed unleashes a cascade of compounding large and small scale stream changes which ultimately disrupt the structure and function of aquatic biotic communities. As impervious surface of roads, parking lots and roofs increases, water movement and storage in a watershed is drastically altered resulting in increased runoff, decreased infiltration, and a more extreme water cycle in both time and space (Klein 1979, Poff et al. 1997, Finkenbine et al. 2000, Konrad and Booth 2002, McMahon et al. 2003). These erratic new hydrological patterns lead to a suite of physical stream system disturbances such as erosion from construction sites and stream banks, channel scouring and incision (Booth 1990), sedimentation (Wolman and Schick 1967, Trimble 1997), flooding, droughts, and disconnection of the floodplain from the stream (Haase 2003). As a result of these disturbances, the composition of stream substrates that support many kinds of life is modified (Sponseller et al. 2001, Roy et al. 2003b) and the reach diversity required to sustain complex ecosystems is reduced or eliminated (Winterbourn and Townsend 1991). In addition to physical changes, storm runoff from urban areas is chemically altered, containing increased concentrations of anthropogenic

pollutants such as industrial chemicals, pesticides, and human waste (Hall and Anderson 1988, Pitt et al. 1995, Van Metre et al. 2000, Beck 2005, Mahler et al. 2005, Gilliom et al. 2006). Deforestation which often accompanies urbanization simultaneously destroys riparian zones (Groffman et al. 2003) which are then no longer available to filter and process these pollutants, or provide canopy cover or litter to help manage stream metabolism via nutrient, energy and temperature cycles (Waite and Carpenter 2000, Jacobson 2001, Sprague et al. 2006).

These interacting effects of urbanization on stream hydrology, habitat, and water quality eventually affect the amount and type of biota able to live in urban streams. While community metrics measuring fish and algal conditions show some patterns with significant noise, urbanization is clearly correlated with degraded invertebrate assemblages in urban areas (Jones and Clark 1987, Kennen 1999, Yoder et al. 1999, Ourso and Frenzel 2003). Invertebrate communities can be characterized in many different ways via total taxa diversity, feeding group types, movement types, life stages, or sensitivity to pollution. Typically, as urban land cover increases in a watershed, total invertebrate taxa richness decreases, sensitive species' richness and abundance decrease and abundance of tolerant species increase.

4.2.2. Biological Condition Gradient (BCG)

The amount and composition of benthic macroinvertebrate species found at a stream sample site can tell much about the biological health of that ecosystem. While

there are a variety of metrics which have been developed to attempts to describe macroinvertebrate communities (Lenat and Barbour 1994, Barbour et al. 1999, USEPA 2005), the interpretation of these community metrics is very specific to the region in which they are measured and to the detailed sampling protocol used. This makes it challenging to compare invertebrate community health across regions and sampling programs.

The Biological Condition Gradient (BCG) systematically defines levels of ecosystem health using characteristics of ecosystem structure and function that respond to increasing stress (Davies and Jackson 2006). BCG levels, or Tiers, help to standardize interpretation and communication of the condition of aquatic biota (collected with differing sampling methods and in diverse ecological settings), facilitate detection of incremental changes in ecological condition, and enable linking of management actions to meaningful environmental outcomes. These six BCG Tiers are defined from best (Tier 1) to worst (Tier 6) ecological status in terms of ecosystem structure and function (Table 36).

Table 36: Biological Condition Gradient (BCG) Tier definitions.

	Structure of biotic community	**Ecosystem function**
Tier 1	Natural or native condition	Natural or native condition
Tier 2	Minimal changes	Minimal changes
Tier 3	Evident changes	Minimal changes
Tier 4	Moderate changes	Minimal changes
Tier 5	Major changes	Moderate changes
Tier 6	Severe changes	Major loss

Each tier is envisioned to exist as a result of a particular level of a system stress, such that biological condition decreases as stress to the system increases (Figure 74).

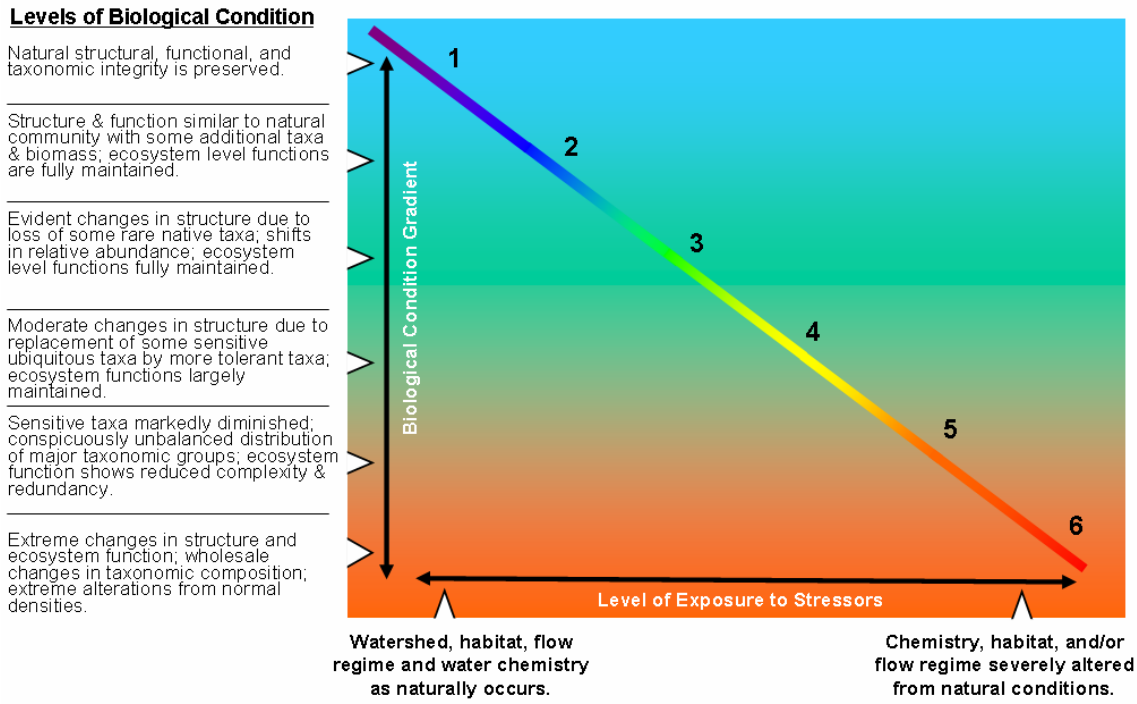

Figure 74: Biological Condition Gradient Tiers relative to increasing stressor gradient.

'Community structure' and 'ecosystem function' are further defined via ten attributes, where Attributes I to VI specify community structure characteristics and Attributes VII to X specify ecological function characteristics (Table 37).

In essence, the BCG is only a set of general descriptions of what each attribute should include and what each Tier should look like based on best professional ecologist and biologist opinion (Jackson and Holdsworth 2007). By design, the BCG directive does not specify exactly how to measure each of these attributes nor what levels of a

Table 37: Biological Condition Gradient (BCG) Attribute definitions.

Attribute	Characteristics
I	Historically documented, sensitive, long-lived or regionally endemic taxa
II	Sensitive-rare taxa
III	Sensitive-ubiquitous taxa
IV	Taxa of intermediate tolerance
V	Tolerant taxa
VI	Nonnative or intentionally introduced taxa
VII	Organism condition
VIII	Ecosystem function
IX	Spatial and temporal extent of detrimental effects
X	Ecosystem 'connectance'

particular measurement would be indicative of each Tier in order to allow for region and sampling program specific differences in what those answers may be. So in order to use the BCG for ecological status classification in a particular region using data from a particular sampling program, BCG users need to determine and quantify: (1) which metrics or variables can be used to measure these ten attributes and (2) which values of these metrics or variables in combination assign a basin to a particular BCG Tier with what certainty. Additionally, the BCG itself offers no functional relationship between BCG and stressor conditions other than stipulating, in general, that biological condition deteriorates with increasing stress. Assignment of BCG tier to a watershed does not automatically link that biological outcome to the causing stressor nor does it recommend management options to mitigate biological condition.

The benefits of BCG discrimination include: enhanced protection for exceptional quality waters, understanding of the type and degree of management to implement for

impaired waters, ability to detect incremental improvement or worsening at a finer scale, ability to identify realistically attainable restoration goals, better ability to diagnose problems and causes with increased detection sensitivity, and early warning detection to find problems and correct them while they are still small (Davies and Jackson 2006). However, to make BCG classification functional, the BCG conceptual framework needs to be translated into a quantifiable, interactive model format which captures the process of stressor impact on biological endpoints and management goals.

4.2.3. Modeling purpose and goals

The two major project objectives are to:

1. Capture multiple interacting environmental and ecological processes affected by urbanization using system-scale Bayesian network model (i.e., approach multiple stressor problem in a new, more realistic manner), and
2. Link urbanization effects on stream biota to management context; describe and quantify how changes in drivers lead to changes in regulatory endpoint (BCG).

4.3. Data

The U.S. Geological Survey (USGS) conducted a series of urban stream studies (Effects of Urbanization on Stream Ecosystems, EUSE) from 1999-2004 as part of the National Water-Quality Assessment (NAWQA) Program. These studies were designed to comprehensively assess the physical, chemical, and biological condition of streams across an urban gradient. using a common study design (McMahon and Cuffney 2000,

Coles et al. 2004, Cuffney et al. 2005, Tate et al. 2005) and consistent measures of urban intensity (Cuffney and Falcone 2008) and sample-collection and processing methods (Fitzpatrick et al. 1998, Moulton et al. 2002). EUSE investigated approximately 30 watersheds of differing levels of urbanization in each of nine different metropolitan regions, one of which was the urban center of Boston and surrounding areas in Massachusetts, New Hampshire, Maine, and Connecticut (New England Coastal Basins, NECB). By substituting space for time, the EUSE study was able to detect the multitude of environmental and ecological stream changes occurring with increasing urbanization.

Urban intensity was measured using land cover and census data to report variables such as percent urban land cover in the watershed basin, percent impervious cover in the watershed basin, population, housing density, and road density. Physical stream characteristics measured include hydrological metrics describing water flow, habitat metrics describing channel shape, light and litter input, soil size and texture, air and water temperature, precipitation, topography and many other measurements. Chemical measurements include various nitrogen and phosphorus nutrient indicators, a suite of pesticide concentrations, and many organic compound measurements using semipermeable membrane devices. Biological variables include community metrics describing fish, invertebrates and algae. EUSE studies have already been used to describe the effects of urbanization on fish (Brown et al. 2009), benthic macroinvertebrates (Cuffney et al. 2009a), algae (Coles et al. 2009), habitat (Faith A.

Fitzpatrick, U.S. Geological Survey, written commun., 2009), and water chemistry (Sprague et al. 2007).

Multilevel hierarchical modeling analysis of all 270 streams in the 9 total metropolitan areas showed that baseline, pre-urbanization conditions and rates of response to urbanization differ among ecoregions (Kashuba et al. 2010). Additionally, each ecoregion has its own endemic flora and fauna composition which indicate biological condition specific to a region. Therefore, system-level urbanization modeling should be conducted at the ecoregion scale to ensure consistency in urbanization patterns and biological condition definitions. In our seminal Bayesian network urbanization modeling methods development, we chose to use the EUSE Southeast United States dataset containing 85 watersheds sampled for the Atlanta, Birmingham, and Raleigh EUSE study areas (Chapter 3). Once methods were developed and vetted on this relatively larger dataset, we chose to focus methods application research on the EUSE Northeast United States dataset (Boston area, 30 watersheds) in order to also be able to utilize the expertise of the Northeast biologists who created the BCG concept (Davies and Jackson 2006).

4.4. Methods

We are interested in modeling the effect of urbanization on macroinvertebrate biological condition in the Northeast United States, ultimately interpreting ecosystem response in terms of the probability of meeting BCG tiers. Challenges include

representing and parameterizing complex, interacting urban stream system dynamics, incorporating current existing ecological knowledge, accounting for system-wide uncertainty from many sources, and ensuring ease of understanding and interactive manipulation in a management context. An innovative Bayesian network modeling construct is used to address these challenges.

4.4.1. Bayesian network model

The major model goal is to describe the interrelationships between all variables in the system simultaneously such that we could know the effect of changing one variable on all the remaining variables. Unfortunately, a functional form for the joint distribution across all variables becomes exponentially more difficult to define as variables are added to the system, becoming intractably large for relatively few variables. A Bayesian network is essentially a compact way of storing information about this entire joint distribution of multiple related variables (Heckerman 1999, Spirtes et al. 2000, Neapolitan 2004, Taroni et al. 2006, Jensen and Nielsen 2007, Kjaerulff and Madsen 2008, Pourret et al. 2008). It consists of a directed acyclic graphical model of nodes representing important concepts and arrows representing probabilistic relations between those concepts. Nodes from which arrows point are called 'parent nodes' and nodes to which arrows point are called 'child nodes'. This modeling framework allows decomposition of a complicated system into sets of conditional probabilities describing the direct effect of a parent node or nodes on a child node. Child nodes are then

themselves modeled as parent nodes of subsequent child nodes, eventually linking the entire system of interest into one statistical representation. In this way, a complicated joint distribution of all variables is deconstructed into the product of more easily specified conditional probabilities *only between connected nodes*. By defining a linked network system in terms of its component conditional probabilities, we greatly reduce the difficulty of parameterizing a model of high dimensionality while retaining the complexity of a joint system representation.

Traditional analysis of EUSE stream ecosystem data attempts to find empirical relationships only between single pairs of environmental concepts at a time using simple regression techniques (Cuffney et al. 2009a), without incorporating the web of additional interconnected environmental variables, uncertainty characterization, or known ecological information about the system. Relative to frequentist modeling constructs, this Bayesian approach is incredibly enticing to aquatic ecologists for its ability to model the whole set of system components that are interacting at different levels in a simple, understandable, probabilistic manner. Many anthropogenic and natural factors affect invertebrates and, rather than investigating each factor individually, using a Bayesian network to describe the effect of urbanization on invertebrate condition acknowledges the complexity of the environmental and ecological processes driving biological response and allows concurrent assessment of all driving factors.

In addition to encouraging systemic understanding in the modeling of scientific processes, Bayesian networks allow systematic incorporation and weighting of both expert knowledge and data, quantify relationships between variables as well as uncertainty in those relationships, reduce input complexity requirements because of their decomposability, adapt flexibly when new data or knowledge becomes available, and provide a graphic interface that is easy to comprehend and manipulate. Utilizing these beneficial qualities, we have developed a modeling approach (Chapter 3) resulting in a fully parameterized and data-updated Bayesian network. Managers can interactively use this model to calculate the probability of attaining desired aquatic ecosystem endpoints (BCG tiers) assuming different levels of urban stressors, environmental conditions and, ultimately, management options.

Bayesian network creation involves two major steps: (1) development of the prior model and distributions and (2) updating the prior with data to obtain the posterior model and distributions. We elicit all aspects of the prior from subject matter experts, including the model structure, variable selection and binning, conditional probabilities and prior weights. This prior model is then updated with the Northeast EUSE data to create the posterior model. Per Bayesian philosophy, we are updating prior knowledge of how a system works upon the availability of additional information about the system. Upon completion, this posterior model can be used to better understand the relationships within the parameterized network of nodes through the propagation of

'evidence', also called predictive or diagnostic probability calculation (Jensen and Nielsen 2007). By setting the value of any parent (driver) node, a Bayesian network can 'predict' the expected distribution of all child (endpoint) nodes down the network from that driver. A Bayesian network can also 'diagnose' the likely distributions of parent (driver) nodes required to achieve a set child (endpoint) node value. In both cases (either following the direction of the arrows or in reverse), using the probabilistic relationships between the nodes, basic statistical principles, graph theory, and Bayes Theorem, the Bayesian network updates what we know about a node based on actual or hypothetical 'evidence' on another node.

4.4.2. Conjugate Dirichlet-multinomial updating

The mechanism of Bayesian network updating, and, hence, the ability to combine expert judgment with data, derives from the principles of Bayes Theorem (Bayes 1763):

$$p(\theta \mid x) \propto p(\theta)\, p(x \mid \theta)$$

(79)

where

$p(\theta)$ is the prior probability distribution of θ, the parameter set of interest,

$p(x \mid \theta)$ is the likelihood function of x, the data, given θ, and

$p(\theta \mid x)$ is the posterior probability distribution of θ, which is a combination of the prior information and the data collected.

In general, Bayesian analysis attempts to gain understanding about the distribution of all possible values of a parameter set of interest, θ. This distribution, $p(\theta)$, is originally constructed from information available prior to the collection of a specific dataset. Once data is available, the likelihood function, $p(x \mid \theta)$, calculates the likelihood of seeing those data given each possible value of θ. To update the prior information with the additional information from the data, the prior distribution is multiplied by the data likelihood function resulting in a posterior distribution of θ conditional on the data observed. This posterior distribution, $p(\theta \mid x)$, is essentially a weighted average of the prior information and the data.

In theory, prior parameter distributions can take any functional form imaginable. However, if no analytical solution for the posterior distribution function is possible, numerical approximation techniques such as Markov Chain Monte Carlo (MCMC) sampling are often computationally laborious and time intensive. If, instead, the prior distribution and the data likelihood function belong to what is called a conjugate family, then a posterior solution of the same functional form as the prior distribution is guaranteed.

In the context of a Bayesian network, each node is divided into a discrete number of categories. The parameter sets of interest, then, are the probabilities that a node's value will fall in each of the categories. For example, say a node representing urban disturbance has three possible categories: high (>31-100% urban land cover), medium

(>7-31%) and low (0-7%). Then the probability that a random Northeast U.S. watershed has high, medium or low urban land cover can be described by parameters θ_1, θ_2, and θ_3, respectively. The possible values of these probabilities can be modeled using a Dirichlet distribution. The Dirichlet distribution form is:

$$p(\theta_1, \theta_2, \theta_3) \propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \theta_3^{\alpha_3 - 1}$$

(80)

$$\sum_{i=1}^{3} \alpha_i = \alpha_0$$

(81)

where

$\theta_1 - \theta_3$ are the probabilities of a sample belonging to categories 1-3, respectively,

$\alpha_1 - \alpha_3$ are coefficients describing the possible values of $\theta_1 - \theta_3$, respectively, and

α_0 is the sum of $\alpha_1 - \alpha_3$, also known as total prior weight or equivalent sample size.

Data collected measuring the urban land cover percentages of a random sample of watersheds, discretized into the three defined urban land cover categories, can be described using the multinomial likelihood function:

$$p(x_1, x_2, x_3 \mid \theta_1, \theta_2, \theta_3) \propto \theta_1^{x_1} \theta_2^{x_2} \theta_3^{x_3}$$

(82)

$$\sum_{i=1}^{3} x_i = n$$

(83)

where

$x_1 - x_3$ are the counts of samples belonging to each of three categories, respectively,

n is the sum of all counts (i.e., total sample size), and

$\theta_1 - \theta_3$ are the probabilities of a sample belonging to categories 1-3, respectively.

The Dirichlet-multinomial family is conjugate. This means if the Dirichlet distribution is used to quantify prior knowledge about the possible values of θ_1, θ_2, and θ_3, and multinomial data, x_1, x_2, and x_3, are collected, then the posterior distribution combining prior knowledge and data will also be of the Dirichlet form. To calculate this posterior distribution, the Dirichlet prior distribution is multiplied by the multinomial likelihood function per Bayes Theorem (Equation 1) resulting in a Dirichlet posterior distribution with new coefficient values:

$$p(\theta_1, \theta_2, \theta_3 \mid x_1, x_2, x_3) \propto \theta_1^{x_1+\alpha_1-1} \theta_2^{x_2+\alpha_2-1} \theta_3^{x_3+\alpha_3-1}$$

(84)

$$\sum_{i=1}^{3} (\alpha_i + x_i) = \alpha_0 + n$$

(85)

where

$\theta_1 - \theta_3$ are the probabilities of a sampled basin belonging to bins 1-3, respectively,

$\alpha_1 - \alpha_3$ are prior coefficients describing the possible values of $\theta_1 - \theta_3$, respectively,

α_0 is the sum of $\alpha_1 - \alpha_3$, also known as total prior weight or equivalent sample size.

$x_1 - x_3$ are the counts of sampled basins belonging to each of three bins, respectively,

and

n is the sum of all counts (i.e., total sample size).

The Dirichlet posterior estimated values for θ_1, θ_2, and θ_3 are a weighted average of prior estimates and data estimates. For example, say the expected value of θ_1 from the Dirichlet prior was 0.20 with a prior weight of 5 (i.e., $\alpha_0 = 5; \alpha_1 = 1$) and data showed that 4 out of 10 samples fell into category 1 (i.e., $x_1 = 4; n = 10$, so maximum likelihood estimate of θ_1 is 0.40). The posterior expected value of θ_1 calculated from the Dirichlet posterior would be the average of 0.20 weighted by $\alpha_0 = 5$ and 0.40 weighted by $n = 10$:

$$E[\theta_1 \mid x_1, n] = \frac{\alpha_1 + x_1}{\alpha_0 + n} = \frac{1+4}{5+10} = 0.33 \tag{86}$$

This posterior expected value of θ_1, 0.33, is the average of 0.20 and 0.40 where 0.40 has twice the weight of 0.20, as specified by the prior weight relative to the data weight (sample size). In this way, prior and data information is combined, appropriately weighted. Another interpretation is that the prior expected value of θ_1, 0.20, is updated with information from data.

4.4.3. Expert elicitation and prior model development

Expert elicitation specifically refers to the translation of expert knowledge into a probabilistic framework (Morgan and Henrion 1990, Frey 1998, Winkler 2003, Reckhow et al. 2005, O'Hagan et al. 2006). Usually scientists prefer to describe the model development process as 'objective' without acknowledging that model and parameter selections are actually made via expert judgment implicitly. Expert elicitation, in contrast, is an explicit incorporation of expert judgment into the modeling process. The goal is to codify expert knowledge, appropriately recognizing the degree of uncertainty in that knowledge, and then use that knowledge directly in model development and parameterization. Using expert elicitation in model development thus focuses on the science of the problem by confirming or denying expected relationships (hypothesis approach), instead of blindly calculating statistics for all possible models and variable combinations.

Expert knowledge is vital to the development of the prior portion of this Bayesian network modeling construct. Unlike data gleaned from a single study, experts provide an integrated estimate of system relationships and uncertainties based on all information synthesized from career's worth of experience. More heavily weighing prior knowledge in which experts have high confidence can also address known problematic data concerns.

To develop this northeast U.S. Bayesian network prior model, multiple groups of experts were recruited with vast career experience in areas of urban planning, water management, water chemistry, basin-scale habitat evaluation and aquatic ecology (Table 38). Following a refined, unique, systematic informed expert prior creation methodology (Chapter 3), model structure, variable selection and binning, conditional probabilities and prior weights were elicited from these groups of subject matter experts.

Model structure refers to the selection and arrangement of Bayesian network nodes and arrows. The goal of structure elicitation is to represent the chain of events through which various aspects of urbanization affect aquatic invertebrate communities. Initial model structure was elicited via textual analysis of the transcript from an open-ended causal narrative interview with a USGS career ecologist. Using this causal map development method (Nadkarni and Shenoy 2004), the expert was asked to describe, step by step, how elements of urbanization contribute to a biologically degraded stream condition. Then, formal textual analysis identified major concepts and causal relationships between concepts which were converted into structure nodes and arrows, respectively. In model structure development, it is important to eliminate reiteration of the same concepts, to clearly define causal pathways, to avoid scale mismatch between concepts, and to balance model complexity with ability to parameterize (Reckhow 1999). Therefore, in an iterative process of expert feedback from all subject matter groups, the initial causal map was pared down into a model structure affirmed by expert consensus

Table 38: Northeast United States urbanization effects model expert teams.

Biologists and aquatic ecologists: • Susan Davies, Maine Department of Environmental Protection • Dave Courtemanch, Maine Department of Environmental Protection • Tom Danielson, Maine Department of Environmental Protection • Susan Jackson, US Environmental Protection Agency, Region 1 • Jeroen Gerritsen, Tetra Tech Inc. • Jim Coles, US Geological Survey, New Hampshire-Vermont Water Science Center • Tom Cuffney, US Geological Survey, North Carolina Water Science Center
Water management assessors: • Peter Weiskel, US Geological Survey, Massachusetts-Rhode Island Water Science Center • Marilee Horn, US Geological Survey, New Hampshire-Vermont Water Science Center • Karen Beaulieu, US Geological Survey, Connecticut Water Science Center • David Armstrong, US Geological Survey, Massachusetts-Rhode Island Water Science Center • Chris Waldron, US Geological Survey, Massachusetts-Rhode Island Water Science Center
Habitat scientists: • Faith Fitzpatrick, US Geological Survey, Wisconsin Water Science Center • Marie Peppler, US Geological Survey, Wisconsin Water Science Center • Barbara Scudder, US Geological Survey, Wisconsin Water Science Center • Amanda Bell, US Geological Survey, Wisconsin Water Science Center
Urban planners and managers: • Paul Sturm, Center for Watershed Protection • Bill Stack, City of Baltimore Department of Public Works, Water Quality Management Service • Kernell G. Ries, US Geological Survey, Maryland Water Science Center • Ronald Bowen, Anne Arundal County, Maryland, Department of Public Works • Janis Markusic, Anne Arundal County, Maryland, Department of Public Works • Chris Victoria, Anne Arundal County, Maryland, Department of Public Works • Joe MacDonald, American Planning Association • Karen Cappiella, Center for Watershed Protection • Hala Flores, Anne Arundal County, Maryland, Department of Public Works

that represents all important processes and causal influences at the selected scale while also being of sufficiently manageable size to allow for parameterization (Figure 75).

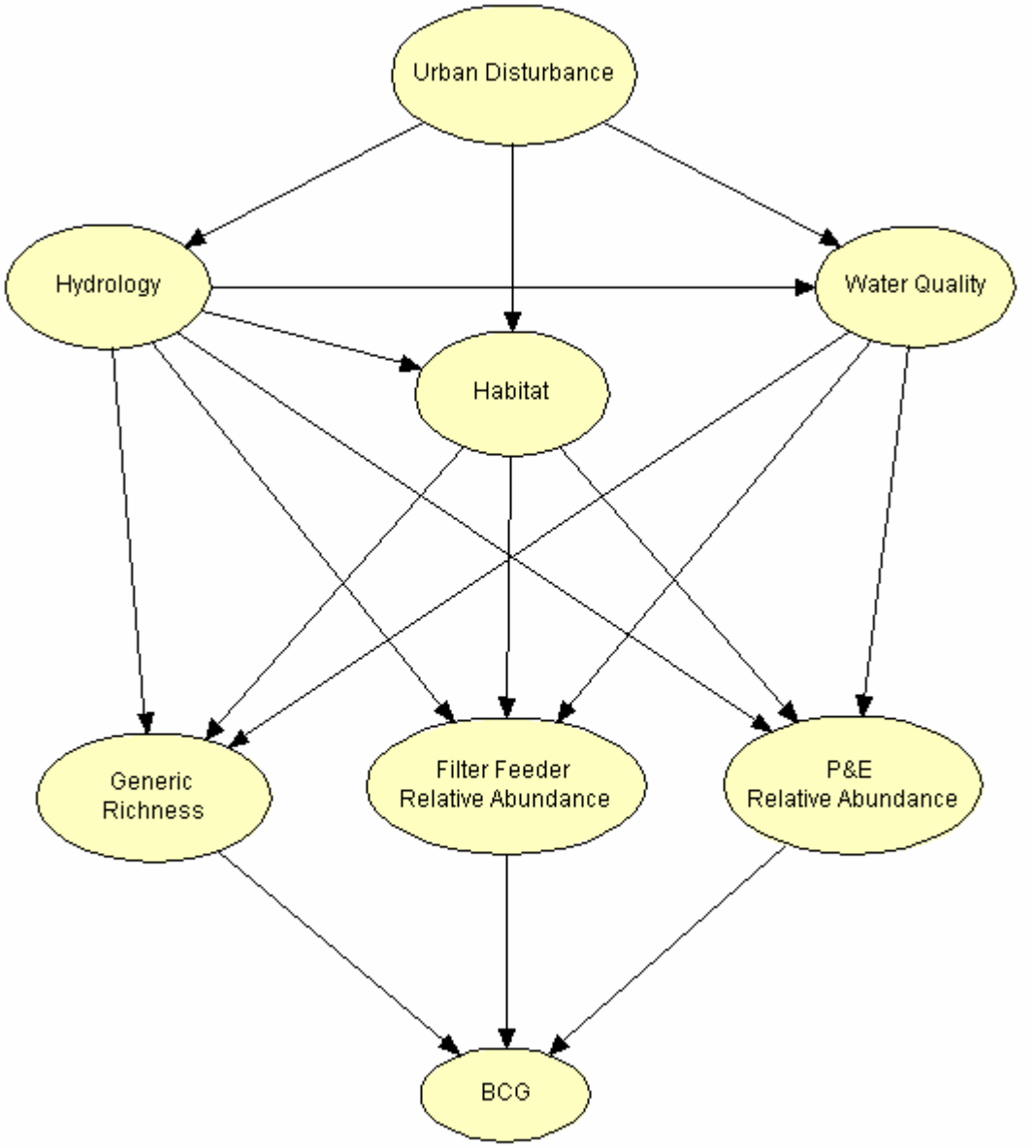

Figure 75: Parsimonious Northeast model structure.

Invertebrate metrics were selected using a two-pronged approach. After the process of BCG Tier assignment (see *BCG Tier assignment* section), experts were asked to

identify the three most vital invertebrate metrics whose values influenced assessment of biological condition. Simultaneously, classification and regression tree (CART) and ordered logistic regression (OLR) models were constructed to evaluate which metrics were most predictive of BCG in the EPA New England Wadeable Streams (NEWS) dataset (Jackson and Holdsworth 2007). The three top invertebrate metrics identified by both experts and statistical analysis were chosen for use in the Bayesian network model (i.e., Generic richness, Filter feeder relative abundance, and Plecoptera and Ephemeroptera relative abundance).

Model structure development involves only identifying major qualitative concepts and their causal connections with each other. To be able to transform such a causal map into a Bayesian network, a quantifiable variable must be assigned to each concept and, as explained above, each continuous variable must be converted into a discrete number of mutually exclusive categories. For each model node, experts rely on career experience, inter-expert discussion and consensus building to select an EUSE dataset variable which best represents that node's intended concept (Table 39). Once the representative variables are identified, the number of categories each node is to be split into must be chosen. This choice involves a tradeoff between the model's discriminatory strength and the accuracy of (potentially complex) parameterization. Having few bins limits the ability of the model to differentiate between different states of a node and to evaluate changes to the variable and the system. However, the more bins a variable is

Table 39: Nodes and variables chosen by experts to represent major system components in Northeast United States Bayesian network model.

Node	Variable: units	Bins					
		low		medium		high	
Urban disturbance	Urban land cover: percent urban land cover in basin area	0-7%		>7-31%		>31-100%	
Hydrology	Flashiness: rises greater than seven times the annual median rise	0		1-3		4 +	
Habitat	Substrate: dominant (>50% of transects) substrate type	fine (sand and smaller)			coarse (gravel and larger)		
Water quality	Conductivity: at low base flow, μsiemens per centimeter at 25°C	0-139		>139-269		>269	
Generic richness	Generic richness: total number of genera	0-14		15-37		38 +	
Filter feeder relative abundance	Filter feeder relative abundance: percent of total abundance that are filter feeders	0-30 %		>30-60%		>60-100%	
P&E relative abundance	P&E relative abundance: percent of total abundance that are Plecoptera or Ephemeroptera	0-5%		>5-20%		>20-100%	
BCG	Biological Condition Gradient: discrete scale of 1 (best) to 6 (worst)	1	2	3	4	5	6

split into, the harder it is to establish dependencies between model variables (Uusitalo 2007). Additionally, based on lessons learned during method development (Chapter 3), we find experts have difficulty distinguishing between the middle two of four categories. Therefore, each continuous variable is split into three bins, with the exception of dominant substrate for 'Habitat' which is binary in nature (either fine or coarse) and BCG whose definition requires six tiers (Table 39).

Bin endpoints for all three-bin variables are then selected. Ideally, each bin should represent a group of values similar to each other and distinct from other bins,

because, after discretization, every value in a bin will be treated in the same way in the model. In practice, this clear distinction between bins is difficult to achieve for continuous variables. Therefore, we adopt an iterative, two-criteria bin selection methodology which starts from equal frequency in each bin based on available data and moves towards ecological significance identification during conditional probability elicitation. In the absence of clear ecological or management dividing lines (such as with substrate and BCG), the 30 northeast EUSE data points for each variable were split such that values less than the 33rd percentile defined the 'low' bin, values between the 33rd and 67th percentiles defined the 'medium' bin, and values greater than the 67th percentile defined the 'high' bin. Given these starting values, experts were asked whether these three categories accurately define what would be considered ecologically 'low', 'medium', or 'high' for each variable and first cut endpoint adjustments were made based on expert feedback. During conditional probability elicitation, experts were able to refine their definitions of 'low', 'medium', or 'high' in the context of describing the behavior of variables in relation to other variables. For example, the endpoint definition for 'low' generic richness was adjusted toward a smaller value during the process of evaluating how low generic richness should affect BCG tier. The experts realized that, as defined, they would expect numbers in the same 'low' category to have different effects on BCG and, therefore, the current definition of 'low' generic richness was too broad and needed to be narrowed.

Just as variable selection and discretization explicitly define model nodes, conditional probability tables (CPTs) are elicited to describe the relationships between the nodes represented by the arrows. Per Bayesian network properties, child nodes (into which arrows point) are conditionally dependent on their parent nodes (from which arrows point). So, the format of CPTs is defined by the model structure, where a CPT must be parameterized for every child node conditional on its parents. CPT elicitation is concerned with obtaining a separate distribution across all possible states of a child node for every possible combination of parent node states. For example, the CPT for the 'Hydrology' node conditional on the 'Urban disturbance' node has three separate distributions across 'high', 'medium', and 'low' flashiness, one for each of the possible states of urban land cover percentage (Table 40).

Table 40: Prior conditional probability table and prior weight (α_0) for "Hydrology" node (flashiness bins in probability units and α_0 in equivalent data points). [Table rows selected for prior weight elicitation highlighted.]

Urban land cover	Flashiness greater than 7 times the median			
	low	medium	high	α_0
low	0.20	0.70	0.10	9.72
medium	0.15	0.55	0.30	9.72
high	0.10	0.40	0.50	9.72

Each of these distributions across child node states is described by its own Dirichlet prior distribution where a θ_i value represents the probability of a sample being in each child node state (where $i = 1$, 2, or 3 for all nodes except 'Habitat' which has two states or 'BCG' which has six states). Marginal nodes with no parents (such as 'Urban

disturbance') are simply assigned an unbiasing flat distribution across all possible discrete states, such that every state has the same probability of occurring. (Dirichlet distributions for the 'Urban disturbance' node are never explicitly reported since urban land cover is simply distributed equally (0.33 per node state), by design, for both the prior and the data and, hence, the posterior.)

CPTs should be elicited following a systematic, prescribed process designed to guide the experts through a rational, thoughtful evaluation of their expertise. There are several documented direct and indirect probability elicitation methods for binomial distributions (Winkler 1967, Morgan and Henrion 1990, Winkler 2003); however, the literature remains sparse on ideal elicitation methods for the multinomial distributions used in this model. Therefore, we developed an adaptation of the direct fixed value elicitation method (Clemen 1991) to apply to multinomial elicitations, following the five steps of the Stanford Research Institute Elicitation Protocol (Spetzler and von Holstein 1975) to attempt to minimize bias and error. To start the elicitation, experts are orientated to the problem context ('motivating') and specific questions of interest ('structuring'), and then asked to summarize and review relevant expert knowledge qualitatively ('conditioning'). After experts are appropriately prepared, prior probabilities are assigned ('encoding') and tested to ensure they correctly capture the expert's opinion ('verifying').

Encoding was conducted in frequency units rather than probability units to facilitate human conceptualization of the information required (Morgan and Henrion 1990). Using amended direct fixed value elicitation, experts were asked to divide a fixed total of events into categories according to their expected likelihood of occurrence. For example, for the first row of the 'Hydrology' child node conditional on the 'Urban disturbance' parent node, the expert is asked: "Assume the watersheds of 100 streams in the northeast United States are known to have 0-7% urban land cover. Assume all other stream features are randomly distributed as if you took a random sample of northeast stream with low urban land cover. How many of those 100 streams would you expect to have low flashiness (no rises greater than seven times the median in a year), medium flashiness (between 1 and 3 rises), or high flashiness (more than 4 rises)?" Experts were encouraged to discuss this question until consensus was reached. This same exercise was then conducted for the remaining possible 'Urban disturbance' parent node state, and for each CPT row of the remaining six child nodes. For the three invertebrate metric nodes and BCG nodes, 1000 total streams were postulated to allow for assignment of very unlikely probabilities less than 1%. Frequency results were converted into probabilities, normalizing where reported total did not equal 100 or 1000 exactly. Complete elicited prior CPTs for all child nodes are reported in Table 40-Table 46.

Table 41: Prior conditional probability table and prior weight (α_0) for "Habitat" node (substrate bins in probability units and α_0 in equivalent data points). [Table rows selected for prior weight elicitation highlighted.]

Urban land cover	Flashiness	Dominant substrate		
		fine	coarse	α_0
low	low	0.15	0.85	7.41
low	medium	0.10	0.90	7.41
low	high	0.05	0.95	7.41
medium	low	0.25	0.75	7.41
medium	medium	0.17	0.83	7.41
medium	high	0.12	0.88	7.41
high	low	0.40	0.60	7.41
high	medium	0.27	0.73	7.41
high	high	0.17	0.83	7.41

Table 42: Prior conditional probability table and prior weight (α_0) for "Water quality" node (conductivity bins in probability units and α_0 in equivalent data points). [Table rows selected for prior weight elicitation highlighted.]

Urban land cover	Flashiness	Conductivity at low base flow			
		low	medium	high	α_0
low	low	0.85	0.10	0.05	14.65
low	medium	0.90	0.06	0.04	14.65
low	high	0.95	0.04	0.01	14.65
medium	low	0.20	0.50	0.30	14.65
medium	medium	0.25	0.50	0.25	14.65
medium	high	0.30	0.50	0.20	14.65
high	low	0.02	0.23	0.75	14.65
high	medium	0.05	0.22	0.73	14.65
high	high	0.10	0.19	0.71	14.65

Table 43: Prior conditional probability table and prior weight (α_0) for "Generic richness" node (richness bins in probability units and α_0 in equivalent data points). [Table rows selected for prior weight elicitation highlighted.]

Flashiness	Substrate	Conductivity	Generic richness			
			low	medium	high	α_0
low	fine	low	0.300	0.600	0.100	1.97
low	fine	medium	0.400	0.500	0.100	1.97
low	fine	high	0.600	0.390	0.010	1.97
low	coarse	low	0.010	0.490	0.500	1.97
low	coarse	medium	0.100	0.350	0.550	1.97
low	coarse	high	0.350	0.550	0.100	1.97
medium	fine	low	0.400	0.400	0.200	1.97
medium	fine	medium	0.500	0.400	0.100	1.97
medium	fine	high	0.700	0.299	0.001	1.97
medium	coarse	low	0.010	0.600	0.390	1.97
medium	coarse	medium	0.100	0.350	0.550	1.97
medium	coarse	high	0.333	0.333	0.333	1.97
high	fine	low	0.500	0.400	0.100	1.97
high	fine	medium	0.600	0.390	0.010	1.97
high	fine	high	0.900	0.099	0.001	1.97
high	coarse	low	0.300	0.600	0.100	1.97
high	coarse	medium	0.450	0.500	0.050	1.97
high	coarse	high	0.590	0.400	0.010	1.97

Table 44: Prior conditional probability table and prior weight (α_0) for "Filter feeder relative abundance" node (relative abundance bins in probability units and α_0 in equivalent data points). [Table rows selected for prior weight elicitation highlighted.]

Flashiness	Substrate	Conductivity	Filter feeder relative abundance			
			low	medium	high	α_0
low	fine	low	0.600	0.390	0.010	1.54
low	fine	medium	0.550	0.440	0.010	1.54
low	fine	high	0.440	0.550	0.010	1.54
low	coarse	low	0.001	0.600	0.399	1.54
low	coarse	medium	0.001	0.299	0.700	1.54
low	coarse	high	0.010	0.290	0.700	1.54
medium	fine	low	0.700	0.299	0.001	1.54
medium	fine	medium	0.700	0.299	0.001	1.54
medium	fine	high	0.500	0.400	0.100	1.54
medium	coarse	low	0.600	0.390	0.010	1.54
medium	coarse	medium	0.250	0.500	0.250	1.54
medium	coarse	high	0.100	0.400	0.500	1.54
high	fine	low	0.800	0.199	0.001	1.54
high	fine	medium	0.800	0.199	0.001	1.54
high	fine	high	0.750	0.200	0.050	1.54
high	coarse	low	0.700	0.290	0.010	1.54
high	coarse	medium	0.300	0.400	0.300	1.54
high	coarse	high	0.200	0.500	0.300	1.54

Table 45: Prior conditional probability table and prior weight (α_0) for "P&E relative abundance" node (relative abundance bins in probability units and α_0 in equivalent data points). [Table rows selected for prior weight elicitation highlighted.]

Flashiness	Substrate	Conductivity	P&E relative abundance			
			low	medium	high	α_0
low	fine	low	0.400	0.550	0.050	5.07
low	fine	medium	0.700	0.290	0.010	5.07
low	fine	high	0.700	0.299	0.001	5.07
low	coarse	low	0.010	0.390	0.600	5.07
low	coarse	medium	0.100	0.600	0.300	5.07
low	coarse	high	0.300	0.600	0.100	5.07
medium	fine	low	0.500	0.400	0.100	5.07
medium	fine	medium	0.550	0.400	0.050	5.07
medium	fine	high	0.699	0.300	0.001	5.07
medium	coarse	low	0.010	0.390	0.600	5.07
medium	coarse	medium	0.010	0.600	0.390	5.07
medium	coarse	high	0.400	0.550	0.050	5.07
high	fine	low	0.490	0.500	0.010	5.07
high	fine	medium	0.700	0.290	0.010	5.07
high	fine	high	0.700	0.299	0.001	5.07
high	coarse	low	0.400	0.400	0.200	5.07
high	coarse	medium	0.500	0.450	0.050	5.07
high	coarse	high	0.600	0.399	0.001	5.07

Table 46: Prior conditional probability table and prior weight (α_0) for "BCG" node (BCG Tier bins in probability units and α_0 in equivalent data points). [Table rows selected for prior weight elicitation highlighted.]

Generic richness	Filter feeder relative abundance	P&E relative abundance	BCG Tier						
			1	2	3	4	5	6	α_0
low	low	low	0.001	0.001	0.001	0.127	0.571	0.299	123.47
low	low	medium	0.001	0.001	0.001	0.127	0.571	0.299	123.47
low	low	high	0.001	0.001	0.001	0.127	0.571	0.299	123.47
low	medium	low	0.001	0.001	0.001	0.225	0.572	0.200	123.47
low	medium	medium	0.001	0.001	0.001	0.299	0.598	0.100	123.47
low	medium	high	0.001	0.001	0.001	0.498	0.498	0.001	123.47
low	high	low	0.001	0.001	0.001	0.001	0.598	0.398	123.47
low	high	medium	0.001	0.001	0.001	0.575	0.384	0.038	123.47
low	high	high	0.001	0.001	0.001	0.498	0.498	0.001	123.47
medium	low	low	0.001	0.001	0.001	0.299	0.598	0.100	123.47
medium	low	medium	0.010	0.196	0.294	0.490	0.010	0.001	123.47
medium	low	high	0.099	0.297	0.396	0.198	0.010	0.001	123.47
medium	medium	low	0.001	0.010	0.296	0.494	0.198	0.001	123.47
medium	medium	medium	0.050	0.300	0.300	0.300	0.050	0.001	123.47
medium	medium	high	0.150	0.399	0.399	0.050	0.001	0.001	123.47
medium	high	low	0.001	0.001	0.050	0.449	0.449	0.050	123.47
medium	high	medium	0.001	0.050	0.449	0.449	0.050	0.001	123.47
medium	high	high	0.001	0.050	0.449	0.449	0.050	0.001	123.47
high	low	low	0.001	0.001	0.001	0.100	0.896	0.001	123.47
high	low	medium	0.105	0.420	0.420	0.053	0.001	0.001	123.47
high	low	high	0.199	0.499	0.299	0.001	0.001	0.001	123.47
high	medium	low	0.001	0.001	0.249	0.499	0.249	0.001	123.47
high	medium	medium	0.001	0.091	0.454	0.408	0.045	0.001	123.47
high	medium	high	0.100	0.200	0.499	0.200	0.001	0.001	123.47
high	high	low	0.001	0.001	0.050	0.449	0.449	0.050	123.47
high	high	medium	0.001	0.001	0.100	0.449	0.449	0.001	123.47
high	high	high	0.050	0.050	0.450	0.350	0.100	0.001	123.47

Finally, prior weights were elicited to know how much to weight prior knowledge relative to data during Bayesian updating. Prior weights are reported in units of equivalent data points to be able to compare with data weight reported in actual

sample size units. After evaluation of three different prior weight elicitation methods, the probability range method (Cowell 1999, Jenson 2009) was found to show lowest within-method variability, most consistent prior weights among nodes, and was most intuitive for experts to report (Chapter 3). This method asks experts to provide a range of values that represents their certainty about a particular parameter. For example, experts reported in CPT elicitation that, of 100 streams with low urban land cover, they would expect 20 to have low flashiness. When asked to provide a certainty range of values, they then said they would not be surprised if there were as little as 5 or as many as 35 streams that had low flashiness in a sample of 100 low urban land cover streams. The premise of the probability range elicitation method is that small reported ranges correspond to high certainty (i.e., high prior weight) and vice versa.

The width of the range has been determined to be approximately two standard deviations long relative to the respective θ_i value (Spiegelhalter et al. 1994). Eliciting a prior probability and an uncertainty range of probabilities for a CPT cell, therefore, provides values for expected mean ($E[\theta_i]$) and expected variance ($\text{var}[\theta_i]$) for the Dirichlet parameter, θ_i. Equations for Dirichlet expected value and Dirichlet expected variance can then be solved for prior weight, α_0, as follows:

$$E[\theta_i] = \frac{\alpha_i}{\alpha_0} \tag{87}$$

$$\text{var}[\theta_i] = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0{}^2(\alpha_0 + 1)} \tag{88}$$

$$\alpha_0 = \frac{E[\theta_i] - E[\theta_i]^2 - \text{var}[\theta_i]}{\text{var}[\theta_i]} \tag{89}$$

In the example reported above, $E[\theta_i] = 0.20$, a range of 0.30 means a standard deviation of 0.15, so $\text{var}[\theta_i] = 0.0225$, resulting in $\alpha_0 = 6.11$ for the low flashiness given low urban land cover CPT cell (Table 47, first row).

Table 47: Probability range prior weight elicitation results for all child state probabilities in each of two randomly selected rows from each conditional probability table.

Child node (variable)	Parent node variable states (bin number)		Child variable state	Prior expected value (μ)	Elicited prior prob. range	Standard deviation from range (σ)	Pior weight (α_0)
Hydrology (Flashiness)	Urban land cover						
	low		low	0.20	0.05-0.35	0.15	6.11
			med	0.70	0.50-0.80	0.15	8.33
			high	0.10	0.02-0.175	0.08	13.98
	high		low	0.10	0.02-0.18	0.08	13.06
			med	0.40	0.30-0.60	0.15	9.67
			high	0.50	0.30-0.65	0.18	7.16
	Mean ± within-method standard deviation						9.72 ± 3.19
Habitat (Substrate)	Urban land cover	Flashiness					
	low	high	fine	0.05	0-0.12	0.06	12.19
			coarse	0.95	0.85-1	0.08	7.44
	high	low	fine	0.40	0.20-0.60	0.20	5.00
			coarse	0.60	0.40-0.80	0.20	5.00
	Mean ± within-method standard deviation						7.41 ± 3.39

Table 47 continued: Probability range prior weight elicitation results for all child state probabilities in each of two randomly selected rows from each conditional probability table.

Child node (variable)	Parent node variable states (bin number)		Child variable state	Prior expected value (μ)	Elicited prior prob. range	Standard deviation from range (σ)	Pior weight (α_0)
Hydrology (Flashiness)	Urban land cover						
	low		low	0.20	0.05-0.35	0.15	6.11
			med	0.70	0.50-0.80	0.15	8.33
			high	0.10	0.02-0.175	0.08	13.98
	high		low	0.10	0.02-0.18	0.08	13.06
			med	0.40	0.30-0.60	0.15	9.67
			high	0.50	0.30-0.65	0.18	7.16
	Mean ± within-method standard deviation						9.72 ± 3.19
Habitat (Substrate)	Urban land cover	Flashiness					
	low	high	fine	0.05	0-0.12	0.06	12.19
			coarse	0.95	0.85-1	0.08	7.44
	high	low	fine	0.40	0.20-0.60	0.20	5.00
			coarse	0.60	0.40-0.80	0.20	5.00
	Mean ± within-method standard deviation						7.41 ± 3.39
Water quality (Conduc-tivity)	Urban land cover	Flashiness					
	low	low	low	0.85	0.75-0.95	0.10	11.75
			med	0.10	0.02-0.20	0.09	10.11
			high	0.05	0-0.10	0.05	18.00
	high	low	low	0.02	0-0.05	0.03	30.36
			med	0.23	0.05-0.30	0.13	10.33
			high	0.75	0.65-0.95	0.15	7.33
	Mean ± within-method standard deviation						14.65 ± 8.48

Table 47 continued: Probability range prior weight elicitation results for all child state probabilities in each of two randomly selected rows from each conditional probability table.

Child node (variable)	Parent node variable states (bin number)			Child variable state	Prior expected value (μ)	Elicited prior prob. range	Standard deviation from range (σ)	Pior weight (α_0)
Generic richness	Flash.	Sub.	Cond.					
	med	coarse	med	low	0.100	0.010-0.200	0.095	8.97
				med	0.350	0.100-0.900	0.400	0.42
				high	0.550	0.300-0.800	0.250	2.96
	high	fine	high	low	0.900	0.500-0.999	0.250	0.45
				med	0.099	0.050-0.800	0.375	-0.37
				high	0.001	0-0.100	0.050	-0.60
	Mean ± within-method standard deviation							1.97 ± 3.65
Filter feeder relative abundance	Flash	Sub.	Cond.					
	low	fine	high	low	0.440	0.200-0.800	0.300	1.74
				med	0.550	0.200-0.800	0.300	1.75
				high	0.010	0.001-0.150	0.075	0.78
	high	coarse	high	low	0.200	0.100-0.600	0.250	1.56
				med	0.500	0.100-0.800	0.350	1.04
				high	0.300	0.100-0.600	0.250	2.36
	Mean ± within-method standard deviation							1.54 ± 0.56
P&E relative abundance	Flash	Sub.	Cond.					
	low	coarse	med	low	0.100	0.050-0.300	0.125	4.76
				med	0.600	0.450-0.900	0.225	3.74
				high	0.300	0.200-0.600	0.200	4.25
	med	coarse	low	low	0.010	0.001-0.100	0.050	3.04
				med	0.390	0.200-0.600	0.200	4.95
				high	0.600	0.500-0.800	0.150	9.67
	Mean ± within-method standard deviation							5.07 ± 2.36

Table 47 continued: Probability range prior weight elicitation results for all child state probabilities in each of two randomly selected rows from each conditional probability table.

Child node (variable)	Parent node variable states (bin number)			Child variable state	Prior expected value (μ)	Elicited prior prob. range	Standard deviation from range (σ)	Pior weight (α_0)
BCG	Gen. rich.	Filter feeder RA	P&E RA					
	low	high	high	1	0.001	0-0.005	0.002	159.48
				2	0.001	0-0.005	0.002	159.48
				3	0.001	0-0.020	0.010	9.03
				4	0.498	0.299-0.697	0.199	5.30
				5	0.498	0.299-0.697	0.199	5.30
				6	0.001	0-0.100	0.050	-0.60
	high	high	high	1	0.050	0.030-0.1510	0.004	12.21
				2	0.050	0.040-0.100	0.001	51.83
				3	0.450	0.350-0.549	0.010	23.80
				4	0.350	0.250-0.450	0.010	21.79
				5	0.100	0.050-0.150	0.002	35.04
				6	0.001	0-0.002	0.000001	999.00
	Mean ± within-method standard deviation							123.47 ± 281.42

Probability ranges were elicited and prior weights were calculated in this way for two random rows of each child node's CPT (noted in yellow in Table 40-Table 46). These cell-level prior weight values were then averaged across a child node for stability (Table 47). Node-averaged prior weights are also reported with elicited CPTs in Table 40-Table 46.

4.4.4. BCG Tier assignment

The assignment of BCG Tier to a sampled site is fundamentally a question of regional expert interpretation. First, experts must decide which measured biological community metrics are indicators of which of the 10 BCG Attribute classes. Then, given a site's specific combination of values for each of these indicators, experts must assess overall biological condition in terms of BGC Tier.

In the northeast U.S., significant effort has been exerted to derive expert consensus on the matching of invertebrate taxa to the first six BCG Attributes relating to community structure (Table 37). Based on consultation with extensive previous classification work conducted in the northeast United States by EPA (Barbour and Yoder 2007, Jackson and Holdsworth 2007), its contractor Tetra Tech (Gerritsen 2008), and the State of Maine (Shelton and Blocksom 2004), we classified some 700 taxa sampled in the EUSE NECB data into appropriate Attribute groups. These taxa groupings describe sensitivity and tolerance of macroinvertebrates specific to the northeast and would need to be re-classified in order to apply to different ecoregions within the United States and beyond. The last four BCG Attributes describing ecological function are much more challenging to measure than those involving community structure and have not yet been well-characterized in classification exercises.

Once EUSE northeast taxa were classified, biologist experts were given 30 spreadsheets (one for each NECB watershed sampled) characterizing each site's

measured invertebrate community structure via invertebrate taxa lists and summaries present in each sample, color-coded by Attribute. In addition to Attribute-sorted metrics, spreadsheets also contained 20 additional standard invertebrate metrics describing richness and abundance of common indicator groups of taxa (EPT, Hydropsyche, Perlidae, etc.). Using their long-term career experience in judging biological condition, experts used this provided information to assign one BCG Tier (1-6) to each NECB site. Tiers were first assigned individually and final assignments were reached through expert group discussion and consensus. These assignments were then considered part of the EUSE dataset used for Bayesian updating to calculate the posterior.

4.5. Results

4.5.1. Prior Bayesian network model

The final elicited Northeast BCG model structure has 8 nodes, 17 arrows (Figure 75) and a total of 381 Dirichlet theta parameters (sum of the product of each child node's parent state combinations and child state combinations). To ensure model parsimony and scale match, only major system components were included.

The final prior Bayesian network combines elicited model structure, variables, CPTs, and prior weights into one representation of node relationships and uncertainties (Figure 76). The model representation shows marginal discrete distributions for each of the nine nodes and arrows causally connecting the nodes. Conditional probability and

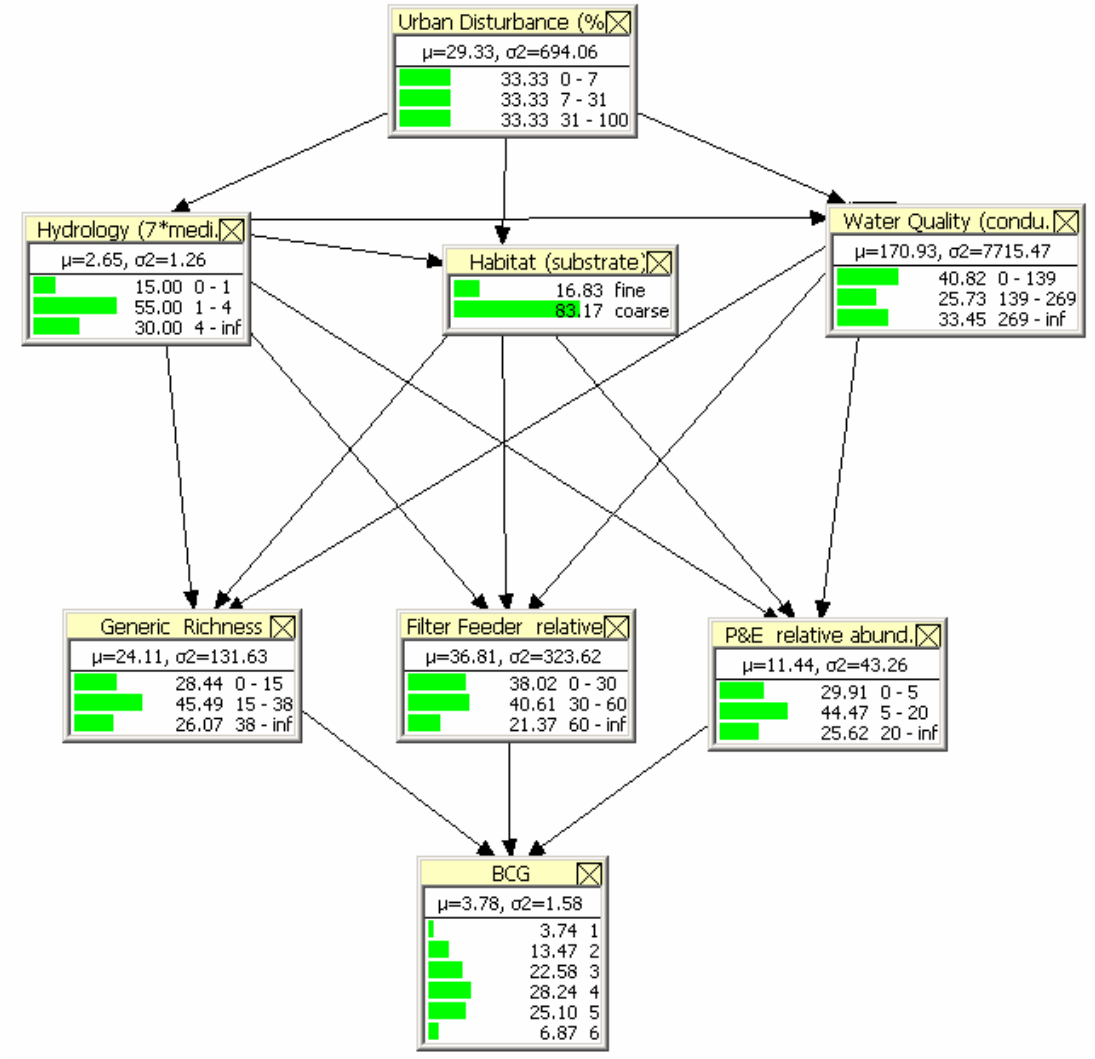

Figure 76: Prior Northeast Bayesian network.

prior weight information is embedded in the arrows pointing from parent to child nodes. For each node, the first column of horizontal bars visually shows the probability of a Northeast watershed sample falling in each node state where longer bars correspond to higher probabilities. The second column provides the numeric probability value. The third column defines the bin endpoints for each state. Mean (μ) and variance (σ^2) values of each node's discrete marginal distribution are also provided

in this model (Hugin 2008) representation. This prior network can provide information such as the probability of observing a 'minimally changed' BCG Tier 2 in a random northeast U.S. watershed is 13.47% (Figure 76).

More usefully, hypothetical 'evidence' can be entered into the network and propagated to predict the effects of changing one variable on the rest of the system. For example, under conditions of low urban land cover, the predictive probability of seeing BCG Tier 2 increases to 22.78% (Figure 77). Additionally, when urban land cover is low, the prior predictive probabilities show that (comparing Figure 76 to Figure 77), in general, hydrology becomes less flashy, habitat becomes more coarse, water quality becomes less conductive, invertebrate generic richness becomes greater, filter feeder relative abundance decreases, and P and E relative abundance increases. These trends are reversed under conditions of high urban land cover, with probability of seeing BCG Tier 2 decreasing to 5.69% (Figure 78). But not only does the Bayesian network illustrate the direction and magnitude of expected change for every node, it does so in the form of changing (discrete) distributions which incorporate system uncertainties about these likely changes.

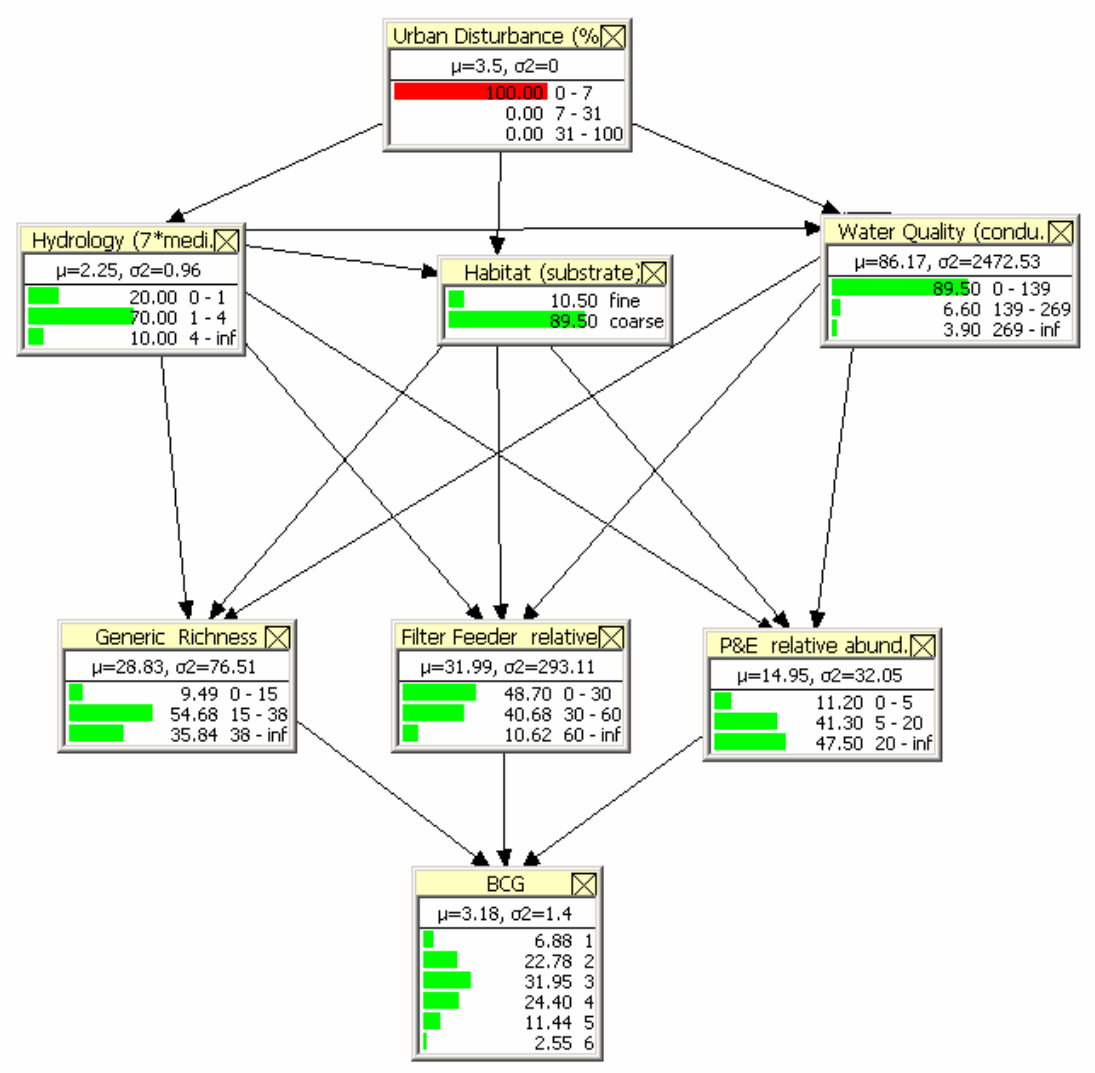

Figure 77: Prior predictive probabilities for low urban disturbance.

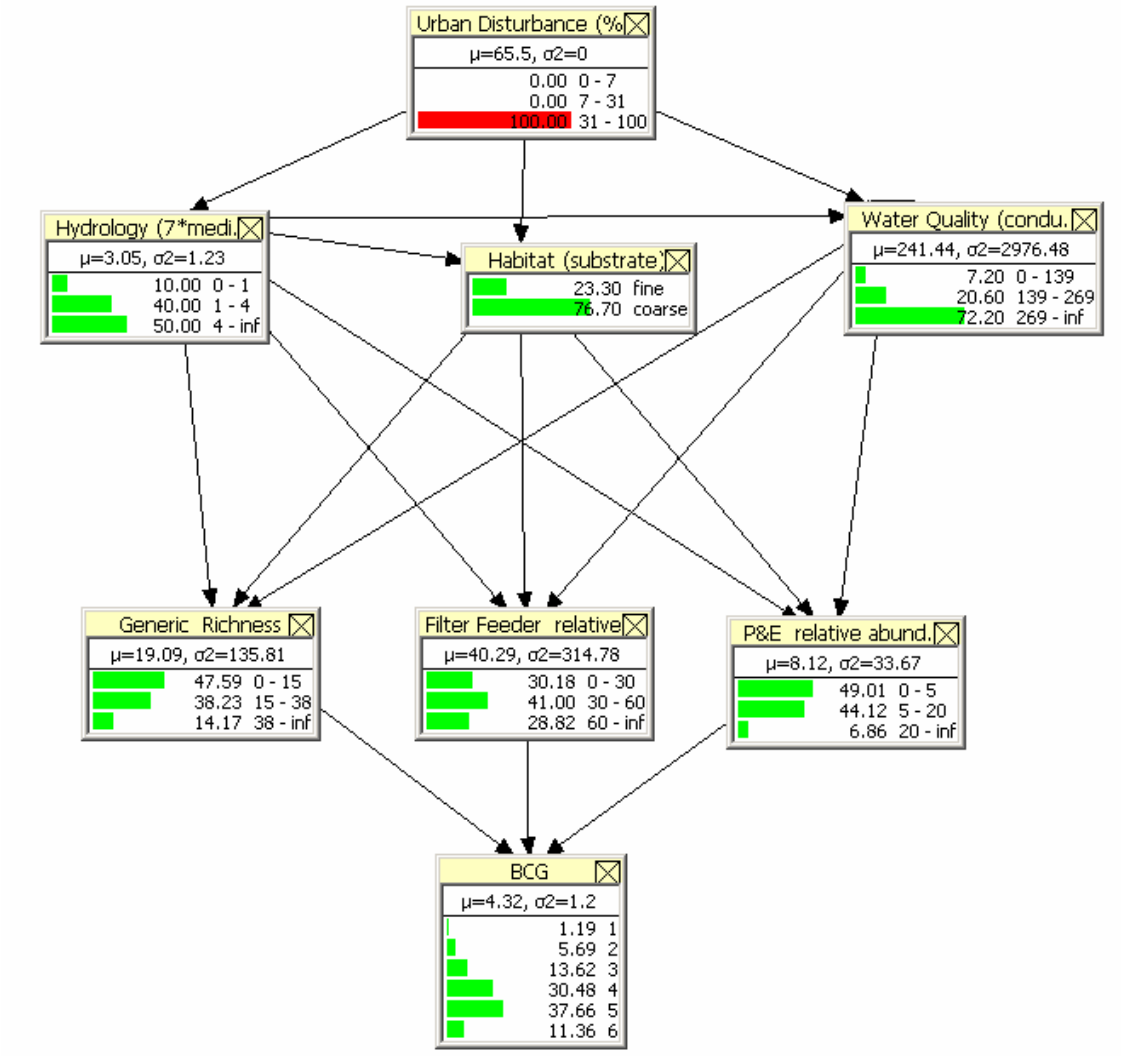

Figure 78: Prior predictive probabilities for high urban disturbance.

4.5.2. Data-only Bayesian network model

Given the elicited model structure, the Bayesian network can be parameterized using only the data sampled from the 30 EUSE northeast watersheds (Table 48-Table 54).

Table 48: Data table for "Hydrology" node (flashiness counts and total number of samples per parent state combination).

Urban land cover	Flashiness greater than 7 times the median			
	low	medium	high	*n*
low	1	8	1	10
medium	1	5	4	10
high	2	4	4	10

Table 49: Data table for "Habitat" node (substrate counts and total number of samples per parent state combination).

Urban land cover	Flashiness	Dominant substrate		
		fine	coarse	*n*
low	low	0	1	1
low	medium	0	8	8
low	high	0	1	1
medium	low	0	1	1
medium	medium	0	5	5
medium	high	0	4	4
high	low	0	2	2
high	medium	0	4	4
high	high	0	4	4

Table 50: Data table for "Water quality" node (conductivity counts and total number of samples per parent state combination).

Urban land cover	Flashiness	Conductivity at low base flow			
		low	medium	high	*n*
low	low	0	1	0	1
low	medium	7	1	0	8
low	high	1	0	0	1
medium	low	0	1	0	1
medium	medium	1	3	1	5
medium	high	1	1	2	4
high	low	0	0	2	2
high	medium	0	2	2	4
high	high	1	1	3	4

Table 51: Data table for "Generic richness" node (richness counts and total number of samples per parent state combination).

Flashiness	Substrate	Conductivity	Generic richness			
			low	medium	high	*n*
low	fine	low	0	0	0	0
low	fine	medium	0	0	0	0
low	fine	high	0	0	0	0
low	coarse	low	0	0	0	0
low	coarse	medium	0	1	1	2
low	coarse	high	0	2	0	2
medium	fine	low	0	0	0	0
medium	fine	medium	0	0	0	0
medium	fine	high	0	0	0	0
medium	coarse	low	0	1	7	8
medium	coarse	medium	0	6	0	6
medium	coarse	high	0	3	0	3
high	fine	low	0	0	0	0
high	fine	medium	0	0	0	0
high	fine	high	0	0	0	0
high	coarse	low	0	0	2	2
high	coarse	medium	0	2	0	2
high	coarse	high	0	5	0	5

Table 52: Data table for "Filter feeder relative abundance" node (relative abundance counts and total number of samples per parent state combination).

Flashiness	Substrate	Conductivity	Filter feeder relative abundance			
			low	medium	high	*n*
low	fine	low	0	0	0	0
low	fine	medium	0	0	0	0
low	fine	high	0	0	0	0
low	coarse	low	0	0	0	0
low	coarse	medium	0	1	1	2
low	coarse	high	0	0	2	2
medium	fine	low	0	0	0	0
medium	fine	medium	0	0	0	0
medium	fine	high	0	0	0	0
medium	coarse	low	4	4	0	8
medium	coarse	medium	0	0	6	6
medium	coarse	high	0	0	3	3
high	fine	low	0	0	0	0
high	fine	medium	0	0	0	0
high	fine	high	0	0	0	0
high	coarse	low	1	1	0	2
high	coarse	medium	0	0	2	2
high	coarse	high	0	1	4	5

Table 53: Data table for "P&E relative abundance" node (relative abundance counts and total number of samples per parent state combination).

Flashiness	Substrate	Conductivity	P&E relative abundance			
			low	medium	high	*n*
low	fine	low	0	0	0	0
low	fine	medium	0	0	0	0
low	fine	high	0	0	0	0
low	coarse	low	0	0	0	0
low	coarse	medium	0	2	0	2
low	coarse	high	1	1	0	2
medium	fine	low	0	0	0	0
medium	fine	medium	0	0	0	0
medium	fine	high	0	0	0	0
medium	coarse	low	0	5	3	8
medium	coarse	medium	3	3	0	6
medium	coarse	high	1	2	0	3
high	fine	low	0	0	0	0
high	fine	medium	0	0	0	0
high	fine	high	0	0	0	0
high	coarse	low	0	1	1	2
high	coarse	medium	2	0	0	2
high	coarse	high	4	1	0	5

Table 54: Data table for "BCG" node (BCG Tier counts and total number of samples per parent state combination).

Generic richness	Filter feeder relative abundance	P&E relative abundance	BCG Tier						
			1	**2**	**3**	**4**	**5**	**6**	***n***
low	low	low	0	0	0	0	0	0	0
low	low	medium	0	0	0	0	0	0	0
low	low	high	0	0	0	0	0	0	0
low	medium	low	0	0	0	0	0	0	0
low	medium	medium	0	0	0	0	0	0	0
low	medium	high	0	0	0	0	0	0	0
low	high	low	0	0	0	0	0	0	0
low	high	medium	0	0	0	0	0	0	0
low	high	high	0	0	0	0	0	0	0
medium	low	low	0	0	0	0	0	0	0
medium	low	medium	0	0	0	0	0	0	0
medium	low	high	0	0	0	0	0	0	0
medium	medium	low	0	0	0	0	0	1	1
medium	medium	medium	0	0	1	0	0	0	1
medium	medium	high	0	0	0	0	0	0	0
medium	high	low	0	0	0	2	6	2	10
medium	high	medium	0	0	2	6	0	0	8
medium	high	high	0	0	0	0	0	0	0
high	low	low	0	0	0	0	0	0	0
high	low	medium	0	1	0	0	0	0	1
high	low	high	0	4	0	0	0	0	4
high	medium	low	0	0	0	0	0	0	0
high	medium	medium	0	3	2	0	0	0	5
high	medium	high	0	0	0	0	0	0	0
high	high	low	0	0	0	0	0	0	0
high	high	medium	0	0	0	0	0	0	0
high	high	high	0	0	0	0	0	0	0

However, because of the small sample size relative to the large number of model parameters, many of the variable state combinations the model attempts to characterize

are not represented in the dataset because they were not measured. In this type of situation, there would be very little confidence in the model if significant expert information were not included in the prior. For comparison sake, an uninformed prior Bayesian network was updated with the data to evaluate the resulting data-only model.

One of the major model components missing from the dataset is fine sediment dominated watersheds. Northeast U.S. watersheds tend to be much more geomorphologically coarse than watersheds in other parts of the country, such as in agricultural runoff dominated regions of the Midwest. Because high levels of fine sediments are rare in the northeast, no such watershed was observed during EUSE data collection. This means there is no information in the data available to characterize any of the CPT probability rows conditional on fine sediment dominated habitat (a total of 81 Dirichlet parameters, Table 51, Table 52, Table 53) nor any of the CPT for the habitat child node (18 Dirichlet parameters, Table 49), despite the fact that expert biologists emphasized the importance of fine sediment as an adverse impactor of invertebrate health. Because of the lack of data, the Habitat node in the data-only model never changes state from 100% 'coarse'.

Similarly, no watersheds with 'low' generic richness (less than 15 genera) were sampled. Since low generic richness is associated with highly urbanized watersheds, this sampling gap may be the result of the NECB EUSE study including no sites with urban land cover greater than 77% of the watershed. In the northeast, many of the most

urbanized streams have been forced into concrete culverts or underground pipes. These streams were consciously not sampled as part of EUSE because they were not considered natural stream systems anymore, which may have resulted in under-sampling of highly urbanized streams. Because of this data gap, any CPT row involving 'low' generic richness as a child or parent state cannot be characterized by the data (27 Dirichlet parameters, Table 51, Table 54).

In addition to these two systematic data gaps, many of the remaining parent state combinations have no data equivalent, with random data gaps increasing as node complexity increases. Particularly, for the BCG node, which has three parents with three states each, only 20 of these 27 parent state combinations have data available (Table 54). This is because some parent state combinations are more likely to occur than others (e.g., 10 of 30 NECB watersheds had parent state combination medium generic richness, high filter feeder relative abundance, and low P&E abundance), and small sample sizes have trouble capturing rare state combinations.

Paradoxically, the effects of parameterizing a Bayesian network with only limited data are extreme marginal distributions with low apparent uncertainty (Habitat and Generic richness nodes, Figure 79), and extreme response to evidence propagation. For example, the data-only model predictive probability of seeing BCG Tier 2 is much greater than the prior model prediction in conditions of low urban land cover (53.86%,

Figure 80), and much less than the prior model prediction in conditions of high urban land cover (0%, Figure 81).

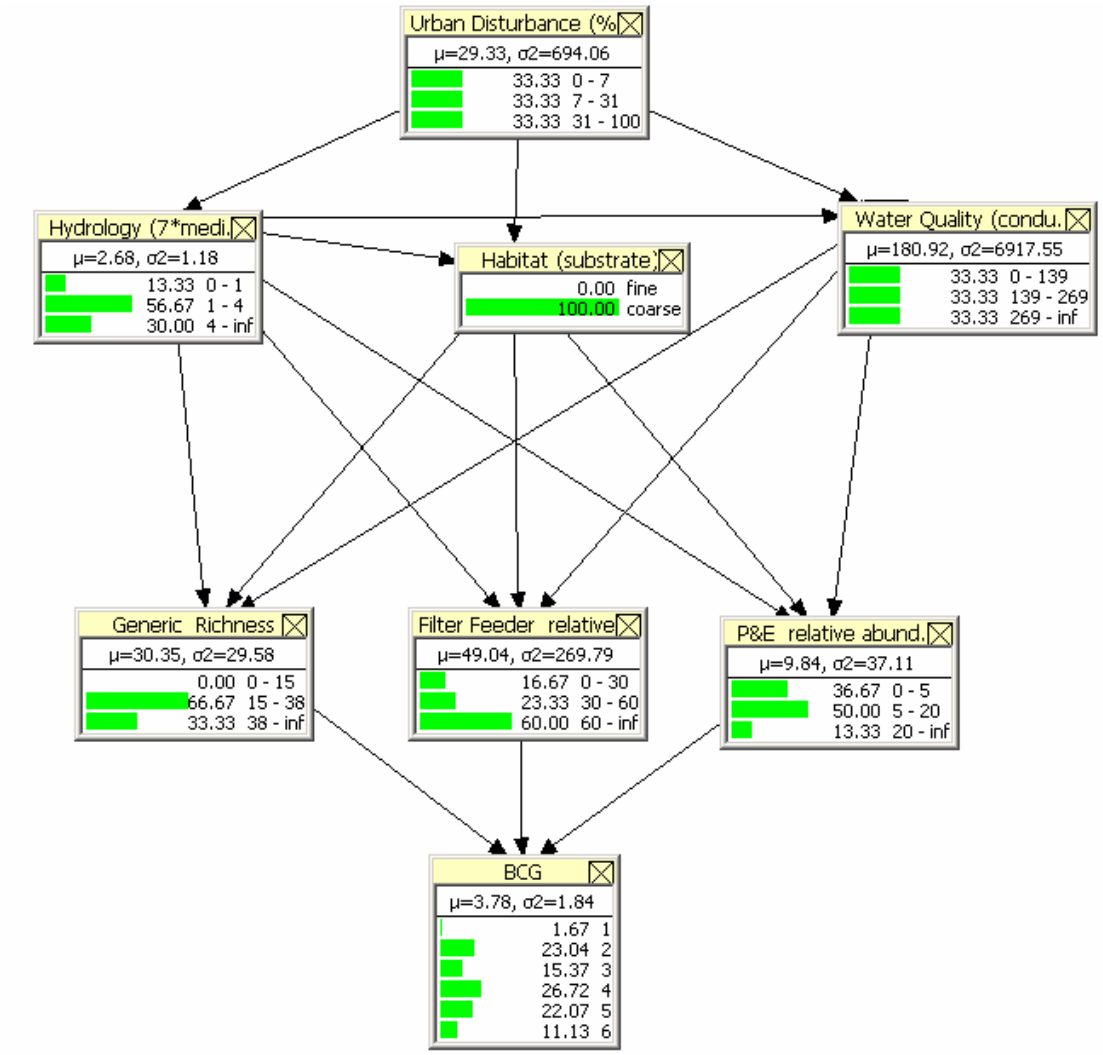

Figure 79: Data-only Northeast Bayesian network (empty prior updated with data).

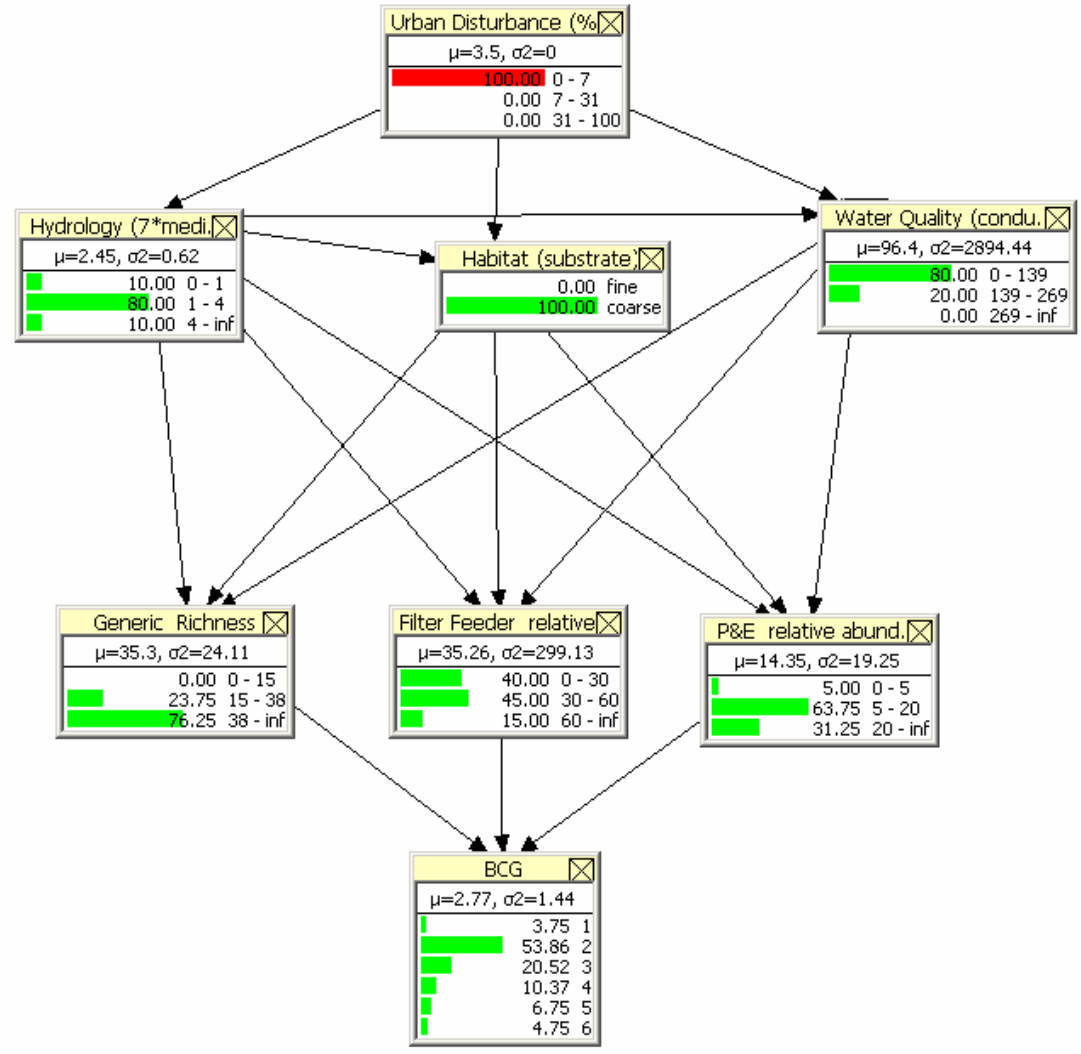

Figure 80: Data-only predictive probabilities for low urban disturbance.

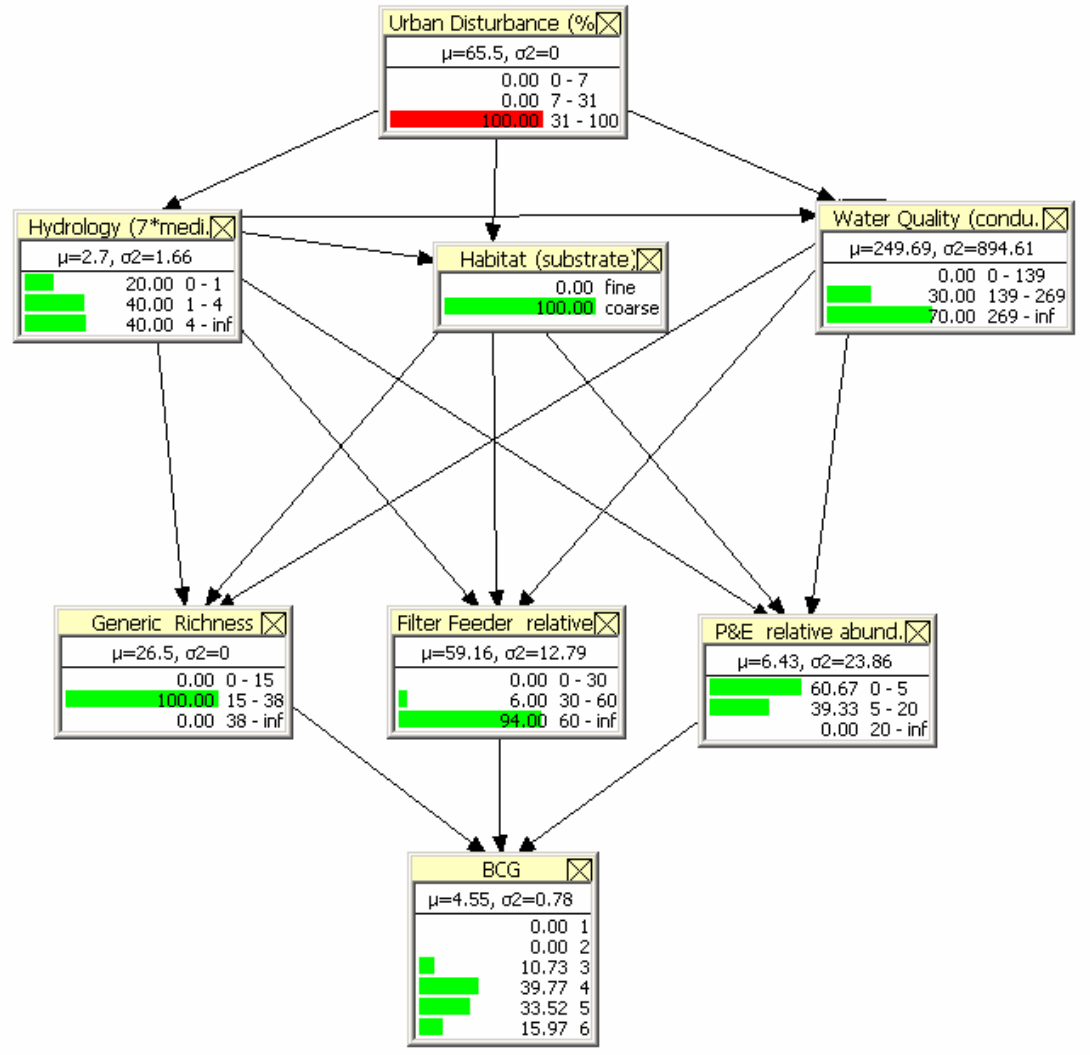

Figure 81: Data-only predictive probabilities for high urban disturbance.

This result can be traced to the abundance of data gap induced CPT zeros in the data model, which are probabilistically interpreted as total certainty of non-occurrence. This false certainty unrealistically distorts prediction magnitude. Although the direction of change is the same in both models, showing that experts and data agree on variable relationships, clearly data alone is insufficient to adequately parameterize this model.

Combining expert knowledge with data will more accurately characterize system functionality.

4.5.3. Posterior Bayesian network model

Using Dirichlet-multinomial updating (as described in the *Conjugate Dirichlet-multinomial updating* section above), the expert elicited prior model was updated with the EUSE NECB data to create the posterior model (Figure 82). Posterior conditional probability tables, which are a weighted average of prior information and data, are calculated for every child node in the model (Table 55- Table 61).

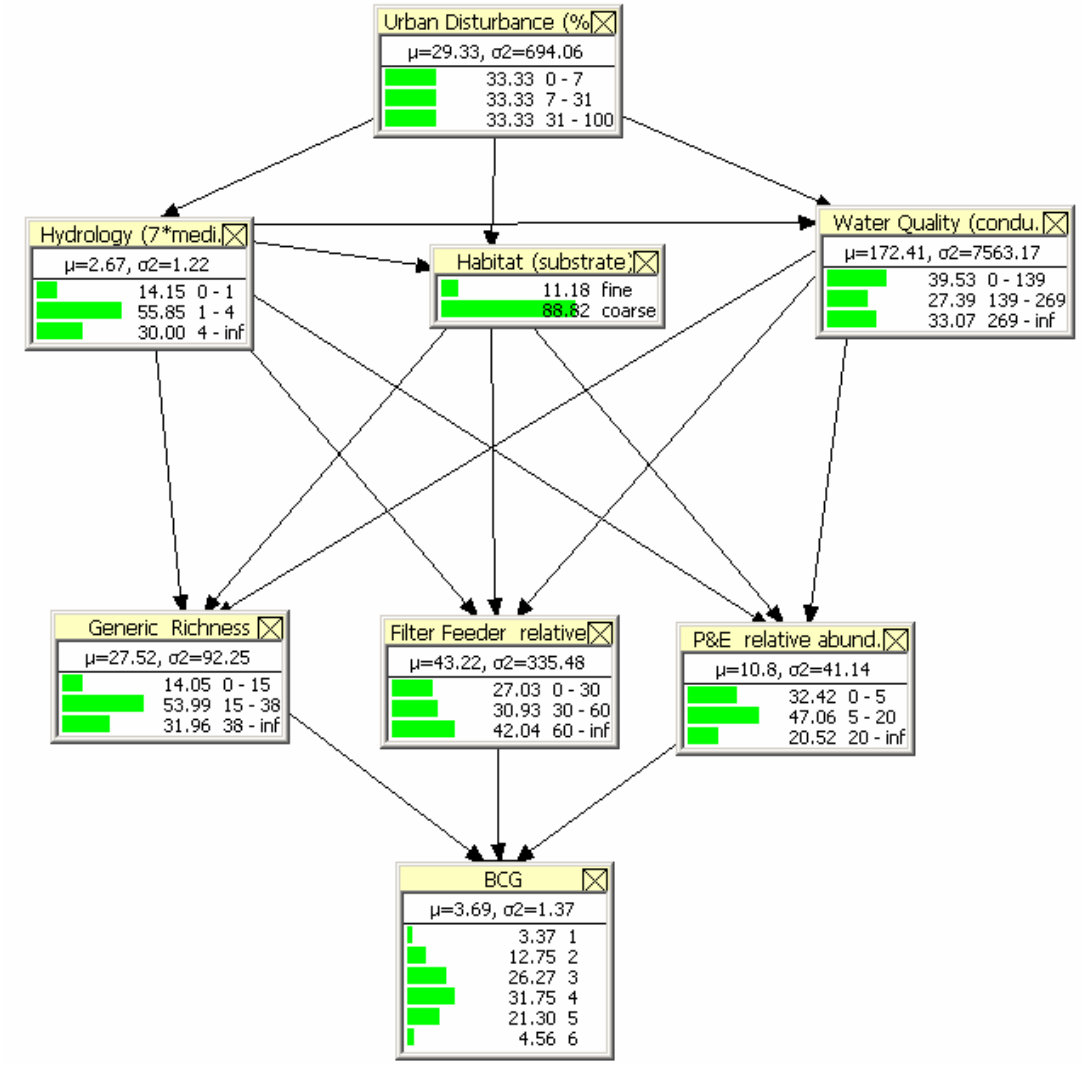

Figure 82: Posterior Northeast Bayesian network.

Table 55: Posterior conditional probability table and posterior weights (α_0+n) for "Hydrology" node (flashiness bins in probability units and α_0+n in equivalent data points).

Urban land cover	Flashiness greater than 7 times the median			
	low	medium	high	α_0+n
low	0.149	0.751	0.100	19.72
medium	0.125	0.525	0.351	19.72
high	0.151	0.400	0.449	19.72

Table 56: Posterior conditional probability table and posterior weights (α_0+n) for "Habitat" node (substrate bins in probability units and α_0+n in equivalent data points).

Urban land cover	Flashiness	Dominant substrate		
		fine	coarse	α_0+n
low	low	0.132	0.868	8.41
low	medium	0.048	0.952	15.41
low	high	0.044	0.956	8.41
medium	low	0.220	0.780	8.41
medium	medium	0.102	0.898	12.41
medium	high	0.078	0.922	11.41
high	low	0.315	0.685	9.41
high	medium	0.175	0.825	11.41
high	high	0.110	0.890	11.41

Table 57: Posterior conditional probability table and posterior weights (α_0+n) for "Water quality" node (conductivity bins in probability units and α_0+n in equivalent data points).

Urban land cover	Flashiness	Conductivity at low base flow			
		low	medium	high	α_0+n
low	low	0.796	0.158	0.047	15.65
low	medium	0.891	0.083	0.026	22.65
low	high	0.953	0.037	0.009	15.65
medium	low	0.187	0.532	0.281	15.65
medium	medium	0.237	0.525	0.237	19.65
medium	high	0.289	0.446	0.264	18.65
high	low	0.018	0.202	0.780	16.65
high	medium	0.039	0.280	0.681	18.65
high	high	0.079	0.203	0.719	18.65

Table 58: Posterior conditional probability table and posterior weights (α_0+n) for "Generic richness" node (richness bins in probability units and α_0+n in equivalent data points).

Flashiness	Substrate	Conductivity	Generic richness			
			low	medium	high	α_0+n
low	fine	low	0.300	0.600	0.100	1.97
low	fine	medium	0.400	0.500	0.100	1.97
low	fine	high	0.600	0.390	0.010	1.97
low	coarse	low	0.010	0.490	0.500	1.97
low	coarse	medium	0.050	0.426	0.525	3.97
low	coarse	high	0.174	0.777	0.050	3.97
medium	fine	low	0.400	0.400	0.200	1.97
medium	fine	medium	0.500	0.400	0.100	1.97
medium	fine	high	0.700	0.299	0.001	1.97
medium	coarse	low	0.002	0.219	0.779	9.97
medium	coarse	medium	0.025	0.839	0.136	7.97
medium	coarse	high	0.132	0.736	0.132	4.97
high	fine	low	0.500	0.400	0.100	1.97
high	fine	medium	0.600	0.390	0.010	1.97
high	fine	high	0.900	0.099	0.001	1.97
high	coarse	low	0.149	0.298	0.553	3.97
high	coarse	medium	0.223	0.752	0.025	3.97
high	coarse	high	0.167	0.830	0.003	6.97

Table 59: Posterior conditional probability table and posterior weights (α_0+n) for "Filter feeder relative abundance" node (relative abundance bins in probability units and α_0+n in equivalent data points).

Flashiness	Substrate	Conductivity	Filter feeder relative abundance			
			low	medium	high	α_0+n
low	fine	low	0.600	0.390	0.010	1.54
low	fine	medium	0.550	0.440	0.010	1.54
low	fine	high	0.440	0.550	0.010	1.54
low	coarse	low	0.001	0.600	0.399	1.54
low	coarse	medium	0.000	0.413	0.587	3.54
low	coarse	high	0.004	0.126	0.869	3.54
medium	fine	low	0.700	0.299	0.001	1.54
medium	fine	medium	0.700	0.299	0.001	1.54
medium	fine	high	0.500	0.400	0.100	1.54
medium	coarse	low	0.516	0.482	0.002	9.54
medium	coarse	medium	0.051	0.102	0.847	7.54
medium	coarse	high	0.034	0.136	0.830	4.54
high	fine	low	0.800	0.199	0.001	1.54
high	fine	medium	0.800	0.199	0.001	1.54
high	fine	high	0.750	0.200	0.050	1.54
high	coarse	low	0.587	0.409	0.004	3.54
high	coarse	medium	0.131	0.174	0.695	3.54
high	coarse	high	0.047	0.271	0.682	6.54

Table 60: Posterior conditional probability table and posterior weights (α0+n) for "P&E relative abundance" node (relative abundance bins in probability units and α0+n in equivalent data points).

Flashiness	Substrate	Conductivity	P&E relative abundance			
			low	medium	high	α_0+n
low	fine	low	0.400	0.550	0.050	5.07
low	fine	medium	0.700	0.290	0.010	5.07
low	fine	high	0.700	0.299	0.001	5.07
low	coarse	low	0.010	0.390	0.600	5.07
low	coarse	medium	0.072	0.713	0.215	7.07
low	coarse	high	0.357	0.572	0.072	7.07
medium	fine	low	0.500	0.400	0.100	5.07
medium	fine	medium	0.550	0.400	0.050	5.07
medium	fine	high	0.699	0.300	0.001	5.07
medium	coarse	low	0.004	0.534	0.462	13.07
medium	coarse	medium	0.276	0.546	0.179	11.07
medium	coarse	high	0.375	0.593	0.031	8.07
high	fine	low	0.490	0.500	0.010	5.07
high	fine	medium	0.700	0.290	0.010	5.07
high	fine	high	0.700	0.299	0.001	5.07
high	coarse	low	0.287	0.428	0.285	7.07
high	coarse	medium	0.641	0.323	0.036	7.07
high	coarse	high	0.699	0.300	0.001	10.07

Table 61: Posterior conditional probability table and posterior weights (α_0+n) for "BCG" node (BCG Tier bins in probability units and α_0+n in equivalent data points).

Generic richness	Filter feeder relative abundance	P&E relative abundance	BCG Tier						
			1	2	3	4	5	6	α_0+n
low	low	low	0.001	0.001	0.001	0.127	0.571	0.299	123.47
low	low	medium	0.001	0.001	0.001	0.127	0.571	0.299	123.47
low	low	high	0.001	0.001	0.001	0.127	0.571	0.299	123.47
low	medium	low	0.001	0.001	0.001	0.225	0.572	0.200	123.47
low	medium	medium	0.001	0.001	0.001	0.299	0.598	0.100	123.47
low	medium	high	0.001	0.001	0.001	0.498	0.498	0.001	123.47
low	high	low	0.001	0.001	0.001	0.001	0.598	0.398	123.47
low	high	medium	0.001	0.001	0.001	0.575	0.384	0.038	123.47
low	high	high	0.001	0.001	0.001	0.498	0.498	0.001	123.47
medium	low	low	0.001	0.001	0.001	0.299	0.598	0.100	123.47
medium	low	medium	0.010	0.196	0.294	0.490	0.010	0.001	123.47
medium	low	high	0.099	0.297	0.396	0.198	0.010	0.001	123.47
medium	medium	low	0.001	0.010	0.294	0.490	0.196	0.009	124.47
medium	medium	medium	0.050	0.297	0.305	0.297	0.050	0.001	124.47
medium	medium	high	0.150	0.399	0.399	0.050	0.001	0.001	123.47
medium	high	low	0.001	0.001	0.046	0.430	0.460	0.061	133.47
medium	high	medium	0.001	0.047	0.437	0.467	0.047	0.001	131.47
medium	high	high	0.001	0.050	0.449	0.449	0.050	0.001	123.47
high	low	low	0.001	0.001	0.001	0.100	0.896	0.001	123.47
high	low	medium	0.104	0.425	0.417	0.052	0.001	0.001	124.47
high	low	high	0.193	0.514	0.290	0.001	0.001	0.001	127.47
high	medium	low	0.001	0.001	0.249	0.499	0.249	0.001	123.47
high	medium	medium	0.001	0.111	0.452	0.392	0.044	0.001	128.47
high	medium	high	0.100	0.200	0.499	0.200	0.001	0.001	123.47
high	high	low	0.001	0.001	0.050	0.449	0.449	0.050	123.47
high	high	medium	0.001	0.001	0.100	0.449	0.449	0.001	123.47
high	high	high	0.050	0.050	0.450	0.350	0.100	0.001	123.47

To evaluate system relationships described by the posterior model, it is easier to interactively manipulate the Bayesian network than to interpret the static, multidimensional series of CPTs. By entering hypothetical evidence on each node in turn, system trends can be understood. For example, as conductivity increases, generic richness decreases, filter feeder relative abundance increases, and P&E relative abundance decreases. This means as water quality becomes impaired, invertebrate communities become less diverse, more tolerant and less sensitive.

The posterior Bayesian network also quantifies the relative magnitudes of system changes. For example, flashiness seems to affect filter feeder relative abundance more than generic richness. This type of relative influence question can be a very useful one from a management perspective. A manager can use a Bayesian network to assess the importance of different model drivers on a desired endpoint, such as BCG, to determine which management actions would lead to the greatest positive impact. In contrast to traditional sensitivity analysis contexts, a Byesian network is able to account for the response of the entire system, acknowledging that variables are related, when evaluating the effect of each driver. This is more realistic than holding all variables but the driver of interest constant since environmental input drivers act in concert to produce a system-wide response.

Additionally, Bayesian network sensitivity analysis incorporates all the uncertainties which are encoded by the network when predicting driver impacts on the endpoint, and

results are reported probabilistically for clearer characterization. Driver assessment sensitivity analysis was performed on the completed posterior Northeast Bayesian network by varying each driver node one state from its 'worst' state associated with negatively impacted environmental conditions to the next 'better' state associated with closer to natural conditions. The likelihood of improving BCG Tier was evaluated for these changes in each driver. For example, under high flashiness, the likelihood of achieving BCG Tier 3 or better is 18.21% (Figure 83). If flashiness were managed to medium level, the likelihood of achieving BCG Tier 3 or better increases to 51.51% (Figure 84). This represents a 33.30% improvement.

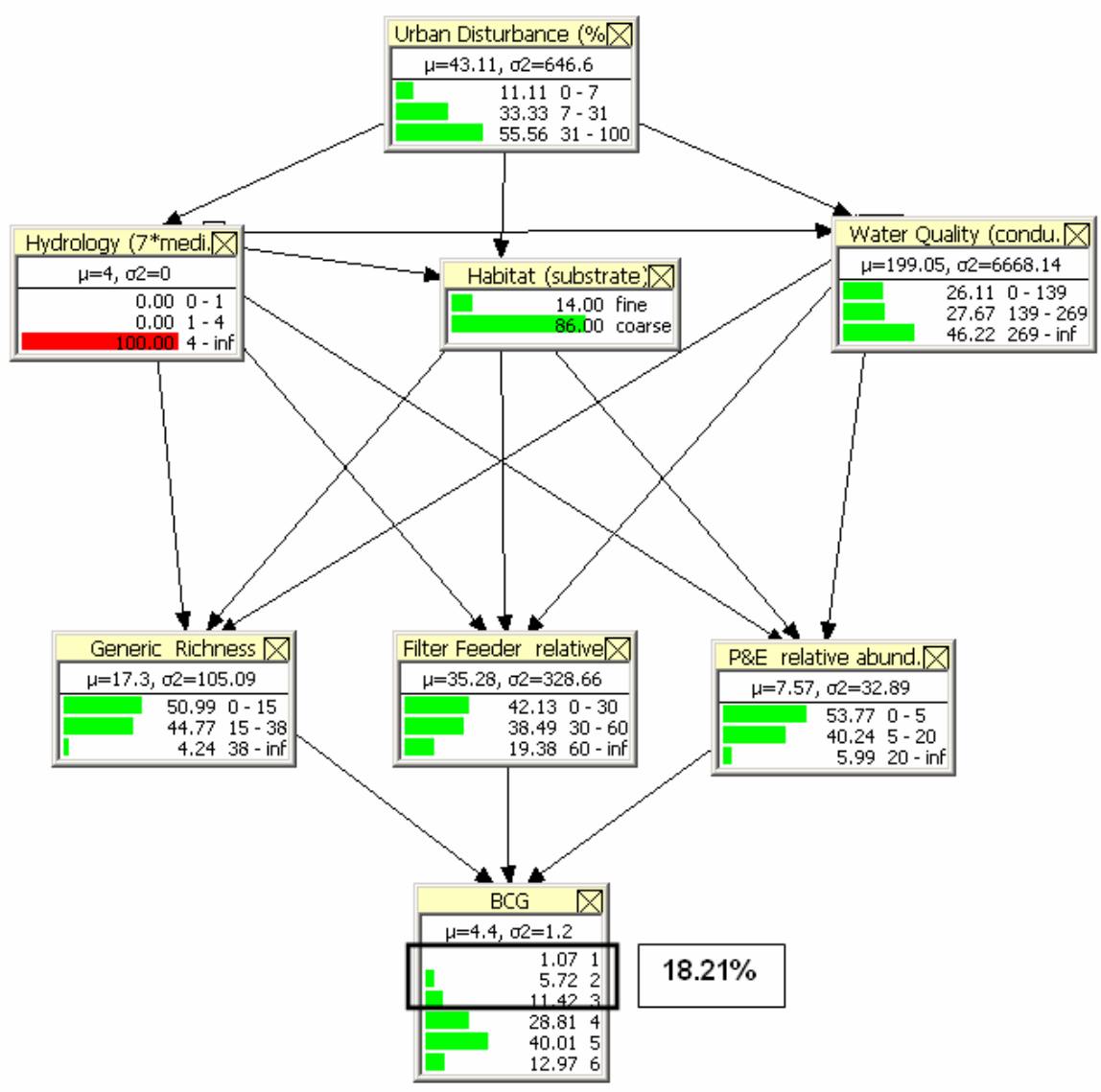

Figure 83: Posterior predictive probabilities for high flashiness.

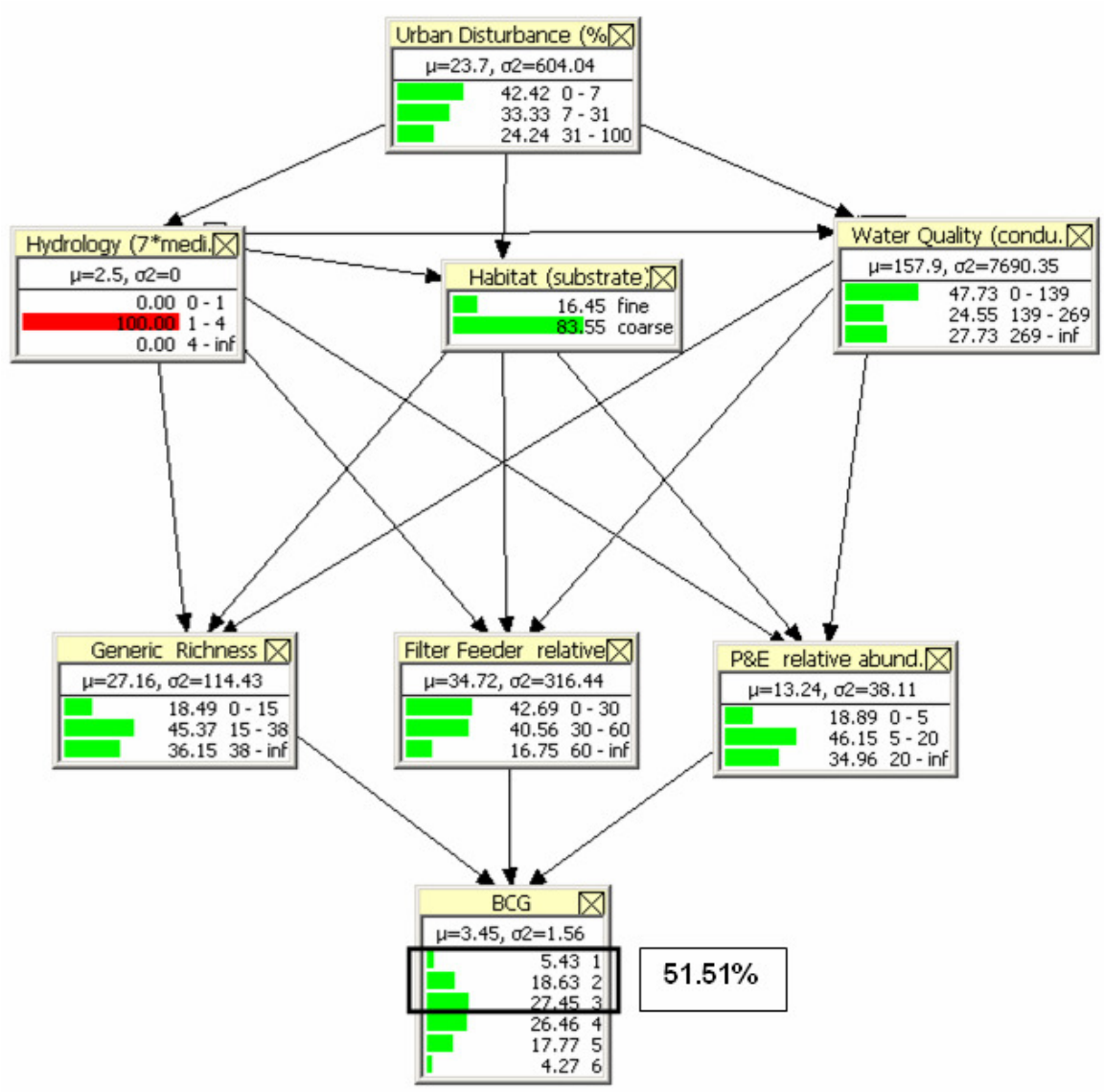

Figure 84: Posterior predictive probabilities for medium flashiness.

The influence of each of the three environmental drivers on BCG Tier was assessed in this manner (Table 62). Improving flashiness had greater effect on 'good' BCG Tier achievement than improving habitat or water quality, both in terms of absolute endpoint likelihood and percent improvement.

Table 62: Assessment of importance of drivers in improving BCG Tier achievement likelihood.

	Possible stream condition improvements increases likelihood of achieving BCG Tiers by:		
	Flashiness high to medium	**Substrate** fine to coarse	**Conductivity** high to medium
BCG Tier 1	1.07% → 5.43% Δ = 4.36%	0.83% → 4.33% Δ = 3.50%	0.65% → 2.87% Δ = 2.22%
BCG Tier >2 (so, 1 & 2)	6.79% → 24.06% Δ = 17.27%	5.56% → 19.57% Δ = 14.01%	4.57% → 13.95% Δ = 9.38%
BCG Tier >3 (so, 1, 2, & 3)	18.21% → 51.51% Δ = 33.30%	14.06% → 45.00% Δ = 30.94%	16.63% → 37.03% Δ = 20.40%

Clearly, before this type of modeling construct can be used for actual environmental system management, it must be specified how exactly flashiness, substrate or conductivity are to be managed. This can be done external to this type of model, where sub-level models are built that link specific management actions such as storm water mitigation to these direct invertebrate drivers. Alternately, specific management nodes can be incorporated internally into the Bayesian network itself. The flexibility of the Bayesian network approach makes these modifications to model structure straightforward.

4.6. Discussion

This Notheast Bayesian network model development and updating accomplished two major objectives. First, an innovative method of modeling multiple

compartment problems was created and implemented. And, second, this multiple compartment system was linked to a management context with a meaningful endpoint. Bayesian network analysis also fills many previous data analysis voids including lack of full-system-level parameterization and understanding, through uncertainty characterization, integration of multiple sources of information, insufficient data sample size, and lack of flexible model assumptions to be able to tailor-make the model to fit data and problem context.

Benefits of this northeast BCG Bayesian network are extensive. In reinterpreting a portion of the comprehensive existing USGS EUSE dataset, the created Bayesian network quantifies the simultaneous influences of multiple aspects of urbanization on macroinvertebrate condition while also quantifying the uncertainty in those relationships. Doing this provides the ability to think of these relationships between variables in terms of an interactive cause-and-effect network rather than a list of model parameters with error bars. We can now model the effect of a change in variable on every other component in the entire network, enabling multidimensional data interpretation in a way not previously possible. It is incredibly useful to have predicted outcomes in terms of the probability of attaining each BCG Tier which gives the model user a more realistic understanding of outcome likelihood rather than an overconfident exact prediction. Managers can ultimately use this Bayesian network to predict the results of possible actions on biological condition and to assess importance of drivers in

a way that realistically takes into account system interactions. Because it links to a meaningful, defined aquatic standard (BCG Tier), this network can eventually function as a decision analysis tool for ecosystem managers.

Due to the adaptive nature of Bayes Theorem, if more data were to be collected about the variables of this Northeast stream urbanization model, the current posteriors would become the new priors and the model could be further informed. This flexible construct allows the model to 'learn' as more data is incorporated. Using this modeling framework, we can quantify the process of scientific hypothesis creation and verification in a transparent manner. This means clearly presenting which expert prior assumptions were included in which iteration of the model, instead of concealing implicit expert input and falsely calling model building 'objective'. Over time, as more data on the system becomes available, they will override incorrect scientific assumptions and fortify correct ones leading to a much more systematic implementation of the scientific process.

Future directions include model expansion to incorporate additional, important system components into the model structure. Once models are developed at the ecosystem level, they can be linked together in a hierarchical Bayesian network construct to develop a multi-tiered understanding of urbanization processes across the country. Ultimately, this work paves the way toward the effective and illuminating use of Bayesian networks in ecological problem solving.

5. Dissertation Conclusions

5.1. Benefits of New Modeling Methods

The new Bayesian modeling methods developed and implemented in this dissertation are successful in solving several previously unanswered questions in USGS EUSE data analysis, and in urbanization modeling, in general. Original basin-level, single predictor regression models (Cuffney et al. 2009a) are unable to accomplish two major goals:

1. Understanding or quantifying differences between regional responses of macroinvertebrate communities to urbanization
2. Addressing the influences of multiple urbanization induced stressors acting simultaneously on an aquatic ecosystem response of interest

The models described in this dissertation effectively address these two dilemmas by creating multilevel hierarchical models and Bayesian network models, respectively.

5.1.1. Multilevel Hierarchical Models

The multilevel hierarchical models (described in Chapter 2) predict effects of urbanization at two scales simultaneously (basin-scale and region-scale). The benefits of such a modeling approach on a hierarchical nested dataset such as USGS EUSE are statistical, scientific and practical. Statistically, multilevel hierarchical models harness the multi-tiered data structure in an efficient manner that minimizes variance by borrowing strength from completely pooled estimates. Scientifically, these models

increase understanding about spatial scale effects in urbanization patterns. While at the basin scale, watershed urbanization impacts stream invertebrate communities in certain negative ways, at a larger scale between basins, different larger scale variables such as climate or previous land use patterns become important in understanding the processes of urbanization within different background conditions. In addition to clarifying understanding of these regional differences, multilevel hierarchical modeling also quantifies these differences to foster interpretation of where and why invertebrates respond less to urbanization and exactly how much less.

And, finally, practically, models that predict lower scale tiers from higher scale tiers can be used to extrapolate to areas that have not yet been sampled at the basin scale from only regional data. This cannot be done with single scale models predicting basin scale response from basin scale drivers because some form of basin scale data is required. However, if the relationships between basin and region scale are modeled and understood, then regional information can be used to predict basin scale response patterns without the need for expensive, time intensive collection of basin scale data.

Future multilevel hierarchical model development should focus on expanding current maximum likelihood estimates into fully Bayesian WinBUGS solutions, incorporating prior information about coefficient distributions. More management oriented system drivers can be modeled at the basin and region levels, and this modeling approach can be expanded to evaluate effects of urbanization on algae and

fish assemblages. Using region-level coefficients, ecological effects can be predicted in regions for which only region-level predictor measures exist and compared to future real world measurements (for example, datasets to be collected in Chicago, Anchorage, and Seattle). Development and implementation of quantitative model evaluation and verification criteria are essential in order to be able to judge the quality of results.

5.1.2. Bayesian Network Models

Bayesian network models (described in Chapters 3 and 4) predict a system of urbanization drivers as they affect biological condition using a nodes-and-arrows graphical interface. A large benefit of this approach over single-predictor regression is its more realistic, thorough representation of a known multiple stressor problem. These Bayesian network models are able to model multiple stressors because they quantify probabilistic relationships between variables in the form of a system of concepts (nodes) and connections between those concepts (arrows). Once such network structure is established, conditional probabilities between nodes are used to model multiple, interacting system components representing the reality that aquatic invertebrates are driven by many environmental factors. The use of these conditional probabilities allow for decomposition of complex overall relationships and description of uncertainty. This Bayesian network construct enables prediction of outcomes at any node, and can incorporate and appropriately weight expert knowledge and data.

Expert knowledge is used implicitly all the time in model development for variable selection and model parameterization. As modelers, we need to acknowledge this use of expert knowledge and systematically, explicitly incorporate it. This modeling philosophy benefits model development by focusing on the science and on confirming or denying expected relationships, as is done when using a hypothesis approach to research. Additionally, expert knowledge incorporation can address problematic data issues, such as the noted water chemistry grab sample scale mismatch.

Ultimately, the Bayesian network models developed in this dissertation increases conceptualization of the overlapping, system scale environmental and ecological processes of urbanization, in a much more comprehensive manner than previously possible. These models enable the analysis of the entire system together and acknowledge and demonstrate that biological response is driven by many factors simultaneously. By conceptualizing the urbanization process in this manner, these Bayesian network models can incorporate management actions and predict the effects of those actions probabilistically. This Bayesian network approach is particularly useful in the study design phase, as it is a good way to evaluate the state of expert knowledge and make data collection decisions in the interest of confirming or denying expected relationships and filling assessed knowledge gaps. This iterative process of Bayesian network model building, using model structure to inform data collection and

subsequent data collection to inform future model structure, is consistent with the iterative nature of the scientific method.

Future Bayesian network research directions include direct incorporation of explicit management nodes to move this modeling methodology closer to management application, and more thorough evaluation of the impacts of model structure size (numbers of nodes and arrows), variable selection, bin discretization, prior weight elicitation, and scale mismatch on model performance.

5.2. Major Contributions to the Literature

In addition to the general application of advanced modeling methods to the prediction of effects of urbanization on environmental endpoints of importance, this dissertation has contributed seven major, unique additions to the Bayesian modeling literature.

First, the use of multilevel hierarchical modeling has provided a statistically more rigorous and ecologically more interpretable methodology for analyzing nested environmental data. By borrowing statistical strength from the whole nine region dataset, uncertainty in regional estimates is reduced and understanding about regional differences in urbanization effects is increased.

Second, it has outlined a specific expert elicitation strategy for use in defining and parameterizing prior Bayesian networks directly from expert knowledge. This strategy has then been implemented in two different network creation examples, and

benefits and drawbacks have been catalogued and evaluated. This comprehensive exercise greatly assists and informs future expert-driven Bayesian network development efforts.

Third, this Bayesian network prior model is actually updated with a complete dataset using conjugate Dirichlet-multinomial distribution forms to fully implement Bayes theorem. In previous environmental modeling literature, only prior Bayesian network models have been developed from a combination of expert information and data. In contrast, this dissertation explicitly separates information available from expert knowledge and data, and, instead, uses data to systematically update each CPT row of an expert prior. This process truly and fully incorporates the principles of Bayes theorem in a way that developing only a prior Bayesian network does not. The trials and implications of a fully Bayesian network are elucidated and explored for the first time in this body of work.

Fourth, this dissertation takes first steps in suggesting and devising methods for comparing prior information to data in a Bayesian network CPT-based posterior. This problem of representing and evaluating how discrete, probabilistic information changes in multiple dimensions has not yet been adequately addressed in the literature. Therefore, the exploration of conclusions drawn from combining mutual information, sensitivity analysis, and weighted score graphing techniques in this dissertation is a seminal contribution to the Bayesian network evaluation discussion.

Fifth, through this thorough Bayesian network development and application, this dissertation explores problem areas of this modeling approach in depth as applied to environmental problem solving. By questioning the implications of variable selection, discretization, scale mismatch, potential prior or data bias, model size, and many other aspects of Bayesian network development, problem areas are identified for future work.

Sixth, this body of work parameterizes the conceptual Biological Condition Gradient (BCG) as applied to urban stressors. The BCG on its own is simply a set of narrative descriptions of six generic tiers of degrading biological conditions. This dissertation specifically defines how each of these tiers probabilistically relates to invertebrate metrics in the Northeast United States, making it a defined endpoint for and urbanization management model.

Finally, a Bayesian network framework is introduced as a method of creating a decision tool to use for management of the effects of urbanization on aquatic ecosystems. This type of model can be used to rank the likelihood of attaining desired management endpoints given different management actions. This flexible modeling construct can incorporate a variety of possible management actions and predict biological effects probabilistically. The Bayesian network model end product is interactive and easy for managers to use without having to calculate all the mathematical distributions directly. These benefits of a Bayesian network framework combine to give this tool enormous potential for use in environmental management decision making.

Appendix A

R Code for Results and Analysis

```
#Analysis section of USGS report

library(arm)

dataDIR <- “z:/EUSE/Data”



## the data

EUSE <- read.csv(paste(dataDIR,”EUSE_USGSReportData.csv”,

sep=”/”), header=T)

attach(EUSE)

n<-9

REGION.name <- as.vector(REGION)   #length(REGION.name)=261

(datapoints)-no Denver outlier

uniqREGION <- unique(REGION.name)

REGIONprecip<-tapply(AnnMeanP,REGION,mean)

REGIONtemp<-tapply(AnnMeanT,REGION,mean)



REGIONbackag<-rep(NA,n)
```

```
P.NLCD78<-P.NLCD7+P.NLCD8

for (i in 1:9){

  REGIONbackag[i]<-mean(P.NLCD78[REGION.UII<=10&REGIONindex==i])

}

REGIONbackag.cat<-rep(NA,n)

for (i in 1:9){

if (REGIONbackag[i]<=50){

  REGIONbackag.cat[i]<-0

  } else {

  REGIONbackag.cat[i]<-1

  }

}



## the models

#set urban predictor

URB<-P.NLCD2/100
```

```
#set eco response variable-- pick one and uncomment it

#REGULAR LINEAR REGRESSION

#ECO<-NMDS1; ECO.name<-"NMDS1"

#ECO<-RICHTOL; ECO.name<-"RICHTOL"



#POISSON GENERALIZED LINEAR REGRESSION

ECO<-RICH; ECO.name<-"RICH"

#ECO<-EPTRICH; ECO.name<-"EPTRICH"



############################################################

## REGULAR LINEAR REGRESSION-- NMDS1 and RICHTOL

#complete pooling

lm.pooled <- lm(ECO ~ URB)      #one slope and one intercept

#no pooling (separate slopes and intercepts)

ab.hat.unpooled <- array (NA, c(n,2))

for (j in 1:n){

  lm.unpooled <- lm (ECO ~ URB, subset=(REGIONindex==j))       #9
(separate) slopes and 9 (separate) intercepts
```

```
  ab.hat.unpooled[j,] <- summary(lm.unpooled)$coef[,1]
}

#partial pooling- varying intercept and slope by group
#MODEL 1:  no group-level predictor
M1 <- lmer(ECO ~ URB + (1+URB|REGION))

#MODEL 2:  precip group-level predictor
M2 <- lmer(ECO ~ URB + REGIONprecip[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION))

#MODEL 3:  temp group-level predictor
M3 <- lmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONtemp[REGION]+(1+URB|REGION))

#MODEL 4:  temp group-level predictor for intercept; precip
predictor for slope
M4 <- lmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION))
```

```
#MODEL 5:  antecedent ag+past group-level predictor

M5 <- lmer(ECO ~ URB + REGIONbackag[REGION] +
URB:REGIONbackag[REGION]+(1+URB|REGION))


#MODEL 6:  categorical ag, precip group-level predictor

M6 <- lmer(ECO ~ URB + REGIONprecip[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION)+(1+URB|REGIONbackag.cat[R
EGION]))


#MODEL 7:  categorical ag, temp group-level predictor

M7 <- lmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONtemp[REGION]+(1+URB|REGION)+(1+URB|REGIONbackag.cat[REG
ION]))


#MODEL 8:  categorical ag, temp group-level predictor for
intercept; precip predictor for slope

M8 <- lmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION)+(1+URB|REGIONbackag.cat[R
EGION]))
```

```
############################################################

## GENERALIZED (POISSON) LINEAR REGRESSION-- RICH and EPTRICH

#complete pooling

lm.pooled <- glm (ECO ~ URB,family=poisson)

#one slope and one intercept

#no pooling (separate slopes and intercepts)

ab.hat.unpooled <- array (NA, c(n,2))

for (j in 1:n){

  lm.unpooled <- glm (ECO ~ URB,
subset=(REGIONindex==j),family=poisson)

  #9 (separate) slopes and 9 (separate) intercepts

  ab.hat.unpooled[j,] <- summary(lm.unpooled)$coef[,1]

}



#partial pooling- varying intercept and slope by group

#MODEL 1:  no group-level predictor

M1 <- glmer(ECO ~ URB + (1+URB|REGION),family=poisson)
```

```
#MODEL 2:  precip group-level predictor

M2 <- glmer(ECO ~ URB + REGIONprecip[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION),family=poisson)


#MODEL 3:  temp group-level predictor

M3 <- glmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONtemp[REGION]+(1+URB|REGION),family=poisson)


#MODEL 4:  temp group-level predictor for intercept; precip
predictor for slope

M4 <- glmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION),family=poisson)


#MODEL 5:  antecedent ag+past group-level predictor

M5 <- glmer(ECO ~ URB + REGIONbackag[REGION] +
URB:REGIONbackag[REGION]+(1+URB|REGION),family=poisson)


#MODEL 6:  categorical ag, precip group-level predictor
```

```
M6 <- glmer(ECO ~ URB + REGIONprecip[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION)+(1+URB|REGIONbackag.cat[R
EGION]),family=poisson)


#MODEL 7:  categorical ag, temp group-level predictor

M7 <- glmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONtemp[REGION]+(1+URB|REGION)+(1+URB|REGIONbackag.cat[REG
ION]),family=poisson)


#MODEL 8:  categorical ag, temp group-level predictor for
intercept; precip predictor for slope

M8 <- glmer(ECO ~ URB + REGIONtemp[REGION] +
URB:REGIONprecip[REGION]+(1+URB|REGION)+(1+URB|REGIONbackag.cat[R
EGION]),family=poisson)


############################################################

#MODEL COEFFICIENTS--can use for all ECO response variables, one
at a time

a.hat.M1 <- coef(M1)$REGION[,1]

b.hat.M1 <- coef(M1)$REGION[,2]
```

```
a.hat.M2 <- coef(M2)$REGION[,1] +
coef(M2)$REGION[,3]*REGIONprecip

b.hat.M2 <- coef(M2)$REGION[,2] +
coef(M2)$REGION[,4]*REGIONprecip

a.se.M2 <- se.ranef(M2)$REGION[,1]

b.se.M2 <- se.ranef(M2)$REGION[,2]



a.hat.M3 <- coef(M3)$REGION[,1] + coef(M3)$REGION[,3]*REGIONtemp

b.hat.M3 <- coef(M3)$REGION[,2] + coef(M3)$REGION[,4]*REGIONtemp

a.se.M3 <- se.ranef(M3)$REGION[,1]

b.se.M3 <- se.ranef(M3)$REGION[,2]



a.hat.M4 <- coef(M4)$REGION[,1] + coef(M4)$REGION[,3]*REGIONtemp

b.hat.M4 <- coef(M4)$REGION[,2] +
coef(M4)$REGION[,4]*REGIONprecip

a.se.M4 <- se.ranef(M4)$REGION[,1]

b.se.M4 <- se.ranef(M4)$REGION[,2]
```

```
a.hat.M5 <- coef(M5)$REGION[,1] +
coef(M5)$REGION[,3]*REGIONbackag

b.hat.M5 <- coef(M5)$REGION[,2] +
coef(M5)$REGION[,4]*REGIONbackag

a.se.M5 <- se.ranef(M5)$REGION[,1]

b.se.M5 <- se.ranef(M5)$REGION[,2]



M6.fixef <- fixef(M6)

M6.ranef <- ranef(M6)

M6.seranef <- se.ranef(M6)

a.hat.M6 <- M6.fixef[1] + M6.ranef[[1]][,1] +
M6.ranef[[2]][c(1,1,1,2,2,2,1,1,1),1] + M6.fixef[3]*REGIONprecip

b.hat.M6 <- M6.fixef[2] + M6.ranef[[1]][,2] +
M6.ranef[[2]][c(1,1,1,2,2,2,1,1,1),2] + M6.fixef[4]*REGIONprecip

a.se.M6 <- M6.seranef[[1]][,1]

b.se.M6 <- M6.seranef[[1]][,2]



M7.fixef <- fixef(M7)
```

```
M7.ranef <- ranef(M7)

M7.seranef <- se.ranef(M7)

a.hat.M7 <- M7.fixef[1] + M7.ranef[[1]][,1] +

M7.ranef[[2]][c(1,1,1,2,2,2,1,1,1),1] + M7.fixef[3]*REGIONtemp

b.hat.M7 <- M7.fixef[2] + M7.ranef[[1]][,2] +

M7.ranef[[2]][c(1,1,1,2,2,2,1,1,1),2] + M7.fixef[4]*REGIONtemp

a.se.M7 <- M7.seranef[[1]][,1]

b.se.M7 <- M7.seranef[[1]][,2]



M8.fixef <- fixef(M8)

M8.ranef <- ranef(M8)

M8.seranef <- se.ranef(M8)

a.hat.M8 <- M8.fixef[1] + M8.ranef[[1]][,1] +

M8.ranef[[2]][c(1,1,1,2,2,2,1,1,1),1] + M8.fixef[3]*REGIONtemp

b.hat.M8 <- M8.fixef[2] + M8.ranef[[1]][,2] +

M8.ranef[[2]][c(1,1,1,2,2,2,1,1,1),2] + M8.fixef[4]*REGIONprecip

a.se.M8 <- M8.seranef[[1]][,1]

b.se.M8 <- M8.seranef[[1]][,2]
```

```
#GRAPHS#################################################

#ONE 9-REGION x vs. y--- for regular linear regressions (AV1 or
RICHTOL)

#M1: no group-level predictors

OUT<-"z:/EUSE/USGS Report/Inverts/USGS Final report drafts and
comments/Final NMDS1 Figs"

postscript(file=paste(OUT, "USGS.M1.eps",sep="/"), width=8,
height=8.5, horizontal=F, paper="special")

par (mfrow=c(3,3), mar=c(4,4,3,1), oma=c(1,1,2,1))

for (j in 1:n){

    plot (URB[REGIONindex==j]*100, ECO[REGIONindex==j],
    xlim=c(min(URB*100),max(URB*100)),
    ylim=c(min(ECO[!is.na(ECO)==TRUE]),max(ECO[!is.na(ECO)==TRUE]
    )),xlab="URB", ylab=ECO.name, cex=1.5, cex.lab=1.2,
    cex.axis=1.2,main=uniqREGION[j],cex.main=1.5)

    curve (coef(lm.pooled)[1] + coef(lm.pooled)[2]/100*x, lwd=.5,
    lty=2, add=T)        #completely pooled=dashed line

    curve (ab.hat.unpooled[j,1] + ab.hat.unpooled[j,2]/100*x,
    lwd=.5, col="gray10", add=T)   #unpooled=black line
```

```
    curve (a.hat.M1[j] + b.hat.M1[j]/100*x, lwd=1, col=3, add=T)
    #partially pooled- green line
}
dev.off()


#ONE 9-REGION x vs. y--- for Poission generalized linear
regressions (TR or EPT)
#M1: no group-level predictors
OUT<-"z:/EUSE/USGS Report/Inverts/USGS Final report drafts and
comments/Final EPTRICH Figs"
postscript(file=paste(OUT, "USGS.M1.eps",sep="/"), width=8,
height=8.5, horizontal=F, paper="special")
par (mfrow=c(3,3), mar=c(4,4,3,1), oma=c(1,1,2,1))
for (j in 1:n){
    plot (URB[REGIONindex==j]*100, ECO[REGIONindex==j],
    xlim=c(0,100), ylim=c(min(ECO),max(ECO)),xlab="URB",
    ylab=ECO.name, cex=1.5, cex.lab=1.2, cex.axis=1.2,
    main=uniqREGION[j], cex.main=1.5)
    curve (exp(coef(lm.pooled)[1] + coef(lm.pooled)[2]/100*x),
    lwd=.5, lty=2, add=T)        #completely pooled=dashed line
```

```
    curve (exp(ab.hat.unpooled[j,1] +
    ab.hat.unpooled[j,2]/100*x), lwd=.5, col="gray10", add=T)
    #unpooled=black line

    curve (exp(a.hat.M1[j] + b.hat.M1[j]/100*x), lwd=1, col=3,
    add=T)     #partially pooled- green line

}

dev.off()


#NINE estimated intercept and slopes across REGION-level
predictor--for all ECO variables

#M2: precip REGION-level predictor

lower.aM2 <- a.hat.M2 - a.se.M2

upper.aM2 <- a.hat.M2 + a.se.M2

postscript(file=paste(OUT, "USGS.M2.eps",sep="/"), width=11.5,
height=6, horizontal=F,paper="special")

par (mfrow=c(1,2),mgp=c(1.5,0.5,0),tck=-0.02,mar=c(3,3,3,1))

    plot (REGIONprecip, a.hat.M2,
    ylim=range(lower.aM2,upper.aM2+0.1),
```

```
    xlab="REGION-level annual precipitation",
    ylab=expression(paste("Estimated intercept, ", alpha[j])),
    pch=19,xlim=c(35,170), main=paste("Multilevel Model 2
    (",ECO.name,"),
    intercept with precipitation"))
    curve (fixef(M2)["(Intercept)"] +
    fixef(M2)["REGIONprecip[REGION]"]*x, lwd=1, col="black",
    add=TRUE)
    segments (REGIONprecip, lower.aM2, REGIONprecip, upper.aM2,
    lwd=.5, col="gray10")
    text(x=REGIONprecip,y=a.hat.M2,labels=uniqREGION,pos=c(4,4,4,
    4,4,4,4,4,4))


lower.bM2 <- b.hat.M2 - b.se.M2
upper.bM2 <- b.hat.M2 + b.se.M2
    plot (REGIONprecip, b.hat.M2,
ylim=range(lower.bM2,upper.bM2),
    xlab="REGION-level annual precipitation",
    ylab=expression(paste("Estimated slope, ", beta[j])), pch=19,
    xlim=c(35,170),main=paste("Multilevel Model 2(",ECO.name,"),
```

```
    slope with precipitation”))

    curve (fixef(M2)[“URB”] +
    fixef(M2)[“URB:REGIONprecip[REGION]”]*x, lwd=1, col=”black”,
    add=TRUE)

    segments (REGIONprecip, lower.bM2, REGIONprecip, upper.bM2,
    lwd=.5, col=”gray10”)

    text(x=REGIONprecip,y=b.hat.M2,labels=uniqREGION,pos=c(4,4,4,
    4,4,4,4,4,4))

dev.off()


#M3: temp REGION-level predictor

lower.aM3 <- a.hat.M3 - a.se.M3

upper.aM3 <- a.hat.M3 + a.se.M3

postscript(file=paste(OUT, “USGS.M3.eps”,sep=”/”), width=11.5,
height=6, horizontal=F,paper=”special”)

par (mfrow=c(1,2),mgp=c(1.5,0.5,0),tck=-0.02,mar=c(3,3,3,1))

    plot (REGIONtemp, a.hat.M3, ylim=range(lower.aM3,upper.aM3),

    xlab=”REGION-level annual temperature”,
    ylab=expression(paste(“Estimated intercept, “, alpha[j])),
```

```
    pch=19, xlim=c(7,20),main=paste("Multilevel Model 3
    (",ECO.name,"),
    intercept with temperature"))
    curve (fixef(M3)["(Intercept)"] +
    fixef(M3)["REGIONtemp[REGION]"]*x, lwd=1, col="black",
    add=TRUE)
    segments (REGIONtemp, lower.aM3, REGIONtemp, upper.aM3,
    lwd=.5, col="gray10")
    text(x=REGIONtemp,y=a.hat.M3,labels=uniqREGION,pos=c(4,4,4,4,
    4,4,4,2,4))


lower.bM3 <- b.hat.M3 - b.se.M3
upper.bM3 <- b.hat.M3 + b.se.M3
    plot (REGIONtemp, b.hat.M3, ylim=range(lower.bM3,upper.bM3),
    xlab="REGION-level annual temperature",
    ylab=expression(paste("Estimated slope, ", beta[j])), pch=19,
    xlim=c(7,20),main=paste("Multilevel Model 3 (",ECO.name,"),
    slope with temperature"))
```

```
    curve (fixef(M3)["URB"] +
    fixef(M3)["URB:REGIONtemp[REGION]"]*x, lwd=1, col="black",
    add=TRUE)

    segments (REGIONtemp, lower.bM3, REGIONtemp, upper.bM3,
    lwd=.5, col="gray10")

    text(x=REGIONtemp,y=b.hat.M3,labels=uniqREGION,pos=c(4,4,4,4,
    4,4,4,4,4))

dev.off()


#M4:  temp group-level predictor for intercept; precip predictor
for slope

lower.aM4 <- a.hat.M4 - a.se.M4

upper.aM4 <- a.hat.M4 + a.se.M4

postscript(file=paste(OUT, "USGS.M4.eps",sep="/"), width=11.5,
height=6, horizontal=F,paper="special")

par (mfrow=c(1,2),mgp=c(1.5,0.5,0),tck=-0.02,mar=c(3,3,3,1))

    plot (REGIONtemp, a.hat.M4, ylim=range(lower.aM4,upper.aM4),

    xlab="REGION-level annual temperature",
    ylab=expression(paste("Estimated intercept, ", alpha[j])),
```

```
    pch=19, xlim=c(7,20),main=paste("Multilevel Model 4
    (",ECO.name,"),
    intercept with temperature"))
    curve (fixef(M4)["(Intercept)"] +
    fixef(M4)["REGIONtemp[REGION]"]*x, lwd=1, col="black",
    add=TRUE)
    segments (REGIONtemp, lower.aM4, REGIONtemp, upper.aM4,
    lwd=.5, col="gray10")
    text(x=REGIONtemp,y=a.hat.M4,labels=uniqREGION,pos=c(4,4,2,4,
    4,4,4,4,4))


lower.bM4 <- b.hat.M4 - b.se.M4
upper.bM4 <- b.hat.M4 + b.se.M4
    plot (REGIONprecip, b.hat.M4,
ylim=range(lower.bM4,upper.bM4),
    xlab="REGION-level annual precipitation",
    ylab=expression(paste("Estimated slope, ", beta[j])), pch=19,
    xlim=c(35,170),main=paste("Multilevel Model 4(",ECO.name,"),
    slope with precipitation"))
```

```
    curve (fixef(M4)["URB"] +
    fixef(M4)["URB:REGIONprecip[REGION]"]*x, lwd=1, col="black",
    add=TRUE)

    segments (REGIONprecip, lower.bM4, REGIONprecip, upper.bM4,
    lwd=.5, col="gray10")

    text(x=REGIONprecip,y=b.hat.M4,labels=uniqREGION,pos=c(4,4,4,
    4,4,4,4,4,4))

dev.off()


#M5:  antecedent ag+past group-level predictor

lower.aM5 <- a.hat.M5 - a.se.M5

upper.aM5 <- a.hat.M5 + a.se.M5

postscript(file=paste(OUT, "USGS.M5.eps",sep="/"), width=11.5,
height=6, horizontal=F,paper="special")

par (mfrow=c(1,2),mgp=c(1.5,0.5,0),tck=-0.02,mar=c(3,3,3,1))

    plot (REGIONbackag, a.hat.M5,
ylim=range(lower.aM5,upper.aM5),

    xlab="REGION-level antecedent ag+pasture",
    ylab=expression(paste("Estimated intercept, ", alpha[j])),
```

```
    pch=19,xlim=c(-2,100), main=paste("Multilevel Model 5
    (",ECO.name,"),
    intercept with antecedent ag+pasture"))
    curve (fixef(M5)["(Intercept)"] +
    fixef(M5)["REGIONbackag[REGION]"]*x, lwd=1, col="black",
    add=TRUE)
    segments (REGIONbackag, lower.aM5, REGIONbackag, upper.aM5,
    lwd=.5, col="gray10")
    text(x=REGIONbackag,y=a.hat.M5,labels=uniqREGION,pos=c(4,4,4,
    4,4,4,4,4,2))


lower.bM5 <- b.hat.M5 - b.se.M5
upper.bM5 <- b.hat.M5 + b.se.M5
plot (REGIONbackag, b.hat.M5, ylim=range(lower.bM5,upper.bM5),
    xlab="REGION-level antecedent ag+pasture",
    ylab=expression(paste("Estimated slope, ", beta[j])), pch=19,
    xlim=c(-2,100),main=paste("Multilevel Model 5(",ECO.name,"),
    slope with antecedent ag+pasture"))
```

```
    curve (fixef(M5)["URB"] +
    fixef(M5)["URB:REGIONbackag[REGION]"]*x, lwd=1, col="black",
    add=TRUE)

    segments (REGIONbackag, lower.bM5, REGIONbackag, upper.bM5,
    lwd=.5, col="gray10")

    text(x=REGIONbackag,y=b.hat.M5,labels=uniqREGION,pos=c(4,4,4,
    4,4,4,2,4,2))

dev.off()


#M6: categorical ag, precip REGION-level predictor

lower.aM6 <- a.hat.M6 - a.se.M6

upper.aM6 <- a.hat.M6 + a.se.M6

postscript(file=paste(OUT, "USGS.M6.eps",sep="/"), width=11.5,
height=6, horizontal=F,paper="special")

par (mfrow=c(1,2),mgp=c(1.5,0.5,0),tck=-0.02,mar=c(3,3,3,1))

    plot (REGIONprecip, a.hat.M6,
ylim=range(lower.aM6,upper.aM6),

    xlab="REGION-level annual precipitation",
    ylab=expression(paste("Estimated intercept, ", alpha[j])),
```

```
    pch=19, xlim=c(35,170), col=c(4,4,4,2,2,2,4,4,4),
    main=paste("Multilevel Model 6 (",ECO.name,"),
    intercept with ag+grassland and precipitation "))
    curve (fixef(M6)["(Intercept)"] + M6.ranef[[2]][1,1] +
    fixef(M6)["REGIONprecip[REGION]"]*x, lwd=1,col="blue",
    add=TRUE)
    curve (fixef(M6)["(Intercept)"] + M6.ranef[[2]][2,1] +
    fixef(M6)["REGIONprecip[REGION]"]*x, lwd=1,col="red",
    add=TRUE)
    segments (REGIONprecip, lower.aM6, REGIONprecip, upper.aM6,
    lwd=.5, col=c(4,4,4,2,2,2,4,4,4))
    text(x=REGIONprecip,y=a.hat.M6,labels=uniqREGION,pos=c(4,4,4,
    4,4,4,4,4,4),col=c(4,4,4,2,2,2,4,4,4))


lower.bM6 <- b.hat.M6 - b.se.M6
upper.bM6 <- b.hat.M6 + b.se.M6
    plot (REGIONprecip, b.hat.M6,
ylim=range(lower.bM6,upper.bM6),
    xlab="REGION-level annual precipitation",
    ylab=expression(paste("Estimated slope, ", beta[j])),pch=19,
```

```
    xlim=c(35,170),col=c(4,4,4,2,2,2,4,4,4),main=paste("Multileve
    l Model 6 (",ECO.name,"),

    slope with ag+grassland and precipitation"))

    curve (fixef(M6)["URB"] + M6.ranef[[2]][1,2] +

    fixef(M6)["URB:REGIONprecip[REGION]"]*x,lwd=1,col="blue",

    add=TRUE)

    curve (fixef(M6)["URB"] + M6.ranef[[2]][2,2] +

    fixef(M6)["URB:REGIONprecip[REGION]"]*x, lwd=1,

    col="red", add=TRUE)

    segments (REGIONprecip, lower.bM6, REGIONprecip, upper.bM6,

    lwd=.5, col=c(4,4,4,2,2,2,4,4,4))

    text(x=REGIONprecip,y=b.hat.M6,labels=uniqREGION,pos=c(4,4,4,

    4,4,4,4,4,4),col=c(4,4,4,2,2,2,4,4,4))

dev.off()



#M7: categorical ag, temp REGION-level predictor

lower.aM7 <- a.hat.M7 - a.se.M7

upper.aM7 <- a.hat.M7 + a.se.M7
```

```
postscript(file=paste(OUT, "USGS.M7.eps",sep="/"), width=11.5,
height=6, horizontal=F,paper="special")

par (mfrow=c(1,2),mgp=c(1.5,0.5,0),tck=-0.02,mar=c(3,3,3,1))

    plot (REGIONtemp, a.hat.M7, ylim=range(lower.aM7,upper.aM7),

    xlab="REGION-level annual temperature",
    ylab=expression(paste("Estimated intercept, ", alpha[j])),
    pch=19,xlim=c(7,20), col=c(4,4,4,2,2,2,4,4,4),
    main=paste("Multilevel Model 7 (",ECO.name,"),

    intercept with ag+grassland and temperature"))

    curve (fixef(M7)["(Intercept)"] + M7.ranef[[2]][1,1] +
    fixef(M7)["REGIONtemp[REGION]"]*x, lwd=1, col="blue",
    add=TRUE)

    curve (fixef(M7)["(Intercept)"] + M7.ranef[[2]][2,1] +
    fixef(M7)["REGIONtemp[REGION]"]*x, lwd=1, col="red",
    add=TRUE)

    segments (REGIONtemp, lower.aM7, REGIONtemp, upper.aM7,
    lwd=.5, col=c(4,4,4,2,2,2,4,4,4))

    text(x=REGIONtemp,y=a.hat.M7,labels=uniqREGION,pos=c(4,4,4,4,
    4,4,4,4,4),col=c(4,4,4,2,2,2,4,4,4))
```

```
lower.bM7 <- b.hat.M7 - b.se.M7

upper.bM7 <- b.hat.M7 + b.se.M7

    plot (REGIONtemp, b.hat.M7, ylim=range(lower.bM7,upper.bM7),

    xlab="REGION-level annual temperature",
    ylab=expression(paste("Estimated slope, ", beta[j])), pch=19,


    xlim=c(7,20),col=c(4,4,4,2,2,2,4,4,4),main=paste("Multilevel
    Model 7 (",ECO.name,"),

    slope with ag+grassland and temperature"))

    curve (fixef(M7)["URB"] + M7.ranef[[2]][1,2] +
    fixef(M7)["URB:REGIONtemp[REGION]"]*x, lwd=1, col="blue",
    add=TRUE)

    curve (fixef(M7)["URB"] + M7.ranef[[2]][2,2] +
    fixef(M7)["URB:REGIONtemp[REGION]"]*x, lwd=1, col="red",
    add=TRUE)

    segments (REGIONtemp, lower.bM7, REGIONtemp, upper.bM7,
    lwd=.5, col=c(4,4,4,2,2,2,4,4,4))

    text(x=REGIONtemp,y=b.hat.M7,labels=uniqREGION,pos=c(4,4,4,4,
    4,4,4,4,4),col=c(4,4,4,2,2,2,4,4,4))

dev.off()
```

```
#M8: categorical ag, temp REGION-level predictor

lower.aM8 <- a.hat.M8 - a.se.M8

upper.aM8 <- a.hat.M8 + a.se.M8

postscript(file=paste(OUT, "USGS.M8.eps",sep="/"), width=11.5,
height=6, horizontal=F,paper="special")

par (mfrow=c(1,2),mgp=c(1.5,0.5,0),tck=-0.02,mar=c(3,3,3,1))

    plot (REGIONtemp, a.hat.M8, ylim=range(lower.aM8,upper.aM8),

    xlab="REGION-level annual temperature",
    ylab=expression(paste("Estimated intercept, ", alpha[j])),
    pch=19,xlim=c(7,20), col=c(4,4,4,2,2,2,4,4,4),
    main=paste("Multilevel Model 8 (",ECO.name,"),

    intercept with ag+grassland and temperature"))

    curve (fixef(M8)["(Intercept)"] + M8.ranef[[2]][1,1] +
    fixef(M8)["REGIONtemp[REGION]"]*x, lwd=1, col="blue",
    add=TRUE)

    curve (fixef(M8)["(Intercept)"] + M8.ranef[[2]][2,1] +
    fixef(M8)["REGIONtemp[REGION]"]*x, lwd=1, col="red",
    add=TRUE)
```

```
    segments (REGIONtemp, lower.aM8, REGIONtemp, upper.aM8,
    lwd=.5, col=c(4,4,4,2,2,2,4,4,4))

    text(x=REGIONtemp,y=a.hat.M8,labels=uniqREGION,pos=c(4,2,4,4,
    4,4,4,4,4),col=c(4,4,4,2,2,2,4,4,4))


lower.bM8 <- b.hat.M8 - b.se.M8

upper.bM8 <- b.hat.M8 + b.se.M8

    plot (REGIONprecip, b.hat.M8,
ylim=range(lower.bM8,upper.bM8),

    xlab="REGION-level annual precipitation",
    ylab=expression(paste("Estimated slope, ", beta[j])), pch=19,
    xlim=c(35,170),col=c(4,4,4,2,2,2,4,4,4),main=paste("Multileve
    l Model 8 (",ECO.name,"),

    slope with ag+grassland and precipitation"))

    curve (fixef(M8)["URB"] + M8.ranef[[2]][1,2] +
    fixef(M8)["URB:REGIONprecip[REGION]"]*x, lwd=1, col="blue",
    add=TRUE)

    curve (fixef(M8)["URB"] + M8.ranef[[2]][2,2] +
    fixef(M8)["URB:REGIONprecip[REGION]"]*x, lwd=1, col="red",
    add=TRUE)
```

```
    segments (REGIONprecip, lower.bM8, REGIONprecip, upper.bM8,
    lwd=.5, col=c(4,4,4,2,2,2,4,4,4))

    text(x=REGIONprecip,y=b.hat.M8,labels=uniqREGION,pos=c(4,4,4,
    4,4,4,4,4,4),col=c(4,4,4,2,2,2,4,4,4))

dev.off()
```

Appendix B

Table rows selected for probability range prior weight elicitation are highlighted and prior expected values within those rows selected for hypothetical sample prior weight elicitation are shown in bold.

Table 63: Prior marginal probability table and prior weight (α_0) for "Urban disturbance" node (urban land cover bins in probability units and α_0 in equivalent data points).

Urban land cover				
0-12%	**>12-36%**	**>36-66%**	**>66%**	α_0
0.25	0.25	0.25	0.25	1

Table 64: Prior marginal probability table and prior weight (α_0) for "Hydrologic modifications" node (dam density bins in probability units and α_0 in equivalent data points).

Dam density			
0 dams/100 km²	**>0-6 dams/100 km²**	**>6 dams/100 km²**	α_0
0.40	0.30	0.30	1

Table 65: Prior conditional probability table and prior weight (α_0) for "Change in generation of flow" node (impervious surface bins in probability units and α_0 in equivalent data points).

Urban land cover	Impervious surface				
	0-2 %	**>2-8%**	**>8-16%**	**>16%**	α_0
0-12%	0.85	0.13	0.01	0.01	7.99
>12-36%	0.09	**0.60**	0.30	0.01	7.99
>36-66%	0.05	0.15	0.70	0.10	7.99
>66%	0.01	0.09	**0.10**	0.80	7.99

Table 66: Prior conditional probability table and prior weight (α_0) for "Habitat" node (channel ratio bins in probability units and α_0 in equivalent data points).

Urban land cover	Flashiness	Channel width:depth ratio				
		0-25 w:d	>25-33 w:d	>33-44 w:d	>44 w:d	α_0
0-12%	0-40 risings	0.05	0.10	0.30	0.55	7.30
>12-36%	0-40 risings	0.05	0.15	0.50	0.30	7.30
>36-66%	0-40 risings	0.10	0.45	0.35	0.10	7.30
>66%	0-40 risings	0.60	0.20	0.15	0.05	7.30
0-12%	41-63 risings	0.10	0.15	0.35	0.40	7.30
>12-36%	41-63 risings	0.10	0.20	0.60	0.10	7.30
>36-66%	41-63 risings	0.15	0.50	0.30	0.05	7.30
>66%	41-63 risings	**0.60**	0.30	0.09	0.01	7.30
0-12%	64-94 risings	0.20	0.50	0.29	0.01	7.30
>12-36%	64-94 risings	0.25	0.60	0.14	0.01	7.30
>36-66%	64-94 risings	0.55	0.39	0.05	0.01	7.30
>66%	64-94 risings	0.64	0.30	0.05	0.01	7.30
0-12%	95+ risings	0.20	0.40	0.39	0.01	7.30
>12-36%	95+ risings	0.25	0.45	0.29	0.01	7.30
>36-66%	95+ risings	0.25	0.50	**0.24**	0.01	7.30
>66%	95+ risings	0.60	0.34	0.05	0.01	7.30

Table 67: Prior conditional probability table and prior weight (α_0) for "Water Quality" node (conductivity bins in probability units and α_0 in equivalent data points).

Urban land cover	Flashiness	Conductivity at low base flow				
		0-92 µs	92-129 µs	129-281 µs	>281 µs	α_0
0-12%	0-40 risings	**0.80**	0.10	0.05	0.05	8.16
>12-36%	0-40 risings	0.70	0.20	0.05	0.05	8.16
>36-66%	0-40 risings	0.50	0.25	0.20	0.05	8.16
>66%	0-40 risings	0.30	0.30	0.30	0.10	8.16
0-12%	41-63 risings	0.80	0.10	0.05	0.05	8.16
>12-36%	41-63 risings	0.70	0.20	0.05	0.05	8.16
>36-66%	41-63 risings	0.50	0.25	0.20	0.05	8.16
>66%	41-63 risings	0.30	0.30	0.30	0.10	8.16
0-12%	64-94 risings	0.70	0.15	0.10	0.05	8.16
>12-36%	64-94 risings	0.65	0.20	0.10	0.05	8.16
>36-66%	64-94 risings	0.40	0.30	0.20	0.10	8.16
>66%	64-94 risings	0.20	0.35	**0.35**	0.10	8.16
0-12%	95+ risings	0.70	0.15	0.10	0.05	8.16
>12-36%	95+ risings	0.65	0.20	0.10	0.05	8.16
>36-66%	95+ risings	0.40	0.30	0.20	0.10	8.16
>66%	95+ risings	0.20	0.35	0.35	0.10	8.16

Table 68: Prior conditional probability table and prior weight (α_0) for "EPT taxa richness" node (EPT taxa bins in probability units and α_0 in equivalent data points).

Flashiness	Channel ratio	Conductivity	EPT taxa richness				
			0-4 taxa	5-7 taxa	8-12 taxa	>13 taxa	α_0
0-40 risings	0-25 w:d	0-92 µs	0.05	0.25	0.40	0.30	9.24
41-63 risings	0-25 w:d	0-92 µs	0.10	0.30	0.40	0.20	9.24
64-94 risings	0-25 w:d	0-92 µs	0.20	0.40	0.30	0.10	9.24
95+ risings	0-25 w:d	0-92 µs	0.30	0.45	0.20	0.05	9.24
0-40 risings	25-33 w:d	0-92 µs	0.01	0.19	0.30	0.50	9.24
41-63 risings	25-33 w:d	0-92 µs	0.05	0.25	0.40	0.30	9.24
64-94 risings	25-33 w:d	0-92 µs	0.10	0.30	0.40	0.20	9.24
95+ risings	25-33 w:d	0-92 µs	0.20	0.40	0.30	0.10	9.24
0-40 risings	33-44 w:d	0-92 µs	0.01	0.04	0.24	0.70	9.24
41-63 risings	33-44 w:d	0-92 µs	0.01	0.19	0.30	0.50	9.24
64-94 risings	33-44 w:d	0-92 µs	0.05	0.25	0.40	0.30	9.24
95+ risings	33-44 w:d	0-92 µs	0.10	0.30	0.40	0.20	9.24
0-40 risings	>44 w:d	0-92 µs	0.01	0.01	0.10	0.88	9.24
41-63 risings	>44 w:d	0-92 µs	0.01	0.05	0.24	0.70	9.24
64-94 risings	>44 w:d	0-92 µs	0.01	0.19	0.30	0.50	9.24
95+ risings	>44 w:d	0-92 µs	0.05	0.25	0.40	0.30	9.24
0-40 risings	0-25 w:d	92-129 µs	0.10	0.30	0.40	0.20	9.24
41-63 risings	0-25 w:d	92-129 µs	0.20	0.40	0.30	0.10	9.24
64-94 risings	0-25 w:d	92-129 µs	0.30	0.45	0.20	0.05	9.24
95+ risings	0-25 w:d	92-129 µs	0.55	0.35	0.09	**0.01**	9.24
0-40 risings	25-33 w:d	92-129 µs	0.05	0.25	0.40	0.30	9.24
41-63 risings	25-33 w:d	92-129 µs	0.10	0.30	0.40	0.20	9.24
64-94 risings	25-33 w:d	92-129 µs	0.20	0.40	0.30	0.10	9.24
95+ risings	25-33 w:d	92-129 µs	0.30	0.45	0.20	0.05	9.24
0-40 risings	33-44 w:d	92-129 µs	0.01	0.19	0.30	0.50	9.24
41-63 risings	33-44 w:d	92-129 µs	0.05	0.25	0.40	0.30	9.24
64-94 risings	33-44 w:d	92-129 µs	0.10	0.30	0.40	0.20	9.24
95+ risings	33-44 w:d	92-129 µs	0.20	0.40	0.30	0.10	9.24
0-40 risings	>44 w:d	92-129 µs	0.01	0.05	0.24	0.70	9.24
41-63 risings	>44 w:d	92-129 µs	0.01	0.19	0.30	0.50	9.24
64-94 risings	>44 w:d	92-129 µs	0.05	0.25	0.40	0.30	9.24
95+ risings	>44 w:d	92-129 µs	0.10	0.30	0.40	0.20	9.24

Table 68 continued: Prior conditional probability table and prior weight (α_0) for "EPT taxa richness" node (EPT taxa bins in probability units and α_0 in equivalent data points).

Flashiness	Channel ratio	Conductivity	EPT taxa richness				
			0-4 taxa	5-7 taxa	8-12 taxa	>13 taxa	α_0
0-40 risings	0-25 w:d	129-281 µs	0.20	0.40	0.30	0.10	9.24
41-63 risings	0-25 w:d	129-281 µs	0.30	0.45	0.20	0.05	9.24
64-94 risings	0-25 w:d	129-281 µs	0.55	0.35	0.09	0.01	9.24
95+ risings	0-25 w:d	129-281 µs	0.65	0.30	0.04	0.01	9.24
0-40 risings	25-33 w:d	129-281 µs	0.10	0.30	0.40	0.20	9.24
41-63 risings	25-33 w:d	129-281 µs	0.20	0.40	0.30	0.10	9.24
64-94 risings	25-33 w:d	129-281 µs	0.30	0.45	0.20	0.05	9.24
95+ risings	25-33 w:d	129-281 µs	0.55	0.35	0.09	0.01	9.24
0-40 risings	33-44 w:d	129-281 µs	0.05	0.25	0.40	0.30	9.24
41-63 risings	33-44 w:d	129-281 µs	0.10	0.30	0.40	0.20	9.24
64-94 risings	33-44 w:d	129-281 µs	0.20	0.40	0.30	0.10	9.24
95+ risings	33-44 w:d	129-281 µs	0.30	0.45	0.20	0.05	9.24
0-40 risings	>44 w:d	129-281 µs	0.01	0.19	0.30	0.50	9.24
41-63 risings	>44 w:d	129-281 µs	0.05	0.25	0.40	0.30	9.24
64-94 risings	>44 w:d	129-281 µs	0.10	0.30	0.40	0.20	9.24
95+ risings	>44 w:d	129-281 µs	0.20	0.40	0.30	0.10	9.24
0-40 risings	0-25 w:d	>281 µs	0.30	0.45	0.20	0.05	9.24
41-63 risings	0-25 w:d	>281 µs	0.55	0.35	0.09	0.01	9.24
64-94 risings	0-25 w:d	>281 µs	**0.65**	0.30	0.04	0.01	9.24
95+ risings	0-25 w:d	>281 µs	0.80	0.18	0.01	0.01	9.24
0-40 risings	25-33 w:d	>281 µs	0.20	0.40	0.30	0.10	9.24
41-63 risings	25-33 w:d	>281 µs	0.30	0.45	0.20	0.05	9.24
64-94 risings	25-33 w:d	>281 µs	0.55	0.35	0.09	0.01	9.24
95+ risings	25-33 w:d	>281 µs	0.65	0.30	0.04	0.01	9.24
0-40 risings	33-44 w:d	>281 µs	0.10	0.30	0.40	0.20	9.24
41-63 risings	33-44 w:d	>281 µs	0.20	0.40	0.30	0.10	9.24
64-94 risings	33-44 w:d	>281 µs	0.30	0.45	0.20	0.05	9.24
95+ risings	33-44 w:d	>281 µs	0.55	0.35	0.09	0.01	9.24
0-40 risings	>44 w:d	>281 µs	0.05	0.25	0.40	0.30	9.24
41-63 risings	>44 w:d	>281 µs	0.10	0.30	0.40	0.20	9.24
64-94 risings	>44 w:d	>281 µs	0.20	0.40	0.30	0.10	9.24
95+ risings	>44 w:d	>281 µs	0.30	0.45	0.20	0.05	9.24

Table 69: Prior conditional probability table and prior weight (α_0) for "Richness-weighted tolerance" node (tolerance bins in probability units and α_0 in equivalent data points).

Flashiness	Channel ratio	Conductivity	Richness-weighted tolerance				
			<5 tol.	5-5.6 tol.	5.6-6.2 tol.	>6.2 tol.	α_0
0-40 risings	0-25 w:d	0-92 µs	0.30	0.40	0.25	0.05	7.66
41-63 risings	0-25 w:d	0-92 µs	0.20	0.40	0.30	0.10	7.66
64-94 risings	0-25 w:d	0-92 µs	0.10	0.30	0.40	0.20	7.66
95+ risings	0-25 w:d	0-92 µs	0.05	0.20	0.45	0.30	7.66
0-40 risings	25-33 w:d	0-92 µs	0.50	0.30	0.19	0.01	7.66
41-63 risings	25-33 w:d	0-92 µs	0.30	0.40	0.25	0.05	7.66
64-94 risings	25-33 w:d	0-92 µs	0.20	0.40	0.30	0.10	7.66
95+ risings	25-33 w:d	0-92 µs	0.10	0.30	0.40	0.20	7.66
0-40 risings	33-44 w:d	0-92 µs	0.70	0.24	0.05	0.01	7.66
41-63 risings	33-44 w:d	0-92 µs	0.50	0.30	0.19	0.01	7.66
64-94 risings	33-44 w:d	0-92 µs	0.30	0.40	0.25	0.05	7.66
95+ risings	33-44 w:d	0-92 µs	0.20	0.40	0.30	0.10	7.66
0-40 risings	>44 w:d	0-92 µs	**0.88**	0.10	0.01	0.01	7.66
41-63 risings	>44 w:d	0-92 µs	0.70	0.24	0.05	0.01	7.66
64-94 risings	>44 w:d	0-92 µs	0.50	0.30	0.19	0.01	7.66
95+ risings	>44 w:d	0-92 µs	0.30	0.40	0.25	0.05	7.66
0-40 risings	0-25 w:d	92-129 µs	0.20	0.40	0.30	0.10	7.66
41-63 risings	0-25 w:d	92-129 µs	0.10	0.30	0.40	0.20	7.66
64-94 risings	0-25 w:d	92-129 µs	0.05	0.20	0.45	0.30	7.66
95+ risings	0-25 w:d	92-129 µs	0.01	0.09	0.35	0.55	7.66
0-40 risings	25-33 w:d	92-129 µs	0.30	0.40	0.25	0.05	7.66
41-63 risings	25-33 w:d	92-129 µs	0.20	0.40	0.30	0.10	7.66
64-94 risings	25-33 w:d	92-129 µs	0.10	0.30	0.40	0.20	7.66
95+ risings	25-33 w:d	92-129 µs	0.05	0.20	0.45	0.30	7.66
0-40 risings	33-44 w:d	92-129 µs	0.50	0.30	0.19	0.01	7.66
41-63 risings	33-44 w:d	92-129 µs	0.30	0.40	0.25	0.05	7.66
64-94 risings	33-44 w:d	92-129 µs	0.20	0.40	0.30	0.10	7.66
95+ risings	33-44 w:d	92-129 µs	0.10	0.30	0.40	0.20	7.66
0-40 risings	>44 w:d	92-129 µs	0.70	0.24	0.05	0.01	7.66
41-63 risings	>44 w:d	92-129 µs	0.50	0.30	0.19	0.01	7.66
64-94 risings	>44 w:d	92-129 µs	0.30	0.40	0.25	0.05	7.66
95+ risings	>44 w:d	92-129 µs	0.20	0.40	0.30	0.10	7.66

Table 69 continued: Prior conditional probability table and prior weight (α_0) for "Richness-weighted tolerance" node (tolerance bins in probability units and α_0 in equivalent data points).

Flashiness	Channel ratio	Conductivity	Richness-weighted tolerance				
			<5 tol.	5-5.6 tol.	5.6-6.2 tol.	>6.2 tol.	α_0
0-40 risings	0-25 w:d	129-281 µs	0.10	0.30	0.40	0.20	7.66
41-63 risings	0-25 w:d	129-281 µs	0.05	0.20	0.45	0.30	7.66
64-94 risings	0-25 w:d	129-281 µs	0.01	0.09	0.35	0.55	7.66
95+ risings	0-25 w:d	129-281 µs	0.01	0.04	0.30	0.65	7.66
0-40 risings	25-33 w:d	129-281 µs	0.20	0.40	0.30	0.10	7.66
41-63 risings	25-33 w:d	129-281 µs	0.10	0.30	0.40	0.20	7.66
64-94 risings	25-33 w:d	129-281 µs	0.05	0.20	0.45	0.30	7.66
95+ risings	25-33 w:d	129-281 µs	0.01	0.09	0.35	0.55	7.66
0-40 risings	33-44 w:d	129-281 µs	0.30	0.40	0.25	0.05	7.66
41-63 risings	33-44 w:d	129-281 µs	0.20	0.40	0.30	0.10	7.66
64-94 risings	33-44 w:d	129-281 µs	0.10	0.30	0.40	0.20	7.66
95+ risings	33-44 w:d	129-281 µs	0.05	0.20	0.45	0.30	7.66
0-40 risings	>44 w:d	129-281 µs	0.50	0.30	0.19	0.01	7.66
41-63 risings	>44 w:d	129-281 µs	0.30	0.40	0.25	0.05	7.66
64-94 risings	>44 w:d	129-281 µs	0.20	0.40	0.30	0.10	7.66
95+ risings	>44 w:d	129-281 µs	0.10	0.30	0.40	0.20	7.66
0-40 risings	0-25 w:d	>281 µs	0.05	0.20	0.45	0.30	7.66
41-63 risings	0-25 w:d	>281 µs	0.01	0.09	0.35	0.55	7.66
64-94 risings	0-25 w:d	>281 µs	0.01	0.04	0.30	0.65	7.66
95+ risings	0-25 w:d	>281 µs	0.01	0.01	0.18	0.80	7.66
0-40 risings	25-33 w:d	>281 µs	0.10	0.30	0.40	0.20	7.66
41-63 risings	25-33 w:d	>281 µs	0.05	0.20	0.45	0.30	7.66
64-94 risings	25-33 w:d	>281 µs	0.01	0.09	0.35	0.55	7.66
95+ risings	25-33 w:d	>281 µs	0.01	0.04	0.30	0.65	7.66
0-40 risings	33-44 w:d	>281 µs	0.20	0.40	0.30	0.10	7.66
41-63 risings	33-44 w:d	>281 µs	0.10	0.30	0.40	0.20	7.66
64-94 risings	33-44 w:d	>281 µs	0.05	0.20	0.45	0.30	7.66
95+ risings	33-44 w:d	>281 µs	0.01	0.09	0.35	0.55	7.66
0-40 risings	>44 w:d	>281 µs	0.30	0.40	0.25	0.05	7.66
41-63 risings	>44 w:d	>281 µs	0.20	0.40	0.30	0.10	7.66
64-94 risings	>44 w:d	>281 µs	0.10	**0.30**	0.40	0.20	7.66
95+ risings	>44 w:d	>281 µs	0.05	0.20	0.45	0.30	7.66

Table 70: Prior conditional probability table and prior weight (α_0) for "Percent intolerant taxa" node (intolerant taxa bins in probability units and α_0 in equivalent data points).

Flashiness	Channel ratio	Conductivity	Percent intolerant taxa				
			0-3%	3-8%	8-18%	>18%	α_0
0-40 risings	0-25 w:d	0-92 µs	0.05	0.25	0.40	0.30	8.23
41-63 risings	0-25 w:d	0-92 µs	0.10	0.30	0.40	0.20	8.23
64-94 risings	0-25 w:d	0-92 µs	0.20	0.40	0.30	0.10	8.23
95+ risings	0-25 w:d	0-92 µs	0.30	0.45	0.20	0.05	8.23
0-40 risings	25-33 w:d	0-92 µs	0.01	0.19	0.30	0.50	8.23
41-63 risings	25-33 w:d	0-92 µs	0.05	0.25	0.40	0.30	8.23
64-94 risings	25-33 w:d	0-92 µs	0.10	0.30	0.40	0.20	8.23
95+ risings	25-33 w:d	0-92 µs	0.20	0.40	0.30	0.10	8.23
0-40 risings	33-44 w:d	0-92 µs	0.01	0.05	0.24	0.70	8.23
41-63 risings	33-44 w:d	0-92 µs	0.01	0.19	0.30	0.50	8.23
64-94 risings	33-44 w:d	0-92 µs	0.05	0.25	0.40	0.30	8.23
95+ risings	33-44 w:d	0-92 µs	0.10	0.30	0.40	0.20	8.23
0-40 risings	>44 w:d	0-92 µs	0.01	0.01	0.10	0.88	8.23
41-63 risings	>44 w:d	0-92 µs	0.01	0.05	0.24	0.70	8.23
64-94 risings	>44 w:d	0-92 µs	0.01	0.19	0.30	0.50	8.23
95+ risings	>44 w:d	0-92 µs	0.05	0.25	0.40	0.30	8.23
0-40 risings	0-25 w:d	92-129 µs	0.10	0.30	0.40	0.20	8.23
41-63 risings	0-25 w:d	92-129 µs	0.20	0.40	0.30	0.10	8.23
64-94 risings	0-25 w:d	92-129 µs	0.30	0.45	0.20	0.05	8.23
95+ risings	0-25 w:d	92-129 µs	0.55	0.35	0.09	0.01	8.23
0-40 risings	25-33 w:d	92-129 µs	0.05	0.25	0.40	0.30	8.23
41-63 risings	25-33 w:d	92-129 µs	0.10	0.30	0.40	0.20	8.23
64-94 risings	25-33 w:d	92-129 µs	0.20	0.40	0.30	0.10	8.23
95+ risings	25-33 w:d	92-129 µs	0.30	0.45	0.20	0.05	8.23
0-40 risings	33-44 w:d	92-129 µs	0.01	0.19	0.30	0.50	8.23
41-63 risings	33-44 w:d	92-129 µs	0.05	0.25	0.40	0.30	8.23
64-94 risings	33-44 w:d	92-129 µs	0.10	0.30	0.40	0.20	8.23
95+ risings	33-44 w:d	92-129 µs	0.20	0.40	0.30	0.10	8.23
0-40 risings	>44 w:d	92-129 µs	0.01	0.05	0.24	**0.70**	8.23
41-63 risings	>44 w:d	92-129 µs	0.01	0.19	0.30	0.50	8.23
64-94 risings	>44 w:d	92-129 µs	0.05	0.25	0.40	0.30	8.23
95+ risings	>44 w:d	92-129 µs	0.10	0.30	0.40	0.20	8.23
0-40 risings	0-25 w:d	129-281 µs	0.20	0.40	0.30	0.10	8.23

Table 70 continued: Prior conditional probability table and prior weight (α_0) for "Percent intolerant taxa" node (intolerant taxa bins in probability units and α_0 in equivalent data points).

Flashiness	Channel ratio	Conductivity	Percent intolerant taxa				
			0-3%	3-8%	8-18%	>18%	α_0
41-63 risings	0-25 w:d	129-281 µs	0.30	0.45	0.20	0.05	8.23
64-94 risings	0-25 w:d	129-281 µs	0.55	0.35	0.09	0.01	8.23
95+ risings	0-25 w:d	129-281 µs	0.65	0.30	0.04	0.01	8.23
0-40 risings	25-33 w:d	129-281 µs	0.10	0.30	0.40	0.20	8.23
41-63 risings	25-33 w:d	129-281 µs	0.20	0.40	0.30	0.10	8.23
64-94 risings	25-33 w:d	129-281 µs	0.30	0.45	0.20	0.05	8.23
95+ risings	25-33 w:d	129-281 µs	0.55	0.35	0.09	0.01	8.23
0-40 risings	33-44 w:d	129-281 µs	0.05	0.25	0.40	0.30	8.23
41-63 risings	33-44 w:d	129-281 µs	0.10	0.30	0.40	0.20	8.23
64-94 risings	33-44 w:d	129-281 µs	0.20	0.40	0.30	0.10	8.23
95+ risings	33-44 w:d	129-281 µs	0.30	0.45	0.20	0.05	8.23
0-40 risings	>44 w:d	129-281 µs	0.01	0.19	0.30	0.50	8.23
41-63 risings	>44 w:d	129-281 µs	0.05	0.25	0.40	0.30	8.23
64-94 risings	>44 w:d	129-281 µs	0.10	0.30	0.40	0.20	8.23
95+ risings	>44 w:d	129-281 µs	0.20	0.40	0.30	0.10	8.23
0-40 risings	0-25 w:d	>281 µs	0.30	0.45	0.20	0.05	8.23
41-63 risings	0-25 w:d	>281 µs	0.55	0.35	0.09	0.01	8.23
64-94 risings	0-25 w:d	>281 µs	0.65	0.30	0.04	0.01	8.23
95+ risings	0-25 w:d	>281 µs	0.80	0.18	0.01	0.01	8.23
0-40 risings	25-33 w:d	>281 µs	0.20	0.40	0.30	0.10	8.23
41-63 risings	25-33 w:d	>281 µs	0.30	0.45	0.20	0.05	8.23
64-94 risings	25-33 w:d	>281 µs	0.55	0.35	0.09	0.01	8.23
95+ risings	25-33 w:d	>281 µs	0.65	0.30	0.04	0.01	8.23
0-40 risings	33-44 w:d	>281 µs	0.10	0.30	0.40	0.20	8.23
41-63 risings	33-44 w:d	>281 µs	0.20	0.40	0.30	0.10	8.23
64-94 risings	33-44 w:d	>281 µs	0.30	0.45	0.20	0.05	8.23
95+ risings	33-44 w:d	>281 µs	0.55	0.35	**0.09**	0.01	8.23
0-40 risings	>44 w:d	>281 µs	0.05	0.25	0.40	0.30	8.23
41-63 risings	>44 w:d	>281 µs	0.10	0.30	0.40	0.20	8.23
64-94 risings	>44 w:d	>281 µs	0.20	0.40	0.30	0.10	8.23
95+ risings	>44 w:d	>281 µs	0.30	0.45	0.20	0.05	8.23

Table 71: Data table for "Urban disturbance" node (urban land cover counts and *n* total samples).

Urban land cover				
0-12%	**>12-36%**	**>36-66%**	**>66%**	***n***
21	22	21	21	85

Table 72: Data table for "Hydrologic modifications" node (dam density counts and *n* total number of samples).

Dam density			
0 dams/100 km²	**>0-6 dams/100 km²**	**>6 dams/100 km²**	***n***
35	23	27	85

Table 73: Data table for "Change in generation of flow" node (impervious surface counts and *n* total number of samples per parent state).

Urban land cover	Impervious surface				
	0-2 %	**>2-8%**	**>8-16%**	**>16%**	***n***
0-12%	19	2	0	0	21
>12-36%	2	18	2	0	22
>36-66%	0	2	14	4	20
>66%	0	0	5	17	22

Table 74: Data table for "Habitat" node (channel ratio counts and *n* total number of samples per parent state combination).

Urban land cover	Flashiness	Channel width:depth ratio				
		0-25 w:d	>25-33 w:d	>33-44 w:d	>44 w:d	*n*
0-12%	0-40 risings	1	2	2	4	9
>12-36%	0-40 risings	2	2	2	1	7
>36-66%	0-40 risings	2	0	0	1	3
>66%	0-40 risings	0	0	0	3	3
0-12%	41-63 risings	2	3	4	1	10
>12-36%	41-63 risings	0	1	1	1	3
>36-66%	41-63 risings	1	0	0	4	5
>66%	41-63 risings	0	1	0	2	3
0-12%	64-94 risings	0	1	1	0	2
>12-36%	64-94 risings	3	5	1	1	10
>36-66%	64-94 risings	0	0	1	3	4
>66%	64-94 risings	2	2	0	1	5
0-12%	95+ risings	0	0	0	0	0
>12-36%	95+ risings	1	1	0	0	2
>36-66%	95+ risings	6	0	1	1	8
>66%	95+ risings	2	4	3	2	11

Table 75: Data table for "Water Quality" node (conductivity counts and *n* total number of samples per parent state combination).

Urban land cover	Flashiness	Conductivity at low base flow				
		0-92 μs	92-129 μs	129-281 μs	>281 μs	*n*
0-12%	0-40 risings	1	2	5	1	9
>12-36%	0-40 risings	0	0	1	6	7
>36-66%	0-40 risings	0	0	0	3	3
>66%	0-40 risings	0	0	0	3	3
0-12%	41-63 risings	8	2	0	0	10
>12-36%	41-63 risings	1	2	0	0	3
>36-66%	41-63 risings	1	0	1	3	5
>66%	41-63 risings	0	1	1	1	3
0-12%	64-94 risings	2	0	0	0	2
>12-36%	64-94 risings	5	4	1	0	10
>36-66%	64-94 risings	0	1	2	1	4
>66%	64-94 risings	0	2	1	2	5
0-12%	95+ risings	0	0	0	0	0
>12-36%	95+ risings	1	1	0	0	2
>36-66%	95+ risings	2	3	3	0	8
>66%	95+ risings	0	4	5	2	11

Table 76: Data table for "EPT taxa richness" node (EPT taxa counts and *n* total number of samples per parent state combination).

Flashiness	Channel ratio	Conductivity	EPT taxa richness				
			0-4 taxa	5-7 taxa	8-12 taxa	>13 taxa	*n*
0-40 risings	0-25 w:d	0-92 µs	0	0	0	0	0
41-63 risings	0-25 w:d	0-92 µs	0	1	1	1	3
64-94 risings	0-25 w:d	0-92 µs	0	0	1	0	1
95+ risings	0-25 w:d	0-92 µs	0	1	1	0	2
0-40 risings	25-33 w:d	0-92 µs	0	0	0	1	1
41-63 risings	25-33 w:d	0-92 µs	0	0	2	0	2
64-94 risings	25-33 w:d	0-92 µs	1	1	0	3	5
95+ risings	25-33 w:d	0-92 µs	0	0	0	1	1
0-40 risings	33-44 w:d	0-92 µs	0	0	0	0	0
41-63 risings	33-44 w:d	0-92 µs	0	0	3	1	4
64-94 risings	33-44 w:d	0-92 µs	0	0	0	1	1
95+ risings	33-44 w:d	0-92 µs	0	0	0	0	0
0-40 risings	>44 w:d	0-92 µs	0	0	0	0	0
41-63 risings	>44 w:d	0-92 µs	0	0	0	1	1
64-94 risings	>44 w:d	0-92 µs	0	0	1	0	1
95+ risings	>44 w:d	0-92 µs	0	0	0	0	0
0-40 risings	0-25 w:d	92-129 µs	0	0	0	0	0
41-63 risings	0-25 w:d	92-129 µs	0	0	0	0	0
64-94 risings	0-25 w:d	92-129 µs	1	0	2	0	3
95+ risings	0-25 w:d	92-129 µs	1	1	1	0	3
0-40 risings	25-33 w:d	92-129 µs	0	0	1	0	1
41-63 risings	25-33 w:d	92-129 µs	0	1	0	2	3
64-94 risings	25-33 w:d	92-129 µs	0	1	0	0	1
95+ risings	25-33 w:d	92-129 µs	0	2	0	0	2
0-40 risings	33-44 w:d	92-129 µs	0	0	1	0	1
41-63 risings	33-44 w:d	92-129 µs	0	0	1	0	1
64-94 risings	33-44 w:d	92-129 µs	0	0	1	0	1
95+ risings	33-44 w:d	92-129 µs	1	1	0	0	2
0-40 risings	>44 w:d	92-129 µs	0	0	0	0	0
41-63 risings	>44 w:d	92-129 µs	0	1	0	0	1
64-94 risings	>44 w:d	92-129 µs	1	0	0	0	1
95+ risings	>44 w:d	92-129 µs	1	0	0	0	1
0-40 risings	0-25 w:d	129-281 µs	0	0	0	1	1

Table 76 continued: Data table for "EPT taxa richness" node (EPT taxa counts and *n* total number of samples per parent state combination).

Flashiness	Channel ratio	Conductivity	EPT taxa richness				
			0-4 taxa	5-7 taxa	8-12 taxa	>13 taxa	*n*
41-63 risings	0-25 w:d	129-281 µs	0	0	0	0	0
64-94 risings	0-25 w:d	129-281 µs	0	0	0	0	0
95+ risings	0-25 w:d	129-281 µs	1	2	0	0	3
0-40 risings	25-33 w:d	129-281 µs	0	0	0	0	0
41-63 risings	25-33 w:d	129-281 µs	0	0	0	0	0
64-94 risings	25-33 w:d	129-281 µs	1	0	0	0	1
95+ risings	25-33 w:d	129-281 µs	1	0	0	0	1
0-40 risings	33-44 w:d	129-281 µs	0	0	1	1	2
41-63 risings	33-44 w:d	129-281 µs	0	0	0	0	0
64-94 risings	33-44 w:d	129-281 µs	1	0	0	0	1
95+ risings	33-44 w:d	129-281 µs	2	0	0	0	2
0-40 risings	>44 w:d	129-281 µs	0	0	2	1	3
41-63 risings	>44 w:d	129-281 µs	0	1	1	0	2
64-94 risings	>44 w:d	129-281 µs	1	1	0	0	2
95+ risings	>44 w:d	129-281 µs	1	1	0	0	2
0-40 risings	0-25 w:d	>281 µs	2	0	2	0	4
41-63 risings	0-25 w:d	>281 µs	0	0	0	0	0
64-94 risings	0-25 w:d	>281 µs	1	0	0	0	1
95+ risings	0-25 w:d	>281 µs	0	1	0	0	1
0-40 risings	25-33 w:d	>281 µs	0	1	1	0	2
41-63 risings	25-33 w:d	>281 µs	0	0	0	0	0
64-94 risings	25-33 w:d	>281 µs	1	0	0	0	1
95+ risings	25-33 w:d	>281 µs	1	0	0	0	1
0-40 risings	33-44 w:d	>281 µs	1	0	0	0	1
41-63 risings	33-44 w:d	>281 µs	0	0	0	0	0
64-94 risings	33-44 w:d	>281 µs	0	0	0	0	0
95+ risings	33-44 w:d	>281 µs	0	0	0	0	0
0-40 risings	>44 w:d	>281 µs	2	0	1	3	6
41-63 risings	>44 w:d	>281 µs	0	3	1	0	4
64-94 risings	>44 w:d	>281 µs	0	1	0	0	1
95+ risings	>44 w:d	>281 µs	0	0	0	0	0

Table 77: Data table for "Richness-weighted tolerance" node (tolerance counts and *n* total number of samples per parent state combination).

Flashiness	Channel ratio	Conductivity	Richness-weighted tolerance				
			<5 tol.	5-5.6 tol.	5.6-6.2 tol.	>6.2 tol.	*n*
0-40 risings	0-25 w:d	0-92 µs	0	0	0	0	0
41-63 risings	0-25 w:d	0-92 µs	0	2	1	0	3
64-94 risings	0-25 w:d	0-92 µs	0	0	1	0	1
95+ risings	0-25 w:d	0-92 µs	0	0	2	0	2
0-40 risings	25-33 w:d	0-92 µs	1	0	0	0	1
41-63 risings	25-33 w:d	0-92 µs	2	0	0	0	2
64-94 risings	25-33 w:d	0-92 µs	1	1	2	1	5
95+ risings	25-33 w:d	0-92 µs	0	1	0	0	1
0-40 risings	33-44 w:d	0-92 µs	0	0	0	0	0
41-63 risings	33-44 w:d	0-92 µs	1	3	0	0	4
64-94 risings	33-44 w:d	0-92 µs	0	1	0	0	1
95+ risings	33-44 w:d	0-92 µs	0	0	0	0	0
0-40 risings	>44 w:d	0-92 µs	0	0	0	0	0
41-63 risings	>44 w:d	0-92 µs	0	1	0	0	1
64-94 risings	>44 w:d	0-92 µs	0	1	0	0	1
95+ risings	>44 w:d	0-92 µs	0	0	0	0	0
0-40 risings	0-25 w:d	92-129 µs	0	0	0	0	0
41-63 risings	0-25 w:d	92-129 µs	0	0	0	0	0
64-94 risings	0-25 w:d	92-129 µs	0	1	1	1	3
95+ risings	0-25 w:d	92-129 µs	0	0	2	1	3
0-40 risings	25-33 w:d	92-129 µs	0	1	0	0	1
41-63 risings	25-33 w:d	92-129 µs	0	1	2	0	3
64-94 risings	25-33 w:d	92-129 µs	0	0	1	0	1
95+ risings	25-33 w:d	92-129 µs	0	0	1	1	2
0-40 risings	33-44 w:d	92-129 µs	0	0	1	0	1
41-63 risings	33-44 w:d	92-129 µs	0	1	0	0	1
64-94 risings	33-44 w:d	92-129 µs	0	1	0	0	1
95+ risings	33-44 w:d	92-129 µs	0	0	0	2	2
0-40 risings	>44 w:d	92-129 µs	0	0	0	0	0
41-63 risings	>44 w:d	92-129 µs	0	1	0	0	1
64-94 risings	>44 w:d	92-129 µs	0	0	0	1	1
95+ risings	>44 w:d	92-129 µs	0	0	0	1	1
0-40 risings	0-25 w:d	129-281 µs	1	0	0	0	1

Table 77 continued: Data table for "Richness-weighted tolerance" node (tolerance counts and *n* total number of samples per parent state combination).

Flashiness	Channel ratio	Conductivity	Richness-weighted tolerance				
			<5 tol.	5-5.6 tol.	5.6-6.2 tol.	>6.2 tol.	*n*
41-63 risings	0-25 w:d	129-281 µs	0	0	0	0	0
64-94 risings	0-25 w:d	129-281 µs	0	0	0	0	0
95+ risings	0-25 w:d	129-281 µs	0	1	1	1	3
0-40 risings	25-33 w:d	129-281 µs	0	0	0	0	0
41-63 risings	25-33 w:d	129-281 µs	0	0	0	0	0
64-94 risings	25-33 w:d	129-281 µs	0	0	0	1	1
95+ risings	25-33 w:d	129-281 µs	0	0	1	0	1
0-40 risings	33-44 w:d	129-281 µs	2	0	0	0	2
41-63 risings	33-44 w:d	129-281 µs	0	0	0	0	0
64-94 risings	33-44 w:d	129-281 µs	0	0	0	1	1
95+ risings	33-44 w:d	129-281 µs	0	0	0	2	2
0-40 risings	>44 w:d	129-281 µs	0	3	0	0	3
41-63 risings	>44 w:d	129-281 µs	0	1	1	0	2
64-94 risings	>44 w:d	129-281 µs	0	0	0	2	2
95+ risings	>44 w:d	129-281 µs	0	0	1	1	2
0-40 risings	0-25 w:d	>281 µs	0	4	0	0	4
41-63 risings	0-25 w:d	>281 µs	0	0	0	0	0
64-94 risings	0-25 w:d	>281 µs	0	0	1	0	1
95+ risings	0-25 w:d	>281 µs	0	0	1	0	1
0-40 risings	25-33 w:d	>281 µs	1	1	0	0	2
41-63 risings	25-33 w:d	>281 µs	0	0	0	0	0
64-94 risings	25-33 w:d	>281 µs	0	0	0	1	1
95+ risings	25-33 w:d	>281 µs	0	0	0	1	1
0-40 risings	33-44 w:d	>281 µs	0	1	0	0	1
41-63 risings	33-44 w:d	>281 µs	0	0	0	0	0
64-94 risings	33-44 w:d	>281 µs	0	0	0	0	0
95+ risings	33-44 w:d	>281 µs	0	0	0	0	0
0-40 risings	>44 w:d	>281 µs	3	1	0	2	6
41-63 risings	>44 w:d	>281 µs	1	0	3	0	4
64-94 risings	>44 w:d	>281 µs	0	0	0	1	1
95+ risings	>44 w:d	>281 µs	0	0	0	0	0

Table 78: Data table for "Percent intolerant taxa" node (intolerant taxa counts and *n* total number of samples per parent state combination).

Flashiness	Channel ratio	Conductivity	Percent intolerant taxa				
			0-3%	3-8%	8-18%	>18%	*n*
0-40 risings	0-25 w:d	0-92 µs	0	0	0	0	0
41-63 risings	0-25 w:d	0-92 µs	1	0	1	1	3
64-94 risings	0-25 w:d	0-92 µs	0	1	0	0	1
95+ risings	0-25 w:d	0-92 µs	1	1	0	0	2
0-40 risings	25-33 w:d	0-92 µs	0	0	0	1	1
41-63 risings	25-33 w:d	0-92 µs	0	0	1	1	2
64-94 risings	25-33 w:d	0-92 µs	1	2	2	0	5
95+ risings	25-33 w:d	0-92 µs	0	0	1	0	1
0-40 risings	33-44 w:d	0-92 µs	0	0	0	0	0
41-63 risings	33-44 w:d	0-92 µs	0	2	0	2	4
64-94 risings	33-44 w:d	0-92 µs	0	0	1	0	1
95+ risings	33-44 w:d	0-92 µs	0	0	0	0	0
0-40 risings	>44 w:d	0-92 µs	0	0	0	0	0
41-63 risings	>44 w:d	0-92 µs	0	0	1	0	1
64-94 risings	>44 w:d	0-92 µs	0	1	0	0	1
95+ risings	>44 w:d	0-92 µs	0	0	0	0	0
0-40 risings	0-25 w:d	92-129 µs	0	0	0	0	0
41-63 risings	0-25 w:d	92-129 µs	0	0	0	0	0
64-94 risings	0-25 w:d	92-129 µs	2	0	1	0	3
95+ risings	0-25 w:d	92-129 µs	1	2	0	0	3
0-40 risings	25-33 w:d	92-129 µs	0	0	0	1	1
41-63 risings	25-33 w:d	92-129 µs	0	0	2	1	3
64-94 risings	25-33 w:d	92-129 µs	1	0	0	0	1
95+ risings	25-33 w:d	92-129 µs	1	1	0	0	2
0-40 risings	33-44 w:d	92-129 µs	0	0	1	0	1
41-63 risings	33-44 w:d	92-129 µs	0	0	0	1	1
64-94 risings	33-44 w:d	92-129 µs	0	0	1	0	1
95+ risings	33-44 w:d	92-129 µs	0	2	0	0	2
0-40 risings	>44 w:d	92-129 µs	0	0	0	0	0
41-63 risings	>44 w:d	92-129 µs	0	0	1	0	1
64-94 risings	>44 w:d	92-129 µs	1	0	0	0	1
95+ risings	>44 w:d	92-129 µs	1	0	0	0	1
0-40 risings	0-25 w:d	129-281 µs	0	0	0	1	1
41-63 risings	0-25 w:d	129-281 µs	0	0	0	0	0

Table 78 continued: Data table for "Percent intolerant taxa" node (intolerant taxa counts and *n* total number of samples per parent state combination).

Flashiness	Channel ratio	Conductivity	Percent intolerant taxa				
			0-3%	3-8%	8-18%	>18%	*n*
64-94 risings	0-25 w:d	129-281 µs	0	0	0	0	0
95+ risings	0-25 w:d	129-281 µs	1	2	0	0	3
0-40 risings	25-33 w:d	129-281 µs	0	0	0	0	0
41-63 risings	25-33 w:d	129-281 µs	0	0	0	0	0
64-94 risings	25-33 w:d	129-281 µs	1	0	0	0	1
95+ risings	25-33 w:d	129-281 µs	0	1	0	0	1
0-40 risings	33-44 w:d	129-281 µs	0	0	0	2	2
41-63 risings	33-44 w:d	129-281 µs	0	0	0	0	0
64-94 risings	33-44 w:d	129-281 µs	1	0	0	0	1
95+ risings	33-44 w:d	129-281 µs	1	1	0	0	2
0-40 risings	>44 w:d	129-281 µs	0	0	2	1	3
41-63 risings	>44 w:d	129-281 µs	0	1	0	1	2
64-94 risings	>44 w:d	129-281 µs	1	1	0	0	2
95+ risings	>44 w:d	129-281 µs	1	0	1	0	2
0-40 risings	0-25 w:d	>281 µs	0	0	1	3	4
41-63 risings	0-25 w:d	>281 µs	0	0	0	0	0
64-94 risings	0-25 w:d	>281 µs	0	1	0	0	1
95+ risings	0-25 w:d	>281 µs	0	1	0	0	1
0-40 risings	25-33 w:d	>281 µs	0	0	0	2	2
41-63 risings	25-33 w:d	>281 µs	0	0	0	0	0
64-94 risings	25-33 w:d	>281 µs	1	0	0	0	1
95+ risings	25-33 w:d	>281 µs	1	0	0	0	1
0-40 risings	33-44 w:d	>281 µs	0	0	1	0	1
41-63 risings	33-44 w:d	>281 µs	0	0	0	0	0
64-94 risings	33-44 w:d	>281 µs	0	0	0	0	0
95+ risings	33-44 w:d	>281 µs	0	0	0	0	0
0-40 risings	>44 w:d	>281 µs	0	2	0	4	6
41-63 risings	>44 w:d	>281 µs	0	1	2	1	4
64-94 risings	>44 w:d	>281 µs	1	0	0	0	1
95+ risings	>44 w:d	>281 µs	0	0	0	0	0

Table 79: Posterior marginal probability table and posterior weights (α_0+n) for "Urban disturbance" node (urban land cover bins in probability units and α_0+n in equivalent data points).

Urban land cover				
0-12%	**>12-36%**	**>36-66%**	**>66%**	α_0+n
0.2471	0.2587	0.2355	0.2587	86

Table 80: Posterior marginal probability table and posterior weights (α_0+n) for "Hydrologic modifications" node (dam density bins in probability units and α_0+n in equivalent data points).

Dam density			
0 dams/100 km²	**>0-6 dams/100 km²**	**>6 dams/100 km²**	α_0+n
0.4116	0.2709	0.3174	86

Table 81: Posterior conditional probability table and posterior weights (α_0+n) for "Change in generation of flow" node (impervious surface bins in probability units and α_0+n in equivalent data points).

Urban land cover	Impervious surface				
	0-2 %	**>2-8%**	**>8-16%**	**>16%**	α_0+n
0-12%	0.8897	0.1048	0.0028	0.0028	28.99
>12-36%	0.0907	0.7601	0.1466	0.0027	29.99
>36-66%	0.0143	0.1143	0.7000	0.1715	27.99
>66%	0.0027	0.0240	0.1934	0.7800	29.99

Table 82: Posterior conditional probability table and posterior weights (α_0+n) for "Habitat" node (channel ratio bins in probability units and α_0+n in equivalent data points).

Urban land cover	Flashiness	Channel width:depth ratio				
		0-25 w:d	>25-33 w:d	>33-44 w:d	>44 w:d	α_0+n
0-12%	0-40 risings	0.0837	0.1675	0.2571	0.4917	16.3
>12-36%	0-40 risings	0.1654	0.2164	0.3951	0.2231	14.3
>36-66%	0-40 risings	0.2650	0.3189	0.2481	0.1680	10.3
>66%	0-40 risings	0.4252	0.1417	0.1063	0.3267	10.3
0-12%	41-63 risings	0.1578	0.2367	0.3789	0.2266	17.3
>12-36%	41-63 risings	0.0709	0.2388	0.5223	0.1680	10.3
>36-66%	41-63 risings	0.1703	0.2967	0.1780	0.3549	12.3
>66%	41-63 risings	0.4252	0.3097	0.0638	0.2013	10.3
0-12%	64-94 risings	0.1570	0.5000	0.3352	0.0078	9.3
>12-36%	64-94 risings	0.2789	0.5422	0.1169	0.0620	17.3
>36-66%	64-94 risings	0.3553	0.2519	0.1208	0.2719	11.3
>66%	64-94 risings	0.5424	0.3407	0.0297	0.0872	12.3
0-12%	95+ risings	0.2000	0.4000	0.3900	0.0100	7.3
>12-36%	95+ risings	0.3038	0.4608	0.2276	0.0078	9.3
>36-66%	95+ risings	0.5114	0.2386	0.1799	0.0701	15.3
>66%	95+ risings	0.3486	0.3542	0.1839	0.1133	18.3

Table 83: Posterior conditional probability table and posterior weights (α_0+n) for "Water Quality" node (conductivity bins in probability units and α_0+n in equivalent data points).

Urban land cover	Flashiness	Conductivity at low base flow				
		0-92 µs	92-129 µs	129-281 µs	>281 µs	α_0+n
0-12%	0-40 risings	0.4387	0.1641	0.3152	0.0821	17.16
>12-36%	0-40 risings	0.3768	0.1077	0.0929	0.4227	15.16
>36-66%	0-40 risings	0.3656	0.1828	0.1462	0.3054	11.16
>66%	0-40 risings	0.2194	0.2194	0.2194	0.3419	11.16
0-12%	41-63 risings	0.8000	0.1551	0.0225	0.0225	18.16
>12-36%	41-63 risings	0.6014	0.3254	0.0366	0.0366	11.16
>36-66%	41-63 risings	0.3860	0.1550	0.2000	0.2590	13.16
>66%	41-63 risings	0.2194	0.3090	0.3090	0.1627	11.16
0-12%	64-94 risings	0.7591	0.1205	0.0803	0.0402	10.16
>12-36%	64-94 risings	0.5674	0.3101	0.1000	0.0225	18.16
>36-66%	64-94 risings	0.2684	0.2836	0.2987	0.1493	12.16
>66%	64-94 risings	0.2000	0.2930	0.2930	0.2140	13.16
0-12%	95+ risings	0.7000	0.1500	0.1000	0.0500	8.16
>12-36%	95+ risings	0.6205	0.2591	0.0803	0.0402	10.16
>36-66%	95+ risings	0.3257	0.3371	0.2866	0.0505	16.16
>66%	95+ risings	0.0852	0.3578	0.4100	0.1470	19.16

Table 84: Posterior conditional probability table and posterior weights (α_0+n) for "EPT taxa richness" node (EPT taxa bins in probability units and α_0+n in equivalent data points).

Flashiness	Channel ratio	Conductivity	EPT taxa richness				
			0-4 taxa	5-7 taxa	8-12 taxa	>13 taxa	α_0+n
0-40 risings	0-25 w:d	0-92 µs	0.0500	0.2500	0.4000	0.3000	9.24
41-63 risings	0-25 w:d	0-92 µs	0.0755	0.3082	0.3837	0.2327	12.24
64-94 risings	0-25 w:d	0-92 µs	0.1805	0.3609	0.3684	0.0902	10.24
95+ risings	0-25 w:d	0-92 µs	0.2466	0.4589	0.2534	0.0411	11.24
0-40 risings	25-33 w:d	0-92 µs	0.0090	0.1714	0.2707	0.5488	10.24
41-63 risings	25-33 w:d	0-92 µs	0.0411	0.2055	0.5068	0.2466	11.24
64-94 risings	25-33 w:d	0-92 µs	0.1351	0.2649	0.2596	0.3404	14.24
95+ risings	25-33 w:d	0-92 µs	0.1805	0.3609	0.2707	0.1879	10.24
0-40 risings	33-44 w:d	0-92 µs	0.0100	0.0400	0.2400	0.7000	9.24
41-63 risings	33-44 w:d	0-92 µs	0.0070	0.1326	0.4360	0.4245	13.24
64-94 risings	33-44 w:d	0-92 µs	0.0451	0.2256	0.3609	0.3684	10.24
95+ risings	33-44 w:d	0-92 µs	0.1000	0.3000	0.4000	0.2000	9.24
0-40 risings	>44 w:d	0-92 µs	0.0100	0.0100	0.1000	0.8800	9.24
41-63 risings	>44 w:d	0-92 µs	0.0090	0.0451	0.2166	0.7293	10.24
64-94 risings	>44 w:d	0-92 µs	0.0090	0.1714	0.3684	0.4512	10.24
95+ risings	>44 w:d	0-92 µs	0.0500	0.2500	0.4000	0.3000	9.24
0-40 risings	0-25 w:d	92-129 µs	0.1000	0.3000	0.4000	0.2000	9.24
41-63 risings	0-25 w:d	92-129 µs	0.2000	0.4000	0.3000	0.1000	9.24
64-94 risings	0-25 w:d	92-129 µs	0.3082	0.3397	0.3144	0.0377	12.24
95+ risings	0-25 w:d	92-129 µs	0.4969	0.3459	0.1496	0.0075	12.24
0-40 risings	25-33 w:d	92-129 µs	0.0451	0.2256	0.4586	0.2707	10.24
41-63 risings	25-33 w:d	92-129 µs	0.0755	0.3082	0.3020	0.3144	12.24
64-94 risings	25-33 w:d	92-129 µs	0.1805	0.4586	0.2707	0.0902	10.24
95+ risings	25-33 w:d	92-129 µs	0.2466	0.5479	0.1644	0.0411	11.24
0-40 risings	33-44 w:d	92-129 µs	0.0090	0.1714	0.3684	0.4512	10.24
41-63 risings	33-44 w:d	92-129 µs	0.0451	0.2256	0.4586	0.2707	10.24
64-94 risings	33-44 w:d	92-129 µs	0.0902	0.2707	0.4586	0.1805	10.24
95+ risings	33-44 w:d	92-129 µs	0.2534	0.4178	0.2466	0.0822	11.24
0-40 risings	>44 w:d	92-129 µs	0.0100	0.0500	0.2400	0.7000	9.24
41-63 risings	>44 w:d	92-129 µs	0.0090	0.2691	0.2707	0.4512	10.24
64-94 risings	>44 w:d	92-129 µs	0.1428	0.2256	0.3609	0.2707	10.24
95+ risings	>44 w:d	92-129 µs	0.1879	0.2707	0.3609	0.1805	10.24
0-40 risings	0-25 w:d	129-281 µs	0.1805	0.3609	0.2707	0.1879	10.24

Table 84 continued: Posterior conditional probability table and posterior weights (α_0+n) for "EPT taxa richness" node (EPT taxa bins in probability units and α_0+n in equivalent data points).

Flashiness	Channel ratio	Conductivity	EPT taxa richness				
			0-4 taxa	5-7 taxa	8-12 taxa	>13 taxa	α_0+n
41-63 risings	0-25 w:d	129-281 µs	0.3000	0.4500	0.2000	0.0500	9.24
64-94 risings	0-25 w:d	129-281 µs	0.5500	0.3500	0.0900	0.0100	9.24
95+ risings	0-25 w:d	129-281 µs	0.5724	0.3899	0.0302	0.0075	12.24
0-40 risings	25-33 w:d	129-281 µs	0.1000	0.3000	0.4000	0.2000	9.24
41-63 risings	25-33 w:d	129-281 µs	0.2000	0.4000	0.3000	0.1000	9.24
64-94 risings	25-33 w:d	129-281 µs	0.3684	0.4061	0.1805	0.0451	10.24
95+ risings	25-33 w:d	129-281 µs	0.5939	0.3158	0.0812	0.0090	10.24
0-40 risings	33-44 w:d	129-281 µs	0.0411	0.2055	0.4178	0.3356	11.24
41-63 risings	33-44 w:d	129-281 µs	0.1000	0.3000	0.4000	0.2000	9.24
64-94 risings	33-44 w:d	129-281 µs	0.2781	0.3609	0.2707	0.0902	10.24
95+ risings	33-44 w:d	129-281 µs	0.4246	0.3699	0.1644	0.0411	11.24
0-40 risings	>44 w:d	129-281 µs	0.0075	0.1434	0.3899	0.4592	12.24
41-63 risings	>44 w:d	129-281 µs	0.0411	0.2945	0.4178	0.2466	11.24
64-94 risings	>44 w:d	129-281 µs	0.1712	0.3356	0.3288	0.1644	11.24
95+ risings	>44 w:d	129-281 µs	0.2534	0.4178	0.2466	0.0822	11.24
0-40 risings	0-25 w:d	>281 µs	0.3604	0.3140	0.2906	0.0349	13.24
41-63 risings	0-25 w:d	>281 µs	0.5500	0.3500	0.0900	0.0100	9.24
64-94 risings	0-25 w:d	>281 µs	0.6842	0.2707	0.0361	0.0090	10.24
95+ risings	0-25 w:d	>281 µs	0.7219	0.2601	0.0090	0.0090	10.24
0-40 risings	25-33 w:d	>281 µs	0.1644	0.4178	0.3356	0.0822	11.24
41-63 risings	25-33 w:d	>281 µs	0.3000	0.4500	0.2000	0.0500	9.24
64-94 risings	25-33 w:d	>281 µs	0.5939	0.3158	0.0812	0.0090	10.24
95+ risings	25-33 w:d	>281 µs	0.6842	0.2707	0.0361	0.0090	10.24
0-40 risings	33-44 w:d	>281 µs	0.1879	0.2707	0.3609	0.1805	10.24
41-63 risings	33-44 w:d	>281 µs	0.2000	0.4000	0.3000	0.1000	9.24
64-94 risings	33-44 w:d	>281 µs	0.3000	0.4500	0.2000	0.0500	9.24
95+ risings	33-44 w:d	>281 µs	0.5500	0.3500	0.0900	0.0100	9.24
0-40 risings	>44 w:d	>281 µs	0.1615	0.1516	0.3081	0.3787	15.24
41-63 risings	>44 w:d	>281 µs	0.0698	0.4360	0.3547	0.1396	13.24
64-94 risings	>44 w:d	>281 µs	0.1805	0.4586	0.2707	0.0902	10.24
95+ risings	>44 w:d	>281 µs	0.3000	0.4500	0.2000	0.0500	9.24

Table 85: Posterior conditional probability table and posterior weights (α_0+n) for "Richness-weighted tolerance" node (tolerance bins in probability units and α_0+n in equivalent data points).

Flashiness	Channel ratio	Conductivity	Richness-weighted tolerance				
			<5 tol.	5-5.6 tol.	5.6-6.2 tol.	>6.2 tol.	α_0+n
0-40 risings	0-25 w:d	0-92 µs	0.3000	0.4000	0.2500	0.0500	7.66
41-63 risings	0-25 w:d	0-92 µs	0.1437	0.4750	0.3094	0.0719	10.66
64-94 risings	0-25 w:d	0-92 µs	0.0885	0.2654	0.4693	0.1769	8.66
95+ risings	0-25 w:d	0-92 µs	0.0396	0.1586	0.5639	0.2379	9.66
0-40 risings	25-33 w:d	0-92 µs	0.5577	0.2654	0.1681	0.0088	8.66
41-63 risings	25-33 w:d	0-92 µs	0.4449	0.3172	0.1982	0.0396	9.66
64-94 risings	25-33 w:d	0-92 µs	0.2000	0.3210	0.3395	0.1395	12.66
95+ risings	25-33 w:d	0-92 µs	0.0885	0.3808	0.3538	0.1769	8.66
0-40 risings	33-44 w:d	0-92 µs	0.7000	0.2400	0.0500	0.0100	7.66
41-63 risings	33-44 w:d	0-92 µs	0.4142	0.4544	0.1248	0.0066	11.66
64-94 risings	33-44 w:d	0-92 µs	0.2654	0.4693	0.2211	0.0442	8.66
95+ risings	33-44 w:d	0-92 µs	0.2000	0.4000	0.3000	0.1000	7.66
0-40 risings	>44 w:d	0-92 µs	0.8800	0.1000	0.0100	0.0100	7.66
41-63 risings	>44 w:d	0-92 µs	0.6192	0.3278	0.0442	0.0088	8.66
64-94 risings	>44 w:d	0-92 µs	0.4423	0.3808	0.1681	0.0088	8.66
95+ risings	>44 w:d	0-92 µs	0.3000	0.4000	0.2500	0.0500	7.66
0-40 risings	0-25 w:d	92-129 µs	0.2000	0.4000	0.3000	0.1000	7.66
41-63 risings	0-25 w:d	92-129 µs	0.1000	0.3000	0.4000	0.2000	7.66
64-94 risings	0-25 w:d	92-129 µs	0.0359	0.2375	0.4172	0.3094	10.66
95+ risings	0-25 w:d	92-129 µs	0.0072	0.0647	0.4391	0.4890	10.66
0-40 risings	25-33 w:d	92-129 µs	0.2654	0.4693	0.2211	0.0442	8.66
41-63 risings	25-33 w:d	92-129 µs	0.1437	0.3812	0.4032	0.0719	10.66
64-94 risings	25-33 w:d	92-129 µs	0.0885	0.2654	0.4693	0.1769	8.66
95+ risings	25-33 w:d	92-129 µs	0.0396	0.1586	0.4604	0.3414	9.66
0-40 risings	33-44 w:d	92-129 µs	0.4423	0.2654	0.2835	0.0088	8.66
41-63 risings	33-44 w:d	92-129 µs	0.2654	0.4693	0.2211	0.0442	8.66
64-94 risings	33-44 w:d	92-129 µs	0.1769	0.4693	0.2654	0.0885	8.66
95+ risings	33-44 w:d	92-129 µs	0.0793	0.2379	0.3172	0.3656	9.66
0-40 risings	>44 w:d	92-129 µs	0.7000	0.2400	0.0500	0.0100	7.66
41-63 risings	>44 w:d	92-129 µs	0.4423	0.3808	0.1681	0.0088	8.66
64-94 risings	>44 w:d	92-129 µs	0.2654	0.3538	0.2211	0.1597	8.66
95+ risings	>44 w:d	92-129 µs	0.1769	0.3538	0.2654	0.2039	8.66

Table 85 continued: Posterior conditional probability table and posterior weights (α_0+n) for "Richness-weighted tolerance" node (tolerance bins in probability units and α_0+n in equivalent data points).

Flashiness	Channel ratio	Conductivity	Richness-weighted tolerance				
			<5 tol.	5-5.6 tol.	5.6-6.2 tol.	>6.2 tol.	α_0+n
0-40 risings	0-25 w:d	129-281 µs	0.2039	0.2654	0.3538	0.1769	8.66
41-63 risings	0-25 w:d	129-281 µs	0.0500	0.2000	0.4500	0.3000	7.66
64-94 risings	0-25 w:d	129-281 µs	0.0100	0.0900	0.3500	0.5500	7.66
95+ risings	0-25 w:d	129-281 µs	0.0072	0.1226	0.3094	0.5609	10.66
0-40 risings	25-33 w:d	129-281 µs	0.2000	0.4000	0.3000	0.1000	7.66
41-63 risings	25-33 w:d	129-281 µs	0.1000	0.3000	0.4000	0.2000	7.66
64-94 risings	25-33 w:d	129-281 µs	0.0442	0.1769	0.3980	0.3808	8.66
95+ risings	25-33 w:d	129-281 µs	0.0088	0.0796	0.4251	0.4865	8.66
0-40 risings	33-44 w:d	129-281 µs	0.4449	0.3172	0.1982	0.0396	9.66
41-63 risings	33-44 w:d	129-281 µs	0.2000	0.4000	0.3000	0.1000	7.66
64-94 risings	33-44 w:d	129-281 µs	0.0885	0.2654	0.3538	0.2924	8.66
95+ risings	33-44 w:d	129-281 µs	0.0396	0.1586	0.3568	0.4449	9.66
0-40 risings	>44 w:d	129-281 µs	0.3593	0.4970	0.1365	0.0072	10.66
41-63 risings	>44 w:d	129-281 µs	0.2379	0.4207	0.3018	0.0396	9.66
64-94 risings	>44 w:d	129-281 µs	0.1586	0.3172	0.2379	0.2863	9.66
95+ risings	>44 w:d	129-281 µs	0.0793	0.2379	0.4207	0.2621	9.66
0-40 risings	0-25 w:d	>281 µs	0.0328	0.4744	0.2956	0.1971	11.66
41-63 risings	0-25 w:d	>281 µs	0.0100	0.0900	0.3500	0.5500	7.66
64-94 risings	0-25 w:d	>281 µs	0.0088	0.0354	0.3808	0.5749	8.66
95+ risings	0-25 w:d	>281 µs	0.0088	0.0088	0.2747	0.7076	8.66
0-40 risings	25-33 w:d	>281 µs	0.1828	0.3414	0.3172	0.1586	9.66
41-63 risings	25-33 w:d	>281 µs	0.0500	0.2000	0.4500	0.3000	7.66
64-94 risings	25-33 w:d	>281 µs	0.0088	0.0796	0.3096	0.6020	8.66
95+ risings	25-33 w:d	>281 µs	0.0088	0.0354	0.2654	0.6904	8.66
0-40 risings	33-44 w:d	>281 µs	0.1769	0.4693	0.2654	0.0885	8.66
41-63 risings	33-44 w:d	>281 µs	0.1000	0.3000	0.4000	0.2000	7.66
64-94 risings	33-44 w:d	>281 µs	0.0500	0.2000	0.4500	0.3000	7.66
95+ risings	33-44 w:d	>281 µs	0.0100	0.0900	0.3500	0.5500	7.66
0-40 risings	>44 w:d	>281 µs	0.3878	0.2975	0.1402	0.1745	13.66
41-63 risings	>44 w:d	>281 µs	0.2172	0.2628	0.4544	0.0657	11.66
64-94 risings	>44 w:d	>281 µs	0.0885	0.2654	0.3538	0.2924	8.66
95+ risings	>44 w:d	>281 µs	0.0500	0.2000	0.4500	0.3000	7.66

Table 86: Posterior conditional probability table and posterior weights (α_0+n) for "Percent intolerant taxa" node (intolerant taxa bins in probability units and α_0+n in equivalent data points).

Flashiness	Channel ratio	Conductivity	Percent intolerant taxa				
			0-3%	3-8%	8-18%	>18%	α_0+n
0-40 risings	0-25 w:d	0-92 µs	0.0500	0.2500	0.4000	0.3000	8.23
41-63 risings	0-25 w:d	0-92 µs	0.1623	0.2199	0.3822	0.2356	11.23
64-94 risings	0-25 w:d	0-92 µs	0.1783	0.4650	0.2675	0.0892	9.23
95+ risings	0-25 w:d	0-92 µs	0.3391	0.4598	0.1609	0.0402	10.23
0-40 risings	25-33 w:d	0-92 µs	0.0089	0.1694	0.2675	0.5542	9.23
41-63 risings	25-33 w:d	0-92 µs	0.0402	0.2011	0.4196	0.3391	10.23
64-94 risings	25-33 w:d	0-92 µs	0.1378	0.3378	0.4000	0.1244	13.23
95+ risings	25-33 w:d	0-92 µs	0.1783	0.3567	0.3758	0.0892	9.23
0-40 risings	33-44 w:d	0-92 µs	0.0100	0.0500	0.2400	0.7000	8.23
41-63 risings	33-44 w:d	0-92 µs	0.0067	0.2914	0.2019	0.5000	12.23
64-94 risings	33-44 w:d	0-92 µs	0.0446	0.2229	0.4650	0.2675	9.23
95+ risings	33-44 w:d	0-92 µs	0.1000	0.3000	0.4000	0.2000	8.23
0-40 risings	>44 w:d	0-92 µs	0.0100	0.0100	0.1000	0.8800	8.23
41-63 risings	>44 w:d	0-92 µs	0.0089	0.0446	0.3223	0.6242	9.23
64-94 risings	>44 w:d	0-92 µs	0.0089	0.2778	0.2675	0.4458	9.23
95+ risings	>44 w:d	0-92 µs	0.0500	0.2500	0.4000	0.3000	8.23
0-40 risings	0-25 w:d	92-129 µs	0.1000	0.3000	0.4000	0.2000	8.23
41-63 risings	0-25 w:d	92-129 µs	0.2000	0.4000	0.3000	0.1000	8.23
64-94 risings	0-25 w:d	92-129 µs	0.3980	0.3298	0.2356	0.0366	11.23
95+ risings	0-25 w:d	92-129 µs	0.4921	0.4346	0.0660	0.0073	11.23
0-40 risings	25-33 w:d	92-129 µs	0.0446	0.2229	0.3567	0.3758	9.23
41-63 risings	25-33 w:d	92-129 µs	0.0733	0.2199	0.4712	0.2356	11.23
64-94 risings	25-33 w:d	92-129 µs	0.2867	0.3567	0.2675	0.0892	9.23
95+ risings	25-33 w:d	92-129 µs	0.3391	0.4598	0.1609	0.0402	10.23
0-40 risings	33-44 w:d	92-129 µs	0.0089	0.1694	0.3758	0.4458	9.23
41-63 risings	33-44 w:d	92-129 µs	0.0446	0.2229	0.3567	0.3758	9.23
64-94 risings	33-44 w:d	92-129 µs	0.0892	0.2675	0.4650	0.1783	9.23
95+ risings	33-44 w:d	92-129 µs	0.1609	0.5173	0.2413	0.0804	10.23
0-40 risings	>44 w:d	92-129 µs	0.0100	0.0500	0.2400	0.7000	8.23
41-63 risings	>44 w:d	92-129 µs	0.0089	0.1694	0.3758	0.4458	9.23
64-94 risings	>44 w:d	92-129 µs	0.1529	0.2229	0.3567	0.2675	9.23
95+ risings	>44 w:d	92-129 µs	0.1975	0.2675	0.3567	0.1783	9.23
0-40 risings	0-25 w:d	129-281 µs	0.1783	0.3567	0.2675	0.1975	9.23

Table 86 continued: Posterior conditional probability table and posterior weights (α_0+n) for "Percent intolerant taxa" node (intolerant taxa bins in probability units and α_0+n in equivalent data points).

Flashiness	Channel ratio	Conductivity	Percent intolerant taxa				
			0-3%	3-8%	8-18%	>18%	α_0+n
41-63 risings	0-25 w:d	129-281 µs	0.3000	0.4500	0.2000	0.0500	8.23
64-94 risings	0-25 w:d	129-281 µs	0.5500	0.3500	0.0900	0.0100	8.23
95+ risings	0-25 w:d	129-281 µs	0.5654	0.3980	0.0293	0.0073	11.23
0-40 risings	25-33 w:d	129-281 µs	0.1000	0.3000	0.4000	0.2000	8.23
41-63 risings	25-33 w:d	129-281 µs	0.2000	0.4000	0.3000	0.1000	8.23
64-94 risings	25-33 w:d	129-281 µs	0.3758	0.4012	0.1783	0.0446	9.23
95+ risings	25-33 w:d	129-281 µs	0.4904	0.4204	0.0802	0.0089	9.23
0-40 risings	33-44 w:d	129-281 µs	0.0402	0.2011	0.3218	0.4369	10.23
41-63 risings	33-44 w:d	129-281 µs	0.1000	0.3000	0.4000	0.2000	8.23
64-94 risings	33-44 w:d	129-281 µs	0.2867	0.3567	0.2675	0.0892	9.23
95+ risings	33-44 w:d	129-281 µs	0.3391	0.4598	0.1609	0.0402	10.23
0-40 risings	>44 w:d	129-281 µs	0.0073	0.1392	0.3980	0.4555	11.23
41-63 risings	>44 w:d	129-281 µs	0.0402	0.2989	0.3218	0.3391	10.23
64-94 risings	>44 w:d	129-281 µs	0.1782	0.3391	0.3218	0.1609	10.23
95+ risings	>44 w:d	129-281 µs	0.2587	0.3218	0.3391	0.0804	10.23
0-40 risings	0-25 w:d	>281 µs	0.2019	0.3028	0.2164	0.2789	12.23
41-63 risings	0-25 w:d	>281 µs	0.5500	0.3500	0.0900	0.0100	8.23
64-94 risings	0-25 w:d	>281 µs	0.5796	0.3758	0.0357	0.0089	9.23
95+ risings	0-25 w:d	>281 µs	0.7133	0.2688	0.0089	0.0089	9.23
0-40 risings	25-33 w:d	>281 µs	0.1609	0.3218	0.2413	0.2760	10.23
41-63 risings	25-33 w:d	>281 µs	0.3000	0.4500	0.2000	0.0500	8.23
64-94 risings	25-33 w:d	>281 µs	0.5988	0.3121	0.0802	0.0089	9.23
95+ risings	25-33 w:d	>281 µs	0.6879	0.2675	0.0357	0.0089	9.23
0-40 risings	33-44 w:d	>281 µs	0.0892	0.2675	0.4650	0.1783	9.23
41-63 risings	33-44 w:d	>281 µs	0.2000	0.4000	0.3000	0.1000	8.23
64-94 risings	33-44 w:d	>281 µs	0.3000	0.4500	0.2000	0.0500	8.23
95+ risings	33-44 w:d	>281 µs	0.5500	0.3500	0.0900	0.0100	8.23
0-40 risings	>44 w:d	>281 µs	0.0289	0.2851	0.2313	0.4546	14.23
41-63 risings	>44 w:d	>281 µs	0.0673	0.2836	0.4327	0.2164	12.23
64-94 risings	>44 w:d	>281 µs	0.2867	0.3567	0.2675	0.0892	9.23
95+ risings	>44 w:d	>281 µs	0.3000	0.4500	0.2000	0.0500	8.23

References

Abel, P. D. 1989. Water Pollution Biology. Ellis Horwood, Chichester, England.

Alameddine, I., Y. Cha, and K. H. Reckhow. 2010. Structure learning with Bayesian networks: An application to estuarine primary production. (in review).

Alberti, M., D. Booth, K. Hill, B. Coburn, C. Avolio, S. Coe, and D. Spirandelli. 2007. The impact of urban patterns on aquatic ecosystems: An empirical analysis in Puget lowland sub-basins. Landscape and Urban Planning **80**:345-361.

Alfaro, M. E., and M. T. Holder. 2006. The posterior and the prior in Bayesian phylogenetics. Annu. Rev. Ecol. Evol. Syst **37**:19-42.

Alpert, M., and H. Raiffa. 1982. A progress report on the training of probability assessors. Pages 294-305 Judgement under Uncertainty: Heristics and Biases. (eds D. Kahneman, P. Slovic, A. Tversky) Cambridge University Press, Cambridge.

Anderson, J. R., E. E. Hardy, J. T. Roach, and R. E. Witmer. 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964 Page 28.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish, 2nd ed. U.S. Environmental Protection Agency, Office of Water, Washington, D.C., EPA 841-B-99-002.

Barbour, M. T., and C. O. Yoder. 2007. (DRAFT under revision) Critical Technical Elements of a Bioassessment Program. U. S. EPA, Office of Water. Washington, DC. (EPA 200X-XX-XXX).

Bayes, T. 1763. An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philosophical Transactions, Giving Some Account of the Present Undertakings, Studies and Labours of the Ingenious in Many Considerable Parts of the World **53**:370-418.

Beasley, G., and P. Kneale. 2002. Reviewing the impact of metals and PAHs on macroinvertebrates in urban watercourses. Progress in Physical Geography **26**:236-270.

Beck, M. B. 2005. Vulnerability of water quality in intensively developing urban watersheds. Environmental Modelling & Software **20**:381-400.

Boerlage, B. 1992. Link Strengths in Bayesian Networks. Master's thesis, Dept. of Computer Science, The University of British Columbia.

Bolstad, P. V., and W. T. Swank. 1997. Cumulative impacts of landuse on water quality in a Southern Appalachian watershed Journal of the American Water Resources Association **33**:519-533.

Booth, D. B. 1990. Stream-channel incision following drainage-basin urbanization. Journal of the American Water Resources Association **26**:407-417.

Booth, D. B., D. Hartley, and R. Jackson. 2002. Forest cover, impervious-surface area, and the mitigation of stormwater impacts. Journal of the American Water Resources Association **38**:835-845.

Booth, D. B., and C. R. Jackson. 1997. Urbanization of aquatic systems: degradation thresholds, stormwater detection, and the limits of mitigation. Journal of the American Water Resources Association **33**:1077-1090.

Borsuk, M. E., C. A. Stow, and K. H. Reckhow. 2004. A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecological Modelling **173**:219-239.

Bray, J. R., and J. T. Curtis. 1957. An ordination of the upland forest communities of southern Wisconsin. Ecological Monographs **27**.

Bromley, J., N. A. Jackson, O. J. Clymer, A. M. Giacomello, and F. V. Jensen. 2005. The use of Hugin® to develop Bayesian networks as an aid to integrated water resource planning. Environmental Modelling & Software **20**:231-242.

Brown, L. R., M. B. Gregory, and J. T. May. 2009. Relation of urbanization to stream fish assemblages and species traits in nine metropolitan areas of the United States. Urban Ecosystems **12**:391-416.

Brown, M. T., and M. B. Vivas. 2005. Landscape development intensity index. Environmental Monitoring and Assessment **101**:289-309.

Choy, S. L., R. O'Leary, and K. Mengersen. 2009. Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical models. Ecology **90**:265-277.

Clarke, K. R., and R. N. Gorley. 2001. Primer 5: Non metric multi dimensional scaling (MDS) in Primer 5 user manual tutorial. Primer-E Ltd., New York.

Clemen, R. T. 1991. Making Hard Decisions: An Introduction to Decision Analysis. PWS-Kent Publishing Company, Boston, MA.

Coles, J. F., A. Bell, B. Scudder, and K. Carpenter. 2009. The effects of urbanization and other environmental gradients on algal assemblages in nine metropolitan areas across the United States: U.S. Geological Survey Scientific Investigations Report 2009–5022, 18 p.

Coles, J. F., T. F. Cuffney, G. McMahon, and Beaulieu. 2004. The effects of urbanization on the biological, physical, and chemical characteristics of coastal New England streams. U.S. Geological Survey Professional Paper 1695. Denver, CO.

Cowell, R. G. 1999. Probabilistic networks and expert systems. New York : Springer-Verlag.

Craig, P. S., M. Goldstein, A. H. Seheult, and J. A. Smith. 1998. Constructing Partial Prior Specifications for Models of Complex Physical Systems. Journal of the Royal Statistical Society. Series D (The Statistician) **47**:37-53.

Cuffney, T. F. 2003. User's manual for the National Water-Quality Assessment Program Invertebrate Data Analysis System (IDAS) software: Version 3. U.S. Geological Survey Open-File Report 03-172. Raleigh, North Carolina, USA.

Cuffney, T. F., M. D. Bilger, and A. M. Haigler. 2007. Ambiguous taxa: effects on the characterization and interpretation of invertebrate assemblages. Journal of the North American Benthological Society **26**:286-307.

Cuffney, T. F., R. A. Brightbill, J. T. May, and I. A. Waite. 2009a. Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas. Ecological Applications:e-View. doi: 10.1890/1808-1311

Cuffney, T. F., R. A. Brightbill, J. T. May, and I. A. Waite. 2009b. Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas Ecological Applications: e-View. doi: 10.1890/1808-1311

Cuffney, T. F., and J. A. Falcone. 2008. Derivation of nationally consistent indices representing urban intensity within and across nine metropolitan areas of the conterminous United States. U.S. Geological Survey Scientific Investigations Report 2008-5095.

Cuffney, T. F., R. Kashuba, S. Qian, I. Alameddine, Y. Cha, B. Lee, J. F. Coles, and G. McMahon. 2010. Multilevel regression models describing regional patterns of invertebrate and algal responses to urbanization across the United States. (almost submitted).

Cuffney, T. F., M.E. Gurtz, and M.R. Meador. 1993. Methods for collecting benthic invertebrate samples as part of the national water-quality assessment program. Open file report 93-406. U.S. Geologic Survey. Raleigh, NC, USA.

Cuffney, T. F., H. Zappia, E. M. P. Giddings, and J. F. Coles. 2005. Effects of urbanization on benthic macroinvertebrate assemblages in contrasting environmental settings: Boston, Massachusetts; Birmingham, Alabama; and Salt Lake City, Utah. American Fisheries Society Symposium **47**:361-407.

Davies, S. P., and S. K. Jackson. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications **16**:1251-1266.

Daymet. 2005. Numerical Terradynamic Simulation Group: Missoula, MT, University of Montana; accessed in December 2005 at http://www.daymet.org.

de Campos, L. M., and J. G. Castellano. 2007. Bayesian network learning algorithms using structural restrictions. International Journal of Approximate Reasoning **45**:233-254.

deCampos, L. M. 2006. A Scoring Function for Learning Bayesian Networks based on Mutual Information and Conditional Independence Tests. J. Mach. Learn. Res. **7**:2149-2187.

Dlamini, W. M. 2009. A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland. Environmental Modelling & Software **25**:199-208.

Falcone, J., and D. Pearson. 2006. Land-Cover and Imperviousness Data for Regional Areas near Denver, Colorado; Dallas-Fort Worth, Texas; and Milwaukee-Green Bay, Wisconsin - 2001. U.S. Geological Survey Data Series 2006-221, Reston, VA.

Falcone, J., J. Stewart, S. Sobieszczyk, J. Dupree, G. McMahon, and G. Buell. 2007. A Comparison of Natural and Urban Characteristics and the Development of Urban Intensity Indices Across Six Geographic Settings.

Finkenbine, J. K., J. W. Atwater, and D. S. Mavinic. 2000. Stream health after urbanization. Journal of the American Water Resources Association **36**:1149-1160.

Fisher, R. A. 1922. On the mathematical foundations of theoretical statistics. Philosophical transactions of the Royal Society of London, Series A **222**:309-368.

Fitzpatrick, F. A., M. A. Harris, T. L. Arnold, and K. D. Richards. 2004. Urbanization influences on aquatic communities in Northeastern Illinois streams. Journal of the American Water Resources Association **40**:461-475.

Fitzpatrick, F. A., I. A. Waite, P. J. D. Arconte, M. R. Meador, M. A. Maupin, and M. E. Gurtz. 1998. Revised methods for characterizing stream habitat in the National Water-Quality Assessment program. U.S. Geological Survey Water-Resources Investigations Report 98-4052. Page 77.

Foster, D., F. Swanson , J. Aber, B. I., B. N., T. D., and K. A. 2003. The importance of land-use legacies to ecology and conservation BioScience **53**:77-88.

Frey, H. C. 1998. BRIEFING PAPER PART 1: Introduction to Uncertainty Analysis. Department of Civil Engineering, North Carolina State University. http://legacy.ncsu.edu/classes/ce456001/www/Background1.html.

Friedman, N. 1997b. Learning Belief Networks in the Presence of Missing Values and Hidden Variables. Pages 125-133 *in* Fourteenth International Conference on Machine Learning D. Fisher (Ed.); Morgan Kaufmann, San Francisco, CA.

Gelman, A., and J. Hill. 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge; New York, NY.

Gerritsen, J. 2008. The Biological Condition Gradient: Model Development and Calibration. U.S. Environmental Protection Agency, Office of Science and Technology. Prepared by Tetra Tech, Inc., Owings Mills, Maryland.

Giddings, E. M. P., A. H. Bell, K. M. Beaulieu, T. F. Cuffney, J. F. Coles, L. R. Brown, F. A. Fitzpatrick, J. Falcone, L. A. Sprague, W. L. Bryant, M. C. Peepler, C. Stephens, and G. McMahon. 2009. Selected physical, chemical, and biological data used to study urbanizing streams in nine metropolitan areas of the United States, 1999-2004. U.S. Geological Survey Data Series 423.

Gilliom, R., J. Barbash, C. Crawford, P. Hamilton, J. Martin, N. Nakagaki, L. Nowell, J. Scott, P. Stackelberg, G. Thelin, and D. Wolock. 2006. Pesticides in the nation's streams and groundwater, 1992–2001. Circular 1291. U.S. Geological Survey. Reston, VA.

Gower, J. C. 1966. Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika **53**:325-338.

Groffman, P. M., D. J. Bain, L. E. Band, K. T. Belt, G. S. Brush, J. M. Grove, R. V. Pouyat, I. C. Yesilonis, and W. C. Zipperer. 2003. Down by the riverside: urban riparian ecology. Frontiers in Ecology and the Environment **1**:315-321.

Haase, D. 2003. Holocene floodplains and their distribution in urban areas--functionality indicators for their retention potentials. Landscape and Urban Planning **66**:5-18.

Hall, K. J., and B. C. Anderson. 1988. The toxicity and chemical composition of urban stormwater runoff. Can. J. Civ. Eng **15**:98-106.

Harding, J. S., E. F. Benfield, P. V. Bolstad, G. S. Helfman, and E. B. D. Jones. 1998. Stream biodiversity: The ghost of land use past. Proceedings of the National Academy of Sciences of the United States of America **95**:14843-14847.

Heckerman, D. 1999. A Tutorial on Learning with Bayesian Networks.*in* M. Jordan, editor. Learning in Graphical Models. MIT Press, Cambridge, MA.

Hellawell, J. M. 1986. Biological Indicators of Freshwater Pollution and Environmental Management. Elsevier, London.

Hilsenhoff, W. L. 1987. An improved biotic index of organic stream pollution. Great Lakes Entomologist **20**:31-39.

Hilsenhoff, W. L. 1988. Rapid field assessment of organic pollution with a family level biotic index. Journal of the North American Benthological Society **7**:65-68.

Horner, R., D. Booth, A. Azous, and C. May. 1997. Watershed determinants of ecosystem functioning. Pages 251-274 Effects of Watershed Development and Management on Aquatic Ecosystems. C. Roessner, editor. American Society of Civil Engineering, New York, NY.

Horwitz, R. J., T. E. Johnson, P. F. Overbeck, T. K. O'Donnell, W. C. Hession, and B. W. Sweeney. 2008. Effects of riparian vegetation and watershed urbanization on fishes in streams of the Mid-Atlantic Piedmont (USA). Journal of the American Water Resources Association **44**:724-741.

Hosack, G. R., K. R. Hayes, and J. M. Dambacher. 2008. Assessing model structure uncertainty through an analysis of system feedback and Bayesian networks Ecological Applications **18**:1070-1082.

Hottelling, H. 1933. Analysis of a complex of statistical variables into principle components. Journal of Educational Psychology **24**:417-441.

Hugin. 2008. Hugin Educational, API version 7.0. released September 9, 2008., Aalborg, Denmark.

Hunsaker, C. T., and D. A. Levine. 1995. Hierarchical approaches to the study of water quality in rivers. BioScience **45**:193-203.

Huryn, A. D., V. M. Butz Huryn, C. J. Arbuckle, and L. Tsomides. 2002. Catchment land-use, macroinvertebrates and detritus processing in headwater streams: taxonomic richness versus function. Freshwater Biology **47**:401-415.

Jaccard, P. 1912. The distribution of flora in the alpine zone. New Phytologist **11**:37-50.

Jackson, S. K., and S. Holdsworth. 2007. The New England Wadeable Stream Survey (NEWS): Development of Common Assessments in the Framework of the Biological Condition Gradient. USEPA Office of Science and Techonology, USEPA Office of Water., Washington, DC.

Jacobson, M. Z. 2001. GATOR-GCMM: A global- through urban-scale air pollution and weather forecast model 1. Model design and treatment of subgrid soil, vegetation, roads, rooftops, water, sea ice, and snow. J. Geophys. Res. **106(D6)**:5385-5401.

Jensen, F. V., and T. D. Nielsen. 2007. Bayesian Networks and Decision Graphs. Second Edition. Springer, New York, NY.

Jenson, F. 2009. Personal communication.

Johnson, B. L., W. B. Richardson, and T. J. Naimo. 1995. Past, present, and future concepts in large river ecology. BioScience **45**:134-141.

Jones, R. C., and C. C. Clark. 1987. Impact of watershed urbanization on stream insect communities. Journal of the American Water Resources Association **23**:1047-1055.

Karr, J. R., and E. W. Chu. 2000. Sustaining living rivers. Hydrobiologia **422-423**:1-14.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and l. J. Schlosser. 1986. Assessing biological integrity in running waters: A method and its rationale. Special publication 5. Illinois Natural History Survey.

Kashuba, R., Y. Cha, I. Alameddine, B. Lee, and T. F. Cuffney. 2010. Multilevel hierarchical modeling of benthic macroinvertebrate responses to urbanization in nine metropolitan areas across the conterminous United States: U.S. Geological Survey Scientific Investigations Report 2009-5243, 88p.

Kemp, S. J., and J. R. Spotila. 1997. Effects of Urbanization on Brown Trout Salmo trutta, Other Fishes and Macroinvertebrates in Valley Creek, Valley Forge, Pennsylvania. American Midland Naturalist **138**:55-68.

Kennen, J. G. 1999. Relation of macroinvertebrate community impairment to catchment characteristics in New Jersey streams Journal of the American Water Resources Association **35**:939-955.

Kennen, J. G., and M. A. Ayers. 2002. Relation of environmental characteristics to the composition of aquatic assemblages along a gradient of urban land use in New Jersey, 1996-98. U.S. Geological Survey Water-Resources Investigations Report 02-4069. West Trenton, NJ.

Kjaerulff, U. B., and A. L. Madsen. 2008. Bayesian Networks and Influence Diagrams. Springer, New York, NY.

Klein, R. D. 1979. Urbanization and stream quality impairment. Journal of the American Water Resources Association **15**:948-963.

Konrad, C. P., and D. B. Booth. 2002. Hydrologic trends associated with urban development for selected streams in the Puget Sound Basin, Western Washington. U.S. Geological Survey, Water-Resources Investigations Report 02-4040 Tacoma, WA, USA.

Kullback, S. 1968. Information Theory and Statistics. Dover Publication.

Lammert, M., and J. D. Allan. 1999. Assessing biotic integrity of streams: Effects of scale in measuring the influence of land use/cover and habitat structure on fish and macroinvertebrates. Environmental Management **23**:257.

Lauritzen, S. L. 1995. The EM algorithm for graphical association models with missing data. Computational Statistics & Data Analysis **19**:191-201.

LeBlanc, R. T., R. D. Brown, and J. E. FitzGibbon. 1997. Modeling the effects of land use change on the water temperature in unregulated urban streams. Journal of Environmental Management **49**:445-469.

Lenat, D., and J. K. Crawford. 1994. Effects of land use on water quality and aquatic biota of three North Carolina Piedmont streams. Hydrobiologia **294**:185-199.

Lenat, D. R., and M. T. Barbour. 1994. Using benthic macroinvertebrate community structure for rapid, cost-effective, water quality monitoring: rapid bioassessment. Pages 187-215 in S.L. Loeb and A. Spacie, editors. Biological monitoring of aquatic systems. Lewis, Boca Raton, Florida, USA. .

Lenat, D. R., L. A. Smock, and D. L. Penrose. 1980. Use of benthic macroinvertebrates as indicators of environmental quality. Pages 97-112 *in* D. L. Worf, editor. Biological Monotoring for Envirionmental Effects. Heath, Lexington, MA.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Statistics and Computing **10**:325-337.

Mahler, B. J., P. C. VanMetre, T. J. Bashara, J. T. Wilson, and D. A. Johns. 2005. Parking Lot Sealcoat: An Unrecognized Source of Urban Polycyclic Aromatic Hydrocarbons. Environ. Sci. Technol. **39**:5560-5566.

Margalef, R. 1958. Information theory in ecology. General Systems **3**:36-71.

McDonnell, M. J., and S. T. A. Pickett. 1990. Ecosystem structure and function along urban-rural gradients: an unexploited opportunity for ecology. Ecology **71**:1232-1237.

McMahon, G., J. D. Bales, J. F. Coles, E. M. P. Giddings, and H. Zappia. 2003. Use of stage data to characterize hydrologic conditions in an urbanizaing environment. Journal of the American Water Resources Association **39**:1529-1546.

McMahon, G., and T. F. Cuffney. 2000. Quantifying urban intensity in drainage basins for assessing stream ecological conditions. Journal of the American Water Resources Association **36**:1247-1261.

McNay, R. S., B. G. Marcot, V. Brumovsky, and R. Ellis. 2006. A Bayesian approach to evaluating habitat for woodland caribou in north-central British Columbia. Canadian Journal of Forest Research **36**:3117-3133.

Merritt, R. W., K. W. Cummins, and V. H. Resh. 1996. Collecting, sampling, and rearing methods for aquatic insects. Pages 12-28 *in* R. W. Merritt and K. W. Cummins, editors. An introduction to the aquatic insects of North America. 3rd edition. Kendall/Hunt Publishing, Dubuque, Iowa. .

Morgan, M., and M. Henrion. 1990. Uncertainty: A Guide to the Treatment of Uncertainty in Policy and Risk Analysis. Cambridge University Press, Cambridge.

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

Morse, C. C., A. D. Huryn, and C. Cronan. 2003. Impervious surface area as a predictor of the effects of urbanization on stream insect communities in Maine, U.S.A. Environmental Monitoring and Assessment **89**:95-127.

Moscrip, A. L., and D. R. Montgomery. 1997. Urbanization, flood frequency, and salmon abundance in Pudget Lowland Streams. Journal of the American Water Resources Association **33**:1289-1297.

Moulton, S. R., J. L. Carter, S. A. Grotheer, T. F. Cuffney, and T. M. Short. 2000. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory-Processing, taxonomy, and quality control of benthic macroinvertebrate samples. US Geological Survey Open-File Report 00-212. . National Water Quality Laboratory, US Geological Survey Denver, Colorado.

Moulton, S. R. I., J. G. Kennen, R. M. Goldstein, and J. A. Hambrook. 2002. Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program. U.S. Geological Survey Open-File Report 02-150. Page 87.

Mullin, T. M. 1986. Understanding and supporting the process of probabilistic estimation. Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA.

Myllymäki, P., T. Silander, H. Tirri, and P. Uronen. 2002. B-Course: A Web-Based Tool for Bayesian and Causal Data Analysis. International Journal on Artificial Intelligence Tools **11**:369.

Nadkarni, S., and P. P. Shenoy. 2004. A causal mapping approach to constructing Bayesian networks. Decision Support Systems **38**:259-281.

NCDENR. 2006. North Carolina Department of Environment and Natural Resources. Standard Operating Procedures for Benthic Macroinvertbrates, http://www.esb.enr.state.nc.us/BAUwww/benthossop.pdf.

Neapolitan, R. E. 2004. Learning Bayesian Networks. Prentice Hall., Upper Saddle River, NJ.

Nicholson, A., and N. Jitnah. 1998. Using mutual information to determine relevance in Bayesian networks. Pages 399-410 PRICAI'98: Topics in Artificial Intelligence.

NOAA. 2005. Coastal Change Analysis Program (CCAP) homepage, National Oceanic and Atmospheric Administration, at http://www.csc.noaa.gov/crs/lca/ccap.html. accessed December 2005.

O'Hagan, A., C. E. Buck, A. Daneshkhah, J. R. Eiser, P. H. Garthwaite, D. J. Jenkinson, J. E. Oakley, and T. Rakow. 2006. Uncertain Judgements: Eliciting Experts' Probabilities. Wiley, Chichester, UK.

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

Ourso, R. T., and S. A. Frenzel. 2003. Identification of linear and threshold responses in streams along a gradient of urbanization in Anchorage, Alaska. Hydrobiologia **501**:117-131.

Paul, M. J., and J. L. Meyer. 2001. Streams in the urban landscape. Annual Review of Ecology and Systematics **32**:333-365.

Pinkham, C. F. A., and J. G. Pearson. 1976. Applications of a new coefficientof similarity to population surveys. Journal of the Water Pollution Control Federation **48**:717-723.

Pitt, R., R. Field, M. Lalor, and M. Brown. 1995. Urban stormwater toxic pollutants: assessment, sources, and treatability. Water Environment Research **67**:260-275.

Poff, N. L., J. D. Allan, M. B. Bain, J. R. Karr, K. L. Prestegaard, B. D. Richter, R. E. Sparks, and J. C. Stromberg. 1997. The Natural Flow Regime. BioScience **47**:769-784.

Pourret, O., P. Naim, and B. Marcot. 2008. Bayesian Networks: A practical Guide to Applications. Wiley, Chichester, UK.

Qian, S. S. 2010. Environmental and Ecological Statistics with R. Chapman & Hall/CRC, Boca Raton, FL.

R. 2008. R: A language and environment for statistical computing. . R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Reckhow, K. H. 1999. Water quality prediction and probability network models. Can. J. Fish. Aquat. Sci. **56**:1150-1158.

Reckhow, K. H. 2009. Bayesian Networks for the Assessment of the Effect of Urbanization on Stream Macroinvertebrates. Proceedings of the 43rd Hawaii International Conference on System Sciences. http://www.nicholas.duke.edu/people/faculty/reckhow/KHR%20PDF%20publications/HICSS-43%20paper-revised-KHR-1.pdf.

Reckhow, K. H., G. B. Arhonditsis, M. A. Kenney, L. Hauser, J. Tribo, C. Wu, K. J. Elcock, L. J. Steinberg, C. A. Stow, and S. J. McBride. 2005. A Predictive Approach to Nutrient Criteria. Environ. Sci. Technol. **39**:2913-2919.

Richards, C., and G. Host. 1994. Examining land use influences on stream habitats and macroinvertebrates: A GIS approach. Journal of the American Water Resources Association **30**:729-738.

Rieman, B., J. T. Peterson, J. Clayton, P. Howell, R. Thurow, W. Thompson, and D. Lee. 2001. Evaluation of potential effects of federal land management alternatives on trends of salmonids and their habitats in the interior Columbia River basin. Forest Ecology and Management **153**:43-62.

Ross, L. T., and D. A. Jones. 1979. Biological aspects of water quality in Florida. Technical Series Volume 4, no. 3. Florida Department of Environmental Regulation, Tallahassee.

Roth, N., J. Allan, and D. Erickson. 1996. Landscape influences on stream biotic integrity assessed at multiple spatial scales. Landscape Ecology **11**:141-156.

Roy, A. H., C. L. Faust, M. C. Freeman, and J. L. Meyer. 2005a. Reach-scale effects of riparian forest cover on urban stream ecosystems. Canadian Journal of Fisheries and Aquatic Sciences **62**:2312-2329.

Roy, A. H., M. C. Freeman, B. J. Freeman, S. J. Wenger, W. E. Ensign, and J. L. Meyer. 2005b. Investigating hydrologic alteration as a mechanism of fish assemblage shifts in urbanizing streams. Journal of the North American Benthological Society **24**:656-678.

Roy, A. H., A. D. Rosemond, D. S. Leigh, M. J. Paul, and J. B. Wallace. 2003a. Habitat-specific responses of stream insects to land cover disturbance: Biological consequences and monitoring implications. Journal of the North American Benthological Society **22**:292-307.

Roy, A. H., A. D. Rosemond, M. J. Paul, D. S. Leigh, and J. B. Wallace. 2003b. Stream macroinvertebrate response to catchment urbanization (Georgia, U.S.A.). Freshwater Biology **48**:329-346.

Saltelli, A., K. Chan, and E. M. Scott. 2000. Sensitivity Analysis. John Wiley and Sons, Chichester.

Schueler, T. R. 1994. The Importance of Imperviousness. Watershed Protection Techniques **1**:100-111.

Schueler, T. R., and J. Galli. 1992. Environmental Impacts of Stormwater Ponds. Watershed Restoration Sourcebook. Anacostia Restoration Team, editors. Metropolitan Washington Council of Governments, Washington, DC.

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal **27**:379-423, 623-656.

Shelton, A. D., and K. A. Blocksom. 2004. A Review of Biological Assessment Tools and Biocriteria in New England States.U.S. Environmental Protection Agency, Cincinnati, Ohio.

Simpson, E. H. 1949. Measurement of diversity. Nature **163**:688.

Sinokrot, B. A., and H. G. Stefan. 1993. Stream temperature dynamics: Measurements and modeling. Water Resources Research ; Vol/Issue: 29:7:Pages: 2299-2312.

Spetzler, C. S., and S. von Holstein. 1975. Probability Encoding in Decision Analysis. Management Science **22**.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. v. d. Linde. 2002. Bayesian measures of model complexity and fit. Journal Of The Royal Statistical Society Series B **64**:583-639.

Spiegelhalter, D. J., N. L. Harris, K. Bull, and R. C. G. Franklin. 1994. Empirical Evaluation of Prior Beliefs About Frequencies: Methodology and a Case Study in Congenital Heart Disease. Journal of the American Statistical Association **89**:435-443.

Spirtes, P., C. Glymour, and R. Scheines. 2000. Causation, Prediction, and Search. MIT Press, Cambridge, MA.

Sponseller, R. A., E. F. Benfield, and H. M. Valett. 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. Freshwater Biology **46**:1409-1424.

Sprague, L. A., D. A. Harned, D. W. Hall, L. H. Nowell, N. J. Bauch, and K. D. Richards. 2007. Response of stream chemistry during base flow to gradients of urbanization in selected locations across the conterminous United States, 2002–04. U.S. Geological Survey Scientific Investigations Report 2007–5083, Reston, Virginia, USA.

Sprague, L. A., R. E. Zuelig, and J. A. Dupree. 2006. Effects of Urbanization on Stream Ecosystems in the South Platte River Basin, Colorado and Wyoming.

Steck, H. 2001. Constraint-based structural learning in Bayesian networks using finite data sets, PhD thesis- University of Munich, Munich, Germany.

Stewart-Koster, B., S. E. Bunn, S. J. Mackay, N. L. Poff, R. J. Naiman, and P. S. Lake. 2009. The use of Bayesian networks to guide investments in flow and catchment restoration for impaired river ecosystems. Freshwater Biology **55**:243-260.

Strayer, D. L., R. E. Beighley, L. C. Thompson, S. Brooks, C. Nilsson, G. Pinay, and R. J. Naiman. 2003. Effects of land cover on stream ecosystems: Roles of empirical models and scaling issues. Ecosystems **6**:407-423.

Sweeney, B. W., and R. L. Vannote. 1978. Size Variation and the Distribution of Hemimetabolous Aquatic Insects: Two Thermal Equilibrium Hypotheses. Science **200**:444-446.

Taroni, F., C. Aitkin, P. Garbolino, and A. Biedermann. 2006. Bayesian Networks and Probabilistic Inference in Forensic Science. Wiley, Chichester, UK.

Tate, C., T. Cuffney, G. McMahon, E. Giddings, J. Coles, and H. Zappia. 2005. Use of an Urban Intensity Index to Assess Urban Effects on Streams in Three Contrasting Environmental Settings Pages 291-315 Effects of Urbanization on Stream Ecosystems. L. R. Brown, R. H. Gray, R. M. Hughes, and M. R. Meador, editors. American Fisheries Society, Symposium 47, Bethesda, Maryland.

Ticehurst, J. L., L. T. H. Newham, D. Rissik, R. A. Letcher, and A. J. Jakeman. 2007. A Bayesian network approach for assessing the sustainability of coastal lakes in New South Wales, Australia. Environmental Modelling & Software **22**:1129-1139.

Trimble, S. W. 1997. Contribution of Stream Channel Erosion to Sediment Yield from an Urbanizing Watershed. Science **278**:1442-1444.

Tversky, A., and D. Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. Science **185**:1124-1131.

Tversky, A., and D. J. Koehler. 1994. Support theory: A nonextensional representation of subjective probability. Psychological Review **101**:547-567.

USEPA. 1997. State source water assessment and protection programs: U.S. Environmental Protection Agency, Office of Water, EPA 816-R-97-009.

USEPA. 2005. DRAFT: Use of Biological Information to Better Define Designated Aquatic Life Uses in State and Tribal Water Quality Standards: Tiered Aquatic Life Uses. U.S. Environmental Protection Agency, Office of Science and Technology.

USGS. 2005. National Land Cover Database 2001 (NLCD 2001), U.S. Geological Survey, at http://www.mrlc.gov/mrlc2k_nlcd.asp. accessed December 2005.

Uusitalo, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. Ecological Modelling **203**:312-318.

Van Metre, P. C., B. J. Mahler, and E. T. Furlong. 2000. Urban Sprawl Leaves Its PAH Signature. Environ. Sci. Technol. **34**:4064-4070.

Vinson, M. R., and C. P. Hawkins. 1998. Biodiversity of stream insects: Variation at local, basin, and regional scales. Annual Review of Entomology **43**.

Vølstad, J. H., N. E. Roth, G. Mercurio, M. T. Southerland, and D. E. Strebel. 2003. Using Environmental Stressor Information to Predict the Ecological Status of Maryland Non-tidal Streams as Measured by Biological Indicators. Environmental Monitoring and Assessment **84**:219-242.

Waite, I. R., and K. D. Carpenter. 2000. Associations among Fish Assemblage Structure and Environmental Variables in Willamette Basin Streams, Oregon. Transactions of the American Fisheries Society **129**:754-770.

Wallace, J. B., J. W. Grubaugh, and M. R. Whiles. 1996. Biotic indices and stream ecosystem processes: Results from an experimental study. Ecological Applications **6**:140-151.

Walls, L., and J. Quigley. 2001. Building prior distributions to support Bayesian reliability growth modelling using expert judgement. Reliability Engineering & System Safety **74**:117-128.

Wallsten, T. S., and D. V. Budescu. 1983. Encoding Subjective Probabilities: A Psychological and Psychometric Review. Management Science **29**:151-173.

Walser, C. A., and H. L. Bart. 1999. Influence of agriculture on in-stream habitat and fish community structure in Piedmont watersheds of the Chattahoochee River System. Ecology of Freshwater Fish **8**:237-246.

Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan Ii. 2005. The urban stream syndrome: Current knowledge and the search for a cure. Journal of the North American Benthological Society **24**:706-723.

Walsh, C. J., A. K. Sharpe, P. F. Breen, and J. A. Sonneman. 2001. Effects of urbanization on streams of the Melbourne region, Victoria, Australia. I. Benthic macroinvertebrate communities. Freshwater Biology **46**:535-551.

Wang, L., J. Lyons, P. Kanehi, R. Bannerman, and E. Emmons. 2000. Watershed urbanization and changes in fish communities in Southeastern Wisconsin streams. Journal of the American Water Resources Association **36**:1173-1189.

Wichert, G. A., and D. J. Rapport. 1998. Fish community structure as a measure of degradation and rehabilitation of riparian systems in an agricultural drainage basin. Environmental Management **22**:425-443.

Winget, R. N., and F. A. Mangum. 1979. Biotic condition index: Integrated biological, physical, and chemical stream parameters for management. Intermountain Region, U.S. Department of Agriculture, Forest Service, Ogden, Utah.

Winkler, R. L. 1967. The assessment of prior distributions in Bayesian analysis. J. Am. Statist. Ass. **67**:776-800.

Winkler, R. L. 2003. An introduction to Bayesian inference and decision, Second edition. Gainesville, Fla. : Probabilistic Publishing.

Winkler, R. L. 2009. Personal communication.

Winterbourn, M., and C. Townsend. 1991. Streams and rivers: One way flow systems. Page 270 Fundamentals of aquatic ecology. Barnes R and Mann K, editors. Blackwell Scientific Publications, Oxford UK.

Wolman, M. G., and A. P. Schick. 1967. Effects of Construction on Fluvial Sediment, Urban and Suburban Areas of Maryland. Water Resour. Res **3**:451-464.

Yoder, C. O., R. J. Miltner, and D. White. 1999. Assessing the status of aquatic life designated uses in urban and suburban watersheds. Pages 16-28. In A. Everson, S. Minamyer, J. Dye, P. Heimbrock, and S.Wilson, editors. National Conference on Retrofit Opportunities for Water Resource Protection in Urban Environments. U.S. Environmental Protection Agency, EPA/625/R-99/002, Washington, DC.

Yoder, C. O., and E. T. Rankin. 1996. Assessing the Condition and Status of Aquatic Life Designated Uses in Urban and Suburban Watersheds. Pages 201-227 Effects of Watershed Development and Management on Aquatic Ecosystems,. L. A. Roesner, editor. American Society of Civil Engineers, New York , NY.

Biography

Roxolana Oresta Kashuba was born in Chicago, IL on August 5, 1979. She has received a Bachelor of Arts in Chemistry from Northwestern University in Evanston, IL in June 2001, a Master of Science in Environmental Science from the University of Illinois School of Public Health in Chicago, IL in August 2003, and a Doctor of Philosophy in Water Quality Modeling from Duke University in Durham, NC in May 2010. Publications include:

Kashuba, R., Cuffney, T., McMahon, G., Reckhow, K., Gerritsen, J. and Davies, S. Parameterizing the Biological Condition Gradient in the northeast United States using a Bayesian network approach. (In review)

Kashuba, R., Cuffney, T., Qian, S., Reckhow, K. Combining expert knowledge with data to develop a Bayesian network describing the impact of urbanization on aquatic macroinvertebrate communities in the southeast United States. U.S. Geological Survey Scientific Investigations Report 2010-XXXX, XXp. (In review)

Cuffney, T.F., **Kashuba, R.**, Qian, Song S., Alameddine, I., Cha, Y., Lee. B., Coles, J.F. and McMahon, G. Multilevel regression models describing regional patterns of invertebrate and algal responses to urbanization across the United States. (In press)

Kashuba, R., Cha, Y., Alameddine, I., Lee, B., and Cuffney, T. Multilevel Hierarchical Modeling of Benthic Macroinvertebrate Responses to Urbanization in Nine

Metropolitan Areas across the Conterminous United States. U.S. Geological Survey Scientific Investigations Report 2009-5243, 88p.

Cuffney, T.F., McMahon, G., **Kashuba, R.**, May, J.T., and Waite, I.A. responses of Benthic Macroinvertebrates toEnvironmental Changes Associated with Urbanization in Nine Metropolitan Areas of the Conterminous United States. In Webb, Richard M.T., and Darius Semmens, eds, Proceedings of the Third Interagency Conference on Research in the Watersheds, September 8-11, 2008.

Kashuba, Roxolana O., Scheff, Peter A. Nonlinear Regression Adjustments of Multiple Continuous Monitoring Methods Produce Effective Characterization of Short-Term Fine Particulate Matter. J Air Waste Manag Assoc. 2008 Jun; 58(6):812-20.

Kashuba, Roxolana O., Scheff, Peter A., Rizzo, Michael. Characterization of Nephelometer- Measured Short- Term Fine Particulate Matter in Region 5. Proc. AWMA Symposium on Air Quality Measurement Methods and Technology, Durham, NC (April 2004).

Roxolana Kashuba has been awarded United States Geological Survey Ph.D. Student Research Salary Support (2008-2010), National Institute for Occupational Safety and Health (NIOSH) Graduate Traineeship Grant in Industrial Hygiene (2001-2003), Research and Conference Travel Grants totaling $10,000 (2003, 2006-2009), and has been inducted into Delta Omega Public Health Honor Society. Her current research interests include environmental statistical methods and uncertainty characterization.