

## Weight-of-Evidence (WOE): Quantitative Estimation of Probability of Impairment for Individual and Multiple Lines of Evidence

Eric P. Smith,\* Ilya Lipkovich, and Keying Ye

Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0439

### ABSTRACT

Environmental decision-making is complex and often based on multiple lines of evidence. Integrating the information from these multiple lines of evidence is rarely a simple process. We present a quantitative approach to the combination of multiple lines of evidence through calculation of weight-of-evidence, with reference conditions used to define a not impaired state. The approach is risk-based with measurement of risk computed as the probability of impairment. When data on reference conditions are available, there are a variety of methods for calculating this probability. Statistical theory and the use of odds ratios provide a method for combining the measures of risk from the different lines of evidence. The approach is illustrated using data from the Great Lakes to predict the risk at potentially contaminated sites.

**Key Words:** Bayesian statistics, odds ratio, hazard ranking, combining information, risk assessment, reference conditions.

### INTRODUCTION

Environmental decision-making is often based on multiple sets of information or lines of evidence. By line of evidence we mean a set of information that pertains to an important aspect of the environment. For example, in the sediment quality triad (Chapman 1996), there are three lines of evidence, the toxicity line, the biological field line and the chemistry line. It is difficult to combine the information from these multiple sources into a single measure for decision-making. Weight-of-evidence (WOE) is sometimes used as an approach for combining the information, however it is rarely used in a quantitative manner. This paper discusses a quantitative approach to WOE. A statistical approach is taken in which the likelihood of the data is calculated under two different scenarios and a decision made based on the ratio of the likelihoods. Our view is that there are two states, and we must decide which of the states is more likely. Examples of pairs of states common to environmental

\* Corresponding author: Tel(voice): 542-231-7929, Tel(fax): 540-231-3863; epsmith@vt.edu



decision making are (impaired, not impaired), (remediate, don't remediate), (list, don't list), *etc.* The view we take is that interest is in a single site, we collect information from a sample at that site and based on the information make a decision that the site is impaired or the site is not impaired. For practical reasons, we assume the simple case that there is ample information on reference conditions and interest is in evaluating a single new location. This gives us the ability to obtain a precise estimate of the probability.

#### ESTIMATING WEIGHT-OF-EVIDENCE (WOE)

A quantitative approach to WOE is based on the concept of statistical weight of evidence. This idea and early applications dates back to work by Alan Turing in World War II (for a more general discussion of history and concepts of statistical weight of evidence see Good 1983, 1985, 1988). In this approach, there are two states and we must decide which state is more likely given the data. We can view the states as the site is or is not impaired. Without observing the information on the site, we may have opinions or insights into the condition of the site. These insights might be based on previous data (condition in previous years) or be from sites that are close in space. This information may be used to form a prior opinion or prior probability of impairment. After the data are collected, we process the data to evaluate the site. This leads to a Bayesian approach in which the data are used to update the prior information. The lack of prior information or unwillingness to use this information leads to a frequentist approach where the data alone are used to make a decision. Either approach may be based on a single line of evidence or multiple lines. The individual lines of evidence are usually evaluated separately and by combining them we hope to make a stronger inference.

Statistical WOE is based on a quantitative evaluation of the data and requires a model that describes the data. In the simplest approach, there are two states and we must decide which state is more likely given the data. Because it is not always easy to describe impairment, an alternative approach is to evaluate the risk of impairment of a site considering the baseline risk of a not impaired site. If we view the possible outcomes as the site is or is not impaired, the risk is then the probability that the site is impaired or  $P(\text{impairment})$  where we evaluate this probability after information is collected.

The odds are a way to evaluate how big the probability is relative to the baseline risk. The odds of impairment are calculated as the ratio of the probability the site is impaired over the probability the site is not impaired. Although this problem is analogous to tossing a coin, estimating the probability of impairment is not easy since "impairment" is not an observable attribute of a sample. The quality or state of the site (impairment or no impairment) must be inferred based on information that is collected on both impaired and unimpaired sites. A reasonable approach is based on Bayes rule (Gelman *et al.* 1995). With no data the probabilities would be estimated based on prior information that may come from previous studies. More generally, we collect data to improve these estimates. Given data we have to calculate the probability of impairment or no impairment. However, even with data we do not have probabilities of impairment, only probabilities associated with observations given a model for impaired sites and not impaired sites. For example, if there is



## Quantitative Weight-of-Evidence for Environmental Assessment

ample information on sites that are not impaired we may compare our data to that and estimate the probability the data come from that distribution. We do not have the probability of no impairment, only the probability that the data come from that distribution. Given data and information about the different groups we can calculate how likely the data are given they are from one of the groups. This probability must be computed based on a statistical model for the data (possibly with different models for each group). For example, we might specify that the model for the impairment sites for dissolved oxygen is normal with mean 4 and for no impairment sites is normal with mean 7. If we obtain a sample with dissolved oxygen equal to 6 and we know the variance of the dissolved oxygen measurements, we can calculate how likely the observation is to have come from each group by calculating the value of the density under each model. Bayes theorem may be used to calculate the ratio in terms of the probability of the data and the prior probabilities. Use of Bayes theorem requires knowledge of the prior probabilities.

If the prior probabilities are taken to be equal the result is the likelihood ratio or Bayes Factor. The quantity measures the likelihood of the data given the site is in the impairment class versus the no impairment class. (More generally, the Bayes Factor also involves parameters that are treated as random and integrated out of calculations; see Kass and Raftery [1995].) For environmental problems there may not be simple approaches for estimating these quantities. Building a model for the impaired or unimpaired sites requires information on how data are distributed for these types of sites and other factors that might influence the observations. Sites classified as unimpaired are often viewed as reference sites. It may not be possible to obtain these sites or there may be covariates that must be considered. Calculation of the likelihood of the data under impairment requires a definition of the impairment or model of the data that we might expect if the observation came from the impairment group. One would have to have different models for different types of impairments (for example, chemical toxicity vs. sedimentation). The models should depend on the strength of the impairment, and may vary in space and time. These models may involve a good deal of work to describe. One approach to calculating the ratio given a lack of information on impairment is to calculate the odds as the ratio of the probability of the data treating the site as impaired relative to the probability of the data treating the site as not impaired.

WOE is a measure of how much an observed feature in the data adds to or subtracts from the evidence of impairment. Numerically, it has been defined (Good 1979) as the log (base 10) of the odds ratio. In our applications, this would correspond to the weight-of-evidence for one line of evidence. A natural consequence of using logs of ratios is that the weight-of-evidence from different lines may be added together to get an overall weight of evidence. Rules for interpreting Bayes Factors are given in Kass and Raftery (1995) and may be adjusted for WOE. These rules are guidelines in much the same way that p-values are guidelines and other authors have suggested alternative views (Good 1983). From a hypothesis testing perspective the weight-of-evidence measures the strength of the evidence against the null hypothesis.

Numerical calculation of WOE is not common to statistics. The reason is that testing of hypotheses and interpretation use the more common approach of likelihood ratio testing and calculation of Bayes Factors. Using the natural log scale and



twice the log of the Bayes Factor leads to the same scale as likelihood ratio testing in general statistical theory (where  $-2\log[\text{likelihood ratio}]$  is used to test hypotheses) and deviance measures in generalized linear models (McCullagh and Nelder 1989). Thus, for applications of WOE in environmental problems a user may choose to summarize results in terms of a WOE measure that is based on the probability of impairment. Alternatively the user may simply calculate and report the actual probability. The value of the use of Bayes Factors and odds is that these may be combined easily over the different lines of evidence.

The actual calculation of WOE for multiattribute environmental studies involves three general stages of analysis. We assume that the researcher has available the information required for making the decision. Thus, decisions have been made about what information needs to be collected or this information has already been collected. The three stages of the analysis are the data preprocessing stage, the processing stage and the combination of the information over the lines of evidence.

### Data Preprocessing

The initial step in the analysis is the preprocessing of the data. This step is required because a probability model is used to calculate the probability of impairment and we must check that the model is reasonable. Preprocessing involves selection of the variables to be used in the analysis, and scaling or transforming these variables. Variables are selected to provide relevant statistical and scientific information on differences between control and impairment. Scaling and transformation are often used to meet assumptions required for analysis. The assumptions needed depend on the model used to calculate the probability. Two methods for this calculation are to assume a model (parametric approach) or to calculate the probability using a nonparametric approach. In the parametric approach we select a probability model for the data. For example, a common model is the normal or Gaussian model. Then there are several assumptions that need to be evaluated for this statistical model to provide a good estimate:

1. Normality of the reference data.
2. Independence of samples in the reference set.
3. Homogeneity of variance in the reference set.

If the normal model is not reasonable then the estimate may be poor and misleading. Problems such as skewness and outliers may lead to inaccuracies in the probability estimate. As environmental data often are not normal, we try to achieve normality via choice of a suitable transformation of the variables or use a method based on the distribution of the data (*i.e.*, logistic regression assumes a binomial distribution). The logarithm is typically used as a transformation with contaminant concentrations. Independence of the reference data is required to provide a good estimate of the variances and covariances in the contaminants. This assumption is best met through choice of the reference locations and sampling occasion. Sites that are spatially close and repeated samples at the same site that are temporally close should be avoided. Homogeneity of the variance in the reference set is required to

12838



## Quantitative Weight-of-Evidence for Environmental Assessment

produce a good estimate of the variances and covariances. An alternative approach would be to allow heterogeneity but include this in the model in some manner. A potential concern here is with multiple sets of reference sites. If the multiple sets are treated as a single set then the estimates of the variances are likely to be smaller than data collected from different sites. Hence, detection of impairment is potentially overly sensitive. A useful strategy with multiple samples from a collection of sites is to try to match the test site with similar reference sites rather than to use all of the sites. This might involve forming clusters of reference sites and matching the test site with a cluster or using an auxiliary set of measurements (such as sediment type). Other potential problems include measuring a single site multiple times and the time of sampling. Selecting reference sites is a difficult and important problem.

The data transformation need not be a transformation of individual contaminants but may also be on the set of measurements. For example, it is common to analyze composites of variables rather than individual variables. Two common approaches are to use principal components (PC) or correspondence analysis (CA) to form new variables. These two methods are useful when the dimensionality of variable space is high and variables are highly correlated. In the case of the sediment quality triad, the PC transformation would typically be applied to the sediment toxicity and metal chemical variables, while correspondence analysis axes can be used for species composition data. Another possible data transformation could be computing some univariate index. For example, composition data can be represented by diversity measures or an index of biological integrity.

When the normal assumption is not valid a possible approach is to use a distance measure and build a nonparametric estimate of the probability of impairment using a jackknife like approach (Dixon 1993). An empirical distribution of distance to reference is computed by removing one reference set of measurements, calculating distance then replacing the measurements. If repeated for all observations in the reference set, a distribution may be estimated and used to evaluate a new set of measurements. For this approach to work we have similar assumptions:

1. The reference site data are from a common distribution
2. The reference sites are independent
3. The distance measure is appropriate for detecting change.

If the data for the reference sites come from a common distribution, then a single distance measure will produce reasonable estimates of how similar the test site is to the reference sites. When there are different sets of reference conditions the distance measure would have to be computed with respect to the different distributions or with respect to the set of reference conditions most similar to the test site. Independence of samples implies that equal weight may be given to each of the samples from the reference sites. The choice of distance measure is a critical step as the distance measure defines the measure of impairment. One important consideration in the selection of the distance measure is the weight given to the variables used in computing the distance (Smith 1998).

The choice of how to preprocess the information is critical to the analysis as it defines the deviations that are of interest. One should be aware of the limitations

12840

associated with these choices. For example, if information on a large number of variables is collected then one has a better chance of detecting a broad scale impairment. If the impairment is only observed through one variable, the other information becomes of low utility for the detection of impairment. Thus there is a need to have a clear idea of what types of impairment are to be detected. For example, if a chemical is only toxic to fish then measuring abundance of benthic macroinvertebrates will not be risk informative. It is critical in selecting an approach to be aware of what types of changes will or will not be well detected in the analysis and how likely the analysis is to detect changes of important magnitude. Methods such as power analysis are useful for evaluating variables and their importance in the decision process. Also, transformation of the variables is often needed. For example, chemical data are often collected in environmental studies. The user needs to decide if the original or standardized data are used. If standardized the method of standardization needs to be chosen. Options might include an overall standardization, standardization relative to a reference group or standardization in terms of toxic units. Choice of standardization will change the magnitude of distances between observations.

#### **Data Processing: Estimating Probability of Impairment**

In the data processing step, the test site is compared with reference conditions in order to obtain a measure of the degree of impairment. This may be an indirect or direct evaluation. For example, with biotic indices (Smoger and Angermeier 1999), there is often a calibration step in which the metrics that make up the index are scaled based on reference conditions. This scaling is an indirect estimation of the distance to the reference condition. A direct evaluation is based on numerically comparing the reference and test measurements. A statistical approach is to compute the probability of impairment through the use of the distance between reference and test measurements. The use of distance in ecological and environmental impairment assessment has a long history that will not be explored here. For example, distance from control forms the basis of outlier detection methods and control chart approaches that view water quality analysis as a quality control problem (Gilbert 1987). Distance also is central to many multivariate methods used to assess ecological change such as correspondence analysis (using chi-square distance, Legendre and Legendre 1998) or multidimensional scaling (Smith *et al.* 1990; Gray *et al.* 1990).

#### **Combining Estimates**

Given estimates of the probability of impairment for each line of evidence the problem now becomes combining these together to produce a single weight-of-evidence estimate. We again assume that the estimate of impairment is based on the reference conditions. There are several options available for making the combined estimate. Two approaches involve combining the probabilities for the lines and combining the odds for each of the lines. There are several possibilities for combining information across the different lines by combining the probabilities. These include using the average probability, the maximum probability, or the product. An alternative approach is to combine the odds. The odds are typically multiplied



## Quantitative Weight-of-Evidence for Environmental Assessment

together or the logarithm of the odds added. From this combined estimate, it is also possible to calculate the combined probability.

### EXAMPLE

As an example we consider data from the Great Lakes that were obtained via the BEAST software (Reynoldson *et al.* 1998, 2000). The reference data consisted of 146 reference samples collected in 1992. Although more data are available on different lines of evidence, these data have information from all three lines of evidence. In addition, there were 25 samples taken from Collingwood Harbour that we use as the predictive or test sample. These 25 samples were taken at nine locations in 1992, 1995, and 1997. Collingwood Harbour is located in the south shore of Nottawasaga Bay, in the southern extension of Lake Huron's Georgian Bay. The site is of interest as it was identified as an Area of Concern (AOC) by the International Joint Commission but was then de-listed in November 1994, following remediation (for details see <http://www.on.ec.gc.ca/glimr/raps/huron/collingwood/intro.html>). Contaminants in the sediment resulted from use of the harbor as a location for ship repair with greatest contamination near the shipyard and in the east and west slips. The nine sampling locations are located as follows: 6703, 6704, and 6705 are located in the harbor with 6703 being farthest from the shipyard and 6705 closest. Sites 6706-6708 are located in the east slip and sites 6709-6711 are located in the west slip. For a map of the locations and additional details on remediation history see <http://www.ijc.org/boards/wqb/cases/collingwood/collingwood.html>.

### Chemical Data

Graphical displays of the chemical data for the reference sites suggested the data were not normal. Distributional plots suggested skewness of the distributions and odd observations. The chemical data were preprocessed using a log transformation for all variables. Figure 1 displays the scatterplot matrix for the transformed data. The probability of impairment was calculated using Mahalanobis distance between the mean of the reference measurements and the measurements for the site (Rencher 1995). Although probability can be computed using a multivariate normal distribution the nonparametric method was used to calculate an empirical distribution from which we calculate the impairment probability for a test site.

### Biological Data

The probability of impairment for the biological data was computed by first calculating new axes using correspondence analysis with the reference data then scoring new sites on these axes. Distances were computed using the scores. Three axes were used in the computations as these produced stable estimates of the probability.

### Toxicity Data

Plots of the toxicity data suggested skewness of the measurements. As with the chemical data, the log transformation greatly reduced the skewness. Empirical distances were used to calculate the probabilities of impairment.



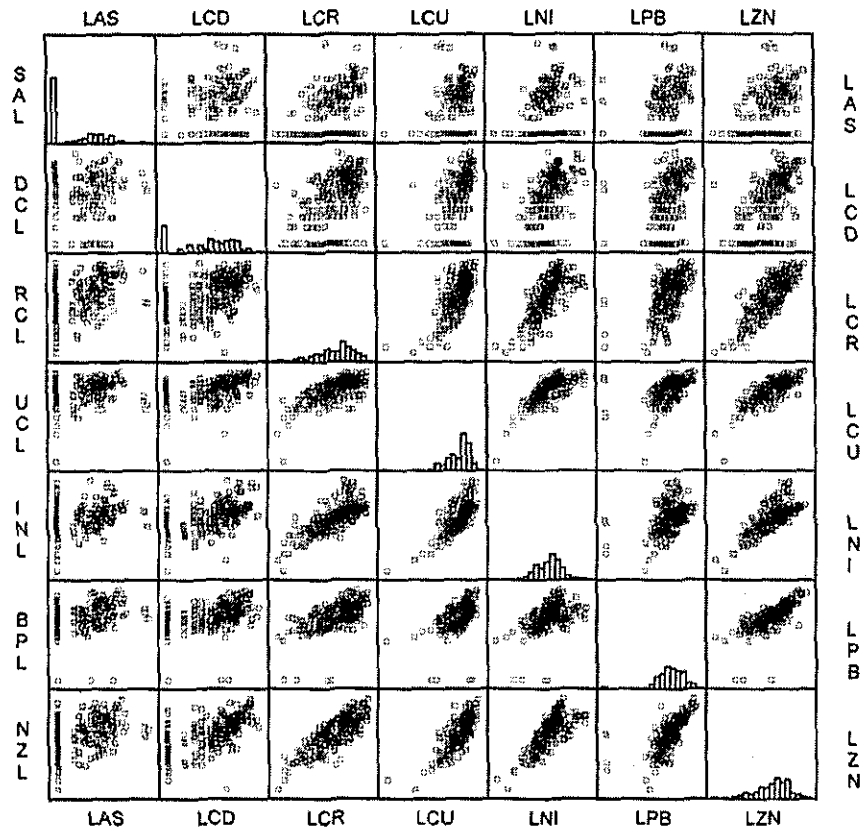


Figure 1. Scatterplot matrix of log transformed chemical reference measurements. The following abbreviations are used: LAS=log (arsenic), LCD=log (cadmium), LCR=log (chromium), LCU=log (copper), LNI=log (nickel), LPB=log (lead), and LZN=log (zinc).



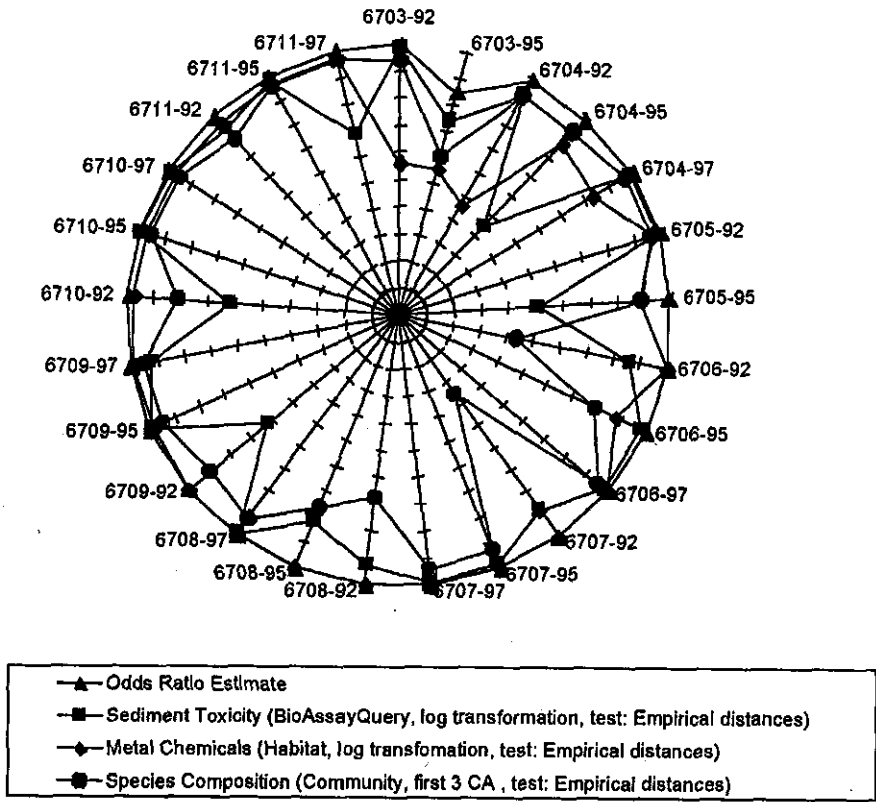


**Combined Estimates**

Figure 2 displays the overall impairment probabilities as well as estimates from individual lines. The probability of impairment is generally high. Individual impairment probabilities are highest for metals. Although there is evidence of biological impairment at some of the sites, the biological impairment probabilities are not generally as high as for metals.

**DISCUSSION**

We have presented an approach for estimation of the probability or risk of impairment for a site based on multiple lines of evidence. Many variations on the approach are possible based on different distances and different summarization of the information. Decisions about distance measures and summarization are best



**Figure 2.** Plot of probabilities for three lines of evidence and overall estimate for sites in test data set. The center corresponds to zero and tick marks represent tenths. The first four numbers in the site/sample label indicate the site number while the last two provide the year of sampling.



made in the initial stages of the study. Although the results of the weight-of-evidence analysis summarize impairment in terms of a single number, this approach is generally restrictive. A single measure attempts to summarize the multivariate degree of impairment. It is certainly possible that several different scenarios will lead to the same or similar measure of impairment. Hence it is important to consider the individual lines of evidence as well as the data themselves. Graphical display of the data is necessary to check for problems and assess assumptions. Biological and environmental evaluation of the collected data is also necessary to verify that the evaluation is scientifically correct as well as statistically valid. The results given here are summaries of the different components of each line of evidence. A further, valuable component of an analysis would be to study which components were important to the individual line of evidence and why. For example, the ten toxicity tests used in the above analysis are all not equally important to the impairment estimate for toxicity. One approach would be to remove individual components then to evaluate the effect on the impairment probability. In this way the components may be evaluated in terms of importance to the probability estimate and give clues as to why the site might be impaired (see for an example Smith *et al.* 1990).

We envision the above approach to be most useful for comparing and ranking sites. As Figure 2 illustrates, it is possible to display the estimates of impairment for different lines and a combined estimate for several sites/times in a single display. The information plotted may be ordered in time or space to look for change in impairment probabilities. This might be useful for restoration/recovery studies. The information for different sites may also be compared to identify sites with greatest risk or trends in impairment. For example, if the sites were located along a toxic gradient, then one would expect to see increases in the toxicity estimate of impairment and this could be displayed on the diagram.

Although we have focused on a statistical approach based on estimation of the probability of impairment, other approaches are available for obtaining a combined estimate of impairment. One common approach is an index-oriented method. In this approach, the numerical values are combined, possibly after a transformation. One example of this approach is Wildhaber and Schmitt (1996). They combine data over biological, toxicological and chemical lines. To do this the chemistry data are standardized by calculating the ratio of the bioavailable component of the contaminant to the chronic toxicity water quality criterion. The toxicological data are standardized by adjusting the test endpoint for the control endpoint. Values may then be averaged over each of the lines of evidence. The biological data are not evaluated in that manner, rather tolerance values are used and a biological index is computed as a tolerance weighted average. The values are then combined over the three lines of evidence. A variation on this approach is given in Soucek *et al.* (2000) and Cherry *et al.* (2001). In these papers the data are not combined within each line of evidence. Rather, important metrics from each line are selected then the metrics are combined into an overall index.

An important aspect of the analysis is the choice of distance measure. Our approach is to use a distance measure that is directly related to a probability distribution. It is therefore important that the assumptions of the distribution be checked so the distance measure is directly related to probability. It is also important that the distance measure reflects impairment in that the farther away from the

12850

## Quantitative Weight-of-Evidence for Environmental Assessment

center, the more impaired the site. One approach to achieve this is to base distance on ordinated data rather than on actual observations. For example, when dealing with chemical data an approach is to calculate principal components for the data from the reference sites. Then decide if the location of the test site is indicative of impairment. A distance measure that reflects a directional distance would be appropriate. Distance might be calculated as zero if the site is on the safe side of the distribution and the ordinary distance measure used if the site is located on the not-safe side. Principal components would be computed based on standardized data to give all the chemicals equal weight in the derivation of the components. Addition details and examples are presented at [www.stat.vt.edu/facstaff/epsmith](http://www.stat.vt.edu/facstaff/epsmith).

### ACKNOWLEDGMENTS

This research was supported in part by funds from the International Lead Zinc Research Organization, Inc. Additional support was provided by a U.S. Environmental Protection Agency's Science to Achieve Results (STAR) grant (No. R82795301). Although the research described in the article has been funded wholly or in part by the U.S. Environmental Protection Agency's STAR programs, it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. The authors are thankful for encouragement from Allen Burton and for comments from Peter Chapman, Valerie Forbes, and Trefor Reynoldson on previous drafts. Their struggles to get statistical minds to think environmentally were greatly appreciated.

### REFERENCES

- Chapman PM. 1996. Presentation and interpretation of Sediment Quality Triad data. *Ecotoxicology* 5:327-39
- Cherry DS, Currie RJ, Soucek DJ, *et al.* 2001. An integrative assessment of a watershed impaired by abandoned mined land discharges. *Environ Poll* 111:377-88
- Dixon PM. 1993. The bootstrap and jackknife: Describing the precision of ecological indices. In: Scheiner SM and Gurevitch J (eds), *Design and Analysis of Ecological Experiments*, pp 290-318. Chapman and Hall, NY, NY, USA
- Gelman A, Carlin JB, Stern HS, *et al.* 1995. *Bayesian Data Analysis*. Chapman and Hall, London, UK
- Gilbert RO. 1987. *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York, NY, USA
- Good IJ. 1979. Studies in the history of probability and statistics, XXXVII. A.M. Turing's statistical work in World War II. *Biometrika* 66:393-6
- Good IJ. 1983. *Good Thinking: The Foundations of Probability and Its Application*. University of Minnesota Press, Minneapolis, MN, USA
- Good IJ. 1985. Weight of evidence: A brief survey. In: Bernardo JM, DeGroot MH, Lindley DV, *et al.* (eds), *Bayesian Statistics 2*, pp 249-70. Elsevier Science Publishers BV, North Holland, Amsterdam, The Netherlands
- Good IJ. 1988. Statistical evidence. In: Kotz S and Johnson NL (eds), *Encyclopedia of Statistics* vol. 8, pp 651-5. John Wiley and Sons, NY, NY, USA
- Gray JS, Clarke KR, Warwick RM, *et al.* 1990. Detection of initial effects of pollution on marine benthos: An example from the Ekofisk and Eldfisk oilfields, North Sea. *Mar Ecol Prog Ser* 66:285-99

12852

- Kass RE and Raftery AE. 1995. Bayes factors. *J Am Stat Assoc* 90:773-95
- Legendre P and Legendre L. 1990. *Numerical Ecology*, 2<sup>nd</sup> Edition. Elsevier, Amsterdam, The Netherlands
- McCullagh P and Nelder JA. 1989. *Generalized Linear Models*, 2<sup>nd</sup> Edition. Chapman and Hall, London, UK
- Rencher A. 1995. *Methods of Multivariate Analysis*. John Wiley & Sons, NY, NY, USA
- Reynoldson TF and Day KE. 1998. *Biological Guidelines for the Assessment of Sediment Quality in the Laurentian Great Lakes*. NWRI Report 98-232. Dept of Fisheries and Oceans Canada, Burlington, ON, Canada
- Reynoldson TB, Day KE, and Pascoe T. 2000. The Development of BEAST: A Predictive Approach for Assessing Sediment Quality in the North American Great Lakes. In: Wright JF, Sutcliffe DW, and Furse MT (eds), *Assessing the Quality of Freshwaters: RIVPACS and Other Techniques*, Chapter 11, pp 165-80. Freshwater Biological Association, Ambleside, UK
- Smith EP. 1998. Randomization methods and the analysis of multivariate ecological data. *Environmetrics* 9:37-51
- Smith EP, Pontasch KW, and Cairns J Jr. 1990. Community similarity and the analysis of multispecies environmental data: A unified statistical approach. *Water Res* 24:507-14
- Smogor RA and Angermeier, PL. 1999. Relations Between Fish Metrics and Measures of Anthropogenic Disturbance in Three IBI<sup>®</sup> Regions in Virginia. In: Simon TP (ed), *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities*, pp 585-610. CRC Press, Boca Raton, FL, USA
- Soucek DJ, Cherry DS, Currie RJ, *et al* 2000. Laboratory to field validation in an integrative assessment of an acid mine drainage-impaired watershed. *Environ Toxicol Chem* 19:1036-43
- Wildhaber MW and Schmitt CJ. 1996. Hazard ranking of contaminated sediments based on chemical analysis, laboratory toxicity tests, and benthic community composition: Prioritizing sites for remedial action. *J Great Lakes Res* 22:639-52

