National Rivers and Streams Assessment 2008–2009 Technical Report

U.S. Environmental Protection Agency
Office of Wetlands, Oceans and Watersheds
Office of Research and Development
Washington, DC 20460

Original - March 2016
Revised - March 2024 Version 1.1
(Version History Page 3)

Acknowledgements

This report resulted from a groundbreaking collaboration to monitor and assess the nation's rivers and streams. The U.S. Environmental Protection Agency (EPA) Office of Water (OW) would like to thank the many participants who contributed to this important effort. Without the collaborative efforts and support from state and tribal environmental agencies, federal agencies, universities and other organizations, this significant assessment of the nation's rivers and streams would not have been possible.

EPA OW would like to thank the field crews, biologists, taxonomists, laboratory staff, data analysts, program administrators, EPA regional coordinators, statisticians, quality control staff, data management staff and many reviewers for their dedication and hard work. Your collective efforts made this report possible. To the many hundreds of participants, EPA expresses its profound thanks and gratitude.

State, Tribal, Territory and Interstate Partners

Alaska Department of Environmental Conservation
Arizona Game and Fish Department
Arkansas Department of Environmental Quality
Bad River Band of Lake Superior Chippewa Indians
Blackfeet Tribe
Bois Forte Band of Chippewa
California Department of Fish and Wildlife
California State Water Resources Control Board
Cheyenne River Sioux Tribe
Colorado Department of Public Health and Environment
Colorado Division of Wildlife
Colville Tribe
Confederated Salish & Kootenai Tribes
Connecticut Department of Energy and Environmental Protection
Delaware Department of Natural Resources and Environmental Control
Delaware River Basin Commission
Fort Peck Assiniboine and Sioux Tribes
Georgia Department of Natural Resources
Idaho Department of Environmental Quality
Illinois Environmental Protection Agency
Iowa Department of Natural Resources
Kansas Department of Health and Environment
Kentucky Division of Water
Leech Lake Band of Ojibwa
Louisiana Department of Environmental Quality
Maine Department of Environmental Protection
Maryland Department of the Environment
Maryland Department of Natural Resources
Menominee Indian Tribe of Wisconsin
Michigan Department of Environmental Quality
Minnesota Pollution Control Agency
Mississippi Department of Environmental Quality
Missouri Department of Conservation
Montana Department of Environmental Quality
Nevada Division of Environmental Protection
New England Interstate Water Pollution Control Commission
New Hampshire Department of Environmental Services
New Jersey Department of Environmental Protection
New Mexico Environment Department
New York Department of Environmental Conservation
Nez Perce Tribe
North Carolina Department of Water Quality
North Dakota Department of Health
Ohio Environmental Protection Agency
Ohio River Valley Water Sanitation Commission
Oklahoma Conservation Commission
Oklahoma Department of Environmental Quality
Oklahoma Water Resources Board
Oregon Department of Environmental Quality
Oregon Department of State Lands
Pennsylvania Department of Environmental Protection
Puerto Rico Department of Natural and Environmental Resources
Rhode Island Department of Environmental Management
South Carolina Department of Health and Environmental Control
South Dakota Department of Environment & Natural Resources
South Dakota Game, Fish and Parks
Susquehanna River Basin Commission
Tennessee Department of Environment and Conservation
Texas Commission on Environmental Quality
Texas Parks and Wildlife Department
Utah Division of Water Quality

Vermont Department of Environmental Conservation
Virginia Department of Environmental Quality
Washington State Department of Ecology
West Virginia Department of Environmental Protection
Wind River
Wisconsin Department of Natural Resources
Wyoming Department of Environmental Quality

Federal Partners

U.S. Department of Agriculture Forest Service
U.S. Department of Interior Bureau of Land Management
U.S. Department of Interior Fish and Wildlife Service
U.S. Department of Interior Geological Survey
U.S. Department of Interior National Park Service
U.S. EPA Office of Research and Development
U.S. EPA Office of Water
U.S. EPA Regions 1-10

Other Partners and Collaborators

Academy of Natural Sciences Philadelphia
Amnis Opes Institute
Arkansas Tech University
Central Plains Center for Bioassessment
Dynamac
EcoAnalysts
Great Lakes Environmental Center, Inc.
Michigan State University
Midwest Biodiversity Institute
Mississippi State University
Oregon State University
SRA International
Tetra Tech, Inc.
University of Arkansas
University of Iowa
University of Mississippi
Utah State University

Version History:
Version 1.1 - March 2024

Section 5.1 on Stressor, Relative and Attributable Risk was updated to correct errors in how the analytical notations (in particular including Probability notations inside of Table 1). Other text in this section was also revised and one additional reference was added.

Because of formatting issues in this older document, text on pages 32-34 of Section 5 was resized to ten point font to maintain page numbering from the original document.

The following people played a pivotal role and lent their expertise to the data oversight and analysis in this project: Ted Angradi, Karen Blocksom, Phil Kaufmann, Tom Kinkaid, Tony Olsen, Steve Paulsen, Dave Peck, John Stoddard, John Van Sickle, and Marc Weber from EPA Office of Research and Development; Richard Mitchell from EPA Office of Water; Daren Carlisle from U.S. Geological Survey; Alan Herlihy from Oregon State University; Jan Stevenson from Michigan State University; Chuck Hawkins from Utah State University; Monty Porter from Oklahoma Department of Environmental Quality; Larry Willis from Virginia Department of Environmental Quality; and Mike Miller from Wisconsin Department of Natural Resources.

The 2008/2009 National Rivers and Streams Assessment survey was led by Ellen Tarquinio with significant programmatic help from Treda Grayson, Susan Holdsworth, Sarah Lehmann, and Richard Mitchell from EPA Office of Water; Steve Paulsen from the EPA Office of Research and Development; and EPA Regional Monitoring Coordinators. The report was written by a team of contributors that included Steve Paulsen from EPA Office of Research and Development; and Susan Holdsworth, Sarah Lehmann, Alice Mayio, Richard Mitchell, and Ellen Tarquinio from EPA Office of Water.

10 "Fish Assemblage" was developed primarily by D. V. Peck (U.S. EPA Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division in Corvallis, OR). K. Blocksom (U.S. EPA Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division in Corvallis, OR) produced the random forest models, managed and updated the various components of the NRSA fish assemblage data, and assisted with reviewing data and computer code to ensure data quality. A. Herlihy (Oregon State University) developed the computer code to produce the candidate FMMIs and select the final FMMI from those candidates. John Van Sickle (U.S. EPA, retired) developed the initial R scripts for producing random forest models of metric response that were subsequently modified for use with the NRSA fish assemblage data. P. Kaufmann (U.S. EPA Office of Research and Development, National Health and Environmental Effects Research Laboratory, Western Ecology Division in Corvallis, OR) provided editorial comments that improved this chapter. The information in this chapter has been funded wholly or in part by the U.S. Environmental Protection Agency. It has been subjected to review by the EPA Office of Water and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

U.S. Environmental Protection Agency. Office of Water and Office of Research and Development. *National Rivers and Streams Assessment 2008-2009 Technical Report* (EPA/841/R-16/008). Washington, DC. March 2016. Updated march 2024. V1.1
http://www.epa.gov/national-aquatic-resource-surveys/nrsa

CONTENTS

Figures

Models

Tables

1 INTRODUCTION

The *National Rivers and Streams Assessment 2008–2009: A Collaborative Survey* ("main report") presents the general overview and results of an unprecedented sampling effort undertaken by the U.S. Environmental Protection Agency and its state and tribal partners. NRSA provides information on the ecological condition of the nation's rivers and streams and the key stressors that affect them, both on a national and an ecoregional scale. It also discusses change in water quality conditions in streams sampled for an earlier study, the Wadeable Streams Assessment (WSA) of 2004.

This document provides the technical details for the main report. This chapter describes the organization of this technical report and identifies major sources of information about the survey operations.

1.1 REPORT ORGANIZATION

The report is organized according to the type of information that is presented. The following sections describe the three main components of the report. Table 1.1 provides a list of abbreviations used throughout the report.

1.1.1 Chapters with Information Applicable to All Data Analyses

The first four chapters present information that applies to the data analyses described in the main report and further detailed in other chapters in this report. Chapters 2 and 3 describe the procedures used to select the probability and reference sites. Chapter 4 describes the findings from EPA's quality assurance throughout the survey. Chapter 5 presents an overview to the NRSA data analyses: the extent and risk assessments; the quantitative evaluations of biological data; and the change analyses used to compare the 2008-2009 findings to the Wadeable Streams Assessment and the 2004-2005 survey.

1.1.2 Chapters Describing Threshold Comparisons

Chapters 6, 7, and 8 describe the threshold evaluations for water chemistry, mercury in fish tissue, and *Enterococci*. In the fish evaluations, EPA compares the concentrations to different levels (thresholds) which, in most cases, were derived from human health concerns. For water chemistry, EPA used NRSA and other data to develop thresholds for its good, fair, and poor designations used in the main report. The chapters describe the thresholds for each indicator.

1.1.3 Chapters Describing Quantitative Biological Indices and Metrics

As stated earlier, Chapter 5 provides a brief overview to developing quantitative biological indices and metrics and observed to expected (O/E) modeling. Chapters 9, 10, and 11 provide a detailed discussion of the application of the quantitative approaches to evaluate benthic macroinvertebrates, fish community assemblage, and physical habitat.

Table 1.1 Abbreviations Used Throughout the Report

Abbreviation	Definition
ANC	Acid neutralizing capacity
CCE	Calibrator Cell Equivalent
CPL	Coastal Plain ecoregion
DII	Dam Influence index
DOC	Dissolved organic carbon
EMAP	EPA's Environmental Monitoring and Assessment Program
EPA	Environmental Protection Agency
FFG	Functional feeding group
FMMI	Fish Multimetric Index
HUC	Hydrologic Unit Codes
IBI	Index of Biotic Integrity
IQR	Interquartile range
km	kilometers
MAHA	Mid-Atlantic Highlands Assessment
MAIA	Mid-Atlantic Integrated Assessment
MMI	Multimetric Index
NAP	Northern Appalachians ecoregion
NAWQA	National Ambient Water Quality Assessment
NLCD	National Land Cover Dataset
NPL	Northern Plains ecoregion
NRSA	National Rivers and Streams Assessment 2008-2009
O/E	Ratio of Observed to Expected
OTU	Operational Taxonomic Unit
PCA	Principal Component Analysis
QA	Quality Assurance
QC	Quality Control
RBS	Relative Bed Stability
RF	Random Forest
RMSE	Root Mean Squared Error
S:N	Signal to Noise ratio
SAP	Southern Appalachians ecoregion
SD	Standard Deviation
SPL	Southern Plains ecoregion
TPL	Temperate Plains ecoregion
UMW	Upper Midwest ecoregion
WMT	Western Mountains ecoregion
WSA	Wadeable Streams Assessment 2000-2004
XER	Xeric Region ecoregion

1.2 ADDITIONAL RESOURCES DESCRIBING PROTOCOLS FOR SURVEY OPERATIONS

The survey developed a series of protocols to ensure consistency throughout the survey operations. The following four additional documents provide the field sampling methods, laboratory procedures, quality measures, and site selection for the 2008-2009 NRSA. Table 1.2 identifies the four documents:

Table 1.2 Four Documents with Protocols Used Throughout the Survey

U.S. EPA. 2007. *National Rivers and Streams Assessment: Field Operation Manual.* EPA/841/B-07/009. Washington, DC: U.S. Environmental Protection Agency.
U.S. EPA. 2007. *National Rivers and Streams Assessment: Laboratory Operations Methods Manual.* EPA/841/B-07/010. Washington, DC: U.S. Environmental Protection Agency.
U.S. EPA. 2007. *National Rivers and Streams Assessment: Quality Assurance Project Plan.* EPA/841/B-07/007. Washington,, DC: U.S. Environmental Protection Agency.
U.S. EPA. 2007. *National Rivers and Streams Assessment: Site Evaluation Guidelines.* EPA/841/B-07/008. Washington, DC: U.S. Environmental Protection Agency.

2 SELECTION OF PROBABILITY SITES

During the summers of 2008 and 2009, more than 85 field crews sampled 1,924 river and stream sites across the country representing nearly 1.2 million miles. Using standardized field methods, they sampled waters as large as the Mississippi River and as small as mountain headwater streams. Sites were selected using a random sampling technique that uses a probability-based design described in this chapter. The following sections describe the statistical objectives, target population, sample frame, survey design, evaluation, and statistical analysis.

2.1 OBJECTIVES

The statistical design requirements for NRSA 2008-2009 were to produce:

- Estimates of the 2008–2009 status of flowing waters nationally and regionally (nine aggregated Omernik ecoregions).

- Estimates of the 2008–2009 status of wadeable streams and non-wadeable rivers nationally and regionally (nine aggregated Omernik ecoregions).

- Estimates of the 2008–2009 status or urban flowing waters nationally.

- Estimates of the change in status in wadeable streams between 2008-2009 and 2004, nationally and regionally (nine aggregated Omernik ecoregions).

A secondary objective was to have each state sample approximately the same number of sites. All states sampled a minimum of 37 to 38 sites with some states opting to sample additional sites as part of enhancement studies.

2.2 TARGET POPULATION

The target population consists of all streams and rivers within the 48 contiguous states that have flowing water during the study index period, excluding portions of tidal rivers up to head of salt. The study index period extends from April/May to September and is generally characterized by low flow conditions. The target population includes the Great Rivers. Run-of-the-river ponds and pools are included while reservoirs are excluded.

2.3 SAMPLE FRAME

The sample frame was derived from the National Hydrography Dataset (NHD), in particular NHD-Plus. Attributes from NHD-Plus and additional attributes added to the sample frame that are used in the survey design include: (1) state, (2) EPA Region, (3) NAWQA Mega Region, (4) Omernik Ecoregion Level 3 (NACEC version), (4) WSA aggregated ecoregions (nine and three regions), (5) Strahler order, (6) Strahler order categories (1^{st}, 2^{nd}, …, 7^{th}, and 8^{th}+), (6) FCODE (defined below), (7) Urban, and (8) Frame07.

The version of NHD-Plus used includes two separate Strahler order calculations, one of which is included in the publicly available NHD-Plus version. The other Strahler order calculation (SO attribute name) more accurately reflects the true Strahler order and is used for the survey design. The StrahCat attribute collapses 8th, 9th, and 10th order rivers into a single category.

The Urban attribute was created by intersecting a modified version of the Census Bureau national urban boundary GIS coverage with NHD-Plus. The Census Bureau's boundaries were buffered 100 meters to include a majority of stream features intersecting and coincident with urban areas. Where this buffer did not completely gather all the river features within the urban areas (rivers intersecting cities are excluded from the Census Bureau's urban areas), the NHD-Plus river area (polygon) features were clipped at a 3-kilometer buffer around the urban areas and combined with the buffered urban area to create the modified urban database. If a stream or river segment was within this boundary, it was designated as "Urban"; otherwise it was "NonUrban."

FCODE came directly from NHD-Plus and was used to identify which segments in NHD were included in the sample frame. The attribute Frame07 identified each segment as either "Include" or "Exclude." Frame07 was created so that segments included in the sample frame could be easily identified. FCODE values included in the GIS shapefile:

Included in FW08 sample frame (Frame07 = "Include"):

33400	Connector
33600	Canal/Ditch
42801	Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = At or Near
46000	Stream/River
46003	Stream/River (Intermittent)
46006	Stream/River (Perennial)
58000	Artificial Path (removed from dataset if coded through Lake/Pond and Reservoirs)

Excluded in FW08 sample frame (Frame07 = "Exclude")

42800	Pipeline
42802	Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Elevated
42803	Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Underground
42804	Pipeline: Pipeline Type = Aqueduct; Relationship to Surface = Underwater
42806	Pipeline: Pipeline Type = General Case; Relationship to Surface = Elevated
4280	Pipeline: Pipeline Type = General Case; Relationship to Surface = Underground
42809	Pipeline: Pipeline Type = Penstock; Relationship to Surface = At or Near
42811	Pipeline: Pipeline Type = Penstock; Relationship to Surface = Underground
42813	Pipeline: Pipeline Type = Siphon
56600	Coastline

Rivers with a Strahler order greater than or equal to 5th order that had FCODE equal to 46003 (intermittent) were included in the FW08 sample frame for all states west of 96 degrees longitude (North Dakota to Texas and states west). This was done to ensure that all large rivers in the more arid west were included regardless of NHD-Plus intermittent code.

2.4 SURVEY DESIGN

The survey design consisted of two major components in order to address the dual objectives of estimating:

(1) Current status for all flowing waters; and
(2) Change in status for wadeable streams from the 2004 Wadeable Streams Assessment.

These two components were the NRSA design and WSA_Revisit design (*i.e.*, sites from the Wadeable Streams Assessment were selected to be sampled during NRSA). A Generalized Random Tessellation Stratified (GRTS) survey design for a linear resource was used for the NRSA design and a GRTS survey design for a finite resource was used for the WSA_Revisit design. The design includes reverse hierarchical ordering of the selected sites.

2.4.1 Stratification

The survey design was explicitly stratified by state for the NRSA design. The original WSA design had several strata (EMAP West, New England, Virginia, Iowa, and remaining eastern states combined). The WSA_Revisit design ignored these strata in the selection of the subset of sites from the WSA to be revisited as part of the current NRSA design.

2.4.2 Multi-Density Categories

A complex unequal probability selection process was used in each of the two components of the survey design. They are described separately in the following sections.

2.4.2.1 NRSA Survey Design

Unequal probability categories were defined separately for wadeable streams (1st to 4th order) and non-wadeable rivers (5th to 10th order). "Wadeable" and "Non-Wadeable" were used to designate Strahler order classes and not to imply that the streams actually would be wadeable or non-wadeable. The expected sample size was 450 for wadeable streams and 900 for non-wadeable rivers.

For the wadeable stream category, within each state, unequal selection probabilities were defined for 1st, 2nd, 3rd, and 4th order streams so that an equal number of sites would occur for each order. Then these unequal selection probabilities were adjusted by WSA nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category.
For the non-wadeable river category, unequal selection probabilities were defined for 5th, 6th, 7th, and 8th+ order rivers so that the expected number of sites would be 350, 275, 175, and 100 sites, respectively. Then these unequal selection probabilities were adjusted by WSA nine aggregated ecoregion categories so that an equal number of sites would occur in each WSA nine aggregated ecoregion category.

Given these initial selection probabilities, the expected number of urban and non-urban sites was calculated to determine if at least 150 urban sites would be selected. Over 150 urban sites were expected, so no additional adjustment was required to satisfy the urban design requirement.

The final adjustment of the selection probabilities was to adjust them to minimize the range in the number of sites across the 48 states while still meeting the other design requirements. Given a total of 1,350 sites for the NRSA design, an objective was to assign each state 28 sites. This could not be achieved, although the range was able to be decreased.

2.4.2.2 WSA_Revisit Design

WSA sampled 1,390 sites between 2000 and 2004. To estimate change, 450 of these sites were visited again as part of the 2008–2009 Rivers and Streams assessment. The revisit design selected the 450 sites using unequal selection probabilities. Initially, all sites were assigned an equal selection probability of 1.

First, four intensification study regions were sampled as part of the WSA. These regions were the Wenatchee Watershed in Washington; Lower John Day and Deschutes watersheds in Oregon; Northern California coastal watersheds; and southern California coastal. The survey design gave the expected number of sites within a study region as if a state-wide survey design was done without intensification.

Second, the density of sites sampled for the EMAP-West portion of the WSA was greater than for the 36 eastern states. The selection probabilities were reduced for EMAP-West states to adjust for this. The density of sites in the Southern Appalachian aggregated ecoregion was less than in other eastern aggregated ecoregions as a result of the site replacement process used in the WSA. The selection probabilities were increased for these sites as well. The latter also ensured that the final weights for these sites were not extreme.

Third, the selection probabilities developed above were adjusted to achieve approximately an equal number of sites across all nine WSA aggregated ecoregions.

Fourth, the overall weight, inverse of selection probability, was calculated by multiplying the original WSA weight by the inverse of the above selection probability. This accounts for the fact that the WSA_Revisit design is a two-stage sample of wadeable streams.

WSA_Revisit design weights and NRSA design weights associated with wadeable streams were later adjusted to account for the fact that they are two independent survey designs of wadeable streams for the 48 states. This was done after the sites were evaluated and sampled.

2.4.2.3 State Designs

For any state that had a current, compatible state-wide probability design that covered all flowing waters, an option was provided to use its sites instead of the flowing water design sites. For the option to be exercised for a state, (1) the state's design must be a probability survey design, (2) its target population of streams and rivers must include the target population for the NRSA, (3) its sample frame must include the NRSA sample frame, and (4) its design must be implemented statewide in 2008–2009. The state must also agree to measure all NRSA indicators using the national field and laboratory protocols.

2.4.3 Oversample

No oversample sites were selected for the WSA_Revisit design. The expectation is that all, or almost all, of the 450 sites selected will be sampled given that they were sampled previously. For the NRSA design, the oversample is nine times the expected sample size within each state. The large oversample size was used to accommodate states that may want to increase the number of sites sampled within their state for a state-level design.

2.4.4 Site Use

Each stream/river selected to be sampled was given a unique site identification (siteID) with two parts: (1) NFW08 that identifies the sites as part of the 2008-9 National Rivers and Streams Assessment and (2) the two-letter state FIPS code followed by a number between 001 and 999 within each state. It was critical that this siteID be used in its entirety to make sure that the stream and river sites were correctly identified.

Sites were organized to be used within each state. If evaluation determined that a stream or river site cannot be sampled, it was replaced by another site within the state. Sites that were coded as 1st, 2nd, 3rd, and 4th were to be replaced by over sample sites that were coded 1st, 2nd, 3rd, or 4th, ignoring order within this range. For example, a 2nd order would be replaced by a 1st, 2nd, 3rd, or 4th order stream. Sites that are coded as 5th, 6th, 7th, 8th, 9th, or 10th order were to be replaced by oversample sites that are coded 5th, 6th, 7th, 8th, 9th, or 10th order, ignoring order within this range. For example, a 5th order river would be replaced by a 5th, 6th, 7th, 8th, 9th, or 10th order river. In each case the next lowest siteID that is within the Strahler order set was used for the replacement.

2.5 Evaluation Process

The survey design weights in the design file assumed that the survey design was implemented as designed. Typically, users prefer to replace sites that cannot be sampled with other sites to achieve the sample size planned. The site replacement process was described above. When sites were replaced, the survey design weights were no longer correct and had to be adjusted. The weight adjustment required knowing what happened to each site in the base design and the oversample sites. EvalStatus (evaluation status) was initially set to "NotEval" to indicate that the site had yet to be evaluated for sampling. When a site was evaluated for sampling, then the EvalStatus for the site was changed. Recommended codes are provided in Table 2.1.

Table 2.1 Recommended Codes for Evaluating Sites

EvalStatus Code	Name	Meaning
TS	Target Sampled	Site was a member of the target population and was sampled
LD	Landowner Denial	Landowner denied access to the site
PB	Physical Barrier	Physical barrier prevented access to the site
NT	Non-Target	Site was not a member of the target population
NN	Not Needed	Site was a member of the oversample and was not evaluated for sampling
Other codes		Other codes were often useful. For example, rather than use NT, the status may include specific codes indicating why the site was non-target.

2.6 Statistical Analysis

Any statistical analysis of the data must incorporate information about the monitoring survey design. In particular, when estimates of characteristics for the entire target population are computed, the statistical analysis must account for any stratification or unequal probability selection in the design. Procedures for doing this are available from the Aquatic Resource Monitoring Web page (http://archive.epa.gov/nheerl/arm/web/html/index.html). A statistical analysis library of functions is available from the Web page to do common population estimates in the statistical software environment R.

2.7 LITERATURE CITED

Diaz-Ramos, S., D. L. Stevens, Jr, and A. R. Olsen. 1996. *EMAP Statistical Methods Manual*. EPA/620/R-96/002, U.S. Environmental Protection Agency, Office of Research and Development, NHEERL-Western Ecology Division, Corvallis, Oregon.

Horn, C.R. and Grayman, W.M. (1993) Water-quality modeling with EPA reach file system. *Journal of Water Resources Planning and Management,* 119, 262-74.

Stevens, D.L., Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. *Environmetrics*, 8:167-95.

Stevens, D.L., Jr. and Olsen, A.R. 1999. Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics*, 4:415-428

Stevens, D. L., Jr., and A. R. Olsen. 2003. Variance estimation for spatially balanced samples of environmental resources. *Environmetrics* 14:593-610.

Stevens, D. L., Jr., and A. R. Olsen. 2004. Spatially-balanced sampling of natural resources in the presence of frame imperfections. *Journal of American Statistical Association*: 99:262-278.

Strahler, A.N. 1957. Quantitative Analysis of Watershed Geomorphology. *Trans. Am. Geophys.* Un. 38, 913-920.

3 METHODOLOGY FOR SELECTION OF REFERENCE SITES

To assess current ecological condition, it is necessary to compare measurements today to an estimate of expected measurements in a less-disturbed situation. Because of the difficulty in establishing pristine conditions for many indicators, NRSA used "Least-Disturbed Condition" as the reference condition. Least-Disturbed Condition can be defined as the best available chemical, physical and biological habitat conditions given the current state of the landscape. NRSA reference thresholds describe the sites whose condition is "the best of what's left." Data from reference sites were used to select metrics for indices of biological integrity (IBIs), develop Observed to Expected ratio (O/E) models, and define the ecoregion-specific condition class thresholds. This chapter describes the methodology used to select the reference site by identifying the sources of reference sites; the screening of the data sources for reference condition; and the final screening for reference sites.

3.1 SOURCES OF REFERENCE SITES

The fish and macroinvertebrate reference sites used in the NRSA came from four major activities:

1. Sites sampled during the NRSA using consistent sampling protocols and analytical methods that were screened to meet ecoregion-specific physical and chemical criteria. These included both sites selected randomly from the probability sample and sites hand-picked by best professional judgment and sampled using NRSA methods as part of the NRSA. Sites sampled as hand-picked, targeted reference sites for the NRSA were identified as reference via a three tiered approach. First, sites throughout the country that were submitted as least-disturbed by states, academics, USGS, and EPA Regions were screened using a quantitative disturbance score for the local watershed (the area draining to the reach segment). Sites were then sent to the EPA Landscape Ecology Lab for a quantitative disturbance score for the cumulative watershed (includes the reach and all upstream reaches). Finally, the top 300 sites that ranked using a visual assessment of disturbance at the 1:24,000 and 1:3,000 scales. 200 sites were selected that covered the nine ecoregions and two resource types and ranked high across all screens.

2. In addition to the sites sampled in the NRSA, as part of the NRSA data analysis process, we obtained possible reference site external data from USGS's National Water-Quality Assessment Program (NAWQA), EPA Region 7, the State of Wisconsin, and the State of Oklahoma. These data included fish and macroinvertebrate assemblage data as well as physical and chemical habitat data.

3. Benthic reference site data from 1,655 wadeable stream sites were available from the 2006 EPA Wadeable Streams Assessment (WSA). In the WSA, reference sites were obtained from two different approaches: first by screening the WSA survey data for physical and chemical criteria in the same manner described in #1 above, and second from macroinvertebrate sample data provided by other agencies, universities, or states from sites that were deemed to be suitable as reference sites by best professional judgment. These sites either were sampled with the same methodology as the WSA or had field and lab protocols with enough similarities that the data analysis group determined that the data were comparable. The reference sites from this second approach were only used in developing a MMI for benthic samples, not for setting any thresholds. The WSA reference site screening process and data sources are described in detail in

Herlihy *et al.* (2008). In Table 3.1, the first two data columns summarize the number of available WSA macroinvertebrate reference sites by ecoregion.

4. Fish reference site data from stream and river sites used by Herlihy *et al.* (2006) in a national analysis of fish assemblage data. The screening process used to define reference sites is described in Herlihy *et al.* (2006) and defined in detail in Appendix 1 of that document. The Herlihy *et al.* (2006) study only used the first two years of data from EMAP-West. The last three years of the data from EMAP-West was also available so that reference fish data was used as well. Final numbers of reference sites and screening used to refine the fish reference population are outlined in 10.

Table 3.1 Macroinvertebrate reference sites available for use in the NRSA

	WSA Activities		NRSA Activities		
Ecoregion	WSA—External	WSA—Screened	NRSA—External	NRSA—Screened	Total
Northern Appalachians (NAP)	114	27	2	37	180
Southern Appalachians (SAP)	370	35	22	38	465
Coastal Plain (CPL)	112	15	3	46	176
Upper Midwest (UMW)	68	12	38	30	148
Temperate Plains (TPL)	124	38	50	22	234
Northern Plains (NPL)	10	18	3	47	78
Southern Plains (SPL)	56	21	51	34	162
Western Mountains (WMT)	335	129	4	40	508
Xeric Region (XER)	132	39	2	33	206
Total	1,321	334	175	327	2,157

3.2 Screening NRSA Data for Reference Condition

To identify reference sites by screening the NRSA data, we used the chemical and physical data collected at each site (*e.g.*, nutrients, turbidity, acidity, riparian condition) to determine whether any given site is in least-disturbed condition for its ecoregion. In the NRSA, eight physical and chemical parameters were used to screen for reference sites, total N, total P, chloride, sulfate, acid neutralizing capacity, turbidity, % fine substrate, and riparian disturbance index. If a site exceeded the screening value for any one stressor it was dropped from reference consideration.

Given that expectations of least-disturbed condition vary across ecoregions, the criteria values for exclusion varied by ecoregion. The nine aggregate level III ecoregions developed for the WSA assessment were used to regionalize reference conditions. Ecoregional specific screening criteria are listed in Table 3.2. The Western Mountains ecoregion was broken into three finer-scale ecoregion subgroups for screening to match EMAP-West's use of a somewhat finer spatial scale.

In addition to the sites sampled in the NRSA, we obtained possible reference site external data from four other agencies. Data from these external surveys were screened for physical and chemical criteria using the same criteria used for NRSA sample sites in Table 3.2 (page 23) using whatever screening data were available in each survey.

All sites in the NRSA (both probability and hand-picked, boatable and wadeable) and the added external data that passed all criteria were considered to be candidate reference sites for the NRSA assessment. The number of reference sites that passed this screening is summarized in Table 3.1. These reference sites include both fish and macroinvertebrate data. Note that the NRSA did not use data on the biological assemblages themselves for any screening as these are the primary components of the stream and river ecosystems being evaluated, and to use them would constitute circular reasoning.

Note that the Rapid Bioassessment Protocol (RBP) physical habitat score was used as a filter in WSA but was not available in the NRSA data to use as a screen. The six ecoregions in the top half of the table were used in WSA and reported in Herlihy *et al.* (2008), the ecoregions in the bottom half of the table were screened using criteria developed in EMAP-West.

3.3 FINAL COMBINED REFERENCE SITE SCREEN

As a final screen, all of the NRSA screened reference sites, and those provided by WSA and any other source, were screened for the influence of dams and adjacent land use. Three additional landscape-GIS screening criteria were applied to the selected physiochemical screened reference sites. These screens included dam influence index, urbanization influence, and agricultural influence.

The dam influence index (DII) was used to assess the influence of upstream dams and the largest reservoir on NRSA reference sites. Any watershed boundaries that had a maximum distance of less than 200 km upstream of the sampling point were completely assessed, any watershed with a distance greater than 200 km upstream of the sample point, had a wedge shaped area assessed until 200 km upstream was reached. For all watersheds and wedges assessed, a calculation of the volume of the largest reservoir, the number of dams, and an index that weighted the maximum reservoir volume within the watershed or wedge by its proximity to the sample point was conducted. Each upstream reservoir was inversely weighted by its upstream flow distance from the sample point as:

$$\mathrm{w_i} = \mathrm{e}^{-\left(\frac{\mathrm{D_{flow}}}{\mathrm{D_{efolding}}}\right)},$$

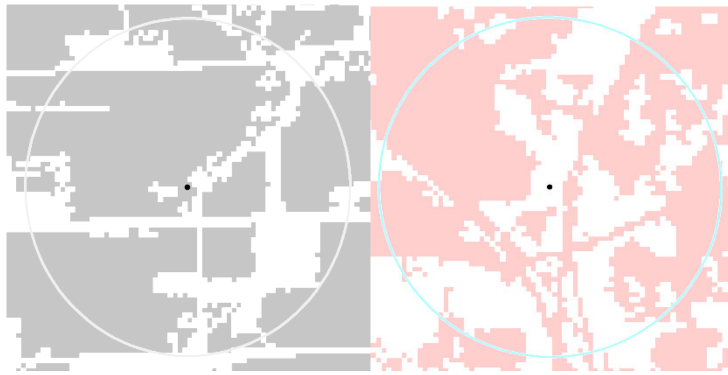
where D_{flow} is the flow distance to the sample site, and $D_{efolding}$ is an e-folding value that determines the rate at which the weight exponentially decreases (here 100 km). DII equals the largest distance-weighted volume within the watershed:

$$DII = \max(\mathrm{w_i} * \mathrm{D_i})$$

where D_i = reservoir volume (km^3). The threshold for dropping a potential reference site was a DII value equal to or greater than one.

Percent urbanization and agricultural influence were assessed within a 1 km^2 area around the mid-point of the sampled stream segment. To conduct this analysis a 1 km^2 radius buffer around the mid-point was overlaid onto the National Land Cover Dataset (NLCD) to calculate the percentage of urban land cover and percent row crop, as defined by the NLCD (Figure 3.1). The threshold for dropping a potential reference sites was any greater than 5% urban land cover and 15% agricultural (row crop) land cover.

Figure 3.1 Examples of percent urban (A, 60%) and row crop (B, 72%) from NLCD.



A B

3.4 LITERATURE CITED

Herlihy, A.T., R.M. Hughes, and J.C. Sifneos. 2006. National clusters of fish species assemblages in the conterminous United States and their relationship to existing landscape classification schemes. pp. 87-112. In R.M. Hughes, L. Wang, and P.W. Seelbach (eds.), Influences of Landscapes on Stream Habitats and Biological Assemblages. *American Fisheries Society Symposium* 48, Bethesda, Maryland.

Herlihy, A.T., S.G. Paulsen, J. Van Sickle, J.L. Stoddard, C.P. Hawkins, and L.L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference condition approach at a continental scale. *Journal of the North American Benthological Society* 27:860-877.

Table 3.2 Criteria for eight chemical and physical habitat filters used to identify the candidate least-disturbed reference sites for each of the nine aggregate ecoregions.

Filter criterion	NAP	SAP	CPL	UMW	TPL	SPL	NPL	XER	WMT-SW[e]	WMT-SRock[e]	WMT-Nrock/Pacific[e]
Total P (µg/L)	>20	>20	>75	>50	>100	>150	>150	>50	>50	>25	>25
Total N (µg/L)	>750	>750	>2500	>1000	>3000	>4500	>4500	>1500	>750	>750	>750
Cl^- (µeq/L)	>250[a]	>200	–	>300	>2000	>1000	>1000	>1000	>300	>200	>200[a]
SO_4^{2-} (µeq/L)	>250	>400	>600	>400	–	–	–	–	–	>200	>200
ANC (µeq/L) + DOC (mg/L)[b]	<50 + <5	<50 + <5	<50 + <5	<50 + <5	<50 + <5	<50 + <5	<50 + <5	<50+<5	<50 + <5	<50 + <5	<50 + <5
Turbidity (NTU)	>5	>5	>10	>5	>50	>50	>50	>25	>5	>5	>5
Riparian Disturbance Index[c]	>2	>2	>2	>2	>2	>2	>2	>1.5	>0.5/>1.5[d]	>1/>1.5[d]	>0.5/>1.5[d]
% fine substrate	>25	>25	>50	>40	>80	>90	>90	>50	>15	>15	>15

Values in red indicate a change from that used in WSA as reported in Herlihy *et al*., (2008).

– indicates filter criterion was not used in that ecoregion.
ANC = acid neutralizing capacity, DOC = dissolved organic C.
[a] Cl^- criterion not applied in Northeastern Coastal Zone (ecoregion 59) or Coast Range (ecoregion 1) sites
[b] Filter was specific for inorganic acidity; site had to exceed both criteria to fail
[c] Riparian disturbance index variable name is *W1_HALL* in physical habitat database (see Chapter 11).
[d] Wadeable stream/Boatable river criteria. Different criteria were used by stream size in the Western Mountains.
[e] To match screening criteria to what was done in the EMAP-West component of WSA, the Western Mountains ecoregion was divided into three subgroups: SW = Southwestern Mountains (Omernik level III codes 8 and 23, Southern California Mts., and Arizona/New Mexico Mts.), SRock = Southern Rockies (Omernik 19 and 21, Southern Rockies and Wasatch/Uintas), and NRock/Pacific = Northern Rockies and Pacific Mountains (all other WMT level III ecoregions).

4 QUALITY ASSURANCE

NRSA successfully implemented and assessed the quality of its operations and data throughout the survey. This chapter documents NRSA's adherence to the requirements of EPA's quality system implemented by the Office of Water as explained in the introduction section below. The following sections describe the quality aspects of the statistical design, field operations, laboratory assessments, data management, and report writing.

4.1 INTRODUCTION

The EPA quality system incorporates a national consensus standard for quality systems authorized by the American National Standards Institute (ANSI) and developed by the American Society for Quality Control (ASQC), ANSI/ASQC E4-2004, *Quality Systems for Environmental Data and Technology Programs – Requirements with Guidance for Use*. EPA Order CIO 2105.0, dated May 5, 2000, requires all of its component organizations to participate in an agency-wide quality system. The EPA Order also requires quality assurance project plans or "equivalent documents" for all projects and tasks involving environmental data.

In accordance with the EPA order, the Office of Water (OW) developed the Office Water Quality Management Plan (QMP; USEPA 2009) to describe OW's quality system that applies to all water programs and activities, including NRSA, collecting or using environmental data. As required by the EPA Order and OW QMP, NRSA developed and abided by its QAPP throughout the survey. The NRSA QAPP contains elements of the overall project management, data quality objectives, measurement and data acquisition, and information management. The QAPP also deals with the data integration necessary between the Wadeable Streams Assessment (WSA), NRSA, and EMAP Western Pilot Study (2001-2004) to create one complete report on the ecological status of the Nation's rivers and streams.

The following companion documents to the QAPP present detailed procedures for every stage of the survey:

- National Rivers and Streams Assessment: Site Evaluation Guidelines, EPA-841-B-07-008
- National Rivers and Streams Assessment: Field Operations Manual, (FOM), EPA-841-B-07-009
- National Rivers and Streams Assessment: Laboratory Methods Manual, (LMM), EPA 841-B-07-010

The four documents together address all aspects of NRSA's data acquisition and evaluation. The LMM and FOM also list measurement quality objectives (MQOs) which were used to evaluate the level of quality attainment for individual survey metrics. Every person involved in NRSA was responsible for abiding by the QAPP and adhering to the procedures specified in its companion documents. Moreover, every NRSA participant was trained in the requirements applicable to the person's role in the survey (*e.g.*, field crews were trained in the FOM procedures and applicable QAPP requirements). For example, field crews attended a combined classroom and hands-on training in field procedures.

4.2 SURVEY DESIGN

NRSA's survey design was based upon statistical concepts that are well accepted by the scientific community. As described in the following sections, the survey design objectives were met by requirements of the statistical design, completeness of implementing the design, and consistency with established procedures.

4.2.1 Statistical Design

There is a large body of statistical literature dealing with sample survey designs which addresses the problem of making statements about many by sampling the few (Kish 1965). Sample surveys have been used in a variety of fields (*e.g.*, monthly labor estimates) to determine the status of populations of interest, especially if the population is too numerous to census or if it is unnecessary to census the population to reach the desired level of precision for describing the population's status. In natural resource fields, probability sampling surveys have been consistently used to estimate the conditions of the entire population. For example, the National Agricultural Statistics Survey (NASS) conducted by the U.S. Department of Agriculture and the Forest Inventory Analysis (FIAT) conducted by the U.S. Forest Service (Bickford *et al*. 1963, Hazard and Law 1989) have both used probability based sampling concepts to monitor and estimate the condition and productivity of agricultural and forest resources from a commodity perspective. The sampling design strategy for NRSA is based on the fundamental requirement for a probability sample of an explicitly defined regional resource population, where the sample is constrained to reflect the spatial dispersion of the population. This design has been documented in peer reviewed literature (Stevens 1994, Stevens and Olsen 1999). By applying the statistical concepts of this design, the survey was able to meet the following overarching data quality objectives:

- In the conterminous U.S., estimate the proportion of river and stream length (± 5 percent) that falls below the designated threshold for good conditions for selected measures with 95 percent confidence.

- For each Omernik Level II Ecoregions, estimate the proportion of river and stream length (±15 percent) that falls below the designated threshold for good conditions for selected measures with 95 percent confidence.

4.2.2 Completeness

To ensure that the implementation of the NRSA sample design resulted in adequate measurements, the survey included completeness requirements for field sampling and laboratory analyses. The QAPP requires that valid data for individual indicators must be acquired from a minimum number of sampling locations to make subpopulation estimates with a specified level of confidence or sampling precision. As the starting place for selecting field sites, EPA used the National Hydrography Database (NHD; http://nhd.usgs.gov/) as the frame representing streams and rivers in the US because it was, and still is, the most complete source available and NRSA encountered few errors, as expected, given the completeness and quality measures undertaken in developing and maintaining NHD. Each participating State and Tribe completed its data collection at the number of sites designated for the State/Tribe by the sample design. Each laboratory met its completeness requirements in analyzing virtually all of the collected samples. Because of the completeness of its sample frame, field sampling, and laboratory work, NRSA met its completeness objective of 95 percent.

4.2.3 Comparability

Comparability is defined as the confidence with which one data set can be compared to another (Stanley and Verner, 1985; Smith *et al.*, 1988). For all indicators, NRSA ensured comparability by the use of standardized sampling procedures, sampling equipment and analytical methodologies by all sampling crews and laboratories. For all measurements, reporting units and format are specified, incorporated into standardized data recording forms, and securely transferred into a centralized information management system. Because NRSA used the same comparability measures to collect data in EMAP West and WSA studies, the data also can be compared across the studies. The following sections on field and laboratory operations describe additional measures to ensure consistency in NRSA.

4.3 Quality Assurance in Field Operations

The Field Operations Manual (FOM) ensured that quality objectives were attainable and survey activities were manageable. As described below, NRSA tested its FOM, trained crews using the FOM, visited crews during the field season, and confirmed fish specimen identifications.

4.3.1 Field Method Pilot Testing

Members of the NRSA steering committee and oversight staff pilot-tested sampling methods and documentation requirements (*e.g.*, field forms) described in the FOM. The pilot study tested the correctness and clarity of the FOM's instructions for executing the procedures and quality steps. The pilot study also tested sampling logistics, sample preparation, and sample shipping instructions. Through lessons learned during the pilot study, NRSA staff corrected and improved the FOM prior to field crew training.

4.3.2 Training of Field Trainers and Assistance Visitors

Before training field crews, members of the NRSA steering committee, oversight staff, contractor trainers, and other experts tested the training materials during an intensive 4-day period that included classroom and hands on training sessions. During the training, the attendees tested the materials to ensure that the instructions were correct and easy to execute. The training materials included the FOM and Quick Reference Guide (QRG). As a result of the training and expert discussions, NRSA staff corrected and improved the FOM and QRG before the field crew training.

4.3.3 Field Crew Training

To ensure consistency across field crews, all field crews were required to attend a 4-day training session prior to visiting any field site. NRSA trainers led seven regional field crew training sessions consisting of classroom and field-based lessons. The lessons included session on conducting site reconnaissance, recording field observations and *in situ* data, collecting field samples, packing jars for shipping, and use of the standardized field forms. The field crew leaders were taught to review every form and verify that all hand-entered data were complete and correct.

4.3.4 Field Assistance Visits

To further assist the crews in correctly implementing the field procedures and quality steps, a NARS staff member or contractor trainer visited every NRSA field crew during the field season. These visits, known as assistance visits (AV), provided an opportunity to observe field crews in the normal course of a field day, assist in correctly applying the procedures, and document the crew's adherence to sampling procedures. If circumstances were noted where a field crew was not conducting a procedure properly, the observer recorded the deficiency, reviewed the appropriate procedure with field team, and assisted the field crew until the procedure was completed correctly.

4.3.5 Revisits of Selected Field Sites

To evaluate comparability within NRSA, both field operations and laboratory assessments, 10 percent of the sites were revisited and samples collected. The primary purpose of the revisits was to allow variance estimates that would provide information on the extent to which the population estimates might vary. The pairs of measurements were used to evaluate signal to noise ratios for each of the indicators described in the main report. Overall, EPA determined that sampling sites at different times had little effect on the overall variability in the data.

4.3.6 Evaluation of Fish Identifications

To ensure consistent naming conventions, field taxonomist and laboratory ichthyologists were required to use commonly accepted taxonomic references to identify fish vouchers. To evaluate their identifications, field taxonomists were required to send the fish vouchers from one or more of its site visits to expert Ichthyologists for a second, independent, identification. Of the 2031 site visits for which field taxonomists were able to collect fish vouchers, 262 (13 percent) were selected for the independent evaluation. Of the 3153 vouchers selected for review, the ichthyologists were able to determine the taxa for 2782 vouchers which were 11 percent of the 25,425 vouchers collected for NRSA. (The remaining 371 vouchers were excluded mainly because the field crews did not provide the vouchers or the identifications.) The NRSA staff compared the taxa identifications by field crews and ichthyologists. For the 2781 vouchers, 87 percent were correct to species and 97 percent to genus. All but three vouchers were identified to family and all were correct to class. On average, the survey more than met the measurement objective for field taxonomists to correctly identify 85 percent of the fish vouchers.

4.4 Laboratory Quality Assurance and Quality Control

The NRSA laboratories used standard methods and/or followed the requirements in the Laboratory Methods Manual (LMM). The QAPP identified the overall quality requirements and the LMM provided methods that could be used to achieve the quality requirements. If a laboratory chose a different method, then it still had to meet the QA requirements as described below.

4.4.1 Basic Capabilities

All laboratories were required to submit documentation of their analytical capabilities prior to analyzing any NRSA sample. NRSA team members reviewed documentation to ensure that the laboratories could meet required measurement quality objectives (MQOs; *e.g.*, reporting limits, detection limits, etc.). National Environmental Laboratory Accreditation Conference (NELAC) certification, satisfactory participation in round-robin or other usual and customary types of evaluations were considered acceptable capabilities documentation.

4.4.2 Benthic Macroinvertebrate Identifications

For benthic macroinvertebrate taxonomy, laboratories were required to use the same taxa lists, conduct regular internal QC checks, and participate in an independent quality check. All participating laboratories identified organisms using the most appropriate technical literature that was accepted by the taxonomic discipline and reflected the accepted nomenclature at the time of the survey. The Integrated Taxonomic Information System (ITIS, http://www.itis.usda.gov) also was used to verify nomenclatural validity and reporting.

Taxonomic accuracy is evaluated by comparing identifications of the same organisms by primary and secondary laboratories. Each primary laboratory provided the organisms from three or more samples, up to 10 percent of its samples, to a secondary laboratory for an independent evaluation. Reconciliation calls were held to allow the taxonomists to come to consensus when organism identification was in question.

Of the 1756 samples identified by seven primary laboratories, 1254 samples had more than 300 organisms, and thus, were eligible for the independent quality check by the secondary laboratory. Of the 1254 samples, the secondary laboratory identified organisms in 130 samples. The mean percent taxonomic disagreement (PTD) between laboratories was 11 percent which more than meets the QAPP's measurement objective of 15 percent. The overall percent difference in enumeration (PDE) was only 1 percent which more than meets the QAPP's measurement objective of 5 percent.

Even though the measurement objectives were met, laboratories implemented recommendations and corrective steps for the QC samples and all other samples with the same organisms. If, for example, it was evident that empty mollusk shells were being identified and recorded in one or more of the QC samples, the laboratories needed to verify that they had not counted empty mollusk shells in their other samples.

4.4.3 Chemical Analyses

For quality assurance of chemical analyses, laboratories used QC samples which are similar in composition to samples being measured. They provide estimates of precision and bias that are applicable to sample measurements. To ensure the ongoing quality of data during analyses, every batch of water samples was required to include QA samples to verify the precision and accuracy of the equipment, reagent quality, and other quality measures. These checks were completed by analyzing blanks or samples spiked with known or unknown quantities of reference materials, duplicate analyses of the same samples, blank analyses, or other appropriate evaluations. The laboratories reported quality assurance results along with each batch of sample results. In addition, laboratories reported holding times. Holding time requirements for analyses ensure analytical results are representative of conditions at the time of sampling. The NARS team reviewed the data and noted any quality failures. The data analysts used the information about quality to determine whether to include or exclude data in the evaluations. As described in the next section, the consolidated NRSA database was further evaluated for quality failures.

4.5 DATA MANAGEMENT AND REVIEW

Information management (IM) is integral to all aspects of the NRSA from initial selection of sampling sites through dissemination and reporting of final, validated data. QA and QC measures implemented for the IM system are aimed at preventing corruption of data at the time of their initial incorporation into the system and maintaining the integrity of data and information after incorporation into the system.

Reconnaissance, field observation and laboratory analysis data were transferred from NRSA survey participants and collected and managed by the NARS IM center. Data and information were managed using a tiered-approach. First, *all* data transferred from a field team or laboratory were physically organized (*e.g.*, system folders) and stored in their original state. Next, NARS IM created a synthesized and standardized version of the data to populate a database that represented the primary source for all subsequent data requests, uses and needs. All samples were tracked from collection to the laboratory.

The IM staff applied an iterative process in reviewing the database for completeness, transcription errors, formatting compatibility, consistency issues and other quality control-related topics. This first-line data review was performed primarily by NARS IM in consultation with the NRSA QA team. A second-phase data quality review consisted of evaluating the quality of data based on MQOs as described in the QAPP. This QA review was performed by the NRSA QA team using a variety of qualitative and quantitative analytical and visualization approaches. Data that met the MQOs were used without restriction. Data that did not meet the MQOs were qualified and further evaluated to determine the extent to which quality control results deviated from the target MQOs. Minor deviations were noted and qualified, but did not prevent data from being used in analyses. Major deviations were also noted and qualified, but data were excluded from the analyses. Data not used for analyses because of quality control concerns account for a subset of the missing data for each indicator analysis and add to the uncertainty in condition estimates.

4.6 DATA ANALYSES

The NRSA team and its expert evaluated the data using standard biological assessment methodologies. After the data analysis was complete, each data analyst documented the assessment procedures and results for use in the technical report (*i.e.*, this document).

4.7 MAIN REPORT

The main report provides a summary of the findings of each of the data analyses and EPA's interpretation of them. After the main report was extensively reviewed in-house by the NRSA team, its partners, and other EPA experts, the report underwent two outside reviews. These outside reviews were the final step in ensuring that the main report and its findings met the quality requirements of the QAPP.

For the first review, EPA contracted with an outside firm to conduct an Independent External Peer Review (IEPR) of the main report. The firm selected three peer reviewers who were experts in water resource monitoring and biological and ecosystem assessments. The firm provided the reviewers with a copy of the main report, along with supporting documentation and a charge that solicited comments specifically on the technical content, completeness and clarity, and scientific integrity of the main report. EPA used the extensive comments from the peer reviewers to refine and review the main report before releasing a draft for public comment. As a result of the numerous and thoughtful public comments, EPA has corrected and revised certain elements of the main report.

4.8 LITERATURE CITED

American National Standards Institute and American Society for Quality Control (ANSI/ASQC), 2004. *Quality Systems for Environmental Data Collection and Environmental Technology Programs: Collection and Evaluation of Environmental Data.* E4-2004. Milwaukee, WI.

Bickford, C.A., C.E. Mayer, and K.D. Water. 1963. An Efficient Sampling Design for Forest Inventory: The Northeast Forest Resurvey. *Journal of Forestry.* 61: 826-833.

Kish, L. 1965. *Survey Sampling.* John Wiley & Sons. New York. 643 pp.

Smith, F., S. Kulkarni, L. E. Myers, and M. J. Messner. 1988. Evaluating and presenting quality assurance data. Pages 157-68 in L.H. Keith, ed. *ACS Professional Reference Book.* Principles of Environmental Sampling. American Chemical Society, Washington, D.C.

Stanley, T.W., and S.S. Verner. 1986. The U.S. Environmental Protections Agency's quality assurance program. pp. 12-19 IN: J.K. Taylor and T.W. Stanley (eds.). *Quality Assurance for Environmental Measurements. ASTM STP 867,* American Society for Testing and Materials, Philadelphia, Pennsylvania.

Stevens Jr., D. L. 1994. Implementation of a National Monitoring Program. *Journal Environmental Management* 42:1-29.

Stevens Jr., D. L., and A. R. Olsen. 1999. Spatially restricted surveys over time for aquatic resources. *Journal of Agricultural, Biological, and Environmental Statistics* 4:415-428.

Stribling, J.B., S.R. Moulton, and G.T. Lester. 2003. Determining the quality of taxonomic data. *Journal of the North American Benthological Society* 22(4):621-631.

USEPA. February 2009. *Office of Water Quality Management Plan.* EPA 821-R-09-001. 3rd Revision. U.S. Environmental Protection Agency, Office of Water. Washington, DC.

USEPA. February 2009. *National Rivers and Streams Assessment: Site Evaluation Guidelines.* EPA-841-B-07-008. U.S. Environmental Protection Agency, Washington, DC.

USEPA. April 2009. *National Rivers and Streams Assessment: Field Operations Manual.* EPA-841-B-07-009. U.S. Environmental Protection Agency, Washington, DC.

USEPA. November 2009. *National Rivers and Streams Assessment: Laboratory Methods Manual.* EPA-841-B-07-010. U.S. Environmental Protection Agency, Washington, DC.

USEPA. December 2010. *National Rivers and Streams Assessment: Integrated Quality Assurance Project Plan, Final Document*. EPA-841-B-07-007. U.S. Environmental Protection Agency, Office of Water and Office of Research and Development, Washington, DC.

USEPA. May 2000. Order CIO 2105.0, Policy and Program Requirements for the Mandatory Agency-wide Quality System.

5 OVERVIEW OF DATA ANALYSES

5.1 Stressor Extent, Relative Risk, and Attributable Risk

The NRSA database contained the field and laboratory data for all sampled sites, whether selected as potential reference sites or from the statistical design. Within each region, least-disturbed sites (i.e., reference sites described in 3) provide a benchmark against which all other sites were compared and classified. The condition classes for each stressor and biological response were determined from data and observations from the least-disturbed sites in each ecoregion and the continuous gradient of observed values at all sites. The resulting three condition classes were defined as follows:

- Good: Not different from the reference sites
- Fair: Somewhat different from the reference sites
- Poor: Markedly different from the reference sites
- Not Assessed: indicator not available for the site

While the "Not Assessed" category was included in the assessment (for instance, if fish were not caught at a site or a sample was damaged) for stressor and response extent analyses, these sites were not utilized in the relative risk or attributable risk analysis.

The condition classes were then used to estimate the extent, relative extent, relative risk, and attributable risk as described in the following sections.

A major goal of NARS is to assess the relative importance of stressors that impact aquatic biota on a national basis. EPA assesses the influence of stressors in three ways: stressor extent, relative risk, and population attributable risk. In NRSA, each targeted and sampled river and stream reach was classified as being in either *Good*, *Fair*, or *Poor* condition, separately for each stressor variable and for each biological response variable. From this data, we estimated the stressor extent (prevalence) of rivers and streams in *Poor* condition for a specified stressor variable. We also estimated the relative risk of each stressor for a biological response. Relative risk is the ratio of the probability of a poor biological condition when the stressor is poor to the probability of a poor biological condition when the stressor is not poor (Van Sickle et al. (2006)). Finally, we estimated the population attributable risk (AR) of each stressor for a biological response. AR combines RR and stressor extent into a single measure of the overall impact of a stressor on a biological response, over the entire population of rivers and streams (Van Sickle and Paulsen (2008)).

5.1.1 Stressor Extent

For each particular stressor, the stressor extent (SE) may be reported as the number of miles, the proportion of miles, or the percent of miles in *Good*, *Fair*, *Poor*, or *Not Assessed* condition. If the SE is reported as the proportion of miles, then it can be interpreted as the probability that a stream chosen at random from the population will be in *Poor* condition for the stressor.

Stressor extent in *Poor* condition is estimated as

(1) SE_p, the sum of the sampling weights for sites that are assessed in *Poor* condition

$$SE_p = \sum_{i=1}^{n_p} w_{pi}$$

,

(2) SEP_p, as the ratio of the sums of the sampling weights for the probability selected sites that are assessed in *Poor* condition divided by the sum of the sampling weights of all the selected sites regardless of condition, i.e.,

$$SEP_p = \frac{\sum_{i=1}^{n_p} w_{pi}}{\sum_{i=1}^{n} w_i}$$

, or

(3) SER_p, the percent of stressor extent in *Poor* condition (i.e., stressor relative extent)

$$SER_p == 100 * SEP_p = 100 * \frac{\sum_{i=1}^{n_p} w_{pi}}{\sum_{i=1}^{n} w_i}$$

where w_{pi} is the weight for the *i*th selected site in the *Poor* condition category, w_i is the weight for the *i*th selected site regardless of condition category, n_p is the number of selected sites that are in *Poor* condition, and n is the total number of sites regardless of their condition category. A stressor condition category may use other terminology to identify if a site is in poor condition but generically, we use the term *Poor*. Note that the extent for a response variable is defined similarly.

5.1.2 Relative Risk and Attributable Risk

To estimate RR and AR, we restrict the sites to those that both the stressor and response variable assessed as *Good*, *Fair*, or *Poor* (or their equivalents). That is, if a site is *Not Assessed* for either the stressor or response variable, it is dropped. Next, for these sites the condition classes are combined to be either *Poor* or *Not Poor* for the stressor and response variables. For example, *Not Poor* combines the *Good* and *Fair* condition classes. Thus, each sampled river or stream was designated as being in either *Poor* (P) or *Not Poor* (NP) condition for each stressor and response variable separately.

To estimate the relative risk and attributable risk for one stressor (S) and one response (B) variable, we compiled a 2x2 table (Table 5.1), based on data from all river and stream sites that were included in the probability sample and that had both the stressor and response variable measured. A separate table must be compiled for each pair of stressor and response variables.

Table 5.1: Extent estimates for response and stressor categories

	Stressor (S)	
Response (B)	Not Poor (NP)	Poor (P)
Not Poor (NP)	$a = \sum_{i=1}^{n_{nn}} w_{nni}$	$b = \sum_{i=1}^{n_{np}} w_{npi}$
Poor (P)	$c = \sum_{i=1}^{n_{pn}} w_{pni}$	$d = \sum_{i=1}^{n_{pp}} w_{ppi}$

Table entries (a, b, c, d) are the sums of the sampling weights of all sampled rivers and streams that were found to have each combination of *Poor* or *Not Poor* condition for stressor and response. For example, $d = \sum_{i=1}^{n_{pp}} w_{ppi}$ where n_{pp} is the number of sites with both the stressor and response in poor condition and w is the weight for the *i*th site. Note that the estimates in Table 5.1 may differ from the stressor extent estimates since both the stressor and response variables must be measured at each site.

Relative Risk

Relative risk (RR) is the ratio of the probability of a *Poor* biological condition when the stressor is *Poor* to the probability of a *Poor* biological condition when the stressor is *Not Poor*. That is,

$$RR = \frac{Pr(B = P|S = P)}{Pr(B = P|S = NP)}$$

Using the simplified notation in Table 5.1, relative risk (RR) is estimated as:

$$RR_{est} = \frac{d/(b + d)}{c/(a + c)}$$

A RR = 1.0 indicates there is no association between the stressor and response. That is, a *Poor* response condition in a river or stream is equally likely to occur whether or not the stressor condition is *Poor*. A RR > 1.0 indicates that a *Poor* response condition is more likely to occur when the stressor is *Poor*. For example, when the RR is 2.0, the chance that a stream is in *Poor* biological (response) condition is twice as likely when the stressor is *Poor* than when the stressor is *Not Poor*.

Further details of RR and its interpretation, including estimation of a confidence interval for RR_{est}, can be found in Van Sickle et al. (2006).

Attributable Risk

Population attributable risk (AR) measures what percent of the extent in *Poor* condition for a biological response variable can be attributed causally to the *Poor* condition of a specific stressor. AR is based on a scenario in which the stressor in *Poor* would be entirely eliminated from the population of river and streams, e.g., by means of restoration activities. That is, all rivers and streams in *Poor* condition for the stressor are restored to the *Not Poor* condition. AR is defined as the proportional decrease in the extent of *Poor* biological response condition that would occur if the stressor were eliminated from the population of rivers and streams. Mathematically, AR is defined as (Van Sickle and Paulsen (2008)).

$$AR = \frac{Pr(B = P) - Pr(B = P|S = NP)}{Pr(B = P)}$$

We estimated AR as

$$AR_{est} = \frac{BEP_p - c/(a + c)}{BEP_p}$$

where

$$BEP_p = \frac{(c + d)}{(a + b + c + d)}$$

and is the estimated proportion of the biological response that is in *Poor* condition. We calculated a confidence interval for AR_{est} following Van Sickle and Paulsen (2008).

An AR can take a value between 0 and 1. A value of 0 indicates either "No association" between stressor and response, or else a stressor has a zero extent, i.e., is not present in the population. A strict interpretation of AR in terms of stressor elimination, as described above, requires one to assume that the stressor-response relation is strongly causal and that stressor effects are reversible. Van Sickle and Paulsen (2008) discuss the reality of these assumptions, along with other issues such as interpreting them when multiple, correlated stressors are present, and using them to express the joint effects of multiple stressors.

However, AR can also be interpreted more informally, as a measure that combines RR and SE into a single index of the overall, population-level impact of a stressor on a response. Van Sickle and Paulsen (2008) show that the population attributable risk can be written as

$$AR = \frac{SEP_p(RR - 1)}{1 + SEP_p(RR - 1)}$$

This shows that the numerator of AR is the product of the SE of *Poor* stressor condition and the "excess" RR, i.e., RR-1, of that stressor. The denominator standardizes this product to yield AR values between 0 and 1. Thus, a high AR for a stressor indicates that the stressor is widely prevalent (has a high SE of *Poor* condition), and the stressor also has a large effect (high RR) in those river and stream reaches where it does have Poor condition.

5.2 QUANTITATIVE BIOLOGICAL ASSESSMENTS

Many countries manage aquatic resources to protect or restore the structure and function that is characteristic of ecosystems with minimal disturbance by human activities. Two approaches, multi-metric indices (MMIs) and predictive models comparing Observed and Expected (O/E) conditions, evaluate potential degradation of river and stream conditions relative to reference sites. The decision of whether traditional or modeled MMIs should be used and whether MMIs for the whole nation or by ecoregion should be used was based on MMI performance and satisfactory model validation. The following two sections describe the two approaches.

5.2.1 Quantitative Biological Metrics and Indices

Multimetric indices of biological condition (MMIs) are commonly used in assessments of aquatic resource condition. MMIs have been used in the U.S. to assess condition based on fish and macroinvertebrate assemblage data (*e.g.*, Karr and Chu, 1999; Barbour *et al.*, 1999; Barbour *et al.*, 1996). As ecological assessments become more common and large scale, accounting for natural variation among aquatic resources has been a challenge for accurate assessment. Ecoregions account for some natural variation in climate, geology, hydrology, and soils among sites, but significant natural variation in size and slope as well as local hydrology and soils can occur within a region and affect the expected characteristics of fish, macroinvertebrate, and algal assemblages at a site.

The multimetric approach involves summarizing various assemblage attributes (*e.g.*, composition, tolerance to disturbance, trophic and habitat preferences) as individual "metrics" or measures of the biological community. Metrics are used to access richness, diversity, and evenness across sites. These metrics are likely to be expressed relative to what is achievable or expected among reference sites that represent near-pristine or best available conditions. Candidate metrics are then evaluated for various aspects of performance and a subset of the best performing metrics are then combined into an index, referred to as a multimetric index or MMI. For example, MMIs might be selected based upon sensitivity of metrics to human disturbance, commonness, and independence of candidate metrics. Sensitivity to human disturbance is measured with t-statistics comparing central tendency and variation between reference and highly disturbed sites. Commonness is determined by the percent of sites with taxa present for calculation of metric values, with 75% as a criterion for acceptability. Metric independence was shown if candidate metrics were not highly correlated ($r2>0.64$) among reference sites, indicating they were relatively independent characterizations of biological condition. In some cases, covarying metrics were included in an MMI if they were in the same metric category and if they contributed to equal weights of metric categories in the MMI. Priority was also given to candidate metrics with consistently high performance across the ecoregions so the same metrics could be used in MMIs for all ecoregions. MMIs for NRSA also were required to clearly distinguish between reference and highly disturbed sites.

NRSA evaluations intentionally excluded metrics calculated with taxa traits based on sensitivity and tolerance to individual pollutants (*e.g.*, nutrients, pH, or conductivity). Such metrics have been highly correlated with metrics for the generalized stressor gradient of human disturbance. For this reason, there is some concern that that traits are uniquely indicative of the pollutant for which they were calculated (Stevenson *et al*. 2008, Stevenson *et al*. submitted). The processes of characterizing biological condition and then identifying pollutants affecting biological condition should be as independent as possible (Stevenson 2006).

For the main report, candidate metrics were recalculated using a mixture of random forest models or t-statistics. Multiple linear regression models adjusted thresholds for natural variation among sites and for later use in modeled MMIs. The modeled metric value was determined as the difference between observed and expected values of a metric at a site. Modeled metric values greater and less than zero were, respectively, greater and less than expected metric values if the site was minimally disturbed. To develop the random forest modeled expected condition, only counts from two-thirds of the reference sites were used. One-third of the reference sites were randomly selected and used later to validate the modeled MMIs. The random forest models used independent variables such as climate, geological, and landscape variables that are affected little by human activities.

5.2.2 Predictive Models

The predictive model approach was initially developed in Europe and Australia, and is becoming more prevalent within the U.S. The approach estimates the expected taxonomic composition of an assemblage in the absence of human stressors (Hawkins *et al*., 2000; Wright, 2000), using a set of "least-disturbed" sites and other variables related natural gradients such as elevation, stream size, stream gradient, latitude, longitude. Site-specific predictions of the expected characteristics of a site can be modeled and enable assessment of ecological condition as the deviation between observed and expected condition.

Expected condition can be minimally disturbed condition or best available condition (*sensu* Stoddard *et al*. 2006), or desired condition (*sensu* Stevenson *et al*. 2004). Predictive models for expected condition can be determined using data from all sites in a regional assessment (Seelbach *et al*. 2002, Baker *et al*. 2005, and Riseng *et al*. 2010) or from selected sites that are assumed to meet management goals, such as reference sites (Clarke *et al*. 1996, Hawkins *et al*. 2000). This approach offers a quantitative way to assess sites that show signs of disturbance or that differ in terms of variety or abundance from what would be considered normal. Candidate metrics are evaluated for aspects of performance and a subset of the best performing metrics are combined into an index known as an Index of Biotic Condition (IBI). This index is then used to assign a ranking of the condition of the resource.

The resulting models are then used to estimate the expected composition at each sampled site if the site had no anthropogenic influences. The expected taxa composition is expressed as "taxa richness." The number actually observed at a site is compared to the total number expected as an observed to expected ratio (O/E index). The O/E ratio predicted by the model for any site expresses the number of taxa found at that site (O), as a proportion of the number that would be expected (E) if the site was in least-disturbed condition. The O/E approach is based on the RIVPACS (River Invertebrate Prediction and Classification System) Models developed by Wright (1995). Ideally, a site in reference condition has O/E = 1.0.

Individual O/E values are most reliably interpreted relative to the entire O/E distribution for reference sites. Departures from a ratio of 1.0 indicate that the sample differs from that expected under less disturbed conditions. An O/E value of 0.70 indicates that 70% of the "expected" taxa at a site were actually observed at the site. This is interpreted as a 30% loss of taxa relative to the site's predicted reference condition. However, O/E values vary among reference sites themselves, around the idealized value of 1.0, because such sites rarely conform to an idealized reference condition, and because of model error and sampling variation. Statistical tests for departures from a ratio of one indicate that the characteristics of the sample differ significantly from that expected under reference conditions.

A random forest model is then built to predict the membership of any site in these classes, using natural environmental features as predictor variables. The predicted occurrence probability of a reference taxon at a site is then predicted to be the weighted average of that taxon's occurrence frequencies in all reference site classes, using the site's predicted group membership probabilities in the classes as weights. Finally, E for any site is the sum, over a subset of reference taxa, of predicted taxon occurrence probabilities. O is the number of taxa in that subset that were observed to be present at the site.

5.2.3 Change Analyses

The NRSA conducted in 2008 to 2009 was the first comprehensive, statistically valid survey of the nation's flowing water resources. It is one in a series of surveys designed to assess the condition of all waters (rivers/streams, lakes, wetlands, and coastal waters). The NRSA is a collaborative effort and partnership between EPA, states, and tribes. Data from this assessment will serve as the baseline for the condition of the nation's rivers and the first change analysis of the nation's wadeable streams.

The sampling design for the NRSA is a probability-based network that provides statistically valid estimates of condition for all rivers and streams with a known confidence. Field crews composed of states, tribes, EPA, USGS, and contractors sampled a total of 2,341 streams and rivers (including reference, base, enhancement and revisit sites) during the summer index period of 2008 and 2009. The survey measures a wide variety of variables intended to characterize the chemical, physical, and biological condition of the nation's flowing waters.

Previously, EPA and partners reported on the condition of all streams in the Wadeable Streams Assessment (WSA). The change analysis examines difference in the population of wadeable streams between the WSA and the NRSA.

5.2.4 Overall Change Analysis

The main report presents the estimated percentage differences in total stream length between surveys as the percentages of total streams length that have been categorized as "good," "fair," and "poor." WSA thresholds were used for each indicator to have a standard threshold for each indicator across surveys. The analysis incorporates a specifically designated set of sampling weights for change to produce regional as well as national estimates.

Sampling locations in both surveys are shown in Figure 5.1. Sites sampled in both surveys have good geographic coverage. There are fewer sites in the west, because fewer streams exist in the west. Note that 359 of the sites (yellow dots) sampled by the first survey were resampled in the second survey.

Our data analysis is based on one water chemistry sample from each site, from each survey. Samples in both surveys were collected during summer low-flow period, when streams are under maximum stress from high temperature and dewatering.

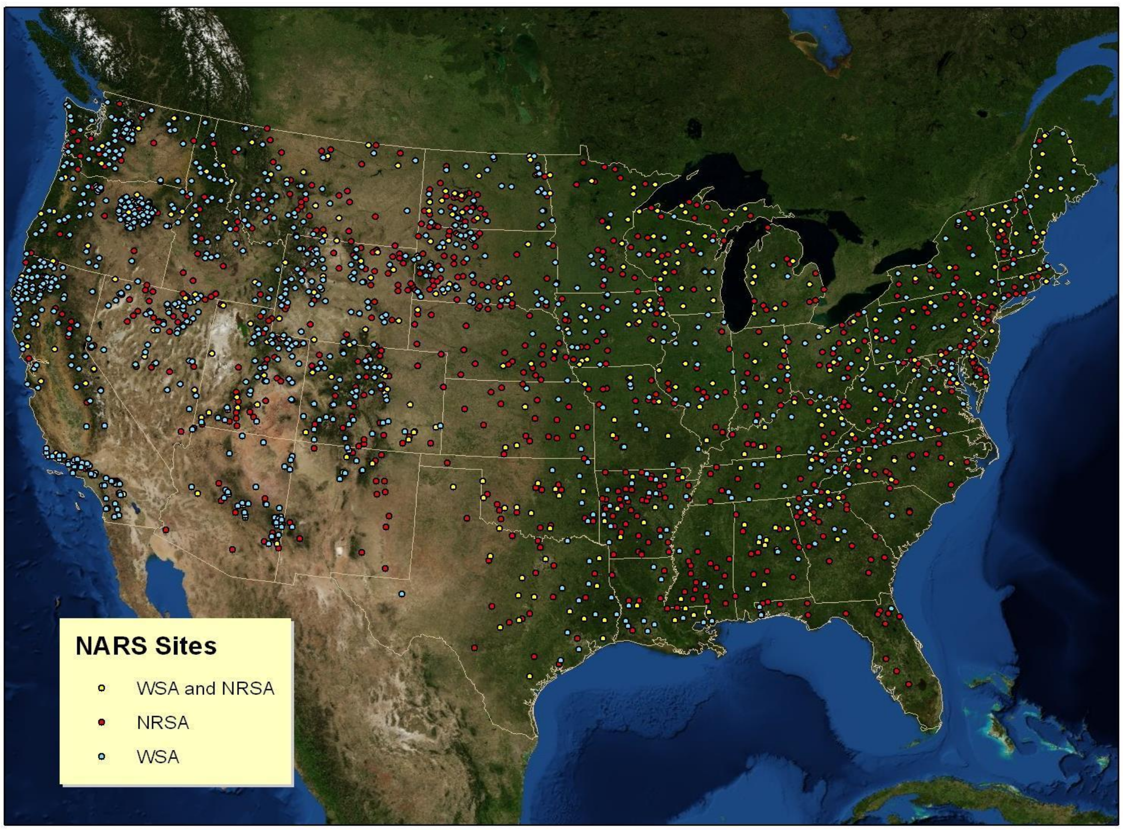


Figure 5.1 Sampling locations for the NRSA and WSA.

5.2.5 Caveats for Interpreting the Change Analysis

The main report provides the first look at changes in wadeable stream sites across the nation using a statistically valid sampling design.

- This analysis does not represent a trend; until additional surveys are implemented, we can only look at differences or "changes."

- While this first assessment of the chemistry data shows an increase in total phosphorus, the cause for that increase is still being explored. EPA scientists are examining whether these differences appear to be related to human effects or represent natural variation (flow, etc.). Preliminary analyses have looked at flow issues as well as whether sampling or lab protocols might explain the difference. None of the preliminary work has provided an explanation for the differences.

These and other possible explanations must still be considered before we draw any conclusion for change from the two surveys.

5.3 Literature Cited

Baker, E. A., K. E. Wehrly, P. W. Seelbach, L. Wang, M. J. Wiley, and T. Simon. 2005. A multimetric assessment of stream condition in the Northern Lakes and Forests ecoregion using spatially explicit statistical modeling and regional normalization. *Transactions of the American Fisheries Society* **134**:697-710.

Barbour, M. T., J. Gerritsen, G. E. Griffith, R. Frydenborg, E. McCarron, J. S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society* 15:185-211.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish*. EPA 841/B-99/002, Office of Water. US Environmental Protection Agency, Washington, DC.

Clarke, R. T., M. T. Furse, J. F. Wright, and D. Moss. 1996. Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. *Journal of Applied Statistics* **23**:311-332.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams. *Ecological Applications* **10(5)**:1456-1477.

Karr, J. R. and E. W. Chu. 2000. Sustaining living rivers. *Hydrobiologia* 422/423:1-14.

Riseng, C. M., M. J. Wiley, P. W. Seelbach, and R. J. Stevenson. 2010. An ecological assessment of Great Lakes tributaries in the Michigan Peninsulas. *Journal of Great Lakes Research* **36**:505-519.

Seelbach, P. W., M. J. Wiley, P. A. Soranno, and M. T. Bremigan. 2002. Aquatic conservation planning: using landscape maps to predict ecological reference conditoins for specific waters. Pages 454-478 *in* K. J. Gutzwiller, editor. *Applying Landscape Ecologyin Biological Conservation.* Springer-Verlag Publishers, New York.

Stevenson, R. J., R. C. Bailey, M. C. Harass, C. P. Hawkins, J. Alba-Tercedor, C. Couch, S. Dyer, F. A. Fulk, J. M. Harrington, C. T. Hunsaker, and R. K. Johnson. 2004. Designing data collection for ecological assessments. Pages 55-84 *in* M. T. Barbour, S. B. Norton, H. R. Preston, and K. W.

Thornton, editors. *Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Society of Environmental Toxicology and Contamination Publication*, Pensacola, Florida.

Stevenson, R. J., R. C. Bailey, M. C. Harass, C. P. Hawkins, J. Alba-Tercedor, C. Couch, S. Dyer, F. A. Fulk, J. M. Harrington, C. T. Hunsaker, and R. K. Johnson. 2004. Interpreting results of ecological assessments. Pages 85-111 *in* M. T. Barbour, S. B. Norton, H. R. Preston, and K. W. Thornton, editors. *Ecological Assessment of Aquatic Resources: Linking Science to Decision-Making. Society of Environmental Toxicology and Contamination Publicatio*n, Pensacola, Florida.

Stevenson, R.J. 2006. Refining diatom indicators for valued ecological attributes and development of water quality criteria. In: Ognjanova-Rumenova, N. And K. Manoylov, eds. *Advances in Phycological Studies*. Pp. 365-383. Pensoft Publishers. Moscow.

Stevenson, R. J., Y. Pan, K. Manoylov, C. Parker, D. P. Larsen, and A. T. Herlihy. 2008. Development of diatom indicators of ecological conditions for streams of the western United States. *Journal of the North American Benthological Society* **27**:1000-1016.

Stevenson, R.J., J. Zalack, J. Wolin. submitted. A multimetric index of lake diatom condition using surface sediment assemblages. *Freshwater Science.*

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* **16**:1267-1276.

Van Sickle, J. and S.G. Paulsen. 2008. Assessing the attributable risks, relative risks, and regional extents of aquatic stressors. *Journal of the North American Benthological Society* 27:920-931.

Wright, J.F., 2000. An introduction to RIVPACS. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.), *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques.* Freshwater Biological Association, Ambleside, UK, pp. 1-24.

Wright, J.F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology*. 20: 181- 197.

6 WATER CHEMISTRY ANALYSES

The main report summarizes four chemical stressors: total nitrogen (TN), total phosphorus (TP), acidity, and salinity. Criteria values and class definitions for acidity and salinity were identical to those used in the Wadeable Streams Assessment (WSA) as described below. TP and TN thresholds were calculated using the same procedure used in WSA but were recalculated to include additional nutrient reference sites sampled in NRSA. This roughly doubled the number of nutrient reference sites available in each ecoregion allowing for better estimation of the percentiles used to calculate thresholds for TN and TP. As described in the following sections, thresholds were established for the same nine ecoregions in the NRSA and WSA analyses.

6.1 ACIDITY AND SALINITY THRESHOLDS

For acidity, criteria values were determined based on values derived during the NAPAP program. Sites with acid neutralizing capacity (ANC) less than zero were considered acidic. Acidic sites with dissolved organic carbon (DOC) greater than 10 mg/L were classified as organically acidic (natural). Acidic sites with DOC less than 10 and sulfate less than 300 μeq/L were classified as acidic deposition impacted, while those with sulfate above 300 μeq/L were considered acid mine drainage impacted. Sites with ANC between 0 and 25 μeq/L and DOC less than 10 mg/L were considered acidic-deposition-influenced but not currently acidic. These low ANC sites typically become acidic during high flow events (episodic acidity).

Salinity data values were divided into good, fair, or poor classes. Salinity classes were defined by specific conductance using ecoregional specific values (Table 6.1).

6.2 TOTAL PHOSPHORUS AND TOTAL NITROGEN THRESHOLDS

Total nitrogen and phosphorus were classified into good-fair-poor classes using a method similar to that used for macroinvertebrate IBI classes using deviation from reference site distribution percentiles by aggregate ecoregion (see Herlihy and Sifneos, 2008 for details).

For nutrients, the value at (and below) the 75th percentile of the reference distribution was used for each ecoregion to define the least-disturbed condition class (good–fair boundary). The 95th percentile (and above) of the reference distribution in each ecoregion defines the most disturbed condition class (Table 6.1).

A set of “nutrient reference sites” was defined for this analysis using both WSA and NRSA data. All available WSA and NRSA sample sites were screened for water chemical and physical habitat disturbances using the process described in 3. Sites with screening values exceeding thresholds in Table 3.2 (page 23) were excluded as nutrient reference sites with the exception that TP and TN values were not used as screens (to avoid circularity).

To make up for losing these disturbance screens in defining nutrient reference sites, we added additional screens for land cover disturbance. A single national criteria were used to drop as nutrient reference sites those sites that had watershed %Urban LULC >10%, watershed road density > 3 km/km2, and watershed population density >100 #/km2. For watershed %Agriculture LULC screening, ecoregional specific criteria were used as screens; NAP, WMT, XER (>10%), CPL, NPL, SAP, SPL, UMW (>25%), TPL (>50%). Before calculating ecoregional nutrient reference site percentiles, outliers (values outside 1.5 times the interquartile range above and below the quartiles) were removed.

Table 6.1 Nutrient and Salinity Category Criteria for NRSA Assessment

Ecoregion	Salinity as Conductivity (µS/cm) Good-Fair	Salinity as Conductivity (µS/cm) Fair-Poor	Total N (µg/L) Good-Fair	Total N (µg/L) Fair-Poor	Total P (µg/L) Good-Fair	Total P (µg/L) Fair-Poor
CPL	500	1000	624	1081	55.9	103
NAP	500	1000	345	482	17.1	32.6
SAP	500	1000	240	456	14.8	24.4
UMW	500	1000	583	1024	36.3	49.9
TPL	1000	2000	700	1274	88.6	143
NPL	1000	2000	575	937	64.0	107
SPL	1000	2000	581	1069	55.8	127
WMT	500	1000	139	249	17.7	41.0
XER	500	1000	285	529	52.0	95.9

6.3 Literature Cited

A.T. Herlihy and J.C. Sifneos. 2008. Developing nutrient criteria and classification schemes for wadeable streams in the conterminous US. *Journal of the North American Benthological Society* 27:932-948.

7 HUMAN HEALTH FISH TISSUE INDICATOR -- MERCURY

Fish are time-integrating indicators of persistent pollutants, and contaminant bioaccumulation in fish tissue has important human and ecological health implications. Contaminants in fish pose risks to human consumers and to piscivorous wildlife. The NRSA fish tissue indicator provides information on the national distribution of selected persistent, bioaccumulative, and toxic (PBT) chemical residues (*e.g.*, mercury and organochlorine pesticides) in predator fish species from rivers 5th order and greater in size of the conterminous United States. For the main report, only the mercury results are presented. For a wide variety of additional chemicals (including selenium, pesticides, PCBs, and other contaminants of emerging concern), analyses are still underway and will be presented in future publications.

The fish tissue indicator field and analysis procedures described below were based on EPA's National Study of Chemical Residues in Lake Fish Tissue (final version now available) and EPA's *Guidance for Assessing Chemical Contaminant Data for Use in Fish Advisories*, Volume 1 (third edition).

7.1 FIELD FISH COLLECTION

The NRSA crews collected fish for the tissue indicator from rivers 5th order and greater in size. Fish tissue samples consisted of a composite of fish (*i.e.*, five individuals of one predator species) from each site. The fish had to be large enough to provide sufficient tissue for analysis (*i.e.*, 500 grams of fillets, collectively). Additional fish criteria for each composite sample included that fish were:

- Same species (for each site);
- Harvestable size per legal requirements or be of consumable size if there were no harvest limits; and
- Similar size so that the smallest individual in the composite was no less than 75% of the total length of the largest individual.

Crews were provided with a recommended list of target fish species (Table 7.1), though they could choose an appropriate substitute if none of the recommended fish were available.

Table 7.1 Recommended Target Species for Fish Tissue Collection (in Order of Preference)

Predator/Gamefish Species (in order of preference)	Family name	Common name	Scientific name	Length Guideline (Estimated Minimum)
	Centrarchidae	Largemouth bass	*Micropterus salmoides*	~280 mm
		Smallmouth bass	*Micropterus dolomieu*	~300 mm
		Black crappie	*Pomoxis nigromaculatus*	~330 mm
		White crappie	*Pomoxis annularis*	~330 mm
	Percidae	Walleye/sauger	*Sander vitreus/S. canadensis*	~380 mm
		Yellow perch	*Perca flavescens*	~330 mm
	Percichthyidae	White bass	*Morone chrysops*	~330 mm
	Esocidae	Northern pike	*Esox lucius*	~430 mm
	Salmonidae	Lake trout	*Salvelinus namaycush*	~400 mm
		Brown trout	*Salmo trutta*	~300 mm
		Rainbow trout	*Oncorhynchus mykiss*	~300 mm
		Brook trout	*Salvelinus fontinalis*	~330 mm

7.2 Mercury Analysis and Human Health Screening Values

All fish tissue samples were analyzed for total mercury using a commercially available mercury analyzer that requires only a small amount of tissue (about 1 gram) for analysis. In screening-level studies of fish contamination, EPA guidance recommends monitoring for total mercury rather than methylmercury since most mercury in adult fish is in the toxic form of methylmercury. Applying the conservative assumption that all mercury is present in fish tissue as methylmercury is also more protective of human health. The human health screening value used to interpret mercury concentrations in fillet tissue is 0.3 milligrams (mg) of methylmercury per kilogram (kg) of tissue (wet weight) or 300 parts per billion (ppb), which is EPA's tissue-based water quality criterion for methylmercury. This threshold represents the concentration that, if exceeded, can potentially be harmful to human health. Application of this threshold to the fillet data identifies the number and percentage of river miles in the sampled population for this study that exceed the mercury human health screening value. Results are presented for the miles of 5th order and larger rivers that could not be sampled, and the miles that exceed/do not exceed the human health screening value.

7.3 Literature Cited

EPA. November 2000. *Guidance for Assessing Chemical Contaminant Data for Use in Fish Advisories, Volume 1 (Third Edition)*. EPA 823-B-00-007.

EPA. 2001. Water Quality Criterion for the Protection of Human Health: Methylmercury. EPA-823-R-01-001. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

EPA. September 2009. *National Study of Chemical Residues in Lake Fish Tissue*. EPA 823-R-09-006.

8 *ENTEROCOCCI* INDICATOR

The EPA has developed and validated a molecular testing method called quantitative polymerase chain reaction (qPCR) as a rapid analytical technique for the detection of enterococci in recreational water. NRSA used this method to assess the presence and quantity of fecal indicators in the nation's rivers and streams. EPA then applied the draft threshold and recreational criteria to the enterococci data to assess the recreational condition of streams and rivers.

8.1 METHODS

To collect enterococci samples, crews took a water sample for the fecal indicator at the last transect after all other sampling was completed. Using a pre-sterilized 250 mL bottle, they collected the sample approximately 1 m off the bank at about 0.3 m (12 inches) below the water. Following collection, crews placed the sample in a cooler and kept it on ice prior to filtration of four 50 mL volumes. Samples were all filtered and frozen on dry ice within 6 hours of collection. In addition to collecting the sample, crews looked for signs of disturbance throughout the reach that would contribute to the presence of fecal contamination to the waterbody.

This collection and the laboratory method followed EPA's Enterococcus qPCR method A. Method A describes a quantitative polymerase chain reaction (qPCR) procedure for the detection of DNA from enterococci bacteria in ambient water matrices based on the amplification and detection of a specific region of the large subunit ribosomal RNA gene (lsrRNA, 23S rRNA) from these organisms. Method A uses an arithmetic formula, the comparative cycle threshold (CT) method, to calculate the ratio of enterococcus lsrRNA gene target sequences (target sequences) recovered in total DNA extracts from water samples relative to those in similarly prepared extracts of calibrator samples containing a known quantity of enterococcus cells. Mean estimates of the absolute quantities of target sequences in the calibrator sample extracts are then used to determine the absolute quantities of target sequences in the water samples. CT values for sample processing control (SPC) sequences added in equal quantities to both the water filtrate and calibrator samples before DNA extraction are used to normalize results for potential differences in DNA recovery or to signal inhibition or fluorescence quenching of the PCR analysis caused by a sample matrix component or possible technical error.

8.2 APPLICATION OF THRESHOLDS

8.2.1 Thresholds

For the data analysis of the enterococci measurements determined by the qPCR method, EPA used its draft thresholds that have been defined and outlined in the document *Recreational Water Quality Criteria* (EPA-HQ-OW-2011-0466). The document contains the EPA's draft ambient water quality criteria recommendations for protecting human health in marine and freshwaters. The document describes the threshold development and values.

8.2.2 Calibration

Before applying the thresholds to the qPCR data, it was necessary to standardize the implementation and calibration of qPCR methods. Comparison of results based on a calibrator cell equivalent (CCE) reporting unit assumes that target sequence copies (TSC) per calibrator cell is the same for all calibrator cell preparations. If calibrator TSC/cell values are not similar, test sample CCE estimates are not comparable without an adjustment. TSC/calibrator cell values can be determined from standard curves using DNA standards of known TSC concentration. Evidence has been seen in several studies that TSC per cell may vary for different calibrator cell preparations.

Estimates of TSC per calibrator cell for new calibrator cell preparations must also be related back to the values associated with the qPCR criteria values in the pending recreational water quality criteria. Estimates of TSC per calibrator cell generated from standard curves for the National Lakes Assessment and NRSA studies had a mean estimate of 20.41 TSC per cell.

8.3 LITERATURE CITED

Cabelli, V.J., A.P. Dufour, M.A. Levin, L.J. McCabe, and P.W. Haberman. 1979. Relationship of Microbial Indicators to Health Effects at Marine Bathing Beaches. *Am. J. Public Health*. 69: 690-696.

Francy, D.S. and Darner, R.A., 1998, Factors affecting Escherichia coli concentrations at Lake Erie public bathing beaches: U.S. Geological Survey, *Water-Resources Investigations Report* 98-4241, 4 p.

Haugland, R.A., Siefring, S.C., Wymer, L.J., Brenner, K.P. and Dufour, A.P., 2005, Comparison of Enterococcus measurements in freshwater at two recreational beaches by quantitative polymerase chain reaction and membrane filter culture analysis: *Water Research*, v. 39, p. 559-568.

Oshiro, R.K., Chambers, Y., Pope, M., Miller, K., Grunerud, R., and Keller, K., 2007, Assessment of the effects of holding time on enterococci concentrations in fresh and marine recreational waters and Escherichia coli concentrations in fresh recreational waters: Poster presented at of the American Society for Microbiology Annual Meeting, Toronto, Canada.

Pope, M. L., Bussen, M., Feige, M.A., Shadix, L., Gonder, S., Rodgers, C., Chambers, Y., Pulz, J., Miller, K., Connell, K., and Standridge, J., 2003, Assessment of the effects of holding time and temperature on *Escherichia coli* densities in surface water samples: *Applied and Environmental Microbiology*, v. 69, no. 10, p. 6201-6207.

U.S. Environmental Protection Agency, 1986, *Ambient Water Quality Criteria for Bacteria -1986*: Washington, D.C., U.S. EPA 440/5-84-002, 18 p.

U.S. Environmental Protection Agency, 2010. *Method A: Enterococci in Water by TaqMan® Quantitative Polymerase Chain Reaction (qPCR) Assay*. EPA-821-R-10-004.

Wade, T.J., Calderon, R.L., Sams, E., Beach, M., Brenner, K.P., Williams, A.H., and Dufour, A.P., 2006, Rapidly measured indicators of recreational water quality are predictive of swimming-associated gastrointestinal illness: *Environmental Health Perspectives*, v. 114, no. 1, p. 24-28.

9 BENTHIC MACROINVERTEBRATE ASSEMBLAGE

The taxonomic composition and relative abundance of different taxa that make up the benthic macroinvertebrate assemblage present in a stream have been used extensively in North America, Europe, and Australia to assess how human activities affect ecological condition (Barbour *et al.* 1995, 1999; Karr and Chu 1999). As explained in general terms in Section 5.2, two principal types of ecological assessment tools to assess condition based on benthic macroinvertebrates are currently prevalent: multimetric indices and predictive models of taxa richness. The purpose of these indicators is to present the complex community taxonomic data represented within an assemblage in a way that is understandable and informative to resource managers and the public. The following sections provide an overview of the approaches used to develop ecological indicators based on benthic macroinvertebrate assemblages, followed by details regarding data preparation and the process used for each approach to arrive at a final indicator.

9.1 OVERVIEW

Multimetric indicators have been used in the U.S. to assess condition based on fish and macroinvertebrate assemblage data (*e.g.*, Karr and Chu, 1999; Barbour *et al.*, 1999; Barbour *et al.*, 1995). The multimetric approach involves summarizing various assemblage attributes (*e.g.*, composition, tolerance to disturbance, trophic and habitat preferences) as individual "metrics" or measures of the biological community. Candidate metrics are then evaluated for various aspects of performance and a subset of the best performing metrics are then combined into an index, referred to as a multimetric index or MMI. For NRSA, the MMI developed in the WSA was used to generate the population estimates used in the assessment. The WSA MMI is detailed in Stoddard *et al.* (2008).

The predictive model approach was initially developed in Europe and Australia, and is becoming more prevalent within the U.S. The approach estimates the expected taxonomic composition of an assemblage in the absence of human stressors (Hawkins *et al.*, 2000; Wright, 2000), using a set of "least-disturbed" sites and other variables related natural gradients (such as elevation, stream size, stream gradient, latitude, longitude). The resulting models are then used to estimate the expected taxa composition (expressed as taxa richness) at each stream site sampled. The number of expected taxa actually observed at a site is compared to the total number of expected taxa as an observed:expected ratio (O/E index). Departures from a ratio of 1.0 indicate that the taxonomic composition in a stream sample differs from that expected under less disturbed conditions.

9.2 DATA PREPARATION

9.2.1 Standardizing Counts

The number of individuals in a sample was standardized to a constant number to provide an adequate number of individuals that was the same for the most samples and that could be used for both multimetric index development and O/E predictive modeling index. A subsampling technique involving random sampling without replacement was used to extract a true "fixed count" of 300 individuals from the total number of individuals enumerated for a sample (target lab count was 500 individuals). Samples that did not contain at least 300 individuals were used in the assessment because low counts can indicate a response to one or more stressors. Only those sites with at least 250 individuals, however, were used as reference sites.

9.2.2 Operational Taxonomic Units

For the predictive model approach, it was necessary to combine taxa to a coarser level of common taxonomy. This new combination of taxa is termed an "operational taxonomic unit" or OTU, and results in fewer taxa than are present in the initial benthic macroinvertebrate count data.

9.2.3 Autecological Characteristics

Autecological characteristics refer to specific ecological requirements or preferences of a taxon for habitat preference, feeding behavior, and tolerance to human disturbance. These characteristics are prerequisites for identifying and calculating many metrics. A number of state/regional organizations and research centers have developed autecological characteristics for benthic macroinvertebrates in their region. For the WSA and NRSA, a consistent "national" list of characteristics that consolidated and reconciled any discrepancies among the regional lists was needed before certain biological metrics could be developed and calibrated and an MMI could be constructed. The same autecological information used in WSA was used in NRSA.

Members of the data analysis group pulled together autecological information from five existing sources: (1) the EPA Rapid Bioassessment Protocols document; (2) the National Ambient Water Quality Assessment (NAWQA) national and northwest lists; (3) the Utah State University list; (4) the EMAP Mid-Atlantic Highlands (MAHA); and (5) the EMAP Mid-Atlantic Integrated Assessment (MAIA) list. These five were chosen because they were thought to be the most independent of each other and the most inclusive. A single national-level list was developed based on the decision rules described in the following sections.

9.2.3.1 Tolerance Values

Tolerance value assignments followed the convention for macroinvertebrates, ranging between 0 (least tolerant or most sensitive) and 10 (most tolerant). For each taxon, tolerance values from all five sources were reviewed and a final assignment made according to the following rules:

1. If values from different lists were all <3 (sensitive), final value = mean.
    - If values from different lists were all >3 and <7 (facultative), final value = mean.
    - If values from different lists were all >7 (tolerant), final value = mean.
    - If values from different lists spanned sensitive, facultative, and tolerant categories, best professional judgment was used, along with alternative sources of information (if available) to assign a final tolerance value.
    - Tolerance values of 0 to ≤3 were considered "sensitive" or "intolerant." Tolerance values ≥7 to 10 were considered "tolerant," and values in between were considered "facultative."

9.2.3.2 Functional Feeding Group and Habitat Preferences

In many cases, there was agreement among the five data sources identified in Section 9.2.3. When discrepancies in functional feeding group (FFG) or habitat preference ("habit") assignments among the five primary data sources were identified, a final assignment was made based on the most prevalent assignment. In cases where there was no prevalent assignment, the workgroup examined why disagreements existed, flagged the taxon, and used best professional judgment to make the final assignment.

9.3 MULTIMETRIC INDEX DEVELOPMENT

9.3.1 Regional Multimetric Development

The same autecology and taxonomic resolution used in WSA was applied to the NRSA macroinvertebrate 300 fixed count data to calculate the community metrics used to calculate the MMI. In the WSA, a best ecoregional MMI was developed by summing the six metrics that performed best in that ecoregion (the national aggregate nine ecoregions). Each of the six metrics was scored on a 0–10 scale by interpolating metrics between a floor and ceiling value. The six metric 0-10 point scaled scores were then summed and normalized to a 0–100 scale by multiplying by 100/60 to calculate the final MMI. Details of this process are described in Stoddard *et al.* (2008).

The final metrics used in each ecoregion, metric direction, and floor and ceiling values are summarized in Table 9.1. Scoring equations are different depending on if the metric responds positively (high values good) or negatively (high values bad) with disturbance. For positive metrics, values above the ceiling get 10 points, and values below the floor get 0 points. For negative metrics, values above the ceiling get 0 points, and values below the floor get 10 points. The interpolation equations for scoring the 0-10 points for metrics between the floor and ceiling values are:

- Positive Metrics: Metric Points = 10*((metric value-floor)/(ceiling-floor))
- Negative Metrics: Metric Points = 10 * (1 - ((metric value-floor)/(ceiling-floor))).

The MMI used in the NRSA report is identical to the WSA MMI in terms of metrics and scoring. Based on NRSA revisit data, the MMI had a S:N ratio of 2.8 and a pooled standard deviation of 10.0 (out of 0–100).

Table 9.1 Six benthic community metrics, scoring direction, and floor and ceiling values used in calculating the NRSA and WSA MMI in each of the nine aggregate ecoregions.

Ecoregion	Direction	Metric	Floor	Ceiling
CPL	Negative	Non-Insect %Individuals	0.70	73.0
	Positive	Shannon Diversity	1.62	3.31
	Positive	Shredder Taxa Richness	1	9
	Positive	Clinger % Taxa Richness	14.3	54.8
	Positive	EPT Taxa Richness	1	17
	Negative	Tolerant %Taxa Richness	5.56	50.0
NAP	Positive	EPT % Taxa Richness	9.52	57.6
	Negative	%Individuals in Top 5 Taxa	37.2	76.2
	Positive	Scraper Taxa Richness	3	12
	Positive	Clinger % Taxa Richness	28.6	70.0
	Positive	EPT Taxa Richness	3	24
	Positive	PTV 0-5.9 % Taxa Richness	46.2	86.1
NPL	Positive	EPT % Taxa Richness	3.85	50.0
	Positive	Shannon Diversity	1.10	3.07
	Positive	Scraper Taxa Richness	1	6
	Negative	Burrower % Taxa Richness	6.45	35.3
	Positive	Ephemeroptera Taxa Richness	0	7
	Positive	PTV 0-5.9 Taxa Richness	4	28
SAP	Positive	Ephemeroptera % Taxa Richness	5.41	28.6
	Positive	Shannon Diversity	2.05	3.44
	Positive	Scraper Taxa Richness	3	12
	Negative	Burrower % Taxa Richness	3.45	25.0
	Positive	EPT Taxa Richness	5	25
	Negative	Tolerant % Taxa Richness	2.44	27.6
SPL	Positive	EPT % Individuals	0.67	66.0
	Positive	Shannon Diversity	1.16	3.27
	Positive	Scraper Taxa Richness	1	8
	Negative	Burrower % Taxa Richness	5.0	36.1
	Positive	EPT Taxa Richness	1	16
	Positive	Intolerant Taxa Richness	1	8
TPL	Positive	EPT % Individuals	0.67	80.3
	Positive	Shannon Diversity	1.41	3.17
	Positive	Scraper Taxa Richness	1	9
	Positive	Clinger Taxa Richness	3	20
	Positive	Ephemeroptera Taxa Richness	1	11
	Negative	PTV 8-9.9 %Taxa Richness	4.35	33.3
UMW	Negative	Chironomid %Taxa Richness	11.2	50.8
	Positive	Shannon Diversity	2.01	3.56
	Positive	Shredder Taxa Richness	3	10
	Negative	Burrower % Taxa Richness	3.77	28.6
	Positive	EPT Taxa Richness	4	22
	Negative	PTV 8-9.9 %Taxa Richness	2.51	29.5
WMT	Positive	EPT % Taxa Richness	18.5	62.9
	Negative	%Individuals in Top 5 Taxa	40.6	82.3
	Positive	Scraper Taxa Richness	1	8
	Positive	Clinger % Taxa Richness	27.0	69.6
	Positive	EPT Taxa Richness	6	23
	Negative	Tolerant %Taxa Richness	2.27	25
XER	Negative	Non-Insect %Individuals	3.33	36.0
	Negative	%Individuals in Top 5 Taxa	44.7	92.3
	Positive	Scraper Taxa Richness	0	7
	Positive	Clinger % Taxa Richness	15.8	65.8
	Positive	EPT Taxa Richness	1	18
	Negative	Tolerant %Taxa Richness	3.57	36.4

9.3.2 Modeling of MMI Condition Class Thresholds

Previous large-scale assessments have converted MMI scores into classes of assemblage condition by comparing those scores to the distribution of scores observed at least-disturbed reference sites. If a site's MMI score was less than the 5th percentile of the reference distribution, it was classified as in "poor" condition; scores between the 5th and 25th percentile were classified as "fair," and scores in the 25th percentile or higher were classified as "good." This approach assumes that the distribution of MMI scores at reference sites reflects an approximately equal, minimum level of human disturbance across those sites. But this assumption did not appear to be valid for some of the nine WSA regions, which was confirmed by state and regional parties at meetings to review the draft results.

For the WSA, the project team performed a principal components analysis (PCA) of the physical habitat and water chemistry variables (Total P, Total N, pH, Chloride, Sulfate, Turbidity, %Fine Substrate, Riparian Disturbance Index) that had originally been used to screen for biological reference sites as described in 3. The first principal component (Factor 1) of this PCA well represented a generalized gradient of human disturbance. MMI scores at the reference sites, however, were weakly, but significantly, related to this disturbance gradient in some of the aggregate ecoregions. Thus, MMI reference distributions from these regions may be biased downward, because they include somewhat disturbed sites which may have lower MMI scores. As part of the WSA, Herlihy *et al.* (2008) developed a process that used this PCA disturbance gradient to reduce the effects of disturbance on threshold values within the reference site population. The process uses multiple regression modeling to develop adjusted thresholds analogous to the 5th and 25th percentiles of reference sites in each ecoregion based on the slope of the MMI-disturbance relationship in each ecoregion.

These adjusted thresholds were used in the WSA but were based on a fairly small sample size of reference sites. To increase the sample size used in the regression model, the threshold adjustment process was rerun for NRSA using the original WSA reference sites plus the additional NRSA reference sites identified in section 2. As in the WSA analysis and other threshold setting, we used a 1.5*interquartile range (IQR) outlier screening test in each ecoregion to drop MMI outliers from the analysis (sites with values outside the range of Q1-1.5*IQR or Q3+1.5*IQR were dropped). This removed 6 sites from the analysis (all low; 3 in WMT, and 3 in XER). There were a grand total of 647 least-disturbed reference sites used for the threshold regression adjustment modeling and the resulting regression statistics for each ecoregion are shown in Table 9.2. The process for calculating these adjusted thresholds and fitting the regression model is detailed in Herlihy *et al.* (2008). Briefly, the process involves setting the goal for disturbance to the 25th percentile of the Factor 1 disturbance score for reference sites in each ecoregion. The ecoregion MMI value at that goal is predicted from the MMI-disturbance regression as:

$$\text{MMIpred} = (\text{GOAL} * \text{SLOPE}) + \text{INTERCEPT}.$$

Then the percentiles to be used as the adjusted thresholds are calculated assuming there is a normal distribution around this predicted mean using the RMSE of the regression model as the standard error,

Good-Fair 25th threshold = MMIpred - 0.675 * RMSE
Fair-Poor 5th threshold = MMIpred - 1.650 * RMSE.

The resulting adjusted MMI threshold values for the condition classes in each ecoregion used in the NRSA report are given in Table 9.3.

Table 9.2 MMI-Disturbance Regression Model Statistics Used for Setting Thresholds

Ecoregion	Number of Reference Sites	Factor 1 Goal*	Regression RMSE	Regression Slope	Regression Intercept
CPL	32	-0.1501	14.55	0	64.74
NAP	56	-0.5247	14.55	-7.257	61.06
NPL	65	0.8723	14.55	-14.95	79.66
SAP	64	-0.5531	14.55	-7.257	50.78
SPL	43	0.7637	14.55	-7.257	50.84
TPL	49	1.045	14.55	-7.257	57.75
UMW	39	-0.1138	14.55	0	46.74
WMT	209	-1.326	14.55	-7.257	50.27
XER	90	-0.4628	14.55	-7.257	63.44

* The 25th percentile of Factor 1 score was the "goal" on the PCA factor 1 disturbance gradient for hindcasting ecoregional thresholds.

Table 9.3 Threshold Values for the Nine Regional Benthic MMIs.

Ecoregion	Good Threshold *	Poor Threshold *
CPL	≥54.9	<40.7
NAP	≥55.0	<40.9
NPL	≥56.8	<42.6
SAP	≥45.0	<30.8
SPL	≥35.5	<21.3
TPL	≥40.3	<26.2
UMW	≥36.9	<22.7
WMT	≥50.1	<35.9
XER	≥57.0	<42.8

*Any site with an MMI score that was not good or poor was considered "fair."

9.4 Predicted O/E Modeling

In addition to the benthic macroinvertebrate MMI approach, predictive O/E modeling was used to assess benthic macroinvertebrate condition. The O/E model compares the observed benthic assemblage at a site to an expected assemblage derived from a population of reference sites. Stressors and anthropogenic impacts lead to a reduction in the number of taxa that are expected to be present under reference conditions. The predictive model approach is used by several states and is a primary assessment tool of Great Britain and Australia.

The O/E ratio predicted by the model for any site expresses the number of taxa found at that site (O), as a proportion of the number that would be expected (E) if the site was in least-disturbed condition. Ideally, a site in reference condition has O/E = 1.0. An O/E value of 0.70 indicates that 70% of the "expected" taxa at a site were actually observed at the site. This is interpreted as a 30% loss of taxa relative to the site's predicted reference condition. However, O/E values vary among reference sites themselves, around the idealized value of 1.0, because such sites rarely conform to an idealized reference condition, and because of model error and sampling variation. The standard deviation of O/E (Table 9.4) indicates the breadth of O/E variation at reference sites. Thus, the O/E value of an individual site should not be interpreted as (1 – taxa loss) without taking account of this variability in O/E. Individual O/E values are most reliably interpreted relative to the entire O/E distribution for reference sites.

A nationally distributed collection of reference sites was first identified, drawn from a pool of sites whose macroinvertebrates were sampled using EMAP protocols. This pool included only NRSA, WSA, EMAP-West, STAR-Hawkins, USGS NAWQA, and MAHA/MAIA sites. One hundred reference sites were set aside to validate the models, and the remaining reference sites were used to calibrate the models (Table 9.4). Each site contributed a single sampled macroinvertebrate assemblage to model calibration and validation. Each sampled macroinvertebrate assemblage comprising more than 300 identified individuals was randomly subsampled to yield 300 individuals. 300-count subsamples were used to build models and assess all NRSA sites.

The predictive modeling approach assumes that expected assemblages vary across reference sites throughout a region, due to natural (non-anthropogenic) environmental features such as geology, soil type, elevation, and precipitation. To model these effects, the approach first classifies reference sites based on similarities of their macroinvertebrate assemblages (Table 9.4). A random forest model is then built to predict the membership of any site in these classes, using natural environmental features as predictor variables (Table 9.4). The predicted occurrence probability of a reference taxon at a site is then predicted to be the weighted average of that taxon's occurrence frequencies in all reference site classes, using the site's predicted group membership probabilities in the classes as weights. Finally, E for any site is the sum, over a subset of reference taxa, of predicted taxon occurrence probabilities. O is the number of taxa in that subset that were observed to be present at the site. The subset of reference taxa used for any site was defined as those taxa with predicted occurrence probabilities exceeding 0.5 at that site.

Final predictive models performed better than corresponding null models (no adjustment for natural-factor effects), as judged by their smaller standard deviation of O/E across calibration sites (Table 9.4).

Similar to the IBI, two scaled approaches were used to develop the O/E model. A national model was initially developed to predict taxa loss at sites. Three models were developed for NRSA usage, together covering the contiguous USA (Table 9.4). The regional models performed better, and were used in the NRSA to predict taxa loss at the sites.

Table 9.4 Benthic Macroinvertebrate Predictive Models

Model Name	Eastern Highlands	Plains and Lowlands	West
Regions covered	NAP, SAP	CPL, UMW, TPL, NPL, SPL	WMT, XER
Number of calibration sites	297	241	659
Number of validation sites	31	21	48
Number of site classes	17	16	
Random Forest predictor variables	Predicted mean summer stream temperature, watershed area, watershed mean minimum annual temperature, predicted mean annual stream temperature, watershed mean annual temperature, watershed mean minimum precipitation	Predicted mean annual stream temperature, watershed mean date of last freeze, watershed mean soil permeability, watershed mean runoff, watershed maximum elevation	Watershed area, watershed mean annual temperature, watershed mean precipitation accumulation, predicted mean annual stream temperature, watershed mean maximum temperature, watershed mean elevation
Standard deviation of O/E at calibration sites: -- Predictive model -- Null model	 0.18 0.22	 0.23 0.26	 0.18 0.25

9.5 Literature Cited

Barbour, M. T., J. Gerritsen, G. E. Griffith, R. Frydenborg, E. McCarron, J. S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society* 15:185-211.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish*. EPA 841/B-99/002, Office of Water. US Environmental Protection Agency, Washington, DC.

Hawkins, C. P., R. H. Norris, J. N. Hogue, and J. W. Feminella. 2000. Development and Evaluation of Predictive Models for Measuring the Biological Integrity of Streams. *Ecological Applications* **10(5)**:1456-1477.

Herlihy, A.T., S.G. Paulsen, J. Van Sickle, J.L. Stoddard, C.P. Hawkins, and L.L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference condition approach at a continental scale. *Journal of the North American Benthological Society* 27:860-877.

Karr, J. R. and E. W. Chu. 2000. Sustaining living rivers. *Hydrobiologia* 422/423:1-14.

Stoddard, J.L., A.T. Herlihy, D.V. Peck, R.M. Hughes, T.R. Whittier, and E. Tarquinio. 2008. A process for creating multi-metric indices for large scale aquatic surveys. *Journal of the North American Benthological Society* 27:878-891.

Wright, J.F., 2000. An introduction to RIVPACS. In: Wright, J.F., Sutcliffe, D.W., Furse, M.T. (Eds.), *Assessing the Biological Quality of Fresh Waters: RIVPACS and Other Techniques*. Freshwater Biological Association, Ambleside, UK, pp. 1-24.

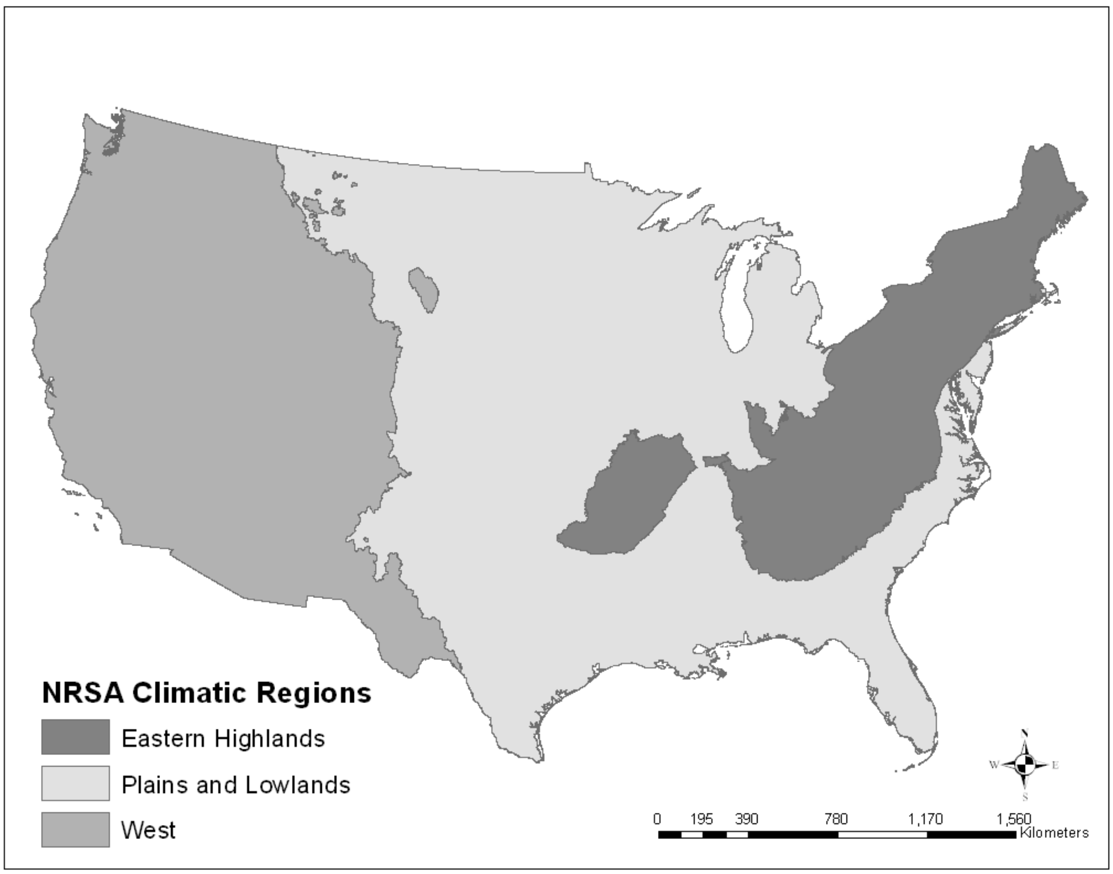
L.L. Yuan, C.P. Hawkins, J. Van Sickle. 2008. Effects of regionalization decisions on an O/E index for the US national assessment. *Journal of the North American Benthological Society* 27:892-905.

10 FISH ASSEMBLAGE

Fish assemblages in streams and rivers offer several unique advantages to assess ecological condition, based on their mobility, longevity, trophic relationships, and socioeconomic importance (Barbour *et al.* 1999, Roset *et al.* 2007). For fish assemblages, assessing ecological condition has generally been based the development and use of multimetric indices (MMIs), which are derivations of the original Index of Biotic Integrity (IBI) developed by Karr and others (Karr 1981, Karr *et al.* 1986, Karr 1991, 1999, Karr and Chu 2000). There are numerous examples of MMIs developed for fish assemblages in smaller streams (*e.g.*, Bramblett *et al.* 2005) as well as for larger rivers (Lyons *et al.* 2001, Emery *et al.* 2003, Mebane *et al.* 2003, Pearson *et al.* 2011).

As described below, EPA developed multimetric indices for fish assemblages (FMMIs) using the combined approach (modeling expected values of metrics). We developed separate FMMIs for each of the three climatic regions (Eastern Highlands, Plains and Lowlands, and West as shown in).

Figure 10.1 Aggregated Omernik ecoregions used for Fish MMI development.



10.1 METHODS

10.1.1 Field Methods

Collection methods for fish are described in the NRSA field operations manual (EPA 2009). Three variants of the basic sampling protocol (using electrofishing) were used depending on the width of the stream and whether or not it was wadeable. For wadeable streams less than 12.5 meters wide, 40 channel widths were sampled for fish. For larger wadeable streams (>12.5 meters wide), 500 meters or 20 channel widths were sampled (or a maximum length of 4 km, whichever was longer). For non-wadeable streams and rivers, at least 20 channel widths were sampled. At large wadeable and non-wadeable sites, sampling continued past the established reach length until 500 individuals were collected.

10.1.2 Counting, Taxonomy, and Autecology

Fish were tallied and identified in the field, then released alive unless used for fish tissue or vouchers. Local voucher specimens were collected if field identification could not be accomplished. Voucher samples for quality assurance (QA) purposes were also collected at 10% of sites for each taxonomist. These voucher samples were sent to an independent taxonomist to assess the accuracy of field identifications. In some cases, field identifications and count data were corrected based on the results from the QA vouchers. All names submitted on field data forms were reviewed and revised when necessary to create a listing of nationally consistent common and scientific names. Where possible, taxonomic names (common and scientific) were based on Nelson *et al.* (2004). The online database FishBase (http://www.fishbase.org) served as a secondary source of taxonomic names. In rare cases, a journal article of a newly described species was used. Collection maps for each taxon were prepared and compared to published maps in Page and Burr (1991). We identified A total of 631 unique taxa, excluding unknowns, hybrids, and amphibians.

Each taxon was characterized for a number of different autecological traits; assignments for each taxon were based on available sources of published information. Traits included habitat guilds (lotic habitat and temperature), trophic guild, reproductive guild, migration pattern, and tolerance to human disturbance. A list of all fish taxa and their associated autecological assignments are available as a tab-delimited data file that will be available as part of the published data for NRSA.

Assignments of native status were made at the scale of 8-digit Hydrologic Unit Codes (HUC). Published sources (USGS Nonindigenous Species database and NatureServe shapefiles of species distributions) were used as the basis for assigning a taxon collected at a particular site as being native or introduced.

Because fish collected at a site cannot always be confidently identified to species, there is a risk of inflating the number of species actually collected. For each sample, we reviewed the list of taxa to determine whether they were represented at more than one level of resolution. For example, if an "unknown catostomus" was collected, and it was the only representative of the genus at the site, we assigned it as distinct. If any other species of the genus were collected, then we considered the unknown as not distinct. We used only the number of distinct taxa in the sample to calculate any metrics based on species richness.

10.2 Fish Multimetric Index Development

EPA used a consistent process to develop a FMMI for each of the three climatic regions. For each candidate metric, we used data from least-disturbed sites to develop a predictive model of metric response based on the set of predictor variables. We constructed our predictive models using random forests (Breiman 2001, Cutler *et al.* 2007, Hawkins *et al.* 2010a). The model provided expected values for the metric (*i.e.*, under least-disturbed conditions) given the particular values of the predictor variables. This approach served to help remove the effects of natural gradients on metric response, which are often confounded with disturbance gradients when expected values for a metric are based solely on a set of regional least-disturbed sites (Hawkins *et al.* 2010a). The predictive model for each metric was then applied to the entire set of sites, and the residuals (deviation from predicted) were used as the response value for the metric. If a sufficient amount of natural variability was accounted for by the modeled metric response, it was considered further and the original metric response was not. We evaluated each metric (modeled or original) for its responsiveness to disturbance, *i.e.*, the ability to discern between least-disturbed (reference sites) and more highly disturbed sites (following Stoddard *et al.* 2008). We then selected metrics representing different dimensions of assemblage structure or function to include in the FMMI based on responsiveness and lack of correlation with other metrics, again following Whittier *et al.* (2007), Stoddard *et al.* (2008), and Van Sickle (2010).

10.2.1 Least Disturbed Sites for Fish

We modified the base list of least-disturbed sites determined for NRSA to eliminate additional fish samples that might not be representative of least-disturbed conditions (**Error! Reference source not found.**). In addition, we identified a random subset of least-disturbed sites (validation sites) within each climatic region and excluded them from model development and FMMI evaluation. We set aside 36 validation sites in the Eastern Highlands, 43 sites in the Plains and Lowlands, and 15 sites in the West region. Calculating FMMI scores for these validation sites should produce a distribution of FMMI scores that are similar to those least-disturbed sites that we used to develop the model.

Table 10.1 Criteria used to select least-disturbed sites for use in developing the FMMI

Criteria		
Start with the base set of NRSA least-disturbed sites		
Keep sites with fish samples		
Drop sites where seining was only method of sampling		
Drop sites with insufficient sampling		
• Wadeable: Less than 50% of reach and < 500 individuals collected		
• Large Wadeable: < 500 m and < 500 individuals collected		
• Boatable: < 20 channel widths sampled		
Drop sites with sufficient sampling where < 30 individuals were collected		
Drop sites with sufficient sampling where nonnative individuals comprised >50% of the total number of individuals collected		
Final Number of Least-disturbed Sites		
Eastern Highlands	155	
Plains and lowlands	178	
West	92	
Total	425	

10.2.2 Candidate Metrics

We calculated 162 candidate metrics representing the following dimensions of fish assemblage structure and function (following Stoddard *et al.* 2008):

- Nonnative species (ALIEN) based on presence in 8-digit USGS Hydrologic Units
- Taxonomic composition (COMP)
- Habitat guild (HABIT)
- Life history/migratory pattern (LIFE)
- Reproductive guild (REPRO)
- Species richness (RICH)
- Tolerance (TOLER) to anthropogenic alterations
- Trophic guild (TROPH)

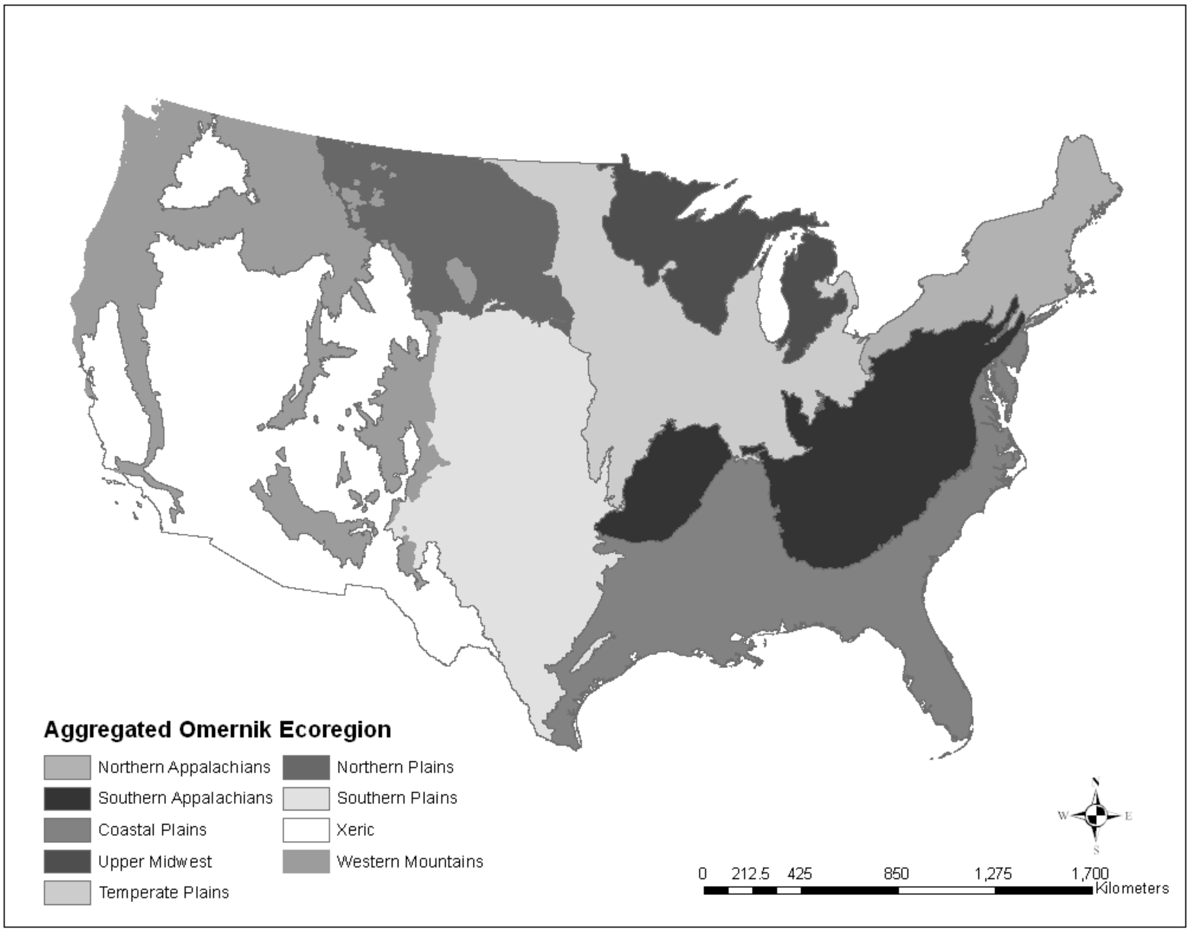
For nearly all metrics, three variants were derived based on all taxa in the sample and for only native taxa in the sample (metname vs. NAT_metname): one based on distinct taxa richness (metnameNTAX), one based on the percent of individuals in the sample (metnamePIND), and one based on the percent of distinct taxa in the sample (metnamePTAX). For some trophic metrics, additional variants were derived using only taxa that were not considered tolerant to disturbance (NTOLmetname). We included only those tolerance metrics based on sensitive and tolerant taxa, because the "intermediate tolerance" assignments included taxa with unknown tolerance.

10.2.3 Predictor Variables

A total of 55 predictor variables were initially provided for all NRSA sites (including handpicked sites). These data were provided to us by Dr. Charles Hawkins and his staff at the Western Center for Monitoring and Assessment of Freshwater Ecosystems, Utah State University, Logan, Utah. These variables (Model 10.1 at the end of the chapter (page 84)) represent the primary natural gradients that are believed to constrain the fish assemblage composition in the absence of human disturbance. The set of predictor variables included those related to watershed area and slope, elevation, latitude and longitude, air temperature, precipitation, and relative humidity. There were also model-derived estimates of flow, runoff, and predicted stream temperature. Many variables were estimated at the point level (at the site coordinates) and the watershed level (all values within a particular watershed were aggregated in some fashion). For the FMMI, we constrained the development of the index to only those sites that had both point and watershed level predictors, as we felt that fish might be more responsive to larger-scale natural driver variables than to more site-specific conditions.

In addition to the predictor variables provided, we calculated potential discharge (Q-POTENT_WS) as the product of runoff and watershed area, and included an aggregated ecoregion variable (AGGR_ECO9_2015; see Figure 10.2). For each region, we screened the set of predictor variables to eliminate redundant variables (keeping one of a set of highly correlated variables), those with large discrepancies in range between the set of least-disturbed sites and all other sites, and those that had missing values for some sites, as a complete set of predictor variables is required to construct the metric models. The final set of 64 predictor variables that we used to develop the FMMI are identified in boldface type in Table 10.2.

Figure 10.2 Nine aggregated Omernik ecoregions (Level III) used as predictor variable for FMMI development and for assigning condition based on FMMI scores in least-disturbed sites.



10.2.4 Random Forest Modeling

Random forests (RF) is a machine learning method that can be applied to produce predictive models for either classification or regression. The algorithm produces many independent regression trees, each being based on a bootstrapped subsample of sites (*i.e.*, sampled with replacement) and a randomly-selected subset of the available set of predictor variables. Each tree is built from a different subset of sites and a different subset of predictor variables. At each node of each tree, a recursive split is made based on the value of a single predictor variable that minimizes the sum of squares in the metric response value. The predicted metric response for each node is estimated as the sample mean of the sites in each group created by the split. This process is repeated at lower nodes using predictor variables that still remain available, and continues until either all predictor variables have been used, or where further splits no longer continue to minimize the sum of squares, or until a split results in a group with less than a minimal number of sites (*e.g.*, 5). The final results of each tree (*i.e.*, the final predicted mean metric response) are then aggregated across all trees by averaging to produce a final predictive model for the metric response based on the set of predictors (Breiman 2001, Prasad *et al.* 2006).

We used the R statistical package (version 3.1.2; R Development Core Team 2014), and the package randomForest (version 4.6-6; Liaw and Weiner 2002). We used the set of least-disturbed sites (minus those set aside for validation) to develop a predictive model for each candidate metric. For each regression tree, 1/3 of the predictor variables were selected at random, and a sample of approximately 0.632 times the number of least-disturbed sites was selected (with replacement). Sites not selected as part of the bootstrapped sample were used to estimate the "pseudo-R2" value assess the performance of the final predictive model. We set the number of independent regression trees to generate at 500. The purpose of the modeling effort was to minimize the potential bias in the FMMI due to the influence of natural environmental gradients on the response of individual candidate metrics (Pont *et al.* 2009, Hawkins *et al.* 2010a).

We followed Hawkins *et al.* (2010a) and Vander Laan and Hawkins (2014) in deciding whether or not to use the modeled responses for a particular metric. For metrics having a pseudo-R2 value >0.10, we applied the predictive model for the metric to the entire set of sites, and retained the residual values as the modeled metric response value. For metrics with pseudo-R2 values ≤0.10, we retained the original response value. Requiring that a predictive model needed to explain at least 10% of the variation in metric values kept the number of potential predictor variables to a reasonable level and reduced the chances of including predictor variables that had little if any effect on the predicted metric response values.

10.2.5 Final Metric Selection

We reduced the number of candidate metrics using a series of screening procedures, following Stoddard *et al.* (2008). The original metric response values were evaluated for range. Richness metrics with range < 4, and percentage metrics with a range < 10%, or with a 90th percentile value=0 were not considered further. To evaluate repeatability, we calculated Signal:Noise (S:N) for each metric following Kaufmann *et al.* (1999), to compare the variance observed at revisit sites (within the index period) with the total variance observed across all sites. Metrics with S:N values < 1.25 were not considered further. For modeled metrics, the S:N value was calculated after modeling to remove the effects of natural variability from the "signal", as suggested by Esselman *et al.* (2013). For both original and modeled metrics, the mean response values of the set of least-disturbed sites and the set of more highly disturbed sites were compared with two-sample t-tests (assuming unequal variances). Stoddard *et al.* (2008) present the advantages of using t values over other statistics as an indicator of metric responsiveness to disturbance. We did not consider metrics with t values < 1.73 further.

Metrics that passed these screens were then sorted by metric category and t-value. In cases where the "native only" variant was similar in t-value to the "all species" variant, only one was retained (usually the all species variant unless there was a sizable difference in the S:N value), and then both variants were retained in the final list of candidate metrics.

10.2.6 Metric Scoring

We rescaled response values for each of the final suite of metrics to a score ranging between 0 and 10. For "positive" metrics (those having higher values in least-disturbed sites) we used the 5th percentile of all sites to set the "floor" (below which a score of 0 was assigned), and the 95th percentile of least-disturbed sites to set the "ceiling" (above which a score of 10 was assigned) following Stoddard *et al.* (2008) and as described by Blocksom (2003). For "negative" metrics (where values were higher in the more disturbed sites), the floor was set at the 5th percentile of least-disturbed sites and the ceiling was set at the 95th percentile of all sites. We assigned a score to response values between the floor and ceiling using linear interpolation.

We summed the eight metric scores for each site to derive the FMMI score. We then multiplied the FMMI score by (10/number of metrics) to rescale the score to range between 0 and 100 points.

10.2.7 Selection of Final FMMIs

Using the final list of candidate metrics, we calculated tens of thousands of candidate FMMIs based on all possible combinations of eight metrics (one from each category), as recommended by Van Sickle (2010). This approach allowed us to evaluate not only the maximum pairwise correlation among a suite of metrics comprising an FMMI, but also the mean pairwise correlation of the suite itself. Indices having low mean correlations among pairs of metrics may perform better than an index containing component metrics selected to minimize redundancy based on a maximum allowable correlation coefficient (Van Sickle 2010).

For each candidate FMMI, we determined:

1. The F value based on comparing the set of least-disturbed vs. the set of more highly disturbed sites. We derived a t-value as $\sqrt{F}$.
    - The difference between the 25th percentile of the set of least-disturbed sites and the 75th percentile of the set of more highly disturbed sites. This value (SEPDIFF) is an estimate of the degree of overlap of the respective boxplots, which has been used as a way to evaluate metric and index performance (Barbour *et al.* 1996).

We calculated 14,400 candidate FMMIs for the Eastern Highlands, 77,760 candidate FMMIs for the Plains and Lowlands, and 34,560 candidate FMMIs for the West. To select the "best" FMMI from these candidates, we input the F values and the SEPDIFF values into a principal components analysis. We selected the FMMI that had the highest score for the first PCA axis for further evaluation. Combining the values for F and SEPDIFF into a single PCA axis score provided a simple, objective, and repeatable way to select an FMMI that had optimal responsiveness to anthropogenic alteration.

Table 10.2 presents the final suites of metrics for each of the three regional FMMIs selected. All FMMIs include eight metrics, so there was at least one metric from each category that passed all of the screens and could be retained in the final list of candidate metrics. The number of modeled metrics included in each FMMI ranges from four in the West to six in the Plains and Lowlands to seven in the Eastern Highlands.

We examined the variable importance plots of for each metric in the final set (produced with the R function varImpPlot) to identify the predictor variables that were the most influential in the model (*i.e.*, that decreased the mean squared error the most). We also examined partial dependence plots of the important predictor variables (produced with the R function partialPlot) to confirm the reasonableness of the relationships between predictor and metric. Table 10.3, Table 10.4, and Table 10.5 summarize the important predictors for each modeled metric included in each of the regional FMMIs.

Table 10.2 Suite of final metrics included in each regional FMMI

Variable names are in parenthesis. Metrics in bold are modeled metrics. Values of t are from comparisons of mean values of least-disturbed and more highly disturbed sites (validation sites have been excluded). Signal:Noise (S:N) values calculated based on Kaufmann et al. (1999), with validation sites excluded. Values for FMMIs are included for ease of comparison.

Metric class	Eastern Highlands FMMI	Plains and Lowlands FMMI	West MMI
Nonnative (ALIEN)	% individuals that are native (NAT_PIND) t=+3.43, S:N=2.9)	**No. of nonnative taxa (ALIENNTAX_RES) t=-3.33, S:N=1.7)**	No. of nonnative taxa (ALIENNTAX) t=-5.48, S:N=2.4
Taxonomic Composition (COMP)	**No. of native centrarchid taxa (NAT_CENTNTAX_RES) t=-5.32, S:N=1.6**	**% taxa that are cyprinids (CYPRPTAX_RES) t=+3.60, S:N=6.4)**	% individuals that are native round-bodied catostomids (NAT_RBCATOPIND) t=-+2.51, S:N=2.7
Habitat guild (HABIT)	% taxa that are native, intolerant rheophils (NAT_INTLRHEOPTAX) t=+7.08, S:N=23.3)	**Number of native rheophil taxa (NAT_RHEONTAX_RES) t=+8.42, S:N=2.7)**	**% taxa that are intolerant and lotic (INTLLOTPTAX_RES) t=+6.78, S:N=17.2)**
Migration Strategy (LIFE)	**% individuals that are native and migratory (NAT_MIGRPIND_RES) t=--2.74, S:N=1.9)**	% taxa that are intolerant and migratory (INTLMIGRPTAX) t=+2.69, S:N=5.3)	**% taxa that are migratory MIGRPTAX_RES t=+7.68, S:N=3.8)**
Reproductive Guild (REPRO)	**% individuals that are lithophilic spawners (LITHPIND_RES) t=+9.66, S:N=5.2)**	**% individuals that are lithophilic spawners (LITHPIND_RES) t=+8.14, S:N=11.8)**	**% taxa that are lithophilic spawners (LITHPTAX_RES) t=+6.35, S:N=4.7)**
Richness (RICH)	**% taxa that are not tolerant (NTOLPTAX_RES) t=-+7.74, S:N=2.0**	**Total Number of native taxa (NAT_TOTLNTAX_RES) t=+4.08, S:N=2.7**	% individuals that are not tolerant (NTOLPIND) t=7.84, S:N=31.5)
Tolerance (TOLER)	**% taxa that are tolerant (TOLRPTAX_RES) t=--8.59, S:N=4.4)**	**% taxa that are not tolerant (NTOLPTAX_RES) t=-4.92, S:N=2.3**	% taxa that are tolerant (TOLRPTAX) t=-6.78, S:N=3.4
Trophic guild (TROPH)	**% taxa that are native invertivores (NAT_INVPTAX_RES) t=+6.28, S:N=2.7**	**% taxa that are native herbivores (NAT_HERBPTAX_RES) t=+4.94, S:N=7.5**	**% individuals that are benthic invertivores (BENTINVPIND_RES) t=+5.22, S:N=7.3**
FMMI	t=+14.84; S:N=4.4	t=+11.68; S:N=8.0	t=+12.53; S:N=11.8

Table 10.3 Important predictor variables of modeled metrics included in the Eastern Highlands FMMI.

Metric	**Predictor Variable Names** (Predictor variables are listed in order of importance as determined by % increase in mean squared error. Total number of candidate predictors was 41.)	**Variable Descriptions**
NAT_PIND	Not Modeled	Not Applicable
NAT_CENTNTAX_RES	MAST, MSST, LON_DD83, TMAX_WS, LWSAREA_NARS, WTDH_WS, WDSUM_WS	Predicted stream temperature, longitude, maximum temperature, watershed area, mean water table height, sum of mean number of "wet days"
NAT_INTLRHEOPTAX	Not Modeled	Not Applicable
NAT_MIGRPIND_RES	LON_DD83, TMEANPW_WS, LAT_DD83, TMEANSY_WS TMEAN_WS, TMIN_WS LWSAREA_NARS, OMH_WS, TMAX_WS	Longitude, previous winter air temperature, latitude, mean air temperature for sampling year, mean air temperature, mean minimum temperature, watershed area, soil organic matter, mean maximum air temperature
LITHPIND_RES	RUNOFF_WS, ELVMAX_WS, AWCH_WS, TMEANSD_WS, MSST, PMAX_WS, LON_DD83, WDMAX_WS	Mean runoff in watershed, maximum elevation, mean SD of air temperature, predicted stream temperature, mean maximum precipitation, longitude, mean maximum number of "wet days"
NTOLPTAX_RES	MSST, TMIN_WS, LAT_DD83, PMAX_WS	Predicted stream temperature, mean minimum air temperature, latitude, mean maximum precipitation
TOLRPTAX_RES	LWSAREA_NARS, ELVMIN_WS, ELVMAX_WS, MSST, PSUMPY_PT	Watershed area, minimum and maximum elevation, predicted stream temperature, sum of monthly precipitation over previous 12 months at point
NAT_INVPTAX_RES	LON_DD83, LWSAREA_NARS, LAT_DD83, OMH_WS, WTDH_WS, TMEANSY_WS	Longitude, watershed area, latitude, soil organic matter, water table height, mean air temperature for sampling year

Table 10.4 Important predictor variables of modeled metrics included in the Plains and Lowlands FMMI

Metric	**Predictor Variable Names** (Predictor variables are listed in order of importance as determined by % increase in mean squared error. Total number of candidate predictors was 32.)	**Variable Descriptions**
ALIENNTAX_RES	ELVSD_WS, AGGR_ECO9_2015	Standard deviation of elevations in watershed, aggregated ecoregion
CYPRPTAX_RES	WDSUM_WS, PMIN_WS, XELEV, MSST, TMEAN_WS, LSTFRZ_WS, MAST	Sum of mean number of "wet days", mean minimum precipitation, elevation at X-site, predicted stream temperature, mean air temperature, mean date of last freeze, predicted mean annual stream temperature
NAT_RHEONTAX_RES	MAFLOWU, MAST, XELEV, PMIN_WS, MSST, OMH_WS WDSUM_WS	Predicted mean annual flow, predicted mean annual stream temperature, mean elevation, mean minimum precipitation, predicted mean summer stream temperature, soil organic matter, sum of mean number of "wet days"
INTLMIGRPTAX	Not Modeled	INTLMIGRPTAX
LITHPIND_RES	MSST, MAST, WSAREA_NARS, WDSUM_WS	Predicted mean summer stream temperature, predicted mean annual stream temperature, watershed area, sum of mean number of "wet days"
NAT_TOTLNTAX_RES	WTDH_PT, LON_DD83, MAFLOWU, ELVMEAN_WS, LWSAREA_NARS, MAST, WTDH_WS, XELEV, PMIN_WS	Mean water table height (point), longitude, predicted mean annual flow, mean elevation, watershed area, predicted mean annual stream temperature, mean water table height (watershed), elevation at X-site, mean minimum precipitation on the watershed
TOLRPTAX_RES	WDSUM_WS, LSTFRZ_PT, LON_DD83, MAST, TMAX_WS, BFI_WS, AGGR_ECO9_2015, TMEANSS_WS, MSST, XELEV, LSTFRZ_WS	Sum of mean number of "wet days", mean date of last freeze (point), longitude, predicted mean annual stream temperature, mean maximum air temperature, USGS base flow index, aggregated ecoregion, mean summer air temperature for sampling year, predicted mean summer stream temperature, elevation at X-site, mean date of last freeze
NAT_HERBPTAX_RES	TMEANPW_PT, TMIN_WS, AGGR_ECO9_2015, LAT_DD83, WTDH_WS, ELVSD_WS, ELVMEAN_WS	Mean air temperature for previous winter (point), mean minimum air temperature, aggregated ecoregion, latitude, mean water table height, standard deviation of elevation, mean elevation

Table 10.5 Important predictor variables of modeled metrics included in the West FMMI.

Metric	**Predictor Variable Names** (Predictor variables are listed in order of importance as determined by % increase in mean squared error. Total number of candidate predictors was 32.)	**Variable Descriptions**
ALIENNTAX	Not Modeled	Not Applicable
NAT_RBCATOPIND	Not Modeled	Not Applicable
INTLLOTPTAX_RES	LWSAREA_NARS, PSUM_WS, RHMEAN_WS, PSUM_PT, PMAX_WS, MSST, LON_DD83, Q_POTENT_WS	Watershed area, sum of mean monthly precipitation for watershed, mean relative humidity, sum of mean monthly precipitation at sampling point, mean maximum precipitation, predicted mean summer stream temperature, longitude, potential discharge
CYPRPTAX_RES	LWSAREA_NARS, WDMIN_WS, PMIN_WS, PMAX_WS	Watershed area, mean minimum number of "wet days", mean maximum precipitation
LITHPTAX_RES	RHMEAN_WS, XELEV, PMIN_WS, WDMIN_WS, WDSUM_WS, LWSAREA_NARS, TMEAN_PT, LON_DD83	Mean relative humidity, elevation at X-site, mean minimum precipitation, mean minimum number of "wet days", sum of mean number of "wet days", watershed area, mean air temperature at sampling point, longitude
NTOLPIND	Not Modeled	Not Applicable
TOLRPTAX	Not Modeled	Not Applicable
BENTINVPIND_RES	MSST, MAST, WSAREA_NARS, WDSUM_WS	Predicted mean summer stream temperature, predicted mean annual stream temperature, watershed area, sum of mean number of "wet days"

10.3 FMMI Performance

We evaluated the performance of the regional FMMIs in several ways (**Error! Reference source not found.**). Comparing the FMMI scores from the set of least-disturbed validation sites to those from the set of least-disturbed sites used to develop each FMMI confirmed that they were behaving as anticipated. For all three regional FMMIs, the mean values of the validation sites and sites used in modeling were not significantly different from those used to develop an FMMI, based on a two-sample t-test (assuming unequal variances).

We evaluated the responsiveness of the regional FMMIs by comparing FMMI scores of the set of least-disturbed sites to the set of more highly disturbed sites (Stoddard *et al.* 2008). Boxplots (Figure 10.3) and two-sample *t* tests (assuming unequal variances) showed that all FMMIs were highly responsive (**Error! Reference source not found.**). The responsiveness of each FMMI was much higher than the responsiveness observed with any of its component metrics (Table 10.2). The SEPDIFF values (**Error! Reference source not found.**; see Section 10.2.7) were positive for the

Eastern Highlands (+5.64) and the West (+6.52), and just slightly negative for the Plains and Lowlands (-0.34).

We estimated precision of the FMMI "models" by calculating the standard deviation of FMMI scores from all least-disturbed sites, after standardizing the scores to a mean of 0. The FMMIs all appear to be very precise, with standard deviation values ranging between 0.1 and 0.2 (**Error! Reference source not found.**). These values are comparable (or better) than many predictive models of taxa loss (Hawkins *et al.* 2010a).

We evaluated the repeatability of the regional FMMIs using a set of sites that were visited at least twice during the course of the NRSA project, typically two times in a single year (Kaufmann *et al.* 1999, Stoddard *et al.* 2008). We used a general linear model (PROC GLM, SAS version 9.2) to obtain estimates of among-site and within-site (from repeat visits) variability. PROC GLM was used because of the highly unbalanced design (only a small subset of sites had repeat visits). We used a nested model (sites within year) where both site and year were random effects. We estimated repeatability by deriving a "signal:noise" (S/N) ratio as $(F-1)/c$, where F is the F-statistic from the ANOVA, and c is an "average" sample size used to estimate the expected mean square (Sokal and Rohlf 1995). If all sites had repeat visits, c would equal 2 (Kaufmann *et al.* 1999). If no sites had repeat visits, c would equal 1. For the Eastern Highlands, $c = 1.0996$, while for the Plains and Lowlands $c = 1.0660$ and for the West $c = 1.0609$. Values of S:N suggest the regional MMIs are repeatable, with values between 4 (Eastern Highlands) and 11.8 (West; **Error! Reference source not found.**). When all regional FMMI scores are combined into a single "national" data set, the calculated S:N value is 10.0 (with c=1.0751).

Table 10.6 Performance statistics for the three regional FMMIs

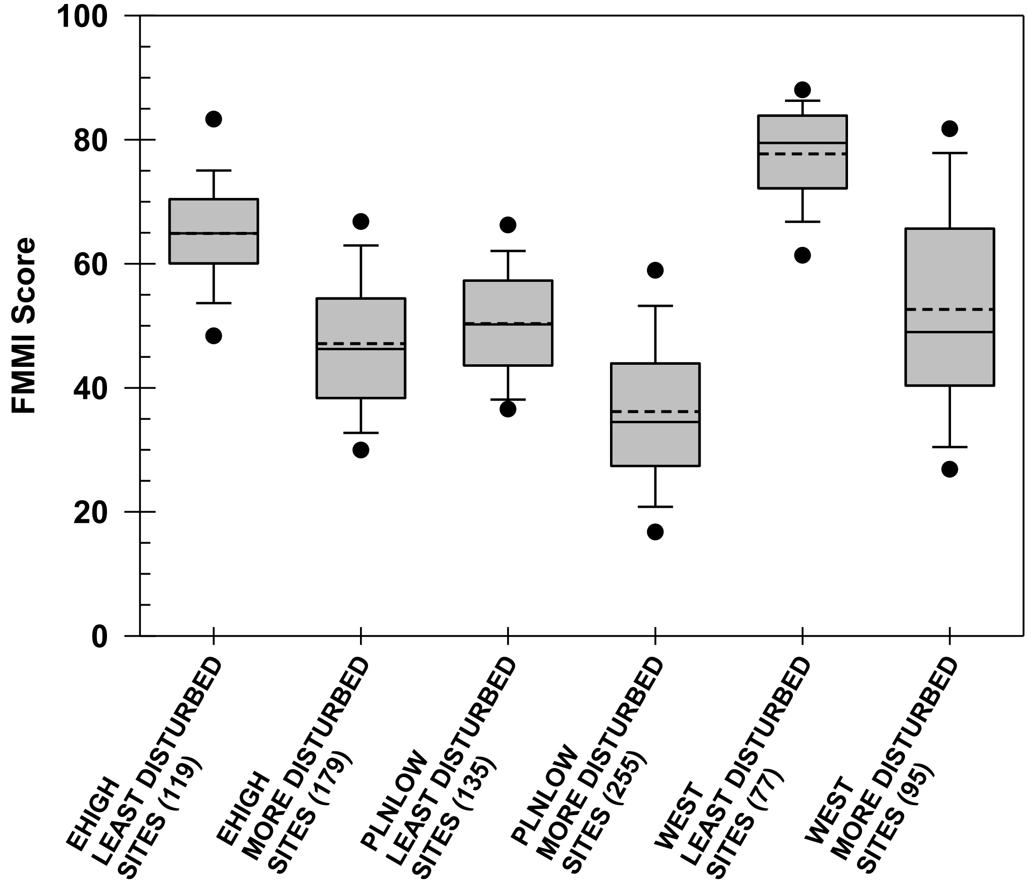
Performance Characteristic	Eastern Highlands FMMI	Plains and Lowlands FMMI	West FMMI
Validation least-disturbed sites vs. least-disturbed sites used in modeling and evaluating metrics and developing candidate FMMIs	t=0.93	t=1.45	t=-1.79
Least-disturbed sites vs. more highly disturbed sites	t=14.84	t=11.68	t=12.53
Difference between 25th percentile of least-disturbed sites and 75th percentile of more highly disturbed sites (*SEPDIFF*)	+5.64	-0.33	+6.52
FMMI model precision (standard deviation of mean-adjusted FMMI scores of least-disturbed sites)	0.144	0.184	0.105
Repeatability (Signal:Noise)	4.4	8.0	11.8

We felt it important to examine the performance of the component metrics across the range of stream sizes sampled for the NRSA. The potential exists for bias in the FMMI due to different fish species pools being available for larger rivers versus smaller streams. We used the set of least-disturbed sites to examine patterns in metric response values across Strahler order categories. For nearly all metrics, the distribution of metric response values and FMMI scores among stream order classes does not indicate a bias due to either stream size or sampling method (Figure 10.4 through Figure 10.6). There is a tendency for total number of native taxa to be low in 1st order sites, and higher in larger order sites (6th order and greater; Figure 10.5 and Figure 10.6).

We looked at the distribution of FMMI scores in least-disturbed sites across Strahler order categories as well. FMMI scores were similar across order categories, although higher order sites in the West tended to have lower FMMI scores (Figure 10.7).

Figure 10.3 Boxplots comparing FMMI scores of least-disturbed sites (index visits of calibration sites) to more highly disturbed sites

Sample sizes are in parentheses. Solid lines indicate median values. Dashed lines indicate mean values. Whiskers indicate the 10th and 90th percentiles. Dots indicate the 5th and 95th percentiles.



We felt it important to examine the performance of the component metrics across the range of stream sizes sampled for the NRSA. The potential exists for bias in the FMMI due to different fish species pools being available for larger rivers versus smaller streams. We used the set of least-disturbed sites to examine patterns in metric response values across Strahler order categories. For nearly all metrics, the distribution of metric response values and FMMI scores among stream order classes does not indicate a bias due to stream size (Figure 10.4 through Figure 10.6). There is a tendency for total number of native taxa to be low in 1st order sites, and higher in larger order sites (6th order and greater; Figure 10.5).

We looked at the distribution of FMMI scores in least-disturbed sites across Strahler order categories as well. FMMI scores were similar across order categories, although higher order sites in the West tended to have lower FMMI scores (Figure 10.7).

Differences across the size range might also result from the different sampling protocols that were used (wadeable, large wadeable, and boatable). We used the set of least-disturbed sites in each of the three climatic regions that were sampled for NRSA to identify potential effects of sampling method on the FMMI scores. Figure 10.9 suggests that the sampling protocol had little effect on the FMMI score.

Finally, we looked at the potential effects of stream temperature on the FMMIs. We classified least-disturbed sites into three temperature categories based on the predicted summer stream temperature (predictor variable MSST; see Table 10.3 through Table 10.5). We considered sites with MSST values ≤ 17 °C as "cold water," and sites with MSST > 20 °C as "warm water." We assigned "cool water" to sites between 17 and 20 °C. Figure 10.8 shows that FMMI scores were similar across the three temperature classes in the Eastern Highlands and the Plains and Lowlands climatic regions. In the West climatic region, warm water sites tended to have lower FMMI scores than either cold or cool water sites.

10.4 SITES WITH LOW FISH ABUNDANCE

The target population of streams and rivers for NRSA includes small headwater streams. Some very small streams may not contain fish even in the absence of human disturbance. We followed the approach described by McCormick *et al.* (2001) and used least-disturbed sites to estimate a watershed area below which the probability was high that no fish would be present (

). This approach uses the relationship between a set of four physical habitat variables that characterize habitat volume and the number of fish collected. This relationship defines a "habitat volume" value below which nearly all sites sampled were devoid of fish. Then this habitat volume index value is related to watershed area to determine the value below which streams are expected to be naturally fishless.

Figure 10.10 shows the results of this analysis. The value for the habitat volume index below which almost all sites are fishless is 0.42. When habitat volume is plotted against watershed area, this value corresponds to a watershed area of approximately 2 km^2. For sites with watershed areas less than 2 km^2 where no fish were collected, we did not report the FMMI score. Otherwise, we assigned an FMMI score of zero to sites with no fish collected.

Figure 10.4 Component metrics of the Eastern Highlands FMMI versus Strahler Order category, based on index visits of least-disturbed sites

Sample sizes are: 1st (25), 2nd (47), 3rd (37), 4th (17), 5th (16), and 6th (4). Whiskers represent 10th and 90th percentiles.

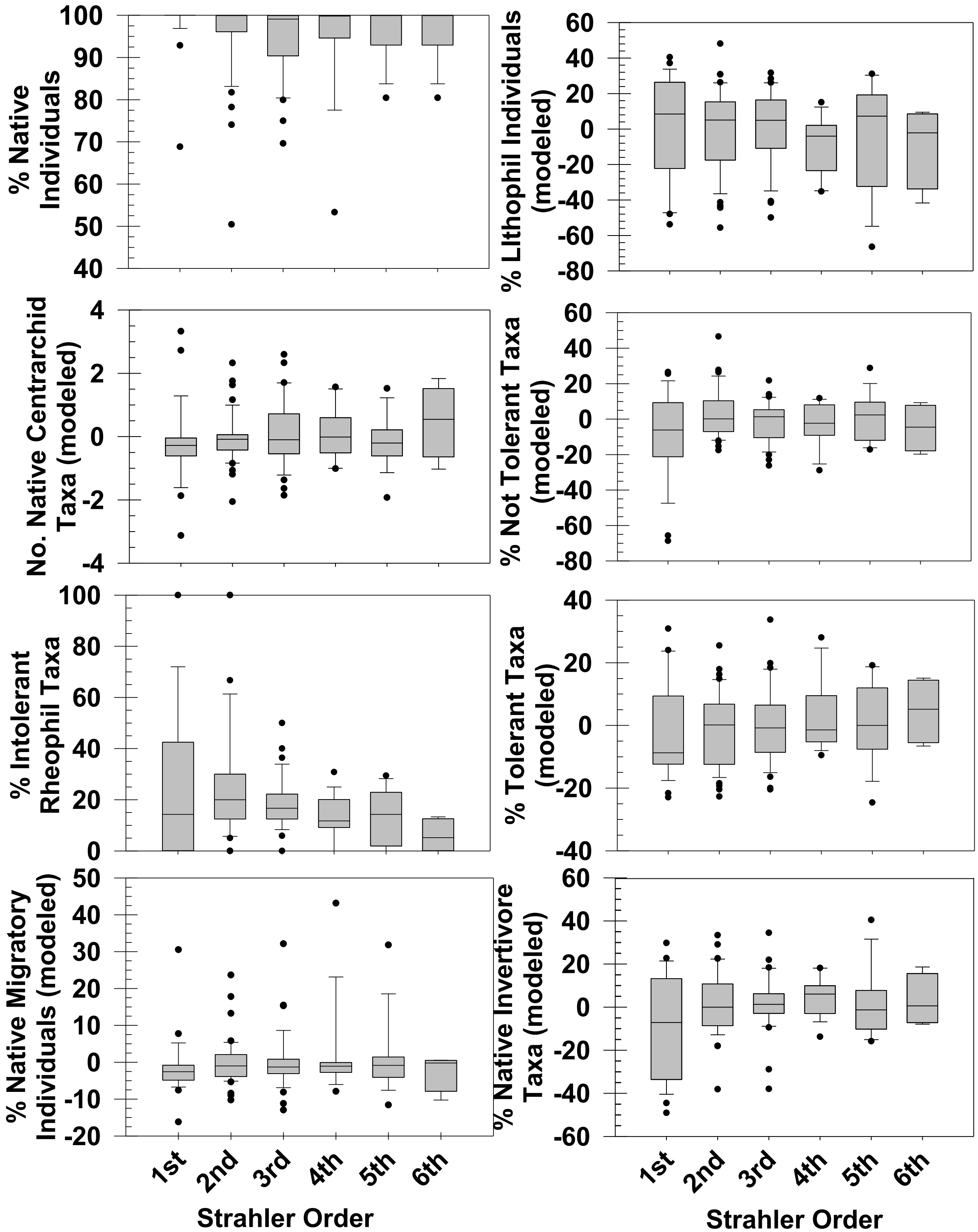


Figure 10.6 Component metrics of the West region FMMI versus Strahler Order category based on index visits of least-disturbed sites

Sample sizes are: 1st (11), 2nd (39), 3rd (34), 4th (31), 5th (25), 6th (21), and 7th and greater (6). Whiskers represent 10th and 90th percentiles.

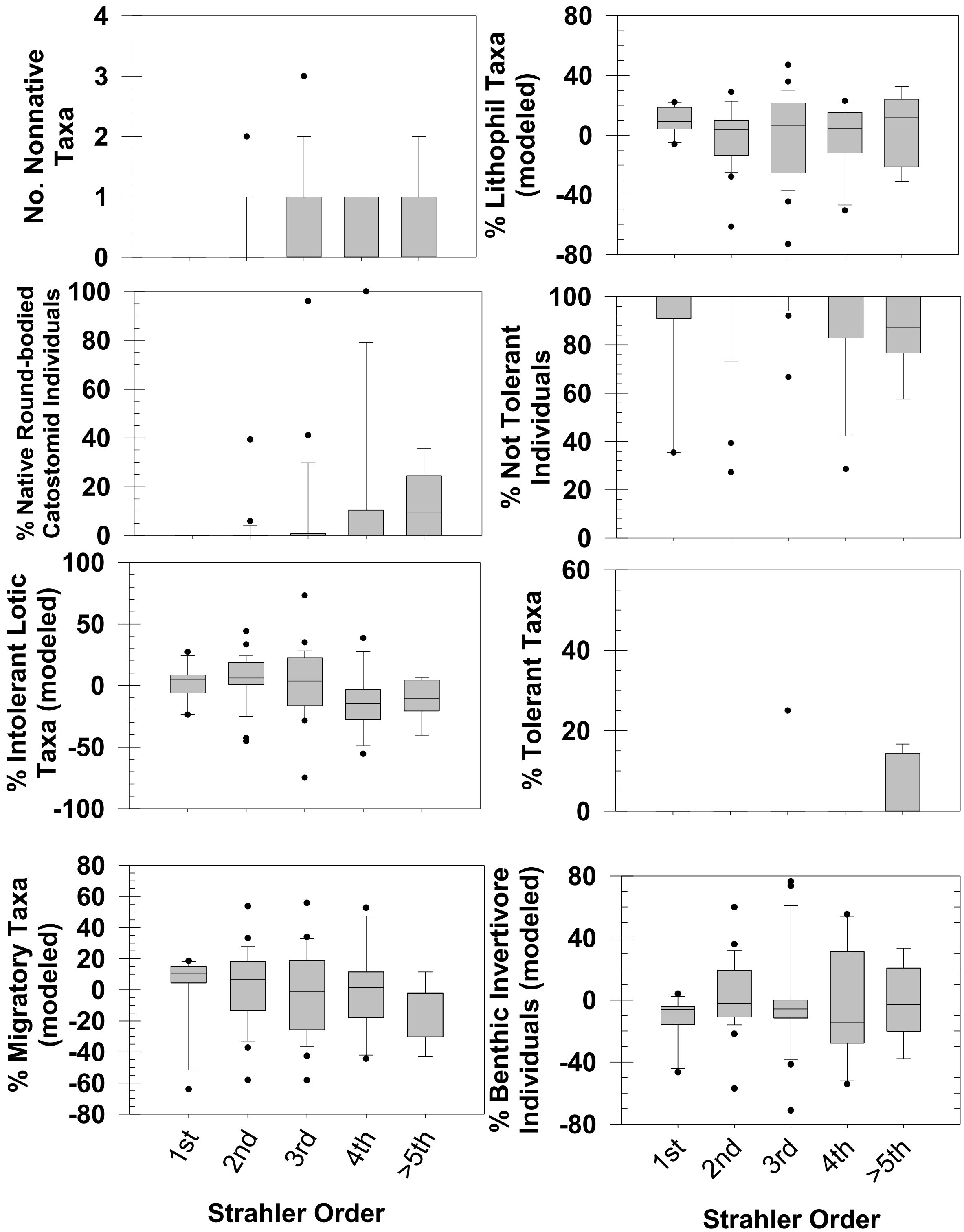


Figure 10.7 FMMI scores of least-disturbed sites versus Strahler order category

Whiskers represent 10th and 90th percentiles. Sample sizes are in parentheses.

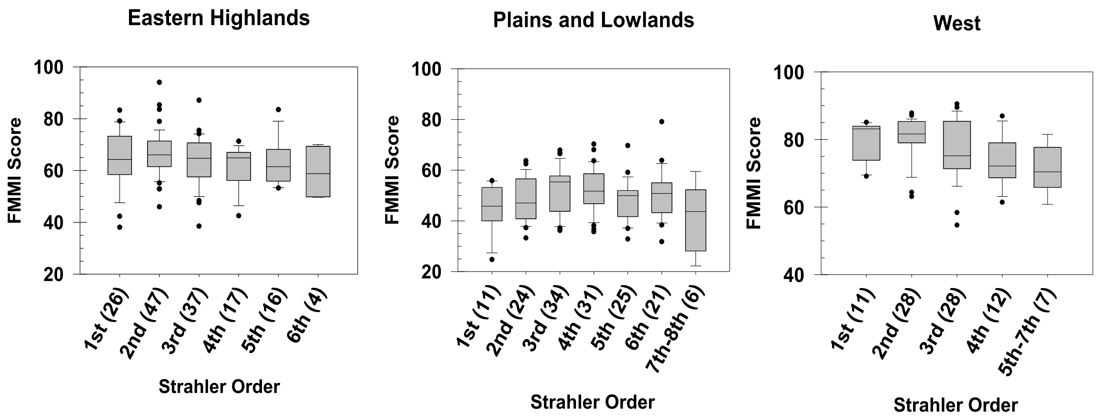


Figure 10.8 Relationship between FMMI scores and fish sampling protocol for index visits to least-disturbed sites

Sample sizes are in parentheses. Solid lines represent median values, dashed lines represent mean values, and whiskers represent the 10th and 90th percentiles.

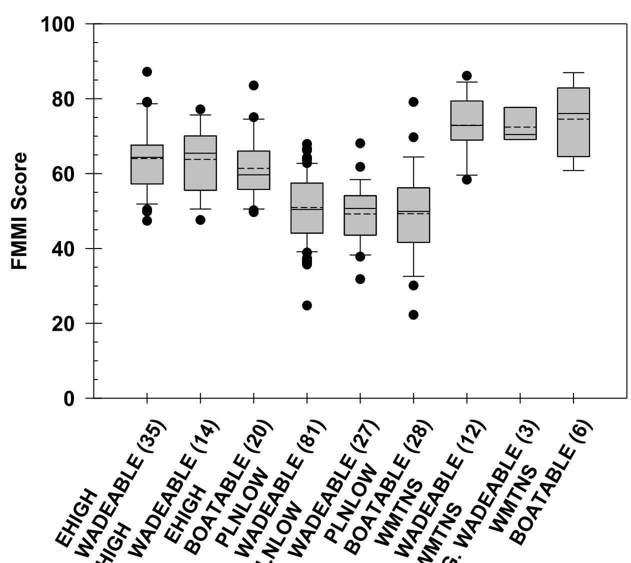


Figure 10.9 Relationship between FMMI scores and stream temperature class (based on predicted mean summer stream temperature [MSST]) for index visits to least-disturbed sites

Cold water sites had MSST ≤ 17 °C, and warm water sites has MSST > 20 °C. Cool water sites had MSST values between 17 and 20 °C. Sample sizes are in parentheses. Solid lines represent median values, dashed lines represent mean values, and whiskers represent the 10th and 90th percentiles

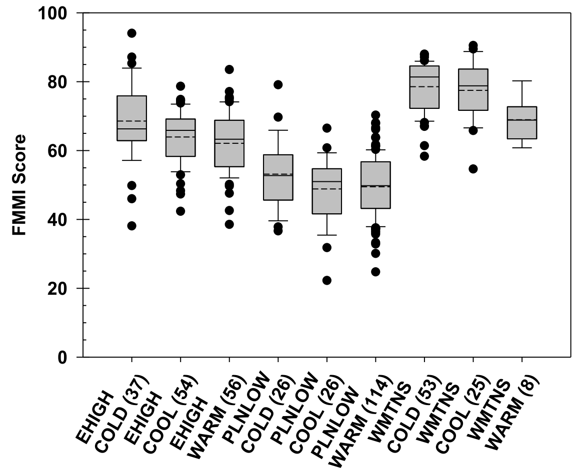
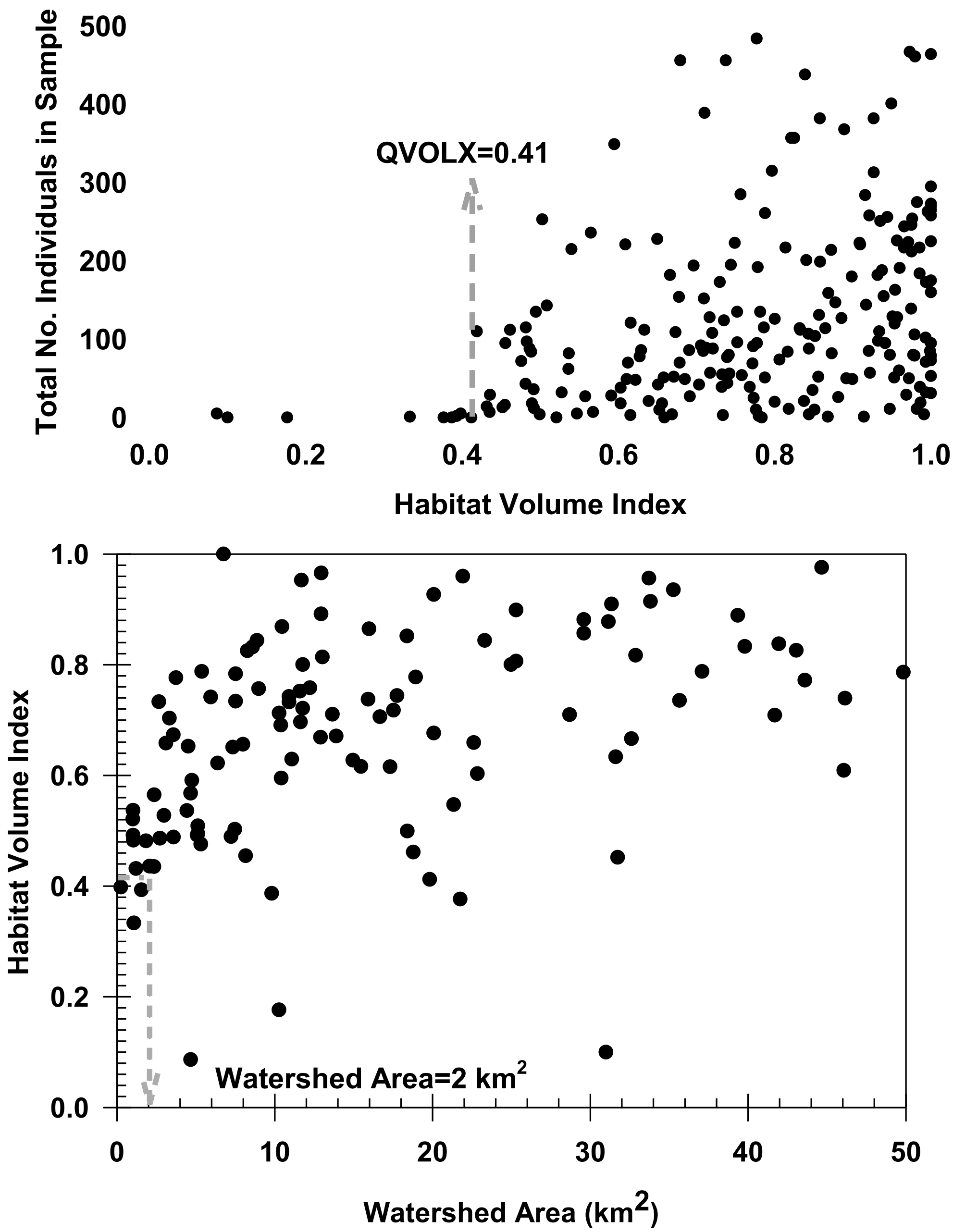


Table 10.7 Determining the minimum watershed area expected to reliably support the presence of fish (adapted from McCormick et al. 2001)

Variable names are from the NRSA database. Scores for each metric between the upper and lower criteria were estimated by linear interpolation.
SET OF SITES
Use least-disturbed sites only (RT_NRSA="R") that were sufficiently sampled for fish (SAMPLED_FISH="YES")to minimize effects of human disturbance **HABITAT VOLUME INDEX**
Percent of support reach length that is dry (PCT_DR)
If PCT_DR< 1%, score=1. If PCT-DR≥ 20%, then score=0.
Log_{10}[(mean wetted width x mean thalweg depth)+0.001] (LXWXD)
If LXWXD> 1, score=1. If LXWXD ≤ -1.4, then score=0
Residual pool depth (RP100)
If RP100 ≥20, then score=1. If RP100 ≤ 0, then score=0
Mean wetted width
If XWIDTH ≥ 6, then score=1. If XWIDTH=0, then score=0
HABITAT VOLUME INDEX=(PCT_DR score + LXWXD score +RP100 score + XWIDTH score)/4
PLOT NUMBER OF FISH COLLECTED (TOTLNIND) VS. HABITAT VOLUME INDEX (QVOLX)
Value for QVOLX below which most sites have no fish=0.41
PLOT HABITAT VOLUME INDEX VS. WATERSHED AREA (WSAREA_NARS)
QVOLX=0.41 corresponds to a watershed area of ~ 2 km^2

Figure 10.10 Relationship between small watershed size, reduced habitat volume, and number of fish collected based on index visits of least-disturbed sites (n=241).

Fish are not likely to be found in streams with watershed areas of <2 km².



10.5 Thresholds for Assigning Ecological Condition

For NRSA, ecological condition is based on the deviation from least-disturbed condition (Stoddard *et al.* 2006, Hawkins *et al.* 2010b). We used the same model-based approach as is described for the benthic macroinvertebrate MMI (Section 9.3.2; Herlihy *et al.* 2008) to develop thresholds for defining "good" condition (similar to least-disturbed) and "Poor" condition (substantially different from least-disturbed) in each of the nine aggregated ecoregions. Using the set of least-disturbed sites, we applied a principal components analysis (PCA) to the suite of stressor variables that were used to select least-disturbed sites to produce a surrogate variable of general disturbance (the site score for the first principal component [PC1]). We then related FMMI scores to PC1 using linear regression to determine the slope and intercept. We compared the slopes of the nine aggregated ecoregions, and used a common slope value for ecoregions with similar slope values. We then identified the intercept at the 25th percentile of the PC1 score based on the slope value. We computed the 5th and 25 percentiles of FMMI score at this value assuming a normal distribution and a constant RMSE value. Table 10.8 presents the regression statistics for each of the nine aggregated ecoregions.

Table 10.9 presents the threshold values developed from the regression models for each of the nine aggregated ecoregions. In addition, we include the percentiles of each set of least-disturbed sites (unmodeled) for comparison. The 25th percentile values were similar for all but the Temperate Plains ecoregion, where the model-derived percentile is nearly 8 points higher. The largest differences between model-derived and unmodeled 5th percentile values are seen in the Northern Appalachians (nearly 4 points higher) and the Coastal Plains (nearly 6 points higher).

We used the model-derived percentiles as thresholds to assign condition. Sites with scores between the two threshold values were assigned a condition class of "fair" (indeterminate). Thresholds for three aggregated ecoregions contain fewer than 30 least-disturbed sites, so the threshold values (particularly those for fair/poor) are based on a small number of sites.

10.6 Discussion

We constructed a FMMI for each climatic region that are responsive to disturbance and repeatable (Table 10.6 and Figure 10.3). Some improvements in the performance of the models might be gained from reducing the number of predictor variable to only those that seemed to be more important (Table 10.3 through Table 10.5), but we do not know if the additional effort required for this would yield a substantial improvement in the performance of the resulting regional FMMI.

The ability to calculate large numbers of candidate MMIs from a set of metrics that m*et al*l of our evaluation criteria is an improvement over stepwise selection of metrics based on correlations with metrics already selected. This approach helps to ensure that the best-performing MMI is selected. Incorporating the difference between the 25th percentile of least-disturbed and 75th percentile of more disturbed sites (SEPDIFF; Table 10.4) and the *F*-score provides a quick and reproducible way of selecting a final FMMI from the tens of thousands of candidate FMMIs that can be generated.

Table 10.8 Statistics for regression models of least-disturbed sites used to determine thresholds for assigning ecological condition for the FMMI

PC1 is the first principal component score. Aggregated ecoregions are shown in Table 10.1*Figure 10.1 and Figure 10.2.*

Aggregated Ecoregion	Slope	Intercept	Goal (P_{25} of PC1)	Predicted FMMI score at Goal	RMSE
Eastern Highlands FMMI					
Northern Appalachians (NAP)	-6.83139	58.11	-0.8917	64.20	8.332
Southern Appalachians (SAP)	-6.83139	60.76	-0.7958	66.20	8.332
Plains and Lowlands FMMI					
Coastal Plains (CPL)	0	50.02	-0.3949	50.02	8.332
Temperate Plains (TPL)	-6.83139	56.92	+0.4902	53.57	8.332
Northern Plains (NPL)	0	50.08	+0.8464	50.08	8.332
Southern Plains (SPL)	0	48.85	+0.6868	48.85	8.332
Upper Midwest (UMW)	0	51.90	-0.09948	51.90	8.332
West FMMI					
Western Mountains (WMT)	0	77.59	-1.5307	77.59	8.332
Xeric West (XER)	-6.83139	76.72	-0.6324	81.04	8.332

Table 10.9 Thresholds for assigning ecological condition based FMMI scores in least-disturbed sites.

Aggregated ecoregions are shown in Table 10.1*Figure 10.1 and Figure 10.2. Sample sizes are in parentheses. Values in bold are used to assign condition classes to NRSA sites.*

Aggregated Ecoregion	Good/Fair (25th percentile)		Fair/Poor (5th percentile)	
	Modeled	**Unmodeled**	**Modeled**	**Unmodeled**
Eastern Highlands				
Northern Appalachians (35)	**58.6**	58.6	**50.5**	46.0
Southern Appalachian (55)	**52.5**	58.3	**52.5**	49.7
Plains and Lowlands				
Coastal Plains (35)	**44.4**	44.8	**36.3**	30.2
Northern Plains (38)	**44.5**	42.5	**36.3**	34.7
Southern Plains (31)	**43.2**	42.7	**35.1**	36.2
Temperate Plains (22)	**47.9**	40.3	**39.8**	37.3
Upper Midwest (24)	**46.3**	43.8	**38.2**	37.2
West				
Western Mountains (59)	**72.0**	72.1	**63.8**	65.8
Xeric West (23)	**75.4**	71.2	**67.3**	63.1

The additional screening of least-disturbed sites for sampling sufficiency and for the presence of nonnative individuals (see Section 10.2.1) reduced sample sizes in all three climatic regions, particularly the West. The availability of least-disturbed sites from other data sources (*e.g.*, the EMAP Western Pilot Study) helped to offset the losses. Other than adjusting the criteria value for the presence of nonnative individuals to something less than 50% (or ignoring nonnatives in selecting least-disturbed sites), nothing can be done to mitigate the presence of nonnatives, but sampling sufficiency can be examined further. Hughes and Herlihy (2007), sampling a set of raftable rivers in Oregon, found that a shorter sampling distance (50 channel widths) and fewer number of individuals collected (120) was sufficient to develop a fish MMI score (Mebane *et al.* 2003) that was similar to the score developed from sampling a reach that was twice as long. Even this distance is much longer than the current minimum length defined for NRSA (20 channel widths [or a maximum length of 4 km] when the large wadeable and boatable fish sampling procedures are used). If the only objective of fish sampling is to develop some type of MMI, then it might be possible to evaluate the data collected to determine if fewer individuals collected would provide FMMIs with similar performance to those developed here. This could potentially increase the number of least-disturbed sites available for FMMI development, and reduce the number of sites that we cannot assess because of insufficient sampling effort.

While shorter reach lengths might be required to obtain an adequate sample of fish to develop an MMI, longer reach lengths might be required to collect the majority of species believed to be present in a particular stretch of stream or river. Hughes *et al.* (2002), again using a set of raftable rivers in Oregon, concluded that longer reach lengths (at least 85 channel widths) were needed to collect the majority of species believed to be present. More complete estimates of species richness are needed to develop robust predictive models of taxa presence (*e.g.*, Carlisle *et al.* 2008, Meador and Carlisle 2009). In 2008, fish were processed by subreach, which allows for some evaluation of the effect of increased reach length on the number of species obtained (*e.g.*, Cao 2001).

The different sampling procedures used do not appear to affect the FMMI (Figure 10.9). There is a tendency for larger least-disturbed sites in the West to have lower FMMI scores (Figure 10.7), although there were no strong patterns observed with stream size in the component metrics (Figure 10.6). The small number of larger least-disturbed streams, despite the number that were sampled in NRSA, constrains developing any alternatives to a national scale for that class alone. A national-scale index might be feasible for larger streams given the advances in available techniques used to construct and evaluate MMIs. There might also be alternative ways to define least-disturbed condition for larger streams (*e.g.*, Angradi *et al.* 2009, Esselman *et al.* 2013).

The FMMIs constructed for NRSA appear to perform well for use in regional-scale assessments of condition, but there are several constraints associated with them. Any MMI based wholly or in part upon the use of predictive models is inherently more difficult for others to try to reproduce or to use with new data (especially new least-disturbed sites), and requires sufficient computer resources and technical expertise to organize the appropriate data sets and modify the computer code. MMIs based on predictive models also require the set of predictor variables, many of which are derived using GIS and large spatial databases (*e.g.*, PRISM), which further requires sufficient computer resources and technical expertise. Alternative approaches exist for constructing MMIs using other types of methods to develop predictive models, such as boosted regression trees (Elith *et al.* 2008, Esselman *et al.* 2013) or algorithms based on information theory (Schoolmaster *et al.* 2012, 2013). These alternative approaches may result in better-performing FMMIs, but will have similar constraints. An important question is whether the performance that is achieved with a predictive-model based MMI is sufficiently improved over that achieved with a more traditional MMI to accept the impacts of the associated constraints. While this is the conclusion from some studies (*e.g.*, Cao *et al.* 2007 for diatom assemblages, Vander Laan and Hawkins 2014 for benthic macroinvertebrate assemblages), it is not clear if it applies to stream fish assemblages, The fish assemblage data from the NRSA provides a means to develop alternative MMIs using a variety of different approaches and compare their performance.

10.7 Literature Cited

Angradi, T. R., M. S. Pearson, T. M. Jicha, D. L. Taylor, D. W. Bolgrien, M. F. Moffett, K. A. Blocksom, and B. H. Hill. 2009. Using stressor gradients to determine reference expectations for great river fish assemblages. *Ecological Indicators* 9:748-764.

Barbour, M. T., J. Gerritsen, G. E. Griffith, R. Frydenborg, E. McCarron, J. S. White, and M. L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. *Journal of the North American Benthological Society* 15:185-211.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish*. EPA 841/B-99/002, Office of Water. US Environmental Protection Agency, Washington, DC.

Blocksom, K. A. 2003. A performance comparison of metric scoring methods for a multimetric index for Mid-Atlantic Highlands streams. *Environmental Management* 31:0670-0682.

Bramblett, R. G., T. R. Johnson, A. V. Zale, and D. G. Heggem. 2005. Development and evaluation of a fish assemblage index of biotic integrity for northwestern Great Plains streams. *Transactions of the American Fisheries Society* 134:624-640.

Breiman, L. 2001. Random forests. *Machine Learning* 45:5-32.

Cao, Y., D. P. Larsen, and R. M. Hughes. 2001. Evaluating sampling sufficiency in fish assemblage surveys— a similarity-based approach. *Canadian Journal of Fisheries and Aquatic Sciences* 58:1782-1793.

Cao, Y., C. P. Hawkins, J. Olson, and M. A. Kosterman. 2007. Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *Journal of the North American Benthological Society* 26:566-585.

Carlisle, D. M., C. P. Hawkins, M. R. Meador, M. Potapova, and J. Falcone. 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society* 27:16-37.

Cutler, D. R., T. C. Edwards, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. Random forests for classification in ecology. *Ecology* 88:2783-2792.

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802-813.

Emery, E. B., T. P. Simon, F. H. McCormick, P. L. Angermeier, J. E. DeShon, C. O. Yoder, R. E. Sanders, W. D. Pearson, G. D. Hickman, R. J. Reash, and J. A. Thomas. 2003. Development of a multimetric index for assessing the biological condition of the Ohio River. *Transactions of the American Fisheries Society* 132:791-808.

Esselman, P. C., D. M. Infante, L. Wang, A. R. Cooper, D. Wieferich, Y.-P. Tsang, D. J. Thornbrugh, and W. W. Taylor. 2013. Regional fish community indicators of landscape disturbance to catchments of the conterminous United States. *Ecological Indicators* 26:163-173.

EPA (United States Environmental Protection Agency). 2009. *National Rivers and Streams Assessment: Field Operations Manual.* EPA 841/B-04/004, Office of Water and Office of Environmental Information, US Environmental Protection Agency, Washington, DC.

Hawkins, C. P. 2006. Quantifying biological integrity by taxonomic completeness: its utility in regional and global assessments. *Ecological Applications* 16:1277-1294.

Hawkins, C. P., Y. Cao, and B. Roper. 2010a. Method of predicting reference condition biota affects the performance and interpretation of ecological indices. *Freshwater Biology* 55:1066-1085.

Hawkins, C. P., J. R. Olson, and R. A. Hill. 2010b. The reference condition: predicting benchmarks for ecological and water-quality assessments. *Journal of the North American Benthological Society* 29:312-343.

Herlihy, A. T., S. G. Paulsen, J. V. Sickle, J. L. Stoddard, C. P. Hawkins, and L. L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. *Journal of the North American Benthological Society* 27:860-877.

Hughes, R. M., P. R. Kaufmann, A. T. Herlihy, S. S. Intelmann, S. C. Corbett, M. C. Arbogast, and R. C. Hjort. 2002. Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. *North American Journal of Fisheries Management* 22:1229-1240.

Hughes, R. M., and A. T. Herlihy. 2007. Electrofishing distance needed to estimate consistent index of biotic integrity (IBI) scores in raftable Oregon rivers. *Transactions of the American Fisheries Society* 136:135-141.

Karr, J. R. 1981. Assessment of biotic integrity using fish communities. *Fisheries* 6:21-27.

Karr, J. R. 1991. Biological integrity: a long neglected aspect of water resource management. *Ecological Applications* 1:66-84.

Karr, J. R. 1999. Defining and measuring river health. Freshwater Biology 41:221-234. Karr, J. R. and E. W. Chu. 2000. Sustaining living rivers. *Hydrobiologia* 422/423:1-14.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing biological integrity in running waters: a method and its rationale. *Special Publication 5, Illinois Natural History Survey,* Champaign, Illinois.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. *Quantifying physical habitat in wadeable streams.* EPA 620/R-99/003, Office of Research and Development, US Environmental Protection Agency, Washington, DC.

Liaw, A. and M. Weiner. 2002. Classification and Regression by randomForest. R News 2:18-22. Lyons, J., R. R. Piette, and K. W. Niermeyer. 2001. Development, validation, and application of a fish-based index of biotic integrity for Wisconsin's large warmwater rivers. *Transactions of the American Fisheries Society* 130:1077-1094.

McCormick, F. H., R. M. Hughes, P. R. Kaufmann, D. V. Peck, J. L. Stoddard, and A. T. Herlihy. 2001. Development of an index of biotic integrity for the Mid-Atlantic Highlands region. *Transactions of the American Fisheries Society* 130:857-877.

Meador, M. R. and D. M. Carlisle. 2009. Predictive Models for Fish Assemblages in Eastern U.S. Streams: Implications for Assessing Biodiversity. *Transactions of the American Fisheries Society* 138:725-740.

Mebane, C. A., T. R. Maret, and R. M. Hughes. 2003. An Index of Biological Integrity (IBI) for Pacific Northwest Rivers. *Transactions of the American Fisheries Society* 132:239-261.

Nelson, J. S., E. J. Crossman, H. Espinosa-Pérez, L. T. Findley, C. R. Gilbert, R. K. Lea, and J. D. Williams. 2004. *Common and Scientific Names of Fishes from the United States Canada and Mexico.* Sixth edition. Special Publication 29, American Fisheries Society, Bethesda, Maryland.

Page, L. M. and B. M. Burr. 1991. *A field guide to freshwater fishes of North America north of Mexico.* Houghton Mifflin, Boston, Massachusetts.

Pearson, M. S., T. R. Angradi, D. W. Bolgrien, T. M. Jicha, D. L. Taylor, M. F. Moffett, and B. H. Hill. 2011. Multimetric Fish Indices for Midcontinent (USA) Great Rivers. *Transactions of the American Fisheries Society* 140:1547-1564.

Pont, D., R. M. Hughes, T. R. Whittier, and S. Schmutz. 2009. A Predictive Index of Biotic Integrity Model for Aquatic-Vertebrate Assemblages of Western U.S. Streams. *Transactions of the American Fisheries Society* 138:292-305.

Pont, D., B. Hugueny, and C. Rogers. 2007. Development of a fish-based index for the assessment of river health in Europe: the European Fish Index. *Fisheries Management and Ecology* 14:427-439.

Prasad, A. M., L. R. Iverson, and A. Liaw. 2006. Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems* 9:181-199.

R Development Core Team. 2014. R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

Roset, N., G. Grenouillet, D. Goffaux, D. Pont, and P. Kestemont. 2007. A review of existing fish assemblage indicators and methodologies. *Fisheries Management and Ecology* 14:393-405.

Schoolmaster, D. R., J. B. Grace, and E. W. Schweiger. 2012. A general theory of multimetric indices and their properties. *Methods in Ecology and Evolution* 3:773-781.

Schoolmaster Jr, D. R., J. B. Grace, E. W. Schweiger, G. R. Guntenspergen, B. R. Mitchell, K. M. Miller, and A. M. Little. 2013. An algorithmic and information-theoretic approach to multimetric index construction. *Ecological Indicators* 26:14-23.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry: *The principles and practice of statistics in biological research.* Third edition. W.H. Freemen and Company, New York.

Stoddard, J. L., A. T. Herlihy, D. V. Peck, R. M. Hughes, T. R. Whittier, and E. Tarquinio. 2008. A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society* 27:878-891.

Stoddard, J. L., D. P. Larsen, C. P. Hawkins, R. K. Johnson, and R. H. Norris. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16:1267-1276.

Van Sickle, J. 2010. Correlated metrics yield multimetric indices with inferior performance. *Transactions of the American Fisheries Society* 139:1802-1917.

Vander Laan, J. J., C. P. Hawkins, J. R. Olson, and R. A. Hill. 2013. Linking land use, in-stream stressors, and biological condition to infer causes of regional ecological impairment in streams. *Freshwater Science* 32:801-820.

Vander Laan, J. J., and C. P. Hawkins. 2014. Enhancing the performance and interpretation of freshwater biological indices: An application in arid zone streams. *Ecological Indicators* 36:470-482.

Whittier, T. R., R. M. Hughes, J. L. Stoddard, G. A. Lomnicky, D. V. Peck, and A. T. Herlihy. 2007. A structured approach for developing indices of biotic integrity--three examples from western streams and rivers in the USA. *Transactions of the American Fisheries Society* 136:718-73.

Model 10.1 (below) presents the initial list of predictor variables that were considered for use in developing predictive models of metric response (Section 10.2.4). The final set of predictor variables that were included in models for at least one of the three regions are indicated in boldface type.

Model 10.1 Predictor Variables Considered for Developing Predictive Models of Metric Responses

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
AGGR_ECO3_2015	Three-level aggregated Omernik ecoregion	Three aggregated Omernik Level III ecoregions based on revisions to ecoregion boundaries as of April 2013, when all Level IV boundaries were defined.
AGGR_ECO9_2015 (PLNLOW)	Nine-level aggregated Omernik ecoregion (9-level)	Nine aggregated Omernik Level III ecoregions based on revisions to ecoregion boundaries as of 2013, when all Level IV boundaries were defined.
AWCH_PT	Available water capacity of soil (sampling point)	Mean high values of available water capacity of soils (fraction) at sampling point derived from State Soil Geographic (STATSGO) Database.
AWCH_WS (EHIGH, PLNLOW)	Available water capacity of soil (watershed)	Mean of the high values of available water capacity (fraction) of soils for the watershed derived from the State Soil Geographic (STATSGO) Database.
BDH_PT	Soil bulk density (sampling point)	Mean high values of soil bulk density of soil types at the sampling point (g/cm^3) from State Soil Geographic (STATSGO) Database.
BDH_WS (EHIGH, PLNLOW, WMTNS)	Soil bulk density (watershed)	Watershed mean of the high values of soil bulk density (g/cm^3) of soils from the State Soil Geographic (STATSGO) Database.
BFI_WS (EHIGH, PLNLOW, WMTNS)	Base flow index for watershed	Base-flow index calculated from the USGS base-flow raster available at: (http://water.usgs.gov/GIS/metadata/usgswrd/XML/bfi48grd.xml). This is the watershed average of the grid as recommended by the USGS to estimate BFI at a sample site.
CUMDRAINAG	Cumulative drainage area (km^2)	Cumulative drainage area at bottom of NHD flowline
DOM_GEOL	Dominant bedrock geology type (watershed)	Geology type with largest percent coverage within the watershed derived from a simplified version of Reed & Bush (2001) - Generalized Geologic Map of the Conterminous United States. See GEOL_PT for codes identifying geology types.
ECO3_PT	Ecoregion	Omernik level III ecoregion at sampling point
ECO4_PT	Ecoregion	Omernik Level IV ecoregion at sampling point
ECOFW_PT	Ecoregion	Freshwater ecoregion at the sampling site (http://www.feow.org).
ELVCV_PT	Elevation	Coefficient of variation of elevations within a radius of 5 digital elevation model cells (30 x 30 meter resolution) of the sample site.

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
ELVMAX_WS (EHIGH, WMTNS)	Maximum elevation in watershed (m)	Derived from National Elevation Dataset.
ELVMEAN_WS (EHIGH, PLNLOW)	Mean elevation in watershed (m)	Derived from National Elevation Dataset.
ELVMIN_WS (EHIGH)	Mean elevation in watershed (m)	Derived from the National Elevation Dataset.
ELVSD_WS (PLNLOW)	Standard deviation of elevations across the watershed (m)	Derived from National Elevation Dataset.
FSTFRZ_PT	Day of first freeze at sampling point (day of year))	GIS raster calculated as $\sum x_i / 30$, where x_i = the modeled day of year (1-365) of first freeze (temperature ≤ 0°C). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
FSTFRZ_WS (EHIGH, PLNLOW)	Mean day of first freeze for watershed (day of year)	Mean of all FstFrz values within the watershed upstream of the sampling point ($\sum FstFrz_i / n$, where i = each of n pixels within the watershed).
GEOL_PT	Dominant bedrock geology type at sampling point	Derived from a simplified version of Reed and Bush (2001) - Generalized Geologic Map of the Conterminous United States. Codes identifying geology types: MAFUL (Mafic-ultramafic), QTRNRY (Quaternary), GNEISS (Gneiss), GRANITIC (Granitic), SDMNTRY (Sedimentary), VOLCANIC (Volcanic).
GIS_LAT	Adjusted latitude for sampling point (decimal degrees)	GIS-corrected latitude (to NHD flowline) used for watershed delineation.
GIS_LONG	Adjusted latitude for sampling point (decimal degrees)	GIS-corrected longitude (to NHD flowline) used for watershed delineation.
GNEISS	Percent of bedrock geology in the watershed classified as gneiss forms	Derived from a simplified version of Reed and Bush (2001) - Generalized Geologic Map of the Conterminous United States.
GRANITIC	Percent of bedrock geology in the watershed classified as granitic forms	Derived from a simplified version of Reed and Bush (2001) - Generalized Geologic Map of the Conterminous United States.
HECTARES	Watershed Area	Watershed area in hectares as calculated by Utah State University.
HUC8_CAT	Hydrologic Unit Code	8-digit HUC (hydrologic unit code) for the USGS defined sub-catchment from which sample was collected (http://water.usgs.gov/GIS/huc.html).

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
HYDR_PT	Modeled discharge at sampling point (cfs)	GIS raster calculated as (MIN[x_i]) / (MAX[x_i]), where x_i = mean monthly discharge for month i for the period of record and x_i includes ≥ 12 months of record. Values were calculated for each of 9,941 USGS gauging stations in the western USA and values for unmeasured locations were interpolated using inverse-distance-squared weighting of the 12 closest gauging stations within 100 km. Each interpolated value represents a 4 x 4 km cell.
HYDR_WS	Mean modeled discharge for watershed (cfs)	Mean of all HYDR values within the watershed upstream of the sampling point (HYDR[i] / n, where i = each of n pixels within the watershed).
INCRFLOWU (EHIGH)	Incremental flow (cfs)	Value for flowline as computed by the Unit Runoff Method
KFCT_PT (EHIGH, WMTNS)	Soil erodibility factor (sampling point)	Soil erodibility factor (no units) of soils at sampling point from the State Soil Geographic (STATSGO) Database.
KFCT_WS (EHIGH, PLNLOW)	Soil erodibility factor (watershed)	Watershed mean of the soil erodibility factor (no units) of soils from the State Soil Geographic (STATSGO) Database.
LAT_DD83 (EHIGH, PLNLOW, WMTNS)	Site latitude (decimal degrees)	Latitude (NAD83 datum) from NRSA integrated design file.
LON_DD83 (EHIGH, PLNLOW, WMTNS)	Site longitude(decimal degrees)	Longitude (NAD83 datum) from NRSA integrated design file.
LSTFRZ_PT	Day of last freeze (sampling point)	GIS raster calculated $\sum x_i$ / 30, where x_i = the modeled day of year (1-365) of last freeze (temperature ≤ 0°C). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
LSTFRZ_WS (EHIGH, WMTNS)	Day of last freeze (watershed)	Mean of all LstFrz values within the watershed upstream of the sampling point ($\sum LstFrz_i$ / n, where i = each of n pixels within the watershed).
LWSAREA_NARS (EHIGH, PLNLOW, WMTNS)	Log-transformed watershed area upstream of sampling point (km^2)	Log_{10}(WSAREA_NARS). Watershed area calculated from sampling point by GIS staff at EPA-Corvallis in 2015. Note values may be different from those calculated by EPA-Duluth.
MAFLOWU (PLNLOW, WMTNS)	Mean annual flow (cfs; unit runoff method))	Value at bottom of flowline as computed by Unit Runoff Method.
MAFLOWV	Mean annual flow (cfs; Vogel method)	Value at bottom of flowline as computed by Vogel Method.

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
MAFUL	Percent bedrock geology classified as mafic or ultramafic forms	Derived from a simplified version of Reed and Bush (2001) - Generalized Geologic Map of the Conterminous United States.
MAST (EHIGH, PLNLOW, WMTNS)	Predicted mean annual stream temperature	Defined as the average of daily USGS measured stream temperature (ST) values over all 12 months. The model was based on 571 reference condition USGS ST sites distributed across the conterminous USA. The model was developed using random forests and a suite of GIS-derived predictors, such as PRISM climate matched to specific years of ST records, BFI_WS (see above), soils, and topography to develop the model.
MAVELU (EHIGH, WMTNS)	Mean annual velocity (fps) based on MAFLOWU	Value at bottom of flowline as computed by Jobson Method (1996) using the flow in MAFlowU.
MAVELV	Mean annual velocity (fps) based on MAFLOWV	Mean Annual Velocity (fps) at bottom of flowline as computed by Jobson Method (1996) using the flow in MAFlowV.
MSST (EHIGH, PLNLOW, WMTNS)	Predicted mean summer stream temperature (°C)	Defined as the average of daily USGS measured stream temperature (ST) values during July and August. The model was based on 571 reference condition USGS ST sites distributed across the conterminous USA. The model was developed using random forests and a suite of GIS-derived predictors, such as PRISM climate matched to specific years of ST records, BFI_WS (see above), soils, and topography to develop the model. The model explained 87% of the variation in MSST and had a RMSE of 1.9 °C.
MWST (EHIGH, WMTNS)	Predicted mean winter stream temperature (°C)	Defined as the average of daily USGS measured stream temperature (ST) values during January and February. The model was based on 484 reference condition USGS ST sites distributed across the conterminous USA. The model was developed using random forests and a suite of GIS-derived predictors, such as PRISM climate matched to specific years of ST records, BFI_WS (see above), soils, and topography to develop the model. The model explained 89% of the variation in MWST and had a RMSE of 1.4 °C.
OMH_PT	Soil organic matter content (high value; % by weight) for sampling point	Derived from State Soil Geographic (STATSGO) Database.
OMH_WS (EHIGH, PLNLOW, WMTNS)	Soil organic matter content (mean of high values; % by weight) for watershed	Derived from State Soil Geographic (STATSGO) Database.

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
PCVPY_PT	Variability in monthly precipitation totals (mm) over the previous year (sampling point)	GIS raster calculated as STANDARD DEVIATION[x_i] / MEAN[x_i], where x_i = the modeled total precipitation (mm) for month i (5-12) of the previous year and month i (1-4) of the sampling year. This value estimates the variability in precipitation for the 12 months prior to the field sampling season. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
PCVPY_WS	Variability in monthly precipitation totals (mm) over the previous year (watershed)	Mean of all PcvPY values within the watershed upstream of the sampling point ($\sum PcvPY_i$ / n, where i = each of n pixels within the watershed).
PMAX_PT	Mean maximum monthly precipitation (mm; sampling point)	GIS raster calculated as $\sum$(MAX[x_i]) / 30, where x_i = the modeled total precipitation (mm) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell (http://www.prism.oregonstate.edu).
PMAX_WS (EHIGH, PLNLOW, WMTNS)	Mean maximum monthly precipitation (mm; watershed)	Mean of all Pmax values within the watershed upstream of the sampling point ($\sum Pmax_i$ / n, where i = each of n pixels within the watershed).
PMIN_PT	Mean minimum monthly precipitation (mm; sampling point)	GIS raster calculated as $\sum$MIN[x_i] / 30, where x_i = the modeled total precipitation (mm) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell (http://www.prism.oregonstate.edu).
PMIN_WS (EHIGH, PLNLOW, WMTNS)	Mean minimum monthly precipitation (mm; watershed)	Mean of all Pmin values within the watershed upstream of the sampling point ($\sum Pmin_i$ / n, where i = each of n pixels within the watershed).
PRMH_PT	Mean soil permeability (high values; in/h) at sampling point	Sampling point mean of the high values of permeability (inches/hour) of soils from the State Soil Geographic (STATSGO) Database.
PRMH_WS	Mean soil permeability (high values; in/h) for watershed	Watershed mean of the high values of permeability (inches/hour) of soils from the State Soil Geographic (STATSGO) Database.
PSUM_PT (WMTNS)	Mean annual precipitation (mm) at sampling point)	Annual sum of the predicted mean monthly precipitation (mm) derived from the PRISM data for the sampling site. Calculated as $\sum X_i$, where X_i = the predicted mean precipitation for month i (1-12) derived from 29 years of record (1961-1990).
PSUM_WS (EHIGH, WMTNS)	Mean annual precipitation (mm) for watershed	Annual sum of the predicted mean monthly precipitation (mm) derived from the PRISM data for the watershed. Calculated as $\sum X_i$, where X_i = the predicted mean precipitation for month i (1-12) derived from 29 years of record (1961-1990).

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
PSUMPY_PT (EHIGH, WMTNS)	Mean annual precipitation (mm) for previous year at sampling point	GIS raster calculated as $\sum x_i$, where x_i = the modeled total precipitation (mm) for month i (5-12) of the previous year and month i (1-4) of the sampling year. This value estimates the total precipitation for the 12 months prior to the field sampling season. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
PSUMPY_WS	Mean annual precipitation (mm) for previous year for watershed	Mean of all PsumPY values within the watershed upstream of the sampling point ($\sum(PsumPY_i)$ / n, where i = each of n pixels within the watershed), where PSUMPY=GIS raster calculated as $\sum(x_i)$, where x_i = the modeled total precipitation (mm) for month i (5-12) of the previous year and month i (1-4) of the sampling year. This value estimates the total precipitation for the 12 months prior to the field sampling season. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
Q_POTENT_WS (WMTNS)	Potential discharge (m^3)	Calculated as the product of runoff (mm) and watershed area (km^2)
QTRNRY	Percent of bedrock geology in the watershed classified as Quarternary forms	Derived from a simplified version of Reed and Bush (2001) - Generalized Geologic Map of the Conterminous United States.
RDH_PT (EHIGH, WMTNS)	Depth to bedrock (inches)at sampling point	High value of depth to bedrock (inches) of soils at the sampling point, derived from the State Soil Geographic (STATSGO) database.
RDH_WS	Mean depth to bedrock (inches)for watershed	Watershed mean of the high values of depth to bedrock (inches) of soils from the State Soil Geographic (STATSGO) Database.
RHMEAN_PT (EHIGH)	Mean monthly relative humidity (%) at sampling point	GIS raster calculated as ($\sum x_i$ / 12) / 30, where x_i = the modeled mean relative humidity (%) for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
RHMEAN_WS (EHIGH, WMTNS)	Mean monthly relative humidity (%) for watershed	Mean of all RHmean values within the watershed upstream of the sampling point ($\sum RHmean_i$ / n, where i = each of n pixels within the watershed).

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
RIVBASIN		Code identifying major drainage basin from which sample was collected: Apalachicola-Chattahoochee Basin (ACB), Arkansas River Basin (ARB), Brazos River Basin (BRB), California Coastal Basins (CCB), Chesapeake Bay Basins (CBB), Colorado River Basin (CORB), Columbia River Basin (CRB), Connecticut River Basin (CNRB), Florida-Alabama Gulf Coast Basins (FGCB), Great Basin (GB), Great Lake Basins (GLB), Great Lakes (GL), Gulf of California Basins (GCAB), Hudson River Basin (HRB), Klamath River Basin (KRB), Lower Mississippi Basin (LMB), Missouri River Basin (MSRB), Mobile-Tombigbee Basin (MTB), Mojave Basin (MVRB), Northeastern Coastal Basins (NECB), Ohio River Basin (ORB), Oregon Closed Basins (OCB), Oregon-Washington Coastal Basins (OWCB), Pee Dee River Basin (PDRB), Red River Basin (RRB), Rio Grande River Basin (RGRB), Sacramento/San Joaquin River Basin (SSRB), Saskatchewan River Basin (SKRB), Savannah River Basin (SVRB), Southeastern Coastal Basins (SECB), St Lawrence Basins (SLB), Susquehanna River Basin (SQRB), Tennessee-Cumberland Basin (TCB), Texas Colorado River Basin (TCRB), Texas Gulf Coast Basins (TGCB), Upper Mississippi River Basin (UMRB).
RUNOFF_WS (EHIGH, PLNLOW, WMTNS)	Mean runoff for watershed (mm)	Watershed average of a raster developed by McCabe and Wolock (2011), calculated at the 8-digit USGS HUC scale. See: McCabe, G. J., and D. M. Wolock. 2011. Century-scale variability in global annual runoff examined using a water balance model. International Journal of Climatology 31:1739-1748.
SDMNTRY (EHIGH)	Percent of bedrock geology in the watershed classified as sedimentary forms	Derived from a simplified version of Reed and Bush (2001) - Generalized Geologic Map of the Conterminous United States.
SHAPE1	Ratio of the watershed area (m^2) to the square of the longest distance (m^2) to the outlet (dimensionless)	Shape factor estimates the elongation of a watershed. Small values indicate round watersheds and large values indicate elongated watersheds.
SHAPE2	Ratio of the watershed area (m^2) to the square of the mean distance (m^2) to the outlet (dimensionless)	Shape factor estimates the elongation of a watershed. Small values indicate round watersheds and large values indicate elongated watersheds.

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
SLOPENHD	Slope of NHD stream segment (m)	Slope (rise/run) of the National Hydrographic Dataset (NHD) line segment derived from digital elevation model estimates of elevation (meters) and NHD estimates of segment length (meters, http://www.horizon-systems.com/nhdplus/).
SQ_KM	Watershed Area (km^2)	Watershed area (km^2) as calculated by Utah State University.
TMAX_PT	Mean annual maximum air temperature at sampling point (°C)	Tmax at the sampling point, where Tmax=GIS raster calculated as $\sum(MAX[x_i]) / 30$, where x_i = the modeled monthly average maximum air temperature (°C) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell. Note that these values are modified from the PRISM annual maximum air temperature grid available at: http://www.prism.oregonstate.edu, that are calculated as $\sum(\sum[x_i] / 12) / 30$, where x_i = the modeled monthly average maximum air temperature (°C) for month i (1-12).
TMAX_WS (EHIGH, PLNLOW, WMTNS)	Mean maximum air temperature for watershed (°C)	Mean of all Tmax values within the watershed upstream of the sampling point ($\sum(Tmax_i) / n$, where i = each of n pixels within the watershed), where Tmax=GIS raster calculated as $\sum(MAX[x_i]) / 30$, where x_i = the modeled monthly average maximum air temperature (°C) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell. Note that these values are modified from the PRISM annual maximum air temperature grid available at: http://www.prism.oregonstate.edu, that are calculated as $\sum(\sum x_i / 12) / 30$, where x_i = the modeled monthly average maximum air temperature (°C) for month i (1-12).
TMAXSD_WS (PLNLOW, WMTNS)	Standard deviation of mean annual maximum air temperature values for the watershed (°C)	Calculated as $(\sqrt{(\sum(\mu - Tmax_i)^2)/(n-1)})$, where i = each of n pixels within the watershed and μ = Tmax_WS).
TMEAN_PT (WMTNS)	Mean monthly air temperature at the sampling point (°C)	GIS raster calculated as $(\sum x_i / 12) / 30$, where x_i = the modeled mean air temperature (°C) for month i (1-12). The modeled monthly mean air temperature (x_i) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell (http://www.prism.oregonstate.edu).

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
TMEAN_WS (EHIGH, PLNLOW)	Mean monthly air temperature for the watershed (°C)	Calculated as ($\sum$(Tmean$_i$) / n, where i = each of n pixels within the watershed), where Tmean=GIS raster calculated as $\sum(\sum[x_i]$ / 12) / 30, where x_i = the modeled mean air temperature (°C) for month i (1-12). The modeled monthly mean air temperature (x_i) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell (http://www.prism.oregonstate.edu).
TMEANPW_PT (PLNLOW)	Mean monthly air temperature for previous winter at sampling point (°C)	This value estimates the average temperature of the winter just prior to when field sampling was done. GIS raster calculated as $\sum(x_i)$ / 3, where x_i = the modeled mean air temperature (°C) for month i (12) of the previous year and months i (1-2) of the sample year. The modeled monthly mean air temperature (xi) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
TMEANPW_WS (EHIGH)	Mean monthly air temperature for previous winter for watershed (°C)	Mean of all TmeanPW values within the watershed upstream of the sampling point ($\sum$ TmeanPW$_i$ / n, where i = each of n pixels within the watershed).
TMEANSD_WS (EHIGH)	Standard deviation of mean monthly air temperature (°C)	Calculated as ($\sqrt{(\sum(\mu - \text{Tmean}_i)^2)/(n-1)}$), where i = each of n pixels within the watershed and μ = Tmean_WS).
TMEANSS_PT (WMTNS)	Mean monthly summer air temperature at sampling point (°C)	This value estimates the average temperature of the specific summer that field sampling was done. GIS raster calculated as $\sum(x_i)$ / 4, where x_i = the modeled mean air temperature (°C) for month i (=6-9) of the sample year. The modeled monthly mean air temperature (x_i) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
TMEANSS_WS (EHIGH, PLNLOW, WMTNS)	Mean monthly summer air temperature for watershed (°C)	This value estimates the average temperature of the specific summer that field sampling was done. Calculated as the mean of all TmeanSS values within the watershed upstream of the sampling point ($\sum$(TmeanSS$_i$) / n, where i = each of n pixels within the watershed), where TmeanSS=GIS raster calculated as $\sum(x_i)$ / 4, where x_i = the modeled mean air temperature (°C) for month I (6-9) of the sample year. The modeled monthly mean air temperature (x_i) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
TMEANSY_PT (WMTNS)	Mean monthly air temperature for sampling year at sampling point (°C)	This value estimates the average temperature of the specific year that field sampling was done. GIS raster calculated as $\sum(x_i)$ / 12, where x_i = the modeled mean air temperature (°C) for month I (1-12) of the sample year. The modeled monthly mean air temperature (x_i) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
TMEANSY_WS (EHIGH, PLNLOW)	Mean monthly air temperature for sampling year for watershed (°C)	This value estimates the average temperature of the specific year that field sampling was done. Calculated as the mean of all TmeanSY values within the watershed upstream of the sampling point ($\sum(TmeanSY_i)$ / n, where i = each of n pixels within the watershed), where TmeanSY=GIS raster calculated as $\sum(x_i)$ / 12, where x_i = the modeled mean air temperature (°C) for month i (1-12) of the sample year. The modeled monthly mean air temperature (x_i) is the average of the minimum and maximum monthly air temperatures (http://www.prism.oregonstate.edu/faq.phtml). Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
TMIN_PT	Mean annual minimum air temperature at sampling point (°C)	GIS raster calculated as $\sum MIN[x_i]$ / 30, where x_i = the modeled monthly average minimum air temperature (°C) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell. Note that these values are modified from the PRISM annual maximum air temperature grid available at: http://www.prism.oregonstate.edu, that are calculated as $\sum(\sum x_i / 12) / 30$, where x_i = the modeled monthly average minimum air temperature (°C) for month i (1-12).
TMIN_WS (EHIGH, PLNLOW, WMTNS)	Mean annual minimum air temperature for watershed (°C)	Mean of all Tmin values within the watershed upstream of the sampling point ($\sum(Tmin_i)$ / n, where i = each of n pixels within the watershed), where Tmin=GIS raster calculated as $\sum(MIN[x_i])$ / 30, where x_i = the modeled monthly average minimum air temperature (°C) for month i (1-12). Values based on 30 years (1971-2000) of PRISM climate estimates. Each value represents a 900 x 900 meter cell. Note that these values are modified from the PRISM annual maximum air temperature grid available at: http://www.prism.oregonstate.edu, that are calculated as $\sum(\sum x_i / 12) / 30$, where x_i = the modeled monthly average minimum air temperature (°C) for month i (1-12).

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
TMINSD_WS	Standard deviation of mean annual minimum air temperature values for the watershed (°C)	Calculated as $(\sqrt{(\sum(\mu - Tmin_i)^2)/(n-1)})$, where i = each of n pixels within the watershed and μ = Tmin_WS).
VOLCANIC	Percent of bedrock geology in the watershed classified as volcanic forms	Derived from a simplified version of Reed and Bush (2001) - Generalized Geologic Map of the Conterminous United States.
WDMAX_PT	Mean maximum number of days with measurable precipitation (wet days) at sampling point (days)	GIS raster calculated as $\sum(\sum x_i / 12) / 30$, where x_i = the modeled maximum number of days with measurable precipitation (i.e., ("wet days") for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
WDMAX_WS (EHIGH, PLNLOW), WMTNS)	Mean maximum number of days with measurable precipitation (wet days) for watershed (days)	Mean of all WDmax values within the watershed upstream of the sampling point ($\sum(WDmax_i) / n$, where i = each of n pixels within the watershed), where WDmax=GIS raster calculated as $\sum(\sum x_i / 12) / 30$, where x_i = the modeled maximum number of days with measurable precipitation (i.e., "wet days") for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
WDMIN_PT	Mean minimum number of days with measurable precipitation (wet days) at sampling point (days)	GIS raster calculated as $(\sum x_i / 12) / 30$, where x_i = the modeled minimum number of days with measurable precipitation (i.e., "wet days") for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
WDMIN_WS (EHIGH, WMTNS)	Mean minimum number of days with measurable precipitation (wet days) for watershed (days)	Mean of all WDmin values within the watershed upstream of the sampling point ($\sum(WDmin_i) / n$, where i = each of n pixels within the watershed), where WDmin=GIS raster calculated as $\sum(\sum x_i / 12) / 30$, where x_i = the modeled minimum number of days with measurable precipitation (i.e., "wet days") for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
WDSUM_PT	Mean number of days per year with measurable precipitation (wet days) at sampling point (days)	GIS raster calculated as $\sum(\sum x_i)/ 30$, where x_i = the modeled mean number of days with measurable precipitation (i.e., "wet days") for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).

Variable (Regional metric response model) *Variables that were included in at least one model are shown in boldface type.*	**Variable Name**	**Description**
WDSUM_WS (EHIGH, PLNLOW)	Mean number of days per year with measurable precipitation (wet days) for watershed (days)	Sum of all WDmean values within the watershed upstream of the sampling point ($\sum(WDsum_i)$ / n, where i = each of n pixels within the watershed), where WDsum=GIS raster calculated as $\sum(\sum x_i)/ 30$, where x_i = the modeled mean number of days with measurable precipitation (i.e., wet days) for month i (1-12). Values based on 30 years (1961-1990) of PRISM climate estimates. Each value represents a 2 x 2 km cell (http://www.prism.oregonstate.edu).
WSAREA_NARS	Watershed Area upstream of sampling point (km^2)	Watershed area calculated from sampling point by GIS staff at EPA-Corvallis in 2015. Note values may be different from those calculated by EPA-Duluth, which included the NHD segment on which a sampling point was located.
WTDH_PT (EHIGH, PLNLOW)	Mean water table depth at sampling point (ft)	Mean high values of seasonally high water table (in feet) at the sampling point. Derived from the State Soil Geographic (STATSGO) Database.
WTDH_WS (EHIGH, PLNLOW)	Mean water table depth for watershed (ft)	Watershed mean of the high values of seasonal water table depth (feet) of soils from the State Soil Geographic (STATSGO) Database.
XELEV **(PLNLOW, WMTNS)**	Elevation (m)	Elevation at the sampling point

11 PHYSICAL HABITAT ASSESSMENT

An assessment of river and stream (fluvial) physical habitat condition is a major component of the National Rivers and Streams Assessment (NRSA). Of many possible general and specific fluvial habitat indicators measured in the NRSA surveys in 2008-2009, the assessment team chose streambed stability & excess fine sediments, instream habitat cover complexity, riparian vegetation, and riparian human disturbances for its assessment. These four indicators are generally important throughout the U.S. Furthermore, the project team had reasonable confidence in factoring out natural variability to determine expected values and the degree of anthropogenic alteration of the habitat attributes represented by these indicators.

In the broadest sense, fluvial habitat includes all physical, chemical, and biological attributes that influence or sustain organisms within streams or rivers. We use the term *physical habitat* to refer to the structural attributes of habitat. NRSA made field measurements aimed at quantifying eight general attributes of physical habitat condition, including direct measures of human disturbance.

- Habitat Volume/Stream Size
- Habitat Complexity and Cover for Aquatic Biota
- Streambed Particle Size
- Bed Stability and Hydraulic Conditions
- Channel-Riparian and Floodplain Interaction
- Hydrologic Regime
- Riparian Vegetation Cover and Structure
- Riparian Disturbance

These attributes were previously identified during EPA's 1992 national stream monitoring workshop (Kaufmann 1993) as those essential for evaluating physical habitat in regional monitoring and assessments. They are typically incorporated in some fashion in regional habitat survey protocols (Platts *et al.* 1983, Fitzpatrick *et al.* 1998, Lazorchak *et al.* 1998, Peck *et al.* 2006, Peck *et al.*, in press, USEPA 2004) and were applied in the previous National Wadeable Streams Assessment (WSA) and the Western Rivers and Streams Pilot (EMAP-W) surveys conducted between 2000 and 2005 (USEPA 2006, Stoddard *et al.* 2005a,b). The major habitat metrics used in those past assessments and considered in NRSA are listed and defined in Table 11.1. Some measures of these attributes are useful measures of habitat condition in their own right (*e.g.*, channel incision as a measure of channel-riparian interaction); others are important controls on ecological processes and biota (*e.g.*, bed substrate size), still others are important in the computation of more complex habitat condition metrics (*e.g.*, bankfull depth is used to calculate Relative Bed Stability [*RBS*]). Like biological characteristics, most habitat attributes vary according to their geomorphic and ecological setting. Even direct measures of riparian human activities and disturbances are strongly influenced by their geomorphic setting. And even within a region, differences in precipitation, stream drainage area channel gradient (slope) lead to variation in many aspects of stream habitat, because those factors influence discharge, flood stage, stream power (the product of discharge times gradient), bed shear stress (proportional to the product of depth and slope), and riparian vegetation. However, all eight of the major habitat attributes can be directly or indirectly altered by anthropogenic activities.

NRSA follows the precedent of EMAP-W and WSA in reporting the condition of fluvial physical habitat condition on the basis of four habitat indicators that are important nationwide, can be reliably and economically measured, and their reference condition under minimal anthropogenic disturbance can be interpreted with reasonable confidence. These are: relative bed stability (RBS) as an indicator of bed sedimentation or hydrologic alteration, the areal cover and variety of fish concealment features as a measure of in-stream habitat complexity, riparian vegetation cover and structure as an indicator of riparian vegetation condition, and a proximity-weighted tally of streamside human activities as an indicator of riparian human disturbances (Paulsen *et al*., 2008).

In this document, we describe the approach taken by NRSA for assessing physical habitat condition in rivers and streams based on the four above-mentioned indicators. We also examine the rationale, importance, and measurement precision of each of these indicators, including the analytical approach for estimating reference conditions for each. Reference conditions for each indicator were interpreted as their expected value in sites having the least amount of anthropogenic disturbance within appropriately stratified regions. In most cases, we also refine the expected values as a function of geoclimatic controlling factors within regions. Finally, we examine patterns of association between physical habitat indicators and anthropogenic disturbance by contrasting habitat indicator values in least- moderate- and most-disturbed sites nationally and within regions.

11.1 METHODS

11.1.1 Physical Habitat Sampling and Data Processing

In the wadeable streams sampled in NRSA, field crews took measurements while wading the length of each sample reach (Peck *et al*. 2006); in non-wadeable rivers, these measurements were made from boats (Peck *et al*. in press, Hughes and Peck 2008). Physical habitat data were collected from longitudinal profiles and from 11 cross-sectional transects and streamside riparian plots evenly spaced along each sampled stream reach (U.S. EPA 2007). The length of each sampling reach was defined proportional to the wetted channel width and measurements were placed systematically along that length to represent the entire reach. Sample reach lengths were 40 times the wetted channel-width (ChW) long in wadeable streams, with a minimum reach length of 150 m for channels less than 3.5 m wide. In non-wadeable rivers, reach lengths were also set to 40 ChW with a maximum length of 2,000 m. Thalweg depth measurements (in the deepest part of channel), habitat classification, and mid-channel substrate observations were made at tightly spaced intervals; whereas channel cross-sections and shoreline-riparian stations for measuring or observing substrate, fish cover (concealment features), large woody debris, bank characteristics and riparian vegetation structure were spaced further apart. Thalweg (maximum) depth was measured at points evenly spaced every 0.4 ChW along these reaches to give profiles consisting of 100 measurements (150 in streams <2.5m wide). The tightly spaced depth measures allow calculation of indices of channel structural complexity, objective classification of channel units such as pools, and quantification of residual pool depth, pool volume, and total stream volume. Channel slope and sinuosity on non-wadeable rivers were estimated from 1:24,000-scale digital topographic maps.

In wadeable streams, wetted width was measured and substrate size and embeddedness were evaluated using a modified Wolman pebble count of 105 particles spaced systematically along 21 equally spaced cross-sections, in which individual particles were classified visually into seven size-classes plus bedrock, hardpan and other (*e.g.*, organic material). The numbers of pieces of large woody debris in the bankfull channel were tallied in 12 size classes (3 length by 4 width classes) along the entire length of sample reaches. Channel incision and the dimensions of the wetted and bankfull stream channel were measured at 11 equally-spaced transects. Bank characteristics and areal cover of fish concealment features were visually assessed in 10 m long instream plots centered on transects, while riparian vegetation structure, presence of large (legacy) riparian trees, non-native (alien) riparian plants, and evidence of human disturbances (presence/absence and proximity) in 11 categories were visually assessed on adjacent 10 × 10 m riparian plots on both banks. In addition, channel gradient (slope) in wadeable streams was measured to provide information necessary for calculating residual pool depth and relative bed stability. In wadeable streams, crews used laser or hydrostatic levels for slopes <2.5%, and optionally were allowed to use hand-held clinometers in channels with slopes >2.5%. Compass bearing between stations were obtained for calculating channel sinuosity. Channel constraint and evidence of debris torrents and major floods were assessed over the whole reach after the other components were completed. Discharge was measured by the velocity-area method at the time of sampling, or by other approximations if that method was not practicable (Peck *et al*. 2006; USEPA 2007). Two-person crews typically completed NRSA habitat measurements in 1.5 to 4 hours of field time, though large, deep streams that were only marginally wadeable took up to several hours longer.

In non-wadeable rivers, NRSA field crews floating downstream in inflatable rafts, or in slower rivers small power boats, measured the longitudinal thalweg depth profile (approximated at mid-channel) using 7.5m telescoping survey rods or SONAR, at the same time tallying snags and off-channel habitats, classifying main channel habitat types, and characterizing mid-channel substrate by probing the bottom. At 11 littoral/riparian plots (each 10m wide x 20m long) spaced systematically and alternating sides along the river sample reach, field crews measured channel wetted width, bankfull channel dimensions, incision, channel constraint. They assessed near-shore, shoreline, and riparian physical habitat characteristics by measuring or observing littoral depths, riparian canopy cover, substrate, large woody debris, fish cover, bank characteristics, riparian vegetation structure, presence of large ("legacy") riparian trees, non-native riparian plants, and evidence of human activities. After all the thalweg and littoral/riparian measurements and observations were completed, the crews estimated the extent and type of channel constraint (see Peck *et al*. in press; USEPA 2007). Channel slope and sinuosity on non-wadeable rivers were estimated from 1:24,000-scale digital topographic maps.

See Kaufmann *et al*. (1999) for calculations of reach-scale summary metrics from field data, including mean channel dimensions, residual pool depth, bed particle size distribution, wood volume, riparian vegetation cover and complexity, and proximity-weighted indices of riparian human disturbances. See Faustini and Kaufmann (2007) for details on the calculation of geometric mean streambed particle diameter, Kaufmann *et al*. (2008, 2009) for calculation of bed shear stress and relative bed stability (modified since published by Kaufmann *et al*. 1999), and Kaufmann and Faustini (2012) for demonstrating the utility of EMAP and NRSA channel morphologic data to estimate transient storage and hydraulic retention in wadeable streams.

11.1.2 Quantifying the Precision of Physical Habitat Indicators

The absolute and relative precision of the physical habitat condition metrics used in NRSA are shown in Table 11.2, based on data from 2113 unique sites and repeat visits to a random subset of 197 of those sites. The RMS_{rep} expresses the precision or replicability of field measurements, quantifying the average variation in a measured value between same-season site revisits, pooled across all sites where measurements were repeated. We calculated RMS_{rep} as the root-mean-square error of repeat visits during the same year, equivalent to the root mean-square error (RMSE) relative to the site means, as discussed Kaufmann *et al*., 1999 and Stoddard *et al*. (2005a). S/N is the ratio of variance among streams ("signal") to that for repeat visits to the same stream("noise") as described by Kaufmann *et al*. (1999).

The ability of a monitoring program to detect trends is sensitive to the spatial and temporal variation in the target indicators as well as the design choices for the network of sites and the timing and frequency of sampling. Sufficient temporal sampling of sites was not available to estimate all relevant components of variance for the entire U.S. However, Larsen *et al* (2004) examined the survey sampling variance components for a number of the EMAP-NARS physical habitat variables, including some of interest in this paper (residual depth, canopy cover, fine sediment, and large wood). Their analysis was based on evaluation on six Pacific Northwest surveys that included 392 stream reaches and 200 repeat visits. These surveys were conducted in Oregon and Washington from 1993 to 1999. Most were from one to three years in duration, but one survey lasted six years. They modeled the likelihood of detecting a 1–2% per year trend in the selected physical habitat characteristics, if such a trend occurs, as a function of the duration of a survey. To calculate the number of years required to detect the defined trends in a monitoring network with a set number of sites, they set the detection probability at >80% with <5% probability of incorrectly asserting a trend if one is not present. We used the same survey data sets to duplicate their analysis for several variables not included in the Larsen *et al*. (2004) publication, including log transformed relative bed stability (*LRBS_BW5)* and riparian vegetation cover complexity (*XCMGW*, the combined cover of three layers of riparian woody vegetation); the results of that trend detection potential is summarized in Table 11.3.

11.2 PHYSICAL HABITAT CONDITION INDICATORS

11.2.1 Relative Bed Stability and Excess Fines

Streambed characteristics (*e.g*., bedrock, cobbles, silt) are often cited as major controls on the species composition of macroinvertebrate, periphyton, and fish assemblages in streams (*e.g*., Hynes 1970, Cummins 1974, Platts *et al*. 1983, Barbour *et al*. 1999, Bryce *et al*., 2008, 2010). Along with bedform (*e.g*., riffles and pools), streambed particle size influences the hydraulic roughness and consequently the range of water velocities in a stream channel. It also influences the size range of interstices that provide living space and cover for macroinvertebrates and smaller vertebrates. Accumulations of fine substrate particles (excess fine sediments) fill the interstices of coarser bed materials, reducing habitat space and its availability for benthic fish and macroinvertebrates (Hawkins *et al*. 1983, Platts *et al*. 1983, Rinne 1988). In addition, these fine particles impede circulation of oxygenated water into hyporheic habitats reducing egg-to-emergence survival and growth of juvenile salmonids (Suttle *et al*. 2004). Streambed characteristics are often sensitive indicators of the effects of human activities on streams (MacDonald *et al*. 1991, Barbour *et al*. 1999, Kaufmann *et al*. 2009). Decreases in the mean particle size and increases in streambed fine sediments can destabilize stream channels (Wilcock 1997, 1998) and may indicate increases in the rates of upland erosion and sediment supply (Lisle 1982, Dietrich *et al*. 1989).

"Unscaled" measures of surficial streambed particle size, such as percent fines or D_{50}, can be useful descriptors of stream bed conditions. In a given stream, increases in percent fines or decreases in

D_{50} may result from anthropogenic increases in bank and hillslope erosion. However, a great deal of the variation in bed particle size among streams is natural: the result of differences in stream or river size, slope, and basin lithology. The power of streams to transport progressively larger sediment particles increases in direct proportion to the product of flow depth and slope. All else being equal, steep streams tend to have coarser beds than similar size streams on gentle slopes. Similarly, the larger of two streams flowing at the same slope will tend to have coarser bed material, because its deeper flow has more power to scour and transport fine particles downstream (Leopold *et al.* 1964, Morisawa 1968). For these reasons, we "scale" bed particle size metrics, expressing bed particle size in each stream as a deviation from that expected as a result of its size, power, and landscape setting (Kaufmann *et al.*, 1999, 2008, 2009).

The scaled median streambed particle size is expressed as Relative Bed Stability (*RBS*), calculated as the ratio of the geometric mean diameter, D_g, divided by D_{cbf}, the critical diameter (maximum mobile diameter) at bankfull flow (Gordon *et al.*, 1992), where D_g is based on systematic streambed particle sampling ("pebble counts") and D_{cbf} is based on the estimated streambed shear stress calculated from slope, channel dimensions, and hydraulic roughness during bankfull flow conditions.

RBS is a measure of habitat stability for aquatic organisms as well as an indication of the potential for economic risk to streamside property and structures from stream channel movement. In many regions of the U.S.A, we may also be able to use *RBS* to infer whether sediment supply is augmented by upslope or bank erosion from anthropogenic or other disturbances, because it can indicate the degree of departure from a balance between sediment supply and transport. In interpreting *RBS* on a regional scale, Kaufmann *et al.* (1999, 2009) argued that, over time, streams and rivers adjust sediment transport to match supply from natural weathering and delivery mechanisms driven by the natural disturbance regime, so that *RBS* in appropriately stratified regional reference sites should tend towards a range characteristic of the climate, lithology, and natural disturbance regime. Values of the *RBS* index either substantially lower (finer, more unstable streambeds) or higher (coarser, more stable streambeds) than those expected based on the range found in least-disturbed reference sites within an ecoregion are considered to be indicators of ecological stress.

Excess fine sediments can destabilize streambeds when the supply of sediments from the landscape exceeds the ability of the stream to move them downstream. This imbalance results from numerous human uses of the landscape, including agriculture, road building, construction, and grazing. Lower-than-expected streambed stability may result either from high inputs of fine sediments (from erosion) or increases in flood magnitude or frequency (hydrologic alteration). When low *RBS* results from fine sediment inputs, stressful ecological conditions result from fine sediments filling in the habitat spaces between stream cobbles and boulders (Bryce *et al.* 2008, 2010). Instability (low *RBS*) resulting from hydrologic alteration can be a precursor to channel incision and arroyo formation (Kaufmann *et al.* 2009). Perhaps less well recognized, streams that have higher than expected streambed stability can also be considered stressed—very high bed stability is typified by hard, armored streambeds, such as those often found below dams where fine sediment flows are interrupted, or within channels where banks are highly altered. Values of *RBS* higher than reference expectations can indicate anthropogenic coarsening or armoring of streambeds, but streams containing substantial amounts of bedrock may also have very high *RBS*, and at this time it is difficult to determine the role of human alteration in stream coarsening on a national scale. For this reason, NRSA reported only on the "low end" of *RBS* relative to reference conditions, generally indicating stream bed excess fine sediments or augmented stormflows associated with human disturbance of stream drainages and riparian zones.

11.2.1.1 *Precision of Sediment and Bed Stability Measurements*

The geometric mean bed particle diameter (D_{gm}) and *RBS* varied over 8 orders of magnitude in the NRSA surveys. Because of this wide variation and the fact that both exhibit repeat-visit variation that is proportional to their magnitude at individual streams, it is useful and necessary to log transform these variables (*LSUB_DMM* and *LRBS_g08*). The RMS_{rep} of *LSUB_DMM* in wadeable streams of the EMAP-W survey was 0.246, similar to that reported by Faustini and Kaufmann (2007) for EMAP-W (0.21). For a D_{gm} = *"y"* mm, the log-based RMS_{rep} of 0.246 translates to an asymmetrical 1SD error bound of 0.57*y* to 1.76*y* mm. The RMS_{rep} of *LRBS_g08* in NRSA wadeable streams was 0.48, approximately 6% of its observed range, but less precise (surprisingly) than that for EMAP-W (RMS_{rep} = 0.365). The log-based RMS_{rep} of 0.48 for NRSA LRBS_g08 translates to an asymmetrical error bound of 0.33*y* to 3.0*y* around an untransformed *RBS* value of *"y"* (Table 2). Compared with the high S/N ratio for *LSUB_DMM* in NRSA (12.4 for wadeable+boatable waters), relative precision for *LRBS_g08* was lower (S/N=5.0), reflecting the reduction in total variance when a large component of natural variability is "modeled out" by scaling for channel gradient, water depth, and channel roughness. Nevertheless, the relative precision of LRBS_g08 is moderately high and easily adequate to make it a useful variable in regional and national assessments (Kaufmann *et al.* 1999, 2008, Faustini and Kaufmann 2007). The transformation of the unscaled geometric mean bed particle diameter D_{gm} to the ratio *RBS* by dividing by the critical diameter reduced the within-region variation by accounting for some natural controlling factors. As a result, we feel that the scaled variable helps to reveal alteration of bed particle size and mobility from anthropogenic erosion and sedimentation (Kaufmann *et al.* 2008, 2009).

We have examined the components of variability of *LRBS* based on earlier surveys and modeled its potential utility in trend detection in the Pacific Northwest region of the U.S. with the same data and procedures as used by Larsen *et al.* (2004), in which all methods were the same as used in EMAP-W and WSA except that bed substrate mean diameter data used by Larsen *et al.* was determined based on 55, rather than 105 particles. (NRSA data differed from data used in that analysis by using laser levels rather than hand-held clinometers to measure wadeable stream slopes <2/5%) That analysis showed that a 50-site monitoring program could detect a subtle trend in *LRBS_BW5* of 2% per year within 8 years, if sites were visited every year (Table 11.3).

11.2.2 Instream Habitat Cover Complexity

Although the precise mechanisms are not completely understood, the most diverse fish and macroinvertebrate assemblages are usually found in streams that have complex mixtures of habitat features: large wood, boulders, undercut banks, tree roots, etc. (see Kovalenko *et al.* 2011). When other needs are met, complex habitat with abundant cover should generally support greater biodiversity than simple habitats that lack cover (Gorman and Karr 1978, Benson and Magnuson 1992). Human use of streams and riparian areas often results in the simplification of this habitat, with potential effects on biotic integrity (Kovalenko *et al.*, 2011). For this assessment, we use a measure (*XFC_NAT* in Kaufmann *et al.*, 1999) that sums the amount of instream habitat consisting of undercut banks, boulders, large pieces of wood, brush, and cover from overhanging vegetation within a meter of the water surface, all of which were estimated visually by NRSA field crews.

11.2.2.1 Quantifying Instream Habitat Complexity

Habitat complexity is difficult to quantify, and could be quantified or approximated by a wide variety of measures. The NRSA Physical Habitat protocols provide estimates for nearly all of the following components of complexity identified during EPA's 1992 stream monitoring workshop (Kaufmann 1993):

- Habitat type and distribution (*e.g.*, Bisson *et al.* 1982, O'Neill and Abrahams 1984, Frissell *et al.* 1986, Hankin and Reeves 1988, Hawkins *et al.* 1993, Montgomery and Buffington 1993, 1997, 1998).
- Large wood count and size (*e.g.*, Harmon *et al.* 1986, Robison and Beschta 1989, Peck *et al.* 2006).
- In-channel cover: Percentage areal cover of fish concealment features, including undercut banks, overhanging vegetation, large wood, boulders (Hankin and Reeves 1988, Kaufmann and Whittier 1997, Peck *et al.* 2006)
- Residual pools, channel complexity, hydraulic roughness(*e.g.*, Kaufmann 1987a, b, Lisle 1987, Stack and Beschta, 1989; Lisle and Hilton 1992, Robison and Kaufmann 1994, Kaufmann *et al.* 1999, Kaufmann *et al.* 2008, Kiem *et al.*, 2002; Kaufmann *et al.* 2011)
- Width and depth variance, bank sinuosity (Kaufmann 1987a, Moore and Gregory 1988, Kaufmann *et al.* 1999, Madej 1999, 2001, Kaufmann *et al.* 2008, Mossop and Bradford 2006; Pearsons and Temple, 2007, 2010; Kaufmann and Faustini, 2012).

Residual depth is a measure of habitat volume, but also serves as one of the indicators of channel habitat complexity, particularly when expressed as a deviation from reference expectations, including the influences of basin size. A stream with more complex bottom profile will have greater residual depth than one of similar drainage area, discharge and slope, but lacks that complexity (Kaufmann 1987a). Conversely, between two streams of equal discharge and slope, the one with greater residual depth (*i.e.*, larger, more abundant residual pools) will have greater variation in cross-sectional area, slope, and substrate size. A related measure of the complexity of channel morphology is the coefficient of variation in thalweg depth, calculated entirely from the thalweg depth profile (*SDDEPTH / XDEPTH*). The thalweg profile is a systematic survey of depth in the stream channel along the path of maximum depth ("thalweg"). In addition to measures of channel morphometric complexity, NRSA physical habitat protocols measure in-channel large wood (sometimes called "large woody debris" or simply "LWD"), and several estimates of the areal cover of various types of fish and macroinvertebrate "cover" or concealment features. The large wood metrics include counts of wood pieces per 100 m of bankfull channel and estimates of large wood volume in the sample reach expressed in cubic meters of wood per square meter of bankfull channel. The "fish cover" variables are visual estimates of the areal cover of single or combined types of habitat features.

NRSA required a general summary metric as a holistic indicator of many aspects of habitat complexity, so used the metric *XFC_NAT*, summing the areal cover from large wood, brush, overhanging vegetation, live trees and roots, boulders, rock ledges, and undercut banks in the wetted stream channel. Habitat complexity and the abundance of particular types of habitat features differ naturally with stream size, slope, lithology, flow regime, and potential natural vegetation. For example, boulder cover will not occur naturally in streams draining deep deposits of loess or alluvium that do not contain large rocks. Similarly, large wood will not be found naturally in streams located in regions where riparian or upland trees do not grow naturally. Though the index

XFC_NAT partially overcomes these differences by summing divergent types of cover, we set stream-specific expectations for habitat complexity metrics in NRSA based on region-specific reference sites and further refined them as a function of geoclimatic controls.

11.2.2.2 Precision of habitat complexity measures

The instream habitat complexity index *XFC_NAT* ranged from 0 to 2.3, or 0% to 230% in NRSA, expressing the combined areal cover of the five cover elements contributing to its sum. The RMS_{rep} of Log(0.01+*XFC_NAT)* was 0.24, meaning that an *XFC_NAT* value of 10% cover at a single stream site has a $\pm$1.0 RMS_{rep} error bound of 6% to 17% (Table 11.2). S/N was relatively low for this indicator (1.87), though higher in wadeable streams (2.29) than in boatable rivers (1.22). Despite its relatively low S/N, the RMS_{rep} for *LXFC_LWD* was 10% of the observed range of *XFC_NAT*. It was retained as a habitat complexity indicator because it contains biologically relevant information not available in other metrics, showed moderate responsiveness to human disturbances, and has precision adequate to discern relatively large differences in habitat complexity.

11.2.3 Riparian Vegetation

11.2.3.1 Quantifying Riparian Vegetation Cover Complexity

The importance of riparian vegetation to channel structure, cover, shading, inputs of nutrients and large wood, and as a wildlife corridor and buffer against anthropogenic disturbance is well recognized (Naiman *et al.* 1988, Gregory *et al.* 1991). Riparian vegetation not only moderates stream temperatures through shading, but also increases bank stability and the potential for inputs of coarse and fine particulate organic material. Organic inputs from riparian vegetation become food for stream organisms and provide structure that creates and maintains complex channel habitat.

The presence of a complex, multi-layered vegetation corridor along streams and rivers is an indicator of how well the stream network is buffered against sources of stress in the watershed. Intact riparian areas can help reduce nutrient and sediment runoff from the surrounding landscape, prevent streambank erosion, provide shade to reduce water temperature, and provide leaf litter and large wood that serve as food and habitat for stream organisms (Gregory *et al.*, 1991). The presence of large, mature canopy trees in the riparian corridor reflects its longevity, whereas the presence of smaller woody vegetation typically indicates that riparian vegetation is reproducing, and suggests the potential for future sustainability of the riparian corridor (Kaufmann and Hughes 2006).

NRSA evaluated the cover and complexity of riparian vegetation based on the metric *XCMGW*, which is calculated from visual estimates made by field crews of the areal cover and type of vegetation in three layers: the ground layer (<0.5m), mid-layer (0.5-5.0 m) and upper layer (>5.0 m). The separate measures of large and small diameter trees, woody and non-woody mid-layer vegetation, and woody and non-woody ground cover are all visual estimates of areal cover. *XCMGW* sums the cover of *woody* vegetation over these three vegetation layers, expressing both the abundance of vegetation cover and its structural complexity. Its theoretical maximum is 3.0 if there is 100% cover in each of the three vegetation layers. *XCMGW* gives an indication of the longevity and sustainability of perennial vegetation in the riparian corridor (Kaufmann *et al.* 1999, Kaufmann and Hughes 2006).

11.2.3.2 Precision of Riparian Vegetation Index

XCMGW ranged from 0 to 2.8 (280% cover), with RMS_{rep} of Log(0.01+*XCMGW*) = 0.146 (Table 11.2), meaning that an *XCMGW* value of 10% at a single stream site has a $\pm$1.0 RMS_{rep} error bound of 7% to 14%. Its S/N ratio was 9.38, indicating very good potential for discerning differences among sites. We examined the components of variability of *XCMGW* and modeled its potential utility in trend detection in the Pacific Northwest region of the U.S. with the same data and procedures as used by Larsen *et al.* (2004). Based on that analysis, a 50-site monitoring program could detect a subtle trend in *XCMGW* of 2% per year within 8 years, if sites were visited every year (Table 11.3).

11.2.4 Riparian Human Disturbances

Agriculture, roads, buildings, and other evidence of human activities in or near stream and river channels may exert stress on aquatic ecosystems and may also serve as indicators of overall anthropogenic stress. EPA's 1992 stream monitoring workshop recommended field assessment of the frequency and extent of both in-channel and near-channel human activities and disturbances (Kaufmann 1993). The vulnerability of the stream network to potentially detrimental human activities increases with the proximity of those activities to the streams themselves. NRSA follows Stoddard *et al.* (2005b) and U.S. EPA (2006) in using a direct measure of riparian human disturbance that tallies 11 specific forms of human activities and disturbances (walls, dikes, revetments or dams; buildings; pavement or cleared lots; roads or railroads; influent or effluent pipes; landfills or trash; parks or lawns; row crop agriculture; pasture or rangeland; logging; and mining) at 22 separate locations along the stream reach, and weights them according to how close to the channel they are observed (*W1_HALL* in Kaufmann *et al.* 1999). Observations within the stream or on its banks are weighted by 1.5, those within the 10 × 10 meter plots are weighted by 1.0, and those visible beyond the plots are weighted by 0.5. The index *W1_HALL* ranged from 0 (no observed disturbance) to ~7 (*e.g.*, equivalent to four or 5 types of disturbance observed in the stream, throughout the reach; or seven types observed within all 22 riparian plots bounding the stream reach). Although direct human activities certainly affect riparian vegetation complexity and layering measured by the Riparian Vegetation Index (previous paragraph), the Riparian Disturbance Index is more encompassing, and differs by being a *direct* measure of observable human activities that are presently or potentially detrimental to streams.

11.2.4.1 Precision of riparian disturbance indicators

The proximity-weighted human disturbance indicator *W1_HALL* ranged from 0 to 7.3 in NARS, and its precision was proportional to the level of disturbance. The RMS_{rep} of *log*(0.1+*W1_HALL)* was 0.186 (Table 2), meaning that a *W1_HALL* value of 1.0 at a single stream site has a $\pm$1.0 RMS_{rep} error bound of 0.65 to 1.53. The relative precision of *Log*(0.1+*W1_HALL*) was moderate (S/N=5.18).

11.3 Estimating Reference Condition for Physical Habitat

11.3.1 Reference Site Screening and Anthropogenic Disturbance Classifications

As part of the routine application of its field and GIS protocols, NRSA obtained various measures of human disturbance associated with each site and its catchment. Site-scale indicators of human disturbance included field observations of various human activities including nearby roads, riprap, agricultural activities, riparian vegetation disturbance, etc., as detailed by Kaufmann *et al.* (1999). These indicators of local scale disturbance were used in combination with water chemistry (Chloride, Total Phosphorus, Total Nitrogen, and Turbidity), as described by Herlihy *et al.* (2008), to screen probability and hand-picked sites and designate them as least- moderately-, and highly-disturbed, relative to other sites within each of the regions of NRSA. In addition, we used basin and sub-basin row crop and urban land use percentages and the density of dams and impoundments as described in the reference technical section to rank sites by disturbance categories, as shown in Table 11.4. To avoid circularity, we did not use any field measures of sediment, in-channel habitat complexity, or riparian vegetation to screen least-disturbed sites used to estimate reference condition for excess streambed fining, instream fish cover, and riparian vegetation. Nor did we use such measures in defining levels of disturbance to use in examining the associations of these habitat metrics with human disturbances. We did, however, use field observations of the level and proximity of streamside human activities in screening reference sites and defining levels of disturbance for evaluating indicator responsiveness. In this article, the designation "**R**" refers to least-disturbed ("reference") sites; "**M**" to moderately-disturbed sites, and "**D**" to the most-disturbed sites within each of the nine aggregate ecoregions discussed herein. We defined these site disturbance categories independent of the habitat indicators we evaluate in this article (other than riparian human disturbances), allowing an assessment of fluvial habitat response to a gradient of human activities and disturbances.

11.3.2 Modeling Expected Reference Values of the Indicators

In the following paragraphs, we describe the conceptual basis for modeling the expected range of values for the each of the physical habitat indicators under least-disturbed (reference) condition. The details of the models are presented in Model 11.1 through Model 11.4 on pages 118 to 124.

For *LRBS*, we modeled expected values based on the distribution of *LRBS* in reference sites within regions or groups of regions. In some regions boatable and wadeable rivers and streams were modeled separately; in others they were combined. Where possible, we used regression models of *LRBS*=f(*W1_Hall*) within reference sites only (*RMD_PHAB*=R), and then set *W1_Hall* (human disturbance) to zero to estimate the central tendency of *LRBS* in the absence of near-stream disturbance regressions (note that zero values of *W1_Hall* were within the regional sets of reference sites). In these cases, the adjusted mean of the reference distribution was defined as the y-intercept of these regressions and the SD about the adjusted reference mean was defined as the RMSE of those regressions. Condition classes were defined based on normal approximation of the 5th and 25th percentiles of the actual or adjusted reference distributions. The definition of "Poor" condition was set as those sites with *LRBS* < the reference mean *LRBS* minus 1.65(SD_{ref}). Sites in "Good" condition with respect to this indicator were those with *LRBS*> the reference mean *LRBS* minus 0.67(SD_{ref}).

For instream fish cover complexity, we estimated expected *XFC_NAT* based on multiple linear regression models predicting Log_{10}(0.01 + *XFC_NAT*) in reference sites from geoclimatic controlling factors within regions or aggregated regions. Because there is a gradient of human disturbance within the set of reference sites in all the regions considered, and it was correlated with

XFC_NAT, we also incorporated field measures of human disturbance into the regressions. Site-specific expected ("E") values of *XFC_NAT* were then calculated by setting the human disturbance metric values to very low values (but never lower than observed among the reference). We then calculated observed/expected (O/E) values of *XFC_NAT* and examined their distribution among reference sites. Because we had modeled-out disturbance to some extent in our calculation of E values, the distribution of O/E in reference sites did not necessarily have a mean of 1/1 (Log=0), although means were very close to 1/1. We set expectations of the O/E values based on the mean and SD of the regional reference distributions, analogous to that described for *LRBS* in the previous paragraph.

For riparian condition (*XCMGW* transformed as *L_xcmgw*= Log_{10}(0.01+*XCMGW*), we estimated expected condition based on simple regional reference site distributions or regression models in which *W1_Hall* was set to zero in regressions prediction *L_xcmgw* as a function of *W1_Hall* within the subset of reference sites (*RMD_PHAB*=R). The adjusted mean *L_xcmgw* for reference sites was defined as the y-intercept and the SD about the reference mean was defined as the RMSE of those regressions.

We did not base thresholds of the riparian human disturbance indicator on the reference distributions, as was done for sediment, habitat complexity and riparian vegetation condition. Rather, the classes for riparian disturbance were set using the same judgement-based criteria for all regions. *W1_HALL*, the database variable name for this indicator, is a direct measure of human disturbance "pressure," unlike the other habitat indicators, which are actually measures of habitat response to human disturbance pressures. It is very difficult to define reference sites without screening sites based on *W1_HALL*. For this reason, we took this different approach for setting riparian disturbance thresholds, defining low disturbance sites as those with *W1_HALL* <0.33 and high riparian disturbance sites as those with *W1_HALL* >1.5; we applied these same thresholds in all ecoregions. A value of 1.5 for a stream means, for example, that at 22 locations along the stream the field crews found an average of one of 11 types of human disturbance within the stream or its immediate banks. A value of 0.33 means that, on average, one type of human disturbance was observed at one-third of the 22 riparian plots along a sample stream or river.

11.4 Response of the Physical Habitat Indicators to Human Disturbance

The Sedimentation and Riparian Vegetation indicators, *LRBS* and *XCMGW* showed modest to strong negative response to human disturbance in most regions and aggregations of regions, as illustrated by t-values (+2.11 to +12.24) comparing differences in means of Reference minus Disturbed sites (Table 11.5). However, the strength of associations with human disturbance tended to be slightly stronger for sediments and much stronger for riparian vegetation in wadeable *versus* boatable sites (Table 11.5, and Figures 11.1 throughFigure 11.14).

Except for the weak contrary response in the Eastern Highlands (t= -1.26), the instream habitat complexity indicator showed moderate response to human disturbance, with t-values ranging from +2.13 to +4.25 (Table 11.5). As for the other habitat indicators, associations were in most cases stronger for wadeable, *versus* boatable sites.

Because the field-obtained measures of riparian disturbance used in the NRSA are themselves direct indicators of human disturbance, and were used to screen reference sites, we did not do t-tests to quantify the strength of relationship between *W1_Hall* and general disturbance class in Figure 11.14. We illustrate the relationship of *W1_HALL* to the human disturbance gradient in these figures to compare the relative magnitudes of *W1_Hall* among disturbed, medium, and relatively undisturbed streams in the various regions of the U.S.

11.5 LITERATURE CITED

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1999. *Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish.* EPA/841-B-99-002, U.S. Environmental Protection Agency, Washington, D.C.

Benson, B. J. and J. J. Magnuson. 1992. Spatial heterogeneity of littoral fish assemblages in lakes: relation to species diversity and habitat structure. *Canadian Journal of Fisheries and Aquatic Sciences* 49:1493-1500.

Bisson, P. A., J. L. Nielsen, R. A. Palmason, and L. E. Grove. 1982. A system of naming habitat types in small streams, with examples of habitat utilization by salmonids during low stream flow. Pages 62-73 In N. B. Armantrout, [editor]. *Acquisition and utilization of aquatic habitat inventory information.* Symposium Proceedings, October 28-30, 1981, Portland, Oregon. The Hague Publishing, Billings, Montana.

Bryce, S.A., Lomnicky, G.A, and P.R. Kaufmann. 2010. Protecting Sediment-Sensitive Aquatic Species in Mountain Streams through the Application of Biologically-Based Criteria Streambed Sediment Criteria. *J. North American Benthological Soc.* 29(2):657-672.

Bryce, S.A., G.A. Lomnicky, P.R. Kaufmann, L.S. McAllister, and T.L. Ernst. 2008. Development of Biologically-Based Sediment Criteria in Mountain Streams of the Western United States. *N. Am. J. Fish. Manage.* 28: 28:1714-1724.

Buffington, J. M. and D. R. Montgomery. 1999a. Effects of hydraulic roughness on surface textures of gravel-bed rivers. *Water Resources Research* 35:3507-3521.

Buffington, J. M. and D. R. Montgomery. 1999b. Effects of sediment supply on surface textures of gravel-bed rivers. *Water Resources Research* 35:3523-3530.

Cummins, K. W. 1974. Structure and function of stream ecosystems. *BioScience* 24:631-641.

Dietrich, W. E., J. W. Kirchner, H. Ikeda, and F. Iseya. 1989. Sediment supply and the development of the coarse surface layer in gravel bed rivers. *Nature* 340:215-217.

Dingman, S. L. 1984. *Fluvial Hydrology.* W.H. Freeman, New York.

Dunne, T. and L. B. Leopold. 1978. *Water in environmental planning.* W. H. Freeman and Co., New York

Faustini, J. M. and P. R. Kaufmann. 2007. Adequacy of Visually Classified Particle Count Statistics From Regional Stream Habitat Surveys1. *Journal of the American Water Resources Association* 43:1293-1315.

Fitzpatrick, F. A., I. R. Waite, P. J. D'Arconte, M. R. Meador, M. A. Maupin, and M. E. Gurtz. 1998. Revised methods for characterizing stream habitat in the National Water-Quality Assessment Program. *Water-Resources Investigations Report 98-4052*, U.S. Geological Survey Reston, Virginia.

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

Gordon, N. D., T. A. McMahon, and B. L. Finlayson. 1992. *Stream hydrology, an introduction for ecologists.* John Wiley & Sons, New York.

Gorman, O. T. and J. R. Karr. 1978. Habitat structure and stream fish communities. *Ecology* 59:507-515.

Gregory, S. V., F. J. Swanson, W. A. McKee, and K. W. Cummins. 1991. An Ecosystem Perspective of Riparian Zones. *BioScience* 41:540-551.

Hankin, D. G. and G. H. Reeves. 1988. Estimating total fish abundance and total habitat area in small streams based on visual estimation methods. *Canadian Journal of Fisheries and Aquatic Sciences* 45:834-844.

Harmon, M. E., J. F. Franklin, F. J. Swanson, P. Sollins, S. V. Gregory, J. D. Lattin, N. H. Anderson, S. P. Cline, N. G. Aumen, J. R. Sedell, G. W. Lienkamper, K. Cormack Jr., and K. W. Cummins. 1986. Ecology of Coarse Woody Debris in Temperate Ecosystems. *Advances in Ecological Research* 15:133-302.

Harrelson, C. C., C. L. Rawlins, and J. P. Potyondy. 1994. Stream channel reference sites: an illustrated guide to field technique. *General Tech. Rep. RM-245*, USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

Hawkins, C. P., M. L. Murphy, and N. H. Anderson. 1983. Density of fish and salamanders in relation to riparian canopy and physical habitat in streams of the northwestern United States. *Canadian Journal of Fisheries and Aquatic Sciences* 40:1173-1186.

Hawkins, C. P., J. L. Kershner, P. A. Bisson, M. D. Bryant, L. M. Decker, S. V. Gregory, D. A. McCullough, C. K. Overton, G. H. Reeves, R. J. Steedman, and M. K. Young. 1993. A Hierarchical Approach to Classifying Stream Habitat Features. *Fisheries* 18:3-12.

Herlihy, A.T., S.G. Paulsen., J. Van Sickle, J.L. Stoddard. C.P. Hawkins, L.L. Yuan. 2008. Striving for consistency in a national assessment: the challenges of applying a reference-condition approach at a continental scale. *J. N. Am. Benthol. Soc.* 27(4):860–877.

Hughes, R. M., and D. V. Peck. 2008. Acquiring data for large aquatic resource surveys: the art of compromise among science, logistics, and reality. *Journal of the North American Benthological Society* 27:837–859.

Hynes, H. B. N. 1970. *The ecology of running waters*. Unversity of Toronto Press, Toronto, Ontario, Canada.

Kappesser, G. B. 2002. A riffle stability index to evaluate sediment loading to streams. *Journal of the American Water Resources Association* 38:1069-1081.

Kaufmann, P. R. 1987a. *Channel morphology and hydraulic characteristics of torrent-impacted forest streams in the Oregon Coast Range,* U.S.A. Ph.D. Dissertation. Oregon State University, Corvallis, Oregon.

Kaufmann, P. R. 1987b. Slackwater habitat in torrent-impacted streams. Pages 407-408 In R. L. Beschta, T. Blinn, G. E. Grant, F. J. Swanson, and G. E. Ice, [editors]. Erosion and Sedimentation in the Pacific Rim. International Association of Hydrologic Science, Pub. No. 165, *Proceedings of an International Symposium,* August 3-7, 1986, Ore. State Univ., Corvallis, OR. International Association of Hydrologic Science.

Kaufmann, P. R. 1993. Physical habitat. Pages 59-69 In R. M. Hughes. *Stream Indicator and Design Workshop*. EPA 600/R-93/138, U.S. Environmental Protection Agency, Office of Research and Development, Corvallis, Oregon.

Kaufmann, P. R. and J. M. Faustini, 2012. Simple measures of channel habitat complexity predict transient hydraulic storage in streams. *Hydrobiologia*. 685:69-95.

Kaufmann, P.R. and R.M. Hughes. 2006. Geomorphic and Anthropogenic Influences on Fish and Amphibians in Pacific Northwest Coastal Streams. In: R.M. Hughes, L. Wang and P.W. Seelbach (Editors), Landscape Influences on Stream Habitats and Biological Assemblages. *American Fisheries Society Symposium* 48:429-455, Bethesda, Maryland.

Kaufmann, P. R. and T. R. Whittier. 1997. Habitat Assessment. Pages 5-1 to 5-26 In J. R. Baker, D. V. Peck, and D. W. Sutton. *Environmental Monitoring and Assessment Program -Surface Waters: Field Operations Manual for Lakes*. EPA/620/R-97/001, U.S. Environmental Protection Agency, Washington, D.C.

Kaufmann, P.R., D.P. Larsen, and J.M. Faustini, 2009. Bed Stability and Sedimentation Associated With Human Disturbances in Pacific Northwest Streams. *J. Am. Water Resources Assoc.* 45(2):434-459.

Kaufmann, P. R., J. M. Faustini, D. P. Larsen, and M. A. Shirazi. 2008. A roughness-corrected index of relative bed stability for regional stream surveys. *Geomorphology* 199:150-170.

Kaufmann, P. R., P. Levine, E. G. Robison, C. Seeliger, and D. V. Peck. 1999. *Quantifying physical habitat in wadeable streams*. EPA/620/R-99/003, U.S. Environmental Protection Agency, Washington, D.C.

Keim, R. F., A. E. Skaugset & D. S. Bateman, 2002. Physical aquatic habitat II. Pools and cover affected by large woody debris in three western Oregon streams. *North American Journal of Fisheries Management* 22: 151–164.

Kovalenko, K.E., S.M. Thomaz, and D.M. Warfe. 2012. Habitat complexity: approaches and future directions – editorial review. *Hydrobiologia* 685:1–17. DOI 10.1007/s10750-011-0974-z

Larsen, D. P., P. R. Kaufmann, T. M. Kincaid, and N. S. Urquhart. 2004. Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. *Canadian Journal of Fisheries and Aquatic Sciences* 61:283-291.

Lazorchak, J.M., D.J. Klemm, and D.V. Peck. 1998. *Environmental Monitoring and Assessment Program-Surface Waters: field operations and methods for measuring the ecological condition of wadeable streams.* EPA/620/R-94/004F, U.S. Environmental Protection Agency, Washington, D.C.

Leopold, L. B. 1994. A View of the River. Harvard University Press, Cambridge, Massachusetts.
Leopold, L. B., M. G. Wolman, and J. P. Miller. 1964. *Fluvial processes in geomorphology*. W.H. Freeman, San Francisco.

Lisle, T. E. 1982. Effects of aggradation and degradation on riffle-pool morphology in natural gravel channels, northwestern California. *Water Resources Research* 18:643-1651.

Lisle, T. E. 1987. Using "residual depths" to monitor pool depths independently of discharge. *Research Note PSW-394*, USDA Forest Service, Pacific Southwest Forest and Range Experimental Station, Berkeley, California.

Lisle, T. E. and S. Hilton. 1992. The volume of fine sediment in pools: an index of sediment supply in gravel-bed streams. *Water Resources Bulletin* 28:371-383.

MacDonald, L. H., A. W. Smart, and R. C. Wismar. 1991. *Monitoring Guidelines to Evaluate Effects of Forestry Activities on Streams in the Pacific Northwest and Alaska*. EPA 910/9-91-001, U.S. Environmental Protection Agency, Region X, Seattle, Washington.

Madej, M. A. 2001. Development of channel organization and roughness following sediment pulses in single-thread, gravel bed rivers. *Water Resources Research* 37:2259-2272.

Montgomery, D. R. and J. M. Buffington. 1993. Channel classification, prediction of channel response, and assessment of channel condition. Washington State Timber/Fish/Wildlife Agreement, *Report TFW-SH10-93-002,* Department of Natural Resources, Olympia, Washington.

Montgomery, D. R. and J. M. Buffington. 1997. Channel-reach morphology in mountain drainage basins. *Geological Society of America Bulletin 109*.

Montgomery, D. R. and J. M. Buffington. 1998. Channel processes, classification, and response. Pages 13-42 In R. Naiman and R. Bilby, [editors]. *River Ecology and Management*. Springer-Verlag, New York.

Moore, K. M. S. and S. V. Gregory. 1988. Summer habitat utilization and ecology of cutthroat trout fry (*Salmo clarki*) in Cascade mountain streams. *Canadian Journal of Fisheries and Aquatic Sciences* 45:1921-1930.

Morisawa, M. 1968. *Streams, their dynamics and morphology*. McGraw-Hill Book Company, New York.

Mossop, B. and M. J. Bradford, 2006. Using thalweg profiling to assess and monitor juvenile salmon (*Oncorhynchus spp*.) habitat in small streams. *Canadian Journal of Fisheries and Aquatic Sciences* 63: 1515–1525.

Naiman, R. J., H. Decamps, J. Pastor, and C. A. Johnston. 1988. The potential importance of boundaries to fluvial ecosystems. *Journal of the North American Benthological Society* 7:289-306.

O'Neill, M. P. and A. D. Abrahams. 1984. Objective identification of pools and riffles. *Water Resources Research* 20:921-926.

Paulsen, S.G., A. Mayio, D.V. Peck, J.L. Stoddard, E.Tarquinio, S.M. Holdsworth, J. Van Sickle, L.L. Yuan, C.P. Hawkins, A.T. Herlihy, P.R. Kaufmann, M.T. Barbour, D.P. Larsen, and A.R. Olsen. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. *J. N. Am. Benthological Soc.* 27(4):812–821.

Pearsons, T. N. and G. M. Temple, 2007. Impacts of early stages of salmon supplementation and reintroduction programs on three trout species. *North American Journal of Fisheries Management* 27: 1–20.

Pearsons, T. N. and G. M. Temple, 2010. Changes to Rainbow Trout abundance and salmonid biomass in a Washington watershed as related to hatchery salmon supplementation. *Transactions of the American Fisheries Society* 139: 502–520.

Peck, D. V., D. K. Averill, A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, M. R. Cappaert, T. Magee, and P. A. Monaco.(in press) *Environmental Monitoring and Assessment Program: Surface Waters Western Pilot Study—field operations manual for nonwadeable streams.* EPA 620/ R-xx/xxx, U.S. Environmental Protection Agency, Washington, D.C.

Peck, D. V., A. T. Herlihy, B. H. Hill, R. M. Hughes, P. R. Kaufmann, D. J. Klemm, J. M. Lazorchak, F. H. McCormick, S. A. Peterson, P. L. Ringold, T. Magee, and M. R. Cappaert. 2006. *Environmental Monitoring and Assessment Program: Surface Waters Western Pilot Study—field operations manual for wadeable streams.* EPA/620/R-06/003, U.S. Environmental Protection Agency, Washington, D.C.

Platts, W. S., W. F. Megahan, and G. W. Minshall. 1983. Methods for evaluating stream, riparian, and biotic conditions. *General Technical Report INT-138*, U.S. Department of Agriculture, Forest Service, Intermountain Forest and Range Experiment Station, Ogden, Utah.

Rinne, J. 1988. Effects of livestock grazing exclosure on aquatic macroinvertebrates in a montane stream, New Mexico. *Great Basin Naturalist* 48:146-153.

Robison, E. G. and R. L. Beschta. 1989. Estimating stream cross sectional area from wetted width and thalweg depth. *Physical Geography* 10:190-198.

Robison, E. G. and P. R. Kaufmann. 1994. Evaluating two objective techniques to define pools in small streams. Pages 659-668 In R. A. Marston and V. A. Hasfurther, [editors]. Effects of Human Induced Changes on Hydrologic Systems. *Summer Symposium proceedings, American Water Resources Association*. June 26-29, 1994, Jackson Hole, Wyo.

Stack, W. R. and R. L. Beschta. 1989. Factors influencing pool morphology in Oregon coastal streams. Pages 401-411 In W. W. Woessner and D. F. Potts, [editors]. *Headwaters Hydrology Symposium. American Water Resources Association*.

Stoddard, J. L., D. V. Peck, A. R. Olsen, D. P. Larsen, J. Van Sickle, C. P. Hawkins, R. M. Hughes, T. R. Whittier, G. Lomnicky, A. T. Herlihy, P. R. Kaufmann, S. A. Peterson, P. L. Ringold, S. G. Paulsen, and R. Blair. 2005a. *Environmental Monitoring and Assessment Program (EMAP): western streams and rivers statistical summary.* EPA 620/R-05/006, U.S. Environmental Protection Agency, Washington, D.C.

Stoddard, J. L., D. V. Peck, S. G. Paulsen, J. Van Sickle, C. P. Hawkins, A. T. Herlihy, R. M. Hughes, P. R. Kaufmann, D. P. Larsen, G. Lomnicky, A. R. Olsen, S. A. Peterson, P. L. Ringold, and T. R. Whittier. 2005b. *An ecological assessment of western streams and rivers.* EPA 620/R-05/005, U.S. Environmental Protection Agency, Washington, D.C.

Suttle, K. B., M. E. Power, J. M. Levine, and C. McNeely. 2004. How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. *Ecological Applications* 14:969–974.

U.S. EPA. 2004. *Wadeable Streams Assessment: field operations manual.* EPA/841/B-04/004, U.S. Environmental Protection Agency, Washington, D.C.

U.S. EPA. 2006. *Wadeable Streams Assessment: a collaborative survey of the Nation's streams.* EPA/641/B-06/002, U.S. Environmental Protection Agency, Washington, D.C.

U.S. EPA 2007. *National Rivers and Streams Assessment: Field Operations Manual* EPA-841-B-07-009. U.S. Environmental Protection Agency, Washington, D.C.

Wilcock, P. R. 1997. The components of fractional transport rate. *Water Resources Research* 33:247-258.

Wilcock, P. R. 1998. Two-fraction model of initial sediment motion in gravel-bed rivers. *Science* 280:410-412.

Table 11.1 Metrics used to characterize the general attributes of stream/river physical habitat

Habitat Volume:
- *LRP100 = log(RP100)* = Log of Mean Residual Depth (cm)

Scaled Habitat Volume:
- *LDVRP100 = log(RP100) – log(Predicted RP100)* = Deviation in Mean Residual Depth from expected value

Habitat Complexity:
- *CVDPTH = SDDEPTH / XDEPTH* = Coefficient of Thalweg Depth Variation
- *C1WM100* = Number of Large Woody Debris pieces/100m of channel.
- *LV1W_MSQ* = log[Volume of Large Woody Debris per m^2 of bankfull channel area (m^3/m^2)].
- *XFC_NAT* = Areal Cover of Woody Debris, Brush, Undercut Banks, Overhanging Vegetation, plus Boulders and Rock Ledges.
- *XFC_NORK* = Areal Cover of Woody Debris, Brush, Undercut Banks, Overhanging Veg.
- *XFC_AQM* = Areal Cover of Aquatic Macrophytes
- *XFC_ALG* = Areal Cover of Filamentous Algae detectable by the unaided eye.

Streambed Particle Size:
- *LSUB_dmm* = log[Streambed surface particle D_{gm} – mm] = log of geometric mean diameter of bed surface sediments in millimeters.
- *PCT_FN* = % Streambed Silt & Finer
- *PCT_SAFN* = % Streambed Sand & Finer
- *XEMBED* = % Substrate Embedded by Sand and Fines

Scaled Streambed Particle Size:
- *DPCT_FN* = Deviation of *PCT_FN* from expected value ("excess Fines")
- *DPCT_SF* = Deviation of *PCT_SAFN* from expected value ("excess Sand+Fines")
- *DEVLSUB* = Deviation of *LSUB_DMM* from expected value (Streambed Fining Index)

Relative Bed Stability:
- *LRBS*= $\log_{10}$ of diameter ratio: Geometric mean bed particle diameter / Critical (mobile) diameter at bankfull flow stage. (LRBS_bw5: see Kaufmann *et al*. 1999; LRBS_g08: see Kaufmann *et al*. 2008, 2009).

Floodplain Interaction:
- *LSINU = Log(SINU)* = Log(Channel Sinuosity).
- *LINCIS_H = log(XINC_H - XBKF_H + 0.1)* = Log of Incision from terrace to bankfull ht (m).
- *LBFWDRAT = log{BKF_W / BKF_H+(XDEPTH/100)}*= log (Bankfull Width/Depth Ratio)
- *LBFXWRAT = log(BKF_W / XWIDTH)*= log (Bankfull Width / Wetted Width) (an index of streamside flood inundation potential)

Hydrologic Regime:
- *LQSLTR_RAT = log{(Qsp+0.0000001)/LTROFF_M}*=log{low flow /annual mean runoff} (~ an inverse index of "droughtiness",
  where: *Qsp = Flow_mps/WSAREAKM= (flow_cfs/35.315)/WSAREAKM*
- *LBFXDRAT =log{(XBKF_H+(XDEPTH/100) / (XDEPTH/100)}*= log(ratio of bankfull depth / wetted depth), a morphometric index of "flashiness".

Riparian Vegetation:
- *XCDENMID*: % Canopy Density measured midstream.
- *XCMG* = Riparian Canopy+Mid-+Ground Layer Vegetation (areal cover proportion)
- *XCMGW* = Riparian Canopy+Mid+Ground Layer Woody Veg.(areal cover proportion)

Riparian Habitat Alteration:
- $QR1=(QRVEG1*QRVEG2*QRDIST1)^{0.3333}$; where:
  if *XCMGW* <=2.00 then *QRVeg1*=.1+(0.9(*XCMGW* / 2.00));
  if *XCMGW* >2.00 then *QRVeg1*=1;
- *QRVeg2*=.1+(0.9(*XCDENBK / 100*)); and *QRDIST1*=1/(1+*W1_HALL*)

Riparian Human Disturbances:
- *W1_HAG* = Riparian & near-Stream Agriculture – all types (proximity-weighted tally)
- *W1H_ROAD* = Riparian & near-Stream Roads (proximity-weighted tally)
- *W1H_CROP* = Riparian & near-Stream Row Crop Agriculture (proximity-weighted tally)
- *W1H_WALL* = Riparian & near-Stream Walls, Dikes, Revetment (proximity-weighted tally)
- *W1_HALL* = Proximity-weighted Index of Human Disturbances of All Types
- *QRDIST1* = 1/(1+*W1_HALL*) = Proximity-weighted Inverse Index of Human Disturbances of All Types

Table 11.2 Sampling revisit precision (repeatability) of the four physical habitat condition indicators

Repeat visits within the summer sampling season were used to calculate RMS_{rep}, which is essentially the standard deviation of repeat sampling pairs to the same stream or river reach. Dividing the square of the RMS_{rep} into the variance among sites gives the S/N variance ratio. (See Kaufmann *et al.* 1999 for ANOVA methods to calculate RMS_{rep} and S/N, where RMS_{rep} is equal to their RMSE.)

Metric	Group	Sites (n)	mean	Repeat pairs (n)	RMS_{rep}	S/N
LRBS_g08	**All Sites**	**1945**	**-0.776**	**177**	**0.482**	**4.97**
	Boatable	711	-0.636	89	0.450	7.36
	Wadeable	1234	-0.860	88	0.512	3.31
	EHIGH	534	-0.397	70	0.500	4.19
	PLNLOW	1002	-1.014	74	0.494	4.67
	WMTNS	409	-0.712	33	0.411	6.07
L_xfc_nat	**All Sites**	**2113**	**-0.590**	**197**	**0.240**	**1.87**
	Boatable	782	-0.575	93	0.242	1.22
	Wadeable	1331	-0.599	104	0.238	2.29
	EHIGH	555	-0.460	73	0.209	0.92
	PLNLOW	1125	-0.675	86	0.263	1.78
	WMTNS	433	-0.545	38	0.241	1.77
L_xcmgw	**All Sites**	**2113**	**-0.286**	**197**	**0.146**	**9.38**
	Boatable	782	-0.175	93	0.155	4.28
	Wadeable	1331	-0.353	104	0.137	13.20
	EHIGH	555	-0.062	73	0.092	6.01
	PLNLOW	1125	-0.381	86	0.174	8.53
	WMTNS	433	-0.340	38	0.162	5.74
L_W1_Hall	All Sites	**2113**	**-0.152**	**197**	**0.186**	**5.18**
	Boatable	782	-0.123	93	0.145	7.99
	Wadeable	1331	-0.170	104	0.216	3.89
	EHIGH	555	-0.108	73	0.184	5.08
	PLNLOW	1125	-0.189	86	0.171	6.02
	WMTNS	433	-0.116	38	0.220	4.00

Table 11.3 Estimated number of years to detect trends in habitat attributes

Number of years required for a 50-site monitoring network to detect 1% and 2% per year trends in habitat attributes with 80% likelihood (*beta,* or power) and *alpha* = 0.05, if specified trends occur, and sites are visited each year. Data were taken from Larsen *et al.* (2004),[a] or calculated using the same data and analytical procedures used in that publication.[b]

Variable	Description	1% trend	2% trend
SDDEPTH[b]	(Std. Deviation of Thalweg Depth)	13 years	8 years
LRP100[a]	(log[Mean Residual Depth])	20	12
PCT_SAFN[a]	(% Sand + Silt)	21	13
XEMBED[b]	(% Embeddedness)	20	12
LRBS_BW5[b]	(log[Rel. Bed Stability])	12	8
LV1W_MSQ[a]	(log[Large Wood Volume/m^2])	27	17
XCMGW[b]	(3-Layer Riparian Woody Veg Areal Cover)	12	8
XCDENMID[a]	(Canopy Density measured midstream)	13	8

Table 11.4 Anthropogenic disturbance screening criteria

Criteria used_to characterize least-disturbed reference (R), moderately disturbed (M), and most-disturbed (D) sample reaches for developing physical habitat condition criteria.

- Values > than those before the slash (/) are EXCLUSION criteria for reference sites.
- Values ≥ those after slash are INCLUSION criteria for most-disturbed sites.
- W, B, and G refer to Wadeable, Boatable, and Great River sites.

Region	PTL	NTL	Cl	SO4	Turb	W1_HALL	W1_HAG Wadeable	W1H_CROP Wadeable	W1H_WALL Wadeable	PCTCROP	PCTURB	DamScreen
NAP	20/100	750/3500	250/10000	250/1000	5/10	2.0/4.0	0.1/0.4	0.05/0.10	0.2/0.4	15/67	5/25	1/1
SAP	20/100	750/3500	200/1000	400/1000	5/20	2.0/4.0	0.1/0.4	0.05/0.10	0.2/0.4	15/67	5/25	1/1
UMW	50/150	1000/5000	300/2000	400/2000	5/30	2.0/4.0	0.15/1.4	0.1/0.4	0.2/0.4	15/67	5/25	1/1
CPL	75/250	2500/8000	999999/ 999999	600/4000	10/50	2.0/4.0	0.15/1.4	0.05/ 0.4	0.2/0.4	15/67	5/25	1/1
TPL	100/500	3000/15000	2000/5000	999999/ 999999	50/100	2.0/4.0	0.67/1.4	0.25/0.48	0.4/0.6	15/67	5/25	1/1
SPL	150/500	4500/10000	1000/5000	999999/ 999999	50/100	2.0/3.0	1.0/1.4	0.15/ 0.25	0.2/0.4	15/67	5/25	1/1
WMT:												
Southwest	50/100	750/1500	300/1000	99999/ 99999	5/10	W:0.5/3.0 B,G:1.5/3.0	0.25/1.4	0.10/0.25	0.2/0.4	15/67	5/25	1/1
S.Rockies	25/100	750/1500	200/1000	200/1000	5/10	W:1.0/3.0 B,G:1.5/3.0	0.3/1.4	0.1/0.25	0.2/0.4	15/67	5/25	1/1
N.Rockies & Pacific	25/100	750/1500	200/1000	200/1000	5/10	W:0.5/3.0 B,G:1.5/3.0	0.3/1.4	0.10/0.25	0.2/0.4	15/67	5/25	1/1
XER	50/150	1500/5000	1000/5000	999999/ 999999	25/75	1.5/3.0	0.6/1.4	0.15/0.25	0.2/0.4	15/67	5/25	1/1

Table 11.5 Responsiveness to levels of human disturbance

Responsiveness of NRSA physical habitat condition metrics to levels of human disturbance, as quantified by *t*-values of the difference between means of least-disturbed reference sites (RMD_PHab=R) minus more-disturbed sites (those screened as RMD_PHab=D). Values shown in red have a sign contrary to expectations.

Metric	Region	t-value R-D (Boatable)	t-value R-D (Wadeable)	t-value R-D (All sites)
LRBS_g08	USA-48	+5.12	+5.82	+7.85
	CPL	+2.36	+0.37	+2.11
	EHIGH (NAP+SAP)	+3.03	+3.01	+4.29
	INTPLNUMW (NPL,SPL,TPL,UMW)	+3.09	+3.68	+4.65
	West (WMT+XER)	+3.19	+5.07	+5.21
LXFC_Nat_OE	USA-48	-3.02	+6.66	+4.25
	CPL	+0.18 (ns)	+3.78	+3.09
	EHIGH (NAP+SAP)	-1.79	-0.12 (ns)	-1.26
	INTPLNUMW (NPL,SPL,TPL,UMW)	-2.30	+3.19	+2.13
	West (WMT+XER)	-1.87	+6.94	+4.17
L_xcmgw	USA-48	+2.38	+12.61	+12.24
	CPL	+4.44	+4.71	+6.10
	EHIGH (NAP+SAP)	+0.08 (ns)	+5.19	+4.11
	INTPLNUMW (NPL,SPL,TPL,UMW)	+0.56 (ns)	+8.66	+8.24
	West (WMT+XER)	+2.16	+6.46	+6.85

Model 11.1 Reference Condition for Channel Bed Sedimentation based on Relative Bed Stability (RBS)

(LRBS_use=LRBS_g08 = *Log₁₀*(RBS_g08) =calculated according to Kaufmann *et al.* (2008)

Following are simple LRBS models (reference distributions) or regression models in which *W1_Hall* is set to zero in regressions on *W1_Hall* within *RMD_PHAB*=R --- Then mean and SD of adjusted ref mean *LRBS* becomes y-intercept and the SD about the reference mean is the RMSE of those regressions.

Coastal Plain (CPL) combined Boatable & Wadeable reference sites model:

LRBS= -0.9855 – 1.0320(*W1_Hall*)

R^2=0.2585, RMSE=0.7123, p=0.0003, df=45 (minus 3 hardpan Boatable sites with LRBS>2)

Condition classes (use y-int -0.67 x & 1.65x RMSE for 25th and 5th %-tiles of the reference distribution)

5th %-tile = - 2.161

25th %-tile= -1.463

Eastern Highlands (NAP & SAP) combined Boatable & Wadeable reference site model:

LRBS= -0.1977 + 0.3311(*W1_Hall*) --------- (note positive slope)

R^2=0.0357, RMSE=0.8445, p=0.0624, df= 97 (NO outliers removed; appears to be subclass of low RBS)

Condition classes (use y-int-0.67x & 1.65x RMSE for 25th and 5th %-tiles of the reference distribution)

5th %-tile = -1.591

25th %-tile= -0.764

Combined Upper Midwest plus Temperate and Southern Plains (UMW, TPL, SPL) combined Boatable & Wadeable reference site model:

LRBS= -0.3126 – 0.8593(*W1_Hall*)

R^2=0.0741, RMSE=1.2239, p=0.0052, df=103

Condition classes (use y-int-0.67x & 1.65x RMSE for 25th and 5th %-tiles of the reference distribution

5th %-tile = -2.332

25th %-tile= -1.133

Northern Plains (NPL) combined Boatable & Wadeable reference site model:

Simple Distrib: ref mean *LRBS*= -0.333 with SD=0.824, df=23 (1 low outlier removed)

Condition classes (use ref mean -0.67x & 1.65x SD for 25th and 5th %-tiles of the reference distribution)

5th %-tile = -1.692

25th %-tile= -0.885

The West (WMT & XER) Separate Boatable and Wadeable reference site models:

Boatable Reference site model:

LRBS= +0.5727 – 0.4064(*W1_Hall*)

R^2=0.0555, RMSE=0.6818, p=0.2677, df=23

(retained 2 low outliers given the small sample size)

(Weak model but scope of *W1_Hall* is small and the same, but stronger, relationship is observed across all sites.)

Condition classes (use y-int-0.67x & 1.65x RMSE for 25th and 5th %-tiles of the reference distribution)

5th %-tile = - 0.5523

25th %-tile= +0.1159

Wadeable Reference site model:

LRBS= -0.5207 -0.4818(*W1_Hall*)

R^2=0.0225, RMSE=0.7006, p=2486, df=60

(Weak model but scope of *W1_Hall* small, and same but stronger relationship across all sites.)

Condition classes (use y-int-0.67x & 1.65x RMSE for 25th and 5th %-iles of the reference distribution)

5th %-tile = -1.677

25th %-tile= -0.990

Model 11.2 - Reference Condition for Instream Fish Cover XFC_NAT (Transformed as $Log_{10}(0.01+XFC\_NAT)$)

Following are aggregated region Observed/Expected (O/E) models with disturbance modeled out where possible (version 6/13/12);

Coastal Plain (CPL) combined Boatable & Wadeable reference site model:
(In Reference Sites, *XFC_NAT* increases slightly in with *QR1* or *QRDIST1* -- over all sites the trend is strong but the ref sites are generally high in fish cover (no need to separate Wadeable and Boatable sites).

$LXFC\_NAT_{ref}$ =-0.6471+0.3820(*QR1*)
R^2=0.0333 RMSE=0.2656 p=0.20, df=49

Expected reference values calculated by setting *QR1* to 0.9 (90^{th} percentile = 0.89)
LXFC_NAT_E = -0.6471 +(0.3820*0.9)
LXFC_NAT_OE= *LXFC_NAT* – *LXFC_NAT_E*

Use mean zero and E-model RMSE as above
LXFC_NAT_OE 25^{th} % tile = -0.67 x 0.2656
LXFC_NAT_OE 5^{th} % tile = -1.65 x 0.2656

Eastern Highlands (NAP & SAP) combined Boatable + Wadeable reference site model:

$LXFC\_NAT_{ref}$= -0.3211 - 0.0348(*Tmax_PT*) +0.1873(*L_ELEV_PTx*) + 0.2545(*RDIST1*) + 0.5455(*QR1*)

R^2=0.2488, RMSE=0.2796, p<0.0001, df=99
Significance values for model predictors:

Tmax_PT:	p=0.0006
L_ELEV_PTx	p=0.0223
RDIST1	p=0.1609
QR1	p=0.0325

Expected reference values calculated by setting *QR1*=0.9:
LXFC_NAT_E= -0.3211 -0.0348(*Tmax_PT*) +0.1873(*L_ELEV_PTx*) +0.2545(0) +0.5455(0.90)

Mean zero and E-model RMSE:
LXFC_NAT_OE 5^{th} % tile = -1.65 x 0.2796
LXFC_NAT_OE 25^{th} % tile = -0.67 x 0.2796

Continuation of Model 11.2 - Reference Condition for Instream Fish Cover *XFC_NAT* (Transformed as Log10(*0.01+XFC*_NAT))

Interior Plains plus Upper Midwest (TPL, NPL, SPL, UMW) combined Boatable & Wadeable reference site model:

$LXFC\_NAT_{ref}$ = 0.7371 -0.0399(*Tmax_PT*) +0.0251(*Tmin_PT*) -0.0005(*Psum_PTx*) +0.7052(*QR1*)

Where:
PSUM_PTx=PsumPY_PT;
IF *PsumPY_PT*=. then *Psum_PTx=PSUM_WSx*;

R^2=0.2369, RMSE=0.3168, p<0.0001, df=128
Significance values for model predictors:

Tmax_PT:	p=0.0153
Tmin_PT:	p=0.0051
PSum_PTx	p<0.0001
QR1	p=0.0018

Expected reference values calculated by setting *QR1* to 0.80 (Note that 0.80 is the 95th percentile of *QR1* among Ref sites in *'INTPLNUMW'*)

LXFC_NAT_E = 0.7371 -0.0399(*Tmax_PT*) +0.0251(*Tmin_PT*) -0.0005(*Psum_PTx*) +0.7052(0.80)

Mean reference site *LXFC_NAT_OE*= -0.0923 with SD= 0.3701
But that SD has disturbance variance in it -- use mean 0.0 and E-model RMSE:

LXFC_NAT_OE 5th % tile = -1.65 x 0.3168
LXFC_NAT_OE 25th % tile = -0.67 x 0.3168

The West (WMT+XER) combined Boatable & Wadeable reference site model):

$LXFC\_NAT_{ref}$ = -1.10469 +0.0153(*LAT_DD83*) -0.0166(*Tmin_PT*) -0.5236(*P_REALM*) -0.5817(*W1_Hall*) +1.1400(*RDIST1*)

R^2=0.5699, RMSE=0.2387, p<0.0001, df=81
Significance values for model predictors:

LAT_DD83	p=0.0291
Tmin_PT:	p=0.0015
P_REALM	p<0.0001
W1_Hall	p=0.0360
RDIST1	p=0.0421

Expected reference values calculated by setting *W1_Hall*=0 and *RDIST1*=0;

LXFC_NAT_E= -1.10469 +0.0153(*LAT_DD83*) -0.0166(*Tmin_PT*) -0.5236(*P_REALM*) -0.5817(0) +1.1400(0)

Mean reference site *LXFC_NAT_OE* = 0.0353 with SD= 0.2683
But that SD has disturbance variance in it -- use E-model RMSE and use mean 0.0.

LXFC_NAT_OE 5th % tile = -1.65 x 0.2387
LXFC_NAT_OE 25th % tile= -0.67 x 0.2387

Model continues on the next page…

Continuation of Model 11.2 - Reference Condition *for Instream* Fish Cover XFC_*NAT* (*Transformed* as
Log$_{10}$(0.01+XFC_NAT))

Condition Class Thresholds for *XFC_NAT*:

Modeled 25th and 5th percentiles of the reference distribution:

LXFC_NAT_OE$_{ref}$ 5th % tile = Mean *LXFC_NAT_OE$_{ref}$* -1.65(SD$_{ref}$)

LXFC_NAT_OE$_{ref}$ 25th % tile = Mean *LXFC_NAT_OE$_{ref}$* -0.67(SD$_{ref}$);

Condition Criteria for *LXFC_NAT*:

If LXFC_NAT_OE < *LXFC_NAT_OE$_{ref}$* 5th % tile then *FCvrCOND* = 'P'

If LXFC_NAT_OE >= *LXFC_NAT_OE$_{ref}$* 5th % tile and *LXFC_NAT_OE* < *LXFC_NAT_OE$_{ref}$* 25th % tile then F*CvrCOND*='M'

If *LXFC_NAT_OE* >= *LXFC_NAT_OE$_{ref}$* 25th % tile then *FCvrCOND*='G';

If *LXFC_NAT_OE* =. then *FCvrCOND*='Z';

Define *RDIST1* and QR1:

We calculated a composite riparian condition index (*QR1*) from the reach summary data describing the cover and structure of riparian vegetation and a proximity-weighted tally of streamside human activities. *QR1* has a theoretical minimum approaching zero where there is no riparian vegetation and very high values of *W1_Hall,* the proximity weighted tally of streamside human land use activities. It approaches 1.0 where there is abundant, complex riparian woody vegetation, high bankside canopy density (measured with densiometer), and no visible human land use activities or channel alterations. It is intended for use in those riparian settings in regions where reference condition is a multi-storied woody vegetation corridor (*XCMGW* approaching 2.0), with bankside canopy density (*XCDENBK*) generally complete (85%-100%) along stream banks, and along rivers above bankfull height. Reference condition is set near zero for the types of riparian human activities identified by the EMAP Physical Habitat field methods (Peck *et al.* 2006; Peck *et al.*, In Press-b). *QR1* is then defined as the geometric mean of three scaled variables as follows (the cube-root is taken to reduce extreme skewness in the product of the three component variables:

QR1= {(*QRVEG1*) (*QRVEG2*) (*QRDIST1*)} $^{0.333}$;

where:

if *XCMGW* <=2.00, then *QRVeg1*=.1+(.9 (*XCMGW*/2.00)),and

if XCMGW >2.00 then *QRVeg1*=1; and

QRVeg2=0.1 + [0.9(*XCDENBK*/100)]; and

QRDIST1=1/ (1+*W1_Hall*);

where:

W1_HALL= distance weighted tally of in-channel, riparian, and near stream human activities.

QR1 decreases with increases in streamside human activities (*W1_Hall*), and increases with increasing riparian woody vegetation complexity (*XCMGW*) and riparian cover density measured at the streambank with a canopy densiometer (*XCDENBK*).

We transformed the variable *W1_Hall,* a proximity-weighted tally of all the targeted types of human activities into an index that is more sensitive at the low end of disturbance and has a range constrained from 0 to 1. The new variable, *RDIST1*=1-{1/ (1+*W1_HallL*)}, with a value of 0 when there are no observable human disturbances, and approaches 1 as the number and extent of human disturbances increases. In the calculation of *QR1* above, we used the *inverse* measure of riparian disturbance variable *QRDIST1*=1/(1+*W1_Hall*), which has a value of 1 when there are no observable human disturbances, and approaches 0 as the number and extent of human disturbances increases.

Model 11.3 Riparian condition XCMGW (Transformed as L_xcmgw= Log10 (0.01+XCMGW)

Following are Simple *L_xcmgw* models (reference distributions) or regression models in which *W1_Hall* is set to zero in regressions on *W1_Hall* within RMD_PHAB=R --- The adjusted mean *L_xcmgw* for reference sites is defined as the y-intercept and the SD about the reference mean is defined as the RMSE of those regressions. Note that the primary reason for excluding outliers is to avoid gross overestimations of the reference SD.

Coastal Plain (CPL) combined Boatable & Wadeable Reference Site model:

L_xcmgw_{ref}= 0.0173 + 0.0846(*W1_Hall*)

(note here the *W1_Hall* slope is POSITIVE; its effect is to prevent overestimating the reference mean and its RMSE (or SD)

R^2=0.0599, RMSE=0.1342, p=0.0835, df= 50

adjusted ref mean *L_xcmgw* = 0.0173 — Antilog = 1.04 – 0.01 = *xcmgw* = 1.03
L_xcmgw SD= *L_xcmgw* RMSE= 0.1342
Est 5th %-tile = ref mean - 1.65(SD) = -0.204 — Antilog = 0.625 – 0.01 = *xcmgw* = 0.615
Est 25th %-tile = ref mean - 0.67(SD) = -0.0726 — Antilog = 0.846 – 0.01 = *xcmgw* = 0.836

Combined Northern and Southern Appalachians (NAP&SAP):

NAP&SAP Boatable Reference Site Model:

L_xcmgw_{ref}= 0.0456 -0.0138(*W1_Hall*) --- virtually the same as the null (simple) model.

R^2=0.0036, RMSE=0.1224, p=0.7347, df = 33 (excludes 3 NAP and 1 NAP low outliers)
null model: Mean *L_xcmgwref*=0.0334 and SD=0.1207, n=34 (excludes same 4 outliers)

adjusted ref mean =0.0456 — Antilog = 1.11 – 0.01 = *xcmgw* = 1.10
SD=RMSE=0.1224
Est 5th %-tile = ref mean - 1.65(SD) = -0.156 — Antilog = 0.698 – 0.01 = *xcmgw* = 0.688
Est 25th %-tile = ref mean - 0.67(SD) = -0.0364 — Antilog = 0.920 – 0.01 = *xcmgw* = 0.910

NAP&SAP Wadeable Reference Site Model:

L_xcmgw_{ref}= 0.0823 -0.2064(*W1_Hall*)

R^2 = 0.2490, RMSE=0.1284, p<0.0001, df = 61

adjusted ref mean =0.0823 — Antilog = 1.21 – 0.01 = *xcmgw* = 1.20
SD=RMSE=0.1284
Est 5th %-tile = ref mean - 1.65(SD) = -0.130 — Antilog = 0.742 – 0.01 = *xcmgw* = 0.732
Est 25th %-tile = ref mean - 0.67(SD) = -0.00373 — Antilog = 0.991 – 0.01 = *xcmgw* = 0.981

Southern Plains (SPL) combined Boatable & Wadeable Reference Site Model:

L_xcmgw_{ref}= -0.3475 +0.2271(*W1_Hall*)

(note here the *W1_Hall* slope is POSITIVE; its effect is to prevent overestimating the ref mean and its RMSE (or SD)

R^2=0.1842, RMSE=0.2565, p=0.018, df = 29 (2 very low outliers removed)

adjusted ref mean = -0.3475 — Antilog = 0.449 – 0.01 = *xcmgw* = 0.439
SD=RMSE= 0.2565
Est 5th %-tile = ref mean - 1.65(SD) = -0.771 — Antilog = 0.169 – 0.01 = *xcmgw* = 0.159
Est 25th %-tile = ref mean - 0.67(SD) = -0.519 — Antilog = 0.303 – 0.01 = *xcmgw* = 0.293

Continuation of Model 11.3 Riparian condition *XCMGW* (Transformed as *L_xcmgw= Log10* (0.01+XCMGW)

<u>Combined Upper Midwest, Northern Plains and Temperate Plains (UMW, NPL TPL)</u>:

<u>UMW,NPL, &TPL Boatable reference site model</u>:

L_xcmgw_{ref}= -0.0526 – 0.2840(W1_Hall)
R^2= 0.2098, RMSE=0.2664, p=0.0125, df = 28 (excludes 2 low outliers)

adjusted ref mean= -0.0526 — Antilog = 0.886 – 0.01 = *xcmgw* = 0.876
SD=RMSE=0.2664
Est 5th %-tile = ref mean - 1.65(SD) = -0.492 — Antilog = 0.322 – 0.01 = *xcmgw* = 0.312
Est 25th %-tile = ref mean - 0.67(SD) = -0.231 — Antilog = 0.587 – 0.01 = *xcmgw* = 0.577

<u>UMW,NPL, &TPL Wadeable reference site model</u>:

L_xcmgw_{ref}= -0.1210 - 0.3276(*W1_Hall*)
R^2=0.2016, RMSE=0.2607, p<0.0001, df= 67 (excludes 2 low outliers)

adjusted ref mean= -0.1210 — Antilog = 0.757 – 0.01 = *xcmgw* = 0.747
SD=RMSE=0.2607
Est 5th %-tile = ref mean - 1.65(SD) = -0.551 — Antilog = 0.281 – 0.01 = *xcmgw* = 0.271
Est 25th %-tile = ref mean - 0.67(SD) = -0.296 — Antilog = 0.506 – 0.01 = *xcmgw* = 0.496

<u>Western Mountain (WMT) combined Boatable & Wadeable reference site model</u>:

L_xcmgw_{ref}= -0.1033 - 0.1586(*W1_Hall*)
R^2= 0.0716, RMSE=0.2356, p=0.0461, df= 55

adjusted ref mean = -0.1033 — Antilog = 0.788 – 0.01 = *xcmgw* = 0.778
SD=RMSE=0.2356
Est 5th %-tile = ref mean - 1.65(SD) = -0.492 — Antilog = 0.322 – 0.01 = *xcmgw* = 0.312
Est 25th %-tile = ref mean - 0.67(SD) = -0.261 — Antilog = 0.548 – 0.01 = *xcmgw* = 0.538

<u>Xeric region (XER) combined Boatable & Wadeable reference site model</u>

L_xcmgw_{ref}= -0.1006 – 0.2467(*W1_Hall*)
R^2=0.1528, RMSE=0.2083, p=0.0397, df= 27 (1 low outlier removed)

adjusted ref mean = -0.1006 — Antilog = 0.793 – 0.01 = *xcmgw* = 0.783
SD=RMSE=0.2083
Est 5th %-tile = ref mean - 1.65(SD) = -0.444 — Antilog = 0.360 – 0.01 = *xcmgw* = 0.350
Est 25th %-tile = ref mean - 0.67(SD) = -0.240 — Antilog = 0.575 – 0.01 = *xcmgw* = 0.565

Condition Class Thresholds for XCMGW:

Modeled 25th and 5th percentiles of the reference distribution:
L_xcmgw_{ref} 5th % tile = Mean L_xcmgw_{ref} -1.65(SD_{ref})
L_xcmgw_{ref} 25th % tile = Mean L_xcmgw_{ref} -0.67(SD_{ref});

Condition Criteria for XCMGW:

If *L_xcmgw* < $Lxcmgw_{ref}$ 5th % tile then *RIPCOND* = 'P'
If *L_xcmgw* >= $Lxcmgw_{ref}$ 5th % tile and *L_xcmgw* < $Lxcmgw_{ref}$ 25th % tile then *RIPCOND*='M'
If *L_xcmgw* >= Lxcmgw_{ref} 25th % tile then *RIPCOND*='G';
If *L_xcmgw* =. then *RIPCOND* ='Z'

Model 11.4 Condition Thresholds for Riparian Human Disturbances (***RDIST_COND*** based on ***W1_Hall***).

We applied uniform condition thresholds used nationwide. The Low (L), Medium (M), and High (H) disturbance levels are analogous to the Good, Fair, Poor condition classification used for the other indicators.

If *W1_Hall* < 0.33 then *RDIST_COND*= 'L';

If *W1_Hall* >= 0.33 and *W1_Hall* < 1.5 then *RDIST_COND*= 'M';

If *W1_Hall* >= 1.5 then *RDIST_COND* = 'X';

If *W1_Hall*=. then *RDIST*_COND= 'Z';

Figures 11.1 through 11.5 on following pages compare **Indicator Responses for Reference and Disturbed Sites Nationally and for the 4 combined modelling regions used in Table 11.5.** Box and whisker plots in each panel show medians and 5th, 25th, 75th, and 95th percentiles for least-disturbed reference (R), moderately-disturbed (M), and highly-disturbed (D) sample sites. Plots show t values for the differences between R and D means for three indicators. Values shown in red have a sign contrary to expectations. The plots are shown separately for boatable and wadeable rivers and streams. The indicators are:

- LRBS_use = LRBS_g08, the indicator of relative bed stability and excess streambed fine sediments.
- LXCF_Nat_OEregdis = Log10 of observed/expected XFC_NAT, the indicator of instream habitat cover complexity.
- L_xcmgw = Log10(0.01 + XCMGW), the indicator of riparian vegetation cover and structure.
- W1_HALL = the proximity weighted indicator of riparian and near-shore human disturbance intensity (no t-test shown because W1_HALL was used in defining R, M, and D sites.

Figure 11.1 Indicator Responses for Reference and Disturbed Sites: National

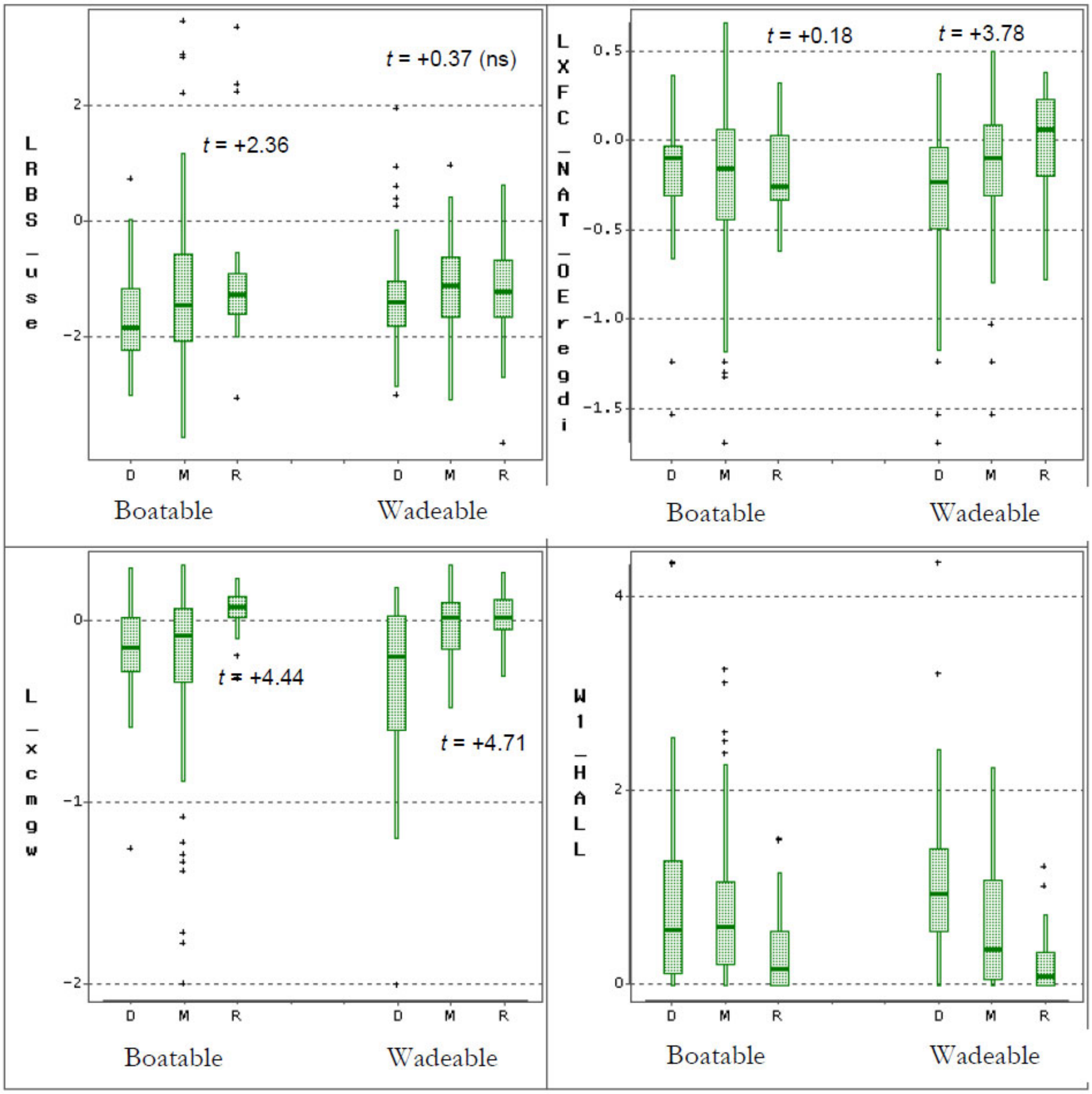


Figure 11.2 Indicator Responses for Reference and Disturbed Sites: Coastal Plains

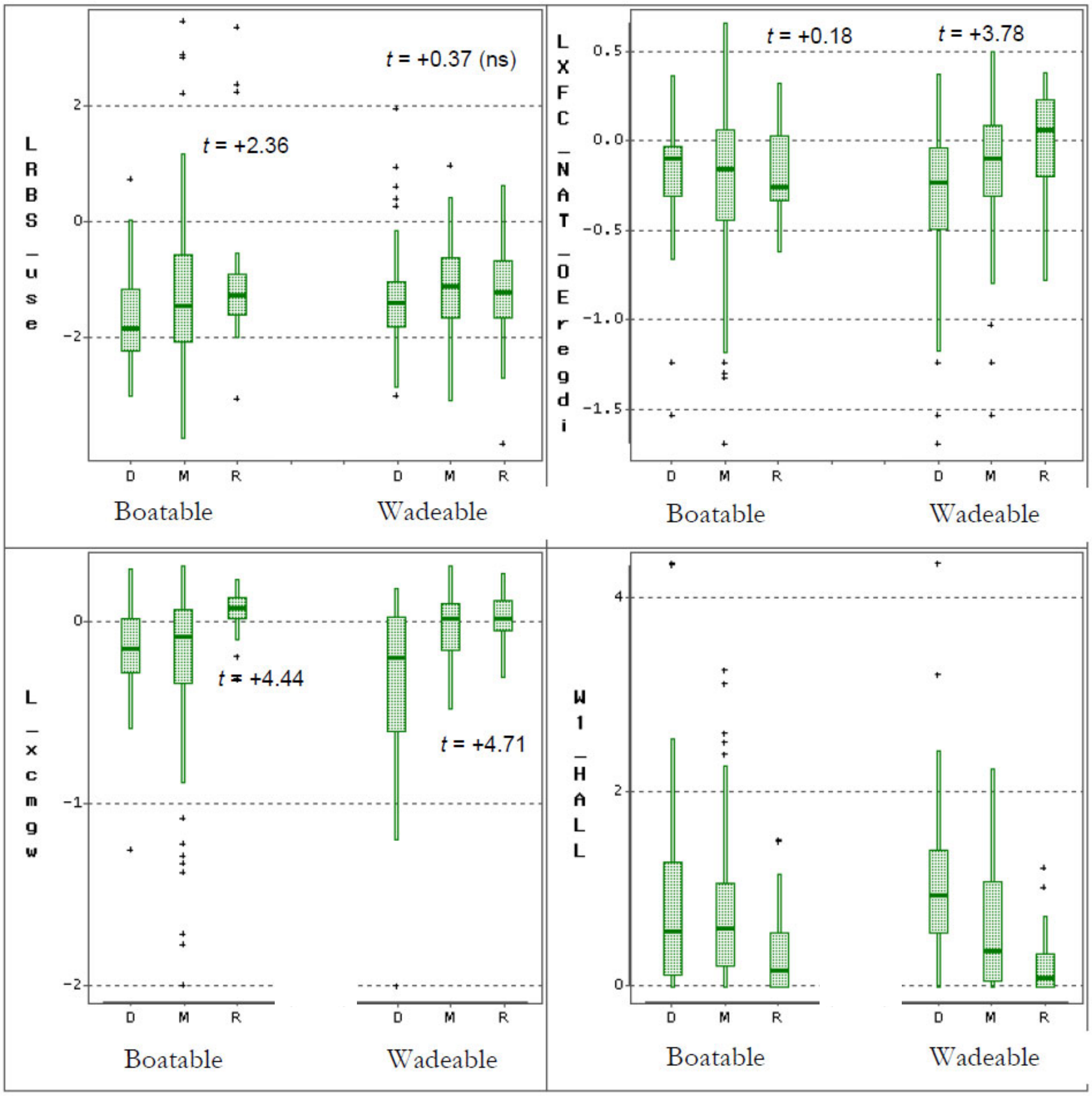


Figure 11.3 Indicator Responses for Reference and Disturbed Sites: Eastern Highlands

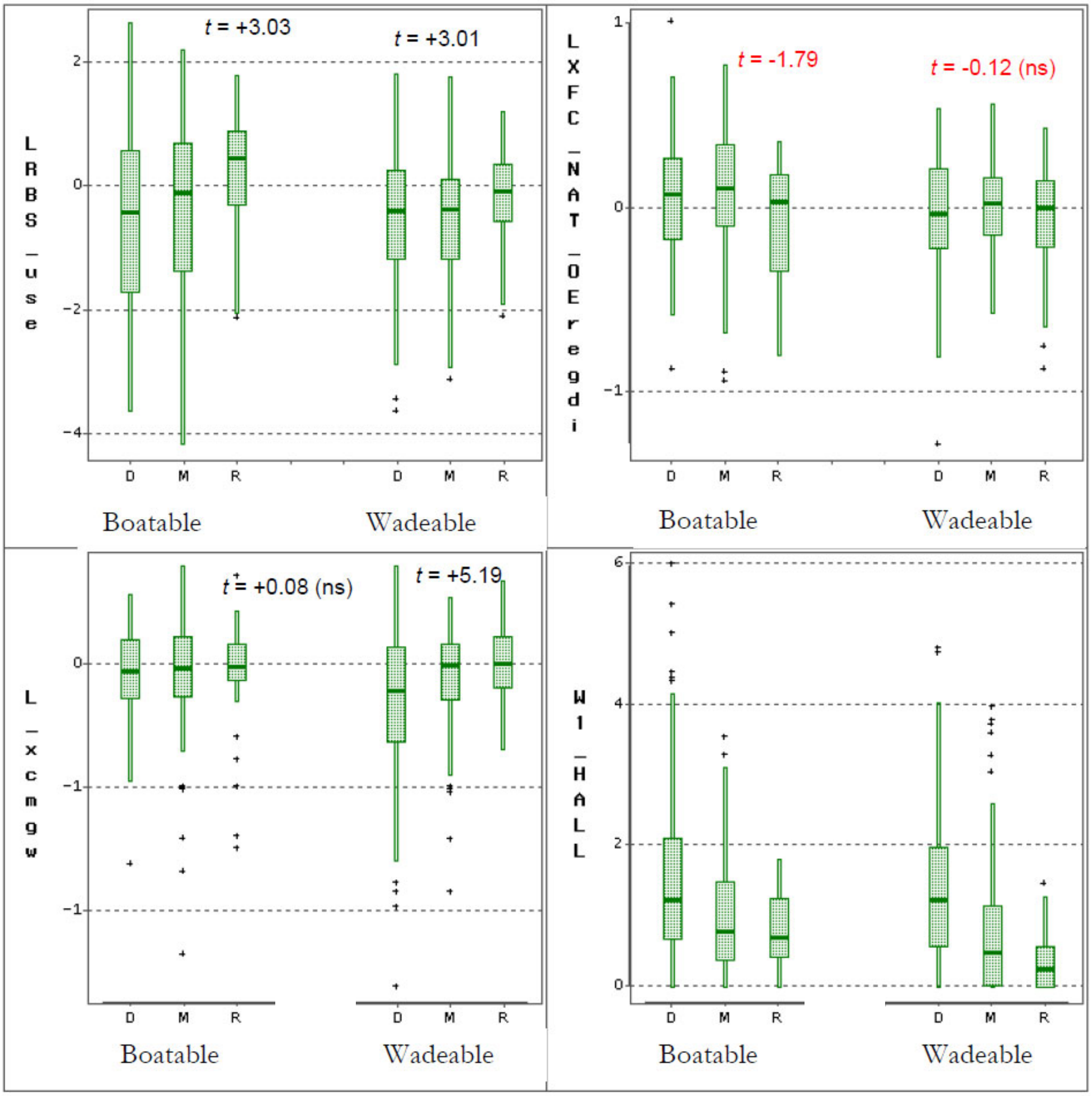


Figure 11.4 Indicator Responses for Reference and Disturbed Sites: Interior Plains + Upper Midwest

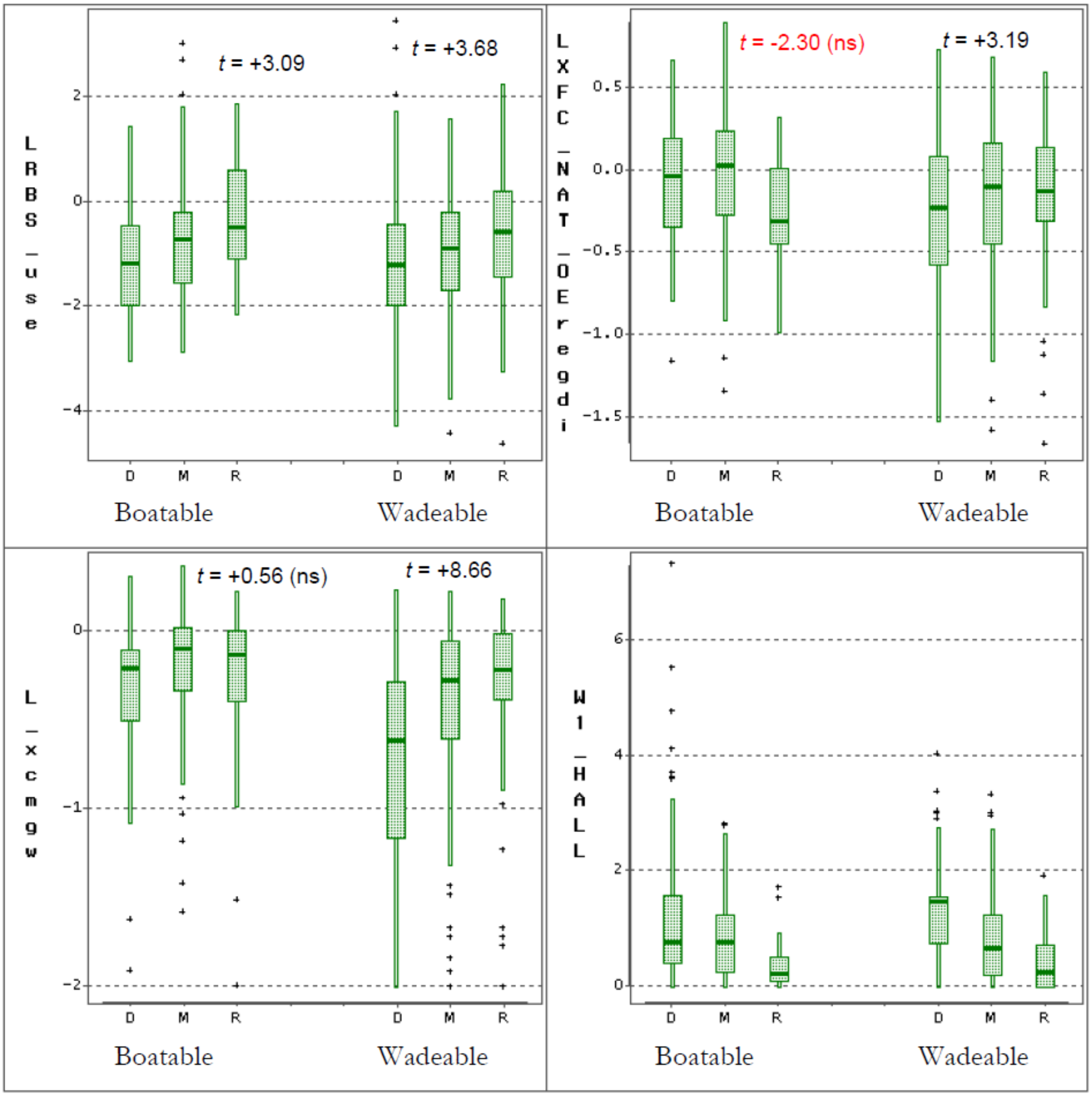
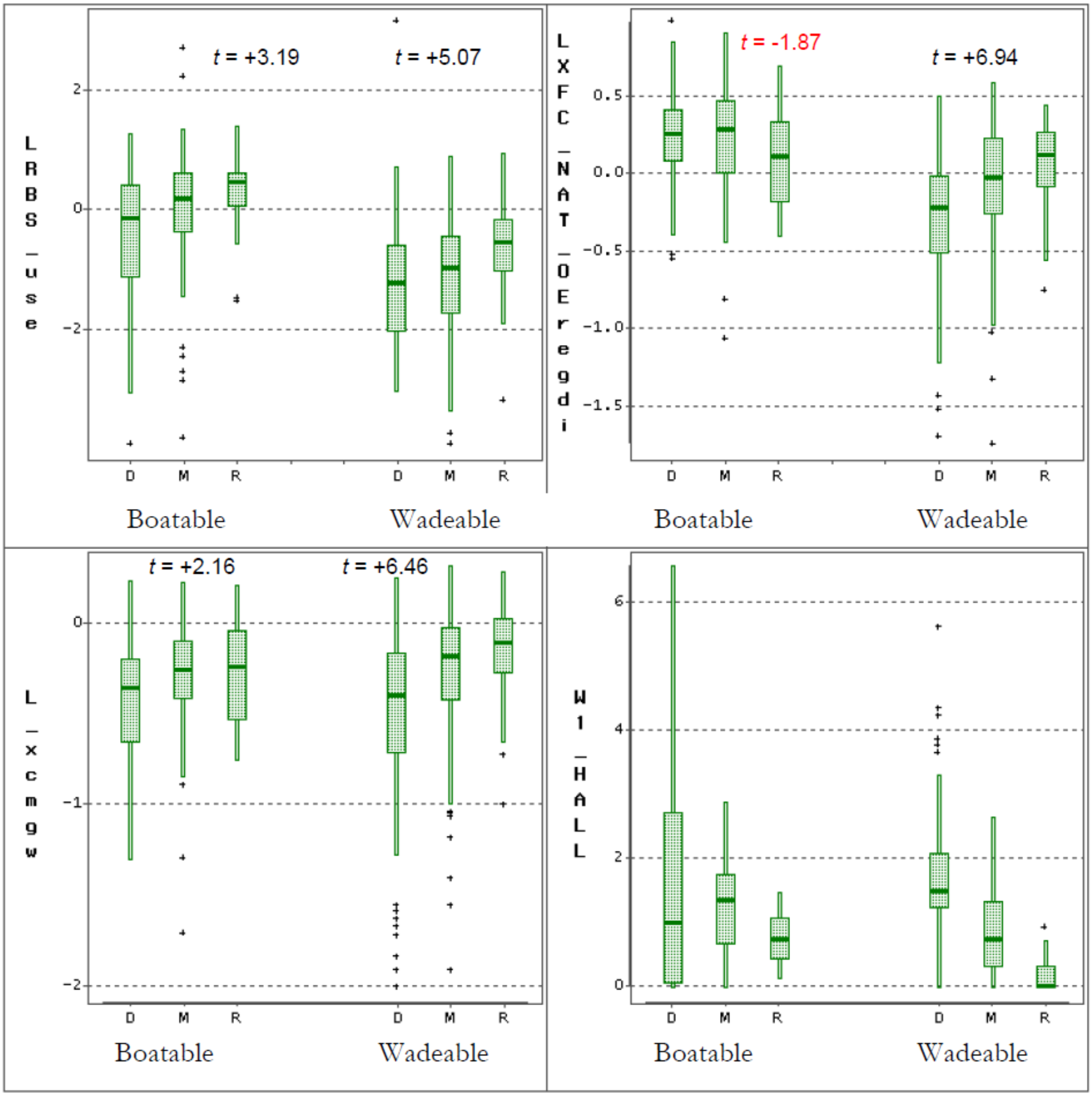


Figure 11.5 Indicator Responses for Reference and Disturbed Sites: West (Mountains and Xeric)



Figures 11.6 through 11.14 on following pages compare **Indicator Responses for Reference and Disturbed Sites in the nine aggregated Omernik Level III ecoregions (Eco9) that are shown in Figure 10.2.** Box and whisker plots in each panel show medians and 5th, 25th, 75th, and 95th percentiles for least-disturbed reference (R), moderately-disturbed (M), and highly-disturbed (D) sample sites. Plots show t values for the differences between R and D means for three indicators. Values shown in red have a sign contrary to expectations. The plots are shown separately for boatable and wadeable rivers and streams. The indicators are:

- LRBS_use = LRBS_g08, the indicator of relative bed stability and excess streambed fine sediments.
- LXCF_Nat_OEregdis = Log10 of observed/expected XFC_NAT, the indicator of instream habitat cover complexity.
- L_xcmgw = Log10(0.01 + XCMGW), the indicator of riparian vegetation cover and structure.
- W1_HALL = the proximity weighted indicator of riparian and near-shore human disturbance intensity (no t-test shown because W1_HALL was used in defining R, M, and D sites.

Figure 11.6 Indicator Response for Ecoregion (Eco9): Coastal Plain

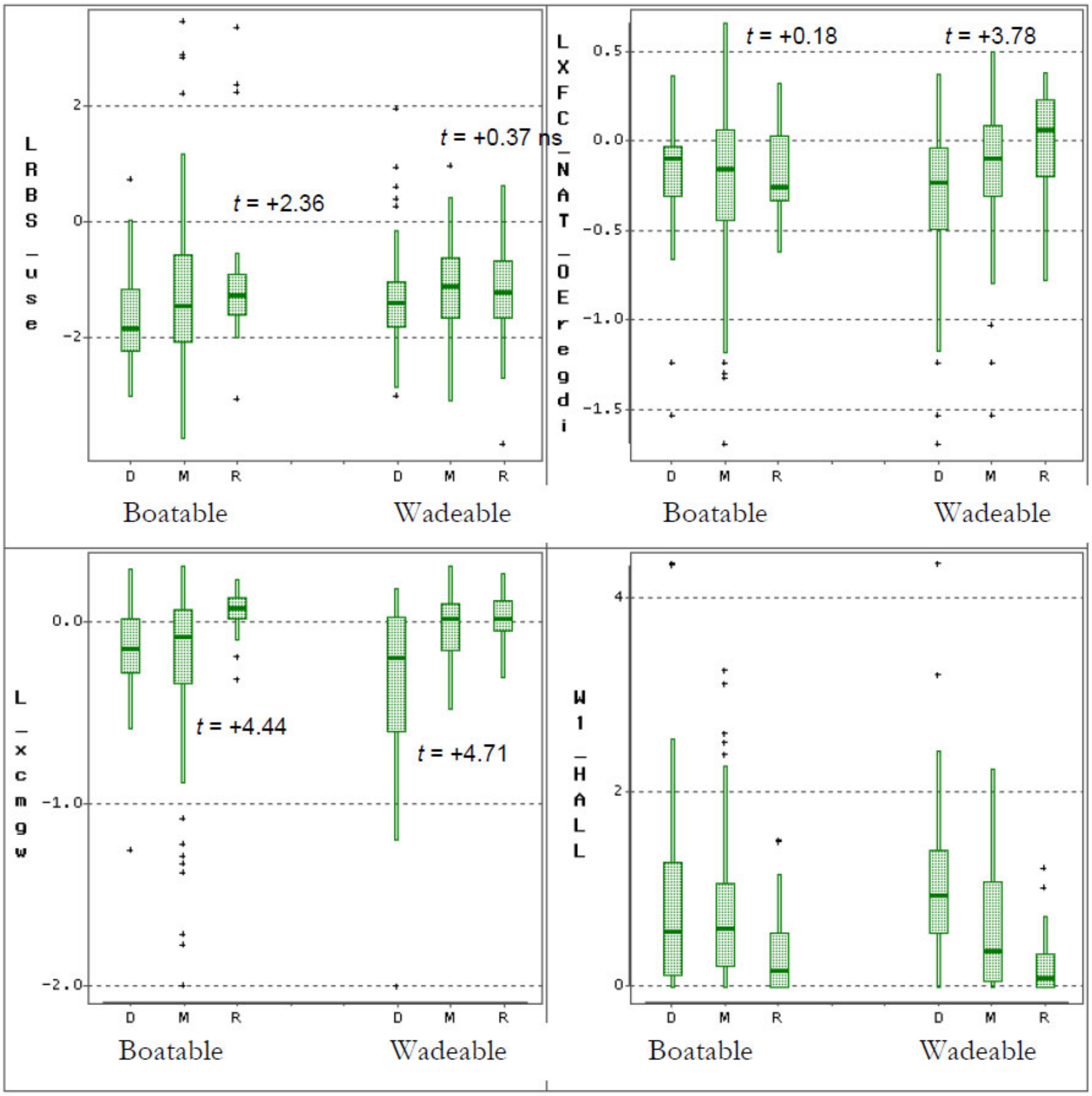


Figure 11.7 Indicator Response for Ecoregion (Eco9): Northern Appalachians

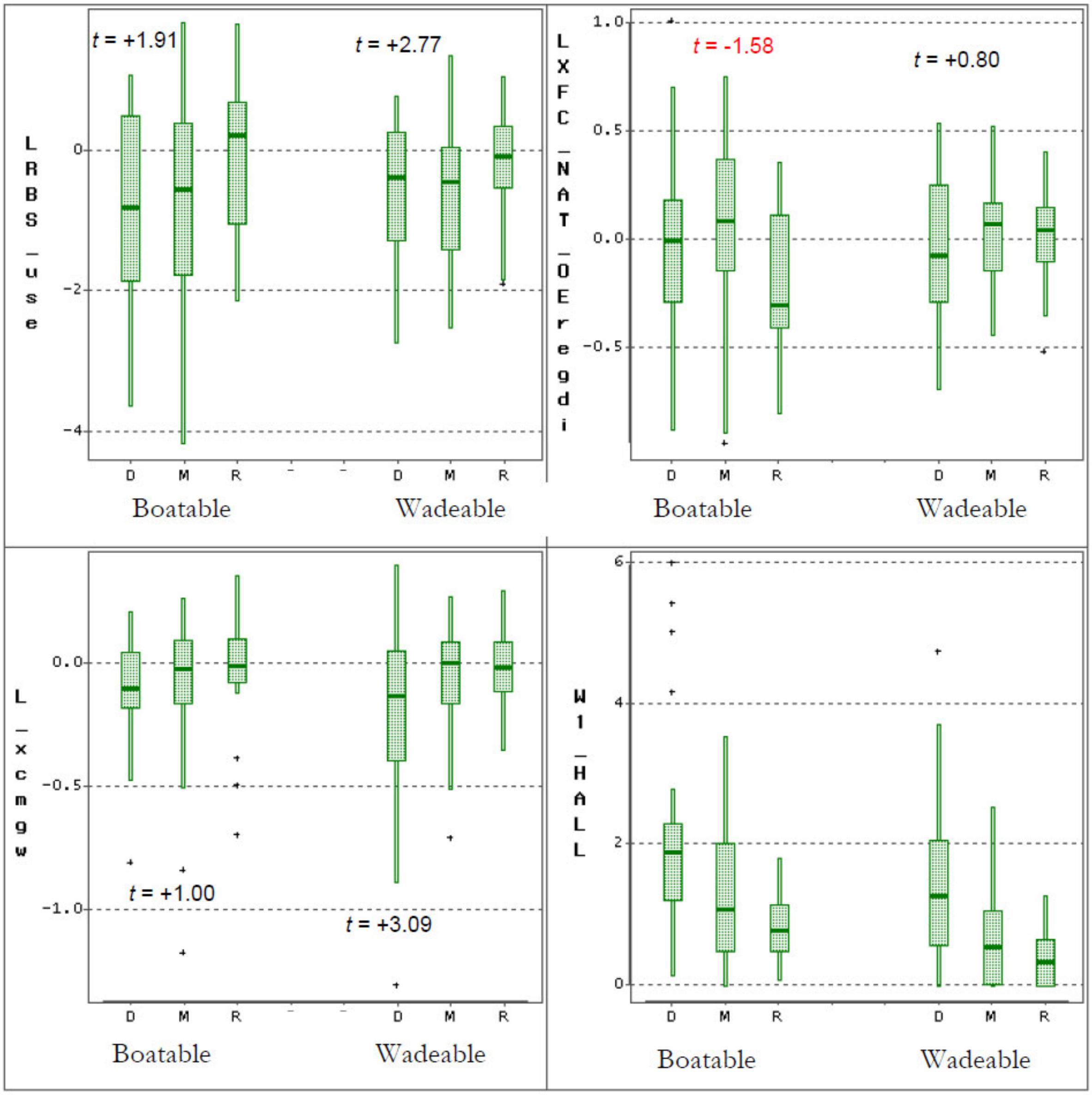


Figure 11.8 Indicator Response for Ecoregion (Eco9): Northern Plains

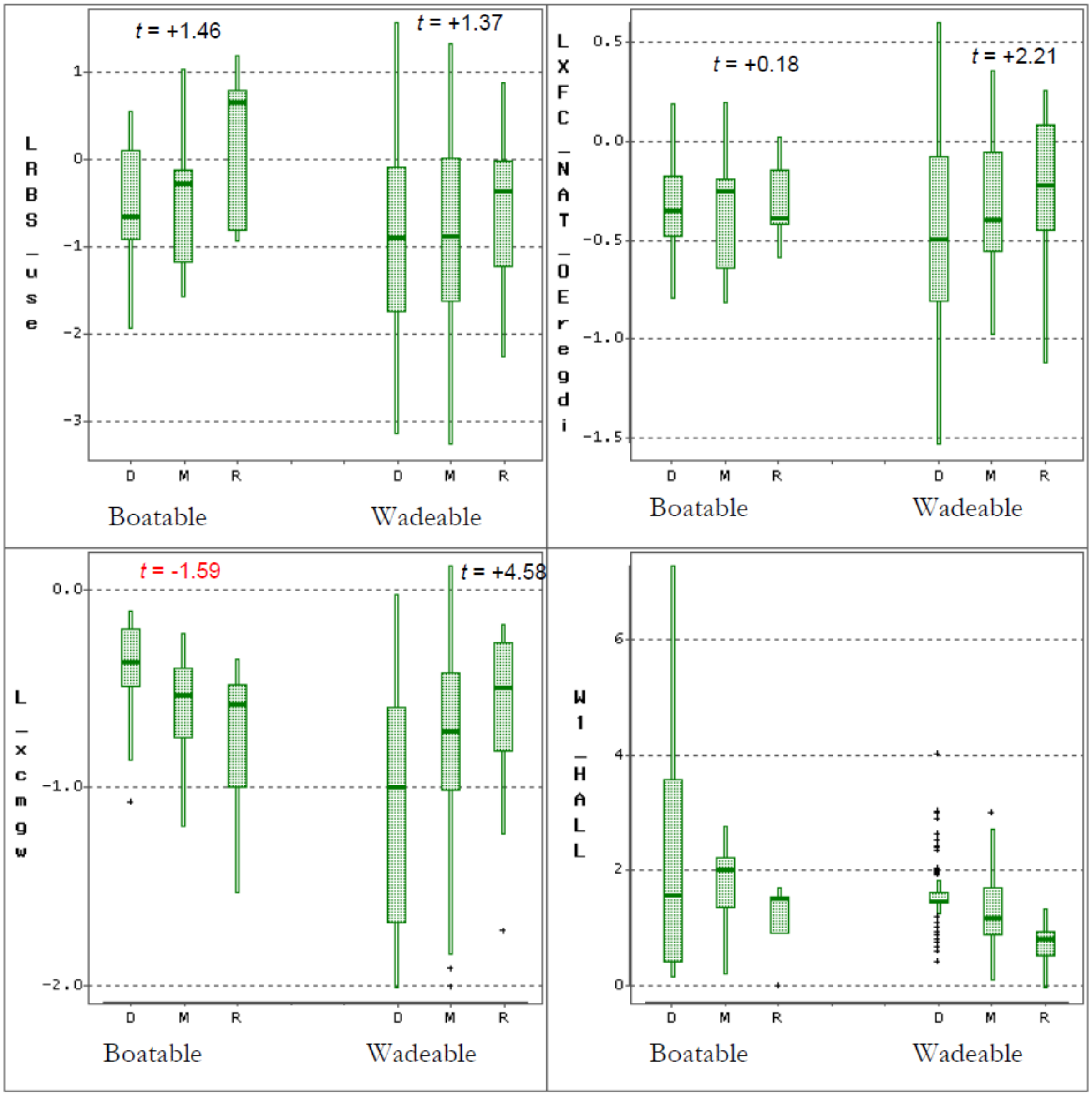


Figure 11.9 Indicator Response for Ecoregion (Eco9): Southern Appalachians

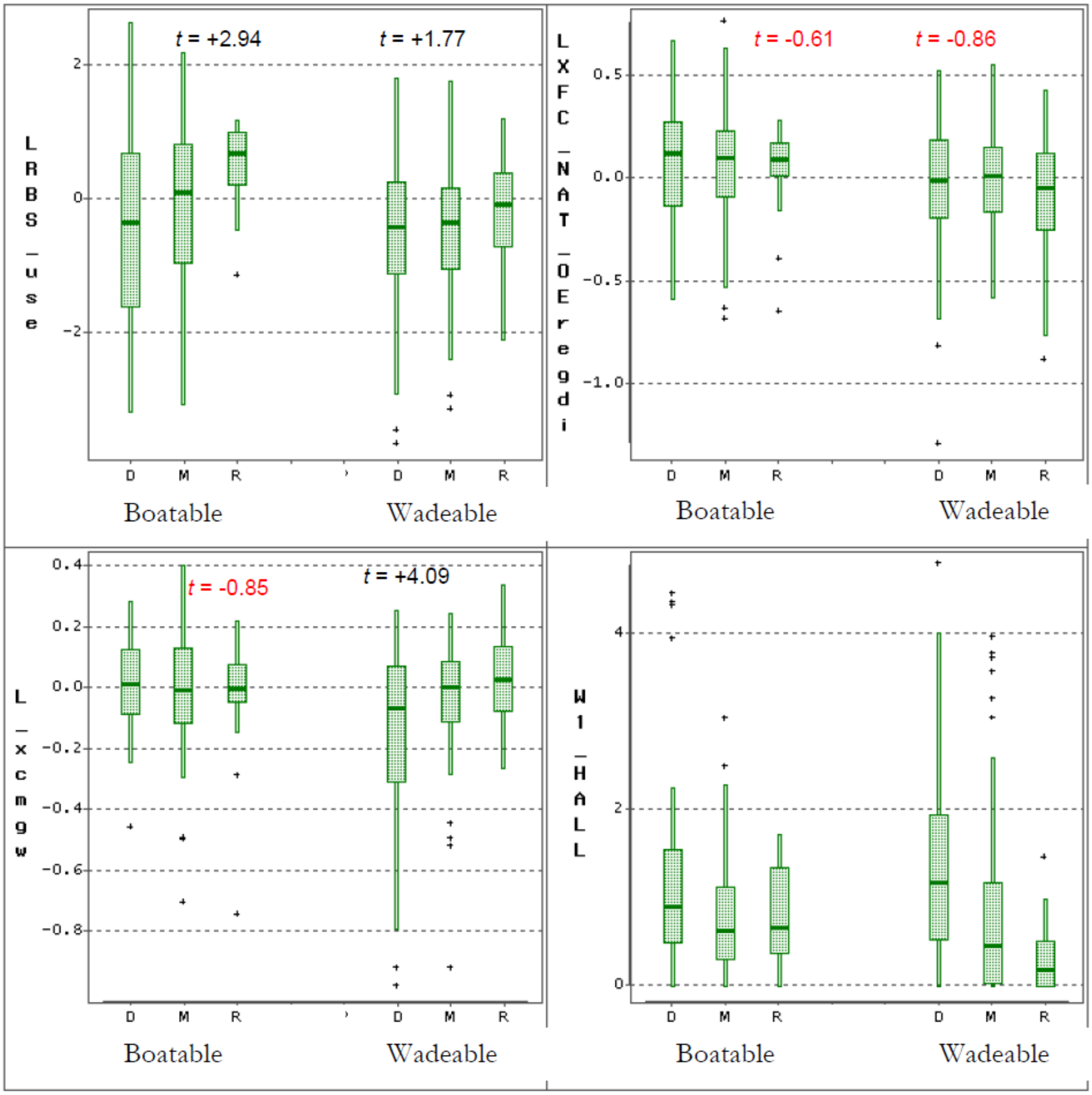


Figure 11.10 Indicator Response for Ecoregion (Eco9): Southern Plains

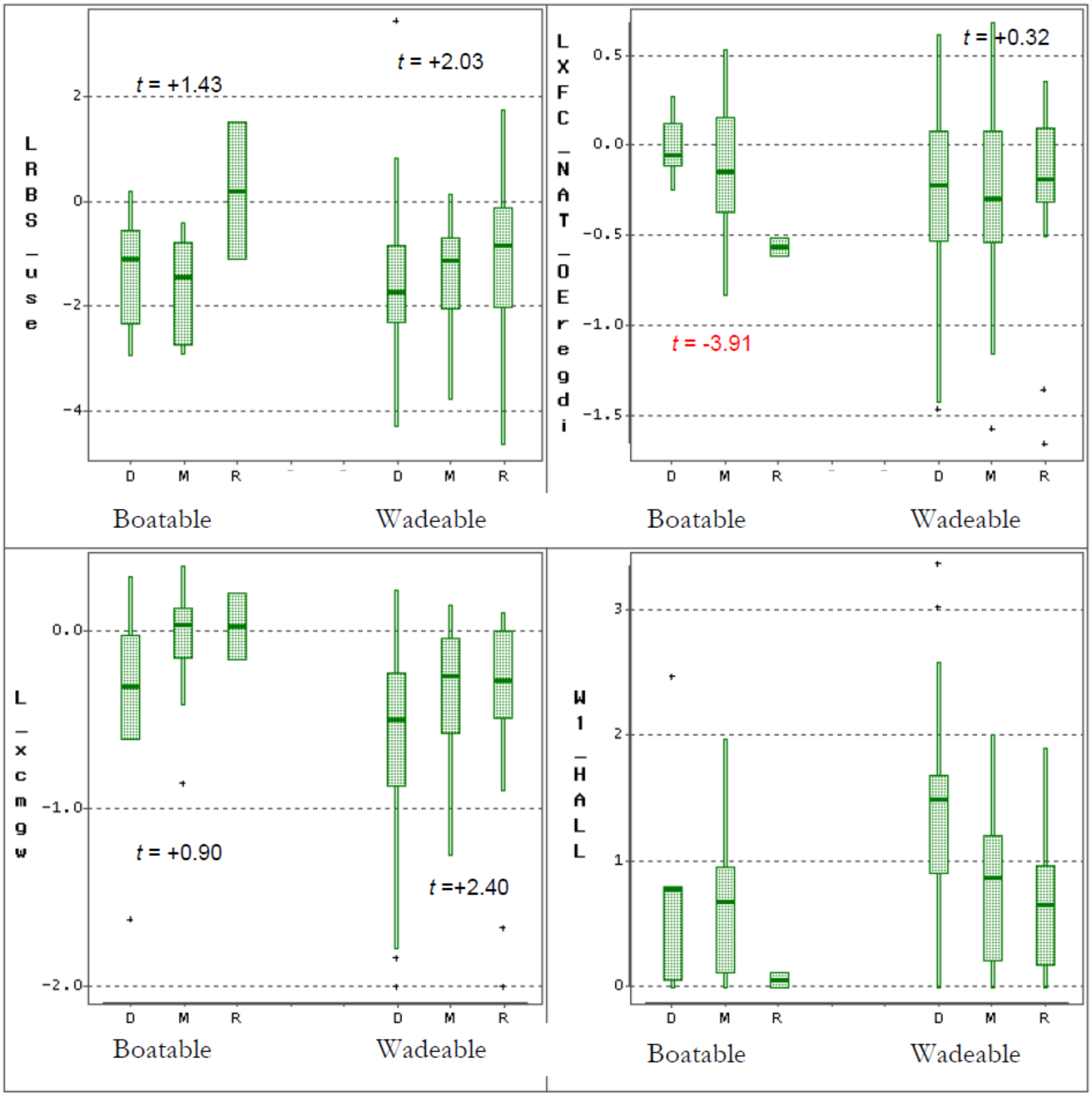


Figure 11.11 Indicator Response for Ecoregion (Eco9): Temperate Plains

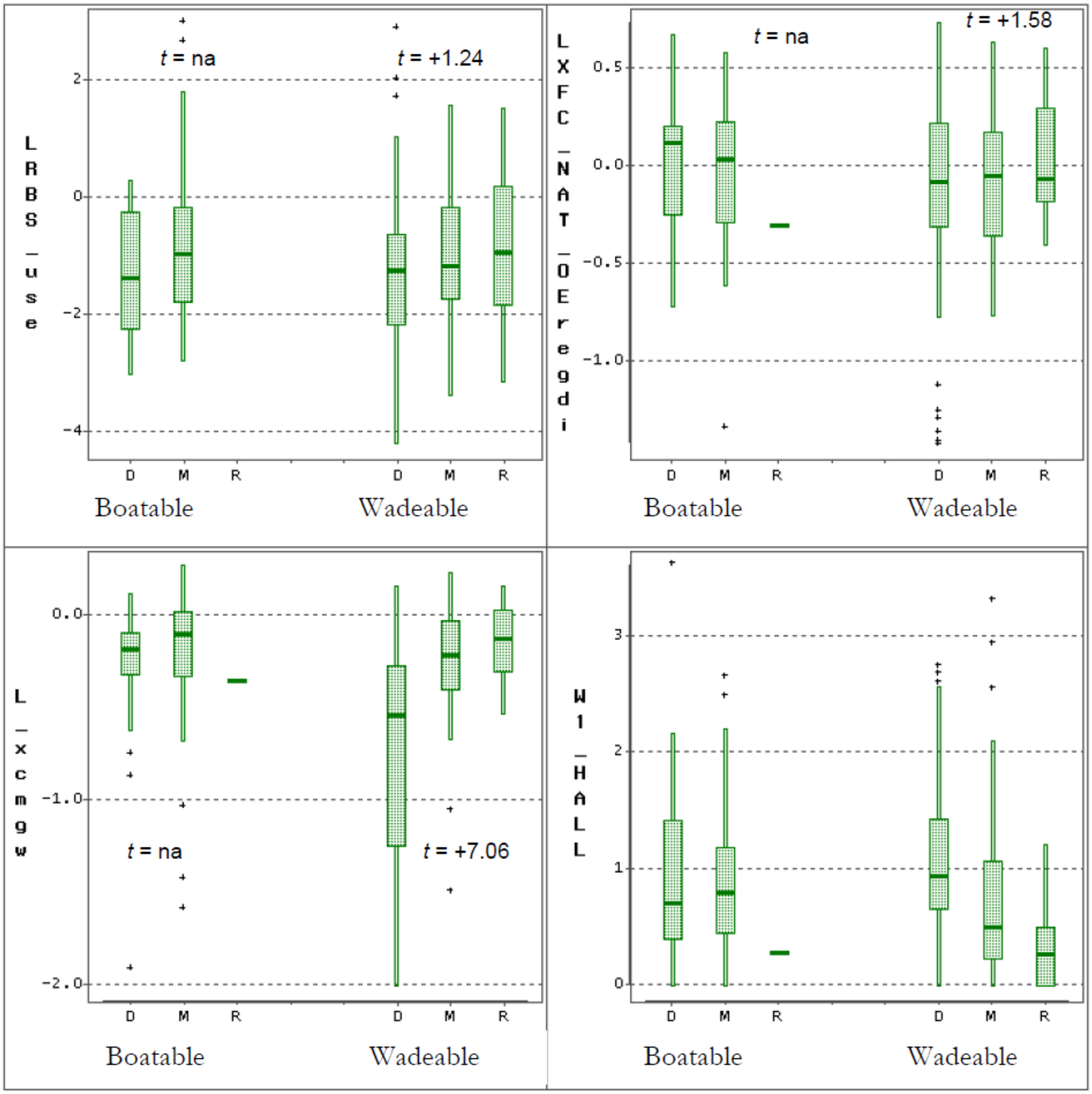


Figure 11.12 Indicator Response for Ecoregion (Eco9): Upper Midwest

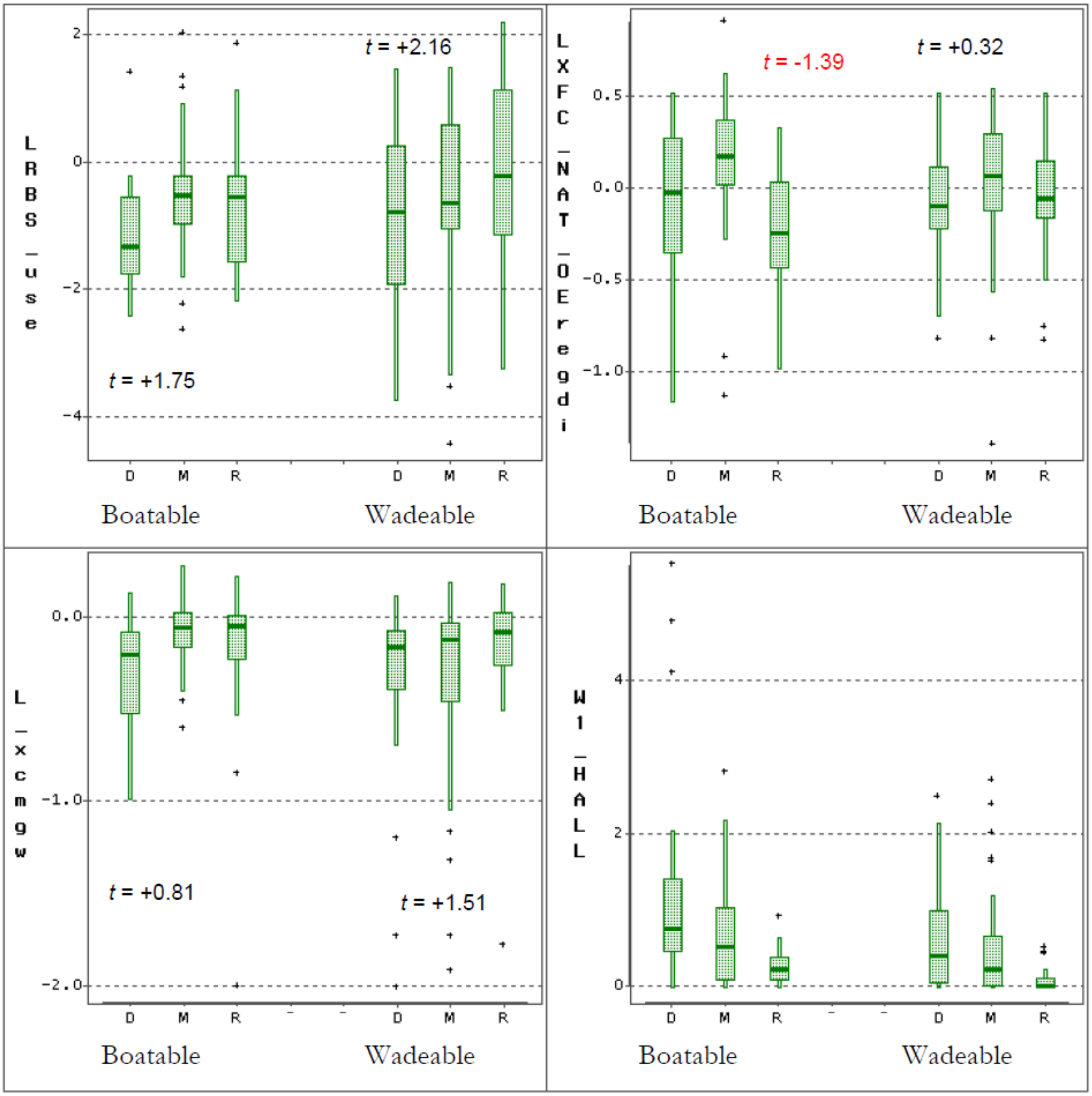


Figure 11.13 Indicator Response for Ecoregion (Eco9): Western Mountains

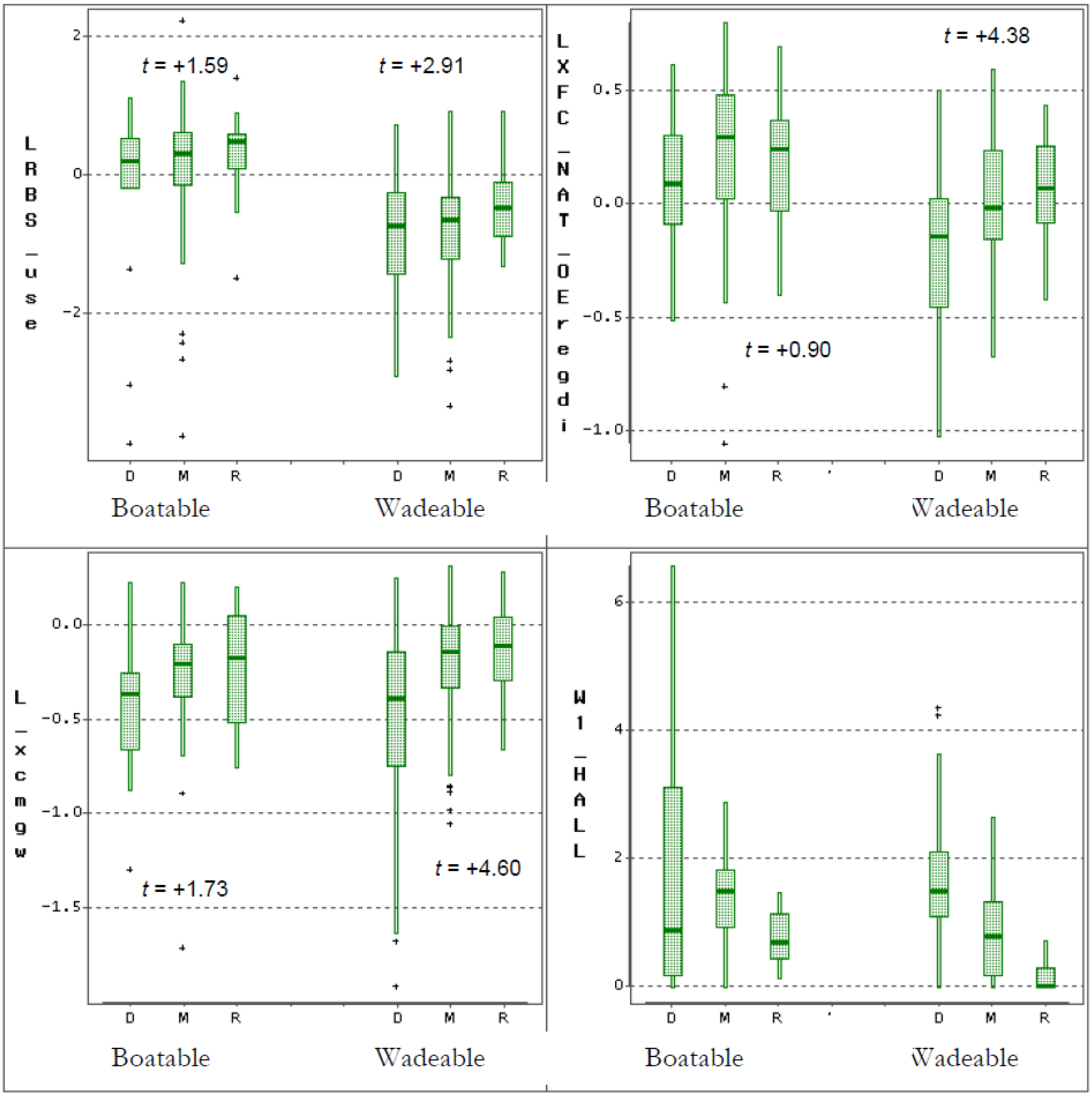


Figure 11.14 Indicator Response for Ecoregion (Eco9): Xeric West