

ABSTRACT

Title of Thesis: FIELD-MEASURED VERSUS DERIVED: WHAT ARE THE MOST EFFECTIVE PREDICTOR VARIABLES IN STREAM BIODIVERSITY MODELS?

Miriam Rebecca Johnston, Master of Science, 2014

Thesis directed by: Professor Matthew C. Fitzpatrick
University of Maryland Center for Environmental Science

Statistical models used to predict and map patterns of biodiversity require environmental variables with full coverage across an area of interest. By necessity, these variables are derived from GIS, remote sensing, or via interpolation, and may not be as physiologically relevant to biota or as representative of on-the-ground conditions as field-measured variables. This research used generalized dissimilarity modeling and occurrence data for freshwater fish and benthic invertebrates in Maryland to examine differences in explanatory power, predictive ability, and management inference yielded by derived and field-measured variables. Across the state and for both taxa, models fit with field-measured variables were superior in explanation and prediction, and nearly always more parsimonious. However, there was little difference between the variable sets in ability to predict management-related indices. Results suggest that field-measured variables are preferred over derived variables overall, but their absence from predictive models may not have a large effect on management inference.

FIELD-MEASURED VERSUS DERIVED: WHAT ARE THE MOST EFFECTIVE
PREDICTOR VARIABLES IN STREAM BIODIVERSITY MODELS?

by

Miriam Rebecca Johnston

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2014

Advisory Committee:
Professor Matthew C. Fitzpatrick, Chair
Professor Andrew J. Elmore
Professor Robert H. Hilderbrand

©Copyright by
Miriam Rebecca Johnston
2014

ACKNOWLEDGMENTS

Thanks go first to Matt Fitzpatrick, whose scientific acuity, dedication, and perennially open door were the foundation of this thesis. Working with such a mentor is a true privilege. Thanks also to my committee, Andrew Elmore and Bob Hilderbrand, for their consistent encouragement and insightful comments.

This research would not have been possible without the Maryland Biological Stream Survey and the generous guidance of staff at the Maryland Department of Natural Resources. Thanks especially to Jason Cessna, Andy Becker, Michael Kashiwagi, Scott Stranko, and Ron Klauda.

For analysis support and responding to my endless barrage of questions with equally endless good humor, thank you to Matt Lisk and Steven Guinn. I'm very grateful to Bob Gardner and Diego Nieto Lugilde for much stimulating conversation, and to the greater community at the Appalachian Lab – what a tremendous place to learn.

This project was funded by Maryland Sea Grant.

TABLE OF CONTENTS

FIGURE LIST	v
TABLE LIST	v
INTRODUCTION	1
METHODS	6
Study region	6
Species occurrence data	9
Environmental predictor variables	11
Field-measured variables	11
Derived variables	12
Community-level modeling	17
Evaluation of field-measured and derived variable sets	21
Management inference: predicting indices of biotic integrity	23
RESULTS	25
Evaluation of field-measured and derived variable sets	25
Explanatory power	25
Parsimony	27
Predictive ability	30
Important individual variables	32
Management inference: predicting indices of biotic integrity	35
DISCUSSION	37
Evaluation of field-measured and derived variable sets	37
Explanatory power	37
Parsimony	40
Predictive ability	41
Important individual variables	41
Management inference: predicting indices of biotic integrity	45
Future model improvements	46

Conclusions.....	47
ADDENDUM.....	48
APPENDICES	50
Appendix A. Fish considered native and non-native	50
Appendix B. Intra-group variable correlations	53
Appendix C. Inter-group variable correlations	56
Appendix D. AICs for all seven variable sets.....	57
Appendix E. Generalized dissimilarity model plots	58
Appendix F. Statistical significance of prediction results	59
Appendix G. Variables included in final generalized dissimilarity models	61
Appendix H. Total number of models for which variables were selected.....	65
LITERATURE CITED.....	66

FIGURE LIST

Figure 1. Maps of study region.....	7
Figure 2. Generalized dissimilarity model deviance explained values.....	26
Figure 3. Deviance uniquely explained by each variable set.....	27
Figure 4. Variable set ability to predict compositional dissimilarity.....	31
Figure 5. Relative importance of predictor variables.....	33
Figure 6. Variable set ability to predict indices of biotic integrity.....	36

TABLE LIST

Table 1. Field-measured variables.....	11
Table 2. Derived variables.....	16
Table 3. Generalized dissimilarity model AICs.....	29
Table 4. Variable selection chi-square test results.....	34

INTRODUCTION

Choosing relevant environmental predictor variables is one of the most universal challenges in modeling geographic patterns of biodiversity (Araújo and Guisan 2006, Synes and Osborne 2011, Williams et al. 2012). Though model output is fundamentally driven by predictor variable input (Araújo and Guisan 2006), the drivers of biotic distributions may be unknown (Dudgeon et al. 2006) or unavailable (Austin 2002). In this context, considerable research has examined predictor variable spatial scale (Peterson et al. 2011, Wang et al. 2003, Morley and Karr 2002, Mackey and Lindenmayer 2001, Austin and Van Niel 2011), variable selection techniques (Mac Nally 2000, Pearce and Ferrier 2000), and methods of managing inter-variable relationships (Braunisch et al. 2013). Comparatively poorly studied but also potentially highly influential is the immediacy with which environmental predictor variables are related to biotic distributions (Austin 1980).

Broadly, predictor variables can be classified as “proximal” or “distal” based on their physiological relevance to biota (Austin 1980, Franklin 1995, Guisan and Zimmermann 2000, Austin 2002). Proximal variables describe direct physiological influences on biotic distributions, such as nutrient availability for plants or water temperature for aquatic organisms. In contrast, distal variables do not have a direct physiological influence, but can indirectly drive biotic distributions if they are correlatively or causally related to proximal variables (Austin 2002, Williams et al. 2012, Elith and Leathwick 2009). Elevation and latitude are classical distal variables. From a theoretical perspective, models based on proximal variables are preferable to models based on distal variables because the relationship between proximal drivers and distal

surrogates may be weak and/or spatially or temporally inconsistent (Austin 2002, Elith and Leathwick 2009, Williams et al. 2012).

Though in theory proximal variables should produce more robust biodiversity models, in practice they are often not used in model fitting. Most proximal variables must be collected by field measurement and are therefore unavailable in the full-coverage layers necessary for predictive mapping (Austin 2002). Instead, full-coverage environmental predictors are typically derived from GIS, remote sensing, or via interpolation of point data. The majority of GIS and remote sensing-derived variables are distal and may not characterize habitat at the spatial scales most relevant to biotic life histories (Cord et al. 2013). Interpolated variables are estimates based on discrete measurements and may not be indicative of on-the-ground conditions. In short, the predictors most relevant to biology are rarely available for spatial prediction, while those available for mapping may have only indirect or approximated influence on biotic distributions.

How much explanatory and predictive power is lost as a result of the exclusion of largely proximal field-measured variables from biodiversity models, and how does that alter management inference? Though the literature generally supports the theory that proximal variables are more effective predictors than distal variables (Leathwick and Whitehead 2001, Zimmermann et al. 2007, Kristensen et al. 2012), comparisons have been sparse on account of the relative difficulty of identifying and collecting sufficient proximal information (Austin 2002) and the great interest in predictive mapping. Such comparisons are particularly lacking in streams and rivers, for which differences between models built with field-measured versus derived variables may be especially pronounced

due to dendritic habitat geometry (Grant et al. 2007), rapidly changing environmental conditions (Dettinger and Diaz 2000), and integration of upstream influences (Nelson et al. 2009).

Despite the theoretical predictions, the importance of derived predictor variables in stream biodiversity models could be greater than the importance of field-measured variables. Conditions in streams change quickly through time and across space, integrating diverse drainage basin processes and possibly reducing the relevance of one-time field measurements (Sala 2000, Benda et al. 2004, Dudgeon et al. 2006, Nelson et al. 2009).

It is more likely, however, that the unique attributes of streams will decrease the explanatory and predictive abilities of derived variables relative to field-measured variables. Possible reasons for decreased importance of derived variables in stream biodiversity models are manifold. First, the branching and hierarchical geometry of streams makes interpolation of point measurements difficult (Benda et al. 2004, Grant et al. 2007). As a result, available interpolated variables are often measured on land rather than in the channel (Hijmans et al. 2005) and may not capture channel conditions. Additionally, given the connected nature of stream networks (Fisher 1997), stream biota respond to both local and upstream drivers (Morley and Karr 2002, Kratzer et al. 2006, Urban et al. 2006, Stanfield and Kilgour 2013). While derived variables can provide a surrogate for upstream influences by characterizing drainage areas and flow pathways, field-measured variables collected in streams reflect the true effects of those influences. Relatedly, a large proportion of a stream network is comprised of small, potentially ephemeral or intermittent headwaters (Freeman et al. 2007), which are often largely

unmapped (Elmore et al. 2013) and narrower than the resolution of environmental grids (30 meters, in the case of Landsat). Without knowing the location and characteristics of these ecologically important stream channels (Freeman et al. 2007), it is difficult to quantify the effects of various watershed land covers, which are attenuated differently across land and through channels (Johnson et al. 2007, Van Sickle and Johnson 2008, Walsh and Kunapo 2009).

In this study, I focused on comparisons between field-measured and derived variables as predictors in Maryland stream biodiversity models. While not perfect proxies, field-measured and derived variables are management-relevant analogs to proximal and distal variables. I used a community-level modeling method – generalized dissimilarity modeling (GDM; Ferrier et al. 2007) – to model pairwise compositional dissimilarity as a function of environmental and geographic distances. GDM considers all species in an assemblage, regardless of rarity, and provides a robust method of statistically selecting, weighting, and transforming candidate environmental variables such that they best represent biological patterns (Ferrier et al. 2007). GDM also lends itself to the development of biological classification systems and mapped patterns, which may be particularly useful for management (Leathwick et al. 2011, Snelder et al. 2012).

The primary goal of this research was to assess the effects of excluding field-measured variables from predictive stream biodiversity models. Specifically, I aimed to (1) understand the relative merits of using field-measured versus derived variables as predictors in stream biodiversity models, (2) identify the individual environmental variables with the greatest influence on fish and invertebrate compositional turnover in Maryland streams, thereby providing information that can inform future variable

measurement and derivation, and (3) quantify the difference in management inference yielded by biodiversity models built with field-measured versus derived variables.

METHODS

Study region

This research involves first through fourth order streams in Maryland west of the Chesapeake Bay. The region is characterized by relatively long gradients in population density, land cover, elevation, and geology and a broad array of stream habitat types. Population density ranges from 2604.7 people per square kilometer in Baltimore City to 17.7 people per square kilometer in Garrett County (World Media Group 2014), with land cover of the eastern portion of the study region being predominantly urban and the western portion largely forested (Homer et al. 2007, Fig. 1a). Cultivation of crops and livestock is prevalent in much of the central study area (Homer et al. 2007). Elevation ranges from sea level at the Chesapeake Bay in the east to 1024 m in the western part of the state (Reger and Cleaves 2008).

The study region spans five physiographic provinces, frequently grouped by the Maryland Department of Natural Resources (MDNR) into three regions: the Coastal Plain (C), the Piedmont (P), and the Highlands. The Coastal Plain is characterized by low relief and unconsolidated sand, gravel, silt, and clay, the Piedmont by rolling hills and hard igneous and metamorphic rock, and the Highlands by comparatively steep terrain and faulted and folded sedimentary and metamorphic rock (Schmidt 1993, Reger and Cleaves 2008). For the purposes of these analyses, I separated the Highlands into two components: the Chesapeake (H) and Youghiogheny (Y) basins (Fig. 1a, b). Streams in the Youghiogheny basin are the only in Maryland that drain to the Gulf of Mexico, and the separation acknowledges the Youghiogheny's distinct biotic communities. I modeled each of these four areas separately, recognizing the role of their unique geologic character

and geomorphic history in governing biotic distributions and influential habitat variables (Melles et al. 2014).

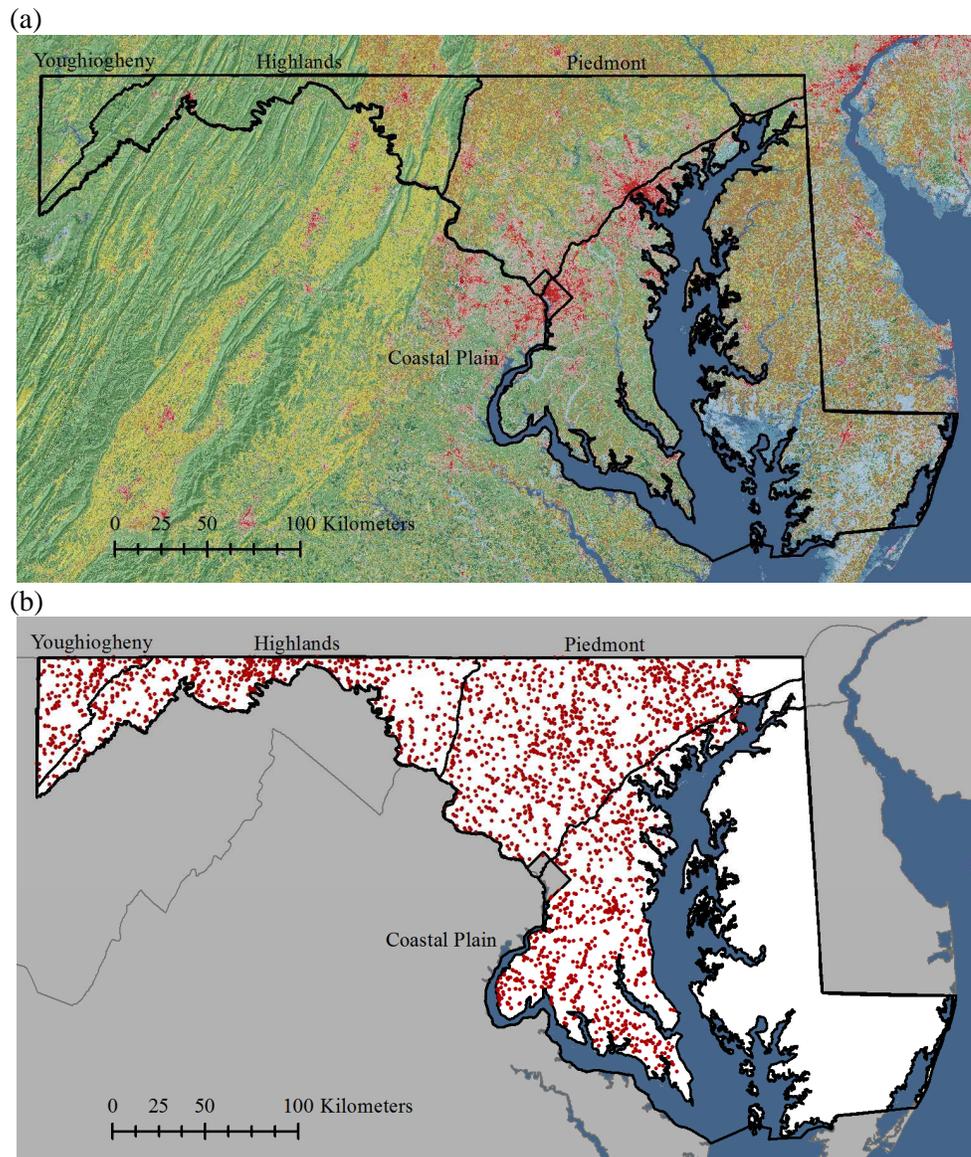


Figure 1. Maps of study region showing (a) land cover from the 2001 National Land Cover Database (Homer et al. 2007) and topography, and (b) stream survey locations visited by MDNR as part of the Maryland Biological Stream Survey. Red and pink polygons in the upper panel depict development, greens are forest, blues are wetlands or open water, and yellows are agriculture.

Maryland is an ideal study area for three reasons. First, there is considerable interest in the conservation and restoration of the Chesapeake Bay, the largest estuary in the United States. Maintaining and promoting the integrity of Maryland's freshwater streams is critical to the Bay's ecological functioning (Goetz et al. 2004). Second, the MDNR has surveyed fish and benthic invertebrates in Maryland streams for almost two decades (Fig. 1b). The resultant dataset provides excellent spatial coverage and is coupled with field-measured habitat variables (Stranko et al. 2007). Finally, Elmore et al. (2013) recently produced a detailed stream map for the Potomac River watershed and several adjacent watersheds, covering the entirety of the study area.

Elmore et al.'s (2013) map was produced using maximum entropy (MaxEnt, Phillips et al. 2006) and terrain and soil variables to predict stream presence at 10m resolution. Predictions included previously unmapped channels, notably headwaters and streams lost to urbanization, and suggested that the National Hydrography Database (NHD) underestimates stream network density in Maryland by up to 250%. Using Elmore et al.'s (2013) map allowed more precise snapping of biological survey locations to stream channels and more accurate measurement of the flow path distance between land and stream pixels, which was used in several of the derived environmental variables (see below). However, it is also important to note that Elmore et al.'s (2013) models were trained on stream presences collected from fully forested watersheds, and the map therefore reflects *potential* stream density rather than existing channels. Areas with altered land use may have either higher or lower stream density than predicted, depending on the type of development (Elmore et al. 2013). Still, despite its potential to misclassify

current channels, Elmore et al.'s (2013) map is considerably more consistent in its accuracy across physiographic provinces than the NHD.

Species occurrence data

Species occurrence data were collected as part of the Maryland Biological Stream Survey (MBSS), a program originated and administered by the MDNR (Stranko et al. 2007). The MBSS follows standardized protocols to sample fish and benthic invertebrates in 75 m segments of non-tidal first, second, third, and fourth order streams (Stranko et al. 2007). The majority of the sample sites were selected randomly, using a probability-based design, while some were selected to answer specific management or research questions.

Briefly, MBSS sampling proceeded as follows. Benthic invertebrate sampling was conducted in March and April using a 450 μm mesh D-net. Twenty square feet of habitat were sampled at each site by choosing 20 locations for the D-net and manually disturbing the substrate in the square foot immediately upstream. Locations were chosen to be a proportional representation of the habitats likely to contain the most diverse benthic invertebrate communities (described in Stranko et al. 2007). Taxa were identified and quantified in the lab, where each benthic sample was spread over a gridded tray and a random number was chosen to determine which 5 x 5 cm grid cell should be sampled (Boward and Friedman 2011). All invertebrates in the chosen cell were identified. If the total number of individuals identified was equal to or greater than 120, analysis of that sample was complete. If there were fewer than 120 individuals in the cell, another cell was randomly selected.

Field sampling of fish was conducted between June and September, the low-flow period, via double-pass electrofishing (Stranko et al. 2007). To ensure that fish neither exited nor entered the site during electrofishing, block nets were positioned at the top and bottom of the 75 m segment, as well as at tributary confluences. All caught fish with body length over 30 mm were censused. Fish were counted and identified to species in the field, when possible.

In three major sampling periods over 18 years (1994-2011), the MBSS collected data for approximately 100 fish and 600 invertebrate taxa. This study uses data from the 86 native fish identified to species and the 581 invertebrate taxa identified to family or genus at 2,165 unique site-years. To ensure that all survey points coincided with Elmore et al.'s (2013) mapped streams, site locations were auto-snapped to the nearest stream cell by flow direction (i.e. points were never snapped to cells upstream of their reported locations). Points farther than 300 m from stream cells or located on flat terrain without clear flow direction were individually examined and, where possible, were manually moved to a stream location according to their site description recorded on MBSS datasheets.

I considered only native fish species because the distributions of non-native species are likely to be dominated by factors other than environmental conditions (e.g., Christmas et al. 2001, Kilian et al. 2012). I compiled separate Maryland native fish lists for each region of interest (Y, H, P, and C) using the USGS database for Nonindigenous Aquatic Species (United States Geological Survey 2014), maps from the NatureServe non-profit organization (NatureServe 2014), publications by MDNR employees and contractors (Southerland et al. 2005, Stranko et al. 2010), and expert knowledge from a

specialist involved with MBSS sampling (Appendix A). In cases when sources conflicted, I favored USGS data.

Environmental predictor variables

Field-measured variables

Field-measured environmental variables were collected by the MBSS at survey locations coincident with faunal sampling (Stranko et al. 2007). They include stream measurements relating to flow and gradient, in-stream habitat, and water chemistry (Table 1). I omitted unordered categorical variables and variables with five or fewer ordered categories to accommodate GDM's use of environmental distances in model fitting (Ferrier et al. 2007). To maintain sample sizes, I also omitted any field-measured variable lacking measurements at >1000 site-years.

Table 1. Field-measured variables collected by the MBSS at survey locations and considered as environmental predictor variables.

Variable code	Description
Flow and gradient	
ST_GRAD	Stream gradient (%), measured from the downstream boundary of the sample segment to the upstream boundary with an inclinometer (1995-2004) and a level (2007-2009)
DischargeCFS	Summer stream flow (cfs), standard transect method
Habitat	
INSTRHAB	In-stream fish habitat structure rating (0-20)
EPI_SUB	Epifaunal substrate rating (benthic invertebrate habitat, 0-20)
VEL_DPTH	Velocity/depth diversity rating (0-20)
POOLQUAL	Pool/glide/eddy quality rating (0-20)
RIFFQUAL	Riffle/run quality rating (0-20)
EMBEDDED	Embeddedness: percentage that gravel, cobble, and boulder particles are surrounded by sediment or flocculent material
SHADING	Percentage of segment that is shaded
AESTHET	Trash rating (0-20)
MAXDEPTH	Maximum depth in sample reach (cm)
AVGWID	Average wetted width of the 1, 25, 50, and 75 m points of the sample segment (m)
AVGTHAL	Average thalweg depth of the 1, 25, 50, and 75 m points of the sample segment (cm)
AVG_VEL	Average velocity of the 1, 25, 50, and 75 m points of the sample segment (m/s)

Variable code	Description
Water chemistry	
PH_LAB	Spring pH (pH units), measured in the lab
COND_LAB	Spring conductance ($\mu\text{mho/cm}$), lab
ANC_LAB	Acid neutralizing capacity ($\mu\text{eq/L}$), lab
DOC_LAB	Dissolved organic carbon (mg/L), lab
SO4_LAB	Sulfate (mg/L), lab
NO3_LAB	Nitrate nitrogen (mg/L), lab
TEMP_FLD	Summer in-situ water temperature ($^{\circ}\text{C}$)
DO_FLD	In-situ dissolved oxygen (mg/L)
PH_FLD	In-situ summer pH (pH units)
COND_FLD	In-situ summer conductance ($\mu\text{mho/cm}$)

Derived variables

Derived environmental variables were calculated using GIS and remote sensing data or downloaded from online databases (e.g., WorldClim, Hijmans et al. 2005). They emphasize local and watershed-scale physical characteristics considered important to biotic distributions and include measures of topography, hydrography, soils, climate, and land use (Austin 2002, Pease et al. 2011, Allan 2004, Table 2). All derived variables are temporally-invariant and were created at or resampled to 10 m resolution to match the scale of the stream network maps.

Topographic variables attempt to characterize landscape shape and, by extension, stream channel shape and related stream characteristics, such as flow speed and substrate type (Melles et al. 2014). They were derived using a 10 m digital elevation model (DEM) from the National Elevation Dataset (NED, Gesch et al. 2002, see Julian et al. 2012 for details).

Hydrographic variables capture flow and network characteristics. Eight-way flow direction and flow accumulation (i.e. watershed size) were derived using the Terrain Analysis Using Digital Elevation Models toolset (TauDEM, Tarboton 2014). Stream

length and network density were measured from Elmore et al.'s (2013) stream map, and attempt to characterize the residence time of water in watersheds, the relative importance of terrestrial and aquatic inputs, and the area of available connected habitat (Elmore et al. 2013). A set of “stream burial” variables were derived that quantify the extent to which stream segments have been paved over or directed into culverts, pipes, or concrete-lined ditches (Elmore and Kaushal 2008). Burial-related variables were calculated using the National Land Cover Database impervious surface map (Homer et al. 2007) and USGS 30-cm aerial photography (Elmore and Kaushal 2008). Stream burial is highly correlated with urbanization and disproportionately affects sensitive headwater species (Elmore and Kaushal 2008).

Soils data were derived from the Natural Resources Conservation Service’s Soil Survey Geographic Database (SSURGO, Soil Survey Staff n.d.). Soils variables were mapped at broad scales (1:12,000 to 1:63,360) and were collected on land over many years. However, many of the SSURGO variables potentially have a strong influence on in-stream conditions: silt-clay % influences runoff potential and flashiness, soil erodibility and bulk density are related to stream sediment load, and soil pH and bedrock depth can affect water chemistry. Derived annual mean surface air temperature is also extrapolated from measurements taken on land (Hijmans et al. 2005). Stream and air temperatures often are closely related, but the relationship may not be linear (Mohseni and Stefan 1999) and varies regionally (Hilderbrand et al. 2014).

Land use and land cover metrics are some of the most commonly-used predictors of water quality and biotic assemblages (Harding et al. 1998, Allan 2004, Van Sickle and Johnson 2008, Utz et al. 2010). Forest, canopy cover, wetland, agriculture, and

impervious surface comprise the land covers available from the NLCD that are relevant to the mid-Atlantic United States (Homer et al. 2007). Impervious surface and canopy cover are mapped as continuous percentages; others are considered either present or absent. Forest, canopy cover, and wetlands are unaltered land covers that tend to have neutral or positive effects on water quality and stream biota. Indeed, use of riparian buffers to mitigate harmful effects of development on waterways has become gospel in both science (Mander et al. 1997) and law (Lee et al. 2004), and wetlands are sufficiently effective contaminant sinks that humans are engineering them (Hansson et al. 2005). In contrast, agricultural land cover is the most pervasive cause of stream impairment in the United States, frequently leading to eutrophication (Boesch et al. 2001) and sedimentation (Lenat 1984). Proportionally, urbanization may have even greater deleterious effects on stream systems than agriculture (Paul and Meyer 2001, Poff et al. 2006): streams draining impervious surface areas tend to suffer higher contaminant levels, increased temperature and erosion, and greater hydrologic irregularity than streams draining undeveloped areas (Allan 2004, Walsh et al. 2005).

Many of the derived variables described above were developed as multiple versions differing in their spatial attributes: (1) local, (2) nonspatial accumulated, and/or (3) spatially-explicit accumulated. Local derived variables reflect the region containing or directly adjacent to a survey site. For non-hydrographic variables, such as land use, local variables were calculated in a 30 x 30 meter (3 x 3 pixel) window around the site. For example, the variable “local canopy cover” is the average percent canopy cover on land in nine 10 x 10 meter pixels, with the site location as the central pixel. If the variable was

hydrographic, calculating a local variable involved only stream cells in 30 x 10 meter window (1 upstream and 1 downstream pixel and the site location as the central pixel).

Accumulated variables, both nonspatial and spatially-explicit, consider the characteristics of a stream pixel's upstream basin, as defined by a D8 flow path and flow accumulation raster derived using TauDEM (Tarboton 2014). Nonspatial accumulated variables measure the average value of a variable for the entire upstream basin. For example, nonspatial accumulated canopy cover is the average percent canopy cover in all upstream pixels that flow into a given location. In contrast, spatially-explicit accumulated variables attempt to accommodate the fact that each upstream pixel is unlikely to have the same influence on a particular downstream pixel (Johnson et al. 2007, Van Sickle and Johnson 2008, Peterson et al. 2011, Sheldon et al. 2012). Spatially-explicit versions of the five land covers (forest, wetland, agriculture, canopy cover, and impervious surface) were calculated using the following inverse-distance weighting scheme adapted from Peterson et al. (2011):

$$\%LU = \frac{\sum_{i=1}^n I(k)W_iFA_i}{\sum_{i=1}^n W_iFA_i} \times 100.$$

LU is the land use of a given class (e.g., forest). W_i is the weight given to an upstream pixel, i , according to its distance from the nearest stream, here $(\text{distance}+1)^{-1}$. FA_i is the number of pixels that flow into pixel i (flow accumulation weight). In the case of discretely classified land uses, $I(k)$ is an index equal to one for the pixels classified as the land use of interest and zero for all other pixels. In the case of land uses with continuous scores, $I(k)$ is equal to the continuous value. Thus, I weighted each pixel of those five

land uses by (1) flow path length between it and a stream and (2) the number of pixels that flow into it. Closer land pixels (along flow paths) were more heavily weighted than farther land pixels; land pixels through which more water flows were more heavily weighted than those through which less water flows. I measured distances along TauDEM-derived flow paths between a land use pixel and the closest stream pixel.

For the purposes of this study, I grouped nonspatial and spatially-explicit accumulated variables together as “accumulated derived.” Therefore, three broad variable sets are compared, differing in measurement location and technique: field-measured variables are collected by the MBSS in-situ, local derived variables are interpolated or derived and estimated at or directly adjacent to the survey point, and accumulated derived variables are interpolated or derived and estimated upstream of the survey point (Table 2).

Table 2. Derived variables considered as predictors in generalized dissimilarity models. Local versions characterize conditions directly adjacent to stream pixels (1 x 3 pixels or 3 x 3 pixels); accumulated versions characterize either mean upstream conditions (“accumulated”) or upstream conditions where each pixel is weighted by flow path distance to the stream and number of contributing pixels (“spatial acc.”). All variables were created at or resampled to 10 m.

Derived variable code	Versions	Description/Units/Source
Topographic		
slp	local accumulated	Slope (degrees), NED DEM.
plan	local accumulated	Transverse curvature at cell, perpendicular to flow direction (1/100 elevation units), NED DEM
prof	local accumulated	Longitudinal curvature at cell, parallel to flow direction (1/100 elevation units), NED DEM
Hydrographic		
dem10mp	local	8 Direction flow raster 1= East, 2=SE, etc., TauDEM
dem10mad8	accumulated	Number of 10x10 meter pixels that flow into the pixel, TauDEM
str_len	local accumulated	Length of stream (km), Elmore et al. (2013) map

Derived variable code	Versions	Description/Units/Source
str_den	accumulated	Upstream network density (km/km ²), Elmore et al. (2013) map
confluence_num	accumulated	Number of stream segments that come together on a pixel, Elmore et al. (2013) map.
str_blen	local accumulated	Length of stream burial (km), burial probability classified using NLCD 2001 30 m ISA layer
bp_2001	local	Burial probability (0-1), determined using NLCD 2001 30 m ISA layer
str_bp	accumulated	Burial probability accumulated (0-1), determined using NLCD 2001 30 m ISA layer
str_blen_den	accumulated	Upstream burial density (km/km ²) determined using NLCD 2001 30 m ISA layer
Land Use/Land Cover		
isa	local accumulated	Proportion impervious surface area, 0-1. NLCD 2001 impervious surface data, 30 m
isa_fls_nor	spatial acc.	
cc	local accumulated	Proportion canopy cover, 0-1. NLCD 2001 canopy cover data, 30 m
cc_fls_nor	spatial acc.	
for	local accumulated	Proportion forest presence, 0-1. NLCD 2001 forest data, 30 m, forest = {41,42,43}
for_fls_nor	spatial acc.	
ag	local accumulated	Proportion agriculture presence, 0-1. NLCD 2001 LULC data, 30 m, agriculture = {81,82}
ag_fls_nor	spatial acc.	
wet	local accumulated	Proportion wetland presence, 0-1. NLCD 2001 LULC data, 30 m, wetlands = {90,95}
wet_fls_nor	spatial acc.	
Soils		
sicl	local accumulated	Proportion of soil volume (0-1) that is below 63 μm in texture, SSURGO, 0.6 ha
kfw	local accumulated	Soil erodibility (K value), SSURGO, 0.6 ha
bd	local accumulated	Bulk density indicator of soil compaction (g/cm ³), SSURGO, 0.6 ha
brd	local accumulated	Distance from soil surface to top of bedrock layer (cm), SSURGO, 0.6 ha
ph	local accumulated	Relative acidity or alkalinity of a soil sample (pH units), SSURGO, 0.6 ha
Climate		
sat	local accumulated	Annual mean temperature (°C*10), WorldClim 2.5 arc-minutes, Bioclim variable 1

Community-level modeling

I used community-level models because of their ability to rapidly analyze assemblages with large numbers of species, make use of data for infrequently-recorded

taxa, and extrapolate patterns beyond sampled communities (Ferrier and Guisan 2006, Fitzpatrick et al. 2011, Jones et al. 2013) – all potential boons for management applications. Additionally, community-level models are valuable tools for variable assessment because they assimilate the responses of many species, so conclusions may be less sensitive to atypical species/environment relationships.

Specifically, I compared field-measured, local derived, and accumulated derived environmental predictor variables using generalized dissimilarity modeling (GDM), a community-level, nonlinear matrix regression technique that models pairwise site compositional dissimilarity (beta diversity) as a function of environmental and geographic distance (Ferrier et al. 2007). GDM accommodates two nonlinearities common in large ecological datasets: variation in the rate of compositional turnover along environmental gradients (non-stationarity), and the curvilinear relationship between compositional dissimilarity and environmental/geographic distance (Ferrier et al. 2007, Allan 2004, Dodds et al. 2010).

GDM addresses non-stationarity in rates of species turnover along gradients by using maximum likelihood estimation to fit flexible, positively monotonic I-splines to each predictor variable (Ferrier et al. 2007). By default and to avoid over-fitting, each I-spline has three knots, and the segments between them are modeled as quadratics. The shape of the I-spline indicates the rate of biological turnover at each position along the gradient (Ferrier et al. 2007). The amplitude of the I-spline, quantified by the sum of its coefficients, corresponds to the relative importance of the predictor variable in contributing to biological turnover between pairs of sites, holding all other variables constant (Ferrier et al. 2007). Thus, the I-splines are partial regression fits that provide a

biologically-supported relationship between environmental and geographic distance and compositional dissimilarity (Fitzpatrick et al. 2013). I-splines thereby provide a means to select, weight, and transform environmental predictor variables objectively such that they best represent biological patterns (Ferrier et al. 2007, Leathwick et al. 2011, Williams et al. 2012). GDM accommodates the asymptotic nature of compositional similarity metrics by transforming the scaled relationship between predictors and compositional dissimilarity using a generalized linear model (GLM) with an exponential link function (Ferrier et al. 2002, 2007).

Fitting GDMs requires a site \times taxa table and a corresponding site \times environment table. The first table comprises the response variable, once data are converted into pairwise site biological distances (in this case Bray-Curtis dissimilarity). The second table, once converted into environmental/geographic distances, comprises the predictor variables. GDM uses these data to derive I-splines and apply the GLM transformation. All GDM analyses were performed in R version 3.0.1 (R Core Team 2013) using the GDM package available from R-Forge (<http://r-forge.r-project.org>, Manion et al. 2014). For this research, I converted MBSS abundance data to presence/absence, which is considered more reliable for the un-censused benthic invertebrates (Boward and Friedman 2011). I omitted pairwise comparisons between sites surveyed in different years to minimize the effects of unmeasured yearly environmental variation, and I included geographic distance as a predictor in all models.

I fit a total of 56 GDMs using all available occurrence data from the MBSS. These included separate models for each combination of region (4), taxon (fish, benthic invertebrates), and variable set (seven combinations of three variable groups). To select

from the candidate predictor variables, I first removed highly correlated variables. I tested Pearson and Spearman correlations within each variable set for each region and removed variables in pairs or groups correlated $>|0.7|$, retaining those variables of correlated sets that I considered most biologically relevant (Williams et al. 2012, Austin 2002). I also tested for correlations across variable sets. In the few cases in which variables were correlated across sets (i.e. local derived with accumulated derived, accumulated derived with field-measured, or local derived with field-measured), I retained the variable deemed most biologically relevant. Manually choosing candidate variables in cross-group comparisons could influence conclusions regarding which variable sets are most effective. However, cross-group correlations were rare, and only three variables were omitted as a result: accumulated surface air temperature, accumulated soil erodibility, and flow accumulation (Appendix C).

After removing correlated variables, I tested for statistical significance of predictor variables in each model using a custom backward selection procedure. First, I built a GDM with all candidate predictor variables and removed predictors with a sum of I-spline coefficients equal to zero, as they had no relationship to biological turnover. I re-fit the GDM without the irrelevant predictors (“full model”) and fit a third GDM (“reduced model”) omitting the predictor associated with the least compositional change, i.e. the variable with the lowest sum of I-spline coefficients (Fitzpatrick et al. 2011). I subtracted the deviance explained of the reduced model from the deviance explained of the full model. To reduce computational burdens associated with matrix randomization and numerous model fitting routines, I omitted the variable if the difference was less than two percent of the models’ mean deviance explained. If the difference was greater, I

randomized species relationships 500 times and created 500 GDM null model pairs, one model with the full predictor set and one with the reduced set. If at least 5% of the null model differences in deviance explained were greater than the true difference in deviance explained (i.e. $p < 0.05$, the variable in question was no more explanatory than a random variable), I omitted the variable in question and repeated the procedure for the next least-relevant predictor. Alternatively, if the true difference in deviance explained was among the top 5% of the null differences, I retained the variable and the selection procedure ended (Fitzpatrick et al. 2011).

Evaluation of field-measured and derived variable sets

I compared models within region and taxon, so the only difference between them was predictor variable set, and I evaluated them in three complementary ways: explanatory power, parsimony, and predictive ability. I measured how well models were able to explain the data using deviance explained, the method GDM uses to assess fit. I also used variation partitioning (strictly speaking, deviance partitioning) to determine the amount of deviance uniquely explained by each of the three variable groups for each region-taxon combination (Borcard et al. 1992, Whittaker 1984, Jones et al. 2013). Deviance partitioning quantified the extent to which different variable sets explained redundant or complementary biological information. Briefly, the procedure involved subtraction of deviance explained values from models fit with different variable sets to determine how much deviance explained was attributable to each set of variables individually. For example, the proportion of deviance explained (DE) attributable to field-measured variables alone is:

$$DE_{\text{field measured variables unique}} = DE_{\text{all three variable sets}} - DE_{\text{local and accumulated derived variables}}$$

Unlike deviance explained, Akaike Information Criterion (AIC) penalizes models for additional parameters, thereby assessing model complexity in conjunction with fit (i.e. parsimony, Akaike 1974). I assessed parsimony using AIC weights. To calculate AIC, which GDM does not report, I fit GDMs in R as log linear binomial generalized linear models (GLMs). Biological similarity was a function of environmental distance, with a log link function and binomial observation error (Millar et al. 2011). I extracted log-likelihood (L) from GLM model objects and defined number of parameters (k) as the number of non-zero I-spline coefficients, plus one for the intercept term. Then, I calculated AIC as $(2*k)-(2*L)$, penalizing models with a greater number of parameters. AIC weights, reflecting which model had the most support from a parsimony standpoint, were calculated using the R package MuMIn (Barton 2014).

Beyond being explanatory and parsimonious, models useful for management must also have the ability to make reasonable predictions to unsurveyed locations. I measured predictive ability by assessing how well models fit with training data were able to predict withheld test data. For each region-taxon/variable set combination, I randomly partitioned site pairs 50 times into training (70%) and testing (30%) sets. I fit GDMs using the training data and the variables pre-selected by the backward selection procedure and then used the fitted models to predict compositional dissimilarity to the withheld 30% of site pairs. I assessed predicted compositional dissimilarity in two ways: (1) using Spearman correlations to show the correspondence between predicted and observed dissimilarities, and (2) using median percent error calculations to determine the magnitude of the

difference between predicted and observed dissimilarities. I compared correlations and median percent errors within region-taxon pairs using Kruskal-Wallis rank sum tests followed by multiple comparison tests.

Additionally, because a different number of variables from each group was available to the models, I used chi-square tests to determine whether models with all variables available to them preferentially selected from particular sets.

Management inference: predicting indices of biotic integrity

Stream management decisions are often informed by metrics such as Indices of Biotic Integrity (IBIs, Karr 1981, Southerland et al. 2007). To determine the difference between the environmental variable sets in the context of managerial inference, I used GDM-transformed environmental variables and the ensemble machine learning method random forests (Breiman 2001) to predict two sets of IBIs: those based on field-measured variables and those based on derived variables. I then compared predicted and MDNR-calculated IBIs to assess the disparity between the variable sets from a management perspective.

For each region-taxon pair, I trained GDMs built with field-measured variables only and GDMs built with derived variables only (local and accumulated together, as this is likely the set that managers would use) on 50 random partitions (70%) of survey locations. I used the I-splines to transform predictors from environmental space (I-spline x-axis) into biological space (I-spline y-axis), leveraging GDM's ability to weight environmental gradients so that they best represent biological patterns. I then trained a random forest model on the transformed predictors with observed IBI as a response, used

GDM to transform environmental gradients at the 30% of sites withheld, and predicted IBI for that holdout with the random forest model. I calculated percent error for each pairwise comparison and used Spearman correlations to compare observed and predicted IBIs for each random partition. Both percent errors and Spearman correlations were compared using Welch's t-tests.

RESULTS

Evaluation of field-measured and derived variable sets

Explanatory power

GDM explained between 3.8 and 43.4% of the deviance in compositional turnover, depending on variable set, region, and taxon (Fig. 2). The least explanatory model was built with only local derived variables (L) for Youghiogheny fish. The most explanatory models, with the same deviance explained, were built for Youghiogheny benthic invertebrates with variables from all groups (FAL) and Youghiogheny benthic invertebrates with field-measured and accumulated derived variables (FA). Youghiogheny fish had the greatest range in deviance explained values between models fit with different variable sets (35.7%), followed by Youghiogheny benthic invertebrates (33.1%). Models for the Coastal Plain benthic invertebrates had the smallest range in deviance explained values (10.9%). The average deviance explained for benthic invertebrate and fish models built with derived variables (AL) was 23.0% and 21.2%, respectively. Average deviance explained for benthic invertebrate and fish models built with field-measured variables (F) was 29.6% and 28.2%, and average deviance explained for models selecting from all variable sets (FAL) was 32.7% and 32.6%.

Within each region-taxon pair, the model built with only local derived variables (L) always had the lowest deviance explained, the four models that included field-measured variables (F, FL, FA, FAL) always had the highest deviance explained, and models built with accumulated derived variables (A, AL) were intermediate. For all regions and both taxa, the effect of adding local derived variables to another variable set (e.g., addition of L variables to A, F, or FA sets) was negligible.

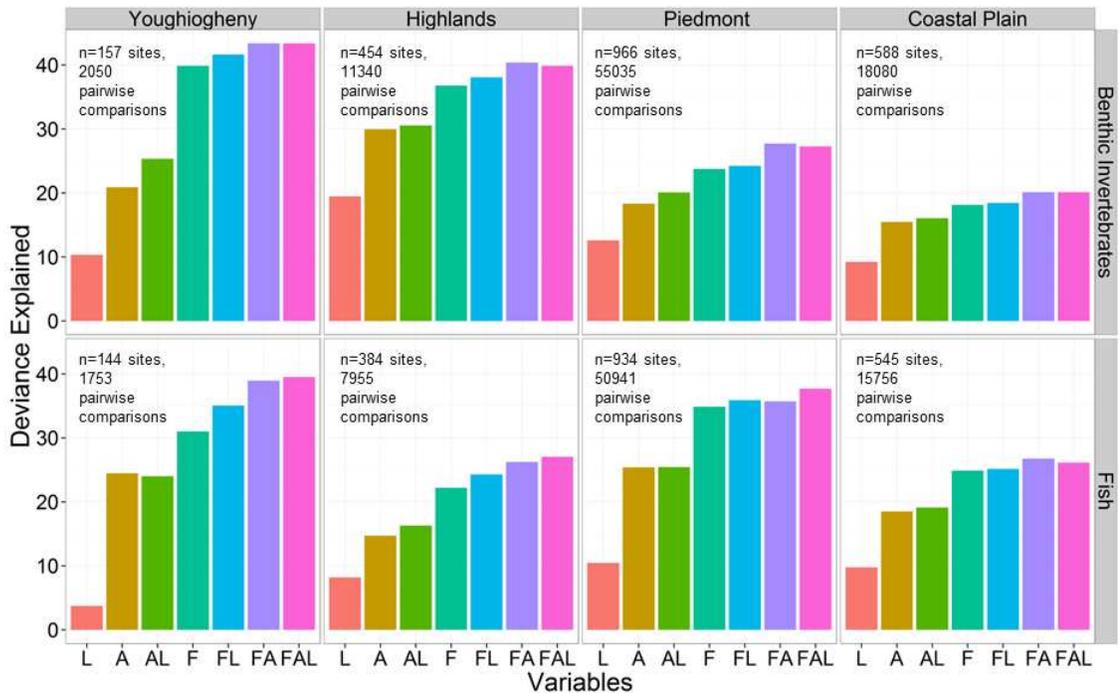


Figure 2. Deviance explained values for generalized dissimilarity models built with all available MBSS data in four regions and for both fish and benthic invertebrates. Each panel is a region-taxon pair. F =field-measured, L = local derived, and A = accumulated derived variables.

Across regions and taxa, local derived variables uniquely explained between zero and 2.0% of the deviance in biological turnover, i.e. almost nothing that the other variable sets did not also explain (Fig. 3). Accumulated derived variables uniquely explained between 1.0 and 4.7% of the deviance, and field-measured variables uniquely explained between 4.1 and 18.0%. Single variable sets (L, A, and F) used in Coastal Plain models tended to uniquely explain a lower proportion of the deviance than variable sets in other models, reflecting the fact that Coastal Plain models in general had low deviance explained values.

Field-measured variables explained the greatest amount of unique deviance in the Youghiogheny region, which usually had higher overall deviance explained values than

the other regions. Of the deviance explained by the Youghiogheny FAL fish model (39.5%), two-fifths (15.5%, red cross, Fig. 3) was uniquely attributable to field-measured variables. Of the deviance explained by the Youghiogheny FAL benthic invertebrate model (43.4%), over two-fifths (18.0%, red triangle, Fig. 3) was uniquely attributable to field-measured variables.

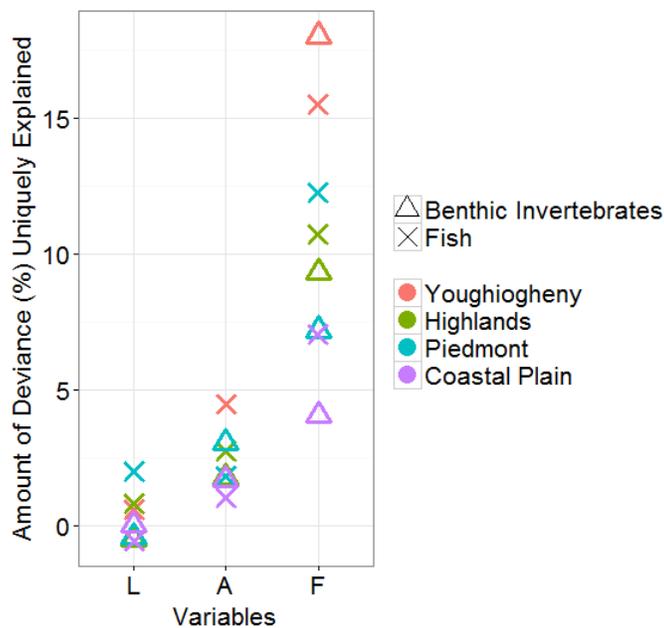


Figure 3. Deviance uniquely explained by each variable set (F =field-measured, L = local derived, A = accumulated derived) for region-taxon pairs. Note that some of the local derived percentages are negative because models were not perfectly nested.

Parsimony

In all region-taxon pairs except Coastal Plain benthic invertebrates, models including field-measured variables as predictors (either FAL models or FA models) were most supported by AIC (Appendix D). In the case of the Coastal Plain benthic invertebrates, the model built with only local derived variables was best supported, likely

because all models were comparatively poor in this region-taxon pair (Fig. 2) and had similar AICs, such that number of parameters weighed heavily in the evaluation.

It is most useful to calculate AIC weights considering only models built with L, A, AL, and F variable sets so that the weights clearly reflect variable set differences. For all region-taxon pairs except the Coastal Plain benthic invertebrates, the model built with field-measured variables had AIC weight >0.98 (Table 3).

Table 3. Generalized dissimilarity model AICs, considering only models built with local derived (L), accumulated derived (A), local and accumulated derived (AL), and field-measured (F) variable sets. Δ AIC shows the difference between the current model and the best model, which is marked with a star (*). AIC_w are AIC weights, and K is the number of parameters.

Models	Youghiogheny				Highlands				Piedmont				Coastal Plain				
	AIC	Δ AIC	AIC_w	K	AIC	Δ AIC	AIC_w	K	AIC	Δ AIC	AIC_w	K	AIC	Δ AIC	AIC_w	K	
Fish	L	2376.1	333.88	0.000	4	9731.0	673.94	0.000	14	65892.2	6579.23	0.000	9	17985.5	1069.5	0.000	17
	A	2157.3	115.07	0.000	20	9399.9	342.81	0.000	9	61527.9	2214.94	0.000	13	17312.3	396.2	0.000	25
	AL	2172.3	130.07	0.000	19	9362.5	305.36	0.000	14	61417.2	2104.26	0.000	12	17290.0	373.9	0.000	28
	F	2042.2*	0.00	1.000	12	9057.1*	0.00	1.000	18	59312.9*	0.00	1.000	17	16916.1*	0.0	1.000	27
Benthic Inverts	L	1653.4	25.88	0.000	8	7901.1	101.01	0.000	15	36853.1	175.12	0.000	14	9593.0*	0.0	0.974	10
	A	1637.7	10.17	0.006	14	7825.7	25.67	0.000	12	36704.2	26.21	0.000	17	9601.9	8.9	0.012	22
	AL	1636.2	8.60	0.013	16	7818.4	18.39	0.000	11	36692.7	14.68	0.001	22	9602.2	9.2	0.010	21
	F	1627.6*	0.00	0.981	17	7800.1*	0.00	1.000	15	36678.0*	0.00	0.999	26	9603.6	10.6	0.005	28

Predictive ability

Fitted relationships between predicted and observed compositional dissimilarity for all variable sets, taxa, and regions were close to the one-to-one line (Appendix E). However, Kruskal-Wallis rank sum tests showed that within each region-taxon pair, there were significant differences in how effectively models fit with different variable sets predicted Bray-Curtis dissimilarities between withheld site-pairs (Fig. 4a, b). Within region-taxon pairs, the only models that did not have significantly different Spearman correlations were built with only derived variable groups (A vs. L models and A vs. AL models) or both included field-measured variables (F vs. FL, FL vs. FAL, etc., Appendix F). Similarly, except in the case of two Coastal Plain benthic invertebrate model comparisons (A vs. F and AL vs. F), the only non-significant differences in percent error were between models that both included or both did not include field-measured variables (Appendix F).

Overall, the pattern in predictive ability was identical to the pattern in explanatory power (Fig. 2): models built with only local derived variables had the lowest predictive ability (average Spearman correlation between observed and predicted Bray-Curtis dissimilarities was 0.30, average median percent error was 18.3 %), followed by models built with accumulated derived variables (A and AL models both had an average $\rho = 0.47$ and an average median percent error of 16.4%). Models built with field-measured variables had the best predictive ability (F models had average $\rho = 0.54$, average median percent error = 15.3%).

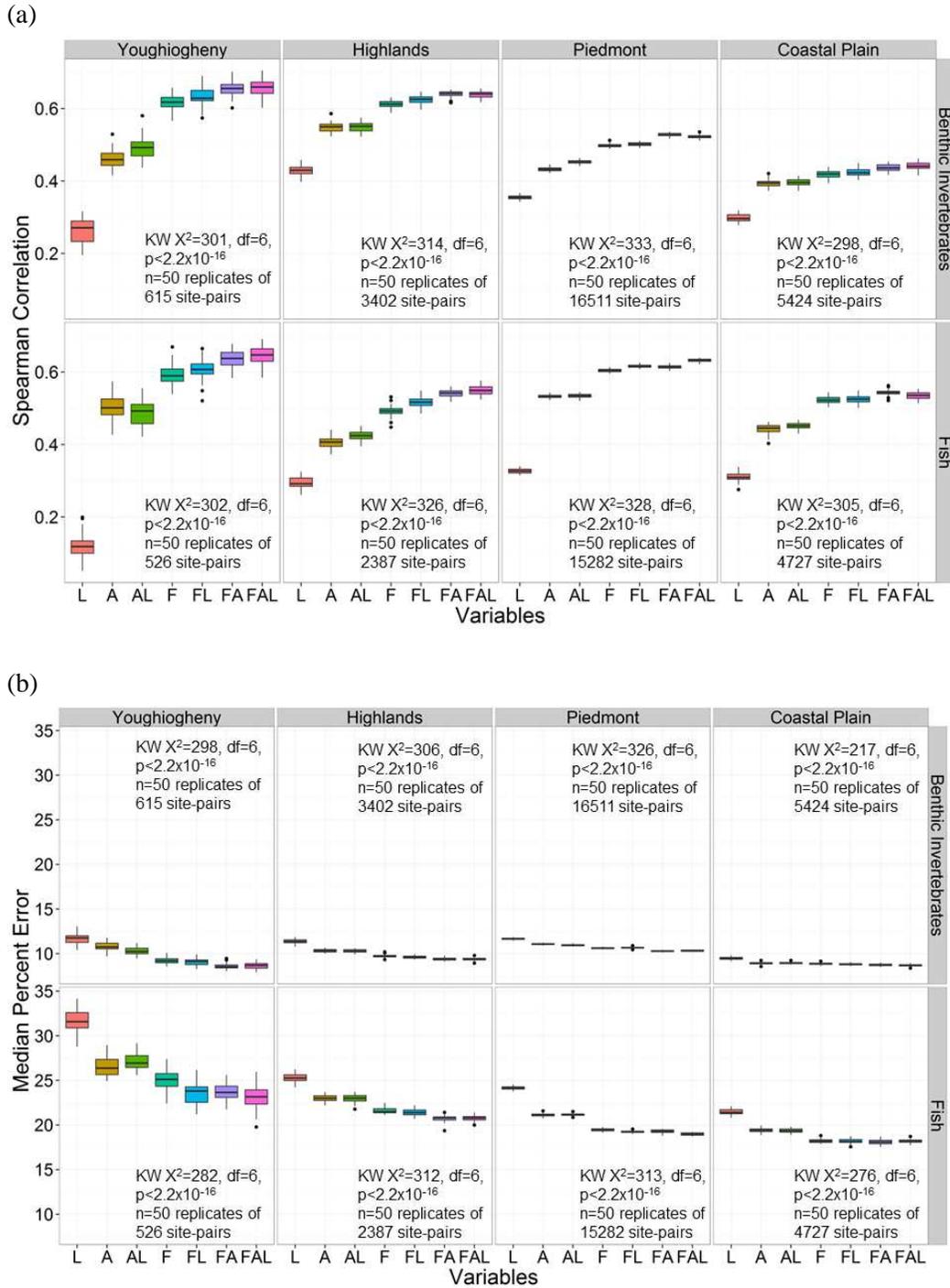


Figure 4. Boxplots of (a) Spearman correlations between observed and predicted Bray-Curtis dissimilarities, where each box summarizes 50 correlation coefficients, and (b) percent error of predicted Bray-Curtis dissimilarities, where each box summarizes 50 median percent errors. F =field-measured, L = local derived, and A = accumulated derived variables. For significant differences between boxes within region-taxon pairs, see Appendix F.

Important individual variables

Once correlated variables were removed (Appendices B, C) and the backward selection process was completed, the final GDMs retained between one and 16 environmental predictor variables (Appendix G). For benthic invertebrate models, the five most common environmental predictors were related to land use, substrate, and gradient, and included weighted impervious surface area (*isa_fls_nor*, 12 models), accumulated derived bedrock depth (*brd_acc*, 12 models), field-measured stream gradient (*ST_GRAD*, 12 models), local derived burial probability (*bp_2001*, 11 models), and field-measured riffle quality (*RIFFQUAL*, 11 models, Appendix H). The most common environmental predictors of fish turnover were related to network position, substrate, temperature and gradient, and included field-measured stream gradient (*ST_GRAD*, 14 models), local derived surface air temperature (*sat_loc*, 13 models), accumulated derived bedrock depth (*brd_acc*, 13 models), flow accumulation (*dem10mad8*, 12 models), and field-measured discharge (*DischargeCFS*, 12 models, Appendix H).

In models for which field-measured, accumulated derived, and local derived variables were all candidates (FAL models), field-measured conductance (*COND_FLD*) and weighted impervious surface area (*isa_fls_nor*) were most strongly related to benthic invertebrate compositional turnover across regions (Fig. 5a). Field-measured stream gradient (*ST_GRAD*) was most strongly related to fish compositional turnover and was chosen as a predictor in all regions except the Coastal Plain (Fig 5b).

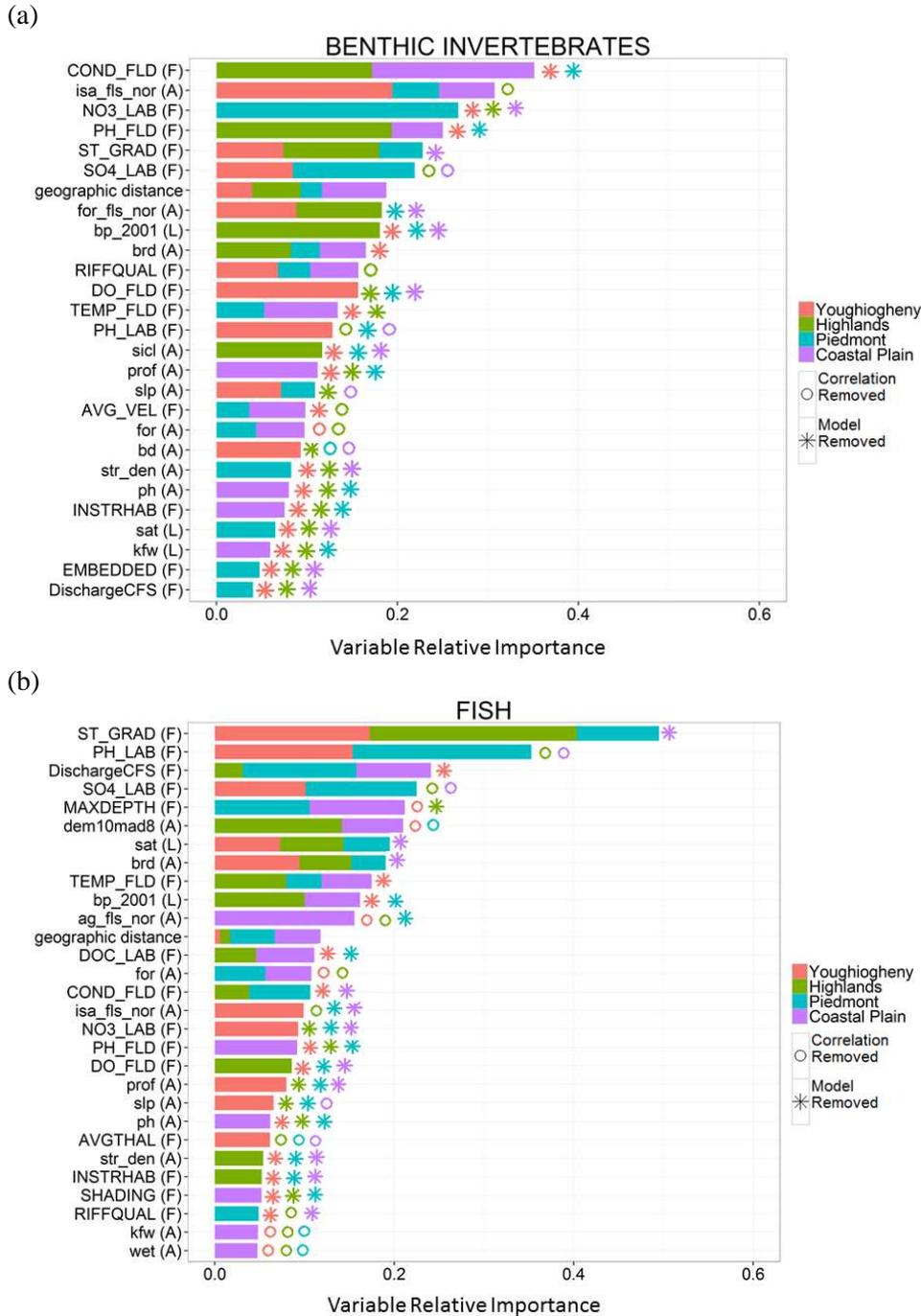


Figure 5. Relative importance of predictor variables selected for (a) benthic invertebrate and (b) fish generalized dissimilarity models when field-measured (F), accumulated derived (A), and local derived (L) variables were candidates. Relative importance is quantified as the sum of I-spline coefficients and normalized so that the length of the bars representing a single region sums to one. Symbols next to the bars show whether unselected variables were removed as a result of (●) correlation analysis or (*) by backward selection.

Though the field-measured variables as a set were almost always superior to derived variables in terms of explanation, prediction, and parsimony, chi-square tests showed that field-measured variables were not chosen by the backward selection procedure out of proportion to their availability (Table 4). This is true both when derived variables were considered as a single category (L + A), in which case there were many more derived than field-measured variables available (Table 4a), and also when derived variables were separated, in which case the Highlands and Coastal Plain had approximately the same number of variables in each set and the Youghiogheny and Piedmont had a greater number of field-measured variables (Table 4b).

Table 4. Chi-square test results showing whether predictors from particular variable sets were selected for inclusion in generalized dissimilarity models out of proportion to their availability. Variable sets tested were (a) AL vs. F and (b) L vs. A vs. F, where F = field-measured, L = local derived, and A = accumulated derived variables. Significant results are starred (*).

Region	Taxon	Chosen Variables in FAL models				Number of Candidate Variables				Chi Square Statistic	p-value
		L	A	AL	F	L	A	AL	F		
(a)											
Youghiogheny	Fish			5	5			27	19	0.312	0.577
Youghiogheny	Benthic Inverts			4	5			27	19	0.754	0.385
Highlands	Fish			5	7			26	14	2.872	0.090
Highlands	Benthic Inverts			4	3			26	14	0.190	0.663
Piedmont	Fish			3	8			27	17	5.392	0.020*
Piedmont	Benthic Inverts			6	8			27	17	2.022	0.155
Coastal Plain	Fish			7	6			29	15	0.842	0.359
Coastal Plain	Benthic Inverts			6	6			29	15	1.352	0.245
(b)											
Youghiogheny	Fish	1	4		5	15	12		19	2.493	0.288
Youghiogheny	Benthic Inverts	0	4		5	15	12		19	4.540	0.103
Highlands	Fish	2	3		7	14	12		14	3.119	0.210
Highlands	Benthic Inverts	1	3		3	14	12		14	1.367	0.505
Piedmont	Fish	1	2		8	15	12		17	5.659	0.059
Piedmont	Benthic Inverts	1	5		8	15	12		17	4.589	0.101
Coastal Plain	Fish	1	6		6	15	14		15	4.052	0.132
Coastal Plain	Benthic Inverts	1	5		6	15	14		15	3.592	0.166

Management inference: predicting indices of biotic integrity

Field-measured variables were not consistently better predictors of IBIs than derived variables (Fig. 6a, b). Rather, predictions of IBIs using derived variables had a lower mean percent error and were more strongly correlated with MDNR-calculated IBIs for five of the eight region-taxon pairs. Without a correction for multiple comparisons, IBI predictions using field-measured variables were significantly more correlated with MDNR-calculated IBIs for three region-taxon pairs and predictions using derived variables were significantly more correlated with MDNR-calculated IBIs for two. Conversely, use of field-measured variables yielded predictions with significantly lower mean percent errors for two region-taxon pairs, and derived variables yielded predictions with significantly lower mean percent errors for three. With a Bonferroni correction, each variable set produced predictions with a significantly higher mean Spearman correlation for one region-taxon pair a significantly lower mean percent error for two region-taxon pairs. Neither derived nor field-measured variables showed regional or taxon-based patterns in ability to predict IBIs.

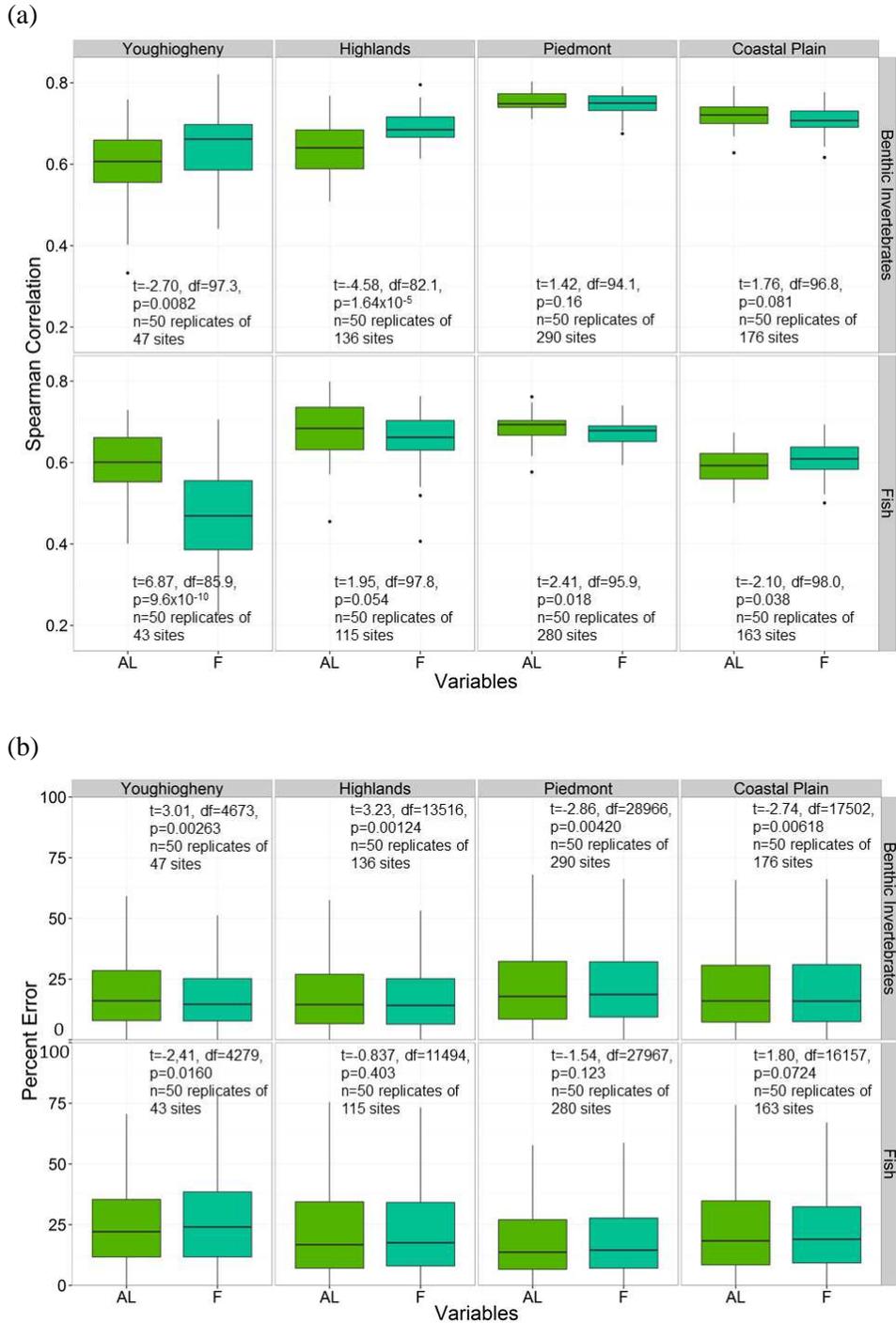


Figure 6. Boxplots of (a) Spearman correlations between MDNR-calculated and predicted IBIs, where each box summarizes 50 correlation coefficients, and (b) percent errors of predicted IBIs, where each box summarizes errors in all replicates and outliers have been suppressed for visualization purposes. F =field-measured and AL = accumulated and local derived variables, transformed using GDM prior to random forest modeling.

DISCUSSION

If statistical models are to be used to map patterns of biodiversity, they must be fit using derived, full-coverage environmental variables (Austin 2002). However, derived variables are often only indirectly related to taxon physiology (i.e. distal), and are therefore predicted by theory to be less effective than proximal variables, which are typically field-measured (Austin 1980, Franklin 1995, Austin 2002). This research had three main goals: to compare stream biodiversity models fit using field-measured and/or derived variables, to identify the major environmental drivers of Maryland fish and invertebrate community turnover, and to determine whether modeling with field-measured versus derived variables leads to different management inference. Results showed that field-measured environmental variables almost always produced models that were more explanatory, had greater predictive ability, and were more parsimonious than derived variables. Field-measured variables also explained a larger amount of unique deviance. However, though performance of derived variables as a group was comparatively poor, several of the individual derived predictors were among the most important, and the predictive superiority of field-measured variables did not persist when the dimensionality of the data was reduced to management-relevant indices.

Evaluation of field-measured and derived variable sets

Explanatory power

The amount of deviance explained by the models developed for this study is on par with that reported by other studies employing GDM in stream networks (Leathwick et al. 2011, Snelder et al. 2012). Leathwick et al.'s (2011) models explained an average of 18.5% of the deviance in benthic invertebrate community turnover and 16.3% of

deviance in fish community turnover in New Zealand. They used only derived variables, including several measures of temperature, flow, slope, and habitat, and considered variables at multiple spatial scales. Snelder et al. (2012) used GDM to model benthic invertebrate and fish beta diversity in France, and explained 41% and 35% of deviance, respectively. They used a relatively modest set of derived variables which included site and watershed slope and temperature, watershed precipitation, site altitude, and several physical and geographical watershed characteristics.

Snelder et al.'s (2012) comparatively high deviance explained values may in part be attributable to the particularly long environmental gradients across continental France (Murphy 2010, Snelder et al. 2012), while Leathwick et al.'s (2011) comparatively low deviance explained values were possibly related to the omission of human land use variables (e.g., impervious surface). However, the fact that deviance explained values for the Maryland freshwater taxa models were lower than Snelder et al.'s (2012) and higher than Leathwick et al.'s (2011) could also be attributable to the different sample sizes (Guisan and Zimmermann 2000). Deviance explained values tend to decrease with a greater number of observations and a smaller number of parameters (Guisan and Zimmermann 2000).

For this study, deviance explained values within region-taxon pairs are directly comparable, as models with different variable sets were fit using the same site pairs. The consistent increase in deviance explained when models included field-measured variables supports the hypothesis that field-measured variables yield more explanatory models than derived variables. Indeed, the observed higher deviance explained of models built with field-measured variables is especially compelling because the group of derived variables

used in this study was comparatively large and sophisticated, and represents an extension of variables previously used in community-level stream biodiversity models (e.g., Leathwick et al. 2011, Snelder et al. 2012). In particular, the base stream map is more realistic than the NHD (Elmore et al. 2013), and the stream burial metrics (Elmore and Kaushal 2008) and land use inverse weighting (Peterson et al. 2011) are relatively novel.

The theoretical explanation for the greater explanatory power of field-measured variables is that they tend to be more physiologically relevant and therefore more closely related to taxon distributions. In contrast, most derived variables are relevant to biota via potentially varying relationships with more proximal drivers. The nature of stream networks might also have contributed to the higher explanatory power of field-measured variables: the dendritic shape and integrative nature of streams makes deriving variables challenging. Specifically, interpolation of stream measurements must take into consideration directionality (Peterson et al. 2013), confluence locations (Benda et al. 2004), and connectivity (Grant et al. 2007); myriad small, ephemeral, or intermittent channels have been traditionally difficult to map using remote sensing tools alone (Elmore et al. 2013); and predictor variables must characterize both upstream and local drivers of biotic distributions (Morley and Karr 2002, Kratzer et al. 2006, Urban et al. 2006, Stanfield and Kilgour 2013). The importance of environmental influences on biotic distributions at multiple scales, in particular, is highlighted by the poor explanatory power of local derived variables compared to accumulated derived variables.

Parsimony

AIC demonstrated that not only do field-measured variables result in more explanatory models, they also typically produce more parsimonious models. The only exception was the Coastal Plain benthic invertebrates, for which the model fit with local derived variables (L) had the greatest AIC support. This model had lower deviance explained and predictive ability than its companion models, but also had many fewer parameters – ten compared to other models with more than twice that number – and AIC and deviance explained values among the models were relatively close.

The Coastal Plain is the region of Maryland with least relief (Reger and Cleaves 2008). As a result, it is the region where the stream map, which depended heavily on flow direction modelling based on a hydrologically-corrected DEM, is likely to be least accurate. Map inaccuracies make drainage areas more difficult to identify and accumulated derived variables less reliable. Additionally, the Coastal Plain encompasses some of the most urbanized areas in Maryland. The combination of high impervious surface cover and elevated hydrological sensitivity to imperviousness (Utz et al. 2011) likely results in greater surface runoff and increased flashiness (Paul and Meyer 2001, Walsh et al. 2005), and field-measured variables, which represent only the moment of sampling, are not likely to characterize extremes. With both accumulated derived and field-measured variables potentially explaining less biological turnover in the Coastal Plain, local variables could explain comparatively more. This logic does not extend to Coastal Plain fish models, for which a larger number of local derived variables was selected and the accumulated derived and field-measured variables performed comparatively better.

Predictive ability

In the context of management, conservation, and restoration planning, the ability of models to make predictions to unsurveyed locations is among their most important attributes, as reliable predictions can ameliorate expensive, time consuming monitoring programs that are the cornerstone of many contemporary management decisions (Kristensen et al. 2012). Predictive ability of the models in this study exhibited the same pattern as deviance explained, with models including field-measured variables being most predictive. The average correlations between observed and predicted Bray-Curtis dissimilarities produced by models in this study were slightly lower than those achieved by other stream modeling studies ($r = 0.37$ to 0.66 for Kristensen et al. [2012], $r = 0.64$ for Usio [2007], $\rho_L = 0.30$, $\rho_{A,AL} = 0.47$ and $\rho_F = 0.54$ for this study), potentially because others used Pearson correlations to assess classification or single species predictions rather than Spearman correlations to assess community dissimilarity. Still, correlation coefficients for A, AL, and F models were sufficiently high and percent errors sufficiently low that these models could be useful in helping to prioritize conservation/restoration and in targeting monitoring efforts. Conversely, models built with only local derived variables may not be as useful.

Important individual variables

In addition to comparing field-measured and derived variables as sets, this research illuminated some of the individual environmental characteristics that drive fish and benthic invertebrate turnover in Maryland. Despite the consistent superiority of models built with field-measured variables in deviance explained, AIC, and predictive

ability, some individual derived variables also performed well. Derived variables were among the most common chosen by the backward selection procedure, and a derived variable was the second most important predictor of benthic invertebrate community turnover.

The mixture of field-measured and derived variables most frequently selected by GDM corresponds to the four regimes identified by Melles et al. (2014) as “critical to aquatic ecosystem functioning and diversity:” the flow regime, the temperature regime, the nutrient regime, and the sediment regime. In terms of variables in this study, flow regime was characterized by gradient and hydrographic network position measures, temperature regime by derived temperature, and sediment regime by soils and substrate variables. Land use characterized all four regimes. The differences between variables chosen for the greatest number of fish and benthic invertebrate models were slight, but network position and temperature were chosen more for fish models, while land use was chosen for more invertebrate models.

Network position (i.e. flow accumulation, dem10mad8) may have been chosen for a greater number of fish models than benthic invertebrate models because most fish, unlike benthic invertebrates (Bilton et al. 2001) are unable to disperse across land at any stage in their life cycle. Additionally, fish are more likely to be limited by physical channel size, given their greater mass. One would expect temperature and land use, however, to be frequent predictors for turnover in both taxa. Both Snelder et al. (2012) and Leathwick et al. (2011) identified temperature (mean annual watershed temperature and segment air temperature, respectively) as a variable that strongly contributed to benthic macroinvertebrate classifications, and Hawkins et al. (2000) concluded that the

importance of latitude, longitude, and elevation in their models suggested temperature as a main driver of invertebrate composition. From a mechanistic standpoint, Anderson and Cummins (1979) found that temperature influenced benthic invertebrate metabolism, food quality and quantity. Likewise, land use has been found to be an important influence on fish habitat via its effects on nutrients, contaminants, hydrology, sedimentation, and cover (Allan 2004, Weijters et al. 2009, Pease et al. 2011).

Though they were not among the most frequently selected variables, however, temperature was still a relatively common predictor of invertebrate turnover, and land use was a relatively common predictor of fish turnover. Local derived surface air temperature was chosen for five benthic invertebrate models and TEMP_FLD, the field-measured version, was chosen for eight (Appendix H). Derived weighted impervious surface area was chosen for seven fish models. These variables may not have been selected more frequently because they are an imperfect approximation true conditions rather than because the environmental characteristic they attempt to represent is not important: air temperature only approximates stream temperature (Mohseni and Stefan 1999, Hilderbrand et al. 2014), a single temperature measurement only approximates the regime, and 2001 data from the National Land Cover Database only approximates cover at the time of biotic sampling.

In addition to the quantifying the number of models for which predictors were chosen, relative variable importance can be measured by summing normalized I-spline coefficients across regions (Fig. 5, Fitzpatrick et al. 2011). Considering only FAL models, the majority of the most important predictors of community turnover were field-measured. For benthic invertebrates, the three most important variables were field-

measured conductance (COND_FLD), weighted accumulated derived impervious surface area (isa_fls_nor), and field-measured nitrate (NO3_LAB). For fish, field-measured stream gradient (ST_GRAD), field-measured pH (PH_LAB), and field-measured stream discharge (DischargeCFS) were most important. Gradient and pH were characterized by both field-measured and derived variables; that models selected the field-measured versions provides additional evidence for the overall superiority of the field measurements and corroborates the results of the model comparisons.

Derived variables can also be important predictors, however, even when they are “competing” with field-measured variables in models. For example, weighted impervious surface area had the second greatest sum of I-spline coefficients of variables included in benthic invertebrate models, corroborating published findings that impervious surface is a strong predictor of mid-Atlantic stream assemblages (King et al. 2005, Utz et al. 2009), sensitive aquatic species (Stranko et al. 2008), and “stream health” more generally (Goetz et al. 2004).

Nonetheless, it is important to note that variable importance can be contingent on gradient length (Murphy 2010). For example, sites in the Youghiogheny region are almost all highly forested, resulting in a short forest gradient (Homer et al. 2007). If forest cover does not emerge as an important predictor in the Youghiogheny, it could be because there is little forest cover turnover. Additionally, variable importance can be affected by the variable selection routine, and a backward selection procedure based on I-splines fit holding all other variables constant ignores variable interactions. If a predictor’s influence on biotic distributions is in part dependent on its association with

another predictor (i.e. sedimentation and nutrient/contaminant addition [Lemly 1982, Magbanua et al. 2013]), the predictor may be erroneously eliminated.

Management inference: predicting indices of biotic integrity

When modeling IBIs instead of community dissimilarity, the pervasive superiority of the group of field-measured variables disappeared. Depending on the region and taxon and with no discernable trend, field-measured variables were better, worse, or equivalent to derived variables in their ability to predict IBIs. To the extent that management decisions rely on IBIs, these results suggest that the difference between field-measured and derived variables would not have an appreciable effect on management inference. With mean Spearman correlations between MDNR-calculated and predicted IBIs ranging from 0.47 (F variables, Youghiogheny fish) to 0.76 (AL variables, Piedmont benthic invertebrates) and mean percent errors from 20.7 % (F variables, Youghiogheny benthic invertebrates) to 35.9 % (F variables, Youghiogheny fish), both models built with field-measured variables and with derived variables could be useful in informing management decisions. In the future, it is possible that models could yield even higher correlations between calculated and predicted IBIs if modeled regions corresponded with the regions for which Maryland IBIs were calibrated (including both warmwater and coldwater Highlands for Maryland fish IBIs, etc., Southerland et al. 2005).

In a similar study that modeled a management-relevant metric, Kristensen et al. (2012) compared the abilities of field-measured (“in-stream”) and derived (“cost-effective GIS-derived”) variables to predict the occurrence of fish assemblages in Danish streams. Kristensen et al.'s (2012) derived variables were comprised of only land use

variables and measured at three spatial scales; their field-measured variables were comparable to those in this study. Though Kristensen et al. (2012) found that models including field-measured variables produced a slightly higher number of correct classifications, they concluded that cost-effective derived variable models were adequate for targeting management efforts. Overall, results were similar: when predicting data aggregated to a management-relevant scale, derived variables performed similarly to field variables.

Future model improvements

How can the explanatory power, predictive ability, and parsimony of future stream biodiversity models be improved? The results suggest that the inclusion of field-measured variables, or perhaps derived variables that better approximate field-measured variables, would have a positive effect on each of these metrics. Additionally, models could likely be improved with the inclusion of candidate variables reflecting past land use, temporal trends, or extreme conditions (Harding et al. 1998, Zimmermann et al. 2007), as well as modeled stream temperature (e.g., Hilderbrand et al. 2014). Previous research has considered multiple methods of weighting upstream land use (e.g., Peterson et al. 2011) and suggested that weighting schemes should consider in-stream as well as over-land flow distance (Van Sickle and Johnson 2008).

Beyond inclusion of additional covariates, models could be improved with increased taxonomic resolution: benthic invertebrates in this study were identified to family or genus, but species within those aggregations could respond to different environmental drivers (Hawkins et al. 2000) and even within species, individuals are

likely to exhibit varying tolerances (Clark et al. 2011). Multiple visits to survey sites would be more likely to characterize typical assemblages, as fish and benthic invertebrates are mobile taxa and may change locations associated with juvenile dispersal (Schlosser 1998, Jackson et al. 1999, Verberk et al. 2008), seasonal migration (Todd and Rabeni 1989, Brönmark et al. 2008), daily cycles (Schloss and Haney 2006, Kobler et al. 2012), or within a season and without clear, predictable pattern (Macneale et al. 2005, Belica and Rahel 2008). Additionally, incorporation of dispersal abilities (Nekola and White 1999, Grant et al. 2007) and biotic interactions (Hutchinson 1957, Araújo and Rozenfeld 2013), frequently neglected in biodiversity models (Pearson and Dawson 2003), could improve model explanatory and predictive ability.

Conclusions

Collectively, this research represents one of the few empirical tests of the proximal/distal theory in stream systems. It elucidates some of the drivers of stream beta diversity in Maryland and provides a basis for further, predictive Maryland stream biodiversity modeling. Field-measured variables were superior to derived variables in explanation and prediction of fish and benthic invertebrate community turnover, and they almost always produced more parsimonious models. However, some derived variables were also important, and there was little difference between the variable sets in terms of prediction of aggregated, management-relevant indices of biotic integrity. The results broadly corroborate the theory that proximal variables are more robust predictors than distal predictors, but also support the use of biodiversity models built with derived variables in IBI-based management decisions.

ADDENDUM

I believe this work will be a useful addition to Maryland stream biodiversity research and the broader predictor variable literature. My largest struggle in completing it was with the optimization algorithm *DynamicFOAM* (the Dynamic Framework for Occurrence Allocation in Metacommunities). *DynamicFOAM* uses predictions from an alpha diversity model, predictions from a beta diversity model (e.g., GDM), a value of gamma diversity, and available community composition data to predict biotic composition at all locations in the study region (Mokany et al. 2011). I hoped to apply *DynamicFOAM* because I felt that predictions of community composition would be a useful addition to predictions of compositional turnover from a management standpoint. However, *DynamicFOAM* had never been applied to a dendritic system before, and results, when obtained, were nearly nonsensical. It is unclear whether this was a failure in the input files or the algorithm, but it was impossible to troubleshoot thoroughly from a graphical user interface. All *DynamicFOAM* results were omitted from the final document.

Another struggle was conceptually to unite the proximal/distal theory with the reality of field-measured and derived variable sets. It is true that these are not exactly the same, but I believe that field-measured and derived variables represent the most management-relevant analogs to proximal and distal variables, and that focusing on them provided more useful results in terms of model assessment and future variable derivation efforts.

Were I to begin this project again, I would make several small adjustments in the methods. First, I would use multiple community-level modeling techniques to assure that

the results demonstrate true differences between variable sets rather than an artifact of the modeling framework. I would also not have included Euclidean distance as a predictor, as it is likely less useful in streams than network distance (Rouquette et al. 2013). I would have assessed variable correlations all at once and across the entire state rather than by variable group and within region, so that all models would have started with the same candidate variable pool and been more comparable. The drawback of a comprehensive correlation assessment is that correlations could be regionally-dependent, but interpretation would have been simpler. I would consider partitioning beta diversity into its species turnover and nestedness components, which would allow assessment of the underlying nature of community turnover (Baselga 2010). Finally, I would explore options for variable selection techniques beyond backward selection, which is not an optimal procedure (Duarte Silva 2001). There are too many variables to try all possible subsets (Hocking 1976), but perhaps Furnival and Wilson's (1974) branching and bounding algorithm could be applied in a GDM framework.

I do not expect that these improvements would change the conclusions of this research because I do not believe that they would alter the directional relationships between models built with different variable sets. Rather, they could refine understanding of mechanism, facilitate interpretation, and increase the total explanatory and predictive power of the models.

APPENDICES

Appendix A. Fish considered native and non-native in analysis regions, based on the USGS database for Nonindigenous Aquatic Species (United States Geological Survey 2014), NatureServe (2014), Southerland et al. 2005, Stranko et al. 2010, and expert advice. “1” denotes native and “0” denotes non-native.

Species	Youghiogheny	Highlands	Piedmont	Coastal Plain
ALEWIFE	0	1	1	1
AMERICAN BROOK LAMPREY	0	0	1	1
AMERICAN EEL	0	1	1	1
AMERICAN SHAD	0	0	1	1
ATLANTIC NEEDLEFISH	0	0	1	1
BANDED DARTER	0	0	0	0
BANDED KILLIFISH	0	1	1	1
BANDED SUNFISH	0	0	0	1
BAY ANCHOVY	0	0	0	0
BLACK CRAPPIE	1	0	0	0
BLACKBANDED SUNFISH	0	0	1	1
BLACKNOSE DACE	1	1	1	1
BLUE RIDGE SCULPIN	0	1	1	1
BLUEBACK HERRING	0	0	1	1
BLUEGILL	1	0	0	0
BLUESPOTTED SUNFISH	0	0	1	1
BLUNTNOSE MINNOW	1	1	1	1
BROOK TROUT	1	1	1	1
BROWN BULLHEAD	1	1	1	1
BROWN TROUT	0	0	0	0
CENTRAL STONEROLLER	1	1	1	1
CHAIN PICKEREL	0	1	1	1
CHANNEL CATFISH	1	0	0	0
CHECKERED SCULPIN	0	1	1	1
COMELY SHINER	0	1	1	1
COMMON CARP	0	0	0	0
COMMON SHINER	1	1	1	1
CREEK CHUB	1	1	1	1
CREEK CHUBSUCKER	1	1	1	1
CUTLIP MINNOW	0	1	1	1
CUTTHROAT TROUT	0	0	0	0
EASTERN MOSQUITOFISH	0	0	1	1
EASTERN MUDMINNOW	0	1	1	1

Species	Youghiogheny	Highlands	Piedmont	Coastal Plain
EASTERN SILVERY MINNOW	0	1	1	1
FALLFISH	0	1	1	1
FANTAIL DARTER	1	1	1	1
FATHEAD MINNOW	0	0	0	0
FLATHEAD CATFISH	0	0	0	0
FLIER	0	0	0	1
GIZZARD SHAD	0	1	1	1
GLASSY DARTER	0	0	1	1
GOLDEN REDHORSE	0	0	1	0
GOLDEN SHINER	1	1	1	1
GOLDFISH	0	0	0	0
GREEN SUNFISH	1	0	0	0
GREENSIDE DARTER	1	0	0	0
INLAND SILVERSIDE	0	0	1	1
IRONCOLOR SHINER	0	0	0	1
JOHNNY DARTER	1	0	0	0
LARGEMOUTH BASS	1	0	0	0
LEAST BROOK LAMPREY	0	0	1	1
LOGPERCH	0	0	1	1
LONGEAR SUNFISH	1	0	0	0
LONGNOSE DACE	1	1	1	1
LONGNOSE GAR	0	0	1	1
MARGINED MADTOM	0	1	1	1
MOTTLED SCULPIN	1	1	1	1
MUD SUNFISH	0	0	1	1
MUMMICHOG	0	0	0	1
NORTHERN HOGSUCKER	1	1	1	1
ORIENTAL WEATHERFISH	0	0	0	0
PEARL DACE	1	1	1	1
PIRATE PERCH	0	0	1	1
POTOMAC SCULPIN	0	1	1	0
PUMPKINSEED	1	1	1	1
QUILLBACK	0	1	1	1
RAINBOW DARTER	1	1	1	0
RAINBOW TROUT	0	0	0	0
REDBREAST SUNFISH	0	1	1	1
REDEAR SUNFISH	0	0	0	0
REDFIN PICKEREL	0	0	1	1
RIVER CHUB	1	1	1	1

Species	Youghiogheny	Highlands	Piedmont	Coastal Plain
ROCK BASS	1	0	0	0
ROSYFACE SHINER	1	1	1	1
ROSYSIDE DACE	0	1	1	1
SATINFIN SHINER	0	1	1	1
SEA LAMPREY	0	1	1	1
SHIELD DARTER	0	1	1	1
SHORTHEAD REDHORSE	0	1	1	1
SILVERJAW MINNOW	1	1	1	1
SMALLMOUTH BASS	1	0	0	0
SPOTFIN SHINER	0	1	1	1
SPOTTAIL SHINER	0	1	1	1
STONECAT	1	0	0	0
STRIPEBACK DARTER	0	1	1	1
STRIPED BASS	0	1	1	1
STRIPED SHINER	1	0	0	0
SWALLOWTAIL SHINER	0	1	1	1
SWAMP DARTER	0	0	1	1
TADPOLE MADTOM	0	0	1	1
TESSELLATED DARTER	0	1	1	1
WALLEYE	0	0	0	0
WARMOUTH	0	0	0	0
WHITE CATFISH	0	1	1	1
WHITE CRAPPIE	1	0	0	0
WHITE PERCH	0	1	1	1
WHITE SUCKER	1	1	1	1
YELLOW BULLHEAD	1	1	1	1
YELLOW PERCH	1	1	1	1

Appendix B. Intra-group Pearson and Spearman variable correlations $> |0.7|$. Outlines delineate groups of correlated pairs that have variables in common. Chosen variables are in italicized bold font, others were omitted.

Field-measured variables

Youghiogheny, Field-measured Variables		Piedmont, Field-measured Variables	
<i>COND_FLD</i>	COND_LAB	AVGWID	<i>DischargeCFS</i>
<i>RIFFQUAL</i>	EPI_SUB	AVGTHAL	AVG_WID
MAXDEPTH	<i>POOLQUAL</i>	POOLQUAL	VEL_DPTH
<i>AVGTHAL</i>	MAXDEPTH	<i>MAXDEPTH</i>	VEL_DPTH
MAXDEPTH	<i>VEL_DEPTH</i>	<i>MAXDEPTH</i>	POOLQUAL
ANC_LAB	<i>PH_LAB</i>	AVGTHAL	<i>MAXDEPTH</i>
AVGWID	<i>DischargeCFS</i>	VEL_DPTH	<i>DischargeCFS</i>
Highlands, Field-measured Variables		AVGTHAL	<i>DischargeCFS</i>
<i>PH_FLD</i>	PH_LAB	AVGTHAL	VEL_DPTH
<i>PH_FLD</i>	ANC_LAB	POOLQUAL	AVGTHAL
<i>COND_FLD</i>	COND_LAB	EPI_SUB	<i>INSTRHAB</i>
<i>COND_FLD</i>	SO4_LAB	ANC_LAB	<i>PH_LAB</i>
ANC_LAB	PH_LAB	ANC_LAB	COND_LAB
ANC_LAB	COND_LAB	<i>COND_FLD</i>	COND_LAB
SO4_LAB	COND_LAB	<i>SO4_LAB</i>	ANC_LAB
<i>COND_FLD</i>	ANC_LAB	<i>COND_FLD</i>	ANC_LAB
VEL_DPTH	<i>INSTRHAB</i>	Coastal Plain, Field-measured Variables	
POOLQUAL	<i>INSTRHAB</i>	<i>PH_FLD</i>	PH_LAB
RIFFQUAL	<i>INSTRHAB</i>	ANC_LAB	PH_LAB
POOLQUAL	VEL_DPTH	ANC_LAB	COND_LAB
<i>MAXDEPTH</i>	VEL_DPTH	SO4_LAB	COND_LAB
AVGTHAL	VEL_DPTH	<i>COND_FLD</i>	SO4_LAB
<i>MAXDEPTH</i>	POOLQUAL	<i>COND_FLD</i>	COND_LAB
AVGTHAL	POOLQUAL	<i>COND_FLD</i>	ANC_LAB
AVGTHAL	<i>MAXDEPTH</i>	EPI_SUB	<i>INSTRHAB</i>
AVGTHAL	AVGWID	POOLQUAL	<i>INSTRHAB</i>
VEL_DPTH	<i>DischargeCFS</i>	POOLQUAL	VEL_DPTH
RIFFQUAL	<i>DischargeCFS</i>	<i>MAXDEPTH</i>	POOLQUAL
AVGWID	<i>DischargeCFS</i>	AVGTHAL	POOLQUAL
AVGTHAL	<i>DischargeCFS</i>	AVGTHAL	<i>MAXDEPTH</i>
AVG_VEL	<i>DischargeCFS</i>	<i>MAXDEPTH</i>	VEL_DPTH
		AVGWID	<i>DischargeCFS</i>
		AVGTHAL	AVGWID

Local derived variables

Youghiogheny, Local Derived Variables	
<i>ph_loc</i>	bd_loc
str_blen_loc	isa_loc
<i>bp_2001</i>	isa_loc
<i>bp_2001</i>	str_blen_loc
<i>str_len_loc</i>	str_blen_loc
Highlands, Local Derived Variables	
<i>ph_loc</i>	bd_loc
<i>for_loc</i>	cc_loc
str_blen_loc	isa_loc
<i>bp_2001</i>	isa_loc
<i>bp_2001</i>	str_blen_loc
<i>str_len_loc</i>	str_blen_loc
Piedmont, Local Derived Variables	
<i>ph_loc</i>	bd_loc
str_blen_loc	isa_loc
<i>bp_2001</i>	isa_loc
<i>bp_2001</i>	str_blen_loc
<i>str_len_loc</i>	str_blen_loc
Coastal Plain, Local Derived Variables	
<i>ph_loc</i>	bd_loc
str_blen_loc	isa_loc
<i>bp_2001</i>	isa_loc
<i>bp_2001</i>	str_blen_loc
<i>str_len_loc</i>	str_blen_loc

Accumulated derived variables

Youghiogheny, Accumulated Derived Variables	
cc_acc	<i>ag_acc</i>
for_acc	<i>ag_acc</i>
ag_fls_nor	<i>ag_acc</i>
for_acc	cc_acc
<i>for_fls_nor</i>	cc_acc
<i>for_fls_nor</i>	for_acc
<i>for_fls_nor</i>	ag_fls_nor
<i>for_fls_nor</i>	cc_fls_nor
str_bp_acc	isa_acc
<i>isa_fls_nor</i>	isa_acc
str_blen_den	isa_acc
<i>isa_fls_nor</i>	str_bp_acc
str_blen_den	str_bp_acc
str_blen_den	<i>isa_fls_nor</i>
<i>str_den</i>	str_bp_acc
<i>str_den</i>	str_blen_den
<i>sicl_acc</i>	kfw_acc
<i>prof_acc</i>	plan_acc
str_len_acc	str_blen_acc
<i>dem10mad8</i>	str_blen_acc
<i>dem10mad8</i>	str_len_acc
<i>wet_fls_nor</i>	wet_acc

Accumulated derived variables, continued

Highlands, Accumulated Derived Variables	
cc_acc_loc	ag_acc
for_acc_loc	ag_acc
<i>ph_acc</i>	ag_acc
<i>slp_acc</i>	ag_acc
ag_fls_nor	ag_acc
cc_fls_nor	ag_acc
<i>for_fls_nor</i>	ag_acc
for_acc	cc_acc
<i>slp_acc</i>	cc_acc
ag_fls_nor	cc_acc
cc_fls_nor	cc_acc
<i>for_fls_nor</i>	cc_acc
<i>slp_acc</i>	for_acc
ag_fls_nor	for_acc
cc_fls_nor	for_acc
<i>for_fls_nor</i>	for_acc
cc_fls_nor	ag_fls_nor
<i>for_fls_nor</i>	ag_fls_nor
<i>for_fls_nor</i>	cc_fls_nor
isa_acc	for_acc
<i>for_fls_nor</i>	isa_acc
isa_fls_nor	isa_acc
isa_fls_nor	<i>for_fls_nor</i>
<i>sicl_acc</i>	kfw_acc
<i>prof_acc</i>	plan_acc
str_len_acc	str_blen_acc
<i>dem10mad8</i>	str_blen_acc
<i>dem10mad8</i>	str_len_acc
<i>wet_fls_nor</i>	wet_acc
<i>str_blen_den</i>	isa_fls_nor
str_bp_acc	<i>sat_acc</i>
<i>str_blen_den</i>	str_bp_acc

Piedmont, Accumulated Derived Variables	
str_bp_acc	isa_acc
<i>isa_fls_nor</i>	isa_acc
str_blen_den	isa_acc
<i>isa_fls_nor</i>	str_bp_acc
str_blen_den	str_bp_acc
str_blen_den	<i>isa_fls_nor</i>
str_len_acc	str_blen_acc
<i>dem10mad8</i>	str_blen_acc
<i>dem10mad8</i>	str_len_acc
<i>wet_fls_nor</i>	wet_acc
<i>for_fls_nor</i>	cc_fls_nor
<i>ag_fls_nor</i>	ag_acc
<i>ph_acc</i>	bd_acc
<i>for_acc</i>	cc_acc
plan_acc	<i>prof_acc</i>
Coastal Plain, Accumulated Derived Variables	
<i>ph_acc</i>	bd_acc
<i>for_acc</i>	cc_acc
isa_acc	<i>for_acc</i>
str_bp_acc	isa_acc
<i>isa_fls_nor</i>	isa_acc
isa_acc_loc	cc_acc
str_blen_den	isa_acc
<i>isa_fls_nor</i>	str_bp_acc
str_blen_den	<i>isa_fls_nor</i>
str_blen_den	str_bp_acc
<i>prof_acc</i>	plan_acc
str_len_acc	str_blen_acc
<i>dem10mad8</i>	str_blen_acc
<i>dem10mad8</i>	str_len_acc
wet_fls_nor	<i>for_fls_nor</i>
wet_fls_nor	<i>wet_acc</i>
<i>ag_fls_nor</i>	ag_acc
<i>str_den</i>	slp_acc

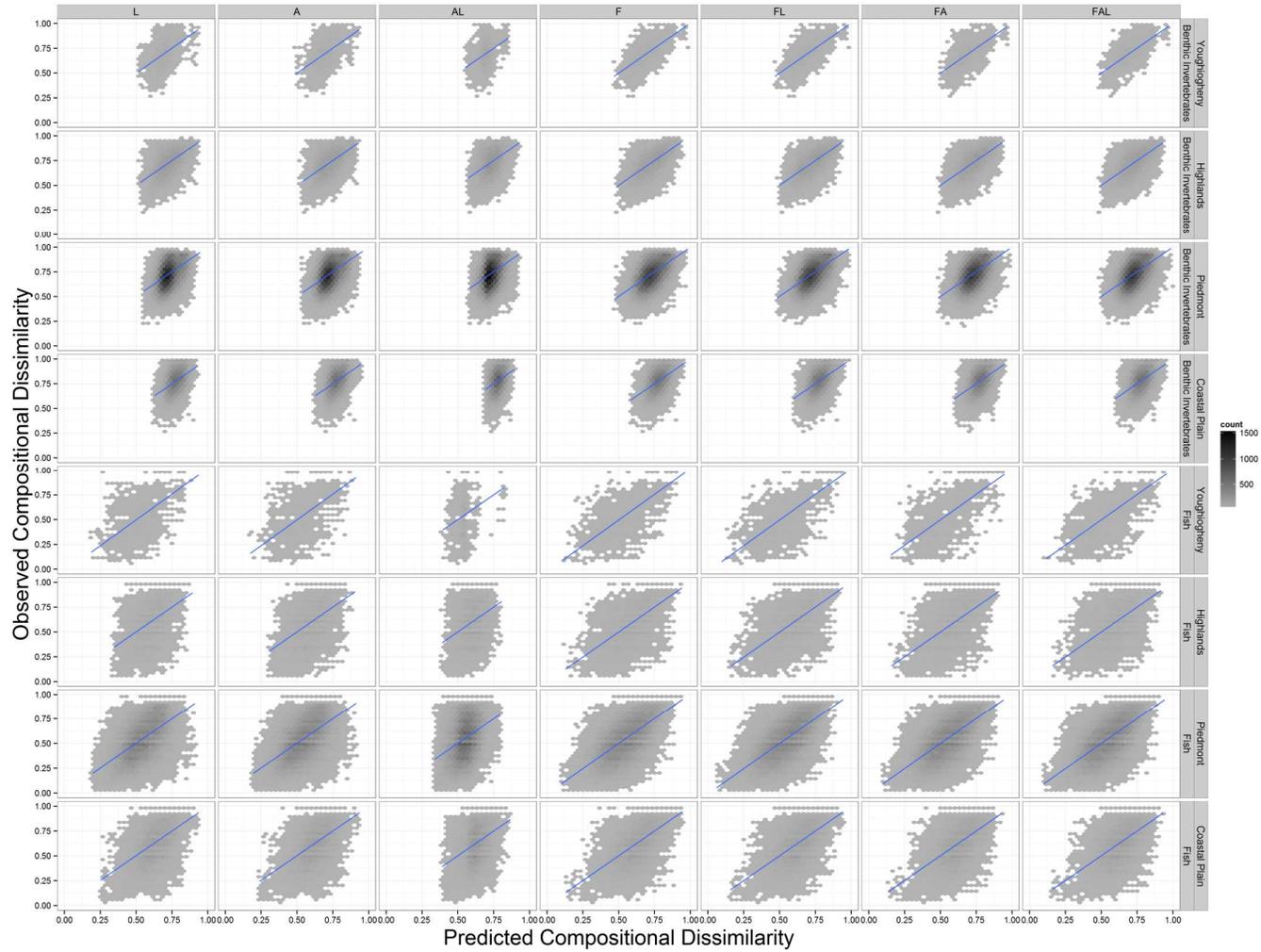
Appendix C. Inter-group Pearson and Spearman variable correlations $> |0.7|$. A candidate variable change of NA indicates that despite the cross-group correlation, no action was taken. This occurred when the less-relevant correlate in the pair had previously been omitted due to presence in another correlated group.

		Correlate 1	Correlate 2	Candidate Variables Change
Local & Accumulated Derived Variables	Youghiogheny	isa_fl_s_nor	str_blen_loc	NA
		isa_fl_s_nor	isa_loc	NA
	Highlands	sat_acc	sat_loc	omit sat_acc
		cc_fl_s_nor	cc_loc	NA
	Piedmont	sat_acc	sat_loc	omit sat_acc
		kfw_acc	kfw_loc	omit kfw_acc
	Coastal Plain	sat_acc	sat_loc	omit sat_acc
		sat_acc	sat_loc	omit sat_acc
Local Derived & Field-measured Variables	Youghiogheny	NA	NA	NA
	Highlands	NA	NA	NA
	Piedmont	NA	NA	NA
	Coastal Plain	NA	NA	NA
Accumulated Derived & Field-measured Variables	Youghiogheny	DischargeCFS	dem10mad8	omit dem10mad8
	Highlands	NA	NA	NA
	Piedmont	DischargeCFS	dem10mad8	omit dem10mad8
	Coastal Plain	NA	NA	NA

Appendix D. AICs compared for GDMs with all seven variable sets. Δ AIC shows the difference between the current model and the best model, marked with a star (*). AIC_w are AIC weights, and K is the number of parameters for each model. Variable codes: L = local derived, A = accumulated derived, F = field-measured.

Models	Youghiogheny				Highlands				Piedmont				Coastal Plain				
	AIC	Δ AIC	AIC _w	K	AIC	Δ AIC	AIC _w	K	AIC	Δ AIC	AIC _w	K	AIC	Δ AIC	AIC _w	K	
Fish	L	2376.1	393.7	0.00	4	9731.0	849.3	0.00	14	65892.2	7471.7	0.00	9	17985.5	1190.0	0.00	17
	A	2157.3	174.9	0.00	20	9399.9	518.2	0.00	9	61527.9	3107.4	0.00	13	17312.3	516.7	0.00	25
	AL	2172.3	189.9	0.00	19	9362.5	480.7	0.00	14	61417.2	2996.8	0.00	12	17290.0	494.4	0.00	28
	F	2042.2	59.8	0.00	12	9057.1	175.3	0.00	18	59312.9	892.5	0.00	17	16916.1	120.5	0.00	27
	FL	2026.2	43.8	0.00	19	8997.2	115.5	0.00	24	58986.3	565.8	0.00	20	16910.0	114.4	0.00	29
	FA	1988.3	5.9	0.05	26	8885.8	4.0	0.12	25	58467.0	46.6	0.00	25	16795.6*	0.0	1.00	34
	FAL	1982.4*	0.0	0.95	24	8881.8*	0.0	0.88	28	58420.5*	0.0	1.00	25	16828.7	33.1	0.00	31
Benthic Inverts	L	1653.4	31.0	0.00	8	7901.1	133.6	0.00	15	36853.1	232.3	0.00	14	9593.0*	0.0	0.97	10
	A	1637.7	15.3	0.00	14	7825.7	58.2	0.00	12	36704.2	83.4	0.00	17	9601.9	8.9	0.01	22
	AL	1636.2	13.7	0.00	16	7818.4	51.0	0.00	11	36692.7	71.8	0.00	22	9602.2	9.2	0.01	21
	F	1627.6	5.1	0.03	17	7800.1	32.6	0.00	15	36678.0	57.2	0.00	26	9603.6	10.6	0.01	28
	FL	1625.1	2.7	0.11	18	7798.6	31.1	0.00	18	36675.4	54.6	0.00	25	9609.4	16.4	0.00	30
	FA	1622.4*	0.0	0.43	19	7767.5*	0.0	0.90	16	36620.8*	0.0	0.94	29	9607.0	13.9	0.00	28
	FAL	1622.4*	0.0	0.43	19	7771.8	4.3	0.10	13	36626.4	5.6	0.06	28	9613.8	20.7	0.00	29

Appendix E. Predicted vs. observed compositional dissimilarity for all generalized dissimilarity models, plotted with best-fit line



Appendix F. Statistically significant differences in GDM ability to predict Bray-Curtis dissimilarity measured using (a) Spearman correlations between predicted and observed dissimilarities and (b) percent error. Hatched fill denotes model pairs that were significantly different. Variable codes: L = local derived, A = accumulated derived, F = field-measured.

(a)

Comparison	Youghiogheny		Highlands		Piedmont		Coastal Plain	
	Fish	Benthic Inverts	Fish	Benthic Inverts	Fish	Benthic Inverts	Fish	Benthic Inverts
A-AL								
A-AF								
A-L								
A-F								
A-FAL								
A-FL								
AL-L								
AL-F								
AL-FA								
AL-FAL								
AL-FL								
L-F								
L-FA								
L-FAL								
L-FL								
F-FA								
F-FAL								
F-FL								
FA-FAL								
FA-FL								
FAL-FL								

(b)

Comparison	Youghiogheny		Highlands		Piedmont		Coastal Plain	
	Fish	Benthic Inverts	Fish	Benthic Inverts	Fish	Benthic Inverts	Fish	Benthic Inverts
A-AL								
A-AF								
A-L								
A-F								
A-FAL								
A-FL								

Comparison	Youghiogheny		Highlands		Piedmont		Coastal Plain	
	Fish	Benthic Inverts	Fish	Benthic Inverts	Fish	Benthic Inverts	Fish	Benthic Inverts
AL-L								
AL-F								
AL-FA								
AL-FAL								
AL-FL								
L-F								
L-FA								
L-FAL								
L-FL								
F-FA								
F-FAL								
F-FL								
FA-FAL								
FA-FL								
FAL-FL								

Appendix G. Variables included in final GDMs, ordered by importance as assessed by the sum of I-spline coefficients. V1 is the most important variable. Region codes: Y = Youghiogheny, H = Highlands, P = Piedmont, CP = Coastal Plain. Taxon codes: F = fish, B = benthic invertebrates. Variable codes: L = local derived, A = accumulated derived, F = field-measured.

Region	Y	Y	H	H	P	P	CP	CP
Taxon	F	B	F	B	F	B	F	B
Vars	L	L	L	L	L	L	L	L
V1	plan_loc	sicl_loc	sat_loc	bp_2001	Geographic	bp_2001	bp_2001	Geographic
V2	Geographic	bp_2001	bp_2001	ph_loc	sat_loc	sat_loc	Geographic	bp_2001
V3		Geographic	ph_loc	sat_loc	sicl_loc	sicl_loc	str_len_loc	kfw_loc
V4		cc_loc	sicl_loc	Geographic	brd_loc	Geographic	kfw_loc	str_len_loc
V5			str_len_loc	ag_loc	bp_2001	brd_loc	sicl_loc	cc_loc
V6			Geographic	wet_loc		str_len_loc	ag_loc	
V7			brd_loc	sicl_loc		ph_loc	for_loc	
V8			plan_loc	slp_loc		ag_loc	sat_loc	
V9				brd_loc				
Region	Y	Y	H	H	P	P	CP	CP
Taxon	F	B	F	B	F	B	F	B
Vars	A	A	A	A	A	A	A	A
V1	dem10mad8	isa_fl_s_nor	dem10mad8	sicl_acc	dem10mad8	dem10mad8	ag_fl_s_nor	plan_acc
V2	brd_acc	slp_acc	brd_acc	bd_acc	sat_acc	str_den	dem10mad8	isa_fl_s_nor
V3	prof_acc	bd_acc	str_blen_den	for_fl_s_nor	for_acc	sat_acc	wet_acc	brd_acc
V4	isa_fl_s_nor	ag_acc	sat_acc	str_blen_den	sicl_acc	isa_fl_s_nor	sat_acc	Geographic
V5	sat_acc	wet_fl_s_nor	Geographic	dem10mad8	Geographic	for_acc	isa_fl_s_nor	ph_acc
V6	bd_acc	Geographic		brd_acc	brd_acc	slp_acc	for_acc	dem10mad8
V7	slp_acc			Geographic		ag_fl_s_nor	ph_acc	for_acc
V8	str_den					Geographic	kfw_acc	wet_acc
V9	ag_acc					brd_acc	Geographic	ag_fl_s_nor
V10	Geographic						brd_acc	kfw_acc
V11							confluence_num	
V12							sicl_acc	

V13
V14

plan_acc
for_flis_nor

Region	Y	Y	H	H	P	P	CP	CP
Taxon	F	B	F	B	F	B	F	B
Vars	AL	AL	AL	AL	AL	AL	AL	AL
V1	dem10mad8	isa_flis_nor	dem10mad8	sicl_acc	dem10mad8	str_den	ag_flis_nor	prof_acc
V2	str_den	bd_acc	bp_2001	bp_2001	for_acc	dem10mad8	dem10mad8	brd_acc
V3	brd_acc	slp_acc	sat_loc	str_blen_den	sicl_acc	sat_loc	bp_2001	Geographic
V4	isa_flis_nor	sicl_loc	brd_acc	for_flis_nor	Geographic	for_acc	wet_acc	ph_acc
V5	bd_acc	ag_acc	str_blen_den	dem10mad8	sat_loc	bp_2001	Geographic	isa_flis_nor
V6	prof_acc	wet_flis_nor	sicl_acc	brd_acc	brd_acc	sicl_loc	ph_acc	dem10mad8
V7	plan_loc	Geographic	Geographic	Geographic		slp_acc	for_acc	for_acc
V8	sat_loc	cc_loc				isa_flis_nor	prof_acc	wet_acc
V9	slp_acc					ag_flis_nor	str_len	bp_2001
V10	Geographic					Geographic	kfw_acc	ag_flis_nor
V11						brd_acc	isa_flis_nor	kfw_acc
V12							brd_acc	
V13							sicl_acc	
V14							for_flis_nor	
V15							sat_loc	

Region	Y	Y	H	H	P	P	CP	CP
Taxon	F	B	F	B	F	B	F	B
Vars	F	F	F	F	F	F	F	F
V1	PH_LAB	DO_FLD	ST_GRAD	COND_FLD	PH_LAB	NO3_LAB	DischargeCFS	COND_FLD
V2	ST_GRAD	NO3_LAB	TEMP_FLD	PH_FLD	SO4_LAB	SO4_LAB	MAXDEPTH	Geographic
V3	NO3_LAB	PH_LAB	DischargeCFS	NO3_LAB	DischargeCFS	TEMP_FLD	DOC_LAB	TEMP_FLD
V4	AVGTHAL	ST_GRAD	DO_FLD	ST_GRAD	MAXDEPTH	AESTHET	PH_FLD	INSTRHAB
V5	Geographic	SO4_LAB	DOC_LAB	SHADING	COND_FLD	EMBEDDED	TEMP_FLD	PH_FLD
V6		AESTHET	INSTRHAB	Geographic	ST_GRAD	PH_LAB	Geographic	AVG_VEL
V7		RIFFQUAL	COND_FLD	EMBEDDED	Geographic	COND_FLD	COND_FLD	RIFFQUAL

V8		Geographic	Geographic		RIFFQUAL	DischargeCFS	SHADING	AESTHET
V9						ST_GRAD	AESTHET	DOC_LAB
V10						Geographic	NO3_LAB	NO3_LAB
V11						INSTRHAB	INSTRHAB	DO_FLD
V12						AVG_VEL	ST_GRAD	MAXDEPTH
V13						RIFFQUAL		

Region	Y	Y	H	H	P	P	CP	CP
Taxon	F	B	F	B	F	B	F	B
Vars	FL	FL	FL	FL	FL	FL	FL	FL
V1	PH_LAB	DO_FLD	ST_GRAD	COND_FLD	PH_LAB	NO3_LAB	DischargeCFS	COND_FLD
V2	ST_GRAD	PH_LAB	TEMP_FLD	PH_FLD	DischargeCFS	SO4_LAB	MAXDEPTH	Geographic
V3	NO3_LAB	NO3_LAB	bp_2001	bp_2001	SO4_LAB	EMBEDDED	DOC_LAB	TEMP_FLD
V4	AVGTHAL	bp_2001	DischargeCFS	ST_GRAD	MAXDEPTH	sat_loc	PH_FLD	PH_FLD
V5	SO4_LAB	SO4_LAB	sat_loc	NO3_LAB	COND_FLD	TEMP_FLD	TEMP_FLD	INSTRHAB
V6	sat_loc	ST_GRAD	DO_FLD	ph_loc	ST_GRAD	DischargeCFS	Geographic	AVG_VEL
V7	plan_loc	RIFFQUAL	DOC_LAB	Geographic	Geographic	sicl_loc	COND_FLD	kfw_loc
V8	bp_2001	DOC_LAB	INSTRHAB	SHADING	sat_loc	AESTHET	NO3_LAB	bp_2001
V9	Geographic	AESTHET	COND_FLD	EMBEDDED	RIFFQUAL	INSTRHAB	SHADING	RIFFQUAL
V10		Geographic	MAXDEPTH			ST_GRAD	bp_2001	DOC_LAB
V11			ph_loc			PH_LAB	INSTRHAB	NO3_LAB
V12			Geographic			Geographic	ST_GRAD	AESTHET
V13							AESTHET	DO_FLD
V14								MAXDEPTH

Region	Y	Y	H	H	P	P	CP	CP
Taxon	F	B	F	B	F	B	F	B
Vars	FA	FA	FA	FA	FA	FA	FA	FA
V1	PH_LAB	isa_fls_nor	ST_GRAD	PH_FLD	PH_LAB	NO3_LAB	ag_fls_nor	COND_FLD
V2	ST_GRAD	DO_FLD	dem10mad8	COND_FLD	SO4_LAB	SO4_LAB	MAXDEPTH	TEMP_FLD
V3	SO4_LAB	PH_LAB	DO_FLD	sicl_acc	DischargeCFS	str_den	PH_FLD	INSTRHAB
V4	NO3_LAB	bd_acc	TEMP_FLD	ST_GRAD	MAXDEPTH	sat_acc	DischargeCFS	Geographic

V5	prof_acc	for_fl_s_nor	str_den	for_fl_s_nor	sat_acc	TEMP_FLD	sat_acc	ph_acc
V6	isa_fl_s_nor	SO4_LAB	sat_acc	str_blen_den	ST_GRAD	EMBEDDED	dem10mad8	isa_fl_s_nor
V7	slp_acc	ST_GRAD	DOC_LAB	SHADING	COND_FLD	ST_GRAD	DOC_LAB	PH_FLD
V8	str_den	slp_acc	INSTRHAB	brd_acc	Geographic	for_acc	kfw_acc	AVG_VEL
V9	brd_acc	RIFFQUAL	COND_FLD	Geographic	RIFFQUAL	DischargeCFS	ph_acc	for_acc
V10	sat_acc	Geographic	brd_acc			isa_fl_s_nor	isa_fl_s_nor	RIFFQUAL
V11	AVGTHAL		DischargeCFS			PH_LAB	TEMP_FLD	dem10mad8
V12	Geographic		Geographic			AVG_VEL	wet_acc	brd_acc
V13						slp_acc	SHADING	
V14						AESTHET	for_acc	
V15						RIFFQUAL	INSTRHAB	
V16						brd_acc	Geographic	
V17						Geographic		

Region	Y	Y	H	H	P	P	CP	CP
Taxon	F	B	F	B	F	B	F	B
Vars	FAL							
V1	ST_GRAD	isa_fl_s_nor	ST_GRAD	PH_FLD	PH_LAB	NO3_LAB	ag_fl_s_nor	COND_FLD
V2	PH_LAB	DO_FLD	dem10mad8	bp_2001	DischargeCFS	SO4_LAB	MAXDEPTH	prof_acc
V3	SO4_LAB	PH_LAB	bp_2001	COND_FLD	SO4_LAB	str_den	PH_FLD	TEMP_FLD
V4	isa_fl_s_nor	bd_acc	DO_FLD	sicl_acc	MAXDEPTH	sat_loc	DischargeCFS	ph_acc
V5	brd_acc	for_fl_s_nor	TEMP_FLD	ST_GRAD	ST_GRAD	TEMP_FLD	dem10mad8	INSTRHAB
V6	NO3_LAB	SO4_LAB	sat_loc	for_fl_s_nor	COND_FLD	isa_fl_s_nor	DOC_LAB	Geographic
V7	prof_acc	ST_GRAD	brd_acc	brd_acc	for_acc	EMBEDDED	bp_2001	AVG_VEL
V8	sat_loc	slp_acc	str_den	Geographic	sat_loc	ST_GRAD	ph_acc	isa_fl_s_nor
V9	slp_acc	RIFFQUAL	INSTRHAB		Geographic	for_acc	TEMP_FLD	kfw_loc
V10	AVGTHAL	Geographic	DOC_LAB		RIFFQUAL	DischargeCFS	SHADING	RIFFQUAL
V11	Geographic		COND_FLD		TEMP_FLD	slp_acc	for_acc	PH_FLD
V12			DischargeCFS		brd_acc	AVG_VEL	Geographic	for_acc
V13			Geographic			RIFFQUAL	kfw_acc	brd_acc
V14						brd_acc	wet_acc	
V15						Geographic		

Appendix H. Total number of GDMs for which variables were selected. For each taxon, a variable was a candidate for 16 models (i.e. an accumulated variable could have been in A, AL, FA, and FAL models in four regions). Variable codes: L = local derived, A = accumulated derived, F = field-measured.

Variable		Number of models for which selected		Variable		Number of models for which selected	
		BENTHIC	FISH			BENTHIC	FISH
ST_GRAD		12	14	str_len	local	2	3
brd	accumulated	12	13	str_blen_den	accumulated	3	2
bp_2001	local	11	10	ph	local	3	2
COND_FLD		9	10	MAXDEPTH		2	2
dem10mad8	accumulated	7	12	AVGTHAL		0	4
isa_fls_nor	spatial accumulated	12	7	plan	local	0	4
sat	local	5	13	kfw	local	3	1
TEMP_FLD		8	9	brd	local	2	2
DischargeCFS		4	12	cc	local	3	0
NO3_LAB		10	6	ag	local	2	1
RIFFQUAL		11	4	ag	accumulated	2	1
PH_LAB		7	8	plan	accumulated	1	1
SO4_LAB		8	7	wet_fls_nor	spatial accumulated	2	0
for	accumulated	8	7	slp	local	1	0
INSTRHAB		6	7	confluence_num	accumulated	0	1
PH_FLD		8	4	for	local	0	1
slp	accumulated	8	4	wet	local	1	0
DOC_LAB		3	8	EPL_SUB		0	0
DO_FLD		6	4	VEL_DPTH		0	0
sat	accumulated	2	8	POOLQUAL		0	0
AESTHET		7	2	AVGWID		0	0
str_den	accumulated	4	5	COND_LAB		0	0
sicl	local	6	3	ANC_LAB		0	0
sicl	accumulated	4	5	prof	local	0	0
for_fls_nor	spatial accumulated	6	2	dem10mp	local	0	0
ag_fls_nor	spatial accumulated	4	4	str_len	accumulated	0	0
SHADING		3	4	str_blen	local	0	0
AVG_VEL		7	0	str_ben	accumulated	0	0
prof	accumulated	2	5	str_bp	accumulated	0	0
bd	accumulated	5	2	isa	local	0	0
EMBEDDED		6	0	isa	accumulated	0	0
wet	accumulated	2	4	cc	accumulated	0	0
kfw	accumulated	2	4	cc_fls_nor	spatial accumulated	0	0
ph	accumulated	4	2	bd	local	0	0

LITERATURE CITED

- Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19:716–723.
- Allan, J. D. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 35:257–284.
- Anderson, N. H., and K. W. Cummins. 1979. Influences of diet on the life histories of aquatic insects. *Journal of the Fisheries Board of Canada* 36:335–342.
- Araújo, M. B., and A. Guisan. 2006. Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33:1677–1688.
- Araújo, M. B., and A. Rozenfeld. 2013. The geographic scaling of biotic interactions. *Ecography*:001--010.
- Austin, M., and K. P. Van Niel. 2011. Improving species distribution models for climate change studies: variable selection and scale: Species distribution models for climate change studies. *Journal of Biogeography* 38:1–8.
- Austin, M. P. 1980. Searching for a model for use in vegetation analysis. *Vegetatio* 42:11–21.
- Austin, M. P. 2002. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. *Ecological modelling* 157:101–118.
- Barton, K. 2014. MuMIn: Multi-model inference. R package version 1.10.0. <http://CRAN.R-project.org/package=MuMIn>.
- Baselga, A. 2010. Partitioning the turnover and nestedness components of beta diversity: Partitioning beta diversity. *Global Ecology and Biogeography* 19:134–143.
- Belica, L. A. T., and F. J. Rahel. 2008. Movements of creek chubs, *Semotilus atromaculatus*, among habitat patches in a plains stream. *Ecology of Freshwater Fish* 17:258–272.
- Benda, L., N. L. Poff, D. Miller, T. Dunne, G. Reeves, G. Pess, and M. Pollock. 2004. The network dynamics hypothesis: how channel networks structure riverine habitats. *BioScience* 54:413–427.
- Bilton, D. T., J. R. Freeland, and B. Okamura. 2001. Dispersal in Freshwater Invertebrates. *Annual Review of Ecology and Systematics* 32:159–181.

- Boesch, D. F., R. B. Brinsfield, and R. E. Magnien. 2001. Chesapeake Bay Eutrophication. *Journal of Environmental Quality* 30:303–320.
- Borcard, D., P. Legendre, and P. Drapeau. 1992. Partialling out the Spatial Component of Ecological Variation. *Ecology* 73:1045.
- Boward, D. M., and E. Friedman. 2011. Maryland biological stream survey laboratory methods for benthic macroinvertebrate processing and taxonomy. Maryland Department of Natural Resources.
- Braunisch, V., J. Coppes, R. Arlettaz, R. Suchant, H. Schmid, and K. Bollmann. 2013. Selecting from correlated climate variables: a major source of uncertainty for predicting species distributions under climate change. *Ecography* 36:971–983.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45:5–32.
- Brönmark, C., C. Skov, J. Brodersen, P. A. Nilsson, and L.-A. Hansson. 2008. Seasonal Migration Determined by a Trade-Off between Predator Avoidance and Growth. *PLoS ONE* 3:e1957.
- Christmas, J., R. Eades, D. Cincotta, A. Shiels, R. Miller, J. Siemien, T. Sinnott, and P. Fuller. 2001. History, management, and status of introduced fishes in the Chesapeake Bay basin. Pages 97–116 *in* G. D. Therres, editor. *Proceedings of Conservation of Biological Diversity: A Key to the Restoration of the Chesapeake Bay Ecosystem and Beyond*, May 10-13, 1998. Maryland Department of Natural Resources.
- Clark, J. S., D. M. Bell, M. H. Hersh, M. C. Kwit, E. Moran, C. Salk, A. Stine, D. Valle, and K. Zhu. 2011. Individual-scale variation, species-scale differences: inference needed to understand diversity: Individual-scale variation, species-scale differences. *Ecology Letters* 14:1273–1287.
- Cord, A. F., R. K. Meentemeyer, P. J. Leitão, and T. Václavík. 2013. Modelling species distributions with remote sensing data: bridging disciplinary perspectives. *Journal of Biogeography* 40:2226–2227.
- Dettinger, M. D., and H. F. Diaz. 2000. Global characteristics of stream flow seasonality and variability. *Journal of Hydrometeorology* 1:289–310.
- Dodds, W. K., W. H. Clements, K. Gido, R. H. Hilderbrand, and R. S. King. 2010. Thresholds, breakpoints, and nonlinearity in freshwaters as related to management. *Journal of the North American Benthological Society* 29:988–997.
- Duarte Silva, A. P. 2001. Efficient Variable Screening for Multivariate Analysis. *Journal of Multivariate Analysis* 76:35–62.

- Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, R. J. Naiman, A.-H. Prieur-Richard, D. Soto, M. L. J. Stiassny, and C. A. Sullivan. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* 81:163.
- Elith, J., and J. R. Leathwick. 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics* 40:677–697.
- Elmore, A. J., J. P. Julian, S. M. Guinn, and M. C. Fitzpatrick. 2013. Potential Stream Density in Mid-Atlantic U.S. Watersheds. *PLoS ONE* 8:e74819.
- Elmore, A. J., and S. S. Kaushal. 2008. Disappearing headwaters: patterns of stream burial due to urbanization. *Frontiers in Ecology and the Environment* 6:308–312.
- Ferrier, S., M. Drielsma, G. Manion, and G. Watson. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. *Biodiversity & Conservation* 11:2309–2338.
- Ferrier, S., and A. Guisan. 2006. Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology* 43:393–404.
- Ferrier, S., G. Manion, J. Elith, and K. Richardson. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions* 13:252–264.
- Fisher, S. G. 1997. Creativity, Idea Generation, and the Functional Morphology of Streams. *Journal of the North American Benthological Society* 16:305–318.
- Fitzpatrick, M. C., N. J. Sanders, S. Ferrier, J. T. Longino, M. D. Weiser, and R. Dunn. 2011. Forecasting the future of biodiversity: a test of single- and multi-species models for ants in North America. *Ecography* 34:836–847.
- Fitzpatrick, M. C., N. J. Sanders, S. Normand, J.-C. Svenning, S. Ferrier, A. D. Gove, and R. R. Dunn. 2013. Environmental and historical imprints on beta diversity: insights from variation in rates of species turnover along gradients. *Proceedings of the Royal Society B: Biological Sciences* 280:20131201–20131201.
- Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in Physical Geography* 19:474–499.

- Freeman, M. C., C. M. Pringle, and C. R. Jackson. 2007. Hydrologic Connectivity and the Contribution of Stream Headwaters to Ecological Integrity at Regional Scales. *JAWRA Journal of the American Water Resources Association* 43:5–14.
- Furnival, G. M., and R. W. Wilson. 1974. Regressions by Leaps and Bounds. *Technometrics* 16:499–511.
- Gesch, D., M. Oimoen, S. Greenlee, C. Nelson, M. Steuck, and D. Tyler. 2002. The national elevation dataset. *Photogrammetric engineering and remote sensing* 68:5–11.
- Goetz, S. J., C. A. Jantz, S. D. Prince, A. J. Smith, D. Varlyguin, and R. K. Wright. 2004. Integrated analysis of ecosystem interactions with land use Change: The Chesapeake Bay watershed. Pages 263–275 *in* R. S. DeFries, G. P. Asner, and R. A. Houghton, editors. *Geophysical Monograph Series*. American Geophysical Union, Washington, D. C.
- Grant, E. H. C., W. H. Lowe, and W. F. Fagan. 2007. Living in the branches: population dynamics and ecological processes in dendritic networks. *Ecology Letters* 10:165–175.
- Guisan, A., and N. E. Zimmermann. 2000. Predictive habitat distribution models in ecology. *Ecological modelling* 135:147–186.
- Hansson, L.-A., C. Bronmark, P. Anders Nilsson, and K. Abjornsson. 2005. Conflicting demands on wetland ecosystem services: nutrient retention, biodiversity or both? *Freshwater Biology* 50:705–714.
- 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* 95:14843–14847.
- Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10:1456–1477.
- Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25:1965–1978.
- Hilderbrand, R. H., M. T. Kashiwagi, and A. P. Prochaska. 2014. Regional and Local Scale Modeling of Stream Temperatures and Spatio-Temporal Variation in Thermal Sensitivities. *Environmental Management* 54:14–22.

- Hocking, R. R. 1976. A Biometrics Invited Paper. The Analysis and Selection of Variables in Linear Regression. *Biometrics* 32:1.
- Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham. 2007. Completion of the 2001 national land cover database for the coterminus United States. *Photogrammetric Engineering and Remote Sensing* 73:337–341.
- Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22:145–159.
- Jackson, J. K., E. Mcelravy, and V. I. Resh. 1999. Long-term movements of self-marked caddisfly larvae (Trichoptera: Sericostomatidae) in a California coastal mountain stream. *Freshwater Biology* 42:525–536.
- Johnson, T. E., J. N. McNair, P. Srivastava, and D. D. Hart. 2007. Stream ecosystem responses to spatially variable land cover: an empirically based model for developing riparian restoration strategies. *Freshwater Biology* 52:680–695.
- Jones, M. M., S. Ferrier, R. Condit, G. Manion, S. Aguilar, and R. Pérez. 2013. Strong congruence in tree and fern community turnover in response to soils and climate in central Panama. *Journal of Ecology* 101:506–516.
- Julian, J. P., A. J. Elmore, and S. M. Guinn. 2012. Channel head locations in forested watersheds across the mid-Atlantic United States: A physiographic analysis. *Geomorphology* 177-178:194–203.
- Karr, J. R. 1981. Assessment of Biotic Integrity Using Fish Communities. *Fisheries* 6:21–27.
- Kilian, J. V., R. J. Klauda, S. Widman, M. Kashiwagi, R. Bourquin, S. Weglein, and J. Schuster. 2012. An assessment of a bait industry and angler behavior as a vector of invasive species. *Biological Invasions* 14:1469–1481.
- King, R. S., M. E. Baker, D. F. Whigham, D. E. Weller, T. E. Jordan, P. F. Kazyak, and M. K. Hurd. 2005. Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecological applications* 15:137–153.
- Kobler, A., Y. Humblet, G. Knaepkens, B. Engelen, and M. Eens. 2012. Diel movement of bullhead (*Cottus perifretum*) in a lowland stream: Diel movement of bullhead. *Ecology of Freshwater Fish* 21:453–460.
- Kratzer, E. B., J. K. Jackson, D. B. Arscott, A. K. Aufdenkampe, C. L. Dow, L. A. Kaplan, J. D. Newbold, and B. W. Sweeney. 2006. Macroinvertebrate distribution in relation to land use and water chemistry in New York City drinking-water-

- supply watersheds. *Journal of the North American Benthological Society* 25:954–976.
- Kristensen, E. A., A. Baattrup-Pedersen, and H. E. Andersen. 2012. Prediction of stream fish assemblages from land use characteristics: implications for cost-effective design of monitoring programmes. *Environmental Monitoring and Assessment* 184:1435–1448.
- Leathwick, J. R., T. Snelder, W. L. Chadderton, J. Elith, K. Julian, and S. Ferrier. 2011. Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications: River classification for conservation management. *Freshwater Biology* 56:21–38.
- Leathwick, J. R., and D. Whitehead. 2001. Soil and atmospheric water deficits and the distribution of New Zealand's indigenous tree species. *Functional Ecology* 15:233–242.
- Lee, P., C. Smyth, and S. Boutin. 2004. Quantitative review of riparian buffer width guidelines from Canada and the United States. *Journal of Environmental Management* 70:165–180.
- Lemly, A. D. 1982. Modification of benthic insect communities in polluted streams: combined effects of sedimentation and nutrient enrichment. *Hydrobiologia* 87:229–245.
- Lenat, D. R. 1984. Agriculture and stream water quality: A biological evaluation of erosion control practices. *Environmental management* 8:333–343.
- Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between—and reconciliation of—“predictive” and “explanatory” models. *Biodiversity & Conservation* 9:655–671.
- Mackey, B. G., and D. B. Lindenmayer. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography* 28:1147–1166.
- Macneale, K. H., B. L. Peckarsky, and G. E. Likens. 2005. Stable isotopes identify dispersal patterns of stonefly populations living along stream corridors. *Freshwater Biology* 50:1117–1130.
- Magbanua, F. S., C. R. Townsend, K. J. Hageman, and C. D. Matthaei. 2013. Individual and combined effects of fine sediment and the herbicide glyphosate on benthic macroinvertebrates and stream ecosystem function. *Freshwater Biology* 58:1729–1744.

- Mander, Ü., V. Kuusemets, K. Lõhmus, and T. Mauring. 1997. Efficiency and dimensioning of riparian buffer zones in agricultural catchments. *Ecological Engineering* 8:299–324.
- Manion, G., M. Fitzpatrick, and M. Lisk. 2014. Gdm01: Create generalised dissimilarity models. R package version 1.0/r63. <http://R-Forge.R-project.org/projects/gdm>.
- Melles, S. J., N. E. Jones, and B. J. Schmidt. 2014. Evaluation of Current Approaches to Stream Classification and a Heuristic Guide to Developing Classifications of Integrated Aquatic Networks. *Environmental Management* 53:549–566.
- Millar, R. B., M. J. Anderson, and N. Tolimieri. 2011. Much ado about nothings: using zero similarity points in distance-decay curves. *Ecology* 92:1717–1722.
- Mohseni, O., and H. G. Stefan. 1999. Stream temperature/air temperature relationship: a physical interpretation. *Journal of Hydrology* 218:128–141.
- Mokany, K., T. D. Harwood, J. M. Overton, G. M. Barker, and S. Ferrier. 2011. Combining α - and β -diversity models to fill gaps in our knowledge of biodiversity: Filling gaps in biodiversity knowledge. *Ecology Letters* 14:1043–1051.
- 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.
- Murphy, J. 2010. Quantifying the relationship between land cover and biological condition of headwater streams. *Freshwater Forum* 28:27–44.
- Natureserve. 2014. NatureServe web service. Arlington, VA. USA. <http://services.natureserve.org>. Accessed Jan 2014.
- Nekola, J. C., and P. S. White. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography* 26:867–878.
- Nelson, K. C., M. A. Palmer, J. E. Pizzuto, G. E. Moglen, P. L. Angermeier, R. H. Hilderbrand, M. Dettinger, and K. Hayhoe. 2009. Forecasting the combined effects of urbanization and climate change on stream ecosystems: from impacts to management options. *Journal of Applied Ecology* 46:154–163.
- Paul, M. J., and J. L. Meyer. 2001. Streams in the Urban Landscape. *Annual Review of Ecology and Systematics* 32:333–365.
- Pearce, J., and S. Ferrier. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling* 128:127–147.

- Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global ecology and biogeography* 12:361–371.
- Pease, A. A., J. M. Taylor, K. O. Winemiller, and R. S. King. 2011. Multiscale Environmental Influences on Fish Assemblage Structure in Central Texas Streams. *Transactions of the American Fisheries Society* 140:1409–1427.
- Peterson, E. E., J. M. Ver Hoef, D. J. Isaak, J. A. Falke, M.-J. Fortin, C. E. Jordan, K. McNyset, P. Monestiez, A. S. Ruesch, A. Sengupta, N. Som, E. A. Steel, D. M. Theobald, C. E. Torgersen, and S. J. Wenger. 2013. Modelling dendritic ecological networks in space: an integrated network perspective. *Ecology Letters* 16:707–719.
- Peterson, E. E., F. Sheldon, R. Darnell, S. E. Bunn, and B. D. Harch. 2011. A comparison of spatially explicit landscape representation methods and their relationship to stream condition. *Freshwater Biology* 56:590–610.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259.
- Poff, N. L., B. P. Bledsoe, and C. O. Cuhaciyan. 2006. Hydrologic variation with land use across the contiguous United States: Geomorphic and ecological consequences for stream ecosystems. *Geomorphology* 79:264–285.
- R Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Reger, J. P., and E. T. Cleaves. 2008. Explanatory text for the physiographic map of Maryland.
- Rouquette, J. R., M. Dallimer, P. R. Armsworth, K. J. Gaston, L. Maltby, and P. H. Warren. 2013. Species turnover and geographic distance in an urban river network. *Diversity and Distributions* 19:1429–1439.
- Sala, O. E., F. S. Chapin III, J. J. Armesto, E. Berlow, J. Bloomfield, R. Dirzo, E. Huber-Sanwald, L. F. Huenneke, R. B. Jackson, A. Kinzig, R. Leemans, D. M. Lodge, H. A. Mooney, M. Oesterheld, N. L. Poff, M. T. Sykes, B. H. Walker, M. Walker, and D. H. Wall. 2000. Global Biodiversity Scenarios for the Year 2100. *Science* 287:1770–1774.
- Schloss, A. L., and J. F. Haney. 2006. Clouds, shadows or twilight? Mayfly nymphs recognise the difference. *Freshwater Biology* 51:1079–1089.

- Schlosser, I. J. 1998. Fish recruitment, dispersal, and trophic interactions in a heterogeneous lotic environment. *Oecologia* 113:260–268.
- Schmidt, M. F. 1993. Maryland's geology. Tidewater Publishers. Centreville, Maryland.
- Sheldon, F., E. E. Peterson, E. L. Boone, S. Sippel, S. E. Bunn, and B. D. Harch. 2012. Identifying the spatial scale of land use that most strongly influences overall river ecosystem health score. *Ecological Applications* 22:2188–2203.
- Van Sickle, J., and C. B. Johnson. 2008. Parametric distance weighting of landscape influence on streams. *Landscape Ecology* 23:427–438.
- Snelder, T., J. B. Ortiz, D. Booker, N. Lamouroux, H. Pella, and U. Shankar. 2012. Can bottom-up procedures improve the performance of stream classifications? *Aquatic Sciences* 74:45–59.
- Soil Survey Staff. n.d. Natural Resources Conservation Service, United States Department of Agriculture. Web Soil Survey. <http://websoilsurvey.nrcs.usda.gov/>. Accessed 2013, 2014.
- Southerland, M. T., M. J. Kline, D. M. Boward, G. M. Rogers, R. P. Morgan II, P. F. Kazyak, R. J. Klauda, and S. A. Stranko. 2005. New biological indicators to better assess the condition of Maryland streams. Maryland Department of Natural Resources.
- Southerland, M. T., G. M. Rogers, M. J. Kline, R. P. Morgan, D. M. Boward, P. F. Kazyak, R. J. Klauda, and S. A. Stranko. 2007. Improving biological indicators to better assess the condition of streams. *Ecological Indicators* 7:751–767.
- Stanfield, L. W., and B. W. Kilgour. 2013. How proximity of land use affects stream fish and habitat. *River Research and Applications* 29:891–905.
- Stranko, S. A., D. M. Boward, J. Kilian, C. Millard, A. J. Becker, R. Gauza, A. Schenk, A. Roseberry-Lincoln, and M. O'Connor. 2007. Maryland biological stream survey sampling manual: field protocols. Maryland Department of Natural Resources.
- Stranko, S. A., S. E. Gresens, R. J. Klauda, J. V. Kilian, P. J. Ciccotto, M. J. Ashton, and A. J. Becker. 2010. Differential effects of urbanization and non-natives on imperiled stream species. *Northeastern Naturalist* 17:593–614.
- Stranko, S. A., R. H. Hilderbrand, R. P. Morgan, M. W. Staley, A. J. Becker, A. Roseberry-Lincoln, E. S. Perry, and P. T. Jacobson. 2008. Brook Trout Declines with Land Cover and Temperature Changes in Maryland. *North American Journal of Fisheries Management* 28:1223–1232.

- Synes, N. W., and P. E. Osborne. 2011. Choice of predictor variables as a source of uncertainty in continental-scale species distribution modelling under climate change: Predictor uncertainty in species distribution models. *Global Ecology and Biogeography* 20:904–914.
- Tarboton, D. 2014. Terrain analysis using digital elevation models (TauDEM). Version 5. <http://hydrology.usu.edu/taudem/taudem5/index.html>. Accessed Jan 2014.
- Todd, B. L., and C. F. Rabeni. 1989. Movement and Habitat Use by Stream-Dwelling Smallmouth Bass. *Transactions of the American Fisheries Society* 118:229–242.
- United States Geological Survey. 2014. Nonindigenous aquatic species database, Gainesville, Florida. <http://nas.er.usgs.gov>. Accessed Jan 2014.
- Urban, M. C., D. K. Skelly, D. Burchsted, W. Price, and S. Lowry. 2006. Stream communities across a rural-urban landscape gradient. *Diversity and Distributions* 12:337–350.
- Usio, N. 2007. Endangered crayfish in northern Japan: Distribution, abundance and microhabitat specificity in relation to stream and riparian environment. *Biological Conservation* 134:517–526.
- Utz, R. M., K. N. Eshleman, and R. H. Hilderbrand. 2011. Variation in physicochemical responses to urbanization in streams between two Mid-Atlantic physiographic regions. *Ecological Applications* 21:402–415.
- Utz, R. M., R. H. Hilderbrand, and D. M. Boward. 2009. Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. *Ecological Indicators* 9:556–567.
- Utz, R. M., R. H. Hilderbrand, and R. L. Raesly. 2010. Regional differences in patterns of fish species loss with changing land use. *Biological Conservation* 143:688–699.
- Verberk, W. C. E. P., H. Siepel, and H. Esselink. 2008. Life-history strategies in freshwater macroinvertebrates. *Freshwater Biology* 53:1722–1738.
- Walsh, C. J., and J. Kunapo. 2009. The importance of upland flow paths in determining urban effects on stream ecosystems. *Journal of the North American Benthological Society* 28:977–990.
- Walsh, C. J., A. H. Roy, J. W. Feminella, P. D. Cottingham, P. M. Groffman, and R. P. Morgan. 2005. The urban stream syndrome: current knowledge and the search for a cure. *Journal of the North American Benthological Society* 24:706–723.

- Wang, L., J. Lyons, and P. Kanehl. 2003. Impacts of Urban Land Cover on Trout Streams in Wisconsin and Minnesota. *Transactions of the American Fisheries Society* 132:825–839.
- Weijters, M. J., J. H. Janse, R. Alkemade, and J. T. A. Verhoeven. 2009. Quantifying the effect of catchment land use and water nutrient concentrations on freshwater river and stream biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems* 19:104–112.
- Whittaker, J. 1984. Model Interpretation from the Additive Elements of the Likelihood Function. *Applied Statistics* 33:52.
- Williams, K. J., L. Belbin, M. P. Austin, J. L. Stein, and S. Ferrier. 2012. Which environmental variables should I use in my biodiversity model? *International Journal of Geographical Information Science* 26:2009–2047.
- World Media Group. 2014. Maryland population density county rank. <http://www.use.com/rank/maryland-state-population-density--county-rank.htm>. Accessed 11 July 2014.
- Zimmermann, N. E., T. C. Edwards, G. G. Moisen, T. S. Frescino, and J. A. Blackard. 2007. Remote sensing-based predictors improve distribution models of rare, early successional and broadleaf tree species in Utah. *Journal of Applied Ecology* 44:1057–1067.