

EPA/600/P-04/116F  
March 2006

# **Estimation and Application of Macroinvertebrate Tolerance Values**

National Center for Environmental Assessment  
Office of Research and Development  
U.S. Environmental Protection Agency  
Washington, DC 20460

## **DISCLAIMER**

This report has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

## **ABSTRACT**

Tolerance values provide a measure of the sensitivity of aquatic organisms to anthropogenic disturbance and have historically provided a useful tool for assessing the biological condition of streams and rivers. However, the tolerance values that are currently available are limited by the geographical areas in which they can be applied, and do not provide diagnostic information as to the causes of the impairment in streams. This report reviews and compares methods for estimating tolerance values from field data and for applying them to assess the biological condition of streams and to diagnose the causes of impairment. The intent of this report is to provide a resource of state, tribal, and regional biologists who wish to use tolerance values to help interpret biological data.

Methods for estimating tolerance values first model (either explicitly or implicitly) the relationship between a given taxon and an anthropogenic stressor gradient (the taxon-environment relationship). Then, a single representative value is extracted from this relationship, which is designated as the tolerance value. Three types of tolerance values are expressed on a continuous scale: (1) the weighted average, (2) cumulative percentiles, and (3) the maximum point of the taxon-environment relationship. A fourth type is categorical and classifies the shape of the taxon-environment relationship. All types of tolerance values provide fairly comparable rankings of tolerance among different taxa.

After tolerance values are estimated, metrics can be computed that summarize the tolerance information of all the taxa observed at a test site. Two types of metrics can be computed: compositional metrics and the mean of observed tolerance values. Compositional metrics apply categorical tolerance values and summarize characteristics of different groups of taxa (e.g., the relative abundance of tolerant taxa). The mean of observed tolerance values can be computed for any continuously valued tolerance value. Virtually all metrics are found to be strongly associated with observed stressor levels. The means of weighted average tolerance values are consistently and strongly associated with observed stressor levels in the data analyzed for this report. Ultimately, metrics values at test sites must be compared with reference distributions to determine whether conditions at test sites differ significantly from expectations.

The main areas of uncertainty with the tolerance value methodology are discussed. Some of these uncertainties and additional implementation issues must be addressed to more fully facilitate the use of tolerance values for the management of the nation's waters.

# CONTENTS

LIST OF TABLES .....	v
LIST OF FIGURES .....	vi
PREFACE .....	viii
AUTHOR AND REVIEWERS .....	ix
1. INTRODUCTION .....	1
2. THEORETICAL UNDERPINNINGS .....	2
2.1. UNIMODAL NICHE MODELS .....	2
2.2. NON-UNIMODAL MODELS .....	5
2.3. DEFINITIONS .....	6
2.4. TYPES OF TOLERANCE VALUES .....	7
3. ESTIMATING TOLERANCE VALUES FROM FIELD DATA .....	8
3.1. DIRECT METHODS .....	8
3.1.1. Central Tendencies (Weighted Averages) .....	8
3.1.2. Environmental Limits (Cumulative Percentiles) .....	9
3.2. INDIRECT METHODS .....	11
3.2.1. Regression Estimates of Taxon-Environment Relationships .....	11
3.2.1.1. Parametric Regressions .....	11
3.2.1.2. Nonparametric Regressions .....	14
3.2.1.3. Model Performance and Overfitting .....	16
3.2.2. Optima .....	18
3.2.3. Curve Classification .....	19
3.3. COMPARING DIFFERENT TOLERANCE VALUES .....	19
3.3.1. Continuous Tolerance Values .....	19
3.3.2. Tolerance Classifications .....	22
3.3.3. Comparisons Between Continuous and Categorical Tolerance Values .....	23
3.3.4. Effects of Regional Characteristics .....	25
4. APPLYING TOLERANCE VALUES IN ASSESSMENT .....	27
4.1. BIOLOGICAL METRICS .....	27
4.1.1. Metrics Based on Categorical Tolerance Values .....	27
4.1.2. Metrics Based on Continuous Tolerance Values .....	31
4.1.3. Performance of Tolerance Value Metrics .....	32
4.1.3.1. Characteristics of the Stressor Gradient .....	32
4.1.3.2. Acceptance Criteria for Tolerance Values .....	33
4.1.3.3. Specificity of Response .....	34
4.1.3.4. Effects of Taxonomic Resolution of Tolerance Values .....	37
4.2. REFERENCE CONDITIONS .....	38

## CONTENTS (continued)

5. AREAS OF UNCERTAINTY AND RESEARCH PRIORITIES .....	39
5.1. CAUSAL RELATIONSHIPS .....	41
5.2. DEFINING STRESSOR GRADIENTS .....	44
5.2.1. General-disturbance Gradients .....	44
5.2.2. Indirect Stressor Gradients .....	45
5.2.3. Sampling Issues .....	46
5.3. FUNDAMENTAL ECOLOGY .....	47
5.3.1. Biological Interactions .....	47
5.3.2. Temporal Characteristics of Stream Communities .....	48
6. IMPLEMENTATION .....	48
6.1. INFORMATION MANAGEMENT .....	48
6.2. TAXONOMIC QUALITY ASSURANCE .....	49
7. CONCLUSIONS AND RECOMMENDATIONS .....	49
REFERENCES .....	51
APPENDIX A: ATTENDEES AT THE WESTERN TOLERANCE VALUES WORKSHOP, CORVALLIS, OREGON .....	54
APPENDIX B: DATA DESCRIPTION .....	56
APPENDIX C: TOLERANCE VALUES .....	58
APPENDIX D: EXAMPLE STATISTICAL SCRIPTS .....	68
D.1. R: BASIC SYNTAX .....	68
D.2. LOADING DATA .....	69
D.3. ESTIMATING TOLERANCE VALUES FROM FIELD DATA .....	70
D.3.1. Weighted Averages .....	70
D.3.2. Cumulative Percentiles .....	71
D.3.3. Parametric Regressions .....	72
D.3.4. Nonparametric Regressions .....	73
D.3.5. Assessing Model Fit .....	74
D.3.6. Optima .....	75
D.3.7. Curve Shape .....	76
D.4. APPLYING TOLERANCE VALUES IN ASSESSMENT .....	77
D.4.1. Richness .....	78
D.4.2. Proportion Total Taxa .....	78
D.4.3. Relative Abundance .....	79
D.4.4. Mean Tolerance Value .....	79

## LIST OF TABLES

Table 1. Observed versus predicted occurrences for <i>Heterlimnius</i> .....	17
Table 2. Correlation coefficients for temperature tolerance values .....	20
Table 3. Correlation coefficients for sediment tolerance values .....	21
Table 4. Confusion matrix for temperature tolerance classifications .....	22
Table 5. Confusion matrix for sediment tolerance classifications .....	22
Table 6. R <sup>2</sup> values for spline fits between metric and stressor level .....	30
Table 7. R <sup>2</sup> values for spline fits between mean tolerance values and observed temperature and sediment in Oregon .....	32
Table 8. R <sup>2</sup> values for models with ROC > 0.65 criteria imposed .....	34
Table 9. R <sup>2</sup> values for family-level tolerance values .....	38

## LIST OF FIGURES

Figure 1. Theoretical unimodal distribution of species occurrence	3
Figure 2. Different unimodal relationships within a species pool	4
Figure 3. Monotonic responses to an environmental gradient	5
Figure 4. Examples of ecological tolerance and biological assessment tolerance	6
Figure 5. Empirical cumulative distribution method for defining tolerance values	10
Figure 6. Relationships between probability of occurrence and temperature for <i>Heterlimnius</i> and <i>Malenka</i>	13
Figure 7. Relationship between relative abundance and temperature for <i>Heterlimnius</i> and <i>Malenka</i>	14
Figure 8. Relationship between probability of occurrence and temperature for <i>Heterlimnius</i> and <i>Malenka</i>	16
Figure 9. Receiver operator characteristic (ROC) curve for <i>Heterlimnius</i>	18
Figure 10. Graphical approach for classifying curve shapes for <i>Heterlimnius</i> , <i>Malenka</i> , and <i>Hydropsyche</i>	20
Figure 11. Comparison of optima tolerance values determined by parametric logistic regression model (GLMMAX) and weighted average tolerance values (WA) for temperature	21
Figure 12. Histogram of weighted average tolerance values for temperature and fine sediment	24
Figure 13. Histogram of weighted average tolerance values for temperature and fine sediment classified by curve shape	25
Figure 14. Illustrations of effect of gradient length	26
Figure 15. Relationship between temperature tolerance metrics and observed temperature in Oregon	29
Figure 16. Relationship between sediment tolerance metric and observed sediment in Oregon	30

**LIST OF FIGURES (continued)**

Figure 17. Temperature tolerance metrics plotted versus percent sands and fines . . . . . 35

Figure 18. Sediment tolerance metrics plotted versus temperature . . . . . 36

Figure 19. Mean weighted average (WA) tolerance values for sediment plotted against  
observed temperature, and mean WA tolerance value for temperature plotted  
against observed sediment . . . . . 37

Figure 20. Comparison of temperature and sediment tolerance metrics at a single test site  
with reference distributions . . . . . 40

Figure 21. Comparison of taxon-environment relationships for *Glossosoma* and *Malenka* . . . 42

## PREFACE

This document, *Estimation and Application of Macroinvertebrate Tolerance Values*, was prepared by the National Center for Environmental Assessment. The document is a technical review of statistical methods for estimating macroinvertebrate tolerance values from field data. It also reviews different methods for applying tolerance values for assessing biological conditions of streams and rivers and for diagnosing the causes of impairment. The purpose of this document is to provide a technical resource for states, tribes, and regions that wish to use tolerance values to interpret stream biological data.

This final document reflects a consideration of all comments received on an External Review Draft dated September 2004 (EPA/600/P-04/116A).

## **AUTHOR AND REVIEWERS**

This document was prepared by the National Center for Environmental Assessment.

### **AUTHOR**

Lester L. Yuan  
National Center for Environmental Assessment  
U.S. Environmental Protection Agency  
Washington, DC 20460

### **REVIEWERS**

Susan B. Norton  
National Center for Environmental  
Assessment  
U.S. Environmental Protection Agency  
Washington, DC 20460

Amina I. Pollard  
National Center for Environmental  
Assessment  
U.S. Environmental Protection Agency  
Washington, DC 20460

John Van Sickle  
Western Ecology Division  
National Health and Environmental Effects  
Research Laboratory  
U.S. Environmental Protection Agency  
200 S.W. 35th Street  
Corvallis, OR 97333-4902

Charles A. Menzie  
Menzie-Cura & Associates Inc.  
2 West Lane  
Severna Park, MD 21146

Mark T. Southerland  
Versar ESM Operations  
9200 Rumsey Road  
Columbia, MD 21045-1934

Jon H. Volstad  
Versar ESM Operations  
9200 Rumsey Road  
Columbia, MD 21045-1934

N. Scott Urquhart  
Department of Statistics  
Colorado State University  
Fort Collins, CO 80523-1877

## 1. INTRODUCTION

Benthic macroinvertebrates are widely collected and used to assess the condition of streams and rivers. One approach for interpreting biological assessment data is to group taxa according to their perceived tolerance or sensitivity to anthropogenic disturbances. After these groups, or tolerance values, are established, the condition of new sites can be assessed on the basis of whether taxa from tolerant or sensitive groups are predominantly collected. Tolerance values have historically played an integral role in biological assessment. One set of tolerance values (the saprobien system) was derived in the early 1900s in central Europe by documenting differences in microorganism assemblages collected at progressively larger distances downstream from sewage outfalls. Other early examples of assigning tolerance values to macroinvertebrates include those of Chutter (1972), who assigned tolerance values to South African macroinvertebrates using best professional judgement, and the Biological Monitoring Working Party (BMWP) (Armitage et al., 1983), which assigned tolerance values to macroinvertebrate families observed in Great Britain. In North America, Hilsenhoff (1987) developed tolerance values with respect to a gradient of organic pollution, and these values are still widely used today. Lenat (1993) refined Hilsenhoff's approach, defining tolerance values with respect to a more comprehensive list of anthropogenic disturbances.

Tolerance values have been used successfully to assess the condition of running waters, but in recent years two issues have arisen regarding the widespread application of these values. First, the application of tolerance values in regions that differ from where they were originally derived has been questioned. For example, most tolerance values (e.g., Hilsenhoff, 1987; Lenat, 1993) have been estimated using data collected in the midwestern and eastern United States, and these tolerance values have then been modified for use elsewhere in the country. However, regional species pools differ substantially in different areas of the United States, and as a result, many of the taxa collected outside of the midwestern and eastern U.S. have not been assigned tolerance values. Furthermore, the stressor gradients commonly observed in other regions can differ from the organic pollution gradients for which the Hilsenhoff tolerance values were originally designed. A more comprehensive examination of the estimation of tolerance values and their general applicability to other regions is required.

The second issue is a growing interest in extending the use of tolerance values from simple assessments of stream condition to diagnoses of the causes of impairment. Over a thousand streams are currently listed on the U.S. Environmental Protection Agency's (EPA's) 303(d) list as being biologically impaired, but the causes of the impairment are unknown, which severely hinders effective management actions. If organisms that are sensitive or tolerant to particular anthropogenic stressors can be identified, the presence or absence of different

organisms in impaired streams can potentially help identify possible sources of impairment. Although most current organism tolerance values have been defined with respect to a single gradient of anthropogenic disturbance (e.g., Hilsenhoff, 1987), the potential for estimating tolerance values with respect to particular stressors has been described in a few studies: Slooff (1983) showed that aquatic macroinvertebrates differ in their sensitivity to different stressors, and in more recent work, Chessman and McEvoy (1998) attempted to estimate tolerance values that discriminated between different types of anthropogenic disturbance. The tolerance values developed by Chessman and McEvoy had limited discriminatory powers, but they did illustrate the potential power of the method.

To address these issues, EPA and the Council of State Governments convened a workshop in February 2004 in Corvallis, OR. Workshop attendees were selected for their experience in deriving and applying tolerance values and included biologists from environmental protection agencies in western states, EPA staff, academics, and industry representatives (Appendix A). The program consisted of a day of presentations of current methods for deriving tolerance values, followed by a day of discussion within smaller breakout groups and half a day of discussion that included the full group.

This report continues the effort initiated by the workshop by synthesizing and expanding on the deliberations of the attendees. In particular, this report (1) reviews the ecological theory that underlies the concept of tolerance values, (2) reviews methods for computing tolerance values from field data, (3) reviews methods for applying tolerance values in stream assessments, (4) identifies uncertainties and research priorities in the derivation and application of tolerance values, and (5) identifies practical issues hindering the broader use of tolerance values in monitoring and assessment programs. The methods discussed in this report can be used to estimate both general disturbance tolerance values (for assessing biological condition) and stressor-specific tolerance values (for diagnosing the causes of impairment). However, the examples provided in the report focus more specifically on the issues pertaining to diagnosis of cause. The intent of this report is to provide a resource for state, tribal, and regional biologists who wish to use tolerance values to help interpret biological data for evaluating biological condition and for diagnosing the causes of stream impairment.

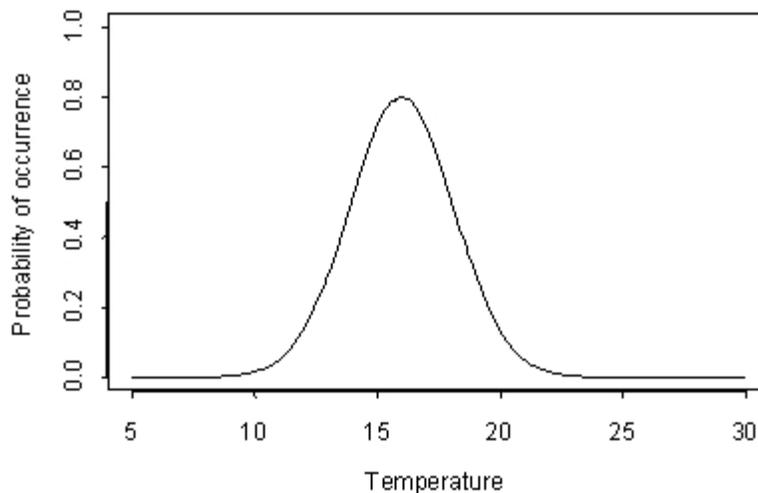
## **2. THEORETICAL UNDERPINNINGS**

### **2.1. UNIMODAL NICHE MODELS**

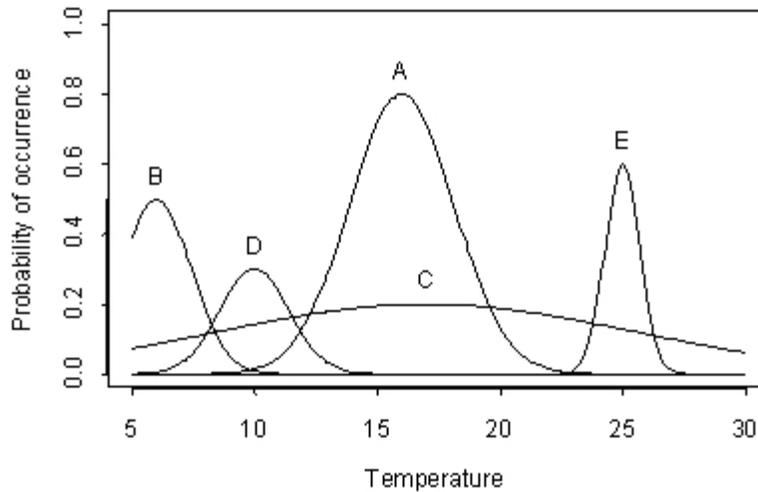
Ecological niche theory suggests that individual species are often unimodally distributed along environmental gradients. This distribution dictates that the probability of a species'

occurrence or abundance has a maximum value at some point along an environmental gradient (e.g., with regard to temperature, Figure 1). This point is known as the species optimum. Probabilities of finding a species and the expected abundance of a species decrease as conditions depart from the species optimum. Unimodal distributions are thought to arise over evolutionary time scales as species specialize to optimally exploit certain habitat conditions or resources through the process of adaptive radiation (niche diversification) (Odum, 1971). Each species is simultaneously affected by a variety of factors, but for simplicity, discussion is restricted here to a single environmental gradient.

Species differ in their optima and in the range of conditions about their optima in which they can persist, so unimodal response functions vary in their location along a particular gradient and in the width of the curves. Furthermore, the frequency of occurrence and abundance of different species can vary greatly, so the maximum values of the response functions will differ among species (Figure 2). Thus, for a given species pool, a diverse set of species–environment relationships are expected along a particular gradient, and if we sample streams at different points along that gradient, we would expect to find a different assemblage of species. Based on the responses shown in Figure 2, we would expect to find species B, C, and D in a 7°C stream, whereas we would expect to find species C and E in a 25°C stream. This shift in



**Figure 1. Theoretical unimodal distribution of species occurrence.**



**Figure 2. Different unimodal relationships within a species pool.** Letters indicated different species.

assemblage composition across environmental gradients is the basic theory that underlies the notion of tolerance values. That is, if we can construct species–environment relationships for environmental gradients that are influenced by human activities, we can potentially predict assemblage changes that are likely to occur as human activity alters conditions at a site. We can also identify those species that are likely to disappear with increased human stress (i.e., sensitive species) and those species that should thrive in anthropogenically stressed sites (i.e., tolerant species).

Many environmental factors vary not only in response to human activities, but also exhibit considerable variability due to natural factors. For example, stream temperature changes naturally with the elevation of the stream, but it also can be changed by the removal of riparian vegetation that shades the stream. Species–environment relationships can be estimated regardless of whether the changes in the environment are due to natural or anthropogenic causes. However, when strong natural and anthropogenic factors influence a particular environmental variable, additional care must be exercised when interpreting the observed changes in species assemblage (see Section 4.2).

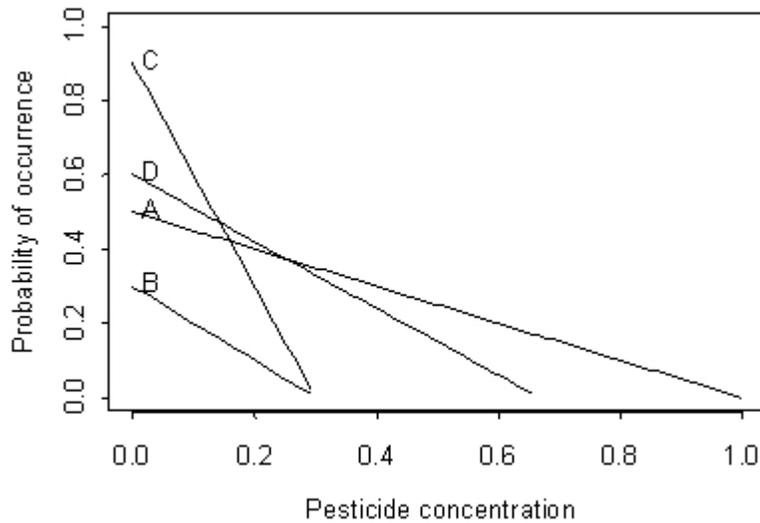
The range of conditions in which a species is observed in the field is an estimate of its *realized niche*. The realized niche is a function of the effects of both environmental gradients and biological interactions. In contrast, the *fundamental niche* defines the range of conditions in which an organism could persist, excluding the effects of any biological interactions. In theory, the fundamental niche is an inherent characteristic of a species and is therefore invariant across all geographical regions. However, species pools change across different regions, so the realized

niche of a particular species may shift as different biological interactions affect its ability to persist. Because many stream ecosystems are structured by disturbance rather than by biological interactions, the realized niches for many species may provide reasonable approximations to the fundamental niches (Allan, 1995). Differences between fundamental and realized niches are discussed further in Section 5.3.1.

## 2.2. NON-UNIMODAL MODELS

The unimodal model may not be applicable to all types of environmental gradients. For certain anthropogenic stressors (e.g., pesticides), a more appropriate model may be one in which the abundance or the probability of observing a particular species decreases monotonically with increasing levels of the stressor. As levels of such stressors increase, the rate at which expected abundances or capture probabilities change may differ among species, but the optimum for all species would be identical, where the stressor level is zero (Figure 3). Therefore, these differences are more subtle than those that would be observed between unimodally distributed species at opposite ends of a gradient (e.g., species B and E in Figure 2).

Non-unimodal relationships also appear when environmental gradients are incompletely sampled. For example, if observations were collected only up to 16°C, then the probability of occurrence for species A in Figure 1 would be observed to be monotonically increasing with respect to temperature.

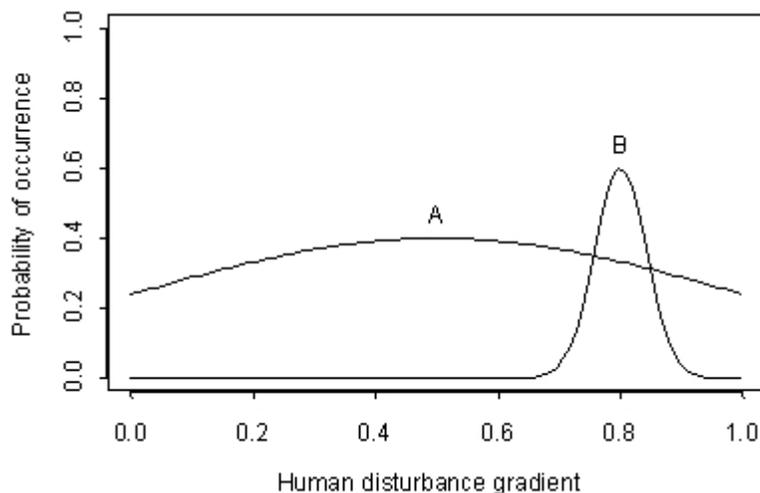


**Figure 3. Monotonic responses to an environmental gradient.** Letters indicate different species.

### 2.3. DEFINITIONS

The use of the term “tolerance” in biological assessments differs from the ecological use of the word. The ecological definition of tolerance refers to the niche breadth, or the range of conditions that an organism can withstand (Odum, 1971). Based on this definition, more tolerant organisms can withstand a broader range of conditions, regardless of the location of their species optimum. To quantify niche breadth, we would measure the range of conditions over which a particular taxon can persist.

In biological assessment, the term tolerance is generally used with respect to a gradient of anthropogenic stress. That is, a tolerant organism is one that is likely to be found in a site that has been highly altered or degraded by human activities. Many species would be considered tolerant by both ecological and biological assessment definitions. For example, species A in Figure 4 would likely be observed in a degraded site, and is therefore tolerant in terms of biological assessment. It is also found in conditions that span the entire observed gradient, and is therefore ecologically tolerant. However, some species thrive in degraded sites but have a narrow niche breadth. These species are tolerant in the context of biological assessment even though they are not tolerant in the ecological sense (species B in Figure 4). For this report we use tolerance strictly in the sense of biological assessment and use niche breadth when referring to the range of conditions within which an organism can persist. Similarly, we define a *sensitive* taxon as one that tends to decline in abundance or occurrence probability as anthropogenic stress increases. We further define a *tolerance value* as a single value that represents the tolerance of a taxon to a particular stressor gradient. This value can have the units and range of the sampled gradient, or it can be rescaled to an arbitrary range (e.g., 1–10).



**Figure 4. Examples of ecological tolerance (A) and biological assessment tolerance (B).** Magnitude of human disturbance increases from left to right.

Both single and composite stressor gradients are applicable when considering tolerance values. A single-stressor gradient is defined by a single environmental attribute (e.g., temperature) that changes with the intensity of human activity. Composite-stressor gradients represent simultaneous changes in many different environmental attributes (e.g., temperature, toxicant concentrations, physical habitat quality) that occur with increased human activity. For example, an aggregate gradient of human disturbance would be considered a composite-stressor gradient. In general, the methods for estimating tolerance values described in this report can be applied to both composite-stressor and single-stressor gradients, although defining an appropriate gradient becomes more difficult with multiple stressors. These issues are explored in greater detail in Section 5.2.1.

## 2.4. TYPES OF TOLERANCE VALUES

A tolerance value can be thought of as any single number that represents the characteristics of a species' relationship with an environmental gradient. Ideally, tolerance values should capture the critical aspects of the entire species–environment relationship, provide a consistent ranking of different species in terms of their tolerance for a particular stressor, and provide the means for analyzing data on species occurrences to derive inferences regarding the environmental conditions at a site. Because species–environment relationships vary greatly in their functional forms, it is difficult to completely characterize the relationship with one value. However, single tolerance values have proven to be a useful tool for biological assessment, where more complete representations of the species–environment relationship are usually too cumbersome to use regularly.

In general, given species–environment relationships for different species, four approaches can be used to derive tolerance values: (1) central tendencies, (2) environmental limits, (3) optima, and (4) curve shapes. Tolerance values expressed in terms of *central tendencies* attempt to describe the average environmental conditions under which a species is likely to occur; tolerance values expressed in terms of *environmental limits* attempt to capture the maximum or the minimum level of an environmental variable under which a species can persist; and tolerance values expressed in terms of *optima* define the environmental conditions that are most preferred by a given species. These three types of tolerance values are expressed in terms of locations on a continuous numerical scale that represents the environmental gradient of interest.

The fourth type of tolerance value relies on a classification of *curve shape* to group species. Species are first classified by the shape of their species–environment relationship into three groups: (1) monotonically increasing, (2) monotonically decreasing, and (3) unimodal (Yuan, 2004). Then, when an increasing value of the environmental variable can be attributed to human activities, those species with monotonically increasing species–environment relationships

are designated as tolerant and those with monotonically decreasing relationships are designated as sensitive. Species with unimodal species–environment relationships are designated as intermediately tolerant. These classifications differ from previously discussed tolerance values because they yield categorical rather than continuous values.

### **3. ESTIMATING TOLERANCE VALUES FROM FIELD DATA**

Given the different types of tolerance values defined in the previous section, how does one estimate these values from field data? Analytical approaches for estimating tolerance values can be divided into two main groups: (1) methods that directly estimate tolerance values from field data (direct methods) and (2) methods that first require estimations of species–environment relationships using regression techniques, and then estimate tolerance values from the regression relationship (indirect methods). Central tendencies and environmental limits can be estimated using methods from the first group, whereas optima and curve shape require methods from the second group. In the following section we describe in more detail the methods for estimating each type of tolerance value.

Until now, discussion has focused on distinct species and their relationship with environmental gradients, because ecological theory hypothesizes that the fundamental niche of a given species is invariant. However, species level identifications are not always available, so as we turn to field data, the focus of the discussion is broadened to consider any level of taxonomic resolution (i.e., taxon–environment relationships). The ramifications of defining tolerance values for higher levels of taxonomy are discussed in Section 4.1.3.4.

Throughout this section methods are illustrated using macroinvertebrate data collected by the EPA’s Environmental Monitoring and Assessment Program–Western Pilot Project (EMAP–West) (see Appendix B). Two representative environmental gradients are considered in particular: stream temperature and the amount of bedded fine sediment. Both stream temperature and the amount of bedded fine sediment vary naturally (e.g., with changes in elevation) and can vary due to human activities (e.g., from logging). These two stressors have been identified as particularly relevant to western streams.

#### **3.1. DIRECT METHODS**

##### **3.1.1. Central Tendencies (Weighted Averages)**

Weighted averaging (WA) has long been used in ecology as a simple, robust approach for estimating the central tendencies of different taxa, or in our case, tolerance values (ter Braak and Looman, 1986). WA and all its variants (e.g., partial least-square WA, ter Braak and Juggins,

1993) operate on the same basic principle. The tolerance value,  $u_{wa}$ , for species  $j$  is estimated by computing the mean of the environmental variable of interest at the sites in which the species is observed:

$$u_{wa,j} = \frac{\sum_{i=1}^N Y_{ij} x_i}{\sum_{i=1}^N Y_{ij}} \quad (1)$$

where  $N$  is the total number of sites and  $x_i$  is the value of the environmental variable of interest at site  $I$ . For presence/absence data,  $Y_{ij}$  is equal to 1 when species  $j$  is present and 0 when species  $j$  is absent, and for abundance data,  $Y_{ij}$  is the abundance of species  $j$  at site  $I$ .

When using weighted averages, a uniform distribution of samples across the environmental gradient is preferred because each location on the gradient will then receive an equal weight. Because environmental gradients are rarely uniformly sampled, weighted average tolerance values are often biased away from the “true” central tendency of the taxon-environment relationship. One solution to this problem is to average samples along the gradient that fall within equal width bins and then to use these binned data to compute the weighted average. However, this procedure is not generally recommended because the true central tendency is rarely of interest. Instead, we are usually interested in the tolerance values of different taxa relative to one another. Within a given data set, all weighted average tolerance values are computed using the same set of environmental data, and therefore, any bias arising from a nonuniform distribution of data will be the same for all taxa and their relative placement along the axis will generally be preserved. Comparisons of weighted average tolerance values across different data sets are more problematic and are discussed in much greater detail in Section 3.3.4.

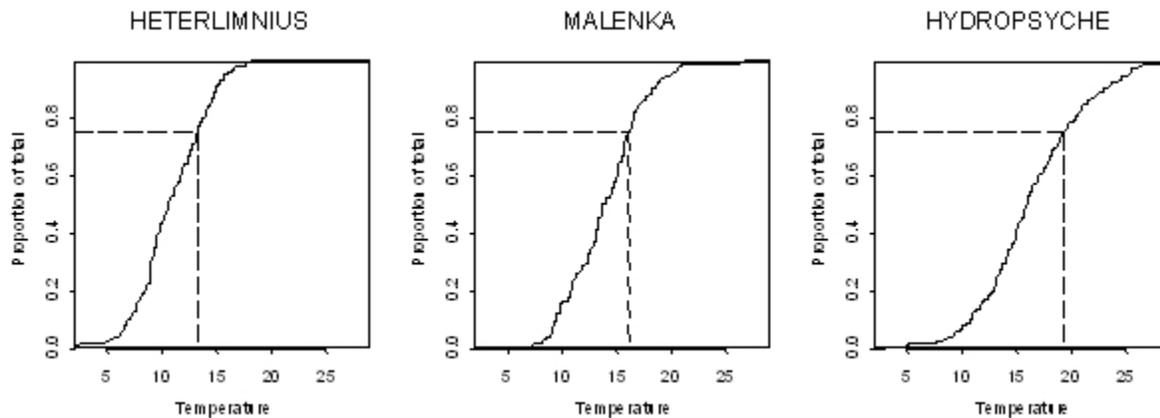
Weighted average tolerance values can be further refined if information regarding the strength of the association between taxon occurrence or abundance and the environmental gradient of interest is available. Then, tolerance values for taxa that are not strongly associated with the gradient can be omitted from future inferences. We return to this topic in Section 4.1.3.2.

### 3.1.2. Environmental Limits (Cumulative Percentiles)

Environmental limits can be estimated by computing cumulative percentiles (CPs) from field data. An empirical CP is estimated for a given value of the environmental variable ( $x_0$ ) as follows:

$$CP(x_0) = \frac{\sum_{i=1}^N Y_{ij} I(x_i < x_0)}{\sum_{i=1}^N Y_{ij}} \quad (2)$$

where  $I = 1$  if  $x_i < x_0$  and  $I = 0$  if  $x_i \geq x_0$  and other variables are as defined in eq 1. For presence/absence data, the numerator is the number of occurrences of taxon  $j$  at sites in which the value of the environmental variable is less than the cutoff value, and the denominator is the total number of occurrences of taxon  $j$ . Plots of CP as a function of  $x_0$  are shown in Figure 5 for three genera. A CP tolerance value would be estimated by fixing CP at a prescribed value and computing the  $x_0$  that corresponds to that value of CP for all taxa. Then,  $x_0$  is an estimate of the tolerance value. To estimate the maximum level of a stress under which a taxon could persist, CP would be fixed at a relatively high value (e.g., 0.75).



**Figure 5. Empirical cumulative distribution method for defining tolerance values.** Cumulative percentile value shown is 0.75. The point at which the dashed line intersects the horizontal axis is the tolerance value for that taxon.

Different practitioners have used different CP values to define tolerance values. Lenat (1993) tried different threshold probabilities and found that a value of 0.75 was most effective. Relyea et al. (2000) used a value that ranged from 0.97 to 0.99, depending on the taxon. The uncertainty in defining  $x_0$  for a given cumulative percentile increases as the percentile value increases, so the selection of the CP value requires that one balance competing factors: higher percentiles may more accurately express the environmental limit of an organism, but the error in empirically defining tolerance values using higher percentiles also increases. In Figure 5, a CP value of 0.75 is shown, which yields a temperature tolerance value of approximately 13°C, 16°C,

and 19°C, respectively, for the three taxa. This same CP value is used in all example computations.

## **3.2. INDIRECT METHODS**

### **3.2.1. Regression Estimates of Taxon-Environment Relationships**

To estimate tolerance values on the basis of taxon optima or curve shape, one first must estimate a taxon-environment relationship for each combination of taxon and environmental variable. In addition to providing a means of computing these additional tolerance values, regression estimates of the taxon-environment relationship provide the additional benefit of allowing one to quantify the strength of the association between a given environmental gradient and changes in the occurrence probability or abundance of a taxon. Combinations of taxon and environmental gradients that are not strongly associated can then be excluded from future inference.

To estimate taxon-environment relationships one usually must apply constraints on the form of the relationship. Two main types of constraints are possible. First, one can assume that the taxon-environment relationship follows a pre-specified functional form. After the functional form is specified, differences in taxon-environment relationships can be summarized by comparing the parameters of the regression relationships. We refer to these methods as *parametric regressions*. Second, one can require only that the taxon-environment relationship vary slowly and smoothly over the range of observations. In this case, the relationship cannot be described using a few simple parameters. We refer to these methods as *nonparametric regressions*.

Ordinary linear regression methods cannot be applied in either of these cases because both taxon presence/absence data and taxon abundance data are not normally distributed. Instead, “generalized” regression methods are used, which adapt linear regression approaches to non-normal distributions (Hastie and Pregibon, 1992). In the case of presence/absence data, the response variable is modeled as a binomial distribution; in the case of abundance data, a negative binomial distribution is often assumed.

#### **3.2.1.1. Parametric Regressions**

A common assumption for taxon-environment relationships is that the distribution of a particular taxon is unimodal with respect to environmental gradients (see Section 2.1). Then, in the case of presence/absence data, a convenient model for the probability of observing a particular taxon, following ter Braak and Looman (1986), is as follows:

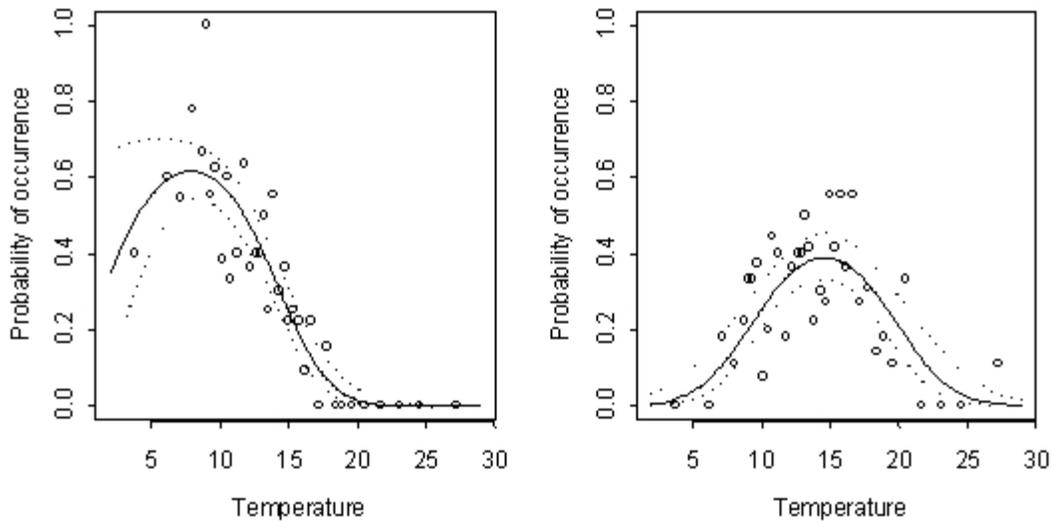
$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1x + b_2x^2 = a - \frac{1}{2} \frac{(x-u)^2}{t^2} \quad (3)$$

where  $p$  is the probability of observing the taxon, and the left side of the equation is the logit transformation of this probability. Additionally,  $x$  is the value of the environmental variable,  $u$  is the species optimum (the point along the environmental gradient where the probability of observing the species is maximized),  $t$  is a measure of the niche breadth, and  $a$  is related to the maximum probability of observation. The constants  $b_0$ ,  $b_1$ , and  $b_2$  can be determined using standard maximum likelihood estimation methods for fitting a curve to observed data and a generalization of linear regression methods (i.e., generalized linear models, GLMs). Then the parameters  $u$ ,  $t$ , and  $a$  can be determined as follows:  $u = -b_1/2b_2$ ,  $t = 1/\sqrt{-2b_2}$ , and  $a = b_0 - b_1^2/2b_2$ . Examples of taxon-environment relationships estimated using eq 3 for two genera, *Heterlimnius* and *Malenka*, with respect to stream temperature are shown in Figure 6. The computed curves closely track the observed capture probabilities for both genera. Confidence limits broaden for *Heterlimnius* as temperatures decrease to the minimum values, because data were sparse in that region.

A similar model can be specified for abundance data as follows:

$$\ln(A) = a - \frac{1}{2} \frac{(x-u)^2}{t^2} \quad (4)$$

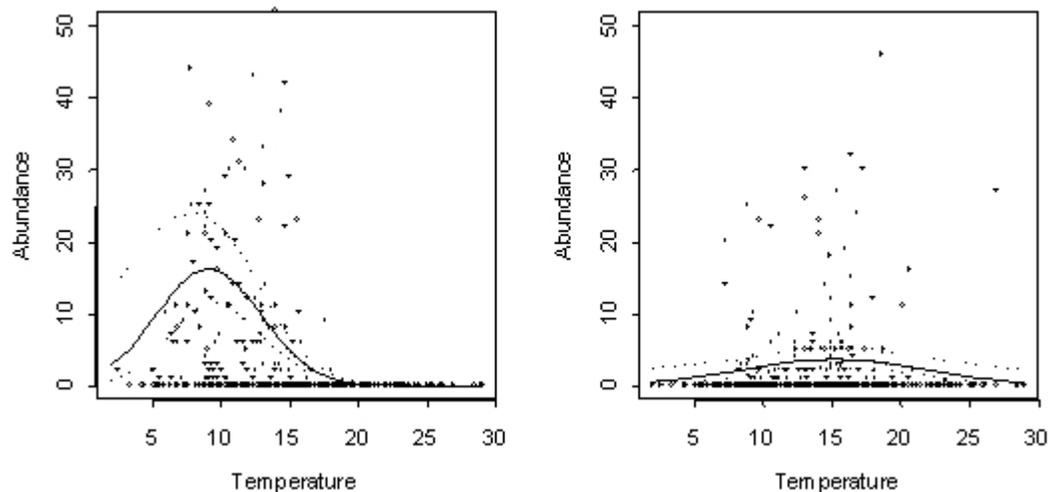
where  $A$  is the observed abundance in a sample and the other variables are defined as above. Abundance,  $A$ , is assumed to follow a negative binomial distribution, with a log-mean value equal to the right hand side of eq 4. With negative binomial distributions, the residual variance of  $A$  is a function of the mean value and one additional parameter that quantifies the degree to which the variance differs from a simple Poisson distribution (White and Bennetts, 1996). Abundance is loosely defined here because this modeling approach can be applied to both absolute and relative abundances. Solving the regression is somewhat more complicated because of the additional parameter, although GLMs can again be used.



**Figure 6. Relationship between probability of occurrence and temperature for *Heterlimnius* (left) and *Malenka* (right).** Solid line is mean relationship between probability of occurrence and temperature determined by logistic regression. Dotted lines are estimated 90% confidence limits about the location of the regression curve. Each open circle represents the average occurrence probability in approximately 10 samples surrounding the indicated temperature.

Examples of taxon-environment relationships estimated using eq 4 and relative abundance data are shown for the same two genera (*Heterlimnius* and *Malenka*) in Figure 7. From the data, it is evident that *Heterlimnius* tends to occur with greater relative abundance than does *Malenka*. Observations of relative abundance data also appear more variable than those of presence/absence data, but notice that each point in Figure 7 represents a single observation, whereas each point in Figure 6 represents the average of approximately 10 samples. Overall, the shapes of the relationship curves and the locations of the taxon optima were virtually identical for models using presence/absence and models using relative abundance, and this similarity held true for most taxa considered in this report. Therefore, attention is focused primarily on presence/absence relationships for the remainder of the report. Furthermore, because of the increased complexity associated with modeling abundance data, the use of presence/absence data for estimating taxon-environment relationships is recommended in most cases.

One potential advantage of explicitly estimating taxon-environment relationships is that the approach allows consideration of more variables than just the stressor of interest. That is, potential covarying factors can be taken into account by adding additional terms to the regression model. Multivariate models present other difficulties, though, so for simplicity, we consider only



**Figure 7. Relationship between relative abundance and temperature for *Heterlimnius* (left) and *Malenka* (right).** Solid line is mean relationship determined between relative abundance and temperature determined by a negative binomial regression. Dotted lines are estimated 90% confidence limits about the location of the mean relationship. Open circles represent observations. Thirteen observations with abundances greater than 50 are not shown for *Heterlimnius* and three observations with abundances greater than 50 are not shown for *Malenka*.

single-variable models in this report. The use of multiple variables is discussed further in Section 5.1.

The use of parametric functions to describe the taxon-environment relationship is both a strength and a weakness of the parametric approach. On the one hand, these functions provide the means to summarize the taxon-environment relationship using a short list of pre-defined parameters. On the other hand, the a priori assumption of a functional form may restrict the taxon-environment relationship to a shape that is not fully supported by field observations. Inspection of plots of observed data and modeled functional fits can help establish whether the assumed functional forms are appropriate.

### 3.2.1.2. *Nonparametric Regressions*

Many researchers have noted that unimodal relationships cannot be expected for all taxa across all gradients (Austin and Meyers, 1996; Oksanen and Minchin, 2002). To address this issue, a modeling approach is often used that requires only that the modeled function vary smoothly and slowly over the modeled range. Here, the distribution of a given taxon is modeled as follows:

$$\ln\left(\frac{p}{1-p}\right) = s_0 + s(x) \quad (5)$$

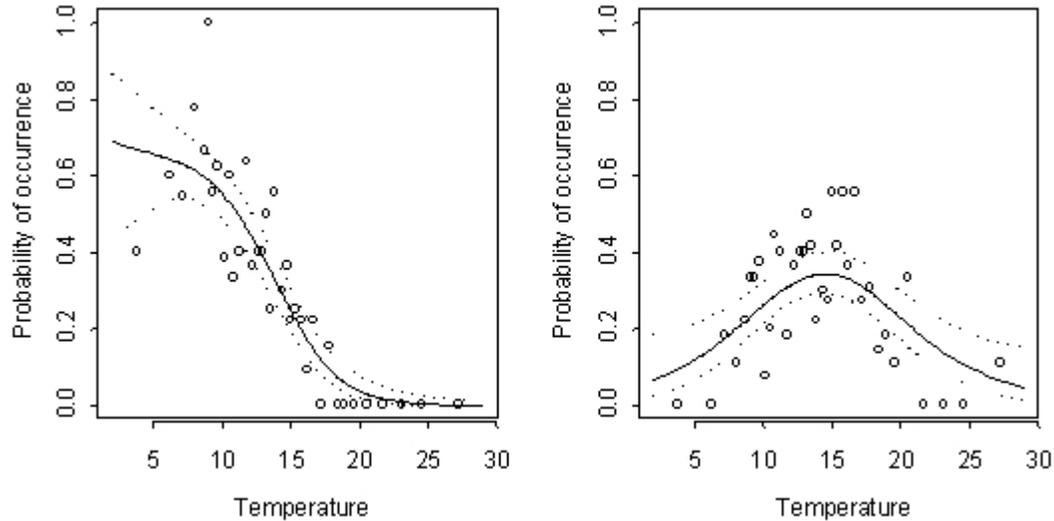
where  $p$  is defined as before,  $s_0$  is a constant, and  $s$  represents a nonparametric smooth curve that is fit through the data.

The locations of the mean responses for each point along a nonparametric curve,  $s$ , are determined through an iterative procedure that uses data in a local neighborhood around each point. The “local” nature of the fit differs fundamentally from that of a parametric model, which computes a best fit based on the entire set of data. Thus, nonparametric responses have the potential to capture smaller scale variations in response. Near the edges of the domain, though, sufficient data do not exist on both sides of the point of interest, and increasing amounts of data must be drawn from within the sampled range; therefore, the width of the neighborhood broadens, and the fit is less local than in the center of the domain (Hastie and Tibshirani, 1999).

This boundary effect is evident in Figure 8, where the response determined by nonparametric regression for *Heterlimnius* at low temperatures differs substantially from the response found by parametric regression (Figure 6). The probability of occurrence, as determined by nonparametric regression, continues to increase all the way to the lowest temperature, whereas the parametric-derived response decreases at the lowest temperature. Because the neighborhood used by the nonparametric curve increases as it approaches low temperatures, it incorporates more of the high-occurrence probabilities at slightly higher temperatures, which may have the effect of maintaining a high-occurrence probability all the way to the boundary of the data. In contrast, the parametric model forces a fit to a unimodal curve and decreases.

In both cases, the confidence limits are very broad at the lower boundary of the data, so the “true” response is impossible to determine from this data set. However, the observed frequency of occurrence at the lowest temperature (the left-most point in Figure 6) suggests that occurrence probabilities for *Heterlimnius* may decrease at the lowest temperature, a feature that is missed by the mean nonparametric response. In general, responses near the edges of the sampled gradient must be interpreted with caution.

A commonly used approach to fitting nonparametric curves is known as the generalized additive model (GAM) (Hastie and Tibshirani, 1999), which allows for more than one explanatory variable, each associated with its own nonparametric smooth curve. For now, though, we consider only a single explanatory variable and again defer discussion of multiple variables to Section 5.1. The flexibility of nonparametric regressions also complicates its use



**Figure 8. Relationship between probability of occurrence and temperature for *Heterlimnius* (left) and *Malenka* (right).** Solid line is mean relationship between probability of occurrence and temperature determined by logistic regression. Dotted lines are estimated 90% confidence limits about the location of the regression curve. Each open circle represents the average occurrence probability in approximately 10 samples surrounding the indicated temperature.

because there are no parameters with which the modeled relationship can be represented. Instead, a numerical representation of the entire curve must be stored for further analysis (e.g., to extract tolerance values).

### 3.2.1.3. Model Performance and Overfitting

Different species will vary in the degree to which their occurrence can be predicted by a particular environmental gradient, and quantifying these differences can be useful for characterizing the performance of different models of taxon-environment relationships. One useful way to quantify the performance of a model is to examine the relationship between the false positive rate and the true positive rate. The true positive rate is the proportion of sites at which a taxon was predicted to be present and sites where it was actually observed. The false positive rate is the proportion of sites at which the taxon was predicted to be present and sites where it was not observed. At a set of test sites, given the values of the environmental gradient and given a model for the taxon-environment relationship, we first compute the predicted probability of occurrence for that taxon. To compare predicted probabilities with actual observations of presence and absence, we then specify a threshold probability above which the

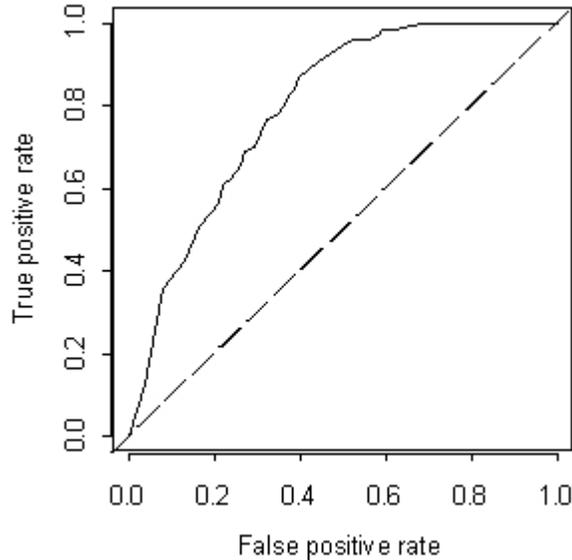
taxon is predicted to be present and below which the taxon is predicted to be absent. An example of this comparison is shown in Table 1 for *Heterlimnius* and a threshold probability of 0.5. The true positive rate in this case is  $71/(49+71)$ , or 0.59, and the false positive rate is  $54/(195+54)$ , or 0.22.

**Table 1. Observed versus predicted occurrences for *Heterlimnius***

<b>Taxon</b>	<b>Predicted absent</b>	<b>Predicted present</b>
Absent	195	54
Present	49	71

As the threshold value of  $p$  is increased, both false and true positive rates increase. We can quantify the trade-off by computing false and true positive rates over a range of threshold values and plotting them against one another (Figure 9). The resulting curve is known as the receiver operating characteristic (ROC) curve, and the area under this curve provides a measure of the classification strength of the model (Manel et al., 2001). The 1:1 line indicates the position of the ROC curve for a model in which the false positive rate is the same as the true positive rate, regardless of the choice of threshold value. Such a model has no classification power, and therefore this area (0.5) is the lowest possible ROC value. The area under the ROC curve approaches 1 as classification strength increases. In the example shown in Figure 9, the area under the ROC curve is 0.8. The minimum ROC value for an “acceptable” model varies with different studies. For this report, we selected a value of 0.55 as a provisional cut-off value. This cut-off value is relatively low compared to the more commonly-used value of 0.7 (Hosmer and Lemeshow, 2000). However, the use of the model here differs from conventional regression models in that we are interested only in characterizing the taxon–environment relationships and not interested in actually predicting the presence or absence of different taxa. The effects of using other cut-off values are explored in Section 4.1.3.2.

Overfitting the data can be an issue when developing taxon–environment relationships. To avoid overfitting regression models, it is generally recommended that at least 10 to 15 observations of the response variable being modeled occur in the data set for each degree of freedom in the explanatory variables. For example, to model the presence or absence of a taxon with an overall frequency of occurrence of 20%, we would require 50–75 samples for each degree of freedom that is contemplated. (Note that for a very common taxon that occurs at a majority of sites, the appropriate definition of an “observation” for this purpose is the absence of



**Figure 9. Receiver operator characteristic (ROC) curve for *Heterlimnius*.** Dashed line shows 1:1 line. Area under the ROC curve is 0.8.

the taxon at a site.) Because most unimodal relationships require at least two degrees of freedom to specify (e.g., the quadratic relationship shown in eq 3), we would require 100–150 samples for each variable. Species–abundance relationships suggest that many taxa are observed infrequently and only a few are relatively common. Thus, the number of taxa for which regression models can be developed may be limited.

### 3.2.2. Optima

After a taxon–environment relationship is modeled by regression, defining the optimum tolerance value is fairly straightforward. The optimum value in parametric regressions is explicitly defined (eqs 3 and 4). For nonparametric regressions, one numerically locates the point of maximum modeled occurrence probability along the regression relationship. Optima for species–environment relationships that increase or decrease monotonically are necessarily located at the edges of the environmental gradient. Thus, all species with monotonically increasing relationships would have the same maximum point, as would all species with monotonically decreasing relationships. In a region containing many such species, it can be difficult to distinguish between their relative tolerances for a given stressor gradient. Furthermore, the edges of the environmental gradient are usually defined only by the range of conditions sampled and can therefore vary between data sets.

### 3.2.3. Curve Classification

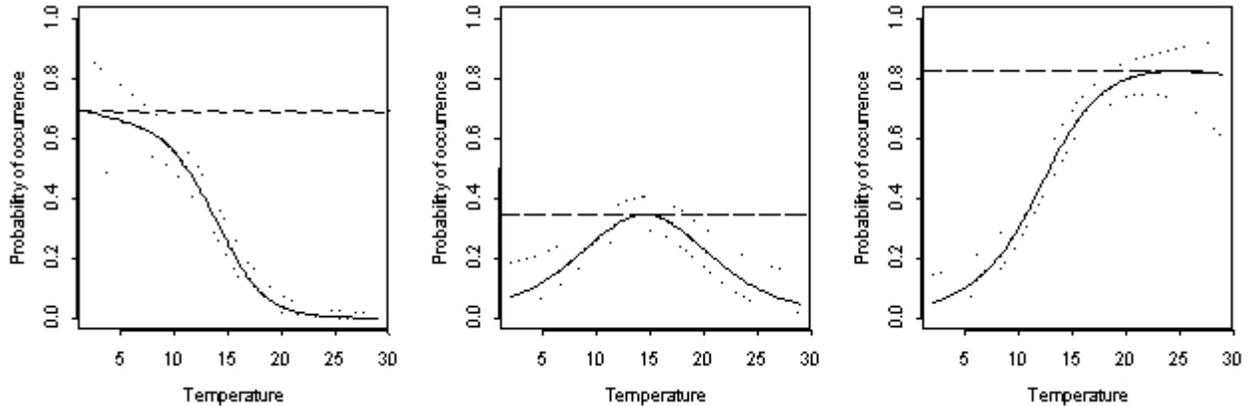
The process of curve classification can be accomplished analytically or graphically, depending on whether parametric (GLM) or nonparametric (GAM) regressions are used to define the taxon-environment relationship. In a parametric regression as defined by eq 3, the curve can be classified according to the statistical significance of different coefficients in the regression. More specifically, if the quadratic term in eq 3 reduces the model deviance by a statistically significant amount, then the relationship is unimodal. Otherwise, the model equation reduces to a linear relationship, and the relationship can be classified as increasing or decreasing, depending upon the sign of linear coefficient ( $b_1$  in eq 3). In rare cases, the quadratic term is statistically significant but the sign of the coefficient ( $b_2$  in eq 3) is positive, indicating that the relationship is concave up. In these cases, the data and the estimated taxon-environment relationship should be examined more carefully to identify reasons for the anomalous behavior.

A graphical approach to curve classification is required for nonparametric regressions, in which the maximum modeled occurrence probability is compared with the confidence limits on either side of the taxon optima. If a horizontal line drawn through the maximum mean occurrence probability deviates from the upper confidence limits on both sides of the taxon optima (e.g., *Malenka* in Figure 10), then the taxon is designated as unimodal. If the line deviates from the upper confidence limit only on the right-hand side of the taxon optimum (e.g., *Heterlimnius* in Figure 10), then the taxon is designated as a decreaser; if it deviates only on the left-hand side (e.g., *Hydropsyche* in Figure 10), then the taxon is designated as an increaser. A parallel set of conditions can be specified with regard to the lower confidence limit. Assuming that human activities cause an increase in the value of the environmental gradient, increasers would be identified as tolerant taxa, decreasers would be identified as intolerant taxa, and unimodal taxa would be designated as intermediately tolerant.

## 3.3. COMPARING DIFFERENT TOLERANCE VALUES

### 3.3.1. Continuous Tolerance Values

We computed tolerance values for sediment and temperature using six different methods: (1) weighted averages (WA), (2) cumulative 75<sup>th</sup> percentile (CP75), (3) parametric regressions combined with the point of maximum occurrence probability (GLMMAX), (4) parametric regressions combined with curve shape classification (GLMCL), (5) nonparametric regressions combined with the point of maximum occurrence probability (GAMMAX), and (6) nonparametric regressions combined with curve shape classification (GAMCL). In both GLM



**Figure 10. Graphical approach for classifying curve shape for *Heterlimnius*, *Malenka*, and *Hydropsyche*.** Solid line is mean response, dotted lines are estimated 90% confidence limits about mean response, dashed line is location of maximum mean response for comparison with confidence limits.

and GAM models, only a single explanatory variable was specified, and it was modeled with two degrees of freedom. Tolerance values were derived only for taxa that occurred in at least 20 sites. Additionally, GLMMAX, GAMMAX, and curve classification tolerance values were computed only for models for which  $ROC > 0.55$ . All tolerance values are tabulated in Appendix C.

All sets of tolerance values except those determined by curve shape classification were compared by computing correlation tables and examining scatter plots. Overall, tolerance values derived by different methods were highly correlated. All types of continuous temperature tolerance values (GLMMAX, GAMMAX, WA, and CD75) were strongly correlated with  $r$ , ranging from 0.88 to 0.97 (Table 2). Correlation coefficients between different types of sediment tolerance values were also high, varying from 0.84 to 0.98 (Table 3). The weakest correlation

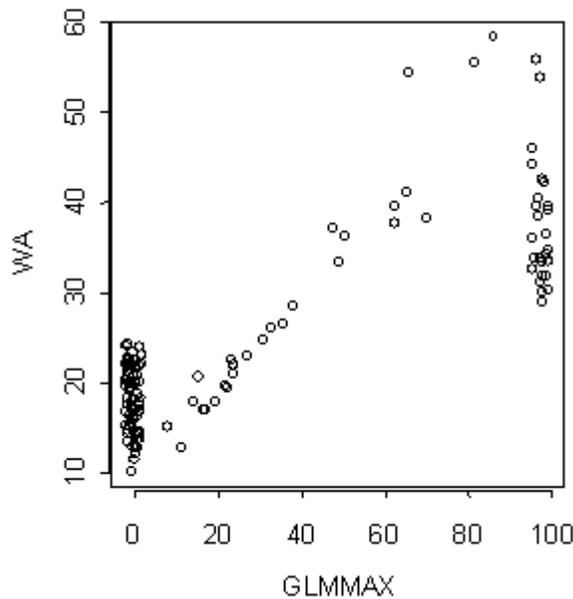
**Table 2. Correlation coefficients for temperature tolerance values**

Method	GAMMAX	WA	CD75
GLMMAX	0.96	0.89	0.88
GAMMAX		0.93	0.91
WA			0.97

**Table 3. Correlation coefficients for sediment tolerance values**

Method	GAMMAX	WA	CD75
GLMMAX	0.98	0.84	0.87
GAMMAX		0.88	0.90
WA			0.97

found was between GLMMAX and WA, primarily because taxon-environment relationships were often monotonic and GLMMAX tolerance values were pinned at either edge of the sampled domain (Figure 11). This same phenomenon was observed for temperature, but a smaller proportion of taxon-environment relationships for temperature were monotonic and the correlation coefficients were not as strongly affected.



**Figure 11. Comparison of optima tolerance value determined by parametric logistic regression model (GLMMAX) and weighted average tolerance values (WA) for temperature.** Each point represents a different taxon. Axis scales are nominally in units of °C. Points have been jittered to more clearly show overlapping points.

The close relationships between different types of tolerance values is not surprising, given that they were estimated from the same set of data. The main reasons for differences likely stem from uncertainties in estimating the point of maximum occurrence probability for gradients that

were only partially sampled. The distribution of samples across the gradient also influences differences between WA and GLMMAX or GAMMAX tolerance values (ter Braak and Looman, 1986).

### 3.3.2. Tolerance Classifications

We compared curve classification tolerance values estimated by different methods using a confusion matrix, in which each position in the matrix corresponds to a combination of GLMCL and GAMCL categories and the number in that position is the number of taxa for which a particular combination of tolerance classifications was found. For example, 35 taxa were identified as sensitive to elevated temperature by both GAMCL and GLMCL and 33 taxa were classified as intermediately tolerant by GLMCL and as sensitive by GAMCL (Table 4). Comparisons of curve classification tolerance values for fine sediment are shown in Table 5.

**Table 4. Confusion matrix for temperature tolerance classifications**

GLM \ GAM	Sensitive	Intermediate	Tolerant
Sensitive	35	0	0
Intermediate	33	16	21
Tolerant	0	0	30

**Table 5. Confusion matrix for sediment tolerance classifications**

GAM \ GLM	Sensitive	Intermediate	Tolerant
Sensitive	69	0	0
Intermediate	16	1	9
Tolerant	0	0	27

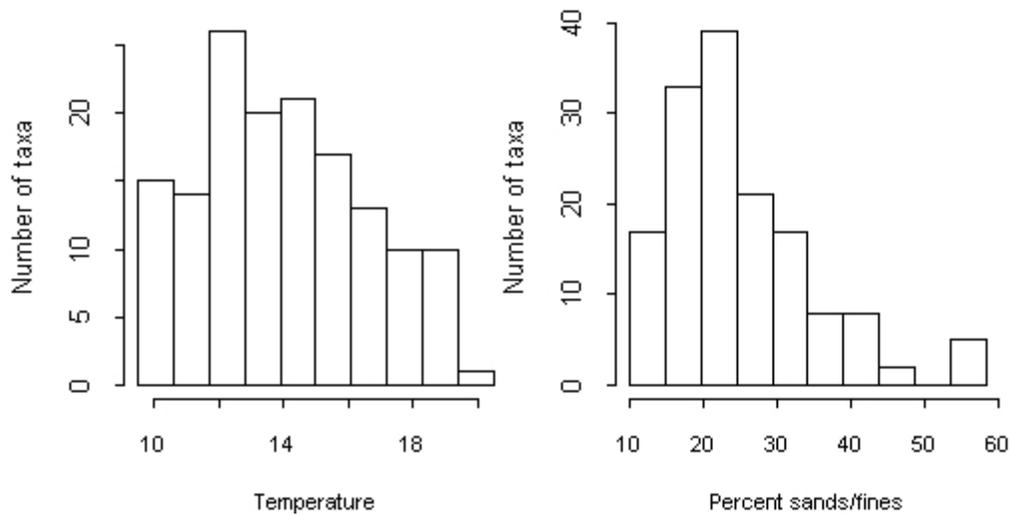
In general, the GLMCL reliance on the statistical significance of the quadratic term yielded many more intermediately tolerant classifications (i.e., unimodal) than did the graphical classification of GAMCL. That is, the quadratic term was often statistically significant, but the confidence intervals surrounding the relationship were too broad to permit a graphical classification of the relationship as unimodal. The taxon-environment relationship for *Heterlimnius* (Figure 6) provides a good illustration of this effect. Here, the unimodal response

was statistically significant, but the confidence intervals at the lowest temperatures were so broad that the graphical method classified this relationship as monotonically decreasing. For our purposes, classifications into tolerant and sensitive categories are more directly useful. Furthermore, the graphical classification method can be applied to both parametric and nonparametric regression results, so we focus only on GAMCL categories for the remainder of this report.

### **3.3.3. Comparisons Between Continuous and Categorical Tolerance Values**

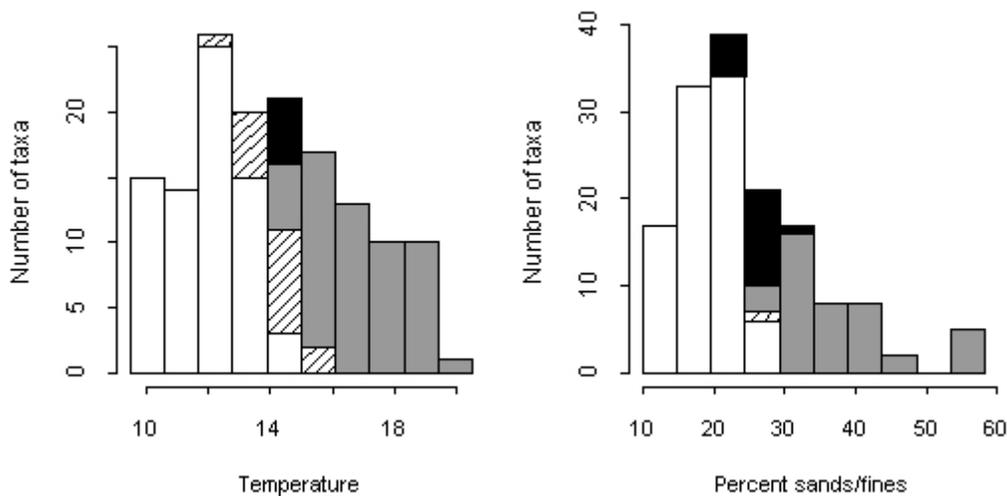
Curve classifications provide a broad characterization of taxa into tolerant or sensitive categories that complement continuous tolerant values (e.g., CP75, WA). In previous analyses, taxa have been classified as tolerant or sensitive on the basis of an existing continuous tolerance value. For example, Klemm et al. (2002) categorized taxa that had already been assigned tolerance values ranging from 0 to 10. In their scheme, taxa with tolerance values less than or equal to 4 were classified as sensitive and taxa with tolerance values greater than or equal to 6 were classified as tolerant. One concern with this approach is that the threshold value used to discriminate between tolerant and sensitive taxa is determined through best professional judgement and may not accurately capture the point at which taxon-environment relationships actually exhibit a substantive change (e.g., a shift from a decreasing relationship to an increasing relationship).

To further explore this issue, consider the distribution of WA optima values for sediment computed from EMAP-West data (Figure 12). We can probably assume that taxa that have low sediment optima are sensitive to excess fine sediment and taxa that have high sediment optima are tolerant. In the middle of the range, though, it is difficult to identify a single WA optimum value below which we are confident that all taxa are sensitive or above which we are confident that all taxa are tolerant. Similarly, taxa that have low temperature optima are likely to be sensitive to elevated temperature and taxa that have high temperature optima are likely to be tolerant to elevated temperature, but taxa with moderate optima are difficult to categorize (Figure 12). A potential solution to this dilemma is to designate taxa with tolerance values falling in the middle of the range as indifferent, following Klemm et al. (2002). However, this approach is conservative, and it is likely that some taxa that are truly sensitive or tolerant will be classified as indifferent.



**Figure 12. Histogram of weighted average tolerance values for temperature (left) and fine sediment (right).**

Tolerance values based on curve shape classifications provide a good alternative to continuously ranked tolerance values. Curve shape classification provides a direct measure of whether a taxon increases or decreases in response to anthropogenic stress, a more definitive classification than can be achieved with continuous tolerance values. We look again at the distribution of WA optima for temperature and sediment, this time color-coded for tolerance classification in terms of curve shape (Figure 13). As expected, taxa classified by curve shape as tolerant are clustered to the right-hand side of the histogram, with high WA optima, and those classified as sensitive are clustered to the left-hand side of the histogram. At moderate WA optima values, though, the correspondence between the value of the optima and the tolerance classification is much weaker.



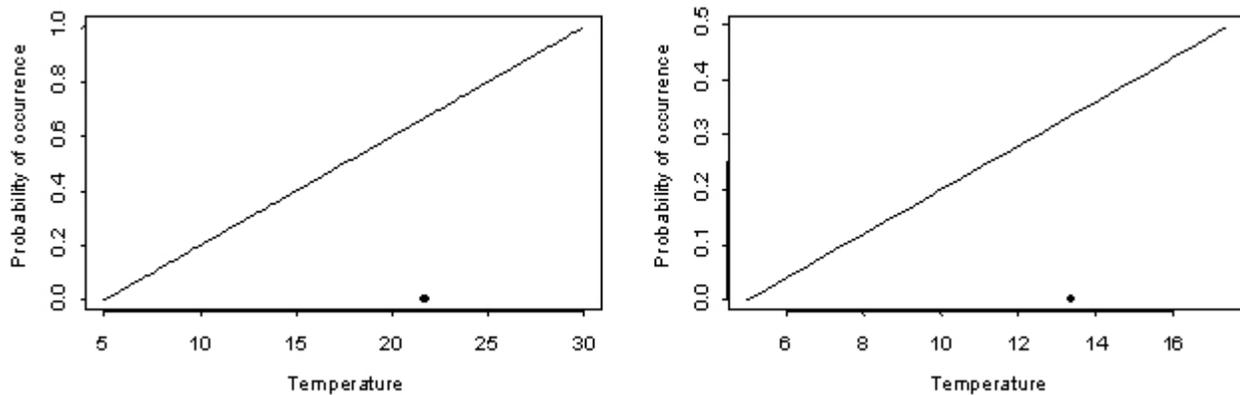
**Figure 13. Histogram of weighted average tolerance values for temperature (left) and fine sediment (right) classified by curve shape.** Shading in bars indicate numbers of taxa within each group classified as sensitive (open), intermediately tolerant (hatched), and tolerant (gray). Black bars indicate taxa for which classifications were not assigned.

### 3.3.4. Effects of Regional Characteristics

All tolerance values discussed above depend on the range of conditions under which they are derived. This dependence stems from the fact that the range of sampled conditions imposes arbitrary limits on the functions used to compute tolerance values. A simple example of this effect can be seen in Figure 14, in which we consider a taxon-environment relationship that increases linearly with temperature. In the left plot, samples are collected across a range of temperatures from 5 to 30°C, whereas in the right plot, samples are collected across half the range. In the first case the weighted average is 22°C and in the second case the weighted average is 13°C. Similar changes in tolerance values would be observed for values derived by cumulative percentile methods.

Optima and curve classification tolerance values are somewhat less susceptible to this effect, but none are immune. For example, optima tolerance values for monotonically increasing or decreasing taxon-environment relationships are located on the edge of the sampled gradient. These values therefore vary with the range of conditions sampled. Curve classification, on the other hand, can be sensitive to the sampled range because species identified as monotonically increasing could be identified as unimodal in a data set that samples a broader range of

conditions. Thus, regardless of the derivation method, very different tolerance values could be derived for a given species, depending on the range of data that are collected.



**Figure 14. Illustrations of effect of gradient length.** Solid circle indicates location of the weighted average tolerance value.

We must consider the effects of regional characteristics when comparing tolerance values across regions or when using tolerance values derived from different study areas in the same assessment. When comparing tolerance values, the absolute values derived in different study areas will always require that the range of conditions sampled in each of the study areas be considered. Alternatively, the relative rankings of sensitivity should be relatively insensitive to the range of sampled conditions, so comparisons of relative rankings across study areas should be fairly straightforward. Furthermore, direct comparisons of species–environment relationships are also possible.

One of the potential strengths of using tolerance values for assessment is that once a tolerance value is derived for a species, it can be used wherever that species is observed. However, as discussed above, the generality of a tolerance value across different regions is influenced by the data set from which it was derived and the range of conditions sampled in that data set. In certain cases, it may be necessary to use tolerance values derived from different data sets within the same assessment. In such cases, tolerance values must be examined to ensure that they are based on the same ranges of environmental conditions.

## **4. APPLYING TOLERANCE VALUES IN ASSESSMENT**

On their own, tolerance values provide valuable information about the relative sensitivity of different taxa to different types of anthropogenic stressors. For example, taxa lists at impaired and reference sites can be compared in terms of the presence or absence of taxa with different tolerance values to infer possible sources of stress. The loss of particularly sensitive taxa could also provide an early indication of impairment. However, to most effectively apply tolerance values for biological assessment, biological metrics that summarize the observations of many different tolerant and sensitive taxa at a given test site are required. These summaries must then be compared with baseline conditions to ascertain whether observed changes are statistically or biologically significant. Therefore, evaluating the efficacy of different tolerance value derivation methods requires that we consider them in conjunction with biological metrics and baseline conditions.

### **4.1. BIOLOGICAL METRICS**

Biological metrics provide the means of summarizing the tolerance values of all of the different taxa observed at a test site. The available types of metrics differ for categorical and continuous tolerance values, so we discuss them in two separate sections.

We assess the performance of different metrics by comparing their values at independent test sites with observations of stressor levels at those same sites. Ideally, metric values should be strongly associated with the observed stressor level, and the variability about this mean response should be low. These characteristics would suggest that small changes in the stressor level could be detected with the biological metric. We compare the performance of biological metric values computed at different sites using data collected in Oregon by the Oregon Department of Environmental Quality (DEQ) (Appendix B). To compute metrics values, taxon abundance data from Oregon streams were combined with tolerance values estimated previously from EMAP-West data (Appendix C). Observations of stream temperature and fine bedded sediment were also available at each of the sites. The Oregon data were collected from a small area within the larger EMAP-West region and constitute a completely independent set of data.

#### **4.1.1. Metrics Based on Categorical Tolerance Values**

Compositional metrics summarize compositional characteristics of a sampled assemblage, and metrics that incorporate aspects of tolerance values (i.e., tolerance value metrics) have frequently been shown to distinguish between degraded and reference streams (Barbour et al., 1999). Examples of tolerance metrics include the relative abundance of tolerant

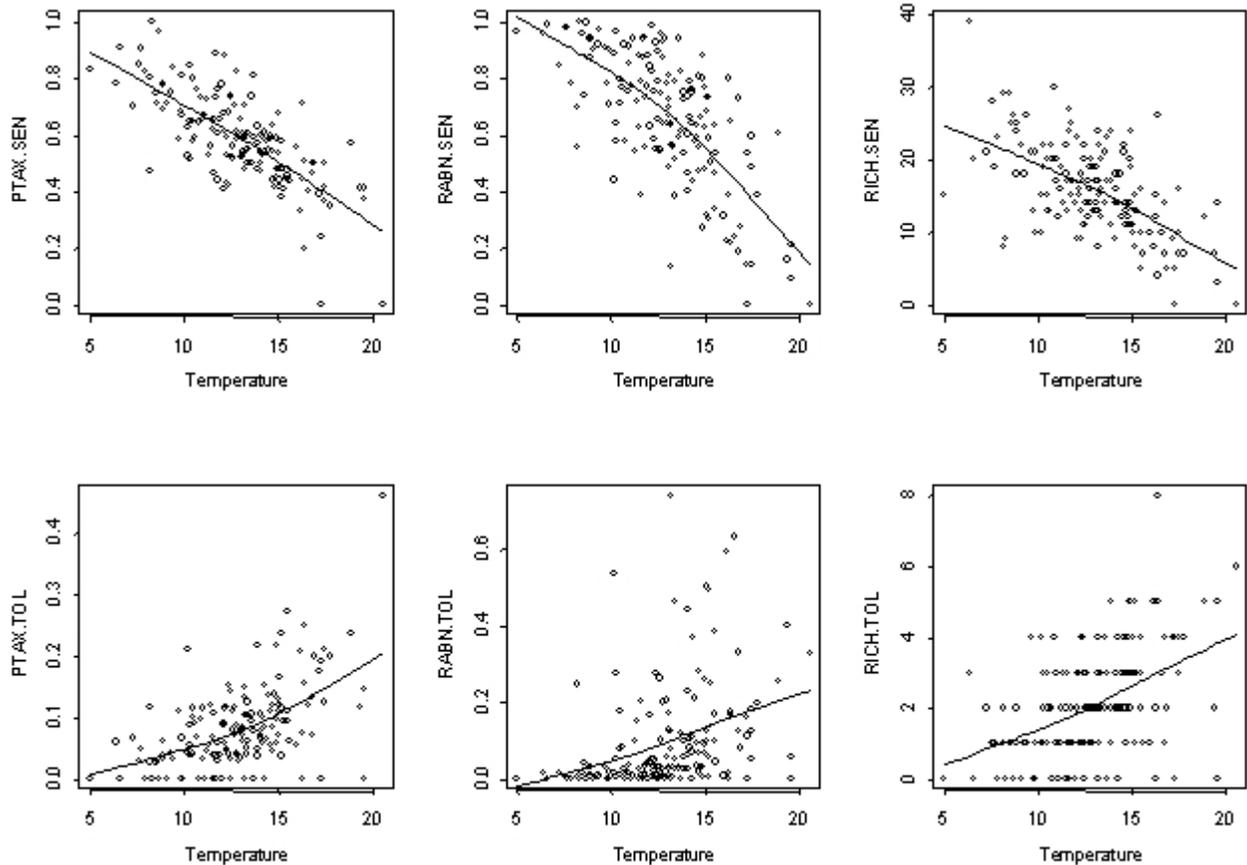
or sensitive taxa, the proportion of total taxa that are tolerant or sensitive, and the richness of tolerant or sensitive taxa.

We used GAMCL classifications estimated from EMAP-West data to group taxa collected in Oregon into sensitive and tolerance categories. Then, we computed values at the Oregon sites for relative abundance of tolerant (RABN.TOL) and sensitive taxa (RABN.SEN), richness of tolerant (RICH.TOL) and sensitive (RICH.SEN) taxa, and proportion of total taxa of tolerant (PTAX.TOL) and sensitive (PTAX.SEN) taxa. Computed metric values were then plotted against observed values of stream temperature and fine sediment. All of the temperature-specific metrics exhibited strong relationships with the observed temperature (Figure 15). The proportion of total taxa that were sensitive appeared particularly strongly related to stream temperature. Many sediment-specific metrics also exhibited reasonably strong relationships with the sediment gradient (Figure 16). As with temperature, the proportion of total taxa that were sensitive exhibited the least variability in its relationship with observed sediment levels.

Metrics quantifying characteristics of tolerant taxa were more weakly associated with observed stressor levels, and the magnitude of variability about the mean values was substantially greater. This higher variability can likely be explained by the relatively low richness of tolerant taxa observed (temperature-tolerant taxon richness increased to only about four taxa at the warmest sites in the study area, and sediment-tolerant taxon richness increased to only about two taxa at sites with high percentages of sands and fines). Low richness, then, contributed to high variability in relative abundance values and proportions of total taxa because the presence or absence of a single taxon could cause large changes in the values of these metrics.

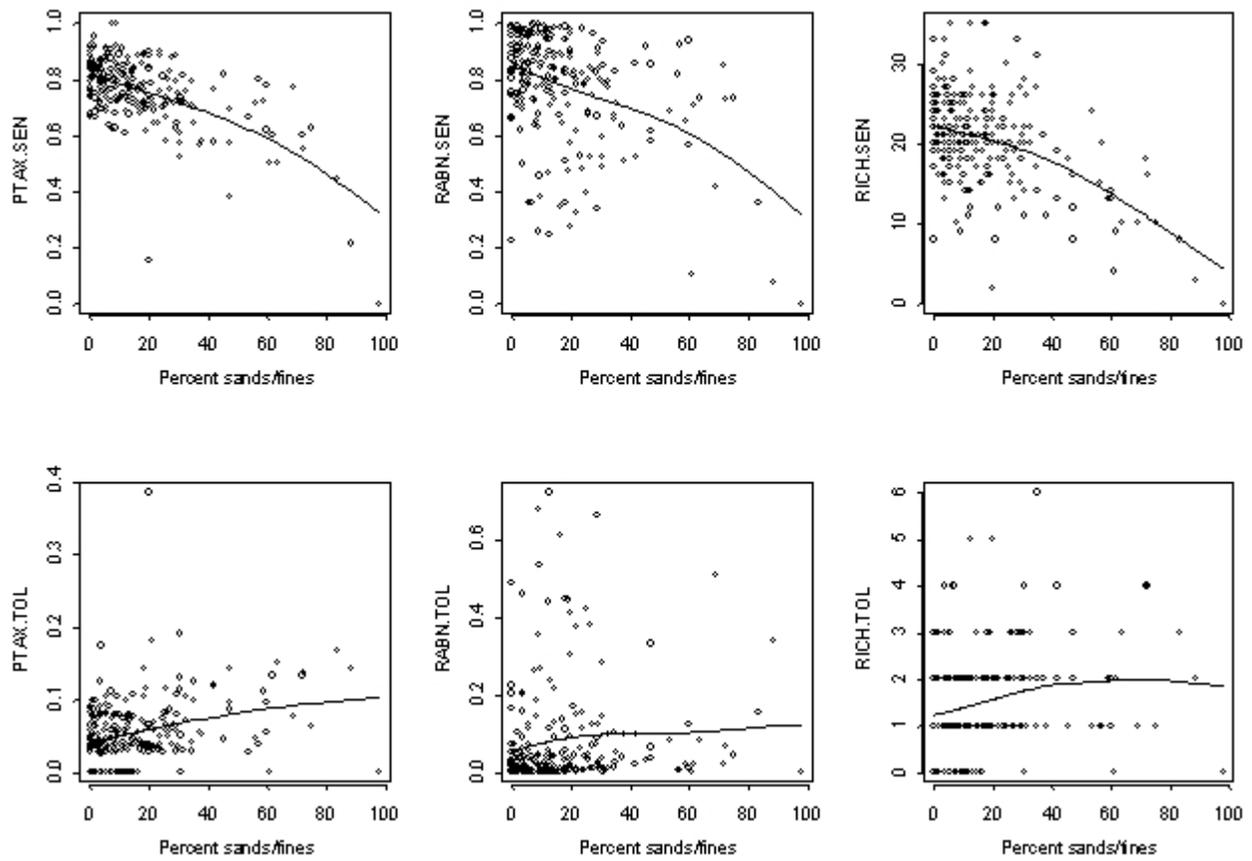
But why is tolerant taxon richness so much lower than sensitive taxon richness? Definitive answers to this question are elusive. One reason may be that fewer taxa were identified as tolerant than as sensitive. For example, a total of 71 of the taxa found in Oregon were classified as sensitive to elevated temperature, whereas only 44 were classified as tolerant. However, the difference in the number of sensitive versus tolerant taxa does not seem large enough to explain the strong differences observed in richness. Another contributing factor may be that sites with elevated temperature and increased fine sediment sites may also be affected by other stressors, the aggregate effect of which is to strongly depress taxon richness. The potential effects of such co-occurring stressors require further study.

To quantify the variability in the association between each metric and the stressor gradient of interest, a nonparametric smoothing spline (as seen in eq 5) was fit to each combination of tolerance value metric and stressor gradient (shown as solid lines in the Figures 15 and 16). Nonparametric curves were chosen because there was no a priori reason to expect that the relationships between metrics and stressor gradient would be linear. An  $R^2$  statistic



**Figure 15. Relationship between temperature tolerance metrics and observed temperature in Oregon.** PTAX.SEN: proportion of total number of taxa classified as sensitive; RABN.SEN: relative abundance of sensitive taxa; RICH.SEN: taxon richness of sensitive taxa; PTAX.TOL: proportion total taxa classified as tolerant; RABN.TOL: relative abundance of tolerant taxa; RICH.TOL: taxon richness of tolerant taxa. Solid line shows position of a smoothing spline fit through the data.

expressing the proportion of variability in the biological metric that was associated with changes in the stressor level was then computed (Table 6). In general, changes in temperature accounted for more variability in the temperature tolerance metrics than did sediment in the sediment tolerance metrics. Also, environmental observations accounted for more variability in metrics related to sensitive taxa than metrics related to tolerant taxa.



**Figure 16. Relationship between sediment tolerance metric and observed sediment in Oregon.** Vertical axis abbreviations and symbols as in Figure 15.

**Table 6.  $R^2$  values for spline fits between metric and stressor level**

<b>Metric</b>	<b>Temperature</b>	<b>Sediment</b>
PTAX.SEN	0.53	0.39
RABN.SEN	0.50	0.19
RICH.SEN	0.35	0.23
PTAX.TOL	0.30	0.09
RABN.TOL	0.12	0.02
RICH.TOL	0.23	0.05

#### 4.1.2. Metrics Based on Continuous Tolerance Values

All of the metrics defined in the previous section can be computed using continuous tolerance values, as long as the continuous tolerance value are first converted to categories (see Section 3.3.3). However, a more natural metric to compute—one that makes use of the full range of values—is the mean tolerance value of taxa observed at a site. If abundance data are available, the tolerance values of each taxon can be weighted by the observed abundance. The computation of these mean tolerance values can be expressed as follows:

$$u_m = \frac{\sum_{j=1}^N Y_{ij} u_j}{\sum_{j=1}^N Y_{ij}} \quad (6)$$

where  $u_m$  is the mean tolerance value,  $N$  is the number of taxa, and  $u_j$  is the tolerance value for taxon  $j$ .  $Y_{ij}$  is defined as previously: for presence/absence data,  $Y_{ij} = 1$  when a taxon  $j$  is present, and  $Y_{ij} = 0$  when taxon  $j$  is absent from site  $I$ ; for abundance data  $Y_{ij}$  is the abundance of taxon  $j$  at site  $I$ . This formula was used by Hilsenhoff (1987) and Lenat (1993) and in weighted average inferences of environmental conditions (e.g., Birks et al., 1990). The only differences between various analyses are the tolerance value that is used and whether abundance or presence/absence data are applied.

To compare the performance of different tolerance values, continuous tolerance values (WA, CP75, GLMMAX, and GAMMAX) were computed from EMAP-West data for temperature and fine sediment. Then, the mean of the tolerance values for taxa observed in each test site in Oregon were computed. A nonparametric smoothing spline was then fit to the relationships between the mean sediment and temperature tolerance values and observed sediment and temperature, and an  $R^2$  statistic computed for each fit. The results are summarized in Table 7.

Overall, the mean sediment and temperature tolerance values were strongly associated with observed sediment levels and temperature. All types of temperature tolerance values accounted for comparable proportions of the variability in the observations, with  $R^2$  ranging from 0.49 to 0.56. For sediment, GAMMAX and GLMMAX performed particularly poorly, with  $R^2 = 0.31$  and 0.28, respectively. For any given type of tolerance value, temperature was more strongly associated with the mean temperature tolerance value than was sediment with the mean sediment tolerance value. WA and CD75 consistently exhibited less variance in their relationships with each environmental variable than did GLMMAX and GAMMAX.

**Table 7. R<sup>2</sup> values for spline fits between mean tolerance values and observed temperature and sediment in Oregon** (One outlier dropped from sediment data.)

Method	Temperature	Sediment
WA	0.56	0.45
CD75	0.56	0.42
GLMMAX	0.49	0.28
GAMMAX	0.49	0.31

Overall, compared with the compositional metrics computed from categorical tolerance values (Table 6), the mean tolerance values exhibited less variable relationships with the observed levels of temperature and sediment. These higher R<sup>2</sup> values likely are a result of two factors. First, the compositional metrics used only two categories to represent tolerance values, whereas the mean tolerance values captured the full range of tolerance values. Second, each mean tolerance value depended on changes in both tolerant and sensitive organisms, so both increasing and decreasing taxa influenced the final mean value. Compositional metrics disaggregate increasing and decreasing responses into distinct groups, which may lessen their ability to represent the overall variability in the system.

#### **4.1.3. Performance of Tolerance Value Metrics**

We observed systematic differences in the performance of tolerance value metrics across different stressors and across different types of tolerance values. In this section we discuss some of the reasons for these differences.

##### **4.1.3.1. Characteristics of the Stressor Gradient**

We observed a consistent difference in the performance of the tolerance value metrics for the two stressors considered as examples in this section: temperature tolerance metrics were always more strongly correlated with temperature observations than were the corresponding comparisons computed for sediment. This performance difference across stressors may reflect (1) differences in the distribution of the sampled values for the two gradients, (2) differences in the mechanisms by which temperature and sediment affect biota, and (3) differences in our ability to measure these gradients.

First of all, a relatively uniform distribution of stream temperatures was sampled in the EMAP-West. In contrast, the percentage of fine sediment in sampled streams was generally low,

and the distribution of fine sediment levels was strongly skewed. A skewed distribution of the underlying environmental gradient has been shown to degrade the accuracy of weighted average optima (ter Braak and Looman, 1986) and would similarly degrade the accuracy of all other types of tolerance values. One way to mitigate the effects of a skewed distribution is to stratify and resample the existing data such that all locations on the environmental gradient are equally represented, but this procedure can introduce errors.

Second, the mechanisms by which stream temperature can influence stream macroinvertebrates are direct and relatively well understood. For cold-blooded organisms such as stream macroinvertebrates, virtually all metabolic processes are linked to environmental temperature, so stream temperature is necessarily a strong determinant of assemblage composition (Allan, 1995). Conversely, the mechanisms by which fine sediment affect stream invertebrates are less well understood and in many cases, less direct. Some proposed mechanisms by which fine sediment can affect aquatic insects include reducing habitat by filling interstitial spaces on stream bottoms and increasing the scour of the stream bottom in high-flow events (Wood and Armitage, 1997). Because of these mechanistic uncertainties, taxon-environment relationships for fine sediment are likely to be less precise than those for temperature.

The final difference between the two stressors is in the precision with which measurements can be collected. Quantifying the composition of a heterogeneous stream bottom is accomplished by measuring a relatively small number of individual particles collected at prescribed locations. Stream temperature is relatively homogeneous throughout the stream. Measurements of substrate composition are therefore likely to be somewhat more variable than measurements of temperature. Note, however, that both temperature and sediment exhibit a high degree of temporal variability, which is measured poorly by one-time samples.

All of the factors discussed above may influence the predictive performance of the tolerance values. Determining the relative contribution of each factor to the observed difference in predictive accuracy requires further study.

#### **4.1.3.2. *Acceptance Criteria for Tolerance Values***

Inherent in the GAM and GLM approaches was a model acceptance criterion whereby models with classification accuracy below a certain limit ( $ROC \leq 0.55$ ) were rejected and no tolerance value was computed. No model acceptance criteria are typically imposed on WA or CD computations, although in our case we did limit WA and CD optima computations to those taxa that occurred in at least 20 sites. In additional tests (Table 8), we explored the effects of tightening the acceptance

**Table 8. R<sup>2</sup> values for models with ROC > 0.65 criteria imposed**

<b>Method</b>	<b>Temperature</b>	<b>Sediment</b>
WA	0.60	0.40
CD75	0.60	0.38
GLMMAX	0.55	0.29
GAMMAX	0.60	0.28

criteria to ROC > 0.65, which omitted additional taxa from the computation of the metric values. We then recomputed average tolerance values for GLMMAX and GAMMAX using this reduced set of taxa. For comparison, we also recomputed average tolerance values for WA and CD75 using this same reduced set of taxa. The effects of the change in acceptance criteria differed for temperature and sediment. For temperature, tightening the ROC criteria slightly increased R<sup>2</sup> values. The R<sup>2</sup> for GAMMAX, WA, and CD75 were all 0.60, while the R<sup>2</sup> for GLMMAX was slightly less. In contrast, for sediment, tightening the criteria actually decreased R<sup>2</sup> values for WA and CD75, while R<sup>2</sup> values for GAMMAX and GLMMAX were essentially unchanged.

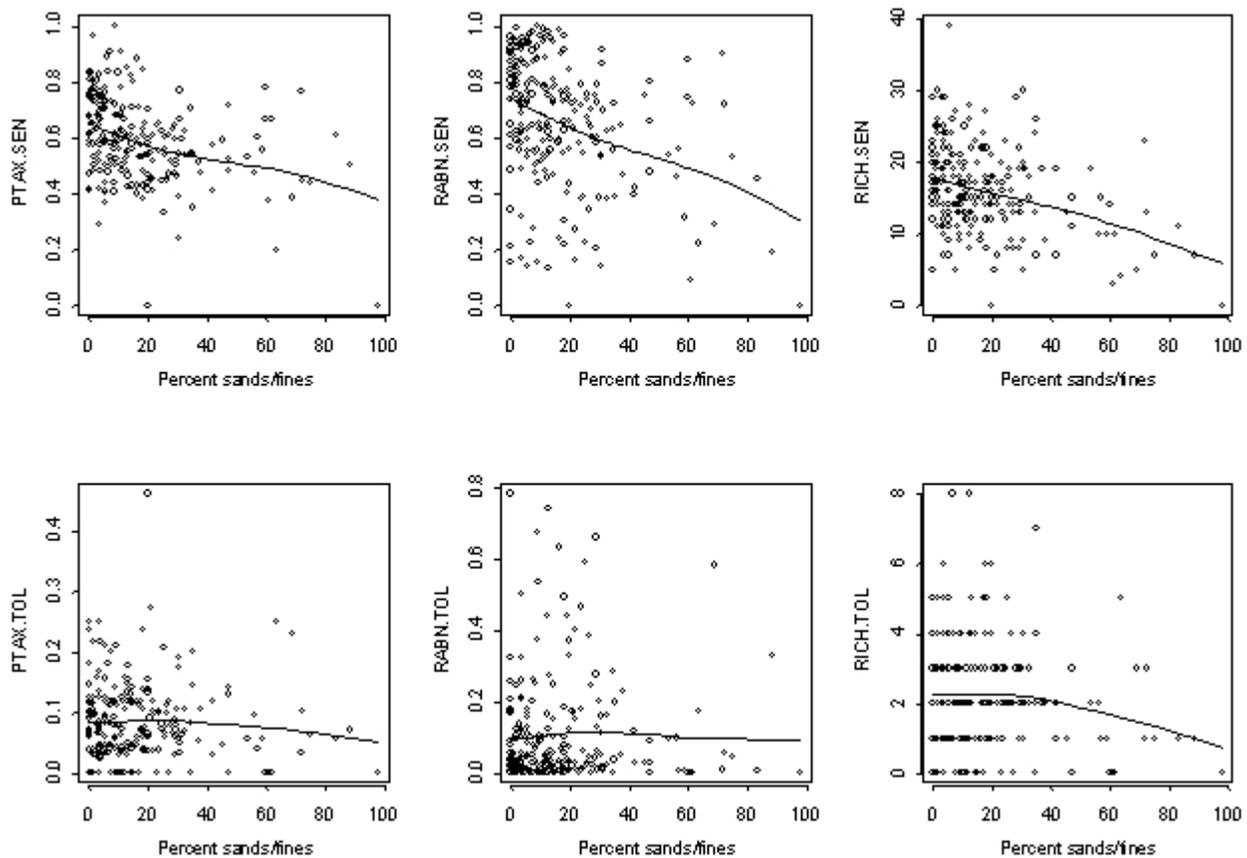
As noted, the distribution of sediment observations in the EMAP-West data set were strongly skewed to the right, which created difficulties for statistical modeling. Many taxa that exhibited strong relationships with the sediment gradient may have failed to meet the acceptance criteria because of this distribution. Thus, for skewed distributions, it seems likely that the ROC > 0.65 criterion was too stringent. On the other hand, for gradients that are sufficiently and evenly sampled (such as temperature), a more restrictive ROC criterion may improve performance.

#### **4.1.3.3. *Specificity of Response***

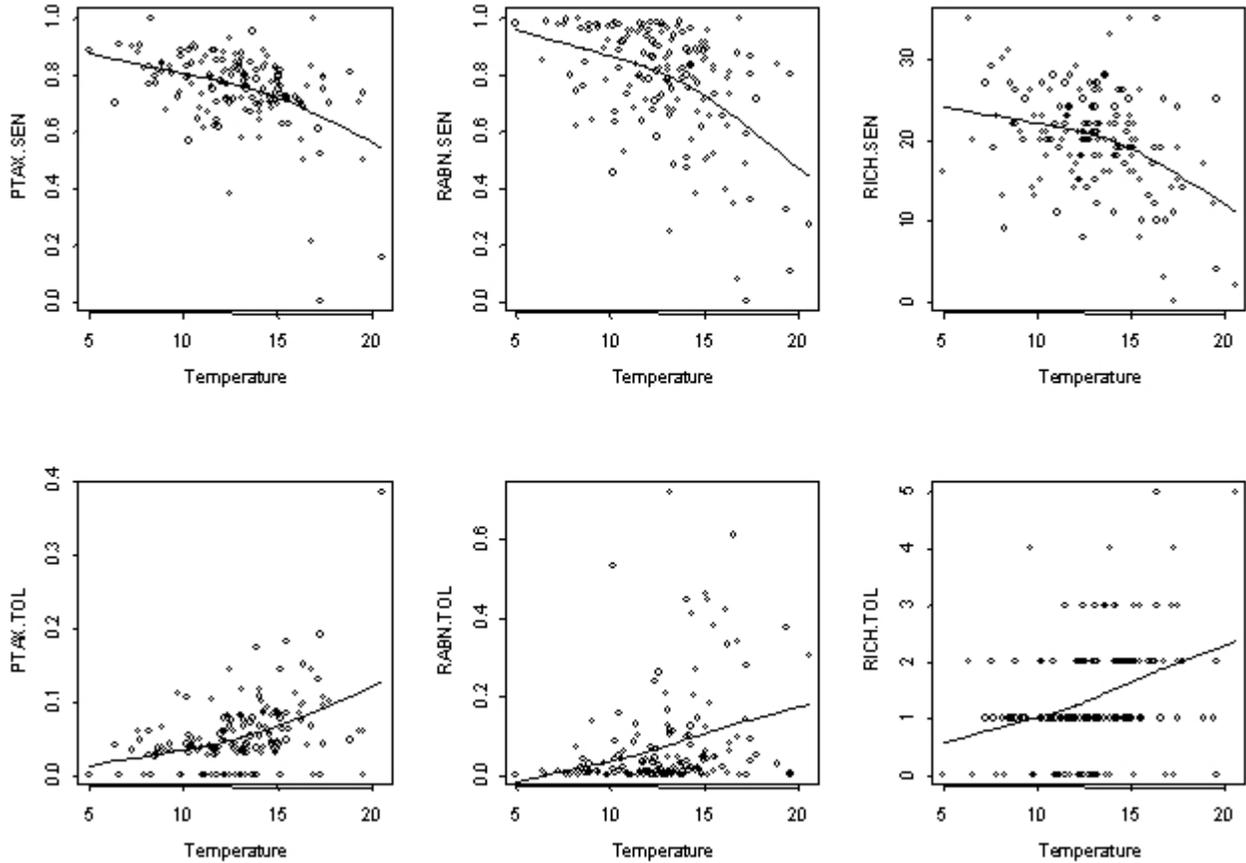
We can test for the specificity of response for different tolerance value metrics by examining their relationship with other environmental gradients. Ideally, a tolerance value metric would exhibit a strong relationship with the stressor for which it was designed and would show no relationship with other gradients. Examples of these comparisons are shown in Figures 17 and 18, where temperature compositional metrics are plotted against the sediment gradient and sediment compositional metrics are plotted against the temperature gradient. Most temperature metrics did not exhibit strong relationships with the sediment gradient. The relative abundance, richness, and proportion of total taxa of temperature-sensitive taxa decreased with increasing fine sediment, but these relationships were less strong and more variable than those

observed between the same metric and stream temperature (Figure 15). Metrics derived from temperature-tolerant taxa all showed no response to the sediment gradient.

Sediment tolerance metrics were less specific to the sediment gradient, as all metrics exhibited a relationship with the temperature gradient. Sediment-sensitive metrics all initially decreased gradually with increasing temperature but then decreased at a faster rate at higher temperatures. Sediment-tolerant metrics all increased with increasing temperature, and  $R^2$  values for these relationships were even higher than those observed for fine sediment.



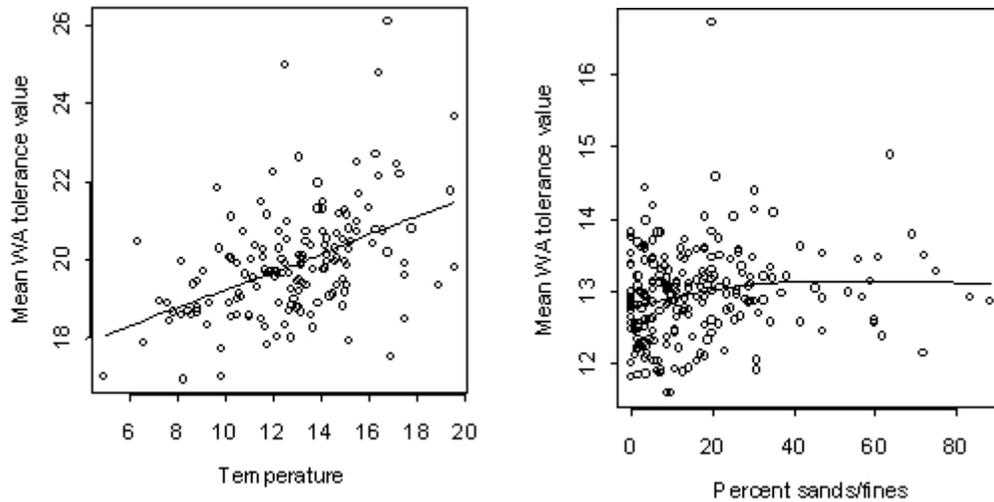
**Figure 17. Temperature tolerance metrics plotted versus percent sands and fines.** Vertical axis abbreviations and symbols as in Figure 15.



**Figure 18. Sediment tolerance metrics plotted versus temperature.** Vertical axis abbreviations and symbols as in Figure 15.

We can perform the same type of comparison using mean tolerance values. In Figure 19, mean WA tolerance values for fine sediment are plotted versus observed stream temperature, and mean WA tolerance values for temperature are plotted versus observed fine sediment. Here, mean sediment tolerance values showed a fairly strong relationship with temperature, whereas mean temperature tolerance values were specific to stream temperature and were not associated with fine sediment.

Results from this exercise suggest that the sediment tolerance classifications are less specific to sediment than the temperature tolerance classifications are to temperature. However, both performed reasonably well, as most metrics and inference indices explained more variability in the stressor of interest than in the covarying stressor. The notable exception to this trend was the sediment-tolerant taxa metrics, and their poor performance can likely be attributed to the distribution of observed sediment values and perhaps to the low richness of sediment-tolerant taxa observed at any given site (see Section 4.1.1).



**Figure 19. Mean weighted average (WA) tolerance values for sediment plotted against observed temperature (left), and mean WA tolerance value for temperature plotted against observed sediment (right).**

Generalizing the present findings to a broader array of stressors is challenging. Covariance in stressors is common, however, and in many cases tolerance values will provide only one line of evidence toward a diagnosis of the most likely cause of impairment. Other lines of evidence will often be needed to supplement the information provided by tolerance values.

#### **4.1.3.4. *Effects of Taxonomic Resolution of Tolerance Values***

Tolerance values for the examples considered thus far have been computed at the genus level. In some monitoring programs, taxa are identified only to the family level. For certain applications, this level of identification may be sufficient; however, as taxonomic resolution coarsens, we can less confidently assume that an invariant fundamental niche is associated with each taxon. A single family may contain numerous genera, each of which may contain numerous species and all of which may have slightly different niches. Conversely, for certain families, we may expect to find only one or two species, and in these cases family-level identification may be sufficient. Furthermore, aggregation to genus or even family is often necessary because of limitations on taxonomic identification accuracy.

To explore the effects of taxonomic resolution, we computed tolerance values in EMAP-West at the family level and used these values to compute average tolerance values in Oregon. The  $R^2$  values for comparisons between average tolerance values and environmental observations are shown in Table 9.

**Table 9. R<sup>2</sup> values for family-level tolerance values**

<b>Method</b>	<b>Temperature</b>	<b>Sediment</b>
WA	0.44	0.34
CD75	0.42	0.34
GLMMAX	0.28	0.30
GAMMAX	0.29	0.31

As expected, a substantial deterioration in predicted accuracy is observed, as R<sup>2</sup> decreased for virtually all methods (compare with values in Table 7). A similar deterioration in performance was observed when tolerance values compositional metrics were computed using family-level identification. These results suggest that for the stressors considered here, taxonomic identification to genus level provides indicators that are more closely associated with observed stressor levels.

#### **4.2. REFERENCE CONDITIONS**

Reference conditions refer to streams or sites that are in their natural condition, unaffected by human activities. These conditions establish the basis of comparison from which we can evaluate the biotic condition of other streams (Bailey et al., 1998). Reference conditions are often approximated within a study area using the “best” sites that are available, that is, sites that are the least disturbed by human activities. Biological characteristics of assemblages in these least-disturbed streams provide baseline expectations, and departures from these expectations are often associated with human activities. With tolerance values, we can infer stressor-specific characteristics of streams; however, we still need to know expected values for these characteristics at the least-disturbed sites within the study area to determine whether changes have occurred. The values of tolerance metrics and inference indices computed at least-impacted sites determine this baseline distribution. Departures from this distribution then can provide evidence for the existence of a particular stressor.

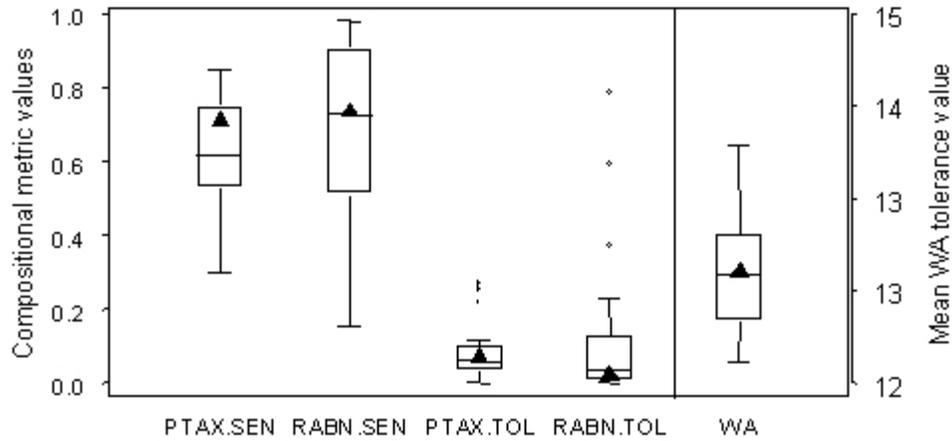
To demonstrate one approach for applying baseline conditions, we used 31 reference sites in western Oregon, supplied by the Oregon DEQ (ODEQ, 2004). These sites were used to compute reference distributions for all of the metrics that we have considered (RABN.TOL, RABN.SEN, PTAX.TOL, PTAX.SEN) for temperature and fine sediment. We also used mean WA tolerance values for sediment and temperature at reference sites. We then examined data from a single, randomly selected site at which biological data had been collected but for which other environmental data were not available. At this test site, we computed metric values for

temperature and sediment and compared them with reference distributions (Figure 20). Test site observations of temperature metric values were all located close to the median value observed within reference sites. In contrast, test site observations of metric values for sediment departed from reference distributions. More specifically, we observed that PTAX.SEN and RABN.SEN were less than reference expectations, whereas PTAX.TOL, RABN.TOL, and WA were all greater than reference expectations.

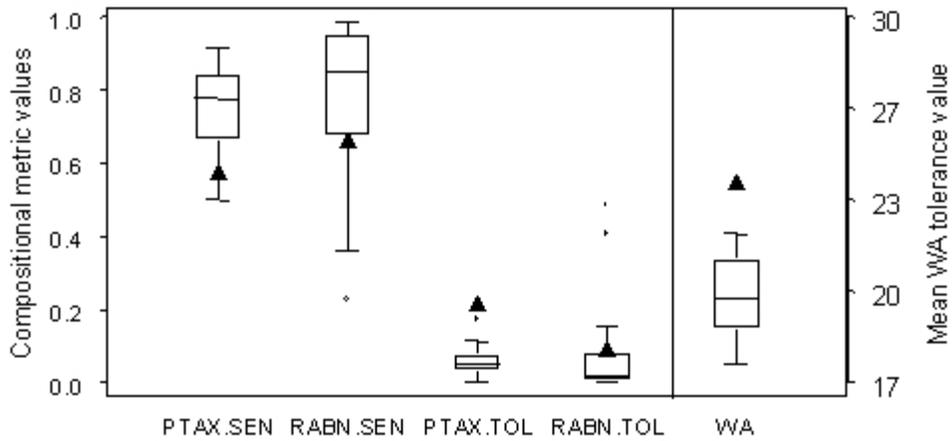
All of these indicators consistently suggest that sediment levels at the test site are elevated and are influencing the macroinvertebrate assemblage. Sediment metrics also can indicate an increase in temperature (as seen in Section 4.1.3.3), but in this case, the complete absence of a change in the temperature-specific metrics seems to suggest that no change in temperature has occurred. A follow-up visit to this site would be necessary to confirm whether these inferences are accurate.

The results are somewhat surprising, given that the method was relatively crude. Note though that this test involved only a single sample, and more samples are required to draw substantive conclusions. Refining the descriptions of least-disturbed sites by grouping biologically similar sites (Clarke et al., 2002) may further reduce the variability of the reference distributions and improve the discriminatory power of tolerance metrics and indices. Also, in cases where the presence of more than one stressor is indicated, we may have to examine the magnitude of the effect of each stressor to identify the stressor that is responsible for the greatest proportion of degradation.

### Temperature metrics



### Sediment metrics



**Figure 20. Comparison of temperature (top) and sediment (bottom) tolerance metrics at a single test site (shown as a solid triangle) with reference distributions (shown as box and whisker plots).** Horizontal lines in the boxes correspond to the 25<sup>th</sup>, 50<sup>th</sup>, and 75<sup>th</sup> percentiles of the reference distribution. Whiskers extend away from the boxes a distance equal to two times the distance between the 25<sup>th</sup> and 75<sup>th</sup> percentiles. Open circles represent samples falling outside the range defined by the whiskers.

## **5. AREAS OF UNCERTAINTY AND RESEARCH PRIORITIES**

On the basis of the results shown in the previous section, stressor-specific tolerance values appear to be a promising tool for monitoring the condition of streams and informing diagnoses of the causes of impairment in streams. Broader application of this approach requires the resolution of a few key areas of uncertainty: causal links between taxon occurrences and environmental gradients must be established for tolerance values, stressor gradients of interest must be clearly defined and measured, and basic ecological questions regarding the response of stream communities to anthropogenic stress must be addressed.

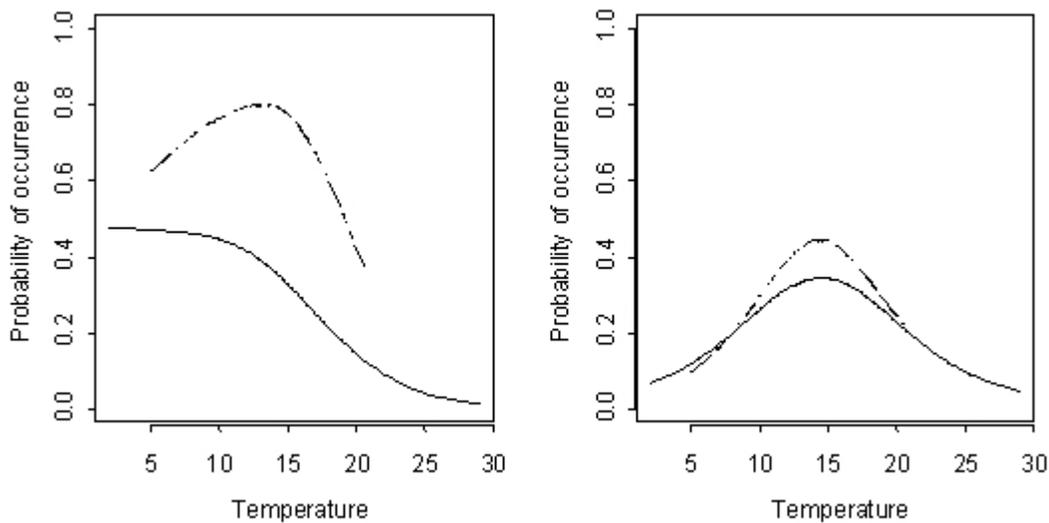
### **5.1. CAUSAL RELATIONSHIPS**

The tolerance values described in this report were all estimated from field data and therefore are subject to the limitations imposed by such data. That is, the analysis of field data provides only correlative relationships and does not necessarily provide any evidence of causation. Thus, a tolerance value defined in terms of one stressor may, in fact, be an indicator for a different, covarying stressor or natural gradient. Covariation among stressors is common. The two stressors discussed extensively in this report—elevated stream temperature and increased fine bedded sediment—can be strongly correlated when they originate from the same human activity. For example, logging removes vegetative cover, which increases sediment loading and also reduces shading to the stream (thus increasing stream temperature). Furthermore, both temperature and fine sediment fraction increase naturally as high-gradient mountain streams transition to low-gradient valley-bottom streams. Tolerance values can be applied most effectively if they are specific to the stressor of interest; therefore, the effects of covariation of variables must be controlled to whatever extent possible.

We briefly addressed the issue of correlated stressors in the previous section by examining the specificity of the derived relationships and plotting relationships between temperature tolerance metrics and observed fine sediment and between sediment tolerance metrics and observed stream temperature. Results were mixed, but both sediment and temperature tolerance metrics exhibited some strong relationships with the correlated stressor gradient. Uncertainties regarding the causal relationships underlying some of the temperature and sediment tolerance values do exist. Furthermore, these tests were limited in scope. It is very possible that other, unsampled environmental gradients also could be correlated with the derived tolerance metrics.

How then do we increase our confidence in the relationship between tolerance values and the stressors for which they were derived? One approach is to compare taxon-environment relationships across different data sets and different regions. We would expect environmental

gradients in different regions to covary in different ways. Furthermore, if a particular taxon-environment relationship truly represented the fundamental niche for a given taxon, then we would expect it to remain similar regardless of location. Thus, if a given taxon-environment relationship remained similar across different regions, we could be more confident that the estimated taxon-environment relationship reflected an actual causal relationship. An example of such a comparison is shown in Figure 21, where the relationship between probability of occurrence of two genera, *Glossosoma* and *Malenka*, and stream temperature are plotted. Two relationships for each genera are shown, one derived using EMAP-West data and one derived using Oregon data. The shapes of the relationships for *Glossosoma* appear different: probability of occurrence reached a maximum in Oregon at about 14°C, whereas it plateaued in EMAP-West at 10°C and remained high for lower temperatures. However, occurrence probabilities in both regions decreased as temperatures increased above 14°C. Taxon-environment relationships for *Malenka* were very similar. Both peaked at approximately 15°C and decreased on either side of the peak.



**Figure 21. Comparison of taxon-environment relationships for *Glossosoma* and *Malenka*.** Dashed line is relationship computed using Oregon data. Solid line computed using data from EMAP-West.

These two examples highlight some of the difficulties inherent in comparing taxon-environment relationships. First, the maximum probability of occurrence of a given taxon is likely to be different across different regions, so taxon-environment relationships will be shifted up or down on the plot relative to one another. Second, the range and distribution of observations differ across regions, so we must attribute different levels of confidence to the different parts of each curve. Further work is required to develop methods for more quantitative comparisons. For now though, qualitative comparisons can provide some evidence of a common

and causal response to a given gradient. For the cases shown in Figure 21, we would probably attribute more confidence to a causal relationship between *Malenka* and temperature than between *Glossosoma* and temperature.

As discussed in Section 3.3.4, the absolute values of tolerance values depend on the data set from which they were derived, and comparing tolerance values across regions probably will not yield insights into the strength of causal relationships. However, if tolerance values are summarized using biological metrics and the relationships of these metrics with gradients in a different region are examined (as with the examples shown in Section 4), these analyses can provide further evidence of causality. Again, the reasoning here is that the covariance structure of stressors differs across regions, and if tolerance metrics and inference indices derived using data from one region are strongly correlated with the stressor of interest in a second region, then it is more likely that the tolerance values reflect a true causal relationship.

Another approach for increasing the confidence assigned to correlative relationships is to attempt to control for other possible sources of covariance by including additional variables in regression models. The main weakness of this approach is that we cannot be sure that all appropriate gradients are included or even that included gradients are appropriately modeled. Furthermore, including some natural gradients may be counterproductive. For example, the size of a stream is known to be strongly correlated with the type of organisms that are observed. However, it is unlikely that stream size is the actual causal agent. Instead, stream size is correlated with factors such as stream temperature and flow regime that directly influence the abundance of different taxa. Thus, including some natural variables in our models can actually impede our ability to accurately define taxon-environment relationships. Completely excluding natural covariates is not a satisfactory solution either. If, for example, the stressor of interest is stream temperature, then including stream size may partially control for the effects of flow regime, which could be useful.

After covariates are included in models, the resulting taxon-environment relationship for the stressor of interest can only be interpreted as being conditional on fixed values of the other covariates. Thus, tolerance values based on these taxon-environment relationships are also conditional on the values of the other covariates. This additional level of complexity in defining tolerance values must be explored in detail and is beyond the scope of this report. In the analyses presented here, we have used only a single explanatory variable in our models.

The strongest approach for increasing our confidence in causal relationships is to use controlled experiments because responses determined through experiment provide powerful evidence of causality. Laboratory experiments, however, are limited in the types of stressors that can be tested and in the types of organisms that can be bred and they do not often replicate the complexity of natural ecosystems. Thus, laboratory dosing tests can reveal that different

macroinvertebrates undergo increased mortality at increased stress and show the different physiological mechanisms that determine whether a particular species is tolerant or sensitive, but these findings do not account for more complex ecological mechanisms (e.g., increased predation, decreased competitive ability) that affect organism survival. A wealth of species-specific sensitivity data that were developed from traditional laboratory toxicity tests are available, and there is a great potential for using these data to inform and complement field-derived tolerance values. However, more research is required to compare and contrast results derived from laboratory experiments and those derived from field studies. Recent work combining microcosm and field experiments have shown potential for bridging these gaps (Clements, 2004).

## **5.2. DEFINING STRESSOR GRADIENTS**

Another necessary step for broadening the applicability of tolerance values is developing tolerance values for all of the stressors that are commonly observed in streams. In this report, we have focused on elevated stream temperature and increased percentages of fine sediments as examples of stressors. Previous efforts to develop tolerance values (e.g., Hilsenhoff, 1987) have focused on nutrient enrichment and organic pollution gradients. Good data sets exist for some stressors (e.g., increased metals and acidification), and the derivation of tolerance values using the methods described in this report should be straightforward. More research will be required to appropriately define and sample other common stressors before tolerance values can be developed.

### **5.2.1. General-disturbance Gradients**

The framework for the biological condition gradient (Davies, 2001) defines a set of narrative descriptions of biological changes in response to a general human-disturbance gradient. The response of sensitive and tolerant taxa figures prominently in this framework, but identifying tolerant or sensitive taxa is hampered by the ambiguity inherent in defining a general human-disturbance gradient. In this report, we have focused on stressor gradients that can be readily quantified by field measurements. Unfortunately, no obvious environmental attribute or set of attributes exists that can accurately define a general human-disturbance gradient. Various proxy measures have been proposed, but all have important issues that must be resolved before they can be applied to tolerance values. It has been suggested, for example, that a land use metric (e.g., percent urban land use in the catchment) provides a good proxy for the human-disturbance gradient; indeed, in many cases, urban land use has served as a useful proxy for disturbance (e.g., Wang et al., 2001). However, the management decisions we hope to inform with tolerance values are often concerned with directing future remediation efforts. If tolerance values are

defined in terms of an urban gradient, then the only logical remediation or mitigation scheme would be removal of the urban land use, an option that is rarely feasible. Thus, defining tolerance values with regard to a land use gradient would not necessarily provide useful suggestions for management. Moreover, the proxy measures of a human-disturbance gradient would change between regions with different sets of dominant stressors; thus, no consistent definition of the gradient or of the associated tolerance values would be possible.

An analytical approach for defining a general disturbance gradient would be to first apply an ordination analysis (e.g., principle components analysis) to stressor data collected from the study area and then estimate tolerance values with respect to the primary axis of the ordination. For this exercise, the use of randomly collected data would be critical to accurately capture the relative frequency with which different anthropogenic stressors affect streams within the study area. The primary axis would then presumably capture a general disturbance gradient, and tolerance values estimated with respect to this gradient would represent generally tolerant and generally sensitive taxa for the study area. This approach would likely under-represent human activities that may cause severe degradation but are limited in spatial extent. Furthermore, natural gradients (e.g., elevation or catchment area) may confound the disturbance gradient described by an ordination approach (as discussed in Section 5.1), so care should be taken to select samples that minimize natural differences between streams.

An alternative to attempting to define a general disturbance gradient is the use of tolerance values that do not require an explicit definition of a gradient. This type of approach has not been described in this report because it differs fundamentally from methods based on taxon-environment relationships. One such approach relies on predictive models (Hawkins, 2003) and quantifies tolerance values from differences between an observed and a modeled frequency of occurrence for a taxon. Thus, the relative tolerance value for a given taxon is based only on how often the taxon is observed across an anthropogenically influenced study area relative to reference expectations. If the tolerance values are derived using data that are collected in a randomized design in a particular region, they could provide an accurate depiction of taxa that increase and taxa that decrease in response to the dominant human activities in a region. In both the predictive modeling and the ordination axis approaches, the general-disturbance tolerance values derived would be unique to a particular region. This restriction seems reasonable, given that dominant stressors differ between different regions. In contrast, the relative rankings of tolerance and sensitivity derived with respect to individual stressor gradients would be expected to be more stable between regions.

### **5.2.2. Indirect Stressor Gradients**

The models described in this report apply most effectively to environmental gradients that directly influence the health and fecundity of a particular organism. These direct gradients would include factors such as the temperature and acidity of the stream. Many other factors can influence species survival but are difficult to measure directly. Usable physical habitat, for example, is a requirement for all organisms, but a complete understanding of how different stream macroinvertebrates utilize physical features in the stream is still lacking. We must therefore resort to indirect measures to quantify many environmental gradients. The indirect gradients for physical habitat may include measures of the percentage of fine sediment on the stream bottom (Wood and Armitage, 1997) or measures of the quantity and distribution of large woody debris (e.g., Lakly and McArthur, 2000). Another important indirect gradient is nutrient concentration. High concentrations of nutrients generally do not directly influence macroinvertebrates; rather, they alter the trophic status of a stream, placing certain insects at a competitive disadvantage (Biggs, 2000). Taxon-environment relationships can be derived with respect to any measured gradient, but the precision of the relationships, as well as their applicability beyond the region in which they were derived, become more uncertain as the mechanisms become less direct.

Resolving these issues requires that we improve our understanding of the mechanisms by which different anthropogenic stressors affect stream biota. Tolerance values can then be defined in terms of the appropriate proximal factor. With nutrients, for example, the optimal solution may be to estimate tolerance values for different periphyton taxa instead of for macroinvertebrates, because periphyton respond directly to nutrient concentrations in streams.

### **5.2.3. Sampling Issues**

The number of stressors for which tolerance values can currently be defined is also limited by the types of data that are readily available. Standard protocols for sampling streams often rely on randomized sampling of riffles and a one-time visit to each stream. These protocols constrain the types of stressors for which tolerance values can be derived. Randomized sampling, while providing a statistically valid sample from which population statistics can be estimated, does not always provide the best data for inferring taxon-environment relationships because stressor gradients are also sampled in proportion to their distribution in the region. This distribution would be suitable for computing tolerance values for some widely distributed stressors. For other stressors that are not as prevalent, a randomized sampling design does not provide enough samples at high levels of stress to accurately compute tolerance values. The distribution of sediment measurements in the EMAP-West data (Section 4.1.3.1) provides a good illustration of this phenomenon. Other sampling designs should be considered for such stressors.

Riffles often contain the greatest density and diversity of macroinvertebrates, but sampling of riffles can reduce our ability to discern the effects of stressors (such as sedimentation) that initially would affect depositional areas in the stream. Depending on the stressor of interest, different habitats may be more appropriate to sample. Transect designs, in which random locations are sampled along prespecified reach transects, can also address this issue.

A one-time visit provides efficient sampling of a broad array of sites, but may not provide information on stressors that have high temporal variability. Pulse additions of sediment, toxicants, and nutrients are poorly quantified by a single visit. Understanding of such stressors would be improved if sampled in temporally resolved studies. Other environmental factors also exhibit high variability but can be effectively measured with a single grab sample as long as the total number of samples is relatively high. For example, stream temperature varies considerably on seasonal and daily cycles, so a single grab sample of temperature does not necessarily provide useful information for a single stream. However, if one considers grab temperature measurements from a large sample of streams, then the seasonal and daily variations often are manifested as random noise, and useful patterns of taxon responses to temperature can be extracted.

There is clearly much to be gained by analyzing existing data sets. However, directed studies toward specific stressors, in which the samples are collected along the full gradient, appropriate habitats are sampled, and temporal resolution of samples designed to accurately characterize the stressor, would greatly enhance existing understanding of a broader array of anthropogenic stressors. Other existing data sets collected with different designs and protocols (e.g., Superfund) can potentially be mined. Controlled manipulations, either in the lab or in the field, can supplement knowledge derived from field data, especially in cases where a well-designed field sampling campaign to capture the same stressors is too costly.

### **5.3. FUNDAMENTAL ECOLOGY**

The derivation and application of tolerance values are closely linked with a fundamental understanding of stream ecology. Many uncertainties regarding tolerance values are also areas of active research in stream ecology. Many examples of related basic ecological questions exist, but here we focus only on two main areas: the role of biological interactions and the temporal characteristics of stream ecosystems.

#### **5.3.1. Biological Interactions**

To derive tolerance values that are applicable beyond the immediate study area (see Section 2.1), we need to assume that biological interactions are relatively unimportant and that realized niches are similar to fundamental niches. These assumptions are broad and are likely

violated in many cases. Different stream types will have different strengths of biological interactions. For example, we would expect communities in streams with stable flow regimes to be more strongly influenced by biological interactions than are communities in streams that are disturbed by frequent floods (Allan, 1995). We would also expect certain taxa to be more subject to biological interactions than others. A vast body of ecological research in both terrestrial and aquatic systems exists that examines these effects (e.g., Ives, 1995). However, for the practical purposes of applying tolerance values, we need specific guidelines on the types of streams and the types of taxa for which biological interactions are likely to be important. Such specific data are not widely available.

### **5.3.2. Temporal Characteristics of Stream Communities**

Disturbance and recovery from disturbance are key factors in the structure of stream communities. Stream communities are dynamic, constantly changing biological systems. Most streams are exposed to strong, natural disturbances, and the stream community observed at any point in time is a manifestation of the trajectory of recovery from the last major disturbance (e.g., flood). Other temporally influenced factors include natural cycles of hatching and emergence and episodic anthropogenic disturbances, such as toxicant pulses. With regard to tolerance values, we need to better understand the extent to which a single snapshot of a stream community reflects a response to the environmental gradient of interest and the extent to which it reflects temporally evolving factors. Other issues to be addressed include the extent to which the conditions observed are a consequence of the initial conditions and zoogeographical history of the stream and to what extent they are a consequence of legacy effects of past human activities. All of these issues can potentially influence the generality of tolerance values derived from a particular study area, and further research is required to identify scenarios in which temporally evolving factors play an important role.

## **6. IMPLEMENTATION**

As an assessment tool, tolerance values are applicable to many different aspects of water quality management. They can help improve assessments of the condition of streams, identify the sources of impairment for effective preparation of total daily maximum loads, and direct and monitor restoration activities. Our intent in preparing this report was first to present and review the methods for deriving tolerance values so that states and tribes can derive tolerance values for use in their own programs. Second, our hope is that this report will lay the foundations for the development of a national or several regional databases of tolerance values that can be used

directly by states and tribes that may not have the resources to derive tolerance values from their own data. For both of these objectives, two implementation issues must be addressed: information management and taxonomic quality assurance.

## **6.1. INFORMATION MANAGEMENT**

The applicability and utility of tolerance values would be vastly enhanced if analyses conducted in different study areas were conveniently accessible. A repository of tolerance values analyses would provide an opportunity for those developing their own tolerance values to compare their results with others from similar locations. These comparisons would also help accumulate evidence of causality, as described in Section 5.1. Finally, a database of tolerance values (e.g., see Appendix C) could potentially be applied directly to assessment questions by entities that do not have the resources to develop their own tolerance values. However, as discussed earlier, a database of tolerance values may not be the most useful format for assessments because absolute tolerance values vary with regional characteristics. The tolerance values listed in Appendix C should be regarded as specific to the western United States and to the range of conditions sampled in the data set. Storing information for the entire taxon-environment relationship derived from different analyses would provide data that are more easily compared between locations. Then, when tolerance values are required for assessment, the stored taxon-environment relationships can be further processed.

A more ambitious vision of regional databases would involve storing and sharing the raw data collected by different states, tribes, or regions. Pooling raw data collected by different entities could lengthen the sampled environmental gradients of interest and increase sample sizes, potentially increasing the accuracy of derived taxon-environment relationships. Combining raw data however, requires that sampling protocols be compatible, and states differ in the types of habitats they sample, the sampling intensity within those habitats, and the number of individuals that they identify from each sample and in many other details of sampling procedures. These differences must be reconciled before raw data can be effectively combined. Methods exist for quantifying sampling differences (e.g., Cao et al., 2002) and reconciling different data sets, so these issues are tractable. Providing a regional repository for raw data may be less immediately useful to state and tribal water quality programs.

## **6.2. TAXONOMIC QUALITY ASSURANCE**

With tolerance values, assessments of site conditions can hinge on the presence or absence of particular taxa, so ensuring the accuracy of taxonomic identifications is critically important. Taxonomists are needed to catalogue the diversity of species in studies of biodiversity (Blackmore, 2002). Unfortunately, fewer people are choosing taxonomy as field of study and

there is a dearth of qualified taxonomists. A certification program for taxonomy could help reinvigorate the field by raising the stature of the profession and establishing minimal levels of expertise for taxonomists involved in biological assessment.

## 7. CONCLUSIONS AND RECOMMENDATIONS

Macroinvertebrate tolerance values have great potential for improving our ability to effectively manage the waters of the United States. Tolerance values derive from fundamental concepts of ecology and effectively represent variations in the relationships between different species and environmental gradients of interest. In our review of methods for deriving and applying tolerance values, we found that WA provides one of the simplest, most robust approaches for estimating the relative sensitivity of different taxa. Weighted averages are therefore recommended as a first estimate of the tolerance values for different taxa.

Other compositional metrics (e.g., relative abundance of sensitive taxa) provide additional information regarding the effects of different stressors on stream assemblages and can enhance diagnoses of the causes of impairment. However, these metrics require that differences between sensitive and tolerant taxa be defined more clearly, and for this purpose, modeling the taxon-environment relationship with GLMs or GAMs provides an invaluable tool. After taxon-environment relationships are defined by regression models, different taxa can be categorized as tolerant or sensitive on the basis of curve classification techniques, and then different compositional metrics can be computed. An added benefit of explicitly modeling the taxon-environment relationship is that the strength of the relationship between a particular taxon and a particular environmental gradient can be quantified. Taxa that are only weakly associated with the gradient of interest can then be excluded from use in metrics.

Tolerance values at both the family and the genus level were examined. Genus-level values accounted for more variability in environmental observations; however, for certain metrics, family-level identifications sufficed, and in all cases, family identifications still provided useful information. We therefore recommend that organisms be identified to the finest taxonomic level possible, but note that tolerance values can provide a useful tool for data collected at coarser taxonomic resolutions.

Comparing the values of tolerance values metrics at test sites with baseline conditions at least-impaired sites is a critical step for the effective application of tolerance values. Without baseline conditions, it is difficult to establish expected values for tolerance value metrics and therefore difficult to determine whether changes have occurred.

To achieve the potential of tolerance values, more research is required to resolve specific uncertainties regarding the importance of biological interactions and to expand our knowledge base to all stressors that are important in streams. Research priorities include comparing and contrasting species sensitivities derived from lab and field studies, mining existing databases to derive tolerance value for other stressor gradients, and performing directed field studies for stressors for which data do not currently exist.

On the practical side, the utility of tolerance values would be broadened if databases of taxon-environment relationships were conveniently accessible, along with tools for deriving different tolerance values from those taxon-environment relationships. Furthermore, a national certification program for taxonomy would greatly strengthen the taxonomic basis for routinely applying tolerance values in water quality management.

## REFERENCES

- Allan, JD. (1995) Stream ecology: structure and function of running waters. London, UK: Chapman and Hall.
- Armitage, PD; Moss, D; Wright, JF; et al. (1983) The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Res* 17:333–347.
- Austin, MP; Meyers, JA. (1996) Current approaches to modeling the environmental niche of eucalyptus: implication for management of forest biodiversity. *Forest Ecology and Management* 85:95–106.
- Bailey, RC; Kennedy, MG; Dervish, MZ; et al. (1998) Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. *Freshwater Biology* 39(4):765–774.
- Barbour, MT; Gerritsen, J; Snyder, BD; et al. (1999) Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. 2nd edition. EPA-841-B-99-002. U.S. Environmental Protection Agency, Washington, DC.
- Biggs, B.J.F. (2000) Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae. *Journal of the North American Benthological Society* 19:17–31.
- Birks, HJB; Line, JM; Juggins, S; et al. (1990) Diatoms and pH reconstruction. *Philos Trans R Soc Lond B Biol Sci* 327(1240):263–278.
- Blackmore, S. (2002) Biodiversity update-progress in taxonomy. *Science* 298:365.
- Cao, Y; Williams, DD; Larsen, DP. (2002) Comparison of ecological communities: the problem of sample representativeness. *Ecological Monographs* 72:41–56.
- Chessman, BC; McEvoy, PK. (1998) Toward diagnostic biotic indices for river macroinvertebrates. *Hydrobiologia* 364:169–182.

- Chutter, FM. (1972) An empirical biotic index of the quality of water in South African streams and rivers. *Water Res* 6:19–30.
- Clarke, RT; Wright, JF; Furse, MT. (2002) RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling* 160(3):219–233.
- Clements, WH. (2004) Small-scale experiments support causal relationships between metal contamination and macroinvertebrate responses. *Ecological Applications* 14(3): 954-967.
- Davies, SP. (2001) Characterizing biological condition categories across gradients of human disturbance. Presented at 49th Annual Meeting of the North American Benthological Society; June 3–8; LaCrosse, WI.
- Hastie, TJ; Pregibon, D. (1992) Generalized linear models. In: Chambers, JM; Hastie, TJ. *Statistical Models in S*, Boca Raton, FL: Chapman & Hall/CRC.
- Hastie, TJ; Tibshirani RJ. (1999) *Generalized additive models*. Washington, DC: Chapman & Hall/CRC.
- Hawkins, CP. (2003) Development of quantitative tolerance values based on predictive model outputs. Presented at 51st Annual Meeting of the North American Benthological Society; May 27-June 1; Athens, GA.
- Hilsenhoff, WL. (1987) An improved biotic index of organic stream pollution. *The Great Lakes Entomologist* 20:31–37.
- Hosmer, DA; Lemshow, S. (2000) *Applied Logistic Regression*. 2<sup>nd</sup> Ed. John Wiley & Sons. New York.
- Ives, AR. (1995) Predicting the response of populations to environmental change. *Ecology* 76(3):926–941.
- Klemm, DJ; Blocksom, KA; Thoeny, WT; et al. (2002) Methods development and use of macroinvertebrates as indicators of ecological conditions for streams in the Mid-Atlantic Highlands region. *Environ Monit Assess* 78:169–212.
- Lakly, MB; McArthur, JV. (2000) Macroinvertebrate recovery of a post-thermal stream: habitat structure and biotic function. *Ecological Engineering* 15:S87–S100.
- Lenat, DR. (1993) A biotic index for the southeastern United States: derivation and list of tolerance values, with criteria for assigning water-quality ratings. *Journal of the North American Benthological Society* 12:279–290.
- Manel, BC; Williams, HC; Ormerod, SJ. (2001) Evaluating presence-absence models in ecology. *Journal of Applied Ecology* 38:921–931.
- Norton, SB. (2000) Can biological assessments discriminate among types of stress? A case study from the eastern cornbelt plains ecoregion. *Environ Toxicol Chem* 19:1113–1119.
- ODEQ (Oregon Department of Environmental Quality) (2004) Selecting reference condition sites: an approach for biological criteria and watershed assessment. Technical Report WAS04-002.
- Odum, EP. (1971) *Fundamentals of ecology*. Philadelphia, PA: W.B. Saunders.
- Oksanen, J; Minchin, PR. (2002) Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling* 157:119–129.
- Relyea, CD; Minshall, GW; Danehy, RJ. (2000) Stream insects as bioindicators of fine sediment. Presented at Watershed Management 2000 Conference; June 21–24; Fort Collins, CO. Water Environment Federation, Alexandria, VA.

- Slooff, W. (1983) Benthic macroinvertebrates and water quality assessment: some toxicological considerations. *Aquat Toxicol* 4:73–82.
- ter Braak, CJF; Juggins, S. (1993) Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. *Hydrobiologia* 269/270:485–502.
- ter Braak, CJF; Looman, CWN. (1986) Weighted averaging, logistic regression and the Gaussian response model. *Vegetatio* 65:3–11.
- Wang, L; Lyons, J; Kanehl, P. (2001) Impacts of urbanization on stream habitat and fish across multiple spatial scales. *Environ Manage* 28(2):255–256.
- White, GC; Bennetts, RE. (1996) Analysis of frequency count data using the negative binomial distribution, *Ecology* 77(8):2549–2557.
- Wood, PJ; Armitage, PD. (1997) Biological effects of fine sediment in the lotic environment. *Environ Manage* 21(2):203–217.
- Yuan, LL. (2004) Assigning macroinvertebrate tolerance classifications using generalised additive models. *Freshwater Biology* 49:662–677.
- Yuan, LL; Norton, SB. (2003) Comparing responses of macroinvertebrate metrics to increasing stress. *Journal of the North American Benthological Society* 22(2):308–322.

**APPENDIX A: ATTENDEES AT THE WESTERN TOLERANCE VALUES WORKSHOP, CORVALLIS, OREGON (FEBRUARY 3–5, 2004)**

<b><u>Name</u></b>	<b><u>Organization</u></b>
Steve Austin	Navajo Environmental Protection Agency
Robert Bennetts	Greater Yellowstone Network
Wease Bollman	Rhithron Associates, Inc.
Darren Brandt	Idaho Department of Environmental Quality
Daren Carlisle	U.S. Geological Survey
Bruce Chessman	Center for Natural Resources, New South Wales Department of Infrastructure, Planning and Natural Resources
Will Clements	Colorado State University
Bob Danehy	Weyerhaeuser
Doug Drake	Oregon Department of Environmental Quality
Dave Feldman	Montana Department of Environmental Quality
Leska Fore	Statistical Design
Jeroen Gerritsen	TetraTech
Rick Hafele	Oregon Department of Environmental Quality
Jim Harrington	California Game and Fish
Chuck Hawkins	Utah State University
Lil Herger	U.S. EPA Region 10
Alan Herlihy	Oregon State University
Shannon Hubler	Oregon Department of Environmental Quality
Dave Huff	Oregon Department of Environmental Quality
Susan Jackson	U.S. EPA Office of Water
Jerry Jacobi	New Mexico Department of Environmental Quality
Phil Kauffman	U.S. EPA Office of Research and Development
Jeff Kershner	U.S. Forest Service
Tina Laidlaw	U.S. EPA Region 8
Phil Larsen	U.S. EPA Office of Research and Development
Dave Lenat	North Carolina Department of Environment and Natural Resources
Gary Lester	EcoAnalysts
Amanda Mays	Council of State Governments

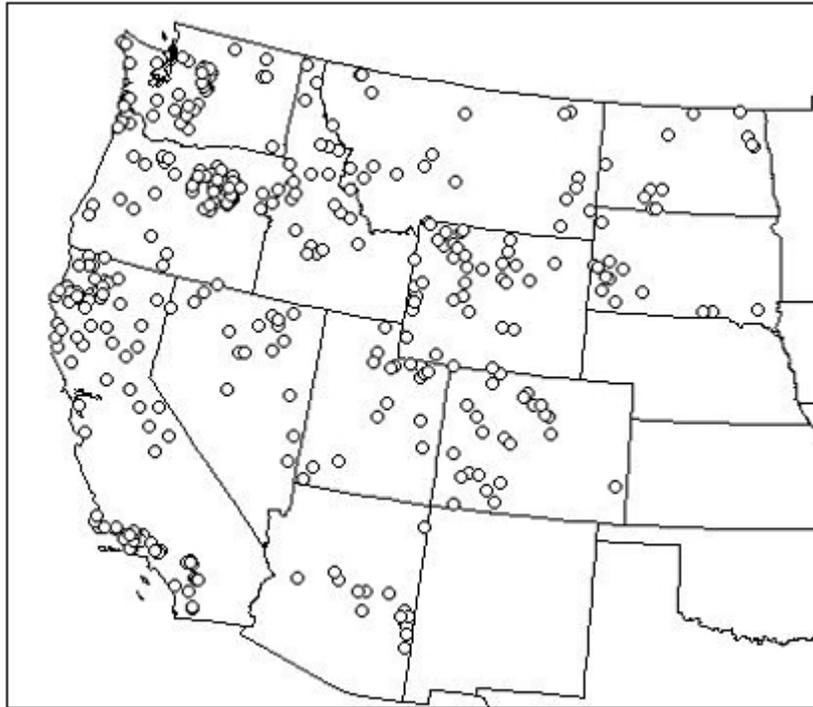
---

<u>Name</u>	<u>Organization</u>
Sue Norton	U.S. EPA Office of Research and Development
Pete Ode	California Game and Fish
Yandong Pan	Portland State University
Mike Paul	Howard University
Rob Plotnikoff	Washington Department of Environmental Quality
Amina Pollard	U.S. EPA Office of Research and Development
Andrew Rehn	California Game and Fish
Christina Relyea	Idaho State University
Bobbie Smith	U.S. EPA Region 9
Patti Spindler	Arizona Department of Environmental Quality
Jan Stevenson	Michigan State University
Glenn Suter	U.S. EPA Office of Research and Development A
Patti Tyler	U.S. EPA Region 8
John Van Sickle	U.S. EPA Office of Research and Development
Ian Waite	U.S. Geological Survey
Lori Winters	U.S. EPA Office of Research and Development
Bob Wisseman	
Lester Yuan	US EPA Office of Research and Development
Jeremy ZumBerge	Wyoming Department of Environmental Quality

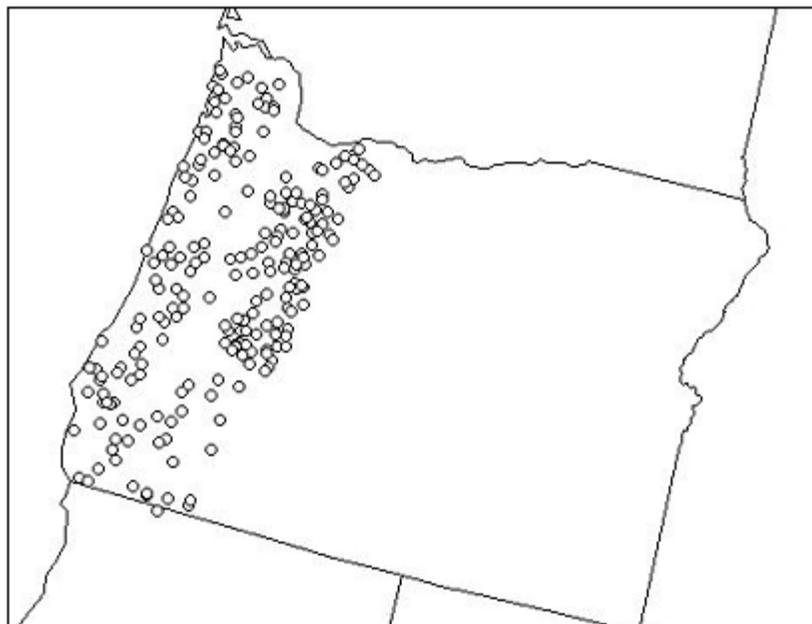
## APPENDIX B: DATA DESCRIPTION

Two data sets were used to illustrate the analysis methods described in this report: one contained data collected by the U.S. Environmental Protection Agency's Environmental Management and Assessment Program-Western Pilot Project (EMAP-West) from 2000 to 2001, and the other contained data collected in western Oregon by the Oregon Department of Environmental Quality from 1999 to 2000 (Figures B-1 and B-2). Both organizations used a similar sampling protocol. A reach 40 times the wetted width of the stream was delineated for sampling. Stream temperature was measured at the time of sampling. Substrate composition was estimated by summarizing the size distribution of particles at five locations on 21 transects. For the EMAP-West, macroinvertebrate samples were collected at eight randomized locations in riffles using a modified D-frame kicknet (500  $\mu\text{m}$  mesh) by disturbing a 1  $\text{ft}^2$  area for 30 seconds. In Oregon, samples were collected by disturbing 2  $\text{ft}^2$  areas at four randomized locations. Samples from both studies were composited and spread on a gridded pan and picked from randomly selected grid squares until at least 500 organisms were collected. Each organism was then identified to the lowest possible taxonomic level (usually genus or species).

A total of 392 complete samples in EMAP-West and 271 complete samples from Oregon were available for analysis.



**Figure B-1. Map of sampling locations for Western Environmental Monitoring and Assessment Program (EMAP-West).**



**Figure B-2. Map of sampling locations in Oregon.**  
**APPENDIX C: TOLERANCE VALUES**

**Table C-1. Temperature tolerance values (°C) derived using EMAP-West data**

Name	GLMMAX	GLMCL	GAMMAX	GAMCL	WA	CD75
<b>TROMBIDIFORMES</b>						
Atractides	.	.	.	.	14.5	17.4
Hygrobates	.	.	.	.	14.7	17.6
Lebertia	2	S	2	S	12.8	15.2
Protzia	11.9	I	10.8	S	12.9	15.2
Sperchon	.		.		14.3	17.8
Sperchonopsis	15.7	I	14.9	T	14.4	16.3
Testudacarus	11	I	9.9	S	12.2	14
Torrenticola	15.1	I	13.8	I	14.2	16.2
<b>VENEROIDA</b>						
Pisidium	17.1	I	17.1	T	15	18.2
<b>BASOMMATOPHORA</b>						
Physa	29.1	T	29.1	T	16.9	19.5
<b>COLEOPTERA</b>						
Agabus	.	.	19.5	T	15.5	19.5
Cleptelmis	14.6	I	14	I	14	15.7
Dubiraphia	29.1	T	27.2	T	16.8	19.8
Eubrianax	16	I	16.5	I	15.4	17.3
Helichus	16.8	I	16.5	T	15.6	17.3
Heterlimnius	7.7	I	2	S	10.9	13.2
Hydraena	.	.	.	.	14.6	17
Lara	11.9	I	11.6	I	12.1	14
Microcylloepus	29.1	T	29.1	T	19.2	22.9
Narpus	14	I	13.5	I	13.7	15.7
Optioservus	16.8	I	17.1	I	14.9	17.5
Ordobrevia	17.1	I	17.1	T	16	18
Oreodytes	16.5	I	16	T	15.2	16.5
Psephenus	22.3	I	29.1	T	18.7	21.3
Zaitzevia	16.5	I	16	I	14.8	17.3
<b>DIPTERA</b>						
Antocha	13.8	I	12.9	I	13.6	15.5
Atherix	29.1	T	19.5	T	16.2	18.2
Atrichopogon	29.1	T	29.1	T	16.9	20.7
Brillia	8.6	I	2	S	12	14.6
Cardiocladius	19.8	I	20.9	T	17	19.6

**Table C-1. Temperature tolerance values (°C) derived using EMAP-West data (continued)**

<b>Name</b>	<b>GLMMAX</b>	<b>GLMCL</b>	<b>GAMMAX</b>	<b>GAMCL</b>	<b>WA</b>	<b>CD75</b>
Chelifera	2	S	4.7	S	12.3	14.6
Chironomus	29.1	T	29.1	T	17.7	22.9
Cladotanytarsus	29.1	T	29.1	T	16.6	19.5
Clinocera	2	S	2	S	10.7	12.7
Corynoneura	10.8	I	9.4	S	13.2	15.3
Cricotopus	29.1	T	29.1	T	15.2	18.8
Cryptochironomus	29.1	T	29.1	T	17.6	21.5
Diamesa	2	S	2	S	11	13.1
Dicranota	2	S	2	S	13	15.6
Dicrotendipes	29.1	T	29.1	T	18.8	20.3
Dixa	12.4	I	11.3	S	13.1	15.2
Eukiefferiella	2	S	2	S	13.4	16.1
Glutops	7.5	I	2	S	10	11.7
Heleniella	2	S	9.1	S	12.6	14.7
Hemerodromia	22.3	I	22.8	T	17.3	19.6
Hexatoma	14.3	I	13.2	I	13.9	16.1
Hydrobaenus	2	S	2	S	9.8	12.7
Krenosmittia	2	S	2	S	10.4	12.4
Larsia	.	.	29.1	T	15	19.6
Limnophila	2	S	2	S	12.6	15.2
Limnophyes	.	.	12.4	T	14.4	17.3
Maruina	29.1	T	29.1	T	17	21
Micropsectra	2	S	2	S	12.6	15.1
Microtendipes	20.3	I	20.9	T	16	18.9
Nanocladius	15.1	I	16.5	S	14.3	18.2
Neoplasta	16.5	I	17.1	T	15.5	18
Nilotanypus	29.1	T	29.1	T	18.6	20.9
Oreogeton	2	S	2	S	10.2	11.9
Orthocladius	2	S	2	S	12.1	14
Pagastia	5.6	I	2	S	11.9	14.7
Parakiefferiella	.	.	2	S	13.4	17.8
Parametriocnemus	11.6	I	10.5	S	13.6	16.3
Paraphaenocladius	2	S	2	S	10.5	12.9
Paratanytarsus	17.6	I	19	T	15.7	19.1
Paratendipes	29.1	T	29.1	T	17.1	22
Parorthocladius	2	S	2	S	10.3	12.5

**Table C-1. Temperature tolerance values (°C) derived using EMAP-West data (continued)**

<b>Name</b>	<b>GLMMAX</b>	<b>GLMCL</b>	<b>GAMMAX</b>	<b>GAMCL</b>	<b>WA</b>	<b>CD75</b>
Pentaneura	21.4	I	24.4	T	17.5	21.3
Pericoma	2	S	2	S	11	12.7
Phaenopsectra	29.1	T	20.6	T	15.6	18.6
Polypedilum	29.1	T	29.1	T	16	19.1
Potthastia	29.1	T	17.9	T	15.4	18.2
Prosimulium	2	S	2	S	9.7	10.6
Pseudochironomus	29.1	T	29.1	T	18.2	22.7
Rheocricotopus	.		11	S	13.9	16.6
Rheotanytarsus	21.2	I	25.3	T	15.7	18.9
Simulium	17.6	I	17.6	T	14.8	17.7
Stempellina	2	S	2	S	11.5	14
Stempellinella	2	S	2	S	12.6	14.8
Synorthocladius	.	.	.	.	14.7	16.4
Tanytarsus	29.1	T	29.1	T	16.7	20.6
Thienemanniella	29.1	T	29.1	T	15	18.1
Thienemannimyia	23.9	I	29.1	T	15.9	18.9
Tipula	2	S	2	S	12.5	16.2
Tvetenia	2	S	2	S	12.8	15.4
Wiedemannia	2	S	8	S	12.3	14.7
<b>EPHEMEROPTERA</b>						
Acentrella	19.5	I	20.6	T	17	19.6
Ameletus	5.8	I	2	S	10.8	13.2
Baetis	14	I	13.8	S	14.1	17
Caenis	29.1	T	29.1	T	18.2	19.8
Caudatella	8.3	I	6.7	S	11.1	13
Cinygma	2	S	2	S	9.6	12.4
Cinygmula	2	S	2	S	10.9	13.2
Dipheter	14	I	13.8	I	13.8	16.3
Drunella	7.7	I	6.4	S	11.9	14
Epeorus	2	S	2	S	12.6	15.1
Ephemerella	2	S	2	S	12.7	15.4
Fallceon	29.1	T	29.1	T	19.4	22.2
Ironodes	11.6	I	11	S	12.4	15.1
Paraleptophlebia	9.1	I	2	S	12.6	15
Rhithrogena	7.5	I	3.4	S	12.3	14.7
Serratella	13.8	I	13.2	I	13.6	16

**Table C-1. Temperature tolerance values (°C) derived using EMAP-West data (continued)**

<b>Name</b>	<b>GLMMAX</b>	<b>GLMCL</b>	<b>GAMMAX</b>	<b>GAMCL</b>	<b>WA</b>	<b>CD75</b>
Tricorythodes	29.1	T	29.1	T	18.5	22.4
<b>HEMIPTERA</b>						
Ambrysus	29.1	T	29.1	T	18.9	22.7
<b>LEPIDOPTERA</b>						
Petrophila	29.1	T	29.1	T	20.5	23.8
<b>ODONATA</b>						
Argia	22.5	I	25	T	18.5	22
Ophiogomphus	29.1	T	29.1	T	18.5	24.5
<b>PLECOPTERA</b>						
Calineuria	15.7	I	16	I	14.8	17
Despaxia	11.3	I	10.8	S	11.8	13.1
Doroneuria	9.1	I	8	S	11	13.2
Hesperoperla	12.9	I	12.7	I	13.2	15.2
Isoperla	.	.	14.9	S	13.8	17.4
Malenka	14.6	I	14.6	I	14	16.1
Megarcys	7.5	I	3.4	S	9.8	11.7
Pteronarcys	13.8	I	13.5	S	13.6	16
Skwala	2	S	11.6	S	13.3	15.6
Suwallia	2	S	6.4	S	11.5	13.2
Sweltsa	5.3	I	2	S	11.4	13.6
Visoka	6.7	I	2	S	10	11.9
Yoraperla	8	I	5.8	S	10.9	13.1
Zapada	2	S	2	S	11.7	14.1
<b>TRICHOPTERA</b>						
Agapetus	15.7	I	15.4	I	14.8	16.3
Amiocentrus	2	S	9.1	S	13	15.1
Anagapetus	2	S	2	S	10.2	12.4
Apatania	11.6	I	11.3	S	12	13.4
Arctopsyche	11.3	I	10.5	S	12.2	13.7
Brachycentrus	.	.	11.9	S	13.3	16
Cheumatopsyche	28.6	I	29.1	T	18.9	22.7
Dicosmoecus	29.1	T	29.1	T	15.7	20.7
Dolophilodes	2	S	7.5	S	11.6	13.1
Eccelisomyia	2	S	2	S	10.3	11.9
Glossosoma	9.7	I	2	S	12.2	14.9
Helicopsyche	29.1	T	29.1	T	17.2	20.4

**Table C-1. Temperature tolerance values (°C) derived using EMAP-West data (continued)**

Name	GLMMAX	GLMCL	GAMMAX	GAMCL	WA	CD75
Hydropsyche	22.3	I	24.7	T	16.4	19.2
Hydroptila	22.5	I	25	T	17.8	20.4
Lepidostoma	12.1	I	11.9	S	13.1	15.7
Micrasema	8.8	I	7.5	S	12.9	15.6
Neophylax	11.9	I	11	S	12.7	14.9
Neothremma	6.4	I	2	S	9.5	10.5
Ochrotrichia	29.1	T	29.1	T	17.5	20.6
Oecetis	29.1	T	29.1	T	16.6	22.9
Oligophlebodes	9.1	I	7.7	S	10.5	12.1
Parapsyche	6.9	I	2	S	11.4	14.2
Polycentropus	29.1	T	29.1	T	17.6	20.4
Rhyacophila	2	S	2	S	11.9	14.8
Wormaldia	17.1	I	16.8	I	15.3	17.8
AMPHIPODA						
Hyaella	20.1	I	21.2	T	16.5	19.1
TRICLADIDA						
Polycelis	2	S	2	S	10.4	12.4

CD75 = cumulative 75<sup>th</sup> percentile

GAMCL = generalized additive model - curve shape class

GAMMAX = generalized additive model - maximum point (optimum)

GLMCL = generalized linear model - curve shape class

GLMMAX = generalized line model - maximum point (optimum)

I = intermediately tolerant

S = sensitive

T = tolerant

WA = weighted averaging

**Table C-2. Sediment tolerance values (percent sands and fines) derived from EMAP-West data**

Name	GLMMAX	GLMCL	GAMMAX	GAMCL	WA	CD75
TROMBIDIFORMES						
Atractides	0	S	15.7	S	22.9	34.3
Hygrobates	.	.	97.1	T	31.2	41
Lebertia	0	S	0	S	22.4	29.8
Protzia	0	S	0	S	14.4	19.2
Sperchon	.	.	.	.	25.9	37.1
Sperchonopsis	0	S	0	S	19.6	25.7
Testudacarus	21.6	I	17.7	S	19.3	27.6
Torrenticola	0	S	0	S	16.1	22.1
VENEROIDA						
Pisidium	97.1	T	97.1	T	42.5	67
BASOMMATOPHORA						
Physa	97.1	T	97.1	T	38.3	57.1
COLEOPTERA						
Agabus	97.1	T	97.1	T	39.4	56.2
Cleptelmis	.	.	35.3	T	26.4	38
Dubiraphia	97.1	T	97.1	T	55.7	74.3
Eubrianax	11.8	I	7.8	S	12.6	17.1
Helichus	.	.	.	.	24.6	35.2
Heterlimnius	0	S	0	S	20.7	29.5
Hydraena	.	.	32.4	S	25.5	36.2
Lara	.	.	.	.	22.7	32.4
Microcylloepus	.	.	32.4	T	29.6	34.3
Narpus	0	S	22.6	S	22	35
Optioservus	34.3	I	30.4	S	26.3	37.1
Ordobrevia	0	S	0	S	10	12.4
Oreodytes	97.1	T	37.3	T	29.9	37.1
Psephenus	0	S	0	S	17.4	24.8
Zaitzevia	0	S	0	S	20	27.6
DIPTERA						
Antocha	7.8	I	0	S	15	21.9
Atherix	0	S	0	S	14.2	16.2
Atrichopogon	25.5	I	19.6	S	21.8	37.1
Brillia	0	S	0	S	20.2	31.4
Caloparyphus	33.4	I	31.4	S	25.9	36.2
Cardiocladius	24.5	I	25.5	S	20.9	29.5

**Table C-2. Sediment tolerance values (percent sands and fines) derived from EMAP-West data (continued)**

Name	GLMMAX	GLMCL	GAMMAX	GAMCL	WA	CD75
Chelifera	23.5	I	21.6	S	22.3	32.4
Chironomus	97.1	T	97.1	T	45.9	58.1
Cladotanytarsus	97.1	T	97.1	T	36	57.1
Clinocera	0	S	0	S	19.2	27.6
Corynoneura	0	S	0	S	23.9	34.3
Cricotopus	97.1	T	97.1	T	31	47.6
Cryptochironomus	80.5	I	86.3	T	55.3	68.6
Diamesa	0	S	0	S	12.7	17.1
Dicranota	.	.	.	.	28.9	40
Dicrotendipes	97.1	T	97.1	T	53.7	74
Dixa	.	.	.	.	23.6	36.2
Eukiefferiella	0	S	0	S	22.7	32.4
Glutops	0	S	17.7	S	20.2	27.6
Heleniella	0	S	0	S	19.6	28.6
Hemerodromia	97.1	T	64.8	T	33.7	55.2
Hexatoma	0	S	0	S	20.7	28.6
Hydrobaenus	.	.	.	.	23.1	33.3
Krenosmittia	.	.	.	.	24	33.7
Larsia	97.1	T	74.6	T	34.1	45.7
Limnophila	97.1	T	97.1	T	31.8	51.4
Limnophyes	97.1	T	97.1	T	33.4	46.7
Maruina	16.7	I	14.7	S	16.8	24.8
Micropsectra	0	S	0	S	21.7	31.4
Microtendipes	.	.	.	.	27.2	36.2
Nanocladius	.	.	60.8	.	28.2	49
Neoplasta	0	S	0	S	19.4	34.3
Nilotanypus	97.1	T	97.1	T	33.8	55.2
Oreogeton	0	S	0	S	16.2	17.1
Orthocladius	.	.	.	.	24.8	38.1
Pagastia	16.7	I	10.8	S	20.5	30.5
Parakiefferiella	62.8	I	70.6	T	37.6	57.1
Parametricnemus	.	.	29.4	S	26.6	35.2
Paraphaenocladius	0	S	0	S	23.2	35
Paratanytarsus	97.1	T	97.1	T	44.1	60
Paratendipes	97.1	T	.	.	33.7	50.5
Parorthocladius	0	S	0	S	15.9	24.8

**Table C-2. Sediment tolerance values (percent sands and fines) derived from EMAP-West data (continued)**

<b>Name</b>	<b>GLMMAX</b>	<b>GLMCL</b>	<b>GAMMAX</b>	<b>GAMCL</b>	<b>WA</b>	<b>CD75</b>
Pentaneura	97.1	T	97.1	T	34.5	54.3
Pericoma	.	.	.	.	25.6	39.7
Phaenopsectra	97.1	T	97.1	T	33.2	54.3
Polypedilum	97.1	T	69.7	T	30.1	50.5
Potthastia	0	S	0	S	12.6	18.1
Prosimulium	0	S	0	S	18	26.9
Pseudochironomus	49.1	I	50	T	33.3	45.7
Rheocricotopus	0	S	0	S	22.5	31.4
Rheotanytarsus	.	.	.	.	27.4	40
Simulium	97.1	T	77.5	T	28.8	40
Stempellina	97.1	T	85.4	T	33.2	55.2
Stempellinella	0	S	0	S	20	26.9
Synorthocladius	21.6	I	22.6	S	19.5	30.5
Tanytarsus	97.1	T	97.1	T	36.3	58.1
Thienemanniella	.	.	.	.	28	38.1
Thienemannimyia	97.1	T	97.1	T	31.7	49
Tipula	65.7	I	73.6	T	40.9	57.1
Tvetenia	0	S	0	S	22.5	32.4
Wiedemannia	0	S	0	S	13.5	20
<b>EPHEMEROPTERA</b>						
Acentrella	.	.	97.1	T	27	45.7
Ameletus	0	S	0	S	19.7	29.5
Baetis	0	S	0	S	24.2	34.3
Caenis	64.8	I	72.6	T	54.2	68.6
Caudatella	0	S	0	S	14.2	21
Cinygma	97.1	T	97.1	T	32.5	46.7
Cinygmula	0	S	0	S	17.8	26.9
Dipheter	0	S	0	S	21.7	32.4
Drunella	0	S	0	S	16.7	25.7
Epeorus	0	S	0	S	13.3	20
Ephemerella	.	.	.	.	23.9	31.4
Fallceon	97.1	T	97.1	T	42.1	60.6
Ironodes	0	S	0	S	15.8	26.1
Nixe	.	.	97.1	.	27.5	36.2
Paraleptophlebia	0	S	0	S	21.7	32.4
Rhithrogena	0	S	0	S	12	16.2

**Table C-2. Sediment tolerance values (percent sands and fines) derived from EMAP-West data (continued)**

Name	GLMMAX	GLMCL	GAMMAX	GAMCL	WA	CD75
Serratella	0	S	0	S	18.1	25.7
Tricorythodes	97.1	T	97.1	T	39.5	60.6
HEMIPTERA						
Ambrysus	48.1	I	47.1	T	37	47.6
LEPIDOPTERA						
Petrophila	.	.	30.4	S	25.1	33.7
ODONATA						
Argia	97.1	T	97.1	T	40.4	66.7
Ophiogomphus	97.1	T	73.6	T	39.1	62.7
PLECOPTERA						
Calineuria	0	S	0	S	12.8	17.9
Despaxia	0	S	0	S	16.1	24.4
Doroneuria	12.8	I	0	S	17.8	26.9
Hesperoperla	0	S	0	S	18.2	27.6
Isoperla	37.3	I	35.3	I	28.4	36.2
Malenka	0	S	0	S	23.9	34.3
Megarcys	0	S	0	S	13.7	15.2
Pteronarcys	0	S	0	S	14.4	26.7
Skwala	0	S	0	S	21.4	30.5
Suwallia	0	S	0	S	12.9	17.1
Sweltsa	0	S	0	S	17.3	24.4
Visoka	0	S	0	S	16.6	26.9
Yoraperla	0	S	0	S	19	27.6
Zapada	0	S	0	S	21.2	31.4
TRICHOPTERA						
Agapetus	0	S	0	S	22	26.7
Amiocentrus	0	S	0	S	19.8	31.4
Anagapetus	0	S	0	S	15	26.1
Apatania	0	S	0	S	15.2	18.1
Arctopsyche	0	S	0	S	11.3	14.3
Brachycentrus	31.4	I	31.4	S	24.6	35.2
Cheumatopsyche	61.8	I	70.6	T	39.4	60.6
Dicosmoecus	19.6	I	17.7	S	17.7	26.7
Dolophilodes	0	S	0	S	14.6	26.1
Ecclisomyia	0	S	0	S	14	20
Glossosoma	0	S	0	S	16.9	26.7

**Table C-2. Sediment tolerance values (percent sands and fines) derived from EMAP-West data (continued)**

Name	GLMMAX	GLMCL	GAMMAX	GAMCL	WA	CD75
Gumaga	0	S	0	S	21.2	28.6
Helicopsyche	.	.	35.3	T	29.9	38.1
Hydropsyche	.	.	.	.	28.1	41
Hydroptila	71.6	I	97.1	T	38.1	55.2
Lepidostoma	0	S	0	S	18.7	27.6
Micrasema	0	S	0	S	21.8	32.4
Neophylax	0	S	0	S	15.3	21.9
Neothremma	18.6	I	18.6	S	16.9	23.5
Ochrotrichia	27.5	I	23.5	S	22.8	33.3
Oecetis	49.1	I	57.9	T	36.1	58.1
Oligophlebodes	0	S	0	S	16.2	20
Parapsyche	0	S	0	S	17.5	26.7
Polycentropus	0	S	0	S	20.1	29.5
Rhyacophila	0	S	0	S	18	26.7
Wormaldia	0	S	12.8	S	20	27.6
AMPHIPODA						
Hyaella	86.3	I	97.1	T	58.3	76.2
TRICLADIDA						
Polycelis	0	S	0	S	19.9	29.5

CD75 = cumulative 75<sup>th</sup> percentile

GAMCL = generalized additive model - curve shape class

GAMMAX = generalized additive model - maximum point (optimum)

GLMCL = generalized linear model - curve shape class

GLMMAX = generalized line model - maximum point (optimum)

I = intermediately tolerant

S = sensitive

T = tolerant

WA = weighted averaging

## APPENDIX D: EXAMPLE STATISTICAL SCRIPTS

This appendix provides short scripts that perform the statistical analyses described in the report. R, a free software package for statistical computations, is used for these computations. More information on R and the package itself are available from <http://www.r-project.org/>. If you wish to run the scripts provided in this section, please visit the R web page and install R on your computer.

### D.1. R: BASIC SYNTAX

Variable names in R can be composed of combinations of letters, numbers, and periods. They are case sensitive.

```
x, y, X, Y, flow.rate
```

(Note that in this and all subsequent sections, R commands can be run by typing text directly into the R Console window. R commands are shown in the Courier font.)

Use the assignment operator, `<-`, to assign a value to a variable.

```
x <- 1                # Assign a single value to the variable x
x <- c(1, 3, 2)       # Assign a vector of numbers to x
x <- c(T, F, T)       # Assign a vector of logical values to x

x <- list(colors = c("red", "blue", "black"), numbers = c(1, 3))
# Assign a list of dissimilar objects to x
```

(Comments are preceded with a “#” and are ignored by R.)

The value of any variable can be examined by typing the variable name, or by using the `print` command:

```
x
print(x)
```

Simple mathematical and statistical operations can be performed on different vectors.

```
x + y                # Addition
x - y                # Subtraction
x * y                # Multiplication
x / y                # Division
mean(x)              # Arithmetic mean
var(x)               # Variance
sum(x)               # The sum of all the elements of x
```

The most commonly used format for storing data is the data frame, which is a list of objects of the same length. Data frames allow one to combine logical, numerical, and factor data in a single data structure.

```
site.name <- c("A", "B", "C", "D")      # A site label stored as a
                                        # factor
pH <- c(7.6, 6.0, 4.0, 8.2)            # Site pH stored as a
                                        # numerical vector
abund.baetis <- c(103, 204, 602, 301)  # Baetis abundance stored as
                                        # a numerical vector
sampled.spring <- c(T, T, F, T)        # Sampling season stored as a
                                        # logical vector

all.data <- data.frame(site.name, pH, abund.baetis, sampled.spring)
                                        # All data combined together
                                        # as a data frame
```

The elements of a vector can be referred to in a variety of ways.

```
x[1]          # The first element of the vector x
x[1:3]        # The first three elements of vector x
x[c(T,T,F)]   # The first two elements of x (assuming that x
              # has three elements)
x[-1]         # All of x except for the first element
```

We can also refer to different subsets of a data frame in a variety of ways.

```
all.data$pH   # The element labelled "pH" from the data frame
              # all.data
all.data[, "pH"] # The same column labelled "pH"
all.data[, 2]   # The second column of the data.frame
all.data[1,]   # The first row of the data.frame
```

Within R, you can access help pages on a particular command by typing,

```
help(<command name>)
```

For example:

```
help(glm)
help(mean)
```

## D.2. LOADING DATA

Data formatted as tab-delimited text can be loaded easily with,

```
data.set <- read.delim(<filename>) # Loads a tab-delimited data file
```

A typical site-environment file would contain the environmental data for each site listed in each row of the file, with each field delimited with tabs:

SITE.ID	TEMP	SED
A	13	40
B	17	20
C	15	10
...		

In the above example, data on stream temperature (TEMP) and fine sediment (SED) is recorded for each site.

A typical site-species file would contain the abundances of different species observed at a site:

SITE.ID	BAETIS	MALENKA	...
A	55	3	...
B	22	10	...
C	4	0	...
...			

Assuming that both site-species and site-environment data are available as tab-delimited text files, they can be loaded into R with the following commands.

```
# Load tab-delimited files
site.species <- read.delim("site.species.txt")
env.data <- read.delim("env.data.txt")

ls()      # Check to see if new data frames are listed
```

The two data sets can be merged into a single data frame (matching environmental and biological data) with the following command:

```
dfmerge <- merge(site.species, env.data, by = "SITE.ID")
```

## D.3. ESTIMATING TOLERANCE VALUES FROM FIELD DATA

### D.3.1. Weighted Averages

The basic formula for a weighted average tolerance value (eqn 1) can be represented in R as follows:

```
WA <- sum(Y*x) / sum(Y)
```

where  $Y$  is a vector containing the abundance of the taxon of interest at each sample.  $Y$  can also contain presence/absence data coded as 1 for present and 0 for absent.  $x$  is vector containing the

value of the environmental variable of interest at each sample, and `sum` computes the sum of the values of a numerical vector.

Use a `for` loop to create weighted averages for many different taxa using previously loaded data. To run this script, first make sure that you have loaded biological data and merged them into a single data frame (see previous section).

First designate the taxa for which you want to compute tolerance values.

```
taxa.names <- c("ACENTRELLA", "DIPHETOR", "AMELETUS")
```

The `for` loop then repeats the computation for each of the selected taxa.

```
WA <- rep(NA, times = length(taxa.names))
# Define a WA to be vector of a length the same as the
# number of taxa of interest

for (i in 1:length(taxa.names)) {
  WA[i] <- sum(dfmerge[,taxa.names[i]]*dfmerge$temp)/
    sum(dfmerge[,taxa.names[i]])
}
names(WA) <- taxa.names
print(WA)
```

### D.3.2. Cumulative Percentiles

Use a `for` loop similar to that used for weighted averages to compute cumulative percentile tolerance values (eqn 2) for different taxa:

```
# Define a storage vector for the cumulative percentile
CP <- rep(NA, times = length(taxa.names))

# Sort sites by the value of the environmental variable
dftemp <- dfmerge[order(dfmerge$temp), ]

# Select a cutoff percentile
cutoff <- 0.75

# Specify three plots per page
par(mfrow = c(1,3), pty = "s")

for (i in 1:length(taxa.names)) {
  # Compute cumulative sum of abundances
  csum <- cumsum(dftemp[,taxa.names[i]])/sum(dftemp[,taxa.names[i]])

  # Make plots like Figure 5
  plot(dftemp$temp, csum, type = "l", xlab = "Temperature",
    ylab = "Proportion of total", main = taxa.names[i])
}
```

```

# Search for point at which cumulative sum is 0.75
ic <- 1
while (csum[ic] < 0.75) ic <- ic + 1

# Save the temperature that corresponds to this
# percentile.
CP[i] <- dftemp$temp[ic]
}
names(CP) <- taxa.names
print(CP)

```

### D.3.3. Parametric Regressions

Parametric regressions for presence/absence of different taxa (eqn 3) can be computed as follows. We store each model in a list.

```

# Create storage list
modlist.glm <- as.list(rep(NA, times = length(taxa.names)))

for (i in 1:length(taxa.names)) {

# Create a logical vector that is true if taxon is
# present and false if taxon is absent.
resp <- dfmerge[,taxa.names[i]]>0

# Fit the regression model and store the results in a list.
# Here, poly(temp,2) specifies that the
# model is fitting using a second order polynomial of the
# explanatory variable. glm calls the function that fits
# Generalized Linear Models. We specify in this case that
# the response variable is distributed binomially.
modlist.glm[[i]] <- glm(resp ~ poly(temp,2), data = dfmerge,
                        family = "binomial")

print(summary(modlist.glm[[i]]))
}

```

Plot model results (similar to Figure 6) as follows.

```

par(mfrow = c(1,3), pty = "s") # Specify 3 plots per page
for (i in 1:length(taxa.names)) {

# Compute mean predicted probability of occurrence
# and standard errors about this predicted probability.
predres <- predict(modlist.glm[[i]], type= "link", se.fit = T)

# Compute upper and lower 90% confidence limits
up.bound.link <- predres$fit + 1.65*predres$se.fit
low.bound.link <- predres$fit - 1.65*predres$se.fit
mean.resp.link <- predres$fit

# Convert from logit transformed values to probability.

```

```

up.bound <- exp(up.bound.link)/(1+exp(up.bound.link))
low.bound <- exp(low.bound.link)/(1+exp(low.bound.link))
mean.resp <- exp(mean.resp.link)/(1+exp(mean.resp.link))

# Sort the environmental variable.
iord <- order(dfmerge$temp)

# Define bins to summarize observational data as
# probabilities of occurrence

# Define the number of bins
nbin <- 20

# Define bin boundaries so each bin has approximately the same
# number of observations.
cutp <- quantile(dfmerge$temp,
probs = seq(from = 0, to = 1, length = 20))

# Compute the midpoint of each bin
cutm <- 0.5*(cutp[-1] + cutp[-nbin])

# Assign a factor to each bin
cutf <- cut(dfmerge$temp, cutp, include.lowest = T)

# Compute the mean of the presence/absence data within each
# bin.
vals <- tapply(dfmerge[, taxa.names[i]] > 0, cutf, mean)

# Now generate the plot,
# Plot binned observational data as symbols.
plot(cutm, vals, xlab = "Temperature",
      ylab = "Probability of occurrence", ylim = c(0,1),
      main = taxa.names[i])

# Plot mean fit as a solid line.
lines(dfmerge$temp[iord], mean.resp[iord])

# Plot confidence limits as dotted lines.
lines(dfmerge$temp[iord], up.bound[iord], lty = 2)
lines(dfmerge$temp[iord], low.bound[iord], lty = 2)

}

```

### D.3.4. Nonparametric Regression

Nonparametric regressions (eqn 5) can be computed with a set of commands similar to those of parameteric regressions.

```

library(gam) # Load GAM library
modlist.gam <- as.list(rep(NA, times = length(taxa.names)))
for (i in 1:length(taxa.names)) {

# Create a logical vector that is true if taxon is
# present and false if taxon is absent.

```

```

resp <- dfmerge[, taxa.names[i]]>0

# Fit the regression model, specifying two degrees of freedom
# to the curve fit.
modlist.gam[[i]] <- gam(resp ~ s(temp, df = 2), data = dfmerge,
                        family = "binomial")

print(summary(modlist.gam[[i]]))
}

```

Plot model results (similar to Figure 8) as follows.

```

par(mfrow = c(1,3), pty = "s")
for (i in 1:length(taxa.names)) {

  predres <- predict(modlist.gam[[i]], type= "link", se.fit = T)

  up.bound.link <- predres$fit + 1.65*predres$se.fit
  low.bound.link <- predres$fit - 1.65*predres$se.fit
  mean.resp.link <- predres$fit

  up.bound <- exp(up.bound.link)/(1+exp(up.bound.link))
  low.bound <- exp(low.bound.link)/(1+exp(low.bound.link))
  mean.resp <- exp(mean.resp.link)/(1+exp(mean.resp.link))

  iord <- order(dfmerge$temp)

  nbin <- 20
  cutp <- quantile(dfmerge$temp, probs = seq(from = 0, to = 1,
      length = 20))

  cutm <- 0.5*(cutp[-1] + cutp[-nbin])
  cutf <- cut(dfmerge$temp, cutp, include.lowest = T)
  vals <- tapply(dfmerge[, taxa.names[i]] > 0, cutf, mean)

  plot(cutm, vals, xlab = "Temperature",
       ylab = "Probability of occurrence", ylim = c(0,1),
       main = taxa.names[i])

  lines(dfmerge$temp[iord], mean.resp[iord])
  lines(dfmerge$temp[iord], up.bound[iord], lty = 2)
  lines(dfmerge$temp[iord], low.bound[iord], lty = 2)
}

```

### D.3.5. Assessing Model Fit

The area under the ROC curve for each model can be computed by first imagining a pair of sites where the species of interest is present at one site and absent at the other. We would expect the probabilities of occurrence predicted by the regression model would be greater at the site where the species is present than at the site where the species is absent. The area

under the ROC is equivalent to the proportion of all such pairwise comparisons in which this condition is satisfied. The following script performs this computation.

```
# Define storage vector for ROC
roc <- rep(NA, times = length(taxa.names))

for (i in 1:length(taxa.names)) {
  # Compute mean predicted probability of occurrence
  predout <- predict(modlist.glm[[i]], type = "response")

  # Generate logical vector corresponding to presence/absence
  resp <- dfmerge[, taxa.names[i]] > 0

  # Divide predicted probabilities into sites where
  # species is present ("x") and sites where the species is
  # absent ("y").
  x <- predout[resp]
  y <- predout[! resp]

  # Now perform all pairwise comparisons of x vs. y
  # and store results in a matrix
  rocmat <- matrix(NA, nrow = length(x), ncol = length(y))
  for (j in 1:length(x)) {
    rocmat[j,] <- as.numeric(x[j] > y)
  }

  # Summarize all comparisons to compute area under ROC
  roc[i] <- sum(rocmat)/(length(x)*length(y))
}
names(roc) <- taxa.names
print(roc)
```

### D.3.6. Optima

The optimum tolerance value can be found by identifying the maximum point on the fitted regression curve.

```
opt <- rep(NA, times = length(taxa.names))
for (i in 1:length(taxa.names)) {
  predout <- predict(modlist.glm[[i]], type = "response")

  # Find the index number of the maximum probability
  imatch <- match(max(predout), predout)

  # The optimum is value of the environmental variable
  # at the maximum probability.
  opt[i] <- dfmerge$temp[imatch]
}
names(opt) <- taxa.names
print(opt)
```

### D.3.7. Curve Shape

Curves shapes can be classified into increasing, decreasing, or unimodal, by comparing the mean responses to the confidence intervals. Note that `unimod.test` is an R function, a series of commands that is executed when `unimod.test` is called.

```
unimod.test <- function(mnr, ubnd, lbnd) {

  # Find the maximum and minimum predicted mean probabilities
  lmax <- max(mnr)
  lmin <- min(mnr)

  # Find index locations for these probabilities
  imax <- match(lmax, mnr)
  imin <- match(lmin, mnr)

  x.out <- F
  y.out <- F

  # Compare mean predicted probability to the left of maximum point
  # with upper confidence bound. Store a T in x.out if
  # any point in the mean response deviates from the
  # upper confidence limit
  if (imax > 1) {
    x.out <- sum(lmax == pmax(lmax, ubnd[1:(imax-1)])) > 0
  }

  # Store a T in y.out if any point in the mean probability
  # to the right of the maximum point deviates from the upper
  # confidence limit
  if (imax < length(ubnd)) {
    y.out <- sum(lmax == pmax(lmax,
      ubnd[(imax+1):length(ubnd)])) > 0
  }

  # Perform same set of tests for lower confidence limit
  a.out <- F
  b.out <- F
  if (imin > 1) {
    a.out <- sum(lmin == pmin(lmin, lbnd[1:(imin-1)])) > 0
  }
  if (imin < length(lbnd)) {
    b.out <- sum(lmin == pmin(lmin,
      lbnd[(imin+1):length(lbnd)])) > 0
  }

  # The information on where the mean curve deviates from the
  # confidence limits tells us its curve shape...
  if (x.out & y.out) {
    return("Unimodal")
  }
  if (a.out & b.out) {
    return("Concave up")
  }
}
```

```

    }
    if (x.out | b.out) {
      return("Increasing")
    }
    if (y.out | a.out) {
      return("Decreasing")
    }
    if (!(x.out | y.out | a.out | b.out)) {
      return(NA)
    }
  }
}

tolcl <- rep("", times = length(taxa.names))
for (i in 1:length(taxa.names)) {
  predres <- predict(modlist.gam[[i]], type= "link", se.fit = T)

  # Compute upper and lower 90% confidence limits
  up.bound.link <- predres$fit + 1.65*predres$se.fit
  low.bound.link <- predres$fit - 1.65*predres$se.fit
  mean.resp.link <- predres$fit

  # Convert from logit transformed values to probability.
  up.bound <- exp(up.bound.link)/(1+exp(up.bound.link))
  low.bound <- exp(low.bound.link)/(1+exp(low.bound.link))
  mean.resp <- exp(mean.resp.link)/(1+exp(mean.resp.link))

  # unimod.test requires that the responses be sorted by
  # the value of the environmental variable.
  iord <- order(dfmerge$temp)

  tolcl[i] <- unimod.test(mean.resp[iord], up.bound[iord],
                        low.bound[iord])
}
names(tolcl) <- taxa.names
print(tolcl)

```

#### D.4. APPLYING TOLERANCE VALUES IN ASSESSMENT

In this section metric values are computed using the same data from which the tolerance values were calculated. Typically, one would develop tolerance value using a calibration data set, and then compute metrics for a second, independent set of test data.

First, expand the list of taxa to include all taxa in the data set that occur in at least 20 sites.

```

# Get names of all taxa in the data set
taxa.names.init <- names(site.species)[-1]

# Compute the number of occurrence of each taxon
getocc <- function(x) sum(x>0)
numocc <- apply(site.species[, taxa.names.init], 2, getocc)

# Save all taxa names that occur in at least 20 sites

```

```
taxa.names <- taxa.names.init[numocc >= 20]
```

Now, recompute nonparametric regression and curve classifications for this expanded list of taxa.

#### D.4.1. Richness

The richness of sensitive taxa can be computed as follows:

```
# Select only taxa for which tolerance values have been computed.
mat1 <- as.matrix(dfmerge[, taxa.names])

# Convert data to presence-absence.
mat2 <- as.numeric(mat1 > 0)
dim(mat2) <- dim(mat1)

# Select sensitive taxa
sens <- as.numeric(tolcl == "Decreasing")

# %*% performs a matrix multiplication, which gives the taxa
# present at a site that were classified as intolerant
sens.rich <- mat2 %*% sens

# Plot resulting metric against observed temperature.
plot(dfmerge$temp, sens.rich, xlab = "Temperature",
     ylab = "Sensitive richness")
```

The same operation can be repeated to compute the richness of tolerant taxa.

#### D.4.2. Proportion Total Taxa

The proportion of observed taxa that are sensitive taxa can be computed using the sensitive richness computed in the last section.

```
tot.rich <- apply(mat2, 1, sum) # Compute total richness

# The command "apply" applies the same operation to the
# rows or columns of a matrix.
# In this case, we compute the sum for each of the rows
# of the matrix.

# The proportion of taxa sensitive is computed with
# simple division.
ptax.sens <- sens.rich/tot.rich

# Plot resulting metric against observed temperature.
plot(dfmerge$temp, ptax.sens, xlab = "Temperature",
     ylab = "Proportion taxa")
```

### D.4.3. Relative Abundance

The relative abundance of sensitive taxa can be computed as follows:

```
# Only select taxa for which tolerance classifications
# have been computed.
mat1 <- as.matrix(dfmerge[, taxa.names])

# Select sensitive taxa
sens <- as.numeric(tolcl == "Decreasing")

# Use matrix multiplication to compute the number of
# sensitive individuals collected
abn.sens <- mat1 %*% sens

# Compute total abundance.
abn.tot <- apply(mat1, 1, sum)

# Compute relative abundance of sensitive taxa
relabn.sens <- abn.sens/abn.tot

# Plot resulting metric against observed temperature.
plot(dfmerge$temp, relabn.sens, xlab = "Temperature",
      ylab = "Relative abundance")
```

The same operation can be repeated to compute the relative abundance of tolerance taxa.

### D.4.4. Mean Tolerance Value

First, compute weighted averages or another continuous tolerance value for the expanded list of taxa.

The mean tolerance value observed at a site can be computed as follows,

```
# Only select taxa for which tolerance values
# have been computed.
mat1 <- as.matrix(dfmerge[, taxa.names])

# First get total abundance
tot.abn <- apply(mat1, 1, sum)

# Use matrix multiplication to compute the sum of all
# observed tolerance values, and then divide by total
# abundance to get the mean tolerance value.
mean.tv <- (mat1 %*% WA)/tot.abn

plot(dfmerge$temp, mean.tv, xlab = "Temperature",
      ylab = "Mean weighted average")
```

In this example, weighted average tolerance values have been used. Other continuous tolerance values such as optima or cumulative percentiles can be substituted in as well. Also, the mean computed here is weighted by abundance.