

Prepared in cooperation with the
Massachusetts Department of Conservation and Recreation, the
Massachusetts Department of Environmental Protection, and the
Massachusetts Department of Fish and Game

Factors Influencing Riverine Fish Assemblages in Massachusetts



Scientific Investigations Report 2011–5193

Cover. Clockwise from the top, blacknose dace, common shiner, white sucker, and brook trout. Images used with permission from the New York State Department of Environmental Conservation.

Factors Influencing Riverine Fish Assemblages in Massachusetts

By David S. Armstrong, Todd A. Richards, and Sara B. Levin

Prepared in cooperation with the
Massachusetts Department of Conservation and Recreation, the
Massachusetts Department of Environmental Protection, and the
Massachusetts Department of Fish and Game

Scientific Investigations Report 2011–5193

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2011

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit <http://www.usgs.gov> or call 1-888-ASK-USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

To order this and other USGS information products, visit <http://store.usgs.gov>

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this report is in the public domain, permission must be secured from the individual copyright owners to reproduce any copyrighted materials contained within this report.

Suggested citation:

Armstrong, D.S., Richards, T.A., and Levin, S.B., 2011, Factors influencing riverine fish assemblages in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2011-5193, 59 p. (Also available at <http://pubs.usgs.gov/sir/2011/5193>)

Acknowledgments

The authors are grateful to the Massachusetts Department of Conservation and Recreation (MDCR), Massachusetts Department of Environmental Protection (MDEP), Massachusetts Department of Fish and Game (MDFG), Executive Office of Energy and Environmental Affairs (EOEEA), members of the Sustainable Water Management Initiative, and other public agencies, private organizations, and individuals for supporting this study. Kathy Baskin and Vandana Rao of the EOEEA; Anne Carroll and Linda Hutchins of the MDCR; Warren Kimball, Thomas Lamonte, and Gerald Szal of the MDEP; and Mark Tisa of the Massachusetts Division of Fisheries and Wildlife (MDFW) are thanked for providing review comments for this report. The authors thank the personnel of MDFW and MDEP who collected the fish samples. The cooperation of public and private landowners who granted permission to access the rivers is appreciated. In addition, the authors thank Naomi Detenbeck and Ralph Abele of the U.S. Environmental Protection Agency, Seth Wenger of the U.S. Forest Service, Donald Orth of the Virginia Polytechnic Institute and State University; Matthew Baker of the University of Maryland-Baltimore County, and Richard Vogel of Tufts University for providing valuable advice during the project.

The authors thank Larry Brown and Chris Konrad of the U.S. Geological Survey (USGS) for technical review for this report and Mary Freeman, Brian Cade, Jim Coles, and Jonathan Kennen of the USGS for providing guidance and assistance during the project. Thanks also go to Lance Ostiguy and Pete Steeves of the USGS, who provided GIS assistance; Ruth Larkins and Dale Simmons, who provided editorial review; and Christine Mendelsohn, Ann Marie Squillacci, Susan Bergin and others in the USGS Pembroke Publishing Service Center for providing figures, tables, and report layout.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

Acknowledgments.....	iii
Abstract.....	1
Introduction.....	2
Purpose and Scope	2
Description of Study Area	3
Previous Studies	3
Factors Limiting Riverine Fishes.....	5
Methods of Determining Factors Affecting Fish Assemblages	6
Fish-Assemblage Characteristics	6
Environmental Factors	7
Simulated Streamflow Data and Measures of Flow Alteration	10
Selection of Environmental Factors and Fish Assemblage Metrics	16
Environmental Factor Selection using Principal Component Analysis and Spearman Rank Correlation	16
Fish Assemblage Classification using Ordination and Hierarchical Cluster Analysis.....	17
Analysis of Associations between Fish Assemblages and Environmental Factors	17
Quantile Regression	18
Generalized Linear Models	19
Factors Affecting Riverine Fish Assemblages	20
Environmental Factors	20
Streamflow-Alteration Indicators	24
Variable Reduction and Selection.....	27
Fish-Assemblage Characteristics	31
Patterns in Fish-Assemblage Structure.....	33
Quantile Regression	35
Generalized Linear Models	40
Fluvial-Fish Species-Richness Models	42
Fluvial-Fish Relative-Abundance Models.....	44
Brook Trout Relative-Abundance Model	46
Model Validation and Limitations	47
Comparison of Study Results to Preliminary Analysis.....	48
Possible Future Modeling Efforts.....	49
Summary and Conclusions.....	50
References Cited	51

Figures

1. Map showing Massachusetts Division of Fisheries and Wildlife fish-sampling sites used to assess environmental and anthropogenic factors influencing fish assemblages in Massachusetts streams.	4
2. Boxplots showing A, percent forest, B, percent wetland area, C, drainage area, D, channel slope, E, percent sand and gravel, and F, mean elevation for contributing areas to 669 fish-sampling sites on Massachusetts streams.	21
3. Boxplots showing A, undammed reach length, B, percent open water, and C, dam density per square mile for 669 fish-sampling sites on Massachusetts streams.	21
4. Graph showing cumulative frequency distributions of impervious cover for contributing areas to 669 fish-sampling sites and 1,429 subbasins and groundwater contributing areas across Massachusetts.	22
5. Map showing impervious cover for contributing areas to 669 fish-sampling sites on Massachusetts streams.	23
6. Graph showing cumulative frequency distributions of net percent alteration of August median flow for 669 fish-sampling sites and 1,429 subbasins and groundwater contributing areas across Massachusetts.	24
7. Map showing net percent alteration of August median flows at 669 fish-sampling sites on Massachusetts streams.	25
8. Scatterplot correlation matrix showing relations between A, environmental factors, and B, flow-alteration variables	28
9. Histograms showing relative abundance of A, fluvial fish, B, cold-water fish, C, brook trout, and D, white sucker in relation to percentage of selected sampling sites on Massachusetts streams.	31
10. Hierarchical cluster analysis dendrogram showing groupings of 21 fish species in Massachusetts.	33
11. Nonmetric multidimensional scaling ordinations of fish-assemblage structure showing a two-dimensional ordination with contours drawn around clusters defined by the 15-percent similarity from the cluster analysis shown in figure 10, for 669 fish-sampling sites on Massachusetts streams.	34
12. Graphs showing quantile regression relations between relative abundance of A, brook trout and B, blacknose dace and percent alteration of August median flow from groundwater withdrawals for selected fish-sampling sites on Massachusetts streams.	36
13. Graphs showing quantile regression relations between A, fluvial-fish relative abundance and B, fluvial-fish species richness and percent alteration of August median flow from groundwater withdrawals for selected fish-sampling sites on Massachusetts streams.	36
14. Graphs showing quantile regression relations between relative abundance of A, blacknose dace and B, white sucker and percent alteration of August median flow from surface-water returns for selected fish-sampling sites on Massachusetts streams.	37
15. Graphs showing quantile regression relations between relative abundance of A, brook trout and B, common shiner and percent impervious cover for the contributing areas to selected fish-sampling sites on Massachusetts streams.....	38
16. Graphs showing quantile regression relations between A, fluvial-fish relative abundance and B, fluvial-fish species richness and percent impervious cover for the contributing areas to selected fish-sampling sites on Massachusetts streams.	38

17. Graphs showing quantile regression relations between A, relative abundance of brook trout and B, fluvial-fish species richness and percent open water for the contributing areas to selected fish-sampling sites on Massachusetts streams.	39
18. Graphs showing quantile regression relations between A, relative abundance of brook trout and B, fluvial-fish species richness and undammed reach length for selected fish-sampling sites on Massachusetts streams.	39
19. Graphs showing generalized linear model output for relations between fluvial-fish species richness and A, percent impervious cover in the percent impervious cover model and B, percent forest in the percent forest model.....	43
20. Graphs showing generalized linear model output for relations between fluvial-fish relative abundance and A, percent alteration of August median flow from groundwater withdrawals in the percent impervious cover model, B, percent impervious cover in the percent impervious cover model, C, percent alteration of August median flow from groundwater withdrawals in the percent forest model, and D, percent impervious cover in the percent forest model	45
21. Graphs showing generalized linear model output for relations between brook trout relative abundance and A, percent impervious cover in the percent impervious cover model and B, percent forest in the percent forest model.	47

Tables

1. Common and scientific names of fishes, species codes, and habitat-use classifications for each species	8
2. Names and variable codes for environmental factors tested for use as explanatory variables.....	9
3. Names and variable codes for land-cover and land-use variables tested for use as explanatory variables.....	12
4. Names and variable codes for flow-alteration variables tested for use as explanatory variables: A, net-flow alteration variables; and B, flow-alteration components	14
5. Descriptions, reference codes, units, and statistical summaries of values for flow-alteration variables	26
6. Descriptions, reference codes, units, and statistical summaries of values for environmental and anthropogenic variables	30
7. Common names of fish, number and percentage of sites sorted by the number of sites at which each species was captured, and number and percentage of individuals sorted by the number of individuals captured at sites used for the assessment of fish assemblages in Massachusetts streams.....	32
8. A, Significant variables and coefficients and B, measures of goodness of fit for generalized linear model equations used to determine relations between environmental and anthropogenic factors and indicator species and fish-assemblage metrics in Massachusetts streams.....	41

Conversion Factors, Datum, and Acronyms

Inch/Pound to SI

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
yard (yd)	0.9144	meter (m)
Area		
acre	4,047	square meter (m ²)
square mile (mi ²)	259.0	hectare (ha)
square mile (mi ²)	2.590	square kilometer (km ²)
Volume		
gallon (gal)	3.785	liter (L)
gallon (gal)	0.003785	cubic meter (m ³)
gallon (gal)	3.785	cubic decimeter (dm ³)
million gallons (Mgal)	3,785	cubic meter (m ³)
cubic foot (ft ³)	0.02832	cubic meter (m ³)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)
cubic foot per second per square mile [(ft ³ /s)/mi ²]	0.01093	cubic meter per second per square kilometer [(m ³ /s)/km ²]
gallon per minute (gal/min)	0.06309	liter per second (L/s)
gallon per day (gal/d)	0.003785	cubic meter per day (m ³ /d)
gallon per day per square mile [(gal/d)/mi ²]	0.001461	cubic meter per day per square kilometer [(m ³ /d)/km ²]
million gallons per day (Mgal/d)	0.04381	cubic meter per second (m ³ /s)
million gallons per day per square mile [(Mgal/d)/mi ²]	1,461	cubic meter per day per square kilometer [(m ³ /d)/km ²]

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$^{\circ}\text{F}=(1.8\times^{\circ}\text{C})+32$$

Temperature in degrees Fahrenheit (°F) may be converted to degrees Celsius (°C) as follows:

$$^{\circ}\text{C}=(^{\circ}\text{F}-32)/1.8$$

Horizontal coordinate information is referenced to the North American Datum of 1983 (NAD 83).

Vertical coordinate information is referenced to the North American Vertical Datum of 1988 (NAVD 88).

Acronyms and Abbreviations

DEM	Digital elevation model
EOEEA	Executive Office of Energy and Environmental Affairs
FD	Fluvial dependent
FS	Fluvial specialist
GLM	Generalized linear model
HUC	Habitat Use Classification
IC	Impervious cover
MDCR	Massachusetts Department of Conservation and Recreation
MDEP	Massachusetts Department of Environmental Protection
MDFG	Massachusetts Department of Fish and Game
MDFW	Massachusetts Division of Fisheries and Wildlife
MG	Macrohabitat generalist
NHD	National Hydrography Dataset
NMDS	Nonmetric multidimensional scaling
NPDES	National Pollutant Discharge Elimination System
OLS	Ordinary least squares
SYE	Sustainable Yield Estimator
TITAN	Threshold Indicator Taxa ANalysis
USGS	U.S. Geological Survey
VIF	Variance inflation factor
WUI	Water-use intensity
YOY	Young-of-the-year
ZINB	Zero-inflated negative binomial

THIS PAGE INTENTIONALLY LEFT BLANK

Factors Influencing Riverine Fish Assemblages in Massachusetts

By David S. Armstrong, Todd A. Richards¹, and Sara B. Levin

Abstract

The U.S. Geological Survey, in cooperation with the Massachusetts Department of Conservation and Recreation, Massachusetts Department of Environmental Protection, and the Massachusetts Department of Fish and Game, conducted an investigation of fish assemblages in small- to medium-sized Massachusetts streams. The objective of this study was to determine relations between fish-assemblage characteristics and anthropogenic factors, including impervious cover and estimated flow alteration, relative to the effects of environmental factors, including physical-basin characteristics and land use. The results of this investigation supersede those of a preliminary analysis published in 2010. Fish data were obtained for 669 fish-sampling sites from the Massachusetts Division of Fisheries and Wildlife fish-community database. A review of the literature was used to select fish metrics—species richness, abundance of individual species, and abundances of species grouped on life history traits—responsive to flow alteration. The contributing areas to the fish-sampling sites were delineated and used with a geographic information system to determine a set of environmental and anthropogenic factors that were tested for use as explanatory variables in regression models. Reported and estimated withdrawals and return flows were used together with simulated unaltered streamflows to estimate altered streamflows and indicators of flow alteration for each fish-sampling site. Altered streamflows and indicators of flow alteration were calculated on the basis of methods developed in a previous U.S. Geological Survey study in which unaltered daily streamflows were simulated for a 44-year period (water years² 1961–2004), and streamflow alterations were estimated by use of water-withdrawal and wastewater-return data previously reported to the State for the 2000–04 period and estimated domestic-well withdrawals and septic-system discharges.

¹Massachusetts Department of Fish and Game, Division of Fisheries and Wildlife.

²A water year is the 12-month period from October 1 to September 30. It is designated by the year in which it ends.

A variable selection process, conducted using principal components analysis and Spearman rank correlation, was used to select a set of 15 non-redundant environmental and anthropogenic factors to test for use as explanatory variables in the regression analyses. Twenty-one fish species were used in a multivariate analysis of fish-assemblage patterns. Results of nonmetric multidimensional scaling and hierarchical cluster analysis were used to group fish species into fluvial and macrohabitat generalist habitat-use classes.

Two analytical techniques, quantile regression and generalized linear modeling, were applied to characterize the association between fish-response variables and environmental and anthropogenic explanatory variables. Quantile regression demonstrated that as percent impervious cover and an indicator of percent alteration of August median flow from groundwater withdrawals increase, the relative abundance and species richness of fluvial fish decrease. The quantile regression plots indicate that (1) as many as seven fluvial fish species are expected in streams with little flow alteration or impervious cover, (2) no more than four fluvial fish species are expected in streams where flow alterations from groundwater withdrawals exceed 50 percent of the August median flow or the percent area of impervious cover exceeds 15 percent, and (3) few fluvial fish remain at high rates of withdrawal (approaching 100 percent) or high rates of impervious cover (between 25 and 30 percent). Three generalized linear models (GLMs) were developed to quantify the response of fluvial fish to multiple environmental and anthropogenic variables. All variables in the GLM equations were demonstrated to be significant (p less than 0.05, with most less than 0.01). Variables in the fluvial-fish relative-abundance model were channel slope, estimated percent alteration of August median flow from groundwater withdrawals, percent wetland in a 240-meter buffer strip, and percent impervious cover. Variables in the fluvial-fish species-richness model were drainage area, channel slope, total undammed reach length, percent wetland in a 240-meter buffer strip, and percent impervious cover. Variables in the brook trout relative-abundance model were drainage area, percent open water, and percent impervious cover. The variability explained by the GLM models, as measured by the pseudo R^2 , ranged from 18.2 to 34.6, and correlations between observed and predicted

2 Factors Influencing Riverine Fish Assemblages in Massachusetts

values ranged from 0.50 to 0.60. Results of GLM models indicated that, keeping all other variables the same, a one-unit (1 percent) increase in the percent depletion of August median flow would result in a 0.9-percent decrease in the relative abundance (in counts per hour) of fluvial fish. The results of GLM models also indicated that a unit increase in impervious cover (1 percent) resulted in a 3.7-percent decrease in the relative abundance of fluvial fish, a 5.4-percent decrease in fluvial-fish species richness, and an 8.7-percent decrease in brook trout relative abundance.

Introduction

Flow alteration and urbanization have been associated with stream-ecosystem degradation in flowing waters, as measured through changes in fish (Karr and Chu, 1999; Freeman and Marcinek, 2006; Zorn and others, 2008; Poff and others, 2010) and macroinvertebrate communities (Coles and others, 2004; Konrad and others, 2008; Kennen and others, 2009). Although streamflow is one of many factors that affect the abundance and distribution of fish, it has been called the master variable because it affects habitat diversity and availability through its influence on stream geomorphology, channel substrate, water depth, velocity, and other factors that, in turn, affect habitat quality, such as water temperature and water quality (Poff and others, 1997; Wilding and Poff, 2008; Poff and Zimmerman, 2010). The likelihood that flow alteration and urbanization will degrade aquatic communities is generally acknowledged by the scientific community and by water managers (Poff and Zimmerman, 2010), but the response of aquatic communities to flow alteration relative to the effects of physical-basin characteristics, land-use, and anthropogenic factors is poorly understood.

Understanding species-stressor and species-environment relations is an important step toward the conservation of riverine fish. Several recent studies of the response of fish assemblages to flow alteration in southern New England streams have determined that as flow alteration increases, the composition of the fish assemblage shifts as fish dependent upon flow during various stages of their life cycles (fluvial fish) decrease in number (abundance) and species richness (number of species) (Armstrong and others, 2001, 2004; Parasiewicz, 2004; Kanno and Vokoun, 2010). Although declines in fluvial-fish metrics are an expected response to increased flow depletion, the level of flow alteration at which responses occur, and the magnitude, rate, and shape of the response curve (for example, linear or nonlinear) are not well understood.

Until 2009, streamflow-simulation tools that could be used to estimate unaltered daily streamflows for ungaged sites and daily streamflows during altered water-use conditions were unavailable except for drainage basins where streamflows could be simulated by use of a hydrologic model. However, the U.S. Geological Survey (USGS),

in cooperation with the Massachusetts Department of Environmental Protection (MDEP), recently developed a computer application—the Sustainable Yield Estimator (SYE, version 1.0; Archfield and others, 2010)—that can be used to estimate altered and unaltered daily streamflows at ungaged sites. Recently (2010), the USGS developed methods for estimating flow alteration indicators and other indicators of anthropogenic alteration for 1,429 subbasins in Massachusetts (Weiskel and others, 2010; referred to as the Massachusetts Water Indicators study in this report). These new capabilities, together with new (2005) geographic information system (GIS) data layers of impervious cover and a comprehensive fish-community database established and maintained by the Massachusetts Division of Fisheries and Wildlife (MDFW), have made possible this statewide assessment of the relations among fish assemblages and environmental and anthropogenic factors.

Massachusetts State agencies are interested in determining the effects of flow alteration on fish assemblages in Massachusetts streams and rivers relative to the effects of other anthropogenic factors, such as impervious cover and dams, and to environmental factors. In 2009, a 3-year cooperative study was initiated for this purpose by the USGS and the Massachusetts Department of Conservation and Recreation (MDCR) with additional cooperation from the MDEP, and the Massachusetts Department of Fish and Game (MDFG). The results of this study are expected to provide information on fish assemblages that could be used by water-resource managers and the Massachusetts Executive Office of Energy and Environmental Affairs (EOEEA) Sustainable Water Management Initiative (SWMI) to make more informed decisions about managing human factors that affect aquatic habitat in Massachusetts.

Purpose and Scope

This report describes associations between stream fish assemblages in Massachusetts and anthropogenic factors, such as indicators of flow alteration, impervious cover, and dams, relative to the effects of basin characteristics and natural environmental conditions. Multivariate analyses that were used to aid in the selection of environmental variables for analysis and to describe patterns in the fish-assemblage structure are presented. The use of quantile regression to assess the effects of anthropogenic factors as limiting factors for fluvial-fish assemblages and fluvial-fish species is described. Generalized linear model (GLM) equations that quantify relations between selected fish-response variables and multiple environmental and anthropogenic explanatory variables are presented.

Small to medium streams across Massachusetts were studied but not streams on Cape Cod, the Islands (Martha's Vineyard, Nantucket), or parts of southeastern Massachusetts where information on simulated streamflows was unavailable. Data on fish samples from the MDFW database collected

during 1998–2008 were used in the analysis of fish assemblages; new fish-sampling data were not collected for this study. Unaltered and altered flows used in the study were simulated for fish-sampling sites using a modeling tool and approaches for estimating water-use developed by the USGS.

Description of Study Area

The study sites are on free-flowing (non-impounded) reaches of small- to medium-sized (wadeable) freshwater streams in Massachusetts that were selected from the MDFW database (fig. 1). The fish-sampling sites include multiple sites nested within drainage basins such that the contributing areas to some fish-sampling sites might also be included in the contributing areas of downstream sampling sites. The contributing areas to some fish-sampling sites in the study include areas in adjacent states. Water use, physical-basin characteristics, and land use for the parts of contributing drainage areas outside of Massachusetts are included in the analysis.

Fish species in Massachusetts are described in Hartel and others (2002). Freshwater fish in the southern New England states are characterized by low-species diversity in comparison to the diversity in the midwestern and southeastern United States (Whitworth, 1996; Kanno and Vokoun, 2008; Meador and Carlisle, 2009; Kanno and others, 2010). Consequently, there are relatively few species in Massachusetts that can be used to detect responses to anthropogenic stressors. The composition of fish assemblages in relatively unaltered rivers in New England and the expected composition of fish assemblages appropriate for restoration of main-stem rivers (referred to as Target Fish Communities) are described in Bain and Meixler (2000, 2008), Meixler, (2011), Parasiewicz (2004), Legros and Parasiewicz (2007), University of New Hampshire and others (2008), Yoder and others (2006, 2009), and Kashiwagi and Richards (2009).

The rivers and landscape of Massachusetts have been altered for hundreds of years, to the extent that almost all streams and rivers in the State reflect some aspect of human disturbance. Human alterations to rivers are extensive and multifaceted and include physical alterations created by dams, road crossings, channelization, and riparian (riverbank) and flood-plain development; flow alterations from water withdrawals, diversions, wastewater-return flows, hydropower, industrial regulation, and urban runoff; and water-quality and sediment alterations caused by wastewater and septic discharges, highway runoff, industrial contaminants, and runoff from urban, suburban, and agricultural lands. These alterations to rivers, coupled with indirect effects of urbanization, impervious cover, and other land-use disturbances across the landscape can have cumulative effects on aquatic biota. Many of these alterations can cause shifts in fish-assemblage attributes. Flow alterations and dams can reduce habitat suitability for some species (lotic and coldwater species, such as several minnow species and trout) and create habitat for others (lentic species, such as sunfish and bass).

This shift has prompted the State to list most native riverine fish species in the Massachusetts Wildlife Action Plan as “Species in Greatest Need of Conservation” (Massachusetts Division of Fisheries and Wildlife, 2006).

Land use in Massachusetts is a mosaic of forest, wetland, and open water interspersed with urban, industrial, suburban, and agricultural land. The distribution of land uses reflects the history of settlement in the region and is closely related to geologic and topographic features, river corridors, and transportation routes (Foster and Aber, 2004; Harrison and Judd, 2011). Even areas classified as mostly forested have some roads, low-density housing, or remnants of historic disturbance to the land and waterways. Although large areas of contiguous forest still exist at high elevations in the Berkshire Mountains in western Massachusetts, many areas in eastern and central Massachusetts are experiencing rapid conversion of forested and agricultural lands to residential and commercial uses (Kennen and others, 2009). As a consequence of the cumulative effect of flow alterations, stream fragmentation, and landscape disturbance, the natural patterns of fish distribution and abundance have been altered in many drainage basins in Massachusetts (Hartel and others, 2002). Despite these historical and ongoing alterations, Massachusetts has areas of high ecological integrity and more than 40 species of native freshwater fish.

Previous Studies

The USGS began a series of studies in 1995 to determine the spatial distribution of, and correlation among, factors related to aquatic habitats, aquatic communities, and flow conditions of Massachusetts streams. For these studies, done in cooperation with the MDCR, Office of Water Resources (formerly the Massachusetts Department of Environmental Management), and the MDEP, several different approaches for determining streamflow requirements for habitat protection were evaluated. Streamflow requirements determined from median daily mean flows for August and from wetted-perimeter calculations determined from discharge measurements are described by Ries (1997) and Mackey and others (1998), respectively. Recurrent low-flow conditions caused by groundwater withdrawals from the Ipswich River in Massachusetts were associated with low abundance and proportions of fluvial fish (fish that require flowing water habitats to complete part or all of their life cycle) (Armstrong and others, 2001). Relations among stream habitat, fish assemblages, and hydrologic conditions were investigated in the Charles and Assabet River Basins by Parker and others (2004), and in the Sudbury River Basin by Zariello and others (2010). Streamflows and methods for determining streamflow requirements for aquatic-habitat protection at index stations in southern New England were studied by Armstrong and others (2004) and Parker and others (2004). Armstrong and others (2008) characterized and classified least-altered streamflows in Massachusetts.

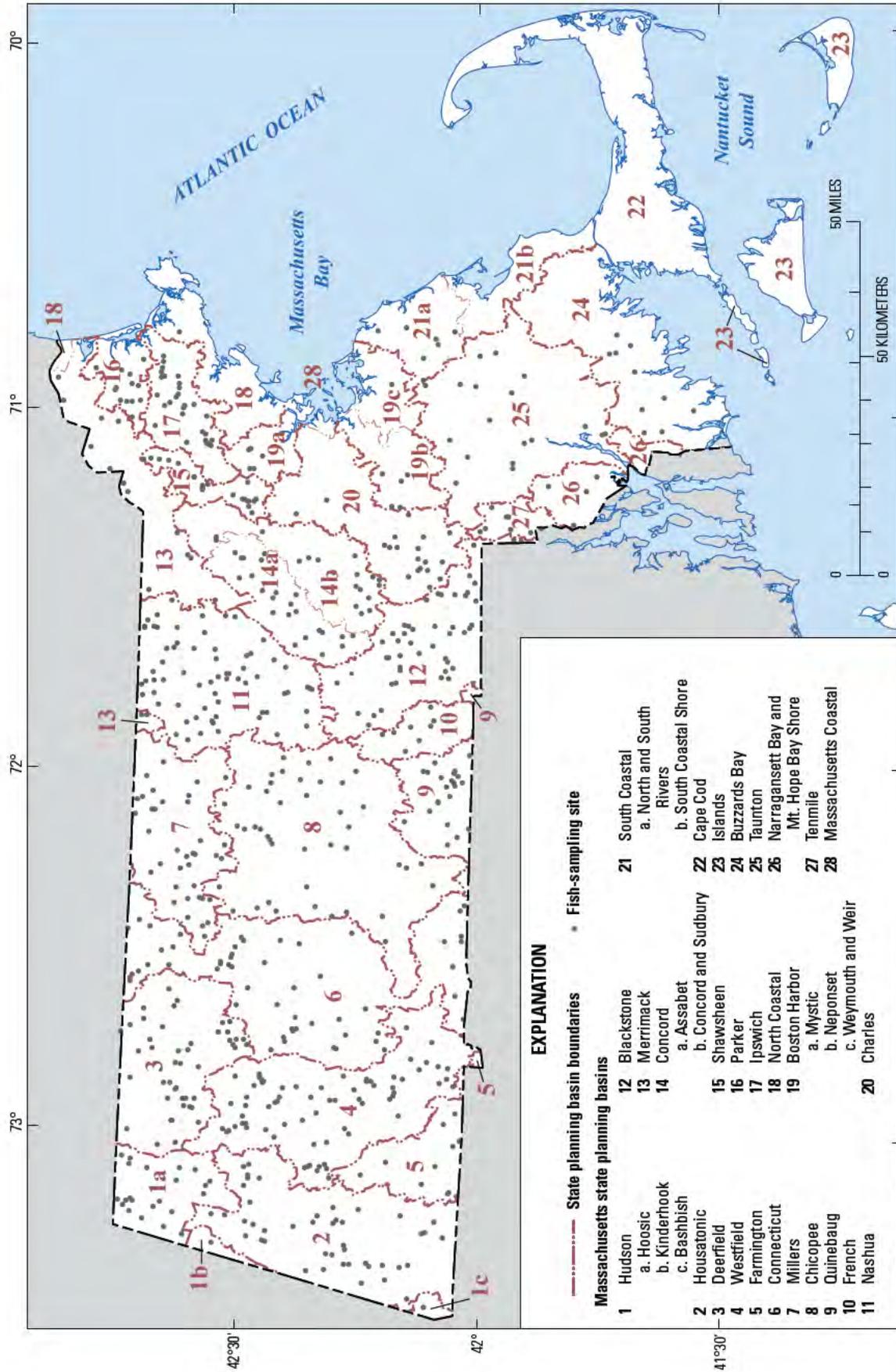


Figure 1. Massachusetts Division of Fisheries and Wildlife fish-sampling sites used to assess environmental and anthropogenic factors influencing fish assemblages in Massachusetts streams. Fish samples were collected from 1998 to 2008.

Several studies in the past decade (2000–10) have described the effects of flow alteration, increased urbanization, and impoundments on fish and macroinvertebrate communities in streams in New England and the northeastern United States. Connecticut streams subjected to high rates of water withdrawal were characterized by low proportions of fluvial fish and benthic insectivores (fish that feed on bottom-dwelling stream insects in riffle habitats) (Kanno and Vokoun, 2010). Urbanization has been shown to be associated with declines in cyprinids (minnows), benthic insectivores, and fluvial dependent fish, as well as declines in invertebrate richness for Ephemeroptera, Plecoptera, and Tricoptera (EPT), dipteran, and midge taxa (Coles and others, 2004; Kennen and others, 2009; Meador and others, 2008; Cuffney and others, 2011). Dams and impoundments alter fish-assemblage structure by shifting the abundance and diversity of fish species from fluvial to generalists (Kinsolving, and Bain, 1993; Bushaw-Newton and others, 2002; Freeman and Marcinek, 2006; Gardner and others, 2011). Flow regulation from hydropower has been shown to reduce the diversity of fish species, particularly those that utilize shallow slow-water habitats (Bain and others, 1988). Fragmentation of stream networks by dams and road crossings make some stream reaches impassable for upstream fish migration and have been associated with reduced upstream abundance and diversity of resident fish (Letcher and others, 2007; Nislow and others, 2011).

In the fall of 2009, Massachusetts State agencies convened the Sustainable Water Management Initiative (SWMI) with the goals of re-instituting the process of determining safe yield, integrating safe yield values into an improved statewide water-allocation program, and developing streamflow criteria to better protect aquatic habitat from the effects of water withdrawals and other stressors. The SWMI prompted the MDEP and MDFG to fund the USGS to accelerate aspects of the project that would lead to the development of streamflow criteria. The accelerated investigation necessitated limiting the scope of data collection and data analysis so that a preliminary report (Armstrong and others, 2010) could be completed as quickly as possible. Generalized linear models were used in the preliminary study to link fluvial-fish relative abundance, fluvial-fish richness, and brook trout relative abundance to a suite of environmental and stressor variables. After the preliminary report was completed, the USGS continued the investigation using a broader suite of fish-assemblage and explanatory variables and a wider range of statistical techniques for variable selection and reduction in order to address the response of fish assemblages to environmental and anthropogenic factors in greater depth.

Factors Limiting Riverine Fishes

Fish assemblages are considered to be good indicators of the condition of the habitat in which they live (Karr and Chu, 1999; Poff and Zimmerman, 2010). The distribution and abundance of fish are determined by environmental variables and anthropogenic factors at multiple scales. Many natural environmental factors play a role in structuring stream-fish assemblages, including flow regime, water quality, stream temperature, stream-channel characteristics, habitat availability and connectivity, physiographic characteristics, land cover, and biotic interactions (Gido and others, 2006; Pritchett and Pyron, 2011). Anthropogenic factors, such as flow and water-quality alterations, dams and impoundments, habitat loss, and urbanization of land, also have major effects on aquatic habitat and fish assemblages (Paul and Meyer, 2001; Coles and others, 2004; Konrad and Booth, 2005; Brown and others, 2009b; Poff and Zimmerman, 2010; Wang and others, 2001a, 2010, 2011; Nislow and others, 2011).

Fish-community responses to flow alterations include loss of sensitive species, reduced diversity, altered assemblages and dominant taxa, reduced abundance, and increases in non-native species (Poff and Zimmerman, 2010). More than three decades of research have illustrated the effects of water withdrawals, the damming and channelization of streams, and urbanization on aquatic communities, including fish (Walsh and others, 2005; Freeman and Marcinek, 2006; Wenger and others, 2009; Carlisle and others, 2010; Poff and Zimmerman, 2010; Meador and Carlisle, 2011). The result has been a widespread call for streamflows that mimic the magnitude, frequency, duration, timing, and rate of change of natural streamflows (the natural flow regime; Poff and others, 1997, 2010; Richter and others, 1997, 2011; Bunn and Arthington, 2002; Annear and others, 2004; Arthington and others, 2006).

Streamflows in Massachusetts typically are highest in spring and lowest in late summer and early fall. Low flows in summer, which maintain critical habitat for fish, also correspond to the period of highest water use and tend to receive the most attention. Nearly natural flows in each season, however, are important for maintaining habitat for fish populations. Moderate flows in winter provide stable habitat during harsh winter conditions. Flood flows in early spring maintain habitat quality by scouring streambeds, modifying channel geomorphology, and importing woody vegetation into stream channels; moderately high flows in late spring and fall facilitate fish migration and provide spawning habitat critical for many species.

Increased urbanization is associated with degraded aquatic communities (Wang and others, 2001a, 2001b; Meador and others, 2005; Roy and others, 2005; Walsh and others, 2005; Wenger and others, 2008, 2009; Brown and others, 2009b; Baker and King, 2010; Cuffney and others, 2011). Urban effects on stream ecosystems are caused by multiple mechanisms, including altered streamflow through increased stormwater runoff, reduced recharge, reduced base flow, and altered stream geomorphology through changes

in sediment supply, erosion, and the filling and piping of headwater channels. Although urbanization can directly alter streamflow through acceleration of the timing and increases in the magnitude of stormwater runoff and decreases in the rates of recharge and evapotranspiration, its effects on aquatic communities are not entirely flow related. Urbanization degrades water quality by increasing concentrations of major ions, nutrients, and toxicants; increasing water temperatures; and decreasing dissolved oxygen in streams. Urbanization also degrades physical habitat by modifying channel bank shape and materials, increasing impervious cover, removing riparian vegetation, decreasing shade, reducing input of woody vegetation, and introducing barriers to movement (Wenger and others, 2009). Urbanization represents the accumulation of many factors that together impair the integrity of aquatic communities.

Methods of Determining Factors Affecting Fish Assemblages

Site-specific fish-assemblage information was evaluated in relation to sub-regional and regional-scale environmental and anthropogenic factors, including indicators of streamflow alteration and impervious cover. Associations between flow-alteration indicators, impervious cover, and ecological response were determined by use of a space-for-time substitution; that is, ecological conditions are characterized at sites along a gradient of alteration (space) rather than tracking changes in individual species or fish-assemblage composition at a given site over multiple years (time) (Pickett, 1989; Wickham and others, 1997; Carter and others, 2009; Brown and others, 2009a). Several statistical techniques, including quantile regression, multivariate analysis, and generalized linear modeling, were applied to determine the association between fish-response variables, and environmental and anthropogenic explanatory variables.

Fish-Assemblage Characteristics

Information on fisheries was evaluated for more than 2,000 fish-sampling sites documented in the MDFW fish database. The database contains information on fish-sampling activities conducted in streams, rivers, lakes, and ponds. For the purposes of this analysis, fish samples were limited to collections made only in free-flowing (non-impounded) reaches of streams. Although the effects of dams in converting river reaches from lotic (riverine) to lentic (ponded) systems are well documented (Poff and Hart, 2002; Poff and others, 2010; Poff and Zimmerman, 2010; Wang and others, 2010), fish-assemblage samples from within reservoirs, lakes, and ponds were not included in the analysis because these settings were not representative of riverine habitat. The fish samples were collected from 1998 to 2007 during the summer

base-flow season (typically mid-June to mid-September) using MDFW protocols designed to gather a cross section of the fish assemblage in wadeable streams (Appendix 1 of Armstrong and others, 2008). The fish-assemblage data in the MDFW fish database were collected, with few exceptions, by the MDFW for statewide monitoring purposes, and samples were not collected to specifically address the potential effects of flow alteration or impervious cover on fish-assemblage structure.

The MDFW fish database was screened for the fish-sampling data used in this study. Fish samples were limited to those collected by use of backpack and barge electrofishing gear (pulsed direct current) during a single upstream pass. Single-pass backpack electrofishing has been determined to be an effective and reliable approach for obtaining information on fish-assemblage structure, such as species richness, when based on consistent sampling protocols (Angermeier and Smogler, 1995; Bertrand and others, 2006; Price and Peterson, 2010; but see Meador and others, 2003). Fish-sampling data were screened for electrofishing sampling efficiency to improve sample consistency and to exclude sites where sampling conditions precluded capture of a representative cross section of the fish assemblage. Sampling sites were removed from the study because of high water, high turbidity, or equipment failure. Stream reaches were evaluated for sample length and length-to-width ratio. Sampling sites in stream reaches of greater than 91 meters (100 yards) were included. Sites in reaches of less than 91 meters were scrutinized for length-to-width ratio to ensure that the habitat was sufficiently sampled. Reaches as short as 30 meters were included only if the average reach width was 1 meter or less. At sites where multiple-pass sampling methods were employed, only fish collected on the first pass were counted.

The MDFW database included many sites that were sampled multiple times over nearly a decade. To reduce pseudoreplication (samples that are not independent) and spatial autocorrelation of environmental variables (similarity among variables based on the proximity of sample sites) (Hurlbert, 1984), sites used in this study were restricted to those that were more than 0.5 kilometers (km) from each other (Wenger and others, 2008). Sites that were closer than 0.5 km were identified as a family of sites. The site with the largest watershed area from each family of sites was included in this study. In the event of a tie, a site was selected randomly.

Only year-round-resident freshwater fish species were included in the fish analysis (table 1). The occurrence of non-resident individuals can be unpredictable and not representative of site-specific environmental conditions (Angermeier and Karr, 1986). Anadromous fish (alewife, blueback herring, American shad, striped bass, sea lamprey) and estuarine fish (mummichog) were excluded. These species are uncommon in the MDFW data base, and all life stages of these species do not occur in the river systems during the entire sampling period. American eel, a catadromous species, was included in the analysis because this species enters freshwater as juveniles and spends most of its adult life (potentially 15 years) in freshwater systems. Trout greater than

200 millimeters (mm) (7.9 inches) were excluded from the analysis to avoid including trout stocked as adults. Although some wild trout greater than 200 mm may have been excluded, their numbers were low and would not have biased the results. All Atlantic salmon were excluded as these fish were stocked as fry as part of the Atlantic salmon restoration effort. Landlocked salmon were not removed, however, because they are produced through natural reproduction. Fish smaller than 40 mm (1.6 inches), representing early life stage young-of-the-year (YOY), were excluded because they are not sampled well by electrofishing, are difficult to identify, have a patchy distribution, and can have exceedingly high year-to-year variability in production in response to natural factors (such as a period of lower than average air temperatures or exceedingly high or low flows) or anthropogenic disturbances. Removal of YOY is a recommended practice in the development of Indexes of Biotic Integrity (IBIs) (Halliwell and others, 1999; Angermeier and Karr, 1986) because YOY numbers can be highly variable through the summer season and because they can occur in large numbers seasonally, potentially skewing a data set (Angermeier and Karr, 1986). Fish-sampling sites were included in this analysis only if they met drainage-area and location criteria required by the Sustainable Yield Estimator (SYE; Archfield and others, 2010), a decision-support tool. Thus sites were limited to those with drainage areas greater than 2 square miles (mi²).

The potential for a large number of zeros in the data set created by retention of infrequently captured (rare) fish species can strongly distort some multivariate analyses (Clark and Warwick, 2001). Accordingly, rare species were omitted for multivariate analyses. This was done by removing species that accounted for less than 0.5 percent of total abundance.

Fish were assigned a habitat-use classification on the basis of published life-history traits (Kinsolving and Bain, 1993; Bain and Meixler, 2000, 2008; Halliwell and others, 1999; Kashiwagi and Richards, 2009). This classification system has been acknowledged to provide insight into anthropogenic alteration in terms of water quantity and has been used in other statewide regional studies, such as Freeman and Marcinek (2006), Yoder and others (2009), Vokoun and Kanno (2009), and Meador and Carlisle (2011). Habitat-use classifications include fluvial specialists, fluvial dependents, and macrohabitat generalists (table 1). Fluvial specialists such as blacknose dace and fallfish require flowing-water habitats throughout their life cycles (Bain and Meixler, 2008). Fluvial dependents, such as common shiner and white sucker, require access to streams or flowing-water habitats for a specific life stage but otherwise can be found in lakes, reservoirs, and rivers. A metric combining fluvial-specialist and fluvial-dependent species into a single fluvial classification was tested. Macrohabitat generalists, such as pumpkinseed and redbfin pickerel, use a broad range of habitat. Macrohabitat generalists include species commonly found in lakes, reservoirs, and rivers and can complete their life cycles in any one of these systems. Additional classifications tested included

tolerance (tolerant, moderately tolerant, and intolerant) and temperature classes (coldwater, coolwater, and warmwater).

Measures of fish relative abundance and richness were determined from fish species information. The relative-abundance metrics for individual fish species classifications were calculated on the basis of counts of the various taxa present at each sampling site. Although some bias may exist in these metrics as a result of the incomplete capture of individual fish and fish species during sampling, any bias is assumed to be constant across all samples. To standardize the sampling effort, relative abundance was expressed in terms of fish catch per unit effort (CPUE), in counts per hour of electrofishing. CPUE calculations were made by dividing abundance by the total time of electrofishing over the reach.

Environmental Factors

Environmental factors tested for use as explanatory variables in this study are mostly sub-regional to regional-scale physical-basin and natural land-cover characteristics that have been identified in the literature as being related to fish-community attributes (Matthews, 1998; Marsh-Matthews and others, 2000; Magalhaes and others, 2002; Diana and others, 2006; Hoeninghaus and others, 2007). Regional scale environmental factors tested include, among others, contributing drainage area, elevation, basin slope, and percent area of sand and gravel, forest, wetland, and open water (table 2). Regional-scale environmental factors are thought to structure stream-fish assemblages through their association with finer-scale habitat features and variables that can have a physiological effect on organisms, such as stream velocity, water quality, water temperature, habitat complexity, habitat space available for different life stages, or other factors, such as environmental stability and ecosystem productivity (Poff and others, 1997; Cereghino and others, 2005; Creque and others, 2005; Gido and others, 2006; Austin, 2007). Environmental factors are not expected to change substantially over short-term periods unless altered by humans.

Regional-scale environmental factors help to determine the expected fish assemblage at a study site. Fish species in large- and medium-sized rivers tend to be more diverse than in small streams and brooks (Goldstein and Meador, 2004). Fish assemblages in streams at higher elevations typically contain fewer species than those at lower elevations (Wickham and others, 1997). Stream gradient generally varies with velocity and substrate and these variables can affect fish-assemblage composition. Some fish species or life stages prefer high-gradient stream habitats, but others prefer low-gradient reaches (Argent and others, 2003). High-gradient streams typically have a greater prevalence of gravel, cobble, boulder, or rock substrate, whereas sand substrates and aquatic vegetation are more common in low-gradient streams. In the glaciated Northeast, a high percentage of sand and gravel areas in a basin generally indicates the presence of aquifers that discharge groundwater to streams (Randall, 2001).

8 Factors Influencing Riverine Fish Assemblages in Massachusetts

Table 1. Common and scientific names of fishes, species codes, and habitat-use classifications for each species.

[Species code: Massachusetts Division of Fisheries and Wildlife species abbreviation. HUC, habitat-use classification (Kashiwage and Richards, 2009): FD, fluvial; FS, fluvial specialist; MG, macrohabitat generalist; --, species not categorized]

Common name	Species code	Family name	Scientific name (genus and species)	HUC
Petromyzontidae (lampreys)				
American brook lamprey	BL		Lampetra appendix	--
Anguillidae (freshwater eels)				
American eel	AE		Anguilla rostrata	MG
Cyprinidae (carp and minnows)				
Goldfish	G		Carassius auratus	MG
Lake chub	LC		Couesius plumbeus	MG
Common carp	C		Cyprinus carpio	MG
Cutlips minnow	CLM		Exoglossum maxillingua	FS
Common shiner	CS		Luxillus cornutus	FD
Golden shiner	GS		Notemigonus crysoleucas	MG
Bridle shiner	BM		Notropis bifrenatus	MG
Spottail shiner	SS		Notropis hudsonius	MG
Bluntnose minnow	BNM		Pimephales notatus	MG
Blacknose dace	BND		Rhinichthys atratulus	FS
Longnose dace	LND		Rhinichthys cataractae	FS
Creek chub	CRC		Semotilus atromaculatus	FS
Fallfish	F		Semotilus corporalis	FS
Catostomidae (suckers)				
Longnose sucker	LNS		Catostomus catostomus	FD
White sucker	WS		Catostomus commersoni	FD
Creek chubsucker	CCS		Erimyzon oblongus	FS
Ictaluridae (bullhead catfishes)				
White catfish	WC		Ameiurus catus	MG
Yellow bullhead	YB		Ameiurus natalis	MG
Brown bullhead	BB		Ameiurus nebulosus	MG
Tadpole madtom	TMT		Noturus gyrinus	FS
Esocidae (pike and pickerels)				
Redfin pickerel	RP		Esox americanus americanus	MG
Northern pike	NP		Esox lucius	MG
Chain pickerel	CP		Esox niger	MG
Umbridae (mudminnows)				
Central mudminnow	CM		Umbra limi	--
Salmonidae (salmon, chars, and trout)				
Rainbow trout	RT		Oncorhynchus mykiss	FS
Landlocked salmon	LLS		Salmo salar	FD
Brown trout	BT		Salmo trutta	FS
Brook trout	EBT		Salvelinus fontinalis	FS
Fundulidae (killifishes)				
Banded killifish	K		Fundulus diaphanus	MG

Table 1. Common and scientific names of fishes, species codes, and habitat-use classifications for each species.—Continued

[Species code: Massachusetts Division of Fisheries and Wildlife species abbreviation. HUC, habitat-use classification (Kashiwaga and Richards, 2009): FD, fluvial; FS, fluvial specialist; MG, macrohabitat generalist; --, species not categorized]

Common name	Species code	Family name	Scientific name (genus and species)	HUC
Cottidae (sculpins)				
Slimy sculpin	SC		Cottus cognatus	FS
Moronidae (striped basses)				
White perch	WP		Morone americana	MG
Centrarchidae (sunfishes and black basses)				
Rock bass	RB		Ambloplites rupestris	MG
Banded sunfish	BS		Enneacanthus obesus	MG
Redbreast sunfish	RBS		Lepomis auritus	MG
Green sunfish	GSF		Lepomis cyanellus	MG
Pumpkinseed	P		Lepomis gibbosus	MG
Bluegill	B		Lepomis macrochirus	MG
Smallmouth bass	SMB		Micropterus dolomieu	MG
Largemouth bass	LMB		Micropterus salmoides	MG
Black crappie	BC		Pomoxis nigromaculatus	MG
Percidae (perches and darters)				
Swamp darter	SD		Etheostoma fusiforme	MG
Tesselated darter	TD		Etheostoma olmstedii	FS
Yellow perch	YP		Perca flavescens	MG

Table 2. Names and variable codes for environmental factors tested for use as explanatory variables.

[Variables in **bold** type were selected for testing in the regression analysis; MA, Massachusetts]

Variable	Code	Units
x Location of sample site, MA State Plane value in meters, divided by 10,000	OutletX	Meters, divided by 10,000
y Location of sample site, MA State Plane value in meters, divided by 10,000	OutletY	Meters, divided by 10,000
Contributing drainage area upstream from sampling site	DA	Square miles
Elevation	ELEV	Meters
Percent sand and gravel in contributing area	SG	Percent
Channel slope	CHSLP	Number
Basin slope	BASSLP	Percent rise
Precipitation	PRECIPin	Inches
Maximum temperature	TEMPMAX	Degrees Celsius
Strahler stream order	Strahler	Strahler stream order
Shreve stream order	Shreve	Shreve stream order
d-link (Shreve stream order of the next downstream reach)	Shreve_ds	Shreve stream order
Difference between Shreve stream order at site and in downstream reach	Shrevediff	Shreve stream order

Groundwater discharge has a moderating affect on stream temperatures and provides critical thermal habitat for species, such as trout, that prefer cooler temperatures in summer and ice-free habitats in winter (Power and others, 1999; Chu and Jones, 2010). Streams that have a high percentage of sand and gravel areas generally also have high base flows, which maintain streamflow during prolonged dry periods between storms (Ries and Friesz, 2000). Large areas of wetland and open water along stream systems can moderate the variability of streamflow, increase temperatures in streams, and provide increased habitat for spawning and recruitment of macrohabitat generalist species into stream systems.

Data on local habitats (stream width, substrate, stream velocity, water depth, and stream temperature) and data on mesohabitats (number, quality, and sequencing of riffles, runs, and pools) were not available for analysis. However, local channel slope, which is strongly related to velocity and substrate composition (Leopold and others, 1995), was estimated using a GIS for each fish-sampling site. To calculate channel slope, elevations were first determined along the stream centerlines using the LongestFlowPath tool in Arc Hydro (ESRI), then extracted by intersecting a 300-meter-diameter buffered circle, centered at the fish-sampling site, with the stream centerline. Channel slope was then calculated for the reach upstream from the fish-sampling site by first determining the difference between the elevation of the point on the centerline at the upstream intersection of the 300-meter circle and the elevation of the sampling site, then dividing by the length of the centerline between these points.

Environmental factors were determined for the contributing areas of each fish-sampling site using data from a variety of GIS sources and clearinghouses (table 2 in Armstrong and others, 2008). Coordinates were determined from a NAD 83 State Plane Massachusetts Mainland, Federal Information Processing standards (FIPS) zone 2001 GIS projection. The location of each fish-sampling site was calculated as the x and y State Plane coordinates (in meters) divided by 10,000. The coordinates increase in value from west to east and south to north. Contributing drainage areas to the fish-sampling sites were delineated from a 10-meter digital elevation model (DEM) in Environmental Systems Research Institute (ESRI) ArcGIS using an automated batch procedure from Arc Hydro Tools (Maidment, 2002). Data on stream networks were obtained from the high-resolution National Hydrography Dataset (NHD) (U.S. Geological Survey, 1999). The contributing drainage areas were used with a GIS to compute basin characteristics and land-use data for each of the sampling sites.

Land-cover and land-use variables, such as percent forest or percent developed area were determined from the National Land-Cover Database (NLCD) (U.S. Geological Survey, 2000) (table 3). Variables based on the amount of land use or land cover in a basin were standardized to percentages by dividing the land-use or land-cover area by the contributing area to the sampling site.

The area of impervious cover is often used as an indicator of urbanization. Impervious surfaces are defined as constructed surfaces, such as buildings, roofs, roads, parking lots, driveways, brick, asphalt, and concrete, and areas of compacted soil or material such as unpaved parking lots that have no vegetation. Pervious surfaces are defined as vegetated areas, natural and altered, water bodies and wetlands, and naturally occurring barren areas such as rocky shores, sand, and bare soil (Massachusetts Office of Geographic and Environmental Information, 2007). The extent of impervious cover in the contributing areas was calculated from 1-meter resolution impervious-cover (IC) data, determined from 2005 infrared orthoimagery (Massachusetts Office of Geographic and Environmental Information, 2007). The impervious-cover data for parts of contributing drainage basins outside of Massachusetts were determined using methods described in Weiskel and others (2010).

Land use in a narrow corridor extending upstream throughout an entire stream network has been used as a predictor of stream ecosystem condition (VanSickle and Johnson, 2008). Consequently, percent land use and the percent area of impervious cover within a buffer strip adjacent to the stream were tested for use as variables in the study (table 3). A 240-meter buffer area width (120 meters outward in both directions from the stream centerline) was used to be consistent with values published in Coles and others (2004).

A suite of measurements representing different features of dams and impoundments was determined for testing as explanatory variables. Dam locations were obtained from a database maintained by the MDFG Division of Ecological Restoration Program. The presence of dams and impoundments in a basin was indicated by dam density per square mile, dam density per stream mile, and percent open water. Open-water area can include areas of impoundments, natural lakes, and natural lakes augmented by dams. The distance of a sampling site from a dam or dams was represented by two sets of variables—one that measured the length of undammed stream reach along the stream centerline upstream and downstream from each sampling site and a second that measured the total length of undammed stream reaches (including tributaries) that were potentially accessible upstream and downstream from each sampling site.

Simulated Streamflow Data and Measures of Flow Alteration

Daily streamflows representing altered and unaltered flows were simulated for each fish-sampling site by use of a modified (unpublished) version of the SYE version 1.0 (Archfield and others, 2010) together with water-use information and approaches developed during the Massachusetts Water Indicator (MWI) study (Weiskel and others, 2010). To estimate streamflows at an ungaged site, the SYE first estimates selected streamflow quantiles for an

unregulated, daily flow-duration curve by solving regression equations that are a function of physical basin characteristics and climate. A continuous daily flow-duration curve was determined by interpolating streamflow between the estimated quantiles. A time series of unregulated daily streamflows during water years 1961 to 2004 was then created by transferring the timing of the daily streamflow at a selected, minimally altered, index gage site to the un-gaged site by equating exceedance probabilities of contemporaneous flow at the two locations (Archfield and others, 2010). Altered daily streamflows were then estimated using a spatially referenced database of surface-water and groundwater withdrawals and discharges during 2000–04 and techniques for estimating withdrawals and return flows associated with private wells and septic systems described in Weiskel and others (2010). Median monthly groundwater withdrawals were calculated from daily withdrawals determined by disaggregating average annual water-use into daily withdrawals on the basis of an average demand curve developed using data from 25 communities (Weiskel and others, 2010). The SYE application does not provide flow estimates for basins with drainage areas less than 2 mi², the main stems of the Merrimack and Connecticut Rivers and parts of southeast Massachusetts, including Cape Cod and the Islands. Consequently, fish samples from streams in these areas were not included in the analysis.

The SYE produces screening-level estimates of daily streamflows (altered flows) by applying monthly mean water-use estimates from a 5-year period (2000–2004) to 44 years of simulated unaltered streamflow conditions (WY 1961–2004) (Weiskel and others, 2010). Comprehensive water-use data for the 1961–2004 period were not available. Consequently, although annual and monthly flow statistics and the range of variability of daily streamflows are well represented, simulated daily streamflow values likely differ from those actually experienced by fish on a daily basis or at the time of sampling.

The sources for the withdrawal and discharge data used to simulate altered flows for this study are the same as those reported previously in the MWI report (Weiskel and others, 2010) and have the same limitations. Withdrawal and discharge data determined from the SYE water-use database include surface-water and groundwater withdrawals for municipal and non-municipal water supply, wastewater returns from sewage-treatment plants, and return flows from point sources permitted by the MDEP and the National Pollutant Discharge Elimination System (NPDES). Flow alterations from groundwater withdrawals from private wells and returns from septic systems were estimated by use of methods described in Weiskel and others (2010). As for the MWI report, the effects of surface-water withdrawals on streamflows downstream from reservoirs are accounted for only in annual flow-alteration statistics and are not accounted for in daily or monthly flow-alteration statistics (Weiskel and others, 2010). The altered streamflows simulated using SYE also do not explicitly account for non-consumptive flow alterations caused by industry, hydroelectric facilities, manipulations from

actively managed private lakes and ponds, or flow alterations caused by altered land use or impervious cover. Because of these limitations, flow alterations in this report are referred to as “flow alteration indicators “ or “estimated flow alterations.”

The streamflow alteration indicators used in this study were calculated as the percent alteration of selected flow statistics representing unaltered streamflows (equation 1).

$$\text{Percent alteration} = \left(\frac{A}{U} - 1 \right) \times 100, \quad (1)$$

where

- A = the altered median monthly or annual mean streamflow, and
 U = unaltered median monthly or annual mean streamflow.

Prior to analysis, the unaltered flows simulated using SYE and flow alterations estimated for the MWI study were used to determine a set of flow alteration indicators. Estimated indicators of flow alteration were determined in two different ways for evaluation during the study. For the first set of flow-alteration indicators, water uses that deplete streamflows and water uses that surcharge streamflows were combined in order to calculate flow-alteration statistics representing net streamflow conditions (table 4A). For the second set of estimated indicators of flow-alteration, water withdrawals and returns were treated separately in order to represent flow alterations that could be attributed to separate water-use components (table 4B). For example, flow alteration indicators were calculated for percent alteration of August median flow from total (municipal + non-municipal + private) groundwater withdrawals, percent alteration of August median flow from surface-water returns, and percent alteration of August median flow from septic returns.

The primary streamflow statistic used for calculating percent flow alteration indicators was the median of monthly median flow (that is, a median flow was determined for each month, then a median was determined from the 44 median monthly flows from 1961 to 2004). Median monthly flows were used because of the inherent uncertainty of simulated SYE streamflow statistics at the daily time scale (Weiskel and others, 2010; Levin and others, 2011). Additional flow-alteration indicators that were evaluated include the ratio of annual return flows to annual unaltered flows (return-flow fraction), the ratio of withdrawals to unaltered flows (withdrawal fraction), mean annual and August median flows normalized to cubic feet per second per square mile, and water-use intensity (WUI). The WUI indicator is a ratio defined as the sum of the absolute value of withdrawals and return flows relative to the long-term average unaltered streamflow from a basin (Weiskel and others, 2007; 2010). The WUI indicator is used to identify basins designated as “churned” basins, where water withdrawals and return flows are similar in magnitude to each other, yet large relative to unaltered streamflows from the basin.

Table 3. Names and variable codes for land-cover and land-use variables tested for use as explanatory variables.

[Variables in **bold** type were selected for testing in the regression analysis; NLCD, National Land Cover Dataset; US, upstream; ACOE, Army Corps of Engineers]

Variable	Code	Unit
Percent impervious cover in contributing area	IC	Percent
Percent impervious cover in a 240-meter buffer, 120 m each side of stream centerline	BFIC	Percent
Population density	PopDsqmi	People per square mile
Total length of roads in contributing area	RoadLkm	Kilometers
ACOE surface area of impoundments	STORarea	Area
Count of dams in upstream contributing area	DamCt	Number of dams in contributing area
Dam density, dam count divided by drainage area	DamDsqmi	Dams per square mile
Dam density per stream mile, dam count divided by US stream length	DamDstm	Dams per stream mile
Length of undammed stream channel upstream of fish sampling site	UdamUSmi	Miles
Length of undammed stream channel downstream of fish sampling site	UdamDSmi	Miles
Total length of undammed stream channel	UdamTmi	Miles
Undammed reach length along longest flow path upstream	DdamUSmi	Miles
Upstream dam	USdamNY	Where 0 = no, 1 = yes
Undammed reach length along longest flow path downstream	DdamDSmi	Miles
Downstream dam	DSdamNYNA	Where = no, 1 = yes, 8 = out of area
Percent area of buffer classified as developed - open space, NLCD class 21	pBDev_o	Percent
Percent area of buffer classified as developed - low intensity, NLCD class 22	pBDev_l	Percent
Percent area of buffer classified as developed - medium intensity, NLCD class 23	pBDev_m	Percent
Percent area of buffer classified as developed - high intensity, NLCD class 24	pBDev_h	Percent
Percent area of buffer classified as developed, NLCD class 21 + 22 + 23 + 24	pBDev_al	Percent
Percent area of buffer classified as barren land (rock/sand/clay), NLCD class 31	pBBar	Percent
Percent area of buffer classified as deciduous forest, NLCD class 41	pBFor_d	Percent
Percent area of buffer classified as evergreen forest, NLCD class 42	pBFor_e	Percent
Percent area of buffer classified as mixed forest, NLCD class 43	pBFor_m	Percent
Percent area of buffer classified as shrub/scrub, NLCD class 52	pBShrub	Percent
Percent area of buffer classified as grassland/herbaceous, NLCD class 71	pBGras	Percent
Percent area of buffer classified as agriculture, pasture/hay, NLCD class 81	pBAg_p	Percent

Table 3. Names and variable codes for land-cover and land-use variables tested for use as explanatory variables.—Continued[Variables in **bold** type were selected for testing in the regression analysis; NLCD, National Land Cover Dataset; US, upstream; ACOE, Army Corps of Engineers]

Variable	Code	Unit
Percent area of buffer classified as agriculture, cultivated crops, NLCD class 82	pBAg_c	Percent
Percent area of buffer classified as open water, NLCD Class 11	pBOW	Percent
Percent area of buffer classified as woody wetlands, NLCD class 90	pBWet_w	Percent
Percent area of buffer classified as emergent herbaceous wetlands, NLCD class 91	pBWet_e	Percent
Percent area of buffer classified as wetland, NLCD class 90 + 91	pBWet_al	Percent
Percent area of buffer classified as agriculture, NLCD class 81 + 82	pBAg_al	Percent
Percent contributing area classified as open water, NLCD Class 11	pCOW	Percent
Percent contributing area classified as developed - open space, NLCD class 21	pCDev_o	Percent
Percent contributing area classified as developed - low intensity, NLCD class 22	pCDev_l	Percent
Percent contributing area classified as developed - medium intensity, NLCD class 23	pCDev_m	Percent
Percent contributing area classified as developed - high intensity, NLCD class 24	pCDev_h	Percent
Percent contributing area classified as barren land (rock/sand/clay), NLCD class 31	pCBar	Percent
Percent contributing area classified as deciduous forest, NLCD class 41	pCFor_d	Percent
Percent contributing area classified as evergreen forest, NLCD class 42	pCFor_e	Percent
Percent contributing area classified as mixed forest, NLCD class 43	pCFor_m	Percent
Percent contributing area classified as shrub/scrub, NLCD class 52	pCShrub	Percent
Percent contributing area classified as grassland/herbaceous, NLCD class 71	pCGras	Percent
Percent contributing area classified as agriculture, pasture/hay, NLCD class 81	pCAG_p	Percent
Percent contributing area classified as agriculture, cultivated crops, NLCD class 82	pCAG_c	Percent
Percent contributing area classified as woody wetlands, NLCD class 90	pCWet_w	Percent
Percent contributing area classified as emergent herbaceous wetlands, NLCD class 91	pCWet_e	Percent
Percent contributing area calculated as the total of low-, medium-, and high-intensity developed + open space, NLCD classes 21+22+23+24	pCDev_al	Percent
Percent contributing area calculated as the total of deciduous + evergreen + mixed forest, NLCD classes 41 + 42 + 43	pFOR_al	Percent
Percent contributing area calculated as the total of shrub, grassland, barren, NLCD 52 + 71 + 31	pShGr	Percent
Percent contributing area calculated as the total of pasture and cultivated agriculture, NLCD classes 81 + 82	pCAG_al	Percent
Percent contributing area calculated as total of wetland classes, NLCD classes 90 + 91	pWET_al	Percent

Table 4. Names and variable codes for flow-alteration variables tested for use as explanatory variables: *A*, net-flow alteration variables; and *B*, flow-alteration components.[Variables in **bold** type were selected for testing in the regression analysis. All variables are in percent]

A. Net-flow alteration variables	
Variable	Code
Net percent alteration of January median flow, surcharged sites	JanNETRp
Net percent alteration of February median flow, surcharged sites	FebNETRp
Net percent alteration of March median flow, surcharged site	MarNETRp
Net percent alteration of April median flow, surcharged sites	AprNETRp
Net percent alteration of May median flow, surcharged sites	MayNETRp
Net percent alteration of June median flow, surcharged sites	JunNETRp
Net percent alteration of July median flow, surcharged sites	JulNETRp
Net percent alteration of August median flow, surcharged sites	AugNETRp
Net percent alteration of September median flow, surcharged sites	SepNETRp
Net percent alteration of October median flow, surcharged sites	OctNETRp
Net percent alteration of November median flow, surcharged sites	NovNETRp
Net percent alteration of December median flow, surcharged sites	DecNETRp
Net percent alteration of January median flow, depleted sites	JanNETWp
Net percent alteration of February median flow, depleted sites	FebNETWp
Net percent alteration of March median flow, depleted sites	MarNETWp
Net percent alteration of April median flow, depleted sites	AprNETWp
Net percent alteration of May median flow, depleted sites	MayNETWp
Net percent alteration of June median flow, depleted sites	JunNETWp
Net percent alteration of July median flow, depleted sites	JulNETWp
Net percent alteration of August median flow, depleted sites	AugNETWp
Net percent alteration of September median flow, depleted sites	SepNETWp
Net percent alteration of October median flow, depleted sites	OctNETWp
Net percent alteration of November median flow, depleted sites	NovNETWp
Net percent alteration of December median flow, depleted sites	DecNETWp
Net percent alteration of mean annual flow, depleted sites	AnnNetWp
Net percent alteration of mean annual flow, surcharged sites	AnnNetRp
Water-use intensity	WUI
Net percent alteration of 7-day minimum flow, depleted sites	7dMinNWp
Net percent alteration of 7-day minimum flow, surcharged sites	7dayMinNRp
Net percent alteration of 7-day maximum flow	7dayMaxPA
Net percent alteration of low-pulse count	LPctPA
Net percent alteration of low-pulse duration	LPdurPA

Table 4. Names and variable codes for flow-alteration variables tested for use as explanatory variables: *A*, net-flow alteration variables; and *B*, flow-alteration components.—Continued[Variables in **bold** type were selected for testing in the regression analysis. All variables are in percent]

B. Flow-alteration components	
Variable	Code
Percent alteration of January median flow from surface-water returns	JANswRp
Percent alteration of February median flow from surface-water returns	FEBswRp
Percent alteration of March median flow from surface-water returns	MARswRp
Percent alteration of April median flow from surface-water returns	APRswRp
Percent alteration of May median flow from surface-water returns	MAYswRp
Percent alteration of June median flow from surface-water returns	JUNswRp
Percent alteration of July median flow from surface-water returns	JULswRp
Percent alteration of August median flow from surface-water returns	AUGswRp
Percent alteration of September median flow from surface-water returns	SEPswRp
Percent alteration of October median flow from surface-water returns	OCTswRp
Percent alteration of November median flow from surface-water returns	NOVswRp
Percent alteration of December median flow from surface-water returns	DECswRp
Percent alteration of January median flow from septic returns	JANsepRp
Percent alteration of February median flow from septic returns	FEBsepRp
Percent alteration of March median flow from septic returns	MARsepRp
Percent alteration of April median flow from septic returns	APRsepRp
Percent alteration of May median flow from septic returns	MAYsepRp
Percent alteration of June median flow from septic returns	JUNsepRp
Percent alteration of July median flow from septic returns	JULsepRp
Percent alteration of August median flow from septic returns	AUGsepRp
Percent alteration of September median flow from septic returns	SEPsepRp
Percent alteration of October median flow from septic returns	OCTsepRp
Percent alteration of November median flow from septic returns	NOVsepRp
Percent alteration of December median flow from septic returns	DECsepRp
Percent alteration of January median flow from groundwater withdrawals	JANgwWp
Percent alteration of February median flow from groundwater withdrawals	FEBgwWp
Percent alteration of March median flow from groundwater withdrawals	MARgwWp
Percent alteration of April median flow from groundwater withdrawals	APRgwWp
Percent alteration of May median flow from groundwater withdrawals	MAYgwWp
Percent alteration of June median flow from groundwater withdrawals	JUNgwWp
Percent alteration of July median flow from groundwater withdrawals	JULgwWp
Percent alteration of August median flow from groundwater withdrawals	AUGgwWp
Percent alteration of September median flow from groundwater withdrawals	SEPgwWp
Percent alteration of October median flow from groundwater withdrawals	OCTgwWp
Percent alteration of November median flow from groundwater withdrawals	NOVgwWp
Percent alteration of December median flow from groundwater withdrawals	DECgwWp
Percent alteration of mean annual flow from surface-water withdrawals from reservoirs	ANNswW
Percent alteration of mean annual flow from withdrawals	ANNWp
Percent alteration of mean annual flow from returns	ANNRp

Selection of Environmental Factors and Fish Assemblage Metrics

The data set collected for this study is a multivariate data set that has numerous sample sites, each described by a set of species abundances and a set of environmental and anthropogenic factors. Multivariate statistical techniques are the standard methods used to simultaneously analyze the variability of multiple variables. These techniques are powerful tools for reducing data complexity and for summarizing trends in ecological communities (Gauch, 1994; Kenkel and others, 2002; Peck, 2010). Several different multivariate techniques were used in this study. Principal components analysis (PCA) supplemented with Spearman rank correlation was used to reduce the number of environmental variables to a subset of variables that accounted for the greatest proportion of variance while minimizing redundancy in the data set. Nonmetric multidimensional scaling (NMDS) and hierarchical cluster analysis were used to explore patterns among fish species and to guide selection of fish-assemblage metrics for use in subsequent analysis of relations between fish and environmental variables.

Environmental and ecological data generally require some form of transformation or standardization prior to use in a multivariate analysis to improve normality and homogeneity of a data set. Environmental data often have different scales of measurement and require standardization to a common scale so that ordinations are not dominated by variables that have large units of measurement. This standardization (alternatively called normalization, relativization by the standard deviate, or transformation to a z-score) is done by subtracting the sample mean from each variable and dividing by the sample standard deviation. Fish species data were square-root transformed as a method of reducing the dominance of highly abundant species in relation to less abundant species (Anderson and others, 2008; Clark and Warwick, 2001). The square root transformation is commonly used for count data (Leps and Smilauer, 2003).

Environmental Factor Selection using Principal Component Analysis and Spearman Rank Correlation

Multicollinearity is the result of strong correlation between explanatory variables, and is a major concern for multivariate regression models. Multicollinearity may result in incorrect signs and magnitudes of regression coefficients, thereby leading to incorrect conclusions about relations between explanatory and response variables. Difficulties may also arise because of spurious relations that result from the confounding effect of spatial autocorrelation or correlation with other variables that were not measured (Mandrak, 1995; Wilkinson and Edds, 2001; Legendre 2002; Graham, 2003). Variables that were highly correlated were identified and were not used together in multivariate regression equations.

PCA and Spearman rank correlation were used to isolate a subset of variables that accounted for the greatest proportion of variance while minimizing redundancy in the data set. Redundant variables were removed; during interpretation, it is common practice to check the variable that was used for modeling to make sure that it was not acting as a proxy for a variable that was removed (Anderson and others, 2008). PCA is an exploratory multivariate ordination technique that is used to reduce the dimensionality of a multivariate data set while accounting for as much of the original variation as possible (Everitt and Hothorn, 2010). Though PCA is generally used to create a smaller number of new variables (called principal components) that can be used as surrogates for the originally large number of variables, it can also be used, as in this study, as a means of reducing the number of variables in the data set (Olden and Poff, 2003; Gillette and others, 2005; Kennen and others, 2008). The PCA analysis was conducted using the correlation matrix to ensure that all the environmental variables contributed equally to the PCA and that the contributions were scale-independent (Legendre and Legendre, 1998). Environmental variables that extracted the dominant patterns of variation were identified by use of the PCA loadings on each significant principal component (PC) axis. The term “loading” refers to the correlation of variables with a PC axis, which provides a measure of how well each variable would substitute for the linear combination of variables that compose the PC axis (James and McCulloch, 1990). Spearman rank correlation was used to evaluate the strength and direction of correlation among environmental variables and to select variables that were not strongly correlated with each other. Variables that are highly correlated contain effectively the same information and are redundant for purposes of analysis (Anderson and others, 2008). Spearman rank correlation was used instead of Pearson correlation because it is robust to outliers and makes no assumptions about linearity in the relation between two variables (Zuur and others, 2009).

The steps in the variable selection process involved identification and retention of the highest loaded variable (positive or negative) from the first PC axis, then elimination of any correlated variables, as determined from a Spearman rank correlation ($\rho > 0.70$). The process was repeated for the next remaining highest-loaded variable and for all variables on the PC axis with loadings greater than 0.32 or less than -0.32 (McGarigal and others, 2000; Tabachnick and Fidell, 2001). The 0.32 loading cutoff value is used as a rule-of-thumb to identify variables that account for more than 10 percent of the variability in a component (the square of a loading value indicates the percentage of the variable’s variance accounted for by the principal component) (McGarigal and others, 2000; Tabachnick and Fidell, 2001). This process was repeated for each significant PC axis with PC axis significance determined by use of the stopping rules of the broken stick eigenvalue model as described in Jackson (1993) and McGarigal and others (2000). Variables that did not have significant loadings on any PC axis were eliminated.

Fish Assemblage Classification using Ordination and Hierarchical Cluster Analysis

Ordination and cluster analysis were used to test the grouping of fish species into habitat-use classifications (fluvial, macrohabitat-generalist). Ordination summarizes the dominant assemblage patterns in multivariate data sets by displaying the patterns in only a few dimensions. Ordination places samples (or variables) on a plot in such an order that samples that lie near to one another will be more similar than samples that are far apart (Shaw, 2003). A cluster analysis is used to assign similar samples (or variables) to groups. The ordination and cluster methods used in this analysis (hierarchical agglomerative clustering and nonmetric multidimensional scaling (NMDS)) were performed with PRIMER v6 (Clarke and Warwick, 2001; Clarke and Gorley, 2006).

Ordination and clustering techniques are based on a comparison of all possible pairs of samples or variables. These pairwise comparisons form a symmetric matrix called a resemblance matrix that is normally presented in a triangular data format similar to that commonly shown for a correlation table. The resemblance coefficients in the matrix are measures of the dissimilarity between samples (or variables). Dissimilarity coefficients (also referred to as distances) are algebraic measures of how close the abundance levels or values for environmental variables are for each sample, where 0 percent represents total similarity and 100 percent represents complete dissimilarity. There are a wide number of dissimilarity measures, each with different assumptions (Warton and others, 2011). The dissimilarity coefficients used in the NMDS ordination and hierarchical cluster analyses in this study follow the recommendations given by Clark and Warwick (2001) and use the Bray-Curtis dissimilarity coefficient to define distance between fish species and Euclidean distance to define distance between environmental variables. Bray-Curtis dissimilarity coefficients measure the degree of dissimilarity in assemblage structure between pairs of samples. Bray-Curtis dissimilarity coefficients range from zero to one, where zero indicates an exact match in assemblage composition and one indicates two samples are completely different with no species in common (Clarke and others, 2006). The Bray-Curtis dissimilarity coefficient is widely used for ecological data because its value is unaffected by the joint absence of species in samples.

Nonmetric multidimensional scaling (NMDS) is a robust nonparametric ordination technique that is appropriate for use with ecological data sets (Clarke, 1993; Clarke and Warwick, 2001; McCune and Grace, 2002). NMDS is considered to be robust because it can handle nonlinear species responses and makes no assumptions about the distribution of variables; it is considered to be nonparametric because it is a rank-based approach. The purpose of an NMDS ordination is to find a configuration of samples (or variables) where the interpoint distances (between samples or variables on the plot) reflect as closely as possible the rank order of the interpoint distances from the original data set (Peck, 2010). The number of axes

for the NMDS ordination (generally 2 or 3) is specified prior to the analysis. NMDS starts from a dissimilarity matrix among variables and generates an ordination solution by using an algorithm to iteratively evaluate the correspondence between ordination distances and rank-order distances. A stress statistic measures the lack of fit between distances in ordination space and dissimilarities; it is also used to indicate how well the NMDS ordination summarizes the observed distances among the samples.

Although the non-parametric nature of NMDS makes it an appropriate ordination technique for ecological data, it has some limitations. NMDS differs from other ordinations, such as PCA, in that it is not an eigenvalue-eigenvector technique, which ordines data such that the first axis explains the greatest amount of variance and each successive axis explains diminishing amounts of variance (Holland, 2008). Because NMDS is non-parametric, it does not provide a quantitative measure for partitioning the variance explained by explanatory variables (Anderson and others, 2008).

The cluster analysis method used in this analysis is hierarchical agglomerative clustering. The method is hierarchical because groups are formed that contain subgroups, so there is a hierarchy of levels. The method is agglomerative because the clustering begins with joining the two most similar objects (samples, species, or variables) to form the first group, then successively combines similar objects or groups to build up a hierarchy of increasingly large clusters until all objects and groups are joined (McGarigal and others, 2000; McCune and Grace, 2002; Peck, 2010). Cluster analysis for the ecological data (fish species) started from a Bray-Curtis resemblance matrix. Next a clustering algorithm or linkage method for defining the distance between groups was selected. The clustering algorithm used was the average linkage, or unweighted-pair group method. The hierarchical classification is displayed in a dendrogram, a tree-like plot that depicts the sequence and dissimilarity level at which each cluster is formed. On a dendrogram, the length of the branches indicates the degree of dissimilarity between the cluster members; similar objects or groups are connected by short branches, and dissimilar objects or groups are connected by long branches. In ecological studies, selection of the number of groups is done by evaluating the change in interpretability of the cluster compositions. The groups identified with cluster analysis were superimposed on NMDS ordinations to check the consistency of the two approaches.

Analysis of Associations between Fish Assemblages and Environmental Factors

Two analytical techniques, quantile regression and GLMs, were applied in this study to evaluate the association between stream fish and environmental and anthropogenic explanatory variables. Each of these techniques has characteristics that make it appropriate for use with ecological data sets. Ecological data sets developed on the basis of

abundances or counts of individual species at sample sites tend to have non-normal distributions that are highly skewed (Anderson, 2001; Ver Hoef and Boveng, 2007). Because species do not occur equally everywhere, ecological data sets, such as fish species relative abundance, typically have occasional samples where some species are locally abundant and also have a high proportion of samples with zero abundance (McArdle and Anderson, 2001; Kenkel and others, 2002; Sileshi and others, 2009; Anderson and others, 2008). Quantile regression is an univariate technique that is considered to be robust with respect to data sparseness and extreme outlying values, and can deal with zero counts (Scharf and others, 1998; Cade and others, 1999; Creque and others, 2005; Vaz and others, 2008). GLMs are mathematical extensions of linear models that allow for some nonlinearity and non-constant variance structure in the data and that can be used for analysis of zero-inflated data (Guisan and others, 2002; Zuur and others, 2009).

Quantile Regression

Quantile regression is a statistical method for fitting a regression line to a quantile of the response as opposed to the mean of the response, as done in ordinary least-squares regression. A quantile regression line can illustrate the upper limits of fish-response variables in relation to stressor gradients and indicate the direction of that association. Although regression quantiles near the upper boundary of a wedge-shaped relation describe potential rather than actual relations between species relative abundance and distribution (Vaz and others, 2008), viewing variables as constraints rather than as correlates can show that changes near the maximum response better represent effects when the measured factor is the limiting constraint (Cade and others, 2005). The ecological concept of limiting factors indicates that species will be abundant if no other factors are limiting, but if one or more factors are limiting, the species will be constrained to lower abundance than expected (Cade and others, 2005). For example, even if flow alterations are nonexistent or minimal, fish habitat may be poor for some species if other requirements such as water-quality conditions conducive to fish survival are not met. Bivariate scatterplots of species and stressor variables illustrate the concept of limiting factors when data display a right-skewed, wedge-shaped relation distributed below an upper bound. The wedge shape of scatterplots indicates that the relations between the plotted variables exhibit high variability. This high variability arises when other environmental or biological variables function as active limiting factors and keep the response variable from reaching the upper constraint line (Thomson and others, 1996; Cade and Noon, 2003; Schooley and Wiens, 2005; Anderson, 2008). The declining upper bound indicates that the explanatory variable can act as a constraint on the given response variable and illustrates the maximum abundance of a species given ideal environmental conditions.

Quantile regression can be used to estimate functional relations between variables for any quantile of a probability distribution (Cade and Noon, 2003; Koenker, 2005), but it is used in this report only to estimate rates of change for functions along or near the upper boundary of the conditional distribution of responses. For a wedge-shaped scatterplot, quantile regression lines that define upper limits typically have different slopes than the regression line that indicates the mean value. For example, a regression line fit through a scatterplot with a high density of points near the origin may have a low slope, indicating little to no relation, even though the upper edge of the data has a strong negative slope. Quantile regression fits a regression line such that a defined proportion of the data points fall below the line and the remainder of the points fall above the line (Cade and Noon, 2003). For example, for the 0.90 quantile (90th quantile), 90 percent of the observations are below the fitted regression line and 10 percent lie above. The 90th quantile has been used as a robust quantile to describe the upper bounds of wedge-shaped relations (Scharf and others, 1998; Wang and others, 2001a, 2001b); it was selected for use in this study over larger quantiles, such as the 95th quantile, to ensure sufficient points were used to estimate the slope of the regression line (Konrad and others, 2008).

Quantile regression models for this study were fit in the software program R (R Development Core Team, 2008) using the *rq* package of Koenker (2005). The *rq* package fits a linear model. Nonlinear quantile regression functions also can be estimated but were not tested for this report. Note that the plots of the linear quantile regression functions in this report show curved lines rather than straight lines. The word “linear” in a linear statistical model is a mathematical term describing the relation between model parameters and the outcome; it does not necessarily imply a straight line relation (Motulsky, 2010; Kery, 2010). For this study, estimates of linear quantile regression were determined for the logs of the response variable (y). A small constant was added before determining the logarithms because there were some zero counts. These linear quantile regressions were returned to a nonlinear form by back-transformation (exponentiating). A back-transformation was possible because regression quantiles retain their statistical properties under any monotonic transformation of y (Cade and Noon, 2003; Cade and Guo, 2000; Koenker and Machado, 1999). That is, the 90th quantile of the transformed data is equivalent to the 90th quantile of the original data. The resulting back-transformed regression line plots as a curve on a scatterplot of the original data, the shape of which is appropriate for modeling relations with a wedge-shaped form (Wang and others, 2001a).

A bootstrap procedure, by which the data set is resampled with replacement within the range of observed and predicted values, was used to obtain standard errors and confidence intervals for the quantile regression line. The direction of the slope of the quantile regression line indicates the nature of the relation between the variables. On the quantile regression plots, a quantile that decreases from left to right indicates

that an increase in the explanatory variable is associated with a decrease in the fish-response variable, and a quantile that increases from left to right indicates that an increase in the explanatory variable is associated with an increase in the fish variable. Confidence intervals for the quantile regression line indicate that the model of the ecological response is imprecise if confidence intervals are relatively wide and indicate the model is reliable if the confidence intervals are narrow. Confidence intervals that encompass a line with a zero slope (a flat line) indicate that the relation between the two variables was insignificant for this data set.

Generalized Linear Models

GLMs were used to develop regression equations that relate fish-response variables to multiple explanatory variables representing environmental and anthropogenic factors. GLMs are frequently used in ecological applications when the data do not meet the assumptions of ordinary least squares (OLS) regression (Zuur and others, 2007, 2009, 2010; Bolker, 2008; Wenger and Freeman, 2008). Fish data in this study violate several assumptions of OLS regression. In particular, fish data are not normally distributed as a result of an excessive number of zeros in the data. Zeros in fish community data sets occur often because all fish species are not present within every habitat, stream system, or sample. A second violation of regression assumptions arises because fish abundance and richness data often do not exhibit a constant variance across all levels of the explanatory variables. A GLM extends linear models by allowing the response variable to have a non-normal distribution and allows for a more flexible variance structure in the fitted model. A GLM also enables models to be developed for response variables that are bounded (as in proportions), cannot be negative (as in counts), or have highly skewed distributions (Cameron and Trivedi, 1998; Ahmadi-Nedushan and others, 2006; Wood, 2006). These characteristics allow GLMs to be applied to a wider range of data than traditional linear models and to include linear regression models as a special case.

GLMs are somewhat analogous to OLS multiple linear regression models in that they predict or model a single response variable on the left side of an equation from a linear combination of one or more explanatory variables on the right side of the equation (the linear predictor). A GLM specifies a link function that allows the expected value of the response variable to depend on a smooth monotonic function of the linear predictor (McCullagh and Nelder, 1989; Wood, 2006). Selection of the link function is dictated by the nature of the response variable. For example, a log link is used for count data that have a Poisson distribution. Once the model is fitted, the expected values of the response variable (the mean value of Y) are obtained by applying the inverse of the link function (for example, exponentiating if a log link is used). A GLM differs from OLS multiple linear regression by using maximum likelihood techniques to determine GLM equation coefficients rather than minimizing the sums-of-squares

(McCullagh and Nelder, 1989). Consequently, although there are tools available for GLMs that can be used for model selection and summary statistics, such as parameter estimates, standard errors, and goodness-of-fit statistics that can be used to evaluate fitted GLMs, these tools are different from the familiar diagnostic statistics used for this purpose in OLS models, such as the coefficient of determination (R^2).

A GLM with a Poisson distribution can be used to model count data, such as species abundance or richness. By definition, the Poisson distribution consists of positive integers with a variance that equals the mean. When the mean is small, the Poisson distribution has a positively skewed shape with a high number of observations equal to zero, making it an appropriate distribution for modeling ecological data. Models for rates, such as catch per unit effort, are fit with an offset to account for differing sampling times.

For a particular set of covariates, the Poisson model is

$$E(Y) = e^{b_0 + b_1x_1 + b_2x_2 + \dots + b_jx_j}, \quad (2)$$

where

- $E(Y)$ = expected mean value of relative abundance (in counts per hour);
- e = Euler's number, the base of the natural logarithm;
- β_j = Poisson model coefficient; and
- x_j = covariate values.

The GLM used for this study was a Poisson model corrected for overdispersion. Overdispersion occurs when there is greater variability in a data set than would be expected on the basis of the probability distribution assumed by the statistical model. Although a Poisson distribution can accommodate skewed data sets, ecological data tend to be overdispersed because of a high number of zero observations and very few high values (Martin, 2005). Overdispersion can also be caused by positive correlations between responses and clustering of data (Zuur and others, 2007; Hilbe, 2011). Outliers, omission of important explanatory predictors or interaction terms, nonlinear relations, or failure to transform predictor variables when required can also create apparent overdispersion. Overdispersion causes the standard errors of model coefficient estimates to be underestimated. This can make a variable appear to be a significant predictor when it is not, leading to poor model selection (Lewin and others, 2010). Overdispersed data can be fit with a quasi-Poisson model which corrects the standard errors on the basis of a dispersion parameter estimated from the model residuals.

The GLM statistical analyses were conducted using the R package for Statistical Computing, ver. 2.8. (R Development Core Team, 2008). The quasi-Poisson models were fit by use of the Stats package in the standard installation of the R statistical software program. Stepwise model selection was performed from a pool of potential explanatory variables, including physical-basin and land-use characteristics, percent impervious cover, and flow-alteration indicators. Selection

of potential fitted models was done by comparison of the residual deviance of the fitted model and the nested model. Pseudo- R^2 values were calculated after Dobson (2002) and Zuur and others (2009) as a measure of the proportion of the variance explained by the fitted model compared to the model with only an intercept. This measure is calculated using deviance residuals, and the interpretation is similar to that of the R^2 used in OLS regression. To further validate the models, predicted values were compared to observed values for each fitted GLM using a remove-one jackknife procedure (Zuur and others, 2009). A simple linear regression of the observed and predicted values provides information about the bias of the fitted GLM (Potts and Elith, 2006; Sileshi and others, 2009). For a simple linear model of the form:

$$Y = b_0 + b_1 \times X, \quad (3)$$

where

$$\begin{aligned} Y &= \text{observed values,} \\ \beta_1 &= \text{linear regression coefficients, and} \\ \hat{X} &= \text{fitted values.} \end{aligned}$$

An intercept of 0 and a slope of 1 for the simple linear regression indicates an unbiased GLM fit. An intercept significantly different from zero indicates a constant or consistent prediction bias, and a slope significantly different from one indicates an inconsistent bias across the range of predictions. Taken together, the pseudo- R^2 and the correlation between observed and predicted observations give a complete picture of the goodness-of-fit and predictive ability of a GLM. Confidence intervals for model predictions were developed on the basis of the standard error of the response.

Factors Affecting Riverine Fish Assemblages

Environmental factors, streamflow alteration indicators, and results of the variable reduction process conducted for the environmental variables using PCA and Spearman rank correlation are described in the following sections. Fish-assemblage compositions for the 669 fish-sampling sites used in this study are presented, along with an assessment of

fish-assemblage structure using hierarchical cluster analysis and NMDS ordination. The association between fish-response variables and the selected environmental and anthropogenic explanatory variables is then evaluated using quantile regression and generalized linear modeling.

Environmental Factors

The 669 fish-sampling sites represent a wide range of environmental and anthropogenic factors. Boxplots illustrate the range for selected environmental factors used in the study (fig. 2). The mean and median contributing drainage areas of the sampling sites were 26 and 7.4 mi², respectively, and drainage areas ranged from 2 to 395 mi². About 57 percent of the fish-sampling sites have contributing drainage areas of less than 10 mi². About 30 percent of sites have drainage areas of 10 to 50 mi², and less than 12 percent of sites have drainage areas greater than 50 mi².

Boxplots illustrate the range of values for variables used in the study to represent dams (fig. 3). The number of dams upstream from the sampling sites ranged from 0 to 182 with a median of 2 upstream dams. Dam density values ranged from 0 to 2.6 with a median of 0.26 dams per square mile, and the length of undammed stream channel along the main stem upstream from the sampling sites ranged from 0 to 33 miles with a median of 3 miles.

The magnitude and distribution of the percent impervious cover represented in the contributing areas to the fish-sampling sites used in this study were compared to those reported in Weiskel and others (2010) for 1,429 subbasins and groundwater contributing areas in Massachusetts (fig. 4). The comparison indicated a similar range of impervious cover; however, the fish-sampling sites were weighted slightly toward less-impervious conditions. About 55 percent of the sites are in basins with less than 5 percent impervious cover, 79 percent are in basins with less than 10 percent impervious cover, and 90 percent are in basins with less than 15 percent impervious cover. Only 4.8 percent of the sites are in basins with greater than 20 percent impervious cover. A map of impervious cover for the fish-sampling sites is shown in figure 5. Impervious-cover values ranged from about 0.3 to 38.1 percent with mean and median values of about 6.5 and 4.2 percent, respectively.

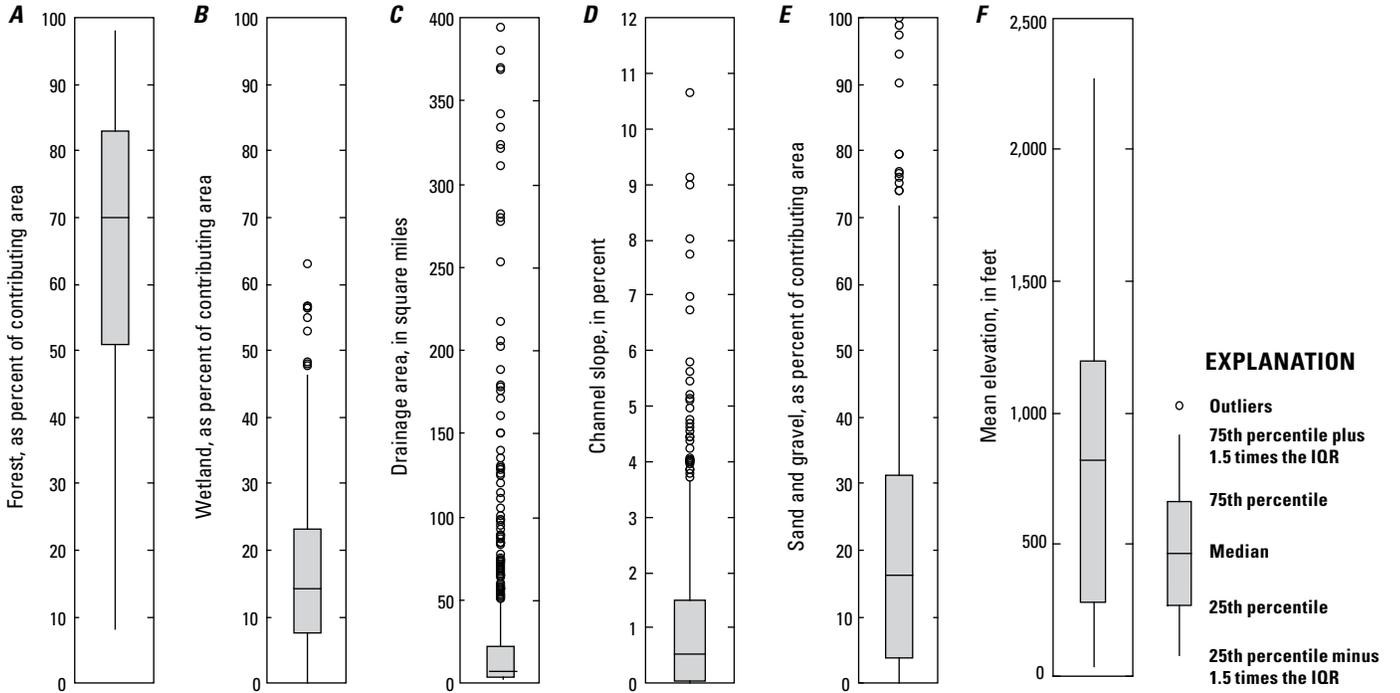


Figure 2. A, Percent forest, B, percent wetland area, C, drainage area, D, channel slope, E, percent sand and gravel, and F, mean elevation for contributing areas to 669 fish-sampling sites on Massachusetts streams. IQR, interquartile range, is the difference between the 75th- and 25th-percentile values.

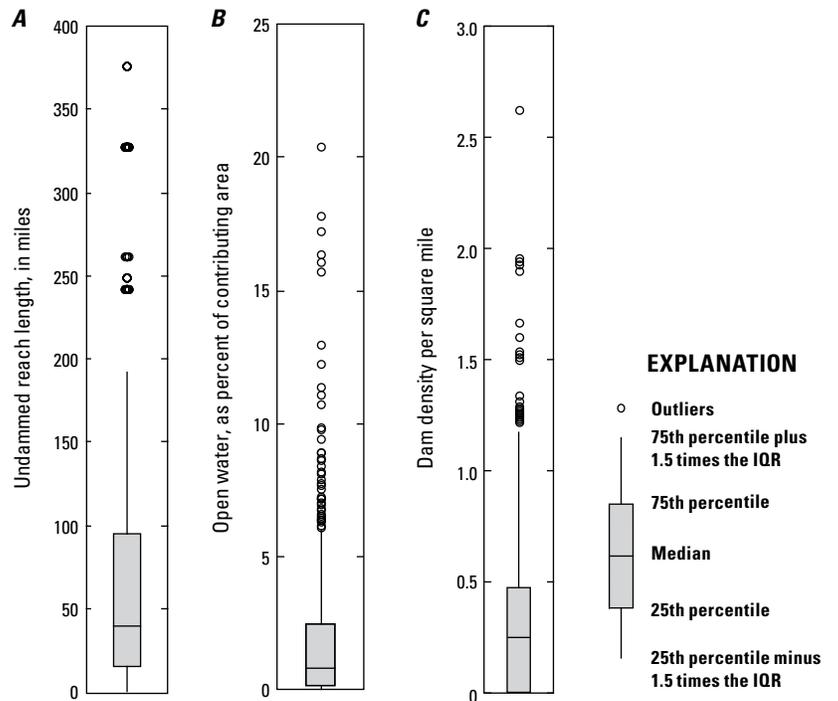


Figure 3. A, Undammed reach length, B, percent open water, and C, dam density per square mile for 669 fish-sampling sites on Massachusetts streams. IQR, interquartile range, is the difference between the 75th- and 25th-percentile values.

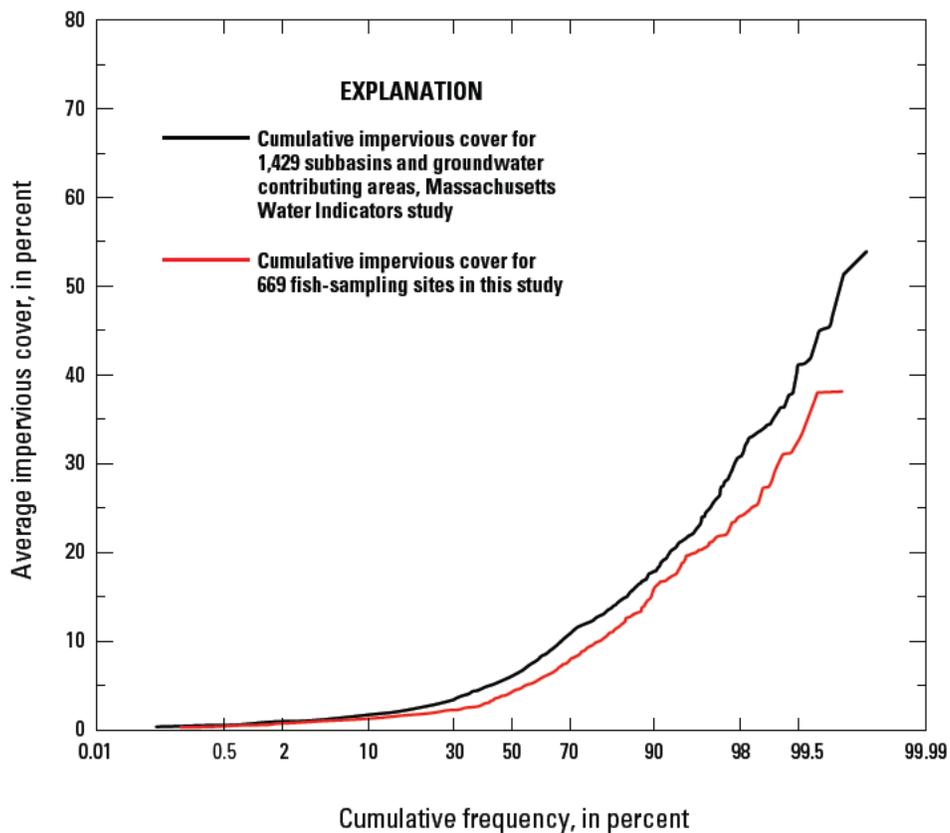


Figure 4. Cumulative frequency distributions of impervious cover for contributing areas to 669 fish-sampling sites and 1,429 subbasins and groundwater contributing areas across Massachusetts. Massachusetts Water Indicators study data from Weiskel and others (2010).

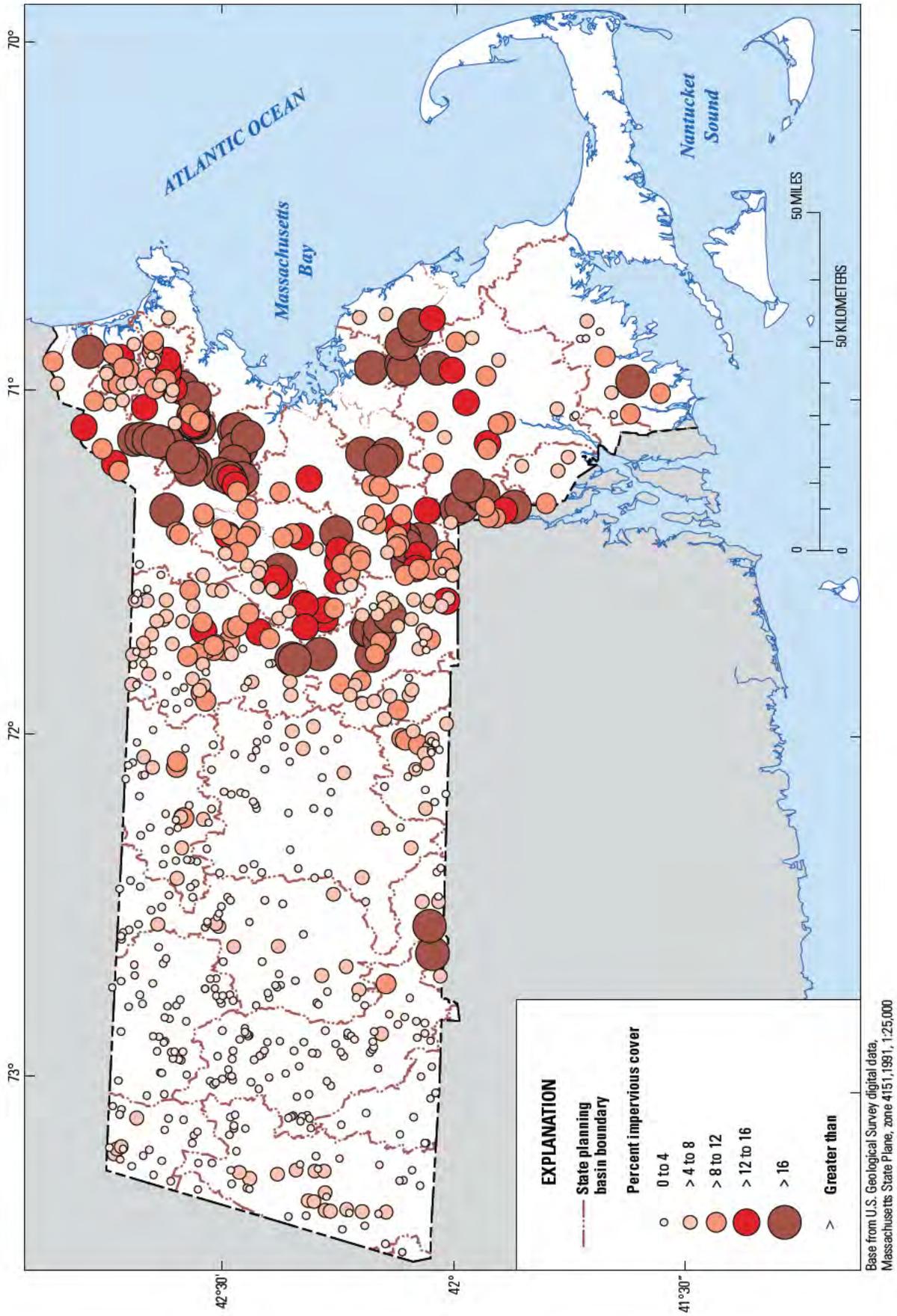


Figure 5. Impervious cover for contributing areas to 669 fish-sampling sites on Massachusetts streams. Impervious cover data from 2005 (Massachusetts Office of Geographic and Environmental Information, 2007).

Streamflow-Alteration Indicators

The magnitude and distribution of the estimated net percent alteration of August median flow represented by the 669 fish-sampling sites used in this study were compared to those reported in Weiskel and others (2010) for 1,429 subbasins and groundwater contributing areas in Massachusetts (fig. 6). The comparison indicated a similar range of percent flow alteration and distribution shape; however, the fish-sampling sites in this study were weighted slightly more toward less-altered conditions for flow-depleted sites and toward slightly more-altered conditions for flow-surcharged sites. The estimated net percent alteration of August median flow at the fish-sampling sites ranged from 100 percent net depleted to 364 percent net surcharged. Of the 669 fish-sampling sites used in the analysis, 67 percent were in net-depleted streams and 33 percent were in net-surcharged streams, under median August conditions. Most sites exhibited only a small amount of net flow alteration. About 14 percent of the sites were indicated to have greater than 10-percent depletion under median August conditions, and about 13 percent of the sites had greater than 10-percent surcharge under these conditions. A map of the percent alteration of August median flow for the fish-sampling sites used for this study (fig. 7) indicates that most sites that have a net surcharge

or net depletion of greater than 10-percent alteration of August median flows are in the highly populated areas just outside the Massachusetts Water Resources Authority (MWRA) Boston-MetroWest water and sewer service area.

Summary statistics representing the percent alteration of August median flow and mean annual flow for the 669 fish-sampling sites used in this study are given in table 5. The flow alteration indicators are presented in two forms: as individual components of flow alteration, such as the percent alteration of August median flow from groundwater withdrawals, surface-water returns, and septic returns; and as indicators of net flow alteration, such as the net percent alteration of August median flow for depleted and surcharged sites. Data indicating the percent alteration from surface-water withdrawals were not available on a monthly basis and are presented only for the mean annual flow statistic. A comparison of the median and 75th percentiles with the mean and maximum values indicates that flow alterations are small for a large proportion of the dataset.

Estimated altered mean annual flows, normalized by drainage area, ranged from 0 to 2.7 ft³/s/mi² with a median value of 2.0 ft³/s/mi². The estimated altered August median flow for the study sites, normalized by drainage area, ranged from 0 to 1.5 ft³/s/mi² with a median value of 0.20 ft³/s/mi². The WUI indicator ranged from 0 percent to 282 percent of unaffected flows.

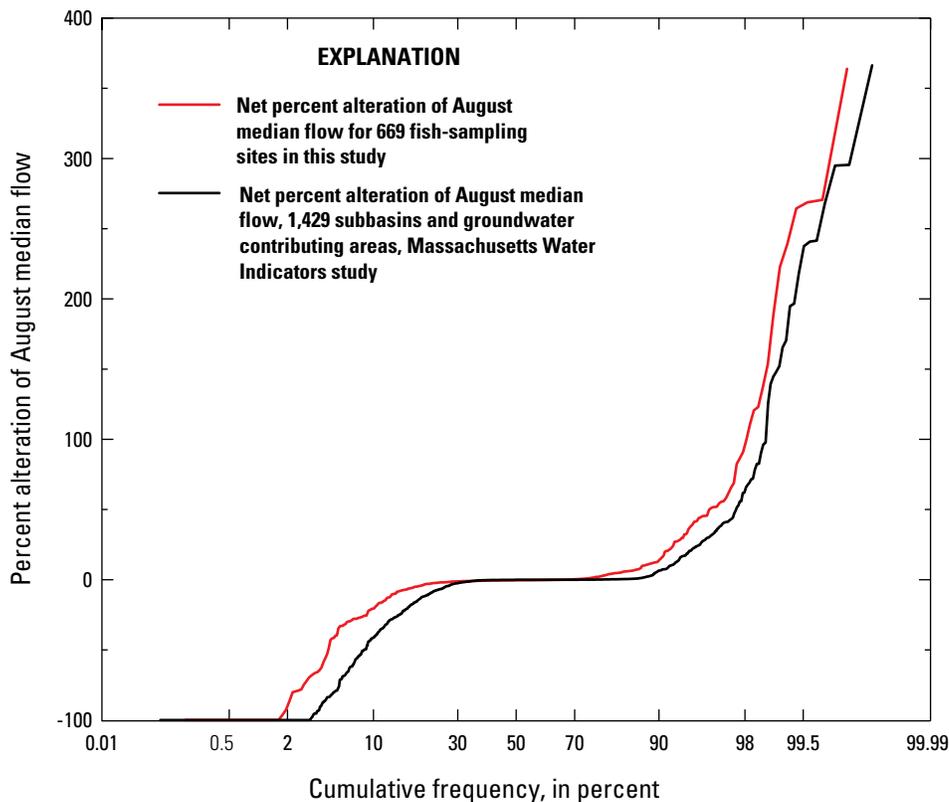


Figure 6. Cumulative frequency distributions of net percent alteration of August median flow for 669 fish-sampling sites and 1,429 subbasins and groundwater contributing areas across Massachusetts. Massachusetts Water Indicators study data from Weiskel and others (2010).

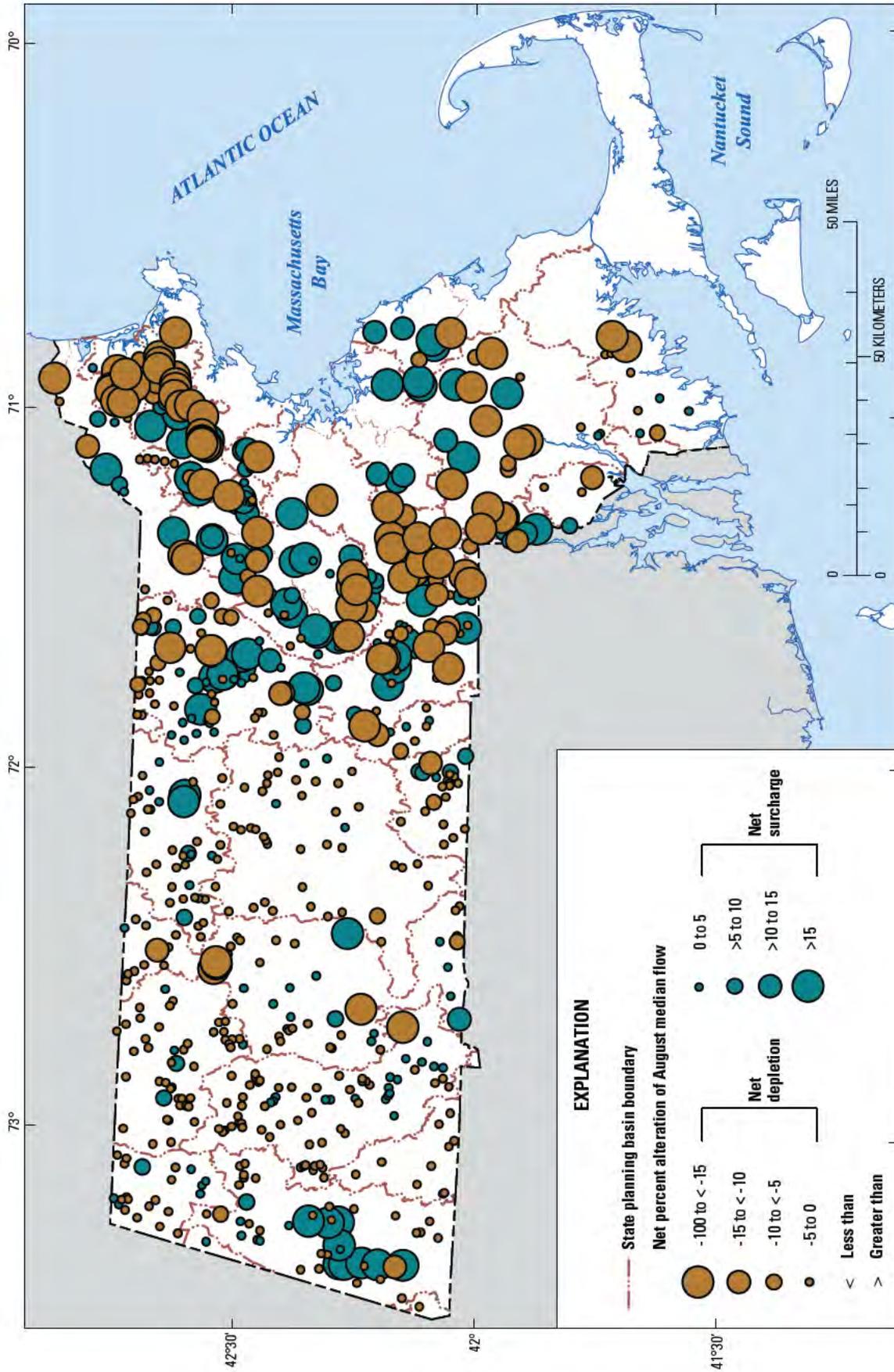


Figure 7. Net percent alteration of August median flows at 669 fish-sampling sites on Massachusetts streams.

Table 5. Descriptions, reference codes, units, and statistical summaries of values for flow-alteration variables.

Variable name	Code	Unit	Summary statistic					Mean
			Minimum	25th percentile	Median	75th percentile	Maximum	
Flow-alteration components								
Percent alteration of August median flow from surface-water returns	AUGswRp	Percent alteration from unaltered August median flow	0.0	0.0	0.0	0.0	372.3	7.5
Percent alteration of August median flow from septic returns	AUGsepRp	Percent alteration from unaltered August median flow	0.0	1.8	4.1	11.9	70.2	8.1
Percent alteration of August median flow from groundwater withdrawals	AUGgwWp	Percent alteration from unaltered August median flow	0.0	2.1	4.3	14.4	100.0	14.2
Percent alteration of mean annual flow from surface-water withdrawals	ANNswW	Percent alteration from unaltered August median flow	0.0	0.0	0.0	0.3	726.0	18.2
Percent alteration of mean annual flow from total withdrawals	ANNWp	Percent alteration from unaltered mean annual flow	0.0	0.2	0.5	3.2	281.7	4.5
Percent alteration of mean annual flow from total returns	ANNRp	Percent alteration from unaltered mean annual flow	0.0	0.2	0.6	1.9	52.5	2.0
Net-flow alteration								
Net percent alteration of August median flow, surcharged sites	AugNETRp	Percent alteration from unaltered August median flow	0.0	0.0	0.0	1.3	363.9	7.9
Net percent alteration of August median flow, depleted sites	AugNETWp	Percent alteration from unaltered August median flow	0.0	0.0	0.4	1.9	100.0	6.7
Water-Use Intensity	WUI	Percent	0.0	0.4	1.4	5.3	281.9	6.5
Net percent alteration of mean annual flow for surcharged sites	AnnNetRp	Percent alteration from unaltered annual mean flow	0.0	0.0	0.0	0.1	32.3	0.7
Net percent alteration of mean annual flow for depleted sites	AnnNetWp	Percent alteration from unaltered annual mean flow	0.0	0.0	0.0	0.5	66.9	2.2

Variable Reduction and Selection

Associations among environmental and anthropogenic factors were initially evaluated by use of Spearman rank correlations and pairwise scatterplots for all pairs of variables (pairplots) (fig. 8). When plotted together, many explanatory variables exhibit wedge-shaped scatterplots, indicating that there are clear limits for certain combinations of natural basin characteristics. For example, few basins have both high basin slope and large areas of sand and gravel or wetlands.

More than 150 explanatory variables were evaluated using a variable reduction process that used PCA and Spearman rank correlation to isolate a subset of variables that accounted for the greatest proportion of variance while minimizing redundancy in the data set. Variables were drawn from six PC axes that were identified as explaining significant proportions of the variance in the data when compared to the broken stick distribution (Jackson, 1993; McGarigal and others, 2000). Variables representing impervious cover, percent forest, percent development, and flow alteration indicators had high loadings on PC axis 1 (PC1). Intercorrelations among the environmental and land-use data were expected and found during the variable reduction process. Results of the Spearman rank correlation indicate that percent impervious cover was correlated with many environmental and land-use variables, including percent developed land (0.97), percent impervious cover in the 240-meter buffer area (0.93), percent forest (-0.92), percent developed land in a 240-meter buffer area (0.89), percent forest in a 240-meter buffer area (-0.87), elevation (-0.77), OutletX (0.76), and basin slope (-0.74). Variables that were correlated with percent impervious cover ($\rho > 0.70$) were removed from the analysis.

The estimated indicators of flow alteration were highly correlated among each of the 12 months (Jan.–Dec.) within most categories of flow alteration. Indicators representing the percent alteration of August median flow were selected for use in the analysis. Monthly median flows and indicators of percent flow alteration for other summer months, such as July or September, were highly correlated with August median flow and could have been used in the analysis with potentially similar results.

Indicators representing the individual components of flow alteration, such as the indicators of percent alteration of August surface-water returns, August septic returns, and August groundwater withdrawals had higher loading values than indicators representing percent alteration from August net returns and net withdrawals; therefore, indicators representing individual components of flow alteration were retained for use in the analysis. Spearman rank correlation also determined

that some measures of flow alteration were highly correlated with percent impervious cover, such as the indicators of percent alteration of annual mean flow from returns (0.85), percent alteration of August median flow from septic returns (0.80), and WUI (0.76). Flow-alteration indicators that were correlated with percent impervious cover ($\rho > 0.70$) were removed from the analysis.

Most variables were selected from PC1 with the exception of percent agriculture in the buffer area (PC4); percent open water (PC5); and drainage area (PC6). The variable for percent sand and gravel was carried forward for statistical analyses but was not selected through this process. Percent sand and gravel had a correlation value ($\rho = 0.71$) near that of the selected correlation cutoff ($\rho = 0.70$) but was retained because it was considered an important variable representing the potential for groundwater contribution to streamflow.

A set of 15 non-redundant environmental and land-use variables and indicators of flow alteration that extract the dominant patterns of variation from the data set were selected as a result of the PCA and Spearman rank correlation variable-reduction process to test for use as explanatory variables in the regression analyses (table 6). These variables are representative of the different types of environmental and stressor variables, and include two variables representing regional-scale physiographic characteristics (drainage area, percent sand and gravel) and one representing a more local measure (channel slope) that is generally indicative of site flow velocity and substrate conditions. Also included are four flow-alteration indicators (percent alteration of August median flow from groundwater withdrawals, percent alteration of August median flow from surface-water returns, and net percent alteration of mean annual flow for depleted sites, percent alteration of annual mean flow from surface-water withdrawals, four measures of dams (total undammed reach length, upstream undammed reach length along the main stem, dam density, and percent open water), and four land-use variables (percent wetland and percent agriculture in a 240-meter buffer along the upstream stream network; and percent impervious cover or percent forest in the contributing area).

Hilborn and Mangel (1997) and Burnham and Anderson (2002) recommend development of multiple candidate models for the exploration of highly correlated variables. Consequently, an alternative data set was created using the above process for variable selection but including percent forest in place of percent impervious cover. The percent impervious cover and percent forest data sets were selected for use to illustrate the manner in which correlated variables could be used in different models with similar results.

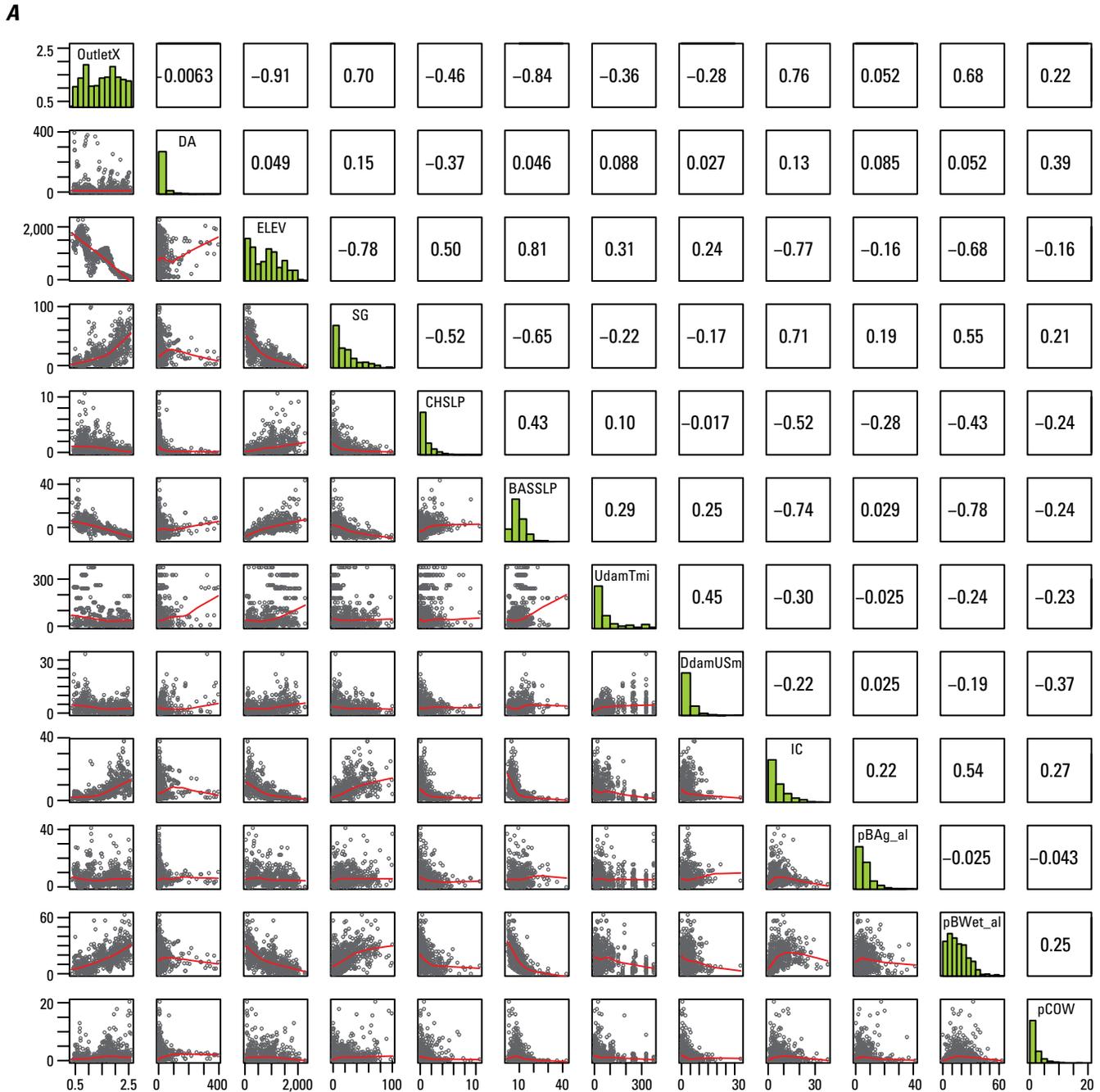


Figure 8. Scatterplot correlation matrix showing relations between A, environmental factors, and B, flow-alteration variables. The main diagonal panels show histograms of each variable. The lower left panels show pair-wise scatterplots with a smoothing LOWESS curve. The upper right panels show Spearman rank correlation coefficients. The headings in the main diagonal panels indicate which variables are plotted along the x and y axes for each panel of the scatterplot matrix and which variables are related in the correlation matrix. Names and units for environmental and flow variables are given in tables 2, 3, and 4.

B

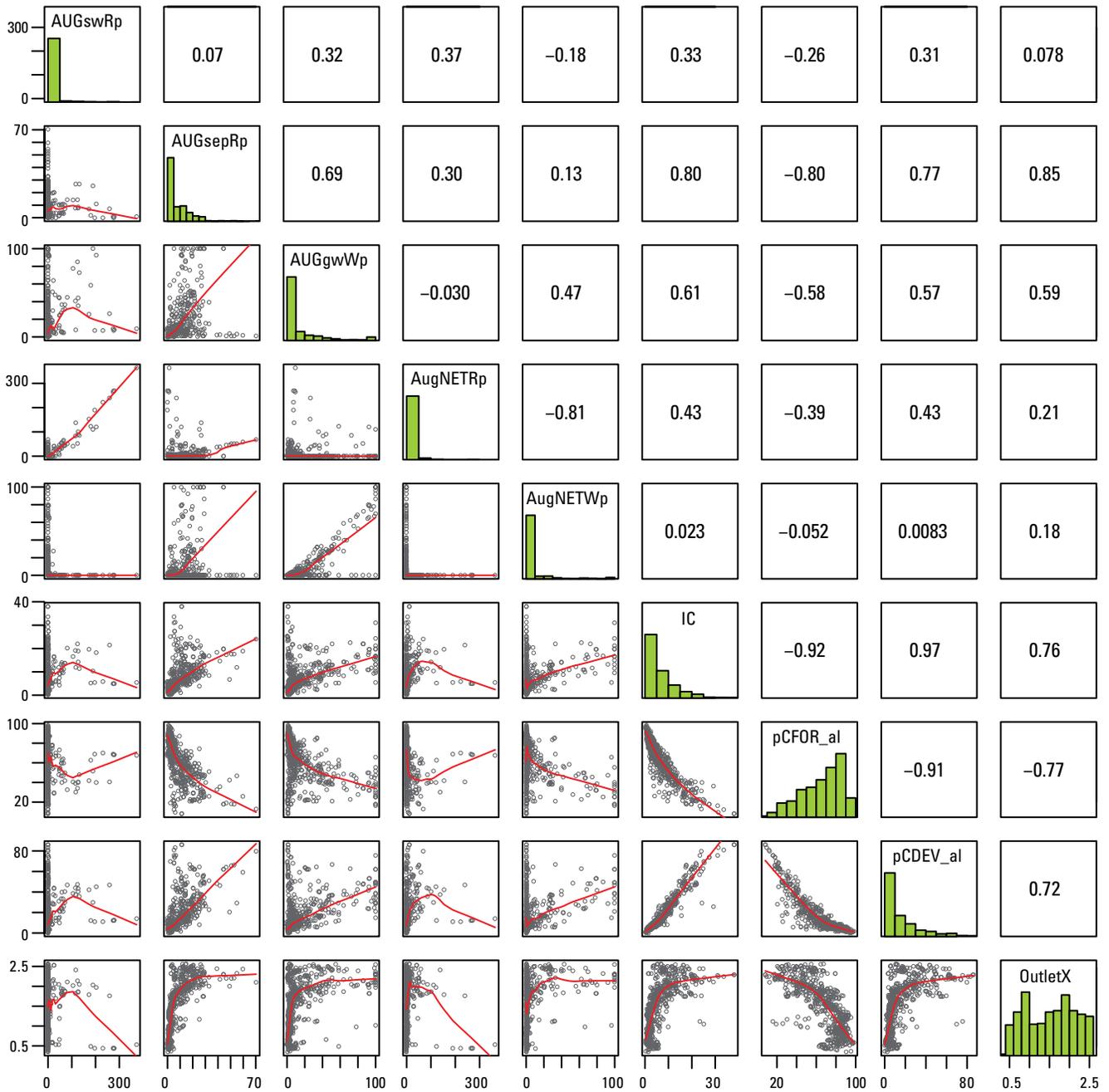


Figure 8. Scatterplot correlation matrix showing relations between A, environmental factors, and B, flow-alteration variables. The main diagonal panels show histograms of each variable. The lower left panels show pair-wise scatterplots with a smoothing LOWESS curve. The upper right panels show Spearman rank correlation coefficients. The headings in the main diagonal panels indicate which variables are plotted along the x and y axes for each panel of the scatterplot matrix and which variables are related in the correlation matrix. Names and units for environmental and flow variables are given in tables 2, 3, and 4. —Continued

Table 6. Descriptions, reference codes, units, and statistical summaries of values for environmental and anthropogenic variables.

Variable name	Code	Unit	Statistical summary of values					Mean
			Minimum	25th percentile	Median	75th percentile	Maximum	
Percent forest	pCFOR_al	Percent area	8.2	51.0	70.1	82.9	98.1	65.6
Percent buffer wetland	pBWet_al	Percent buffer area	0.0	7.6	14.3	23.1	63.0	16.2
Drainage area	DA	Square miles	2.0	3.8	7.4	22.0	394.5	25.9
Channel slope	CHSLP	Percent rise	0.0	0.0	0.5	1.5	10.6	1.0
Percent sand and gravel	SG	Percent area	0.0	3.7	16.1	31.2	100.0	21.2
Percent impervious cover	IC	Percent area	0.2	2.0	4.2	9.1	38.1	6.4
Percent alteration of August median flow from groundwater withdrawals	AUGgwWp	Percent unaltered	0	2.0	4.2	14.5	100	14.1
Undammed reach length	UdamTmi	August median flow	0.1	14.8	38.7	95.1	375.8	78.5
Percent open water	pCOW	Percent area	0	0.1	0.7	2.4	20.3	1.7
Percent alteration of August median flow from surface-water returns	AUGswRp	Percent unaltered August median flow	0	0	0.0	0.0	372.2	7.4
Undammed reach length upstream along mainstem	DdamUSmi	Miles	0	1.0	2.9	4.8	33.3	3.5
Percent buffer agriculture	pBAg_al	Percent buffer area	0	2.8	4.9	8.6	41.1	6.4
Percent alteration of annual surface-water withdrawals	ANNswW	Percent unaltered mean annual flow	0	0	0	0.2	725.9	18.1
Dam density	DamDsqmi	Dams per square mile	0	0	0.2	0.4	2.6	0.3
Net percent alteration of mean annual flow for depleted sites	AnnNetWp	Percent alteration from unaltered annual mean flow	0	0	0	0.5	66.9	2.2

Fish-Assemblage Characteristics

The fish-assemblage data analyzed included information on 83,221 individual fish of 45 species from 669 fish-sampling sites. Histograms of the fish metrics indicate that many distributions were highly non-normal and had a positive or right skew; most observations were at or near zero values (fig. 9). Only 20 of these species were present at greater than 10 percent of the sites. The five most common species in terms of species occurrence among sites were white sucker, blacknose dace, brook trout, pumpkinseed, and longnose dace (table 7). These five species were each present at about 53, 48, 41, 39, and 33 percent of the sites, respectively. The maximum observed species richness at a site was 15 (range 0–15), and the mean and median values for the observed species richness

were each 6 species. In terms of total abundance, 20 species accounted for 95 percent of the total number of fish sampled. The five most common species in terms of abundance were blacknose dace, white sucker, longnose dace, brook trout, and fallfish. Fish relative abundance, measured as the number of fish sampled per hour of electrofishing, ranged from 4 to 2,500 fish per hour.

The number of species used for analysis varied depending on the type of statistical analysis applied. All 45 species were used to determine relative abundance and species richness for the fluvial and macrohabitat-generalist fish-assemblage metrics. With infrequently captured fish species removed, a set of 21 species, shown in bold in table 7, was used for all multivariate analyses.

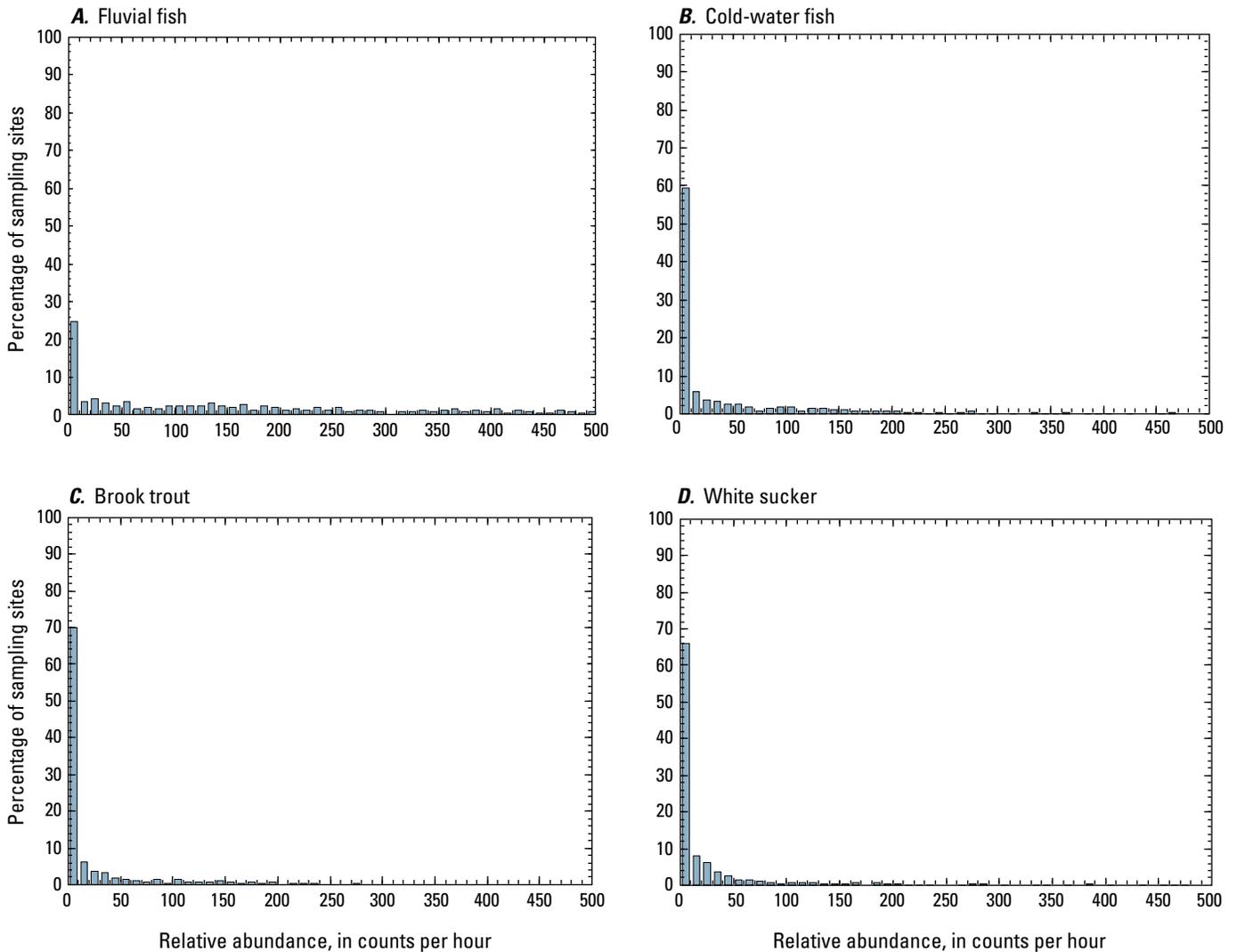


Figure 9. Relative abundance of A, fluvial fish, B, cold-water fish, C, brook trout, and D, white sucker in relation to percentage of selected sampling sites on Massachusetts streams. Fish samples were collected from 1998 to 2008.

32 Factors Influencing Riverine Fish Assemblages in Massachusetts

Table 7. Common names of fish, number and percentage of sites sorted by the number of sites at which each species was captured, and number and percentage of individuals sorted by the number of individuals captured at sites used for the assessment of fish assemblages in Massachusetts streams.

[Fish species in **bold** type were selected for use in the multivariate analysis]

Common name	Number of sites	Percent of sites	Common name	Number of individuals	Percent of individuals
White sucker	353	52.77	Blacknose dace	22,236	26.84
Blacknose dace	318	47.53	White sucker	7,260	8.72
Brook trout	276	41.26	Longnose dace	6,585	7.91
Pumpkinseed	260	38.86	Brook trout	6,554	7.88
Longnose dace	224	33.48	Fallfish	5,059	6.08
Bluegill	211	31.54	Slimy sculpin	4,980	5.98
Fallfish	178	26.61	Common shiner	4,655	5.59
Largemouth bass	172	25.71	Redfin pickerel	3,401	4.09
American eel	171	25.56	American eel	2,905	3.49
Redfin pickerel	171	25.56	Pumpkinseed	2,626	3.16
Chain pickerel	169	25.26	Bluegill	2,351	2.83
Yellow bullhead	167	24.96	Creek chub	1,817	2.18
Common shiner	152	22.72	Tesselated darter	1,614	1.94
Tesselated darter	134	20.03	Brown trout	1,549	1.86
Brown bullhead	124	18.54	Rock bass	963	1.16
Golden shiner	121	18.09	Yellow bullhead	942	1.13
Slimy sculpin	119	17.79	Largemouth bass	936	1.12
Brown trout	116	17.34	Redbreast sunfish	919	1.10
Yellow perch	105	15.70	Golden shiner	861	1.03
Creek chub	93	13.90	Smallmouth bass	743	0.89
Redbreast sunfish	53	7.92	Yellow perch	712	0.86
Banded sunfish	52	7.77	Chain pickerel	686	0.82
Creek chubsucker	43	6.43	Bluntnose minnow	561	0.67
Smallmouth bass	35	5.23	Brown bullhead	457	0.55
Rock bass	33	4.93	Banded sunfish	378	0.45
Swamp darter	29	4.33	Spottail shiner	359	0.43
Longnose sucker	24	3.59	Landlocked salmon	208	0.25
Spottail shiner	14	2.09	Longnose sucker	180	0.22
Bluntnose minnow	13	1.94	Creek chubsucker	152	0.18
Black crappie	12	1.79	American brook lamprey	79	0.09
Green sunfish	11	1.64	Lake chub	63	0.08
Lake chub	8	1.20	Swamp darter	62	0.07
Landlocked salmon	7	1.05	Common carp	46	0.06
American brook lamprey	6	0.90	Rainbow trout	44	0.05
Rainbow trout	6	0.90	Cutlips minnow	40	0.05
Banded killifish	5	0.75	Green sunfish	32	0.04
Bridle shiner	3	0.45	Central mudminnow	26	0.03
Common carp	3	0.45	Black crappie	21	0.03
Central mudminnow	3	0.45	Bridle shiner	17	0.02
Northern pike	3	0.45	White perch	13	0.02
White perch	3	0.45	Banded killifish	10	0.01
Goldfish	2	0.30	Tadpole madtom	7	0.01
Cutlips minnow	1	0.15	Northern pike	5	0.01
Tadpole madtom	1	0.15	White catfish	4	0.00
White catfish	1	0.15	Goldfish	3	0.00

Patterns in Fish-Assemblage Structure

Patterns in fish-assemblage structure among the sites were examined using hierarchical cluster analysis and NMDS. The environmental and fish data were analyzed using multivariate routines available in the PRIMER version 6.1 software package (Clarke and Warwick, 2001; Clarke and Gorley, 2006). Fish species abundance data were square-root transformed, and environmental data were standardized to z-scores by subtracting the sample mean and dividing by the sample standard deviation. Bray-Curtis similarity coefficients were used as distance measures for the fish data, and Euclidean distance measures were calculated to define the distance between environmental variables.

Patterns in relative abundance among fish species were investigated by use of a hierarchical cluster analysis. Samples and groups were joined in the cluster analysis on the basis of an unweighted group-average algorithm. The cluster analysis combines fish species into groups or clusters, which can be illustrated with a dendrogram (fig. 10). Fish species in the clusters show a distinct division between fluvial and generalist species at the 15-percent similarity level. The clustering also

indicates a subdivision between cold-water species (brook trout, brown trout, and sculpin) and the remainder of the fluvial species.

Fish-assemblage structure was also investigated using an NMDS ordination. Multiple NMDS runs were made using 100 restarts and a minimum stress value of 0.001. Runs were made for both two-dimensional and three-dimensional solutions. The fit of the NMDS ordinations are quantified by stress values that indicate how well the among-sample distances in the NMDS ordination preserve the actual sample dissimilarities. Stress values in the ordinations did not change from the outcome of the original run when the analysis was repeated, supporting that the ordinations represent sample dissimilarities accurately (Clarke and Gorley, 2006). Stress values for the two-dimensional (2-D) and three-dimensional (3-D) ordinations were 0.134 and 0.085, respectively. Both stress values are relatively low (< 0.20), indicating that the ordinations provide a reasonable representation of the original distances (Legendre and Anderson, 1999; Clarke and Warwick, 2001). The ordination biplots for both the 2-D (fig. 11) and 3-D solutions indicate distinct differences between fish species classified as fluvial and macrohabitat generalists.

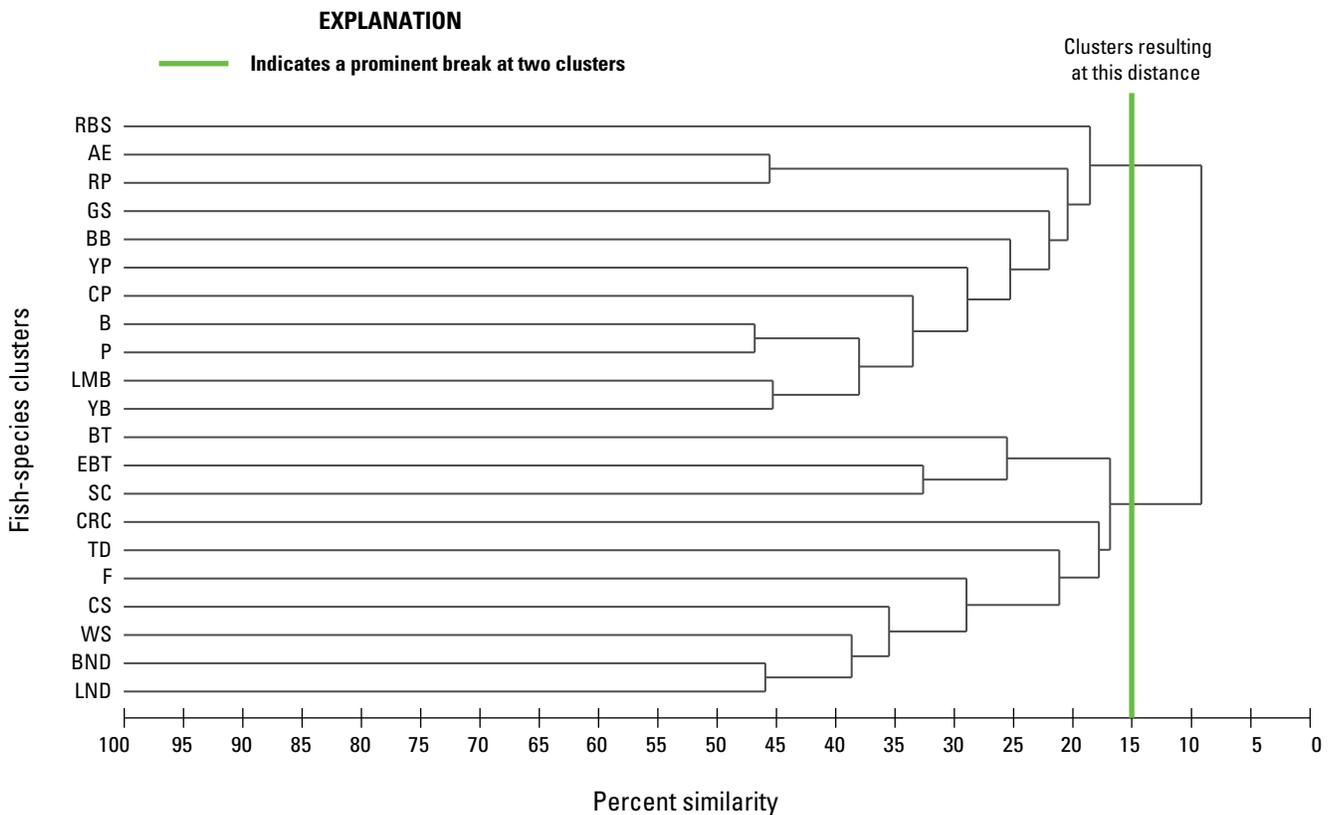


Figure 10. Hierarchical cluster analysis dendrogram showing groupings of 21 fish species in Massachusetts. Clustering developed on the basis of square-root-transformed Bray-Curtis similarity values of fish species using group-average clustering. Names for fish-species codes are given in table 1. Fish samples were collected from 1998 to 2008.

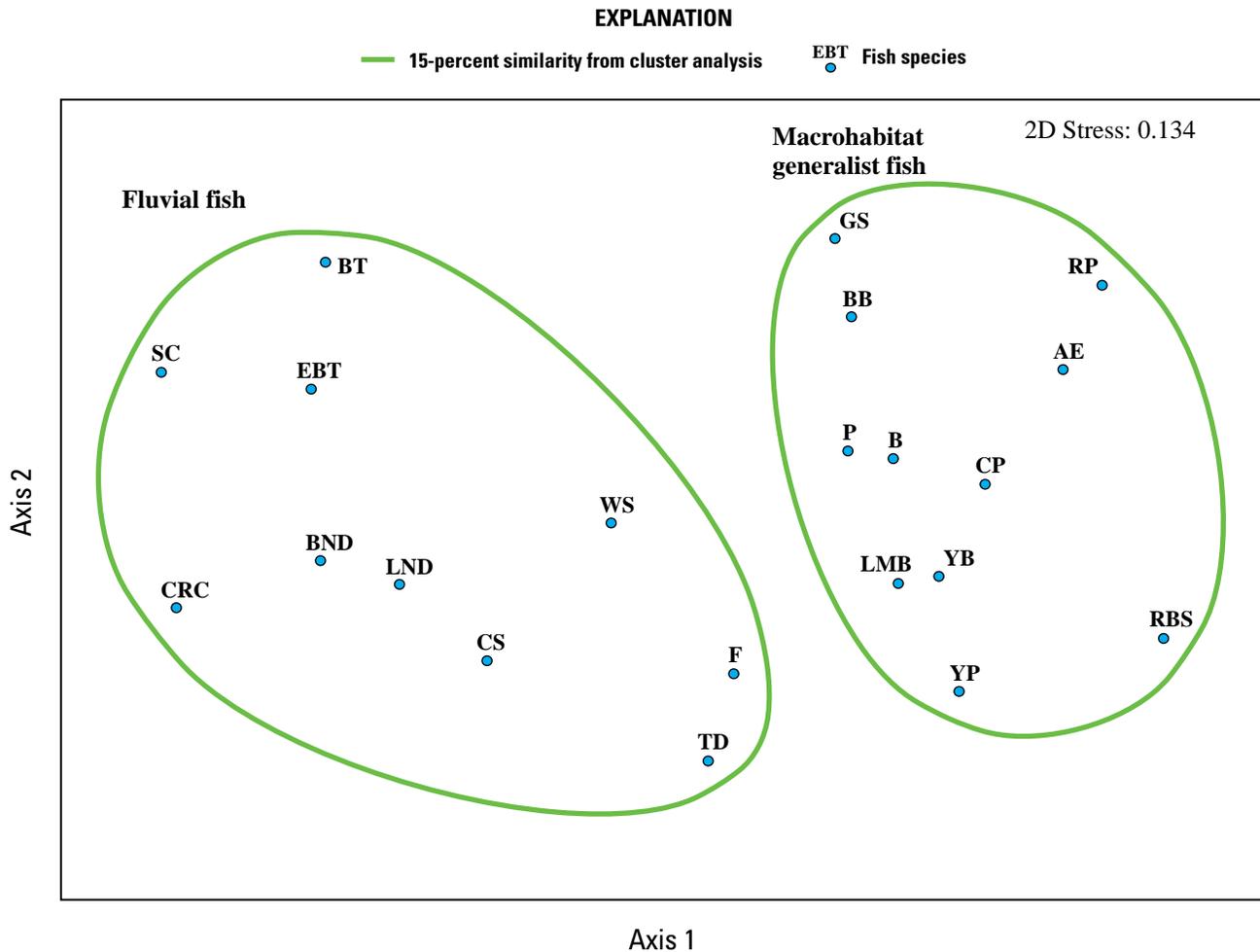


Figure 11. Nonmetric multidimensional scaling ordinations of fish-assemblage structure showing a two-dimensional ordination with contours drawn around clusters defined by the 15-percent similarity from the cluster analysis shown in figure 10, for 669 fish-sampling sites on Massachusetts streams. Ordination developed on the basis of square-root-transformed Bray-Curtis similarity values of fish species. Two dimensional stress of the ordination is included. Names for fish species codes are given in table 1. Fish samples were collected from 1998 to 2008.

Combinations of ordination and clustering can be an effective way of checking the adequacy of both representations (Clarke and Warwick, 2001). Superimposing lines to show the hierarchical clusters of fish species at a similarity level of 15 percent (solid green line) on the NMDS ordination (fig. 11) reveals agreement between the two techniques and clearly shows a division between fluvial and generalist species. This distinction between fluvial and generalist species was used as a basis for combining fluvial-specialist and fluvial-dependent species into a single fluvial classification and to justify the use of fluvial-fish metrics as the fish-response variables for the quantile regressions and GLM regression models discussed in the following sections.

Four commonly occurring species (blacknose dace, brook trout, common shiner, and white sucker) were selected for use as indicator taxa. These species represent a range of sensitivities to anthropogenic factors (pollution tolerance, temperature tolerance). The species also represent a range of biological attributes (maximum body size, migration habits, and home-range sizes) that together provide information on fluvial-fish species statewide. Fish metrics that were tested in addition to the relative abundance of individual taxa include species richness and relative abundance of fluvial and macrohabitat-generalist species, relative abundance of coldwater (stenothermal) fish, and total fish abundance.

Quantile Regression

Plots showing quantile regression lines were constructed to illustrate limiting-factor relations between fish-response variables and anthropogenic factors representing flow alteration, impervious cover, and dams. The quantile regression lines shown are for the 90th quantile with 95-percent confidence intervals. All quantile regressions shown have slopes that are significantly different from zero.

Quantile regressions were performed for several fluvial-fish species, including brook trout, blacknose dace, common shiner, and white sucker. Increases in the estimated percent alteration of August median flow from groundwater withdrawals were associated with decreases in the 90th quantile for the relative abundance of brook trout (fig. 12A) and blacknose dace (fig. 12B). These trends are consistent with life history traits for the various species. Brook trout and blacknose dace are both fluvial specialists and the steep slope of the factor ceilings depicted by the 90th quantile support their sensitivity to flow alteration. Quantile regression plots also demonstrated a negative relation between fluvial fish and the estimated percent alteration of August median flow from groundwater withdrawals. The estimated percent alteration of August median flow from groundwater withdrawals was associated with declines for both fluvial-fish relative abundance and fluvial-fish species richness (fig. 13). The narrow 95-percent confidence intervals around the relations of fluvial-fish relative abundance and fluvial-fish species richness to the estimated percent alteration of August median flow from groundwater withdrawals demonstrates a high degree of confidence in these relations. The quantile regressions demonstrate that increases in flow-alteration indicators or increases in impervious cover are, in general, associated with decreases in fluvial fish and most fluvial-fish species. The quantile regression plots indicate that (1) as many as seven fluvial-fish species are expected in streams with little flow alteration or impervious cover; (2) no more than four fluvial-fish species are expected in streams where flow alterations from groundwater withdrawals exceed 50 percent of the August median or the percent area of impervious cover exceeds 15 percent; and (3) few fluvial fish remain at high rates of withdrawal (approaching 100 percent). The fluvial-fish species-richness plot also illustrates that, for the 90th quantile, an increase in estimated percent alteration of August median flow from groundwater withdrawals from 0 to 14 percent is associated with a loss of roughly one fluvial species.

The quantile regression plots indicate that the strength and direction of the response of individual fish species can vary, depending on the species and variables compared. For example, quantile regression plots for individual species indicate a varied response to surface-water returns. Increases in surface-water returns as a percent of unaltered August median flow were associated with declines in the 90th quantile

for the relative abundance of blacknose dace (fig. 14A) and increases in the 90th quantile for white sucker (fig. 14B). White sucker are considered to be tolerant to contamination, turbidity, and low dissolved oxygen, so this result was not unexpected. Large surface-water returns occur primarily in large river systems. Consequently, when considered on a statewide basis, relatively few of the sample sites had surface-water returns, potentially influencing the strength of the response.

Increases in percent impervious cover were associated with decreases in the 90th quantile for fluvial species, including brook trout (fig. 15A) and common shiner (fig. 15B). Increases in percent impervious cover were associated with strong decreases in the 90th quantile for fluvial-fish relative abundance and species richness (fig. 16). The relatively narrow 95-percent confidence interval around the relation between fluvial-fish species richness and impervious cover indicates a high degree of confidence in this relation. The 90th quantile in the impervious cover plot illustrates that (1) for low levels of impervious cover (less than 10 percent), fluvial-fish species richness is reduced roughly one fluvial species by an increase in impervious cover of approximately 4 percent; (2) no more than four fluvial-fish species are expected in streams where the percent area of impervious cover exceeds 15 percent; and (3) few fluvial fish remain at high rates of impervious cover (25 to 30 percent).

Two different variables indicated a negative relation between dams and fluvial fish. Increases in percent open water were associated with decreases in the 90th quantile for the relative abundance of brook trout (fig. 17A) and fluvial-fish species richness (fig. 17B). The percent area of open water reflects the areal size of impoundments, lakes, and ponds in the contributing area upstream from the sampling site. Although some water bodies represented in the open-water variable do not have dams, most lakes and ponds in Massachusetts do have dams or outlet structures that control water levels. A second dam variable, undammed reach length, indicated a positive relation between fluvial-fish species richness and the length of the undammed stream reaches that are potentially available to fish at the sampling site. Increases in undammed reach length were also associated with increases in the 90th quantile for the richness of fluvial fish (fig. 18).

Quantile regression models are useful for modeling species responses to stressor gradients when those responses are linear, such as the upper bound of a wedge-shaped relation. Although many of the relations between species responses and stressor gradients examined in this study exhibited a wedge shape, some scatterplots indicated no relation or non-linear relations. For example, relations between macrohabitat generalists and impervious cover indicated a potential nonlinear, unimodal response. Linear quantile regression models would not be appropriate for these nonlinear relations.

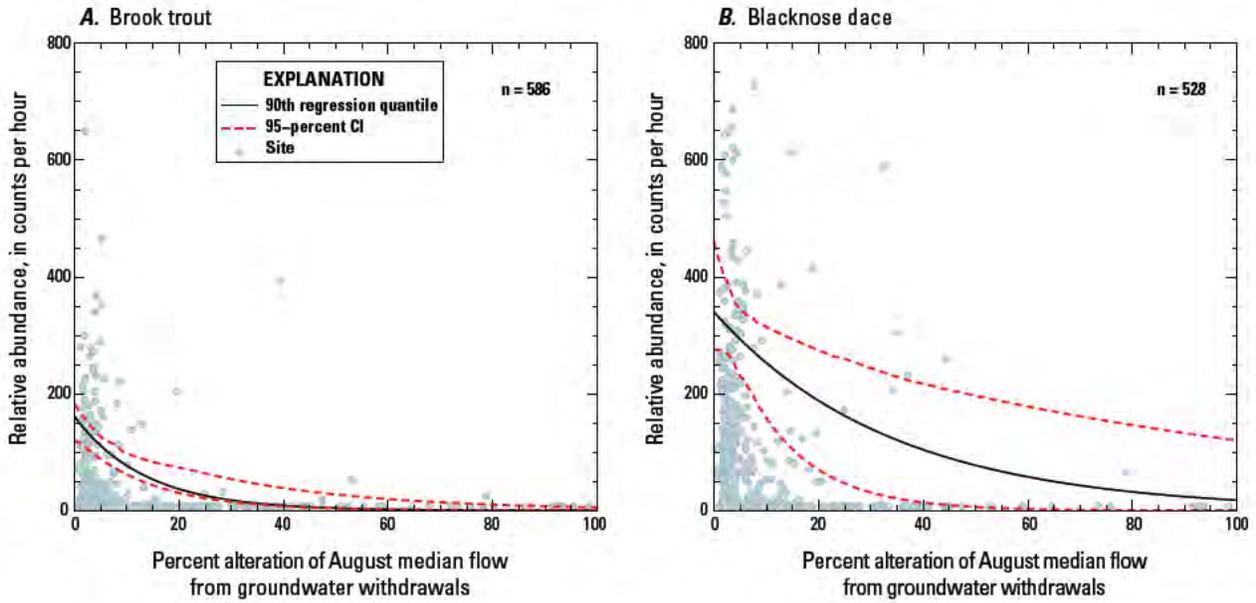


Figure 12. Quantile regression relations between relative abundance of A, brook trout and B, blacknose dace and percent alteration of August median flow from groundwater withdrawals for selected fish-sampling sites on Massachusetts streams. CI, confidence interval; n, number of sites. Fish samples were collected from 1998 to 2008.

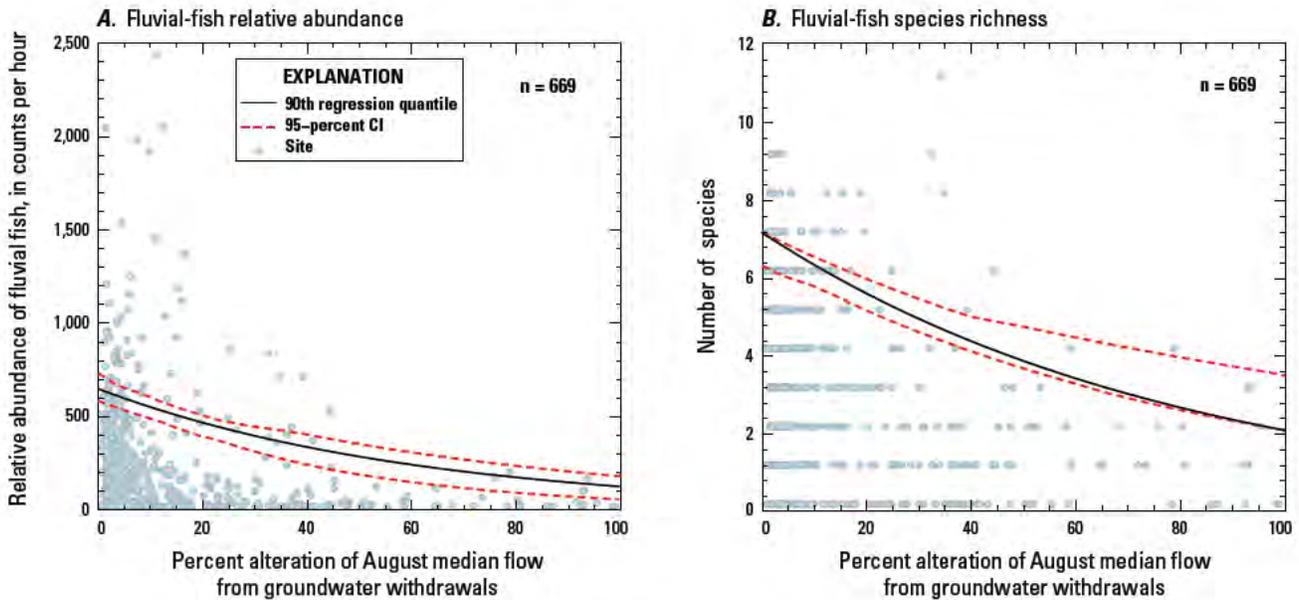


Figure 13. Quantile regression relations between A, fluvial-fish relative abundance and B, fluvial-fish species richness and percent alteration of August median flow from groundwater withdrawals for selected fish-sampling sites on Massachusetts streams. CI, confidence interval; n, number of sites. Fish samples were collected from 1998 to 2008.

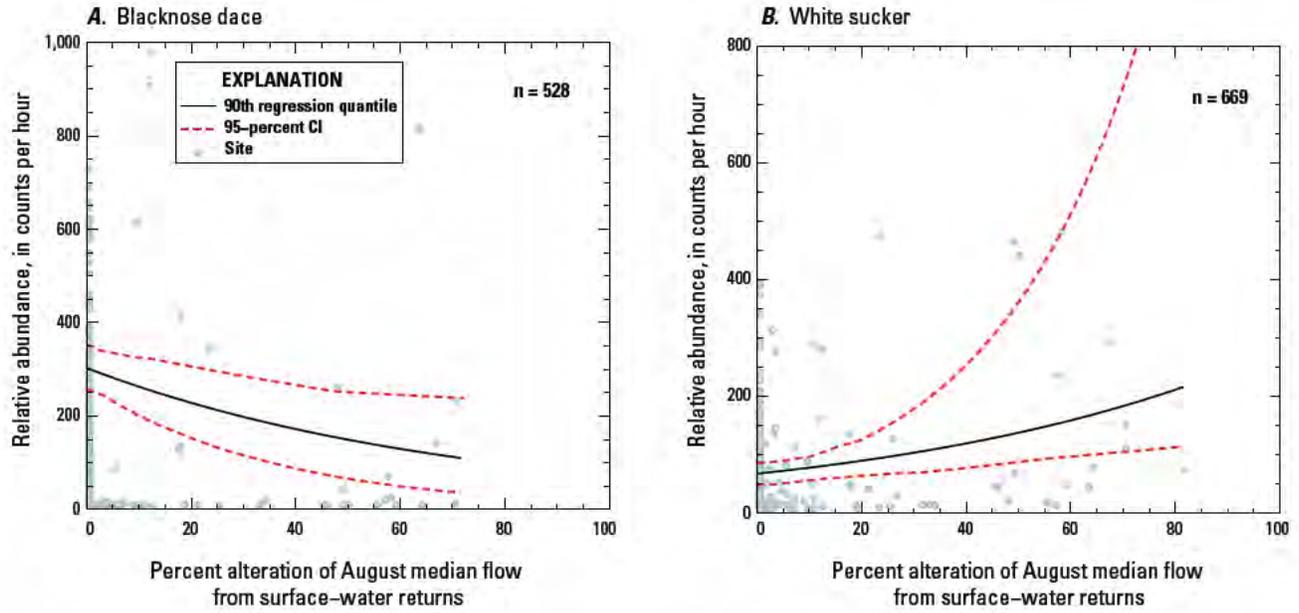


Figure 14. Quantile regression relations between relative abundance of A, blacknose dace and B, white sucker and percent alteration of August median flow from surface-water returns for selected fish-sampling sites on Massachusetts streams. CI, confidence interval; n, number of sites. Fish samples were collected from 1998 to 2008.

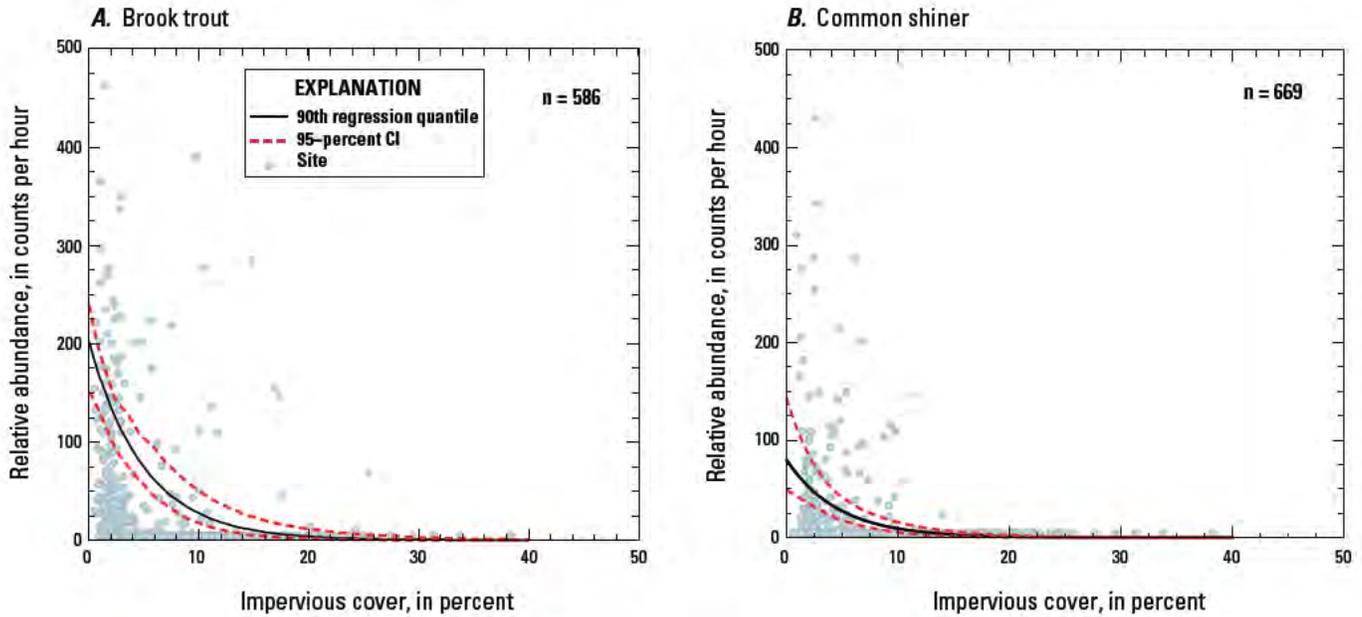


Figure 15. Quantile regression relations between relative abundance of A, brook trout and B, common shiner and percent impervious cover for the contributing areas to selected fish-sampling sites on Massachusetts streams. CI, confidence interval; n, number of sites. Fish samples were collected from 1998 to 2008.

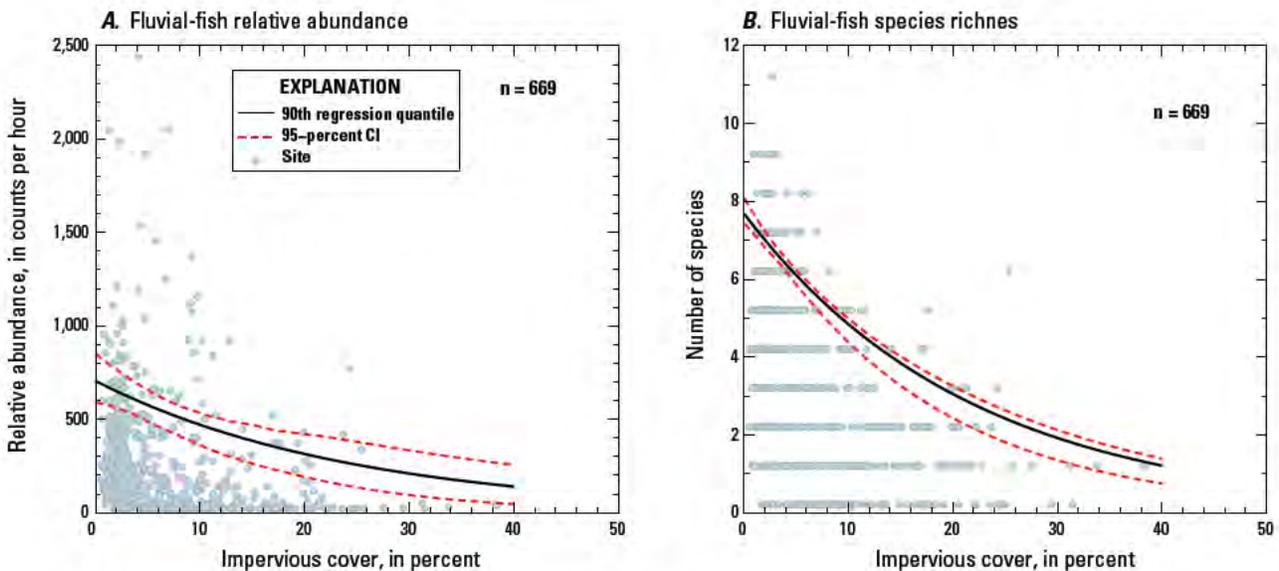


Figure 16. Quantile regression relations between A, fluvial-fish relative abundance and B, fluvial-fish species richness and percent impervious cover for the contributing areas to selected fish-sampling sites on Massachusetts streams. CI, confidence interval; n, number of sites. Fish samples were collected from 1998 to 2008.

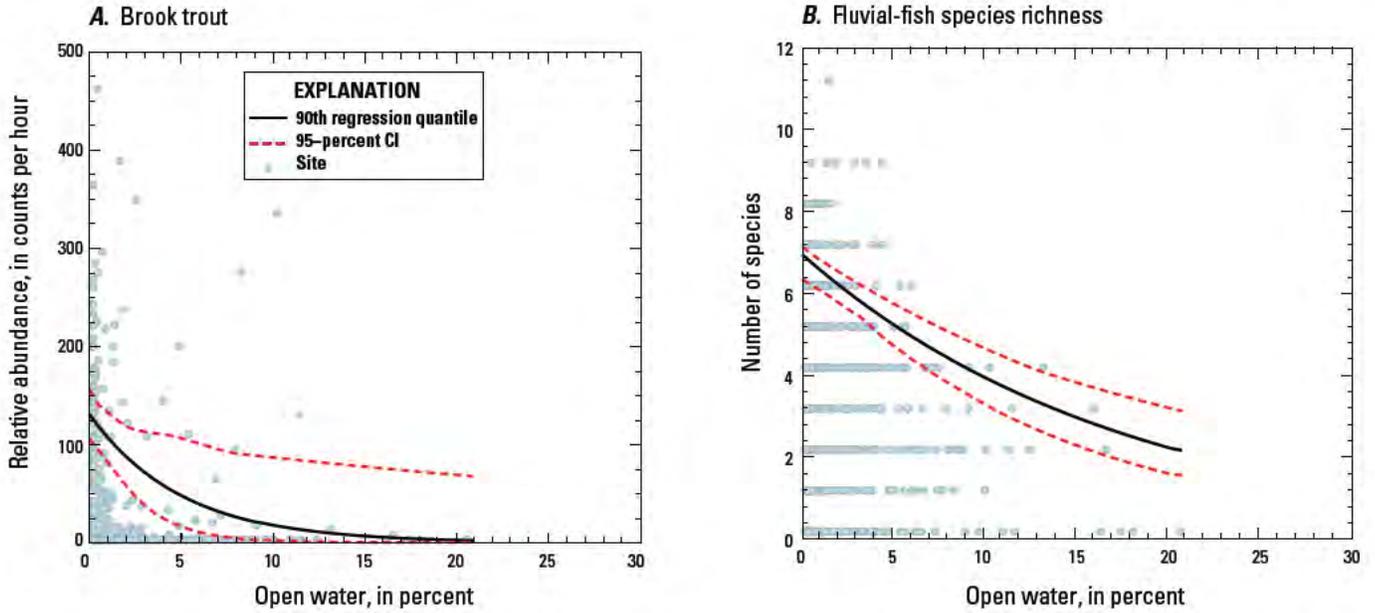


Figure 17. Quantile regression relations between A, relative abundance of brook trout and B, fluvial-fish species richness and percent open water for the contributing areas to selected fish-sampling sites on Massachusetts streams. CI, confidence interval; n, number of sites. Fish samples were collected from 1998 to 2008.

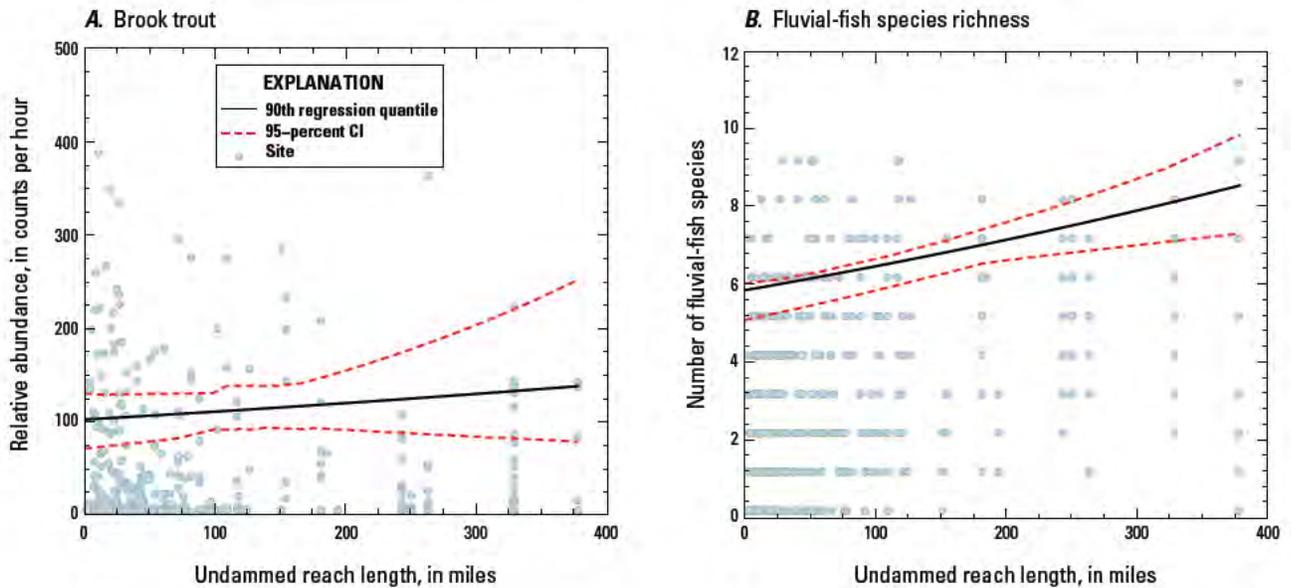


Figure 18. Quantile regression relations between A, relative abundance of brook trout and B, fluvial-fish species richness and undammed reach length for selected fish-sampling sites on Massachusetts streams. CI, confidence interval; n, number of sites. Fish samples were collected from 1998 to 2008.

Generalized Linear Models

Several GLM models were developed to characterize the association between fish-response variables and environmental and anthropogenic explanatory variables. The three strongest GLM models are those for fluvial-fish species richness, fluvial-fish relative abundance, and brook trout relative abundance. The model selection process and each of the GLM models are described in the following sections, and figures illustrating relations between fish metrics and key variables are presented. A range of fish response variables was tested for use in the GLM regression models. Models tested for coldwater fish, macrohabitat generalists, and several additional indicator species (white sucker, longnose dace) were either not significant at the 0.05 level when compared to the null model results or exhibited substantial bias and so are not included in this report.

Significant variables for each model, their coefficients, and measures of goodness of fit for comparisons of observed and predicted values are listed in table 8. Impervious cover was a significant variable for all three of the equations. Percent forest was a significant variable for all three of the equations from the alternate pool of variables that included percent forest in place of percent impervious cover. Other variables in the equations were drainage area, channel slope, percent wetland and percent agriculture in a buffer area, estimated percent

alteration of August flow from groundwater withdrawals (municipal + non-municipal + private), undammed reach length upstream, and percent open water in the contributing area.

Variable coefficients in the models can be interpreted in a manner similar to that of coefficients from a multilinear OLS log-normal regression. For example, negative coefficients in the fluvial-fish models indicate that an increase in the explanatory variable is associated with a decrease in the fish-assemblage response variable, and positive coefficients in the models indicate that an increase in the explanatory variable is associated with an increase in the fish-assemblage response variable. The magnitude of the coefficient can be used to estimate the percent change in the response variable that is associated with a unit change in one covariate, keeping all other variables the same. A constant percent change in the response variable can be determined because the equations are linear in log space. In other words, because the GLM equations have an exponential form, with the explanatory variables in the exponents on the right side of the equation, a unit change in an explanatory variable leads to a multiplicative change in the response variable, which can be expressed as a constant percent increase or decrease. A graphical approach was used to depict the effect of a change in magnitude of particular covariates on the response variable.

Table 8. A, Significant variables and coefficients and B, measures of goodness of fit for generalized linear model equations used to determine relations between environmental and anthropogenic factors and indicator species and fish-assemblage metrics in Massachusetts streams.

[Independent variables and coefficients are for equation 2. Abbreviations for independent variables are given in tables 2, 3, and 4. CI, confidence interval; SE, standard error; <, less than]

A. Significant variables and coefficients						
Model	Independent variable	Coefficient	SE	p-value		
Variable pool for models including percent impervious cover						
Fluvial-fish species richness	Intercept	1.7911	0.0664	<0.0010		
	DA	0.0011	0.0004	0.0037		
	CHSLP	-0.0968	0.0206	<0.0010		
	UdamTmi	0.0008	0.0002	0.0009		
	pBWet_al	-0.0262	0.0029	<0.0010		
	IC	-0.0557	0.0063	<0.0010		
Fluvial-fish relative abundance	Intercept	6.1523	0.0942	<0.0010		
	CHSLP	-0.0840	0.0361	0.0202		
	AUGgwWp	-0.0091	0.0042	0.0286		
	pBWet_al	-0.0289	0.0059	<0.0010		
	IC	-0.0373	0.0132	0.0047		
Brook trout relative abundance	Intercept	4.9336	0.2012	<0.0010		
	DA	-0.1291	0.0256	<0.0010		
	pCOW	-0.2172	0.0948	0.0223		
	IC	-0.0916	0.0305	0.0028		
Variable pool for models including percent forest						
Fluvial-fish species richness	Intercept	-0.1231	0.1996	0.5378		
	DA	0.0011	0.0004	0.0038		
	CHSLP	-0.0862	0.02114	<0.0010		
	UdamTmi	0.0006	0.0002	0.0011		
	pBAg_al	0.0157	0.0047	0.0008		
	pBWet_al	-0.0147	0.0036	<0.0010		
	pCFOR_al	0.0192	0.0021	<0.0010		
Fluvial-fish relative abundance	Intercept	4.9888	0.3596	<0.0010		
	CHSLP	-0.0878	0.0365	0.0164		
	AUGgwWp	-0.0091	0.0041	0.0254		
	pBWet_al	-0.0227	0.0067	<0.0010		
	pCFOR_al	0.0125	0.0040	0.0019		
Brook trout relative abundance	Intercept	2.0466	0.6284	0.0012		
	DA	-0.1380	0.0269	<0.0010		
	pCFOR_al	0.0317	0.0075	<0.0010		
B. Measures of goodness of fit						
Model	Pearson's r	Pseudo-R squared (percent)	Intercept of linear model of observed versus predicted	95-percent CI for intercept	Slope of linear model of observed versus predicted	95-percent CI for slope
Variable pool for models including percent impervious cover						
Fluvial-fish species richness	0.6	33.1	0.15	-0.18 to 0.48	0.95	0.85 to 1.05
Fluvial-fish relative abundance	0.5	18.2	24.82	12.50 to 37.14	0.74	0.64 to 1.38
Brook trout relative abundance	0.50	34.7	0.72	-1.45 to 2.89	0.93	0.80 to 1.06
Variable pool for models including percent forest						
Fluvial-fish species richness	0.59	33.3	0.16	-0.18 to 0.50	0.95	0.85 to 1.05
Fluvial-fish relative abundance	0.51	18.4	24.87	12.67 to 37.07	0.74	0.64 to 1.38
Brook trout relative abundance	0.52	34.6	0.53	-1.60 to 2.66	0.95	0.82 to 1.08

Fluvial-Fish Species-Richness Models

The models that provided the best fit for fluvial-fish species richness contained variables for drainage area, channel slope, total undammed reach length, percent buffer agriculture, percent buffer wetland, and percent forest or percent impervious cover. Species richness was negatively associated with percent impervious cover, percent buffer wetland, and channel slope. Species richness was positively associated with drainage area, percent forest, total undammed reach length, and percent buffer agriculture. The pseudo R^2 value for the fluvial richness model was 33.1 percent for the model that included percent impervious cover, and 33.3 percent for the model that included percent forest. Results of the models were significant at the 0.05 level when compared to results of the null model (intercept only). The correlation coefficients between observed and predicted values for the models, including percent impervious cover and percent forest, were 0.60 and 0.59, respectively. In addition, the intercept of the linear regression line between observed and predicted values was not significantly different from 0, and the slope was not different from 1 for both models, indicating that predictions from the fluvial-fish species-richness equations are relatively unbiased.

The coefficients for the fluvial-fish species-richness models (table 8A) can be used in equations 4 and 5 (below) to determine the relation between fluvial-fish species richness and percent impervious cover or the relation between

fluvial-fish species richness and percent forest, respectively. Keeping all other variables static, a unit (1 percent) increase in percent impervious cover is associated with a 5.4 percent decrease in fluvial-fish species richness, with the 95-percent confidence interval indicating the decrease is between 4.2 and 6.6 percent. Keeping all other variables static, a unit (1 percent) increase in percent forest is associated with a 1.9 percent increase in fluvial-fish species richness, with the 95-percent confidence interval indicating the increase is between 1.5 and 2.4 percent.

The associations between fluvial-fish species richness and percent impervious cover, and between fluvial-fish species richness and percent forest, are depicted in figure 19. On the graphs, fluvial-fish species richness is shown to decrease with increasing percent impervious cover and to increase with increasing percent forest. These relations were determined by applying the fluvial-fish species-richness equations, using the median values from the data set for the environmental factors in the equation (drainage area, channel slope, undammed reach length, and percent wetland) and varying the percent impervious cover from zero to 40 percent. The line on the plot showing percent impervious cover is not extended to 100 percent because the highest value for percent impervious cover in the data set was 40 percent. If the equations were calculated for specific locations, the relation between fluvial-fish species richness and percent impervious cover would differ slightly at each location, depending on the values for the variables in the equation.

Fluvial-fish richness (with IC)

$$E(Y) = e^{1.7911 + 0.0011 DA - 0.0968 CHSLP + 0.0008 UdamTmi - 0.0262 pBWet - 0.0557 IC}, \quad (4)$$

Fluvial-fish richness (with forest)

$$E(Y) = e^{-0.1231 + 0.0011 DA - 0.0862 CHSLP + 0.0006 UdamTmi + 0.0157 pBAg - 0.0147 pBWet + 0.0192 pFOR}, \quad (5)$$

where

- $E(Y)$ = mean number of fluvial species;
- e = Euler's number, the base of the natural logarithm;
- DA = contributing drainage area to fish-sampling site;
- $CHSLP$ = channel slope;
- $UdamTmi$ = total undammed reach length;
- $pBWet$ = percent wetland in a 240 meter buffer; and
- IC = percent impervious cover in the contributing drainage area;
- $pBAg$ = percent agriculture in a 240 meter buffer; and
- $pFOR$ = percent forest in the contributing drainage area.

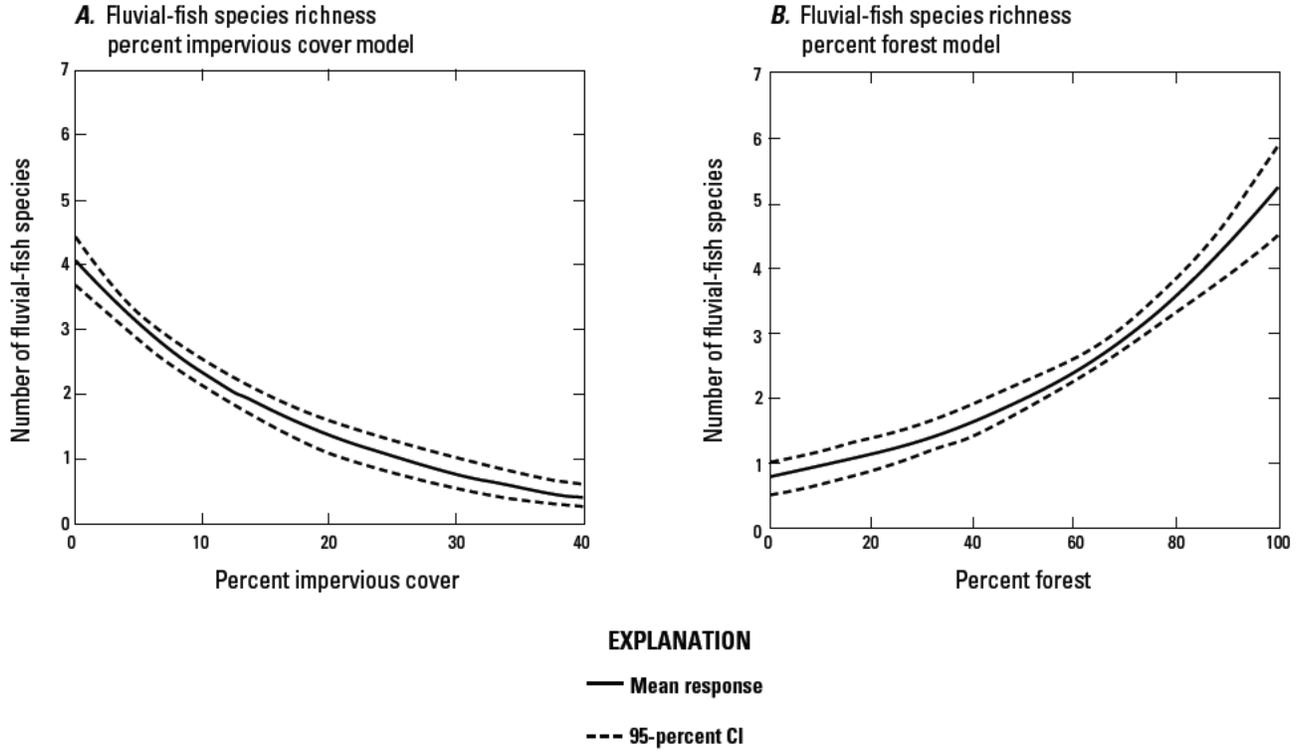


Figure 19. Generalized linear model output for relations between fluvial-fish species richness and A, percent impervious cover in the percent impervious cover model and B, percent forest in the percent forest model. Relations determined using median values for basin characteristics of 669 fish-sampling sites on Massachusetts streams. CI, confidence interval.

Fluvial-Fish Relative-Abundance Models

The models that provided the best fit for fluvial-fish relative abundance contained variables for channel slope, estimated percent alteration of median August groundwater withdrawals, percent buffer wetland, and percent impervious cover (equation 6, below) or percent forest (equation 7, below). Fluvial-fish relative abundance was negatively related to channel slope, estimated percent alteration of August median flow from groundwater withdrawals, percent buffer wetland, and percent impervious cover. Fluvial-fish relative abundance was positively associated with percent forest. The pseudo R^2 value was 18.2 percent for the fluvial abundance model that included percent impervious cover, and 18.4 percent for the model that included percent forest. Results of the models were significant at the 0.05 level when compared to results of the null model (intercept only). The correlation coefficients between observed and predicted values for the models including percent impervious cover and percent forest were 0.50 and 0.51, respectively. Plots of the model-predicted versus observed data indicated that the fluvial-fish relative-abundance equations slightly over-predict the relative abundance of fluvial fish for sites with high abundance.

The coefficients for the fluvial-fish relative abundance models (table 8A) can be used in equation 6 to determine the relations between fluvial-fish relative abundance and percent flow alteration from groundwater withdrawals and percent impervious cover, and in equation 7 for the relation between fluvial-fish relative abundance and percent-flow alteration from groundwater withdrawals and percent forest. Keeping

all other variables static, a unit (one percent) increase in percent impervious cover is associated with a 3.7 percent decrease in fluvial-fish relative abundance with the 95-percent confidence interval, indicating the decrease is between 1.1 and 6.1 percent. Keeping all other variables static, a unit (one percent) increase in percent forest is associated with a 1.3-percent increase in fluvial-fish relative abundance with the 95-percent confidence interval, indicating the increase is between 0.5 and 2.1 percent. A unit (one percent) increase in percent alteration of August median streamflow from groundwater withdrawals is associated with a 0.9-percent decrease in fluvial-fish relative abundance, with the 95-percent confidence interval indicating the decrease is between 0.1 and 1.7 percent.

For purposes of illustration, associations between fluvial-fish relative abundance, flow alteration indicators, percent impervious cover, and percent forest are depicted in fig. 20. On the plots, fluvial-fish relative abundance is shown to decrease with increases in the estimated percent alteration of August median flow from groundwater withdrawals, to decrease with increasing percent impervious cover, and to increase with increasing percent forest. Relations shown in these figures were determined with the fluvial-fish relative-abundance equations using the median values from the data set for the environmental factors and varying the percent flow alteration with either the percent impervious cover or the percent forest values. The ranges for these variables in the dataset were 0 to 100 percent for both the percent alteration of August median flow from groundwater withdrawals and for percent forest, and 0 to 40 percent for impervious cover.

Fluvial-fish relative-abundance model (with IC)

$$E(Y) = e^{6.1523 - 0.0840 CHSLP - 0.0091 AUGgwWp - 0.0289 pBWet - 0.0373 IC}, \quad (6)$$

Fluvial-fish relative-abundance model (with forest)

$$E(Y) = e^{4.9888 - 0.0878 CHSLP - 0.0091 AUGgwWp - 0.0227 pBWet + 0.0125 pFOR}, \quad (7)$$

where

- $E(Y)$ = mean fluvial-fish relative abundance, in counts per hour;
- e = Euler's number, the base of the natural logarithm;
- $CHSLP$ = channel slope;
- $AUGgwWp$ = percent alteration of August median flow from groundwater withdrawals;
- $pBWet$ = percent wetland in a 240 meter buffer;
- IC = percent impervious cover in the contributing drainage area; and
- $pFOR$ = percent forest in the contributing drainage area.

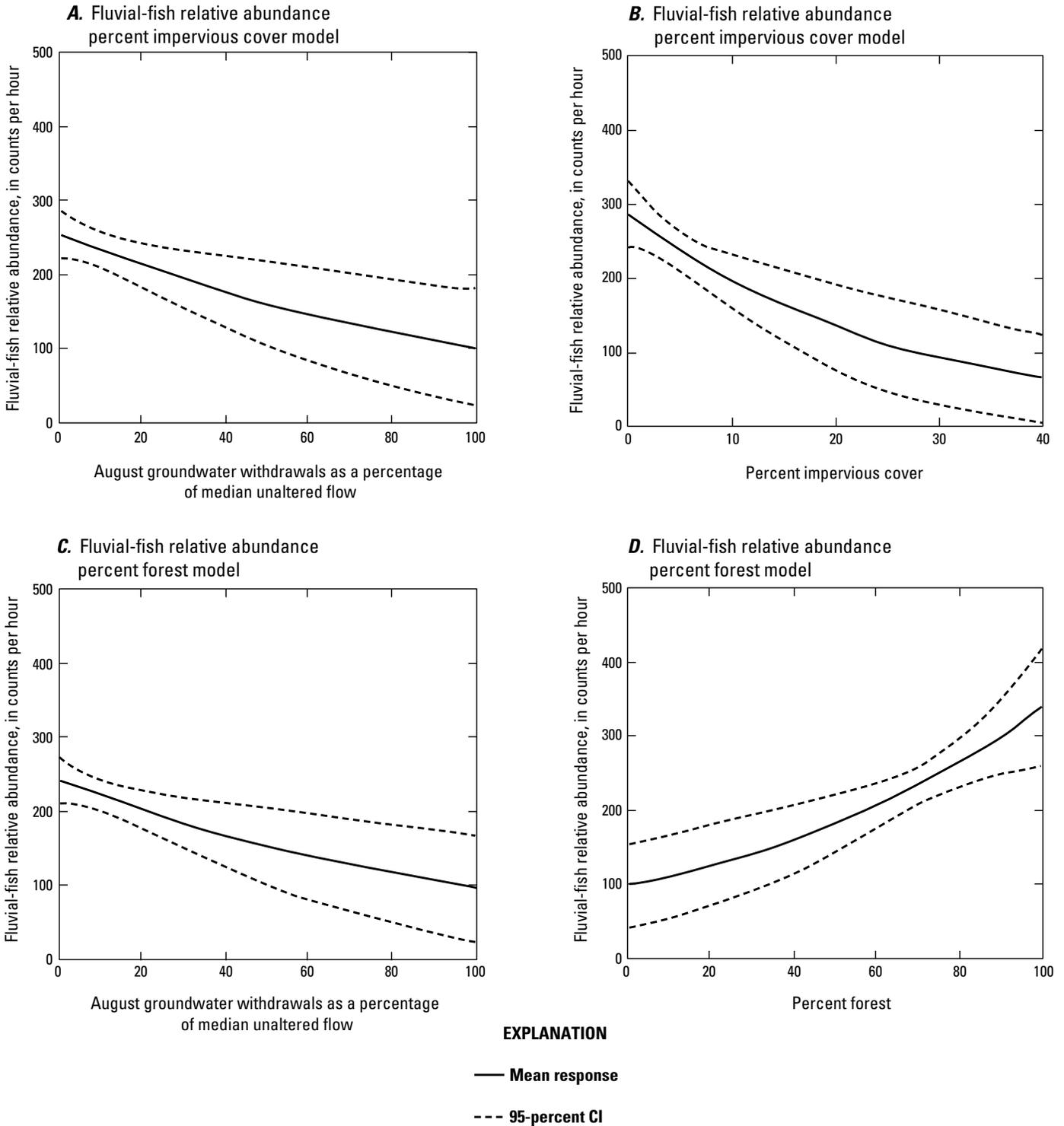


Figure 20. Generalized linear model output for relations between fluvial-fish relative abundance and A, percent alteration of August median flow from groundwater withdrawals in the percent impervious cover model, B, percent impervious cover in the percent impervious cover model, C, percent alteration of August median flow from groundwater withdrawals in the percent forest model, and D, percent impervious cover in the percent forest model. Relations determined using median values for basin characteristics of 669 fish-sampling sites on Massachusetts streams. CI, confidence interval.

Brook Trout Relative-Abundance Model

The models that provided the best fit for brook trout relative abundance contained variables for drainage area, percent open water, and percent impervious cover using equation 8 (below), and drainage area and percent forest using equation 9 (below). Brook trout relative abundance was negatively related to drainage area, percent open water, and percent impervious cover. Brook trout relative abundance was positively related to percent forest. The pseudo R^2 value was 34.7 for the brook trout model that included percent impervious cover and 34.6 for the model that included percent forest. The models were significant at the 0.05 level when compared to the null model (intercept only). The correlation coefficients between observed and predicted values for the models, including percent impervious cover and percent forest, were 0.50 and 0.52, respectively.

The coefficients for the brook trout relative abundance models (table 8A) can be used in equations 8 and 9 to determine the relation between brook trout abundance and

percent impervious cover, or the relation between brook trout abundance and percent forest, respectively. Keeping all other variables static, a unit (one percent) increase in percent impervious cover is associated with an 8.7 percent decrease in brook trout relative abundance with the 95-percent confidence interval, indicating the decrease is between 3.1 and 14.0 percent. Keeping all other variables static, a unit (one percent) increase in percent forest is associated with a 3.2-percent increase in brook trout relative abundance, with the 95-percent confidence interval indicating the increase is from 1.7 to 4.7 percent.

For purposes of illustration, associations between brook trout relative abundance and percent impervious cover and percent forest are depicted in figure 21. On the plots, brook trout relative abundance is shown to decrease with increases in percent impervious cover and to increase with increasing percent forest. Relations shown in these figures were determined using the brook trout relative-abundance equations with the median values from the data set for the environmental factors.

Brook trout relative abundance (with IC)

$$E(Y) = e^{4.9336 - 0.1291 DA - 0.2172 pCOW - 0.0916 IC}, \quad (8)$$

Brook trout relative abundance (with forest)

$$E(Y) = e^{2.0466 - 0.1380 DA + 0.0317 pFOR}, \quad (9)$$

where

- $E(Y)$ = mean brook trout relative abundance, in counts per hour;
- e = Euler's number, the base of the natural logarithm;
- DA = contributing drainage area to fish-sampling site;
- $pCOW$ = percent open water in the contributing drainage area;
- IC = percent impervious cover in the contributing drainage area; and
- $pFOR$ = percent forest in the contributing drainage area.

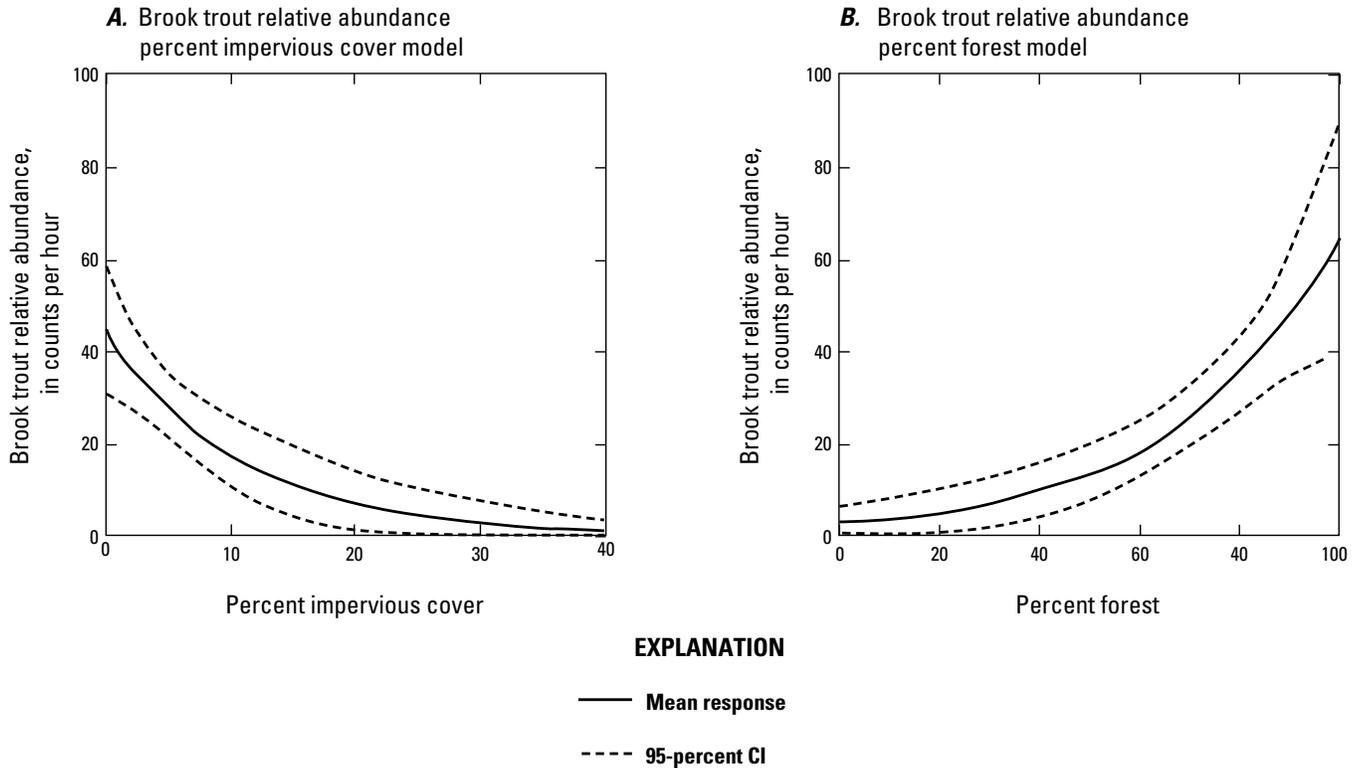


Figure 21. Generalized linear model output for relations between brook trout relative abundance and A, percent impervious cover in the percent impervious cover model and B, percent forest in the percent forest model. Relations determined using median values for basin characteristics of 669 fish-sampling sites on Massachusetts streams. CI, confidence interval.

Model Validation and Limitations

Pseudo R^2 values for the models ranged from 18.2 percent for the fluvial-fish relative-abundance model to 34.7 percent for the brook trout model. Correlation coefficients for observed versus predicted values of fluvial-fish relative abundance and fluvial-fish species richness ranged from 0.50 to 0.60. All three of the models were significant at the 0.05 level when compared to the null model (intercept only). The goodness of fit measures in table 8B indicate that the models in this study are within the range of reported values for similar ecological models (Pearce and Ferrier, 2001; Potts and Elith, 2006; Meador and Carlisle, 2009; Roy and others, 2009; Snelder and Lamouroux, 2010). Unexplained variability may be attributed to many causes, including unmeasured factors, the choice of environmental factors and anthropogenic covariates tested, location of sample sites relative to alterations, errors in determining environmental predictors from a GIS, fish-capture efficiency, and the use of modeled rather than measured flow data. The use of more site-specific habitat, stream temperature, or stream flow data might have improved model performance (Infante and Allan, 2010).

Several additional issues are addressed below to assist with interpretation of the models. The first is that fish-response variables do not necessarily drop to zero when the estimated August withdrawals reach 100 percent of the median August flow. The SYE-simulated median August median flow is a long-term flow statistic, calculated as the median of August median flows from WY1961 to 2004. Daily streamflows in August may be greater or less than the median on any given day, depending on the frequency of storm events, or they may be greater or less than the long-term median for much of the month, depending on whether it is a wet or dry summer. In the model, 100-percent depletion of August median flow may correspond to dry streambeds for times when the median (or less) flow was present in the stream channel. If flows are greater than the August median, water withdrawals would still leave some (reduced) streamflow. Some flow may also be provided by surface-water or septic return flows. In addition, fish-sample information comes from a single sample collected in a single day during a period, month, or year when the flow was not zero. Fish-assemblage sampling does not take place in dry streambeds, so one extreme end of the spectrum of

fish response to flow alteration can be missing from the data set for some sites. Although fluvial-fish relative abundance does not, on average, drop to zero for estimated 100-percent depletion of August median flow, the likelihood of zero values for fluvial fish does increase. For example, the 100 samples with the lowest percent alteration of August median flow from groundwater withdrawals in the data set included only 10 zero values for fluvial-fish relative abundance, whereas the 100 samples with the highest percent alteration of August median flow from groundwater withdrawals included 28 zero values.

An additional confounding factor is that median monthly groundwater withdrawals were calculated from daily withdrawals that were determined by disaggregating average annual water-use into daily withdrawals on the basis of an average demand curve developed for 25 communities (Archfield and others, 2010). Therefore, groundwater withdrawals are better used to compare the relative average depletion between sites than to indicate the flow conditions in the stream at the time the stream was sampled. In addition, surface-water reservoirs used for water-supply can alter streamflows through storage, diversions, and flow releases, and these flow alterations are not included in the simulated daily flows or estimates of monthly flow alterations. Although withdrawals from reservoirs for surface-water supply are estimated at the annual time scale for the percent alteration of annual mean flow statistic, this variable was not significant in the three best GLM models.

The GLM models for fluvial-fish species-richness and brook trout relative-abundance did not explicitly contain flow alteration variables. Although the absence of measures of flow alteration from these models does indicate that a strong relation between the estimated flow alteration measures and the mean values for these fish-assemblage metrics was not detected for this data set, it does not necessarily indicate that there is no relation between flow alteration and fluvial-fish species richness or brook trout relative abundance. Both equations contained an impervious cover variable, and increases in impervious cover have been demonstrated to be associated with flow alterations, such as increased flashiness; alterations of the frequency, magnitude, volume, and duration of stormwater runoff; and reductions in base flow. For some variables, the quantile regression plots indicate relations between fish assemblage measures and explanatory variables that were not depicted by the GLM equations. For example, figure 12A indicates a negative relation between relative abundance of brook trout and the estimated percent alteration of August median flow from groundwater withdrawals.

The GLM model predictions represent the mean response at a site given the specified values for the variables in the equation. Consequently, these models are best employed to compare fish-assemblage response among a set of sites or to look at the change in the mean response associated with a change in a particular covariate. The models are not meant to be used to predict responses at individual sites or extrapolated to define reference conditions. Particular caution is needed when extrapolating to the extremes of any one variable. For

example, because very few fish-assemblage measurements were made at sites having 0- or more than 20-percent impervious cover, the model is less robust at these ends of the impervious-cover spectrum.

Comparison of Study Results to Preliminary Analysis

The results of this analysis supersede those of the preliminary analysis (Armstrong and others, 2010). Although both the preliminary analysis and this study indicate significant relations between fish assemblage-characteristics and environmental and anthropogenic factors, the analyses in this report provide more statistical justification for variable selection without sacrificing explanatory power. Instead of relying on best professional judgment, a variable reduction process (using PCA and Spearman rank correlation) was used to select variables that accounted for the greatest proportion of variance while minimizing redundancy in the data set. An expanded list of environmental variables was tested for this analysis, including detailed land-use variables (in addition to simple impervious cover calculations), several variables representing dams, and a wider range of flow-alteration indicators. One result of the variable reduction process was that indicators of flow alteration representing the individual components of flow alteration were used in this study rather than the indicators representing net flow alteration that were used in the preliminary analysis. Multivariate analyses, including nonmetric multidimensional scaling (NMDS) and hierarchical cluster analysis, were used in this study to explore patterns among fish species and to guide selection of fish-assemblage metrics for use in subsequent analyses of relations between fish and environmental variables.

Both analyses used quantile regression to illustrate the upper limits of fish-response variables in relation to stressor gradients and to indicate the direction of that association, and both analyses used GLMs to quantify fish-assemblage response to environmental and anthropogenic factors. The GLM equations in this report and the preliminary report exhibit some differences. For example, both analyses included variables representing percent wetland and percent alteration of August median flow. However, the percent wetland values in the preliminary report were determined from the contributing areas to the fish-sampling sites, whereas the percent wetland values in this study were determined from a buffer area. The indicator of flow alteration used in the GLM in the preliminary report represented net percent alteration of August median flow, whereas the flow alteration indicator used in this study was percent alteration of August median flow from (public and private) groundwater withdrawals. Flow-alteration indicators representing other components of flow alteration, such as percent alteration of August median flow from surface-water returns or percent alteration of August median flow from septic returns, were analyzed and

were either too highly correlated with other (more highly loaded) variables to include in the GLM equations or were not significant variables in the equations.

The GLMs used in the preliminary report (Armstrong and others, 2010) were zero-inflated negative binomial models (ZINB), which are mixture models that use a negative binomial distribution in conjunction with a logistic regression process in order to model the excessive zeros present in the data set. The quasi-Poisson model used in this report is an alternative to the ZINB model for data sets with many zero values. When re-assessing the fish data for this study, it was determined that the quasi-Poisson framework could adequately account for the excessive zeros in the data set. The quasi-Poisson model has some advantages over the ZINB, including a wider variety of goodness of fit statistics and the capability to estimate confidence intervals of the mean response variable. In addition, the relatively simpler model structure is less prone to over-fitting of the data (including variables as significant that may not actually be significant) and is more easily interpreted. The ZINB and quasi-Poisson models resulted in similar relations between fish metrics, and environmental and anthropogenic factors. A comparison of the correlation between observed and predicted values for both model forms show that the quasi-Poisson model performed as well as or better than the ZINB model in comparing predicted versus observed values. Although the predictive ability of these models is not high, the relations characterized by the coefficients in the model are highly significant. In a model with relatively little bias, these relations can be used to examine how fish-response variables change along a gradient of values for a particular explanatory variable.

Possible Future Modeling Efforts

The GLM regression equations potentially could be applied to each of the 1,429 subbasins delineated in the Massachusetts Water Indicators project (Weiskel and others, 2010). Data on most variables used in the GLM equations are available from the Massachusetts Water Indicators project, with the exception of local channel slope and land-use variables in a buffer area, which could be calculated. Once the equations have been run with the existing conditions (current impervious cover and flow alteration), the subbasins could be grouped into broad categories that could be useful for assessing the effects of human activities—land use and water withdrawals—on aquatic habitat on a statewide basis.

The explanatory power of the GLM models potentially can be improved by identifying additional attributes to be tested while developing new equations. Local habitat features and water quality are known to explain part of the variability of fish assemblages (Gorham and Karr, 1978; Brown and others, 2009b), but data for these variables were not available on a statewide basis for this study. Additional habitat variables that could be tested include local habitat features, such as

substrate, depth, velocity, macrohabitat (pool, riffle, run), habitat complexity, and riparian vegetation. Water-quality variables that could be important include water chemistry, toxicity, nutrients, pesticides, herbicides, specific conductance, temperature, and dissolved oxygen.

Additional information on the effects of dams could also potentially improve the variance explained by the models. Dams alter many aspects of stream habitat that affect stream fish assemblages (temperature, habitat, barriers, flow), and a number of different metrics may be needed to reflect their impacts (Wang and others, 2010). Fish-community data from impounded reaches were intentionally omitted from this work. Comparisons of fish assemblages between impounded and non-impounded reaches could also be useful for developing new equations.

More detailed data for all water uses, such as water use at a daily time step, could prove to be highly beneficial for future modeling efforts. Data documenting the effects of flow releases and water withdrawals on streamflows downstream from reservoirs used for water supply were not available for this study. Several recent studies that specifically targeted the sampling of reaches downstream from reservoirs have documented the effects of water withdrawals and flow-releases from surface-water supply reservoirs on fish assemblages (Freeman and Marcinek, 2006; Kanno and Vokoun, 2010).

Further testing of the interactions of variables and of spatial covariation between variables may allow more variance to be explained when building regression models. For example, some aspects of flow alteration may be included in other variables, such as impervious cover and impoundments. Impervious cover has been demonstrated to be associated with increased flashiness, altering the frequency, magnitude, volume, and duration of stormflows, and reducing baseflow. Storage of streamflow by impoundments may reduce the short- and long-term flow variability of downstream reaches. Spatial relations that exist between natural environmental factors and land use, dams, and flow alterations complicate the task of assessing the effects of water withdrawals on fish at broad spatial scales (Borcard and others, 1992, 2004, 2011; Legendre, 1993, Borcard and Legendre, 1994). Environmental and land-use variables tend to co-vary spatially across Massachusetts. For example, the contributing areas to sample sites in western Massachusetts tend to have higher elevations, basin slopes, and percent forest, and lower percent of sand and gravel and wetland than sites in eastern Massachusetts. Some land-use variables, such as development and impervious cover, also exhibit an east-west gradient. Investigation of spatial structure among variables could potentially improve the regression models.

The diversity of the responses of individual fluvial-fish species to flow alteration and impervious cover highlights the importance of a species-by-species analysis in addition to the evaluation of fish-assemblage metrics. Although a number of species can be classified in a similar fashion (for example, fluvial), this does not necessarily indicate that each species will respond in the same way to all environmental

or anthropogenic factors. Additional tools like Threshold Indicator Taxa ANalysis (TITAN) (Baker and King, 2010; King and Baker, 2010) could be used to determine the nature of response of each species to various stressor gradients. TITAN analyses describe how individual species respond along a particular gradient (for example, impervious cover) and allow direct comparisons of the direction and extent of species responses to different stressor variables. An additional benefit of TITAN is that the method may indicate response thresholds that cannot be detected by the linear methods used in this analysis.

Although several significant relations between explanatory and response variables are described in this report, these relations do not definitively determine causation. For example, impervious cover, a convenient variable to measure, is associated with declines in fluvial-fish relative abundance, but the cause of the decline may be less conveniently measured variables, such as chemical contaminants and alterations to flow and temperature. The use of experimental designs and sampling methods specifically developed to increase model inference could improve predictions of the effects of human disturbances on fish assemblages (Downs and others, 2002; Downs, 2010).

Summary and Conclusions

Fish-sampling data collected from 1998 to 2008 at 669 Massachusetts Fisheries and Wildlife fish-sampling sites in streams and rivers were used to determine a set of fish-assemblage and indicator-species variables. A review of the literature was used to select a set of fish-assemblage variables thought to be responsive to flow alteration, along with a set of environmental and anthropogenic factors thought to be important for determining riverine fish abundance and distribution. Contributing drainage areas to each fish-sampling site were determined using a geographic information system (GIS) and were used to calculate more than 150 environmental and anthropogenic factors for each fish-sampling site to use for testing as explanatory variables in regression equations. Flow-alteration indicators used in the study were determined from streamflows simulated for each fish-sampling site using the USGS Sustainable Yield Estimator with water-use data and approaches developed by the Massachusetts Water Indicator study. Fish metrics tested include fish-assemblage variables, such as species richness and relative abundance of fluvial and macrohabitat-generalist species, and the relative abundance of individual species, such as brook trout, blacknose dace, common shiner, and white sucker.

A statistical process (using principal components analysis and Spearman rank correlation) was used to select 15 environmental and anthropogenic factors to test for use as explanatory variables in the regression analyses. These factors accounted for the greatest proportion of variance

while minimizing redundancy in the data set. Environmental factors selected through the analysis include drainage area, channel slope, percent sand and gravel, percent open water, percent forest, and percent wetland in a buffer area. Variables representing anthropogenic factors include impervious cover, estimated percent alteration of August median flow from groundwater withdrawals, estimated percent alteration of August median flow from surface-water returns, estimated percent alteration of annual mean flow from surface-water withdrawals, net percent alteration of mean annual flow for depleted sites, percent agriculture in a buffer, the undammed reach length upstream from the sampling site along the centerline, the total undammed reach length potentially available from the fish-sampling site, and dam density.

After removal of infrequently captured species from the analysis, a set of 21 fish species was used in a multivariate analysis to examine fish-assemblage patterns. Nonmetric multidimensional scaling and hierarchical cluster analysis indicate that fish species could be naturally grouped into fluvial and generalist habitat-use classes, justifying the use of fluvial-fish metrics as fish-response variables for the quantile regressions and generalized linear models (GLMs).

Bivariate scatterplots show that many relations between fish metrics and anthropogenic factors have a wedge shape, indicating a heterogeneous variance and a strong right skew. The declining upper edge of a wedge-shaped relation can indicate that an explanatory variable is acting as a constraint on organisms. Quantile regression, using 90th regression quantiles, indicated that percent impervious cover and the estimated percent alteration of August median flow from groundwater withdrawals are negatively associated with fluvial-fish relative abundance and fluvial-fish species richness, and that percent area of open water is negatively associated with fluvial-fish relative abundance. Quantile regression analyses using different indicator species demonstrated that the strength of these relations varied between species. The quantile regression plots indicate that (1) as many as seven fluvial-fish species are expected in streams with little flow alteration or impervious cover, (2) no more than four fluvial-fish species are expected in streams where estimated percent alteration of August median flow from groundwater withdrawals exceeds 50 percent or the percent area of impervious cover exceeds 15 percent, and (3) few fluvial fish remain at high rates of withdrawal (approaching 100 percent) or high rates of impervious cover (between 25 and 30 percent).

GLMs with a Poisson distribution corrected for overdispersion were used to relate a suite of explanatory variables to the selected fish-response variables. GLMs are an appropriate analytical tool for non-normally distributed data, count data, and data sets with large numbers of zero values. Numerous models were tested. The three strongest models are those for fluvial-fish species richness, fluvial-fish relative abundance, and brook trout relative abundance. The variables in these GLM equations were all demonstrated

to be significant ($p < 0.05$, with most < 0.01). Variables in the fluvial-fish species-richness model were drainage area, channel slope, total undammed reach length, percent wetland in a 240-meter buffer area, and percent impervious cover. Variables in the fluvial-fish relative-abundance model were channel slope, an indicator of percent flow alteration of August median flow from groundwater withdrawals, percent wetland in a 240-meter buffer area, and percent impervious cover. Variables in the brook trout relative-abundance model were drainage area, percent open water, and percent impervious cover. Correlation coefficients between observed and predicted values of the fitted models ranged from 0.50 to 0.60. The amount of variability in fish-response variables that could be attributed to environmental and anthropogenic factors by the equations was in line with those reported in the literature for similar studies.

Percent impervious cover and percent flow alteration were both determined to be significant variables influencing fish assemblages. Percent impervious cover was determined to be a significant variable for all three GLM equations. The indicator of percent alteration of August median flow from groundwater withdrawals was a significant variable for the fluvial-fish relative-abundance equation. The equations were used to quantify the change in fish-assemblage metrics associated with a unit change in these variables. The fluvial-fish species-richness equation indicated that, keeping all other variables the same, a unit increase in impervious cover is associated with a 5.4 percent decrease in fluvial-fish species richness, with the 95-percent confidence interval indicating the decrease is between 4.2 and 6.6 percent. Results of the fluvial-fish relative-abundance equation indicate that, keeping all other variables the same, a unit increase in the percent alteration of August median streamflow from groundwater withdrawals indicator is associated with a 0.9-percent decrease in relative abundance of fluvial fish, with the 95-percent confidence interval indicating the decrease is between 0.1 and 1.7 percent. Keeping all other variables the same, a unit increase in impervious cover is associated with a 3.7-percent decrease in fluvial-fish relative abundance, with the 95-percent confidence interval indicating the decrease is between 1.1 and 6.1 percent. Results of the brook trout relative-abundance equation indicate that an increase in impervious cover is associated with a decrease in relative abundance of brook trout of 8.7 percent, with the 95-percent confidence interval indicating the decrease is between 3.1 and 14.0 percent. The quantile regressions and GLM equations developed during this study illustrate statewide relations between fish-assemblage metrics and environmental and anthropogenic factors. The relations and equations can be used as tools to assess potential fish responses to different water- and land-management decisions for small- to medium-sized streams in Massachusetts.

References Cited

- Ahmadi-Nedushan, Behrouz, St-Hilaire, Andre, Berrube, Michel, Robichaud, Elaine, Thiemonge, Nathalie, and Bobee, Bernard, 2006, A review of statistical methods for the evaluation of aquatic habitat suitability for instream flow assessment: *River Research and Applications*, v. 22, p. 503–523.
- Anderson, M.J., 2001, A new method for non-parametric multivariate analysis of variance: *Austral Ecology*, v. 26, p. 32–46.
- Anderson, M.J., 2008, Animal-sediment relationships re-visited—Characterizing species distributions along an environmental gradient using canonical analysis and quantile regression splines: *Journal of Experimental Marine Biology and Ecology*, v. 366, p. 16–27.
- Anderson, M.J., Gorley, R.N., and Clarke, K.R., 2008, PERMANOVA+ for PRIMER – Guide to software and statistical methods: Plymouth, United Kingdom, PRIMER-E, 214 p.
- Angermeier, P.L., and Karr, J.L., 1986, Applying an index of biotic integrity based on stream-fish communities—considerations in sampling and interpretation: *American Journal of Fisheries Management*, v. 6, p. 418–429.
- Angermeier, P.L., and Smogler, R.A., 1995, Estimating number of species and relative abundances in stream-fish communities—Effects of sampling effort and discontinuous spatial distributions: *Canadian Journal of Fisheries and Aquatic Sciences*, v. 52, p. 936–949.
- Annear, T., Chisholm, I., Beecher, H., Locke, A., Aarestad, P., Coomer, C., Estes, C., Hunt, J., Jacobson, R., Jobsis, G., Kauffman, J., Marshall, J., Mayes, K., Smith, G., Wentworth, R., and Stalnaker, C., 2004, *Instream flows for riverine resource stewardship* (revised ed.): Cheyenne, Wyoming, Instream Flow Council, 268 p.
- Archfield, S.A., Vogel, R.M., Steeves, P.A., Brandt, S.L., Weiskel, P.K., and Garabedian, S.P., 2010, The Massachusetts Sustainable-Yield Estimator—A decision-support tool to assess water availability at ungaged stream locations in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2009–5227, 41 p.
- Argent, D.G., Bishop, J.A., Stauffer, J.R., Carline, R.F., and Myers W.L., 2003, Predicting freshwater fish distributions using landscape-level variables: *Fisheries Research*, v. 60, p. 17–32.

- Armstrong, D.S., Parker, G.W., and Richards, T.A., 2008, Characteristics and classification of least altered streamflows in Massachusetts: U.S. Geological Survey Scientific Investigations Report 2007–5291, 113 p.
- Armstrong, D.S., Parker, G.W., and Richards, T.A., 2004, Evaluation of streamflow requirements for habitat protection by comparison to streamflow characteristics at index streamflow-gaging stations in southern New England: U.S. Geological Survey Water-Resources Investigations Report 03–4332, 101 p.
- Armstrong, D.S., Richards, T.A., and Brandt, S.L., 2010, Preliminary assessment of factors influencing riverine fish communities in Massachusetts: U.S. Geological Survey Open-File Report 2010–1139, 43 p.
- Armstrong, D.S., Richards, T.A., and Parker, G.W., 2001, Assessment of habitat, fish communities, and streamflow requirements for habitat protection, Ipswich River, Massachusetts, 1998–99: U.S. Geological Survey Water-Resources Investigations Report 01–4161, 72 p.
- Arthington, A.A., Bunn, S.E., Poff, N.L., and Naiman, R.J., 2006, The challenge of providing environmental flow rules to sustain river ecosystems: *Ecological Applications*, v. 16, no. 4, p. 1311–1318.
- Austin, Mike, 2007, Species distribution models and ecological theory—a critical assessment and some possible new approaches: *Ecological Modelling*, v. 200, p. 1–19.
- Bain, M.B., Finn, J.T., and Booke, H.E., 1988, Streamflow regulation and fish community structure: *Ecology*, v. 69, no. 2, p. 382–392.
- Bain, M.B., and Meixler, M.S., 2000, Defining a target fish community for planning and evaluating enhancement of the Quinebaug River in Massachusetts and Connecticut: Ithaca, N.Y., New York Cooperative Fish and Wildlife Research Unit, Cornell University, 20 p.
- Bain, M.B., and Meixler, M.S., 2008, A target fish community to guide river restoration: *River Research and Applications*, v. 24, p. 453–458.
- Baker, M.E., and King, R.S., 2010, A new method for detecting and interpreting biodiversity and ecological community thresholds: *Methods in Ecology and Evolution*, v. 1, no. 1, p. 25–37.
- Bertrand, K.N., Gido, K.B., and Guy, C.S., 2006, An evaluation of single-pass versus multiple-pass backpack electrofishing to estimate trends in species abundance and richness in prairie streams: *Transactions of the Kansas Academy of Science*, v. 109, no. 3/4, p. 131–138.
- Bolker, B.M., 2008, *Ecological models and data in R*: Princeton, N.J., Princeton University Press, 396 p.
- Borcard, Daniel, and Legendre, Pierre, 1994, Environmental control and spatial structure in ecological communities—An example using oribatid mites (Acari, Oribatei): *Environmental and Ecological Statistics* v. 1, p. 37–61.
- Borcard, Daniel, Gillet, Francois, and Legendre, Pierre, 2011, *Numerical ecology with R*: New York, Springer, 306 p.
- Borcard, Daniel, Legendre, Pierre, Avois-Jacquet, Carol, and Tuomisto, Hanna, 2004, Dissecting the spatial structure of ecological data at multiple scales: *Ecology*, v. 85, no. 7, p. 1826–1832.
- Borcard, Daniel, and Legendre, Pierre, and Drapeau, Pierre, 1992, Partialling out the spatial component of ecological variation: *Ecology*, v. 72, p. 1045–1055.
- Brown, L.R., Cuffney, T.F., Coles, J.E., Fitzpatrick, Faith, McMahon, Gerard, Steuer, Jeffrey J., Bell, A.H., and May, J.T., 2009a, Urban streams across the USA—Lessons learned from studies in 9 metropolitan areas: *Journal of the North American Benthological Society*, v. 28, no. 4, p. 1051–1069.
- Brown, L.R., Gregory, M.B., and May, J.T., 2009b, Relation of urbanization to stream fish assemblages and species traits in nine metropolitan areas of the United States: *Urban Ecosystems*, v. 12, p. 391–416.
- Bunn, S.E., and Arthington, A.H., 2002, Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity: *Environmental Management*, v. 30, no. 4, p. 492–507.
- Burnham, K.P., and Anderson, D.R., 2002, *Model selection and multimodel inference—A practical information-theoretic approach*: New York, Springer-Verlag, 488 p.
- Bushaw-Newton, K.L., Hart, D.D., Pizzuto, K.E., Thomson, J.R., Egan, J., Ashley, J.T., Johnson, T.E., Horwitz, R.J., Keeley, M., Lawrence, J., Charles, D., Getenby, C., Kreeger, D.A., Nithengale, T., Thomas, R.L., and Velinsky, D.J., 2002, An integrative approach towards understanding ecological responses to dam removal—The Manatawny Creek study: *Journal of the American Water Resources Association*, v. 38, no. 6, p. 1581–1599.
- Cade, B.S., and Guo, Q., 2000, Estimating effects of constraints on plant performance with regression quantiles: *Oikos*, v. 91, p. 245–254.
- Cade, B.S., and Noon, B.R., 2003, A gentle introduction to quantile regression for ecologists: *Frontiers in Ecology and the Environment*, v. 1, no. 8, p. 412–420.
- Cade, B.S., Noon, B.R., and Flather, C.H., 2005, Quantile regression reveals hidden bias and uncertainty in habitat models: *Ecology*, v. 83, no. 3, p. 786–800.

- Cade, B.S., Terrell, J.T., and Schroeder, R.L., 1999, Estimating effects of limiting factors with regression quantiles: *Ecology*, v. 80, no. 1, p. 311–323.
- Cameron, A.C., and Trivedi, P.K., 1998, *Regression analysis of count data*: New York, Cambridge University Press, 411 p.
- Carlisle, D.M., Wolock, D.M., and Meador, M.R., 2010, Alteration of streamflow magnitudes and potential ecological consequences—A multiregional assessment: *Frontiers in Ecology and Environment*, 2010, accessed May 2, 2011, at http://water.usgs.gov/nawqa/pubs/Carlisleetal_FLowAlterationUS.pdf.
- Carter, Timothy, Jackson, C.R., Rosemond, Amy, Pringle, Cathy, Radcliffe, David, Tollner, William, Maez, John, Leigh, David, and Trice, Amy, 2009, Beyond the urban gradient—Barriers and opportunities for timely studies of urbanization effects on aquatic ecosystems: *Journal of the North American Benthological Society*, v. 28, no. 4, p. 1038–1050.
- Cereghino, R., Santoul, F., Compin, A., Figuerola, J., and Mastrorillo, S., 2005, Co-occurrence patterns of some small bodied freshwater fishes in southern France—Implications for fish conservation and environmental management: *Ambio*; v. 34, no. 6, p. 440–444.
- Chu, C., and Jones, N.E., 2010, Spatial variability of thermal regimes and other environmental determinants of stream fish communities in the Great Lakes Basin, Ontario, Canada: *River Research and Applications*, v. 27, p. 646–662.
- Clarke, K.R., 1993, Non-parametric multivariate analysis of changes in community structure: *Australian Journal of Ecology*, v. 18, p. 117–143.
- Clarke, K.R., and Gorley, R.N., 2006, *PRIMER v6—User manual/tutorial*: Plymouth, United Kingdom, PRIMER-E Ltd, 190 p.
- Clarke, K.R., and Warwick, R.M., 2001, *Change in marine communities—An approach to statistical analysis and interpretation*: Plymouth, United Kingdom, PRIMER-E Ltd, variously paged.
- Clarke, K.R., Somerfield, P.J., and Chapman, M.G., 2006, On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray-Curtis coefficient for denuded assemblages: *Journal of Experimental Marine Biology and Ecology*, v. 330, p. 55–80.
- Coles, J.F., Cuffney, T.F., McMahon, Gerard, and Beaulieu, K.M., 2004, The effects of urbanization on the biological, physical, and chemical characteristics of coastal New England streams: *U.S. Geological Survey Professional Paper 1695*, 47 p.
- Coles, J.F., Cuffney, T.F., McMahon, Gerard, and Rosiu, C.J., 2010, Judging a brook by its cover—The relation between ecological condition of a stream and urban land cover in New England: *Northeast Naturalist*, v. 17, no. 1, p. 29–48.
- Creque, S.M., Rutherford, E.S., and Zorn, T.G., 2005, Use of GIS-derived landscape-scale habitat features to explain spatial patterns of fish density in Michigan Rivers: *North American Journal of Fisheries Management*, v. 25, p. 1411–1425.
- Cuffney, T.F., Brightbill, R.A., May, J.T., and Waite, I.A., 2011, Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas: *Ecological Applications*, v. 20, no. 5, p. 1384–1401.
- Diana, Matt, Allan, J.D., Infante, Dana, 2006, The influence of physical habitat and land use on stream fish assemblages in Southeastern, Michigan: *American Fisheries Society Symposium*, v. 48, p. 359–374.
- Dobson, A.J., 2002, *An introduction to generalized linear models*: Boca Raton, Fla., Chapman and Hall, 225 p.
- Downs, B.J., 2010, Back to the future —Little-used tools and principles of scientific inference can help disentangle effects of multiple stressors on freshwater ecosystems: *Freshwater Biology*, v. 55, Supplement 1, p. 60–79.
- Downs, B.J., Barmuta, L.A., Fairweather, P.G., Faith, D.P., Lake, P.S., Mapstone, B.D., and Quinn, G.P., 2002, *Monitoring ecological impacts*: Cambridge, United Kingdom, Cambridge University Press, 434 p.
- Everitt, B.S., and Hothorn, Torsten, 2010, *A handbook of statistical analyses using R*: Boca Raton, Fla., CRC Press, 355 p.
- Foster, D.R., and Aber, J.D., 2004, *Forests in time—The environmental consequences of 1,000 years of change in New England*: New Haven, Conn., Yale University Press, 477 p.
- Freeman, M.C., and Marcinek, P.A., 2006, Fish assemblage responses to water withdrawals and water supply reservoirs in Piedmont streams: *Environmental Management*, v. 38, no. 3, p. 435–450.

- Gardner, C., Coghlan, S.M., Jr., Zydlewski, J., and Saunders, R., 2011, Distribution and abundance of stream fishes in relation to barriers—Implications for monitoring stream recovery after barrier removal: *River Research and Applications*, accessed September 12, 2011, at <http://onlinelibrary.wiley.com/doi/10.1002/rra.1572/pdf>.
- Gauch, H.G., Jr., 1994, *Multivariate analysis in community ecology*: New York, Cambridge University Press, 298 p.
- Gido, K.B., Falke, J.A., Oakes, R.M., and Haze, K.J., 2006, Fish-habitat relations across spatial scales in prairie streams: *American Fisheries Society Symposium*, v. 48, p. 265–285.
- Gillette, D.P., Tiemann, J.S., Edds, D.E., and Wildhaber, M.L., 2005, Spatiotemporal patterns of fish assemblage structure in a river impounded by low-head dams: *Copeia*, v. 3, p. 539–549.
- Goldstein, R.M., and Meador, M.R., 2004, Comparison of fish species traits from small streams to large rivers: *Transactions of the American Fisheries Society*, v. 133, p. 971–983.
- Gorham, O.T., and Karr, J.R., 1978, Habitat structure and stream fish communities: *Ecology*, v. 59, no. 3, p. 507–515.
- Graham, M.H., 2003, Confronting multicollinearity in ecological multiple regression: *Ecology*, v. 84, p. 2809–2815.
- Guisan, Antoine, Edwards, T.C., and Hastie, Trevor, 2002, Generalized linear and generalized additive models in studies of species distributions—Setting the scene: *Ecological Modelling*, v. 157, p. 89–100.
- Halliwell, D.B., Langdon, R.W., Daniels, R.A., Kurtenbach, J.P., and Jacobson, R.A., 1999, Classification of freshwater fish species of the northeastern United States for use in the development of indices of biological integrity, with regional applications, *in* Simon, T.P., ed., *Assessing the sustainability and biological integrity of water resources using fish communities*: New York, CRC Press, p. 301–335.
- Harrison, Blake, and Judd, R.W., eds., 2011, *A landscape history of New England*: Cambridge, Mass., the MIT Press, 413 p.
- Hartel, K.E., Halliwell, D.B., and Launer, A.E., 2002, *Inland fishes of Massachusetts*: Lincoln, Mass., Massachusetts Audubon Society, 328 p.
- Hilbe, J.M., 2011, *Negative binomial regression*: Cambridge, United Kingdom, Cambridge University Press, 553 p.
- Hilborn, Ray, and Mangel, Marc, 1997, *The ecological detective—Confronting models with data*: Princeton, N.J., 315 p.
- Hoeinghaus, D.J., Winemiller, K.O., and Birnbaum, J.S., 2007, Local and regional determinants of stream fish assemblage structure—Inferences based on taxonomic vs functional group: *Journal of Biogeography*, v. 34, p. 324–338.
- Holland, S.M., 2008, *Non-metric multidimensional scaling (MDS)*: Athens, Ga., University of Georgia, accessed April 27, 2011, at <http://www.uga.edu/strata/software/pdf/mdsTutorial.pdf>.
- Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments: *Ecological Monographs*, v. 54, p. 187–211.
- Infante, D.M., and Allan, J.D., 2010, Response of stream fish assemblages to local-scale habitat as influenced by landscape—A mechanistic investigation of stream fish assemblages, *in* Gido, K.B., and Jackson, D.A. eds., *Community ecology of stream fishes—Concepts, approaches, and techniques*: Bethesda, Md., American Fisheries Society, American Fisheries Society Symposium 73, Ottawa, Ontario, Canada, 19–20 August 2008, p. 371–397.
- Jackson, D.A., 1993, Stopping rules in principal components analysis—A comparison of heuristic and statistical approaches: *Ecology*, v. 74, no. 8, p. 2204–2214.
- James, F.C., and McCulloch, C.E., 1990, *Multivariate analysis in ecology and systematic—Panacea or Pandora's box*: *Annual Review of Ecology and Systematics*, v. 21, p. 129–166.
- Kanno, Yoichiro, and Vokoun, J.C., 2008, Biogeography of stream fishes in Connecticut—Defining faunal regions and assemblage types: *Northeastern Naturalist*, v. 15, no. 4, p. 557–576.
- Kanno, Yoichiro, and Vokoun, J.C., 2010, Evaluating effects of water withdrawals and impoundments on fish assemblages in southern New England streams, USA: *Fisheries Management and Ecology*, v. 17, no. 2, 12 p.
- Kanno, Y., Vokoun, J.C., and Beauchene, M. 2010, Development of dual fish multi-metric indices of biological condition for streams with characteristic thermal gradients and low species richness: *Ecological Indicators*, v. 10, p. 565–571.
- Karr, J.R., and Chu, E.W., 1999, *Restoring life in running waters*: Washington, D.C., Island Press, 206 p.
- Kashiwagi, Michael, and Richards, Todd, 2009, *Development of target fish-community models for Massachusetts mainstem rivers*: Boston, Commonwealth of Massachusetts Department of Fish and Game, Division of Fisheries and Wildlife Technical Report, 85 p.

- Kenkel, N.C., Derkson, D.A., Thomas, A.G., and Watson, P.R., 2002, Multivariate analysis in weed science research: *Weed Science*, v. 50, p. 281–292.
- Kennen, J.G., Chang, Ming, Tracy, B.H., 2005, Effects of landscape change on fish assemblage structure in a rapidly growing metropolitan area in North Carolina, USA: *American Fisheries Society Symposium*, v. 47, p. 39–52.
- Kennen, J.G., Kauffman, L.J., Ayers, M.A., Wolock, D.M., and Colarullo, S.J., 2008, Use of an integrated flow model to estimate ecologically relevant hydrologic characteristics at stream biomonitoring sites: *Ecological Modelling*, v. 211, p. 57–76.
- Kennen, J.G., Riva-Murray, K., and Beaulieu, K.M., 2009, Determining hydrologic factors that influence stream macroinvertebrate assemblages in the northeastern US: *Ecohydrology*, v. 2, no. 1, p. 88–106.
- Kery, Marc, 2010, *Introduction to WinBUGS for ecologists—A Bayesian approach to regression, anova, mixed models, and related analyses*: Burlington, Mass., Academic Press, 302 p.
- King, R.S., and Baker, M.E., 2010, Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients: *Journal of the North American Benthological Society*, v. 29, no. 3, p. 998–1008.
- Kinsolving, A.D., and Bain, M.B., 1993, Fish assemblage recovery along a riverine disturbance gradient: *Ecological Applications*, v. 3, no. 3, p. 531–544.
- Koenker, R., 2005, *Quantile regression*: New York, Cambridge University Press, 349 p.
- Koenker, R., and Machado, J.A.F., 1999, Goodness of fit and related inference processes for quantile regression: *Journal of the American Statistical Association*, v. 94, p. 1296–1310.
- Konrad, C.P., and Booth, D.B., 2005, Hydrologic changes in urban streams and their ecological significance, *in* Brown, L.R., Gray, R.H., Hughes, R.M., and Meador, M.R., eds., *Effects of urbanization on stream ecosystems*: Bethesda, Md., American Fisheries Society Symposium 47, Proceedings of the Symposium The Effects of Urbanization on Aquatic Ecosystems, Quebec City, Quebec, Canada, August 1–12, 2003, p. 157–177.
- Konrad, C.P., Brasher, A.M.D., and May, J.T., 2008, Assessing streamflow characteristics as limiting factors on benthic invertebrate assemblages in streams across the western United States: *Freshwater Biology*, v. 53, no. 10, p. 1917–2131.
- Legendre, Pierre, 1993, Spatial autocorrelation—Trouble or new paradigm?: *Ecology*, v. 74, p. 1659–1673.
- Legendre, Pierre, 2002, The consequences of spatial structure for the design and analysis of ecological field surveys: *Ecography*, v. 25, p. 601–615.
- Legendre, Pierre, and Anderson, M.J., 1999, Distance-based redundancy analysis—Testing multispecies responses in multifactorial ecological experiments: *Ecological Monographs*, v. 69, no. 1, p. 1–24.
- Legendre, Pierre, and Legendre, Louis, 1998, *Numerical ecology*: Amsterdam, The Netherlands, Elsevier Science, 853 p.
- Legros, J.D., and Parasiewicz, Piotr, 2007, Development and analysis of a target fish-community model to assess the biological integrity of the Lamprey Designated River, New Hampshire, and to identify indicator fish species for a MesoHABSIM model: Concord, N.H., New Hampshire Department of Environmental Services, accessed April 1, 2010, at http://des.nh.gov/organization/divisions/water/wmb/rivers/instream/lamprey/documents/tfc_report_legros_21june2007final.pdf.
- Leopold, L.B., Wolman, M.G., and Miller, J.P., 1995, *Fluvial processes in geomorphology*: Mineola, N.Y., Dover Publications, 522 p.
- Leps, Jan, and Smilauer, Peter, 2003, *Multivariate analysis of ecological data using CANOCO*: Cambridge, United Kingdom, Cambridge University Press, 269 p.
- Letcher, B.H., Nislow, K.H., Coombs, J.A., O'Donnell, M.J., and Dubreuil, T.J., 2007, Population response to habitat fragmentation in a stream-dwelling brook trout population, 2007, accessed September 12, 2011, at *PloS ONE* v. 2, no. 11 (Also available at e1139.doi:10.1371/journal.pone.0001139).
- Lewin, W., Freyhof, J., Volker, H., Mehner, T., Wolter, C., 2010, When no catches matter—Coping with zeros in environmental assessments: *Ecological Indicators*, v. 10, p. 572–583.
- Levin, S.B., Archfield, S.A., and Massey, A.J., 2011, Refinement and evaluation of the Massachusetts firm-yield estimator model version 2.0: U.S. Geological Survey Scientific Investigations Report 2011–5125, 49 p., plus CD-ROM.
- Mackey, P.C., Barlow, P.M., and Ries, K.G., III, 1998, Relations between discharge and wetted perimeter and other hydraulic-geometry characteristics at selected streamflow-gaging stations in Massachusetts: U.S. Geological Survey Water-Resources Investigations Report 98–4094, 44 p.

- Magalhaes, M.F., Batalha, D.C., and Collares-Pereira, M.J., 2002, Gradients in stream fish assemblages across a Mediterranean landscape—Contributions of environmental factors and spatial structure: *Freshwater Biology*, v. 47, p. 1015–1031.
- Maidment, D.R., 2002, *Arc Hydro, GIS for water resources*: Redlands, Calif., ESRI Press, 220 p.
- Mandrak, N.E., 1995, Biogeographic patterns of fish species richness in Ontario lakes in relation to historical and environmental factors: *Canadian Journal of Fisheries and Aquatic Sciences*, v. 52, p. 1463–1474.
- Marsh-Matthews, E., and Matthews, W.J., 2000, Geographic, terrestrial, and aquatic factors—Which most influence the structure of stream fish assemblages in the Midwestern United States?: *Ecology of Freshwater Fish*, v. 9, p. 9–21.
- Massachusetts Division of Fisheries and Wildlife, 2006, Commonwealth of Massachusetts Comprehensive Wildlife Conservation Strategy: Massachusetts Division of Fisheries and Wildlife, accessed June 1, 2011, at http://www.mass.gov/dfwele/dfw/habitat/cwcs/pdf/mass_cwcs_final.pdf.
- Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Low-Choy, S.J., Tyre, A.J., and Possingham, P., 2005, Zero tolerance ecology—Improving ecological inference by modeling the source of zero observations: *Ecological Letters*, v. 8, p. 1235–1246.
- Massachusetts Office of Geographic and Environmental Information, 2007, Impervious surface datalayer for Massachusetts: Massachusetts Office of Geographic and Environmental Information, accessed April 1, 2010, at http://www.mass.gov/mgis/impervious_surface.htm.
- Matthews, W.J., 1998, *Patterns in freshwater fish ecology*: New York, Chapman and Hall, 756 p.
- McArdle, B. H., and Anderson, M.J., 2001, Fitting multivariate models to community data—Comment on distance-based redundancy analysis: *Ecology*, v. 82, p. 290–297.
- McCullagh, P., and Nelder, J.A., 1989, *Generalized linear models* (2d ed.): London, Chapman and Hall, 532 p.
- McCune, Bruce, and Grace, J.B., 2002, *Analysis of ecological communities*: Glenden Beach, Oregon, MjM Software Design, 300 p.
- McGarigal, Kevin, Cushman, Sam, and Stafford, Susan, 2000, *Multivariate statistics for wildlife and ecology research*: New York, Springer Verlag, 283 p.
- Meador, M.R., and Carlisle, D.M., 2009, Predictive models for fish assemblages in eastern U.S. streams—Implications for assessing biodiversity: *Transactions of the American Fisheries Society*, v. 138, p. 725–740.
- Meador, M.R., and Carlisle, D.M., 2011, Relations between altered streamflow variability and fish assemblages in eastern USA streams: *River Research and Applications*, accessed May 31, 2011, at <http://onlinelibrary.wiley.com/doi/10.1002/rra.1534/pdf>.
- Meador, M.R., Carlisle, D.M., and Coles, J.F., 2008, Use of tolerance values to diagnose water-quality stressors to aquatic biota in New England streams: *Ecological Indicators*, v. 8, p. 718–728.
- Meador, M.R., and Carlisle, D.M., 2011, Relations between altered streamflow variability and fish assemblage in eastern USA streams: *River Research and Applications*, accessed August 1, 2011, at <http://onlinelibrary.wiley.com/doi/10.1002/rra.1534/pdf>.
- Meador, M.R., Coles, J.F., and Zappia, H., 2005, Fish assemblage responses to urban intensity gradients in contrasting metropolitan areas—Birmingham, Alabama, and Boston, Massachusetts, *in* Brown, L.R., Hughes R.M., Gray R., and Meador, M.R., eds., *Effects of urbanization on stream ecosystems*: Bethesda, Md., American Fisheries Society (AFS) Symposium 47, p. 409–423.
- Meador, M.R., McIntyre, J.P., and Pollack, K.H., 2003, Assessing the efficacy of single-pass backpack electrofishing to characterize fish community structure: *Transactions of the American Fisheries Society*, v. 132, p. 39–46.
- Meixler, Marci, 2011, Application of the target fish community model to an urban river system: *Journal of Environmental Management*, v. 92, p. 1138–1147.
- Motulsky, Harvey, 2010, *Intuitive biostatistics—A nonmathematical guide to statistical thinking*: New York, Oxford University Press, 447 p.
- Nislow, K.H., Hudy, Mark, Letcher, B.H., and Smith, E.P., 2011, Variation in local abundance and species richness of stream fishes in relation to dispersal barrier—Implications for management and conservation: *Freshwater Biology*, accessed August 1, 2011, at <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2427.2011.02634.x/pdf>.
- Olden, J.D., and Poff, N.L. 2003, Redundancy and the choice of hydrologic indices for characterizing streamflow regimes: *River Research and Applications*, v. 19, p. 101–121.
- Parasiewicz, Piotr, 2004, *Ecohydrology study of the Quinebaug River—Final report to the Project Management Committee and the New England Interstate Water Pollution Control Commission*: Ithaca, N.Y., Cornell University Department of Natural Resources, Instream Habitat Program and the NY Cooperative Fish and Wildlife Research Unit, 385 p.

- Parker, G.W., Armstrong, D.S., and Richards, T.A., 2004, Comparison of methods for determining streamflow requirements for aquatic habitat protection at selected sites on the Assabet and Charles Rivers, Eastern Massachusetts, 2000–2002: U.S. Geological Survey Scientific Investigations Report 2004–5092, 72 p.
- Paul, M.J., and Meyer, J.L., 2001, Streams in the urban landscape: *Annual Review of Ecology and Systematics*, v. 32, p. 333–365.
- Pearce, J., and Ferrier S., 2001, The practical value of modeling relative abundance of species for regional conservation planning: *Biological Conservation*, v. 98, p. 33–43.
- Peck, J.E., 2010, *Multivariate analysis for community ecologists—Step-by-step using PC-ORD*: Glenden Beach, Oreg., MjM Software Design, 162 p.
- Pickett, S.T.A., 1989, Space-for-time substitution as an alternative to long-term studies, *in* Likens, G.E., ed., *Long-term studies in ecology—Approaches and alternatives*: New York, Springer Verlag, p. 110–135.
- Poff, N.L., Allen, J.D., Bain, M.B., Karr, J.R., Prestagard, K.L., Richter, B.D., Sparks, R.E., and Stromberg, J.C., 1997, The natural flow regime—A paradigm for river conservation and restoration: *Bioscience*, v. 47, p. 769–784.
- Poff, N.L., and Hart, D.D., 2002, How dams vary and why it matters for the emerging science of dam removal: *Bioscience*, v. 52, no. 8, p. 659–668.
- Poff, N.L., Richter, B.D., Arthington, A.H., Bunn, S.E., Naiman, R.J., Kendy, E., Acreman, M., Apse, C., Bledsoe, B.P., Freeman, M., Henriksen, J., Jacobson, R.B., Kennen, J., Merritt, D.M., O’Keeffe, J., Olden, J.D., Rogers, K., Tharme, R.E., and Warner, A., 2010, The ecological limits of hydrologic alteration (ELOHA)—A new framework for developing regional environmental flow standards: *Freshwater Biology*, v. 55, p. 147–170.
- Poff, N.L., and Zimmerman, J.K.H., 2010, Ecological responses to altered flow regimes—A literature review to inform environmental flows science and management: *Freshwater Biology*, v. 55, p. 194–205.
- Potts, J.M., and Elith, J., 2006, Comparing species abundance models: *Ecological Modelling*, v. 199, p. 153–163.
- Power, G., Brown, R.S., and Imhof, J.G., 1999, *Groundwater and fish—Insights from northern North America: Hydrological Processes*, v. 13, p. 401–422.
- Price, A.L., and Peterson, J.T., 2010, Estimation and modeling of electrofishing capture efficiency for fishes in wadeable warmwater streams: *North American Journal of Fisheries Management*, v. 30, p. 481–498.
- Pritchett, T., and Pyron, M., 2011, Fish assemblages respond to habitat and hydrology in the Wabash River, Indiana: *River Research and Applications*, accessed August 1, 2011, at <http://onlinelibrary.wiley.com/doi/10.1002/rra.1528/pdf>.
- R Development Core Team, 2008, *R—A language and environment for statistical computing*: Vienna, Austria, R Foundation for Statistical Computing, ISBN 3–900051–07–0, accessed April 1, 2010, at <http://www.R-project.org>.
- Randall, A.D., 2001, Hydrogeologic framework of stratified-drift aquifers in the glaciated northeastern United States: U.S. Geological Survey Professional Paper 1415–B, 179 p.
- Richter, B.D., Baumgartner, J.V., Wigington, Robert, and Braun, D.P., 1997, How much water does a river need?: *Freshwater Biology*, v. 37, p. 231–249.
- Richter, B.D., Davis, M.M., Apse, C., and Konrad, C., 2011, A presumptive standard for environmental flow protection: *River Research and Applications*, accessed May 31, 2011, at <http://onlinelibrary.wiley.com/doi/10.1002/rra.1511/pdf>.
- Ries, K.G. III, 1997, August median streamflows in Massachusetts: U.S. Geological Survey Water-Resources Investigations Report 97–4190, 24 p.
- Ries, K.G., III, and Friesz, P.J., 2000, Methods for estimating low-flow statistics for Massachusetts streams: U.S. Geological Survey Water-Resources Investigations Report 00–4135, 81 p.
- Roy, A.H., Freeman, M.C., Freeman, B.J., Wenger, S.J., Ensign, W.E., and Meyer, J.L., 2005, Investigating hydrologic alteration as a mechanism of fish assemblage shifts in urbanizing streams: *Journal of the North American Benthological Society*, v. 24, p. 656–678.
- Roy, A.H., Purcell, A.H., Walsh, C.J., and Wenger, S.J., 2009, Urbanization and stream ecology—Five years later: *Journal of the North American Benthological Society*, v. 28, no. 4, p. 908–910.
- Scharf, F.S., Juanes, Francis, and Sutherland, Michael, 1998, Inferring ecological relationships from the edges of scatter diagrams—Comparison of regression techniques: *Ecology*, v. 79, p. 448–460.
- Schooley, R.L., and Wiens, J.A., 2005, Spatial ecology of cactus bugs—Area constraints and patch connectivity: *Ecology*, v. 86, p. 1627–1639.
- Shaw, P.J.A., 2003, *Multivariate statistics for the environmental sciences*: New York, Oxford University Press, 233 p.
- Sileshi, G., Hailu, G., and Nyadzi, G., 2009, Traditional occupancy-abundance models are inadequate for zero-inflated ecological count data: *Ecological Modelling*, v. 220, p. 1764–1775.

- Snelder, T.N., and Lamouroux, Nicolas, 2010, Co-variation of fish assemblages, flow regimes and other habitat factors in French rivers: *Freshwater Biology*, v. 55, p. 881–892.
- Tabachnick, B.G., and Fidell, L.S., 2001, *Using multivariate statistics*: Boston, Allyn and Bacon, 966 p.
- Thomson, J.D., Weiblen, G., Thomson, B.A., Alfaro, S., and Legendre, P., 1996, Untangling multiple factors in spatial distributions—Lilies, gophers, and rocks: *Ecology*, v. 77, p. 1698–1715.
- University of New Hampshire, University of Massachusetts, and Normandeau Associates, Inc., 2008, Development and analysis of target fish-community models to evaluate the status of the existing fish communities in the Upper and Lower Souhegan River, Appendix 6, *in* Souhegan River Protected Instream Flow Report: Concord, N.H., New Hampshire Department Of Environmental Services, NHDES–R–WD–06–50, accessed April 1, 2010, at <http://des.nh.gov/organization/divisions/water/wmb/rivers/instream/souhegan/study.htm>.
- U.S. Environmental Protection Agency, 2000, National water quality inventory—1998 Report: Washington, D.C., U.S. Environmental Protection Agency Report EPA–841–F–00–006, variously paged.
- U.S. Geological Survey, 1999, The national hydrography dataset: U.S. Geological Survey Fact Sheet 106–99, accessed March 2001 at <http://nhd.usgs.gov/>.
- U.S. Geological Survey, 2000, National land-cover data (NLCD) circa 1992—Completed nationwide September 2000: U.S. Geological Survey, accessed October 27, 2003, at <http://seamless.usgs.gov/>.
- VanSickle, John, and Johnson, C.B., 2008, Parametric distance weighting of landscape influence on streams: *Landscape Ecology*, v. 23, p. 427–438.
- Vaz, Sandrine, Martine, C.S., Eastwood, P.D., Ernande, Bruno, Carpentier, Andre, Meaden, G.J., and Coppin, Frank, 2008, Modelling species distributions using regression quantiles: *Journal of Applied Ecology*, v. 45, p. 204–217.
- Ver Hoef, J.M., and Boveng, P.L., 2007, Quasi-Poisson vs. negative binomial regression—How should we model overdispersed count data?: *Ecology*, v. 88, no. 1, p. 2766–2772.
- Vokoun, J.C., and Kanno, Youchiro, 2009, Evaluating effects of water withdrawals and impoundments on fish assemblages in Connecticut streams: Storrs, Conn., University of Connecticut, 31p.
- Waco, K.E., and Taylor, W.W., 2010, The influence of groundwater withdrawal and land use changes on brook charr (*Salvelinus fontinalis*) thermal habitat in two coldwater tributaries in Michigan, USA: *Hydrobiologia*, v. 650, p. 101–116.
- Walsh, C.J., Roy, A.H., Feminella, J.W., Cottingham, P.D., Groffman, P.M., and Morgan, R.P. II, 2005, The urban stream syndrome—Current knowledge and the search for a cure: *Journal of the North American Benthological Society*, v. 25, no. 3, p. 706–723.
- Wang, Lizhu, Lyons, John, and Kanehl, Paul, 2001a, Impacts of urbanization on stream habitat and fish across multiple spatial scales: *Environmental Management*, v. 28, no. 2, p. 255–266.
- Wang, Lizhu, Lyons, John, and Kanehl, Paul, 2001b, Impacts of urban land cover on trout streams in Wisconsin and Minnesota: *Transactions of the American Fisheries Society*, v. 132, p. 825–839.
- Wang, Lizhu, Infante, Dana, Lyons, John, Stewart, Jana, and Cooper, Arthur, 2011, Effects of dams in river networks on fish assemblages in non-impoundment sections of rivers in Michigan and Wisconsin, USA: *River Research and Applications*, v. 27, p. 473–487.
- Warton, D.I., Wright, S.T., and Wang, Yi., 2011, Distance-based multivariate analyses confound location and dispersion effects: *Methods in Ecology and Evolution*, accessed June 20, 2011, at <http://onlinelibrary.wiley.com/doi/10.1111/j.2041-210X.2011.00127.x/abstract>.
- Weiskel, P.K., Brandt, S.L., DeSimone, L.A., Ostiguy, L.J., and Archfield, S.A., 2010, Indicators of streamflow alteration, habitat fragmentation, impervious cover, and water quality for Massachusetts stream basins: U.S. Geological Survey Scientific Investigations Report 2009–5272, 79 p.
- Weiskel, P.K., Vogel, R.M., Steeves, P.A., Zarriello, P.J., DeSimone, L.A., and Ries, K.G., 2007, Water use regimes—Characterizing direct human interactions with hydrologic systems: *Water Resources Research*, v. 43, 11 p.
- Wenger, S.J., and Freeman, M.C., 2008, Estimating species occurrence, abundance, and detection probability using zero-inflated distributions: *Ecology*, v. 89, no. 10, p. 2953–2959.
- Wenger, S.J., Peterson, J.T., Freeman, M.C., Freeman, B.J., and Homans, D.D., 2008, Stream fish occurrence in response to impervious cover, historic land use, and hydrogeomorphic factors: *Canadian Journal of Fisheries and Aquatic Sciences*, v. 65, p. 1250–1264.

- Wenger, S.J., Roy, A.H., Jackson, C.R., Bernhardt, E.S., Carter, T.L., Filoso, Solange, Gibson, C.A., Hession, W.C., Kaushal, S.S., Marti, Euginia, Meyer, J.L., Palmer, M.A., Paul, M.J., Purcell, A.H., Ramirex, Alonso, Rosemond, A.D., Schofield, K.A., Sudduth, E.B., and Walsh, C.A., 2009, Twenty-six key research questions in urban stream ecology—An assessment of the state of the science: *Journal of the North American Benthological Society*, v. 28, no. 4, p. 1080–1098.
- Whitworth, W.R., 1996, Freshwater fishes of Connecticut: Connecticut Department of Environmental Protection Bulletin 114, 243 p.
- Wickham, J.D., Wu, Jianguo, and Bradford, D.F., 1997, A conceptual framework for selecting and analyzing stressor data to study species richness at large spatial scales: *Environmental Management*, v. 21, no. 2, p. 247–257.
- Wilding, T.K., and Poff, N.L., 2008, Flow ecology relationships for the watershed flow evaluation tool, accessed April 1, 2010, at <http://cwcb.state.co.us/NR/rdonlyres/3ECA858C-9305-4EA7-ACF8-7EE8FA41646E/0/WFETAppB.pdf>.
- Wilkinson, C.D., and Edds, D.R., 2001, Spatial pattern and environmental correlates of a Midwestern stream fish community—Including spatial autocorrelation as a factor in community analysis: *American Midland Naturalist*, v. 146, no. 2, p. 271–289.
- Wood, S.N., 2006, Generalized additive models—An introduction with R: New York, Chapman and Hall, 392 p.
- Yoder, C.O., Center for Applied Bioassessment and Biocriteria, Kulik, Audet, and Bagley, 2006, The spatial and relative abundance characteristics of the fish assemblages in three Maine rivers: Friends of Merrymeeting Bay, accessed May 2, 2011, at http://www.friendsofmerrymeetingbay.org/cybrary/FishWildlife/MAINE.RIVERS.REPORT_FINAL.09.01.2006.pdf.
- Yoder, C.O., Hersha, L.E., and Rankin, E.T., 2009, Fish assemblage and habitat assessment of the Presumpscot River: Columbus, Ohio, Midwest Biodiversity Institute, MBI Technical Report MBI.2008-12-6, accessed May 2, 2011, at http://www.cascobay.usm.maine.edu/pdfs/presumpscot%20_final_revised_report_20090731.pdf.
- Zarriello, P.J., Parker, G.W., Armstrong, A.S., and Carlson, C.S., 2010, Effects of water use and land use on streamflow and aquatic habitat in the Sudbury and Assabet River Basins, Massachusetts—Chapter 1, Simulated effects of water use, and projected water-use and land-use change on streamflow with a precipitation-runoff model; Chapter 2, Fish communities, stream temperature, and assessment of minimum streamflow targets for aquatic habitat at selected sites: U.S. Geological Survey Scientific Investigations Report 2010–5042, 160 p.
- Zorn, T.G., Seelbach, P.W., Rutherford, E.S., Wills, T.C., Cheng, S.-T., and Wiley, M.J., 2008, A regional-scale habitat suitability model to assess the effects of flow reduction on fish assemblages in Michigan streams: Ann Arbor, Mich., Michigan Department of Natural Resources, Fisheries Research Report 2089, 46 p.
- Zuur, A.F., Ieno, E.N., and Elphick, C.S., 2010, A protocol for data exploration to avoid common statistical problems: *Methods in Ecology and Evolution*, v. 1, p. 3–14.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A., and Smith, G.M., 2009, Mixed effects models and extensions in ecology with R: New York, Springer Science and Business Media, LLC, 574 p.
- Zuur, A.F., Ieno, E.N., Smith, G.M., 2007, Analyzing ecological data: New York, Springer Science and Business Media, LLC, 672 p.

THIS PAGE INTENTIONALLY LEFT BLANK

Prepared by the Pembroke and West Trenton Publishing
Service Centers.

For more information concerning this report, contact:

Director
U.S. Geological Survey
Massachusetts-Rhode Island Water Science Center
10 Bearfoot Road
Northborough, MA 01532
dc_ma@usgs.gov

or visit our Web site at:
<http://ma.water.usgs.gov>

