

## An index of compositional dissimilarity between observed and expected assemblages

John Van Sickle<sup>1</sup>

National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency,  
200 SW 35<sup>th</sup> St., Corvallis, Oregon 97333 USA

**Abstract.** The reference-condition approach to bioassessment often uses the observed/expected ( $O/E$ ) ratio to indicate anthropogenic alteration of aquatic macroinvertebrates, fish, or periphyton assemblages. Given a list of taxa found at  $\geq 1$  minimally disturbed reference sites,  $E$  is the number of those taxa that would be expected in a sampled assemblage if the sampled stream were in reference condition, and  $O$  is the number of those taxa observed in the sample. An  $O/E$  value significantly  $< 1.0$  indicates that a stream has lost taxa relative to its reference-condition expectation, possibly because of anthropogenic stress. However, the  $O/E$  index can be insensitive to stress-induced shifts in taxonomic composition that leave assemblage richness unchanged. As an alternative to  $O/E$ , I propose using  $BC$ , an adaptation of Bray–Curtis distance, to measure the compositional dissimilarity between an observed and expected assemblage directly. I compared the performance of  $BC$  and  $O/E$  at 5685 stream and lake sites throughout the contiguous 48 states of the US using 1 of 10 River Invertebrate Prediction and Classification system (RIVPACS)-type models to predict expected assemblages. Percentages of independently determined nonreference sites that were declared to be nonreference by  $BC$  exceeded the percentage declared to be nonreference by  $O/E$  by an average of 6 to 16 percentage points, depending on whether the 2 indices included low-probability taxa, whether a 1-sided or 2-sided  $O/E$  criterion was used to declare nonreference, and whether predictive or null models were used to predict expected assemblages. Correlations between  $BC$  scores and anthropogenic stressor variables were stronger than correlations between  $O/E$  scores and anthropogenic stressor variables in 18 of 25 cases. In contrast to  $O/E$ ,  $BC$  can include low-probability taxa without reducing its power to detect nonreference conditions.

**Key words:** predictive model, bioassessment,  $O/E$ , reference condition, macroinvertebrate assemblage.

A widely used approach to aquatic bioassessment compares a sampled assemblage of macroinvertebrates, fish, or periphyton to the assemblage that one would expect to find if the sampled ecosystem were in a minimally disturbed reference condition (Reynoldson et al. 1997, Bailey et al. 2004, Stoddard et al. 2006). In many applications of this approach, a River Invertebrate Prediction and Classification system (RIVPACS)-type model (Moss et al. 1987, Wright et al. 1993) is used to predict the expected assemblage at a sampling location (site) in the form of a set of occurrence probabilities, one for each taxon (Wright 1995, Parsons and Norris 1996, Hawkins et al. 2000, Joy and Death 2002, Johnson 2003).  $E$  is the number of taxa in the expected assemblage and is equal to the sum (across taxa) of predicted occurrence probabilities, and  $O$  is the number of taxa observed

in the sample. The  $O/E$  ratio of observed to expected taxon richness is used to compare observed and expected assemblages, and a value significantly different from 1.0 indicates that the biota at a site is not in reference condition (Wright 2000).  $O/E$  is a popular bioassessment index partly because of its simplicity and clear interpretation.  $O/E$  measures the taxonomic completeness of the fauna at a site in terms of the number of taxa lost or gained relative to reference conditions (Moss et al. 1987, Wright 2000, Hawkins 2006).

However, most applications of  $O/E$  provide only a partial comparison of observed and expected assemblages in 2 ways. First, the  $O/E$  index is based only on the numbers of taxa in the 2 assemblages, so it might not capture taxon-specific disagreements between them. Recognizing this limitation, Clarke et al. (1996) explored several statistically based indices of taxon-specific disagreement, such as the likelihood

<sup>1</sup> E-mail address: vansickle.john@epa.gov

TABLE 1. Calculation of the ratio of observed/expected ( $O/E$ ) taxa at a site and mean absolute difference for hypothetical observed ( $O_k$ ) and expected ( $P_k$ ) assemblages involving 5 reference-site taxa.

Taxon ( $k$ )	$O_k$	$P_k$	$ O_k - P_k $
1	1	0.1	0.9
2	1	0.3	0.7
3	1	0.8	0.2
4	0	0.9	0.9
5	0	0.9	0.9
$O = \text{sum} = 3$		$E = \text{sum} = 3$	
			Mean = 0.72

that the taxonomic composition (i.e., the taxon list) of the observed assemblage could have been generated by a realization of the expected occurrence probabilities (see also Oberdorff et al. 2001). However, the indices proposed by Clarke et al. (1996) have seen little use in bioassessments, possibly because of interpretation difficulties. Second, investigators often exclude the low-probability taxa at a site (those with predicted occurrence probabilities  $<0.5$ ) when calculating that site's  $O/E$  score (e.g., Hawkins et al. 2000, Simpson and Norris 2000, Hawkins and Carlisle 2001, Van Sickle et al. 2005). Recent research has shown that such exclusions often increase, and do not decrease, the sensitivity of  $O/E$  for detecting non-reference conditions (Ostermiller and Hawkins 2004, Clarke and Murphy 2006, Van Sickle et al. 2007).

I propose a measure of compositional dissimilarity that complements  $O/E$  and summarizes the taxon-specific disparities between observed and expected assemblages. Like  $O/E$ , the dissimilarity measure is a bioassessment index that can be calculated directly from the output of any RIVPACS-type model. I begin with a numerical example in which  $O/E$  fails to represent taxon-specific disparities. The example also shows that informative contributions to  $O/E$  from low-probability and high-probability taxa partially cancel each other, which explains why  $O/E$  is often more sensitive when low-probability taxa are excluded from its calculation than when all taxa are included. Such cancellations do not occur with the dissimilarity index proposed here, and I argue that inclusion of low-probability taxa does not reduce its sensitivity. Last, I compare the performances of  $O/E$  and the dissimilarity index across 10 different RIVPACS-type models applied to a total of 5685 stream and lake assemblages sampled throughout the contiguous US. I assess the performance of each index according to its ability to detect nonreference conditions and its correlations with anthropogenic stressor variables.

### Indices for Comparing Observed and Expected Assemblages

#### $O/E$ index

RIVPACS-type models predict occurrence probabilities,  $P_k$ ,  $k = 1, 2, 3, \dots, M$ , at a site for  $M$  reference-site taxa, i.e., for the list of taxa found at  $\geq 1$  of the sampled reference sites used to calibrate the model. The taxonomic composition of the corresponding observed assemblage at the site can be defined by  $O_k$ ,  $k = 1, 2, 3, \dots, M$ , where  $O_k = 0$  or 1 if reference-site taxon  $k$  was observed to be absent or present, respectively, at the site. The  $O/E$  index is given by Clarke et al. (1996) as

$$\frac{O}{E} = \frac{\sum O_k}{\sum P_k} \tag{1}$$

Table 1 presents hypothetical observed and expected assemblages for a very small assemblage (5 reference-site taxa). In practice,  $O/E$  should not be applied to such a small assemblage because the index would be extremely vulnerable to random sampling variations in the occurrences of only 1 or 2 taxa. However, the table does illustrate how  $O/E$  can fail to represent disagreements between observed and expected assemblages. In Table 1,  $O = E = 3$ , yielding  $O/E = 1.0$ . This  $O/E$  score would imply that the observed assemblage is in reference condition. However, taxa 1 and 2 were each predicted to be unlikely to occur at the site (low probability;  $P_k < 0.5$ ), whereas both were observed to be present (Table 1). In addition, taxa 4 and 5 were predicted to be likely to occur (high probability;  $P_k \geq 0.5$ ), whereas both were observed to be absent. These major disparities are not captured by  $O/E$  because contributions to  $O/E$  from taxa 1 and 2 partially cancel out the contributions from taxa 4 and 5. For example, taxon 1 adds 1 to the numerator of  $O/E$  and 0.1 to its denominator (equation 1), resulting in an increase of  $O/E$ . In contrast, taxon 5 adds 0 to the numerator of  $O/E$  and 0.9 to its denominator, resulting in a net decrease of  $O/E$ . Contributions from these 2 taxa drive  $O/E$  in opposing directions, with the result of a partial cancellation of their contributions to the index. The example in Table 1 suggests that  $O/E$  will be relatively insensitive to stress-induced shifts in taxonomic composition that have little net effect on the number of reference-site taxa (Hawkins et al. 2000, Davies and Jackson 2006).

If low-probability taxa are excluded when calculating  $O/E$ , then they cannot cancel the contributions from high-probability taxa (Hawkins et al. 2000). In addition, high-probability taxa respond more strongly and consistently to nonreference conditions, in terms of their occurrence frequencies over a large number of

TABLE 2. Ten River Invertebrate Prediction and Classification system (RIVPACS)-type predictive models for freshwater ecosystems, showing the number of reference sites used to develop and calibrate each model and the number of nonreference sites to which each model was applied. All models are for aquatic macroinvertebrates unless noted otherwise.

Model name (abbreviation)	Reference sites	Nonreference sites	Source
Mid-Atlantic Highlands (MH)	72	432	Van Sickle et al. (2005)
Colorado (CO)	46	88	C. P. Hawkins (Utah State University, personal communication)
Oregon (OR)	133	460	Van Sickle et al. (2006)
California (CA)	185	194	Hawkins et al. (2000)
Wadeable Streams Assessment:			USEPA (2006)
Appalachians (AW)	217	264	
Plains (PW)	138	444	
West (WW)	518	557	
Ohio fish (OH)	114	1409	Hawkins (2006), de Zwart et al. (2006)
Sierra lakes mixed assemblages (SL)	164	70	Knapp et al. (2005)
Appalachian diatoms (AP)	52	128	Carlisle et al. (2008)
Total	1639	4046	

sites, than do the low-probability taxa (Van Sickle et al. 2007). Together, these 2 observations explain why the exclusion of low-probability taxa from  $O/E$  often increases its ability to differentiate between reference and nonreference conditions (Van Sickle et al. 2007). For example, the exclusion of taxa 1 and 2 in Table 1 results in  $O/E = 0.38$ , a value that more accurately reflects the strong disagreement between observed and expected assemblages for 4 of the 5 taxa. Hawkins et al. (2000) also discussed the opposing contributions of low- and high-probability taxa to  $O/E$  in terms of taxon abundances.

#### BC index

Taxon-specific disagreements can be measured by the absolute value of the difference between each  $O_k$  value and its corresponding  $P_k$  value (Table 1). For the example in Table 1, the mean of these absolute differences is 0.72, which clearly reflects (on a scale of 0 to 1) the strong disagreement seen for 4 of the 5 taxa. Note that a low-probability taxon observed to be present (taxon 1) and a high-probability taxon observed to be absent (taxon 5) both contribute positively and equally to the sum or mean of absolute differences (Table 1).

Initial trials indicated that the mean or sum of absolute differences did not perform any better than  $O/E$  in terms of sensitivity and stressor correlations. However, standardizing the sum of absolute differences relative to the full set of  $O_k$  values and  $P_k$  values yielded an index with markedly improved performance:

$$BC = \frac{\sum |O_k - P_k|}{\sum (O_k + P_k)} \quad [2]$$

I call this the  $BC$  index because equation 2 is the formula for Bray–Curtis dissimilarity adapted to taxonomic composition comparisons of observed and expected assemblages (Bray and Curtis 1957, Faith et al. 1987, Legendre and Legendre 1998). Bray–Curtis dissimilarity has been used widely in community ecology, and its applications have included the creation of reference-site groups for RIVPACS-type models (e.g., Parsons and Norris 1996, Hawkins et al. 2000). However, known properties of Bray–Curtis dissimilarity should not be attributed casually to  $BC$  because some of those properties assume that dissimilarity is calculated from taxon abundances rather than  $O_k$  and  $P_k$  values (Clifford and Stephenson 1975, Faith et al. 1987, Clarke et al. 2006). The  $BC$  index ranges from 0 to 1, and a low  $BC$  value indicates that the observed assemblage has a taxonomic composition similar to its reference-condition expectation.

#### Comparative Performance of $O/E$ and $BC$

##### Methods

*Data, models, and taxon subsets.*—I compared the performance of  $O/E$  and  $BC$  for 10 RIVPACS-type models of freshwater systems sampled within the contiguous US (Table 2). The 10 models were developed from a total of 1639 reference-site assemblages (1 assemblage per sampled stream or lake site), following methods described by Hawkins et al. (2000), Hawkins and Carlisle (2001), and Van Sickle et al. (2005, 2006). The appropriate model was then applied to each of 4046 additional sites, all of which were independently determined to be in nonreference condition based on factors such as water chemistry and physical habitat (Van Sickle et al. 2007).

Three models apply to stream macroinvertebrates in the Mid-Atlantic Highlands region (MH model), which extends from Virginia to New York, in the state of Oregon (OR), or in forested mountain areas of northern and central California (CA) (Table 2). Additional models apply to stream macroinvertebrates in the Rocky Mountains of Colorado (CO), fish in Ohio streams (OH), diatoms in streams of the Appalachian Mountains (AP), and mixed assemblages of amphibians, reptiles, benthic macroinvertebrates, and zooplankton in lakes of the Sierra Mountains in California (SL). Last, 3 models apply to stream macroinvertebrates in 3 regions covering the contiguous US (Appalachians [AW], plains [PW], and West [WW]) based on assemblages sampled between 2000 and 2004 during the US Environmental Protection Agency's Wadeable Stream Assessment project. Details of sampling methods, taxonomic identifications, determination of reference/nonreference status, and model structures are given in Van Sickle et al. (2007), and the sources are listed in Table 2.

I calculated 2 versions of *BC* and *O/E*. One version included only the subset of high-probability taxa ( $P_k \geq 0.5$ ), and the other included all modeled taxa ( $P_k > 0$ ). In addition, I calculated each index from the expected assemblages predicted by RIVPACS-type models and from the expected assemblages predicted by corresponding null models (Van Sickle et al. 2005). The null model predicts a single, fixed expected assemblage for all sites. Thus, the performance features of null-model *BC* and *O/E* estimates cannot be artifacts of the multivariate methods (clustering and discriminant analysis) used by RIVPACS-type models.

Some, but not all, of the 10 models were developed and calibrated using a subset of the reference sites, where the remainder was set aside for model validation. For consistency across models, I used only the calibration sites to estimate reference-site distributions and statistics of *O/E* and *BC*. I also excluded any nonreference site that required model extrapolation to accommodate its set of predictor variable values. Extrapolation was required if a  $\chi^2$  test of the distance between the site's predictor values and centroids of the reference site groups had a *p* value  $< 0.01$  (Moss et al. 1987, Clarke et al. 2003).

*Index performance: sensitivity.*—For each of the 10 models, I estimated the sensitivities of *O/E* and *BC*, i.e., the ability of each index to detect nonreference conditions (Ostermiller and Hawkins 2004). I measured sensitivity by the percentage of test-site scores that lay outside the distribution of reference-site scores (Ostermiller and Hawkins 2004, Van Sickle et al. 2005). Reference-site *O/E* scores generally are normally distributed with a mean near 1.0 (Wright 1995,

Hawkins et al. 2000). However, normal probability plots (not shown) suggested that reference distributions of *BC* scores were positively skewed for 7 of the 10 models. Therefore, I used empirical quantiles of the *BC* and *O/E* reference distributions as thresholds for nonreference conditions.

I assumed that a test site's *BC* score indicated nonreference conditions if it exceeded the 90<sup>th</sup> percentile of the reference *BC* distribution. This 1-sided criterion is appropriate because only a large *BC* score can indicate a departure from reference condition. I estimated sensitivities for *O/E* based on 2-sided and 1-sided criteria because both have been invoked in prior applications of *O/E*. Under a 2-sided *O/E* criterion (e.g., Clarke et al. 2003, Ostermiller and Hawkins 2004), I defined sensitivity as the percentage of test sites with a score  $\leq 5^{\text{th}}$  percentile or  $\geq 95^{\text{th}}$  percentile of the reference distribution. A 1-sided *O/E* criterion is appropriate if only the loss of taxa ( $O/E < 1.0$ ) is taken to indicate stress or impairment (Wright 1995). I used the 10<sup>th</sup> percentile of the reference distribution as the 1-sided *O/E* criterion. The 1-sided *BC* criterion and the 2-sided and 1-sided *O/E* criteria assume that nonreference scores lie outside a range of 90% of the reference scores.

*Index performance: correlations with stressor variables.*—For sites covered by 5 of the 10 models in Table 2, I correlated *BC* and *O/E* scores with water-chemistry, physical-habitat, and watershed landuse variables sampled at the same sites and times as assemblages. The 5 models all apply to stream macroinvertebrate assemblages, and for each model, I selected 5 environmental variables that could represent anthropogenic stresses to the benthic community (Table 3). For the Colorado model (CO), I chose metals-contamination and sedimentation variables that reflect potential mining impacts on streams (C. P. Hawkins, Utah State University, personal communication; Table 3). For the other 4 models (MH, AW, PW, WW), I chose variables representing potential water-quality, sedimentation, physical-habitat, and watershed landuse stresses. I graphically compared the strengths of *BC* and *O/E* correlations with these environmental variables without identifying which variables had the strongest correlations. Statistical inferences about differences between *BC* and *O/E* correlations were prohibited by lack of independence among the 5 correlations for any one model.

## Results

### *Sensitivity*

On average, *BC* sensitivity exceeded that of *O/E* regardless of whether the 2 indices were calculated

TABLE 3. Stream environmental variables correlated with *BC* and *O/E* indices. Correlations with *BC* and *O/E* were calculated separately for each of 5 models. Field and laboratory methods for environmental data are described by Lazorchak et al. (1998) and Kaufmann et al. (1999) for the MH, AW, PW, and WW models, and by C. P. Hawkins (Utah State University, personal communication) for the CO model.

Model (abbreviation)	Environmental variables
Mid-Atlantic (MH)	Rapid Bioassessment Protocol habitat quality, agricultural + urban land cover in watershed, $\text{NO}_3^-$ , $\text{Cl}^-$ , total suspended solids
Colorado (CO)	Dissolved Cu, dissolved Zn, specific conductivity, % substrate embeddedness, % sand and fine particles in substrate
Wadeable Stream Assessment models (AW, PW, WW)	Rapid Bioassessment Protocol habitat quality, watershed population density, total N, $\text{Cl}^-$ , % sand and fine particles in substrate

using  $P_k > 0$  or  $P_k \geq 0.5$ , whether predictive or null models were used to obtain expected assemblages, or whether a 2-sided or 1-sided criterion was used for *O/E* (Table 4). Estimated mean differences between *BC* and *O/E* sensitivity had wide confidence intervals because of the small sample size ( $n = 10$  models) and the control of family-wise confidence level (Table 4). Nevertheless, confidence intervals included a 0 mean difference in only 2 of the 8 cases. When *BC* and *O/E* were based on  $P_k \geq 0.5$ , *BC* sensitivity exceeded *O/E* sensitivity for 9 of 10 models by 1.2 to 28.6 percentage points (Fig. 1A), with a mean difference of 10.9 points (Table 4). When *BC* and *O/E* were based on  $P_k > 0$ , *BC* sensitivity exceeded *O/E* sensitivity by a mean difference of 15.7 percentage points (Table 4, Fig. 1B).

A linear regression of *BC* sensitivity when  $P_k > 0$  vs *BC* sensitivity when  $P_k \geq 0.5$  across the 10 models had an estimated slope of 0.90 (standard error [SE] = 0.13) and intercept of 2.25 (SE = 5.86) (Fig. 2). Simultaneous, Bonferroni-corrected, 95% confidence intervals (Myers 1990) for the intercept and slope of this regression included 0.0 and 1.0, respectively, indicating no significant difference from a 1:1 line. This result suggests that *BC* generally maintained the same level of sensitivity across a range of values from 25% to 70%, whether or not it included low-probability taxa. A similar regression for *O/E* sensitivity had an estimated

slope of 0.42 (SE = 0.08). Simultaneous confidence intervals for the slope and intercept of the *O/E* regression did not include 0.0 and 1.0, respectively, indicating that *O/E* based on  $P_k > 0$  generally lost more sensitivity relative to *O/E* based on  $P_k \geq 0.5$  as sensitivity increased (Fig. 2).

The sensitivities of the *BC* and *O/E* indices ranged between 16 and 68% depending on the data set and the set of included taxa (Fig. 1A, B). Thus, between 32 and 84% of sites that were independently declared to be nonreference had index values indicating that their assemblages were in reference condition. This result illustrates the somewhat noisy relationships between assemblage condition and the factors (water chemistry and habitat) that were used to independently estimate site condition.

#### Correlations with stressor variables

Within each of 5 models, the 5 chosen stressor variables (Table 3) generally had low Spearman correlations among themselves, indicating an overall low level of redundancy. Of a total of 50 such correlations,  $r = -0.77$  in 1 case,  $0.5 < r < 0.7$  in 6 cases, and  $r < 0.5$  in the remaining 43 cases.

When  $P_k \geq 0.5$ , the magnitudes of 18 of the 25 *BC* correlations with stressors exceeded the magnitudes of

TABLE 4. Mean difference between the sensitivities of an adaptation of the Bray–Curtis distance (*BC*) and the observed/expected taxa ratio (*O/E*) across 10 models, expressed in percentage points. Parentheses contain confidence intervals on the mean difference. Confidence intervals were Bonferroni-corrected to achieve a 90% confidence level for the family of 8 nonindependent intervals (Ramsey and Schafer 1997). Mean differences are given for the 1-sided *BC* percentage vs 2-sided or 1-sided *O/E* percentages, for expected assemblages obtained from either a predictive model or a null model, and for indices that include either all taxa ( $P_k > 0$ ) or only high-probability taxa ( $P_k \geq 0.5$ ).

Model type	Included taxa	<i>BC</i> – (2-sided <i>O/E</i> )	<i>BC</i> – (1-sided <i>O/E</i> )
Predictive	$P_k \geq 0.5$	10.9 (1.8–19.9)	3.9 ( 3.0–10.7)
Predictive	$P_k > 0$	15.7 (6.6–24.7)	13.9 (1.9–25.8)
Null	$P_k \geq 0.5$	10.3 (4.1–16.5)	5.6 (1.7–9.5)
Null	$P_k > 0$	10.2 (1.8–18.5)	9.1 ( 0.6–18.8)

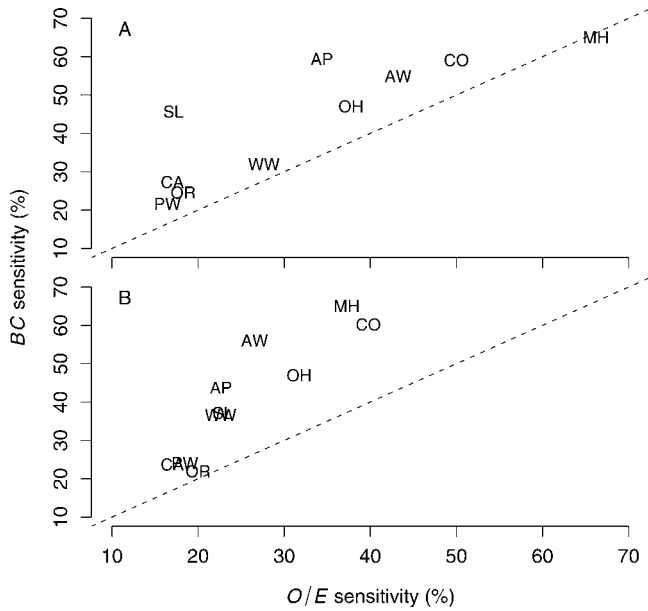


FIG. 1. Relationship between sensitivities of an adaptation of the Bray–Curtis distance (*BC*) and the observed/expected taxa ratio (*O/E*) for 10 models and data sets when indices were calculated with only high-probability taxa ( $P_k \geq 0.5$ ) (A) or all taxa ( $P_k > 0$ ) (B). *O/E* sensitivity was based on a 2-sided criterion, and expected assemblages were obtained from predictive models. Model abbreviations are given in Table 2. Dashes denote 1:1 line.

corresponding *O/E* correlations, although the largest exceedance was only 0.09 correlation units (Fig. 3A). However, when  $P_k > 0$ , the magnitudes of 21 of the 25 *BC* correlations exceeded the magnitudes of corresponding *O/E* correlations by as much as 0.37 units (Fig. 3B). All stressor correlations with either *BC* or *O/E* were  $< 0.6$ , indicating generally weak relationships. However, all correlations had the expected direction, i.e., *O/E* (*BC*) was positively (negatively) correlated with habitat quality and negatively (positively) correlated with all other stressors. All correlations were based on relatively large sample sizes (134 sites for the CO model and  $> 300$  sites for each of the other 4 models).

In 16 of 25 cases, correlations of *BC* with stressors were slightly weaker when  $P_k \geq 0.5$  than when  $P_k > 0$  (Fig. 4). For *O/E*, the opposite pattern was seen. In 19 of 25 cases, correlations of *O/E* with stressors were stronger (often substantially) when  $P_k \geq 0.5$  than when  $P_k > 0$ .

**Discussion**

*BC* showed consistently greater average sensitivity than did *O/E* across 10 RIVPACS-type models and corresponding null models applied to 4046 nonreference sites (Table 4, Fig. 1). The null model results suggest that greater *BC* sensitivity was not an artifact

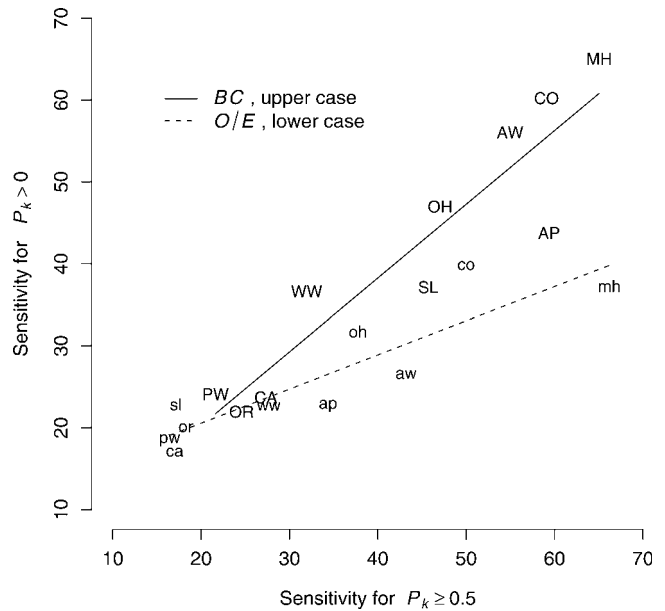


FIG. 2. Relationship between sensitivities of indices calculated with all taxa ( $P_k > 0$ ) and with only high-probability taxa ( $P_k \geq 0.5$ ) for 10 models and data sets. Indices were based on an adaptation of the Bray–Curtis distance (*BC*) and the observed/expected taxa ratio (*O/E*). Regression lines are drawn for *BC* sensitivities ( $r^2 = 0.86$ ) and for *O/E* 2-sided sensitivities ( $r^2 = 0.80$ ). Model abbreviations are given in Table 2.

of the clustering and discriminant function components of the RIVPACS-type models. Thus, I would expect *BC* to be more sensitive than *O/E* for any bioassessment model structure with an expected assemblage that takes the form of predicted taxon occurrence probabilities (e.g., Oberdorff et al. 2001, Joy and Death 2003, Linke et al. 2005). *BC* correlations with stressor variables also were stronger than *O/E* correlations in most cases (Fig. 3).

These results are consistent with the nature of the differences between observed and expected assemblages that are measured by *BC* vs *O/E*. *BC* will respond to differences in taxonomic composition, regardless of the difference in numbers of reference-site taxa. In contrast, *O/E* measures only the difference in reference-site taxon richness and can be relatively insensitive to changes in taxonomic composition. As a result, *BC*, but not *O/E*, will reflect the effects of certain stressors or lower stress levels that might alter the composition, but not the reference-site taxon richness, of an assemblage (Hawkins et al. 2000, Davies and Jackson 2006).

*BC* and *O/E* compare somewhat different attributes of observed and expected assemblages, so it might be useful to report both indices. In many cases, they would give similar assessments. For example, a high

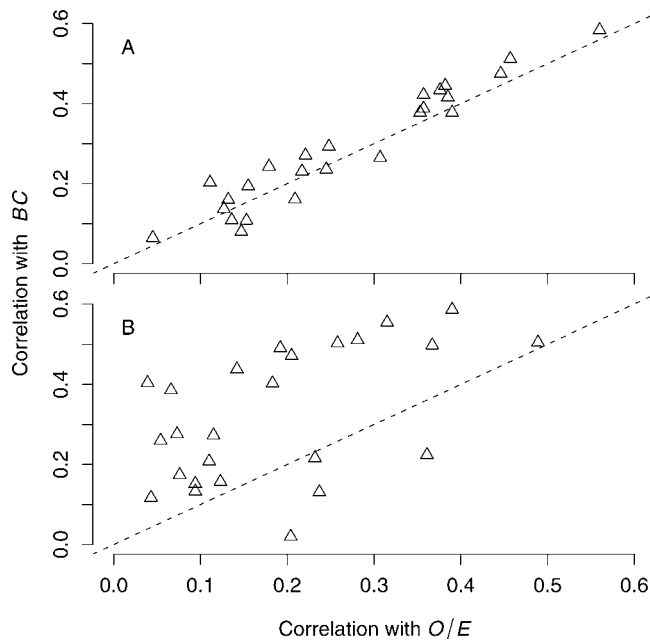


FIG. 3. Magnitudes of Spearman correlation coefficients with environmental stressor variables, for an adaptation of Bray–Curtis distance ( $BC$ ) and the observed/expected ratio ( $O/E$ ). Correlations of 5 stressor variables with  $BC$  and  $O/E$  from each of 5 models and data sets (MH, CO, AW, PW, WW) are shown, where  $BC$  and  $O/E$  were calculated for only the high-probability taxa ( $P_k \geq 0.5$ ) (A) or for all taxa ( $P_k > 0$ ) (B). Dashes denote 1:1 line. Model abbreviations are given in Table 3.

$BC$  value and  $O/E \ll 1.0$  both indicate substantial differences between 2 assemblages. However, in some cases,  $BC$  and  $O/E$  could disagree. For example, the combination of a high  $BC$  value and  $O/E \approx 1.0$  would indicate an observed assemblage with an altered taxonomic composition but little change in reference-taxon richness relative to the expected assemblage (e.g., the case in Table 1). Conversely, the combination of  $O/E < 1.0$  and a low  $BC$  within the reference range could occur when the observed assemblage reflects a net loss of several taxa but otherwise has observed taxon presences/absences that closely match the predicted reference-condition probabilities.

#### Implications

The use of  $BC$  has at least 4 implications for the future development and application of RIVPACS-type models:

1) *Revisiting the role of low-probability taxa.*—Contributions to  $O/E$  from low-probability taxa that are observed to be present at a site partially cancel out the contributions from high-probability taxa that are observed to be absent. In contrast, the corresponding

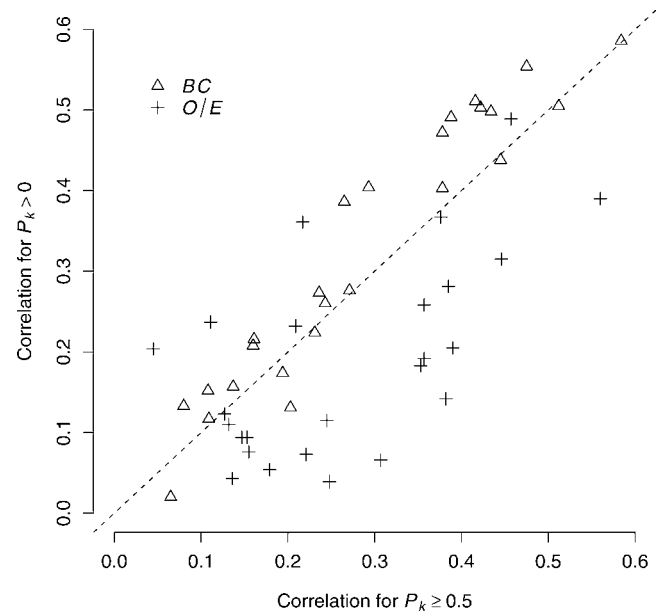


FIG. 4. Magnitudes of Spearman correlation coefficients with environmental stressor variables, for an adaptation of Bray–Curtis distance ( $BC$ ) and the observed/expected ratio ( $O/E$ ), where  $BC$  and  $O/E$  were calculated for only the high-probability taxa ( $P_k \geq 0.5$ ) or for all taxa ( $P_k > 0$ ). Correlations of 5 stressor variables with  $BC$  and  $O/E$  from each of 5 models and data sets (MH, CO, AW, PW, WW) are shown. Dashes denote 1:1 line. Model abbreviations are given in Table 3.

contributions to  $BC$  are positive and do not cancel each other. As a result,  $BC$  has slightly higher sensitivities to nonreference conditions in most cases (Fig. 2) when its calculation includes both groups of taxa ( $P_k > 0$ ). In the examples chosen, most correlations between  $BC$  and stressor variables were also slightly stronger for  $P_k > 0$  than for  $P_k \geq 0.5$  (Fig. 4).

These results add a new dimension to the debate about the utility of low-probability taxa in RIVPACS-type bioassessment models because that debate historically has focused on the behavior of  $O/E$  (e.g., Cao et al. 2001, Marchant 2002, Ostermiller and Hawkins 2004, Clarke and Murphy 2006, Van Sickle et al. 2007). Unlike  $O/E$ , the  $BC$  index will be sensitive to putative impairment signals from low-probability taxa. However, the strength and reliability of those signals remain questionable given the unpredictability of occurrences of the low-probability taxa and their lack of a strong group response to ecosystem stress (Van Sickle et al. 2007).

2) *Model evaluation.*—The distributional statistics of  $O/E$  scores from model calibration and validation sites have been used to assess the performance of a RIVPACS-type model prior to its application at new

sites (Clarke et al. 1996, Hawkins et al. 2000, Van Sickle et al. 2005, 2006). *BC* scores also would be appropriate for this purpose, particularly when *BC* itself will be used as a primary site-assessment index. I suggest using the 90<sup>th</sup> percentile of the *BC* reference distribution to evaluate model performance because that distribution might be skewed. Future research could reveal whether the *O/E* and *BC* measures of model quality tend to select the same best models (Van Sickle et al. 2006).

3) *Use of other dissimilarity measures.*—Measures of compositional dissimilarity other than *BC* could be adapted for comparisons between observed and expected assemblages. For example, the Canberra metric also sums absolute differences, but it uses a different standardization from that of *BC* (Legendre and Legendre 1998). As another example, the Euclidean distance measure would square the differences between  $O_k$  and  $P_k$ , which would increase the relative contributions from the very-low-probability taxa that are observed to be present and the very-high-probability taxa that are observed to be absent. These or other compositional indices can be screened quickly for appropriateness of their responses to key, taxon-specific discrepancies between the observed and expected assemblages, such as those illustrated by Table 1.

4) *Listing the assemblages.*—A list of the  $O_k$  values and corresponding  $P_k$  values at an individual site will reveal which taxa have large discrepancies in their predicted probabilities and observed occurrences (Wright 2000). Such lists are produced by the predictive modeling software available for the UK RIVPACS (<http://www.ceh.ac.uk/products/software/water.html>), for the Australian River Assessment System (AUSRIVAS; <http://ausriv.as.canberra.edu.au/>), and for the US Western Center for Monitoring and Assessment of Freshwater Ecosystems (<http://www.cnr.usu.edu/wmc>). Taxa in these lists usually are sorted either alphabetically or by their  $P_k$  values. As an alternative, I suggest sorting them by values of  $|O_k - P_k|$ . With this sorting, all taxa that show a strong disagreement between predicted probabilities and observed occurrences will be conveniently located at one end of the list, and all taxa that show strong agreement will be at the other end of the list.

### Acknowledgements

This research was funded by the US Environmental Protection Agency (EPA). I thank Chuck Hawkins, Daren Carlisle, and Lester Yuan for use of their models and data sets. I also thank Chuck Hawkins, Phil Larsen, Bruce Chessman, Bob Ozretich, and 2 anyo-

mous referees for valuable suggestions on the manuscript. The manuscript has been subjected to the EPA's peer and administrative review and has been approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

### Literature Cited

- BAILEY, R. C., R. H. NORRIS, AND T. B. REYNOLDS. 2004. Bioassessment of freshwater ecosystems: using the reference condition approach. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- BRAY, J. R., AND J. T. CURTIS. 1957. An ordination of the upland forest communities in southern Wisconsin. *Ecological Monographs* 27:325–349.
- CAO, Y., D. P. LARSEN, AND R. ST-J. THORNE. 2001. Rare species in multivariate analysis for bioassessment: some considerations. *Journal of the North American Benthological Society* 20:144–153.
- CARLISLE, D. M., C. P. HAWKINS, M. R. MEADOR, M. POTAPOVA, AND J. FALCONE. 2008. Biological assessments of Appalachian streams based on predictive models from fish, macroinvertebrate, and diatom assemblages. *Journal of the North American Benthological Society* 27:16–37.
- CLARKE, K. R., P. J. SOMERFIELD, AND M. G. CHAPMAN. 2006. On resemblance measures for ecological studies, including taxonomic dissimilarities and a zero-adjusted Bray–Curtis coefficient for denuded assemblages. *Journal of Experimental Marine Biology and Ecology* 330:55–80.
- CLARKE, R. T., M. T. FURSE, J. F. WRIGHT, AND D. MOSS. 1996. Derivation of a biological quality index for river sites: comparison of the observed with the expected fauna. *Journal of Applied Statistics* 23:311–332.
- CLARKE, R. T., AND J. F. MURPHY. 2006. Effects of locally rare taxa on the precision and sensitivity of RIVPACS bioassessment of freshwaters. *Freshwater Biology* 51:1924–1940.
- CLARKE, R. T., J. F. WRIGHT, AND M. T. FURSE. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling* 160:219–233.
- CLIFFORD, H. T., AND W. STEPHENSON. 1975. Introduction to numerical classification. Academic Press, New York.
- DAVIES, S. P., AND S. K. JACKSON. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications* 16:1251–1266.
- DE ZWART, D. S., D. DYER, L. POSTHUMA, AND C. P. HAWKINS. 2006. Predictive models attribute effects on fish assemblages to toxicity and habitat alteration. *Ecological Applications* 16:1295–1310.
- FAITH, D. P., P. R. MINCHIN, AND L. BELBIN. 1987. Compositional dissimilarity as a robust measure of ecological distance. *Vegetatio* 88:135–141.
- HAWKINS, C. P. 2006. Quantifying biological integrity by taxonomic completeness: evaluation of a potential

- indicator for use in regional- and global-scale assessments. *Ecological Applications* 16:1277–1294.
- HAWKINS, C. P., AND D. M. CARLISLE. 2001. Use of predictive models for assessing the biological integrity of wetlands and other aquatic habitats. Pages 59–83 in D. P. Batzer and S. A. Wissinger (editors). *Bioassessment and management of North American freshwater wetlands*. John Wiley and Sons, New York.
- HAWKINS, C. P., R. H. NORRIS, J. N. HOGUE, AND J. W. FEMINELLA. 2000. Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10:1456–1477.
- JOHNSON, R. K. 2003. Development of a prediction system for lake stony-bottom littoral macroinvertebrate communities. *Archiv für Hydrobiologie* 158:517–540.
- JOY, M. K., AND R. G. DEATH. 2002. Predictive modeling of freshwater fish as a biomonitoring tool in New Zealand. *Freshwater Biology* 47:2261–2275.
- JOY, M. K., AND R. G. DEATH. 2003. Assessing biological integrity using freshwater fish and habitat selection functions. *Environmental Management* 32:747–759.
- KAUFMANN, P. R., P. LEVINE, E. G. ROBISON, C. SEELIGER, AND D. V. PECK. 1999. Quantifying physical habitat in wadeable streams. EPA/620/R-99/003. Office of Research and Development, US Environmental Protection Agency, Washington, DC.
- KNAPP, R. A., C. P. HAWKINS, J. LADUA, AND J. G. McCLORY. 2005. Fauna of Yosemite National Park has low resistance but high resilience to fish introductions. *Ecological Applications* 15:835–847.
- LAZORCHAK, J. M., D. J. KLEMM, AND D. V. PECK (EDITORS). 1998. *Environmental Monitoring and Assessment Program: field operations and methods for measuring the ecological condition of wadeable streams*. EPA/620/R-94/004F. Office of Research and Development, US Environmental Protection Agency, Washington, DC.
- LEGENBRE, P., AND L. LEGENBRE. 1998. *Numerical ecology*. 2<sup>nd</sup> edition. Elsevier, Amsterdam, The Netherlands.
- LINKE, S., R. H. NORRIS, D. P. FAITH, AND D. STOCKWELL. 2005. ANNA: a new prediction method for bioassessment programs. *Freshwater Biology* 50:147–158.
- MARCHANT, R. 2002. Do rare species have any place in multivariate analysis for bioassessment? *Journal of the North American Benthological Society* 21:311–313.
- MOSS, D., M. T. FURSE, J. F. WRIGHT, AND P. D. ARMITAGE. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology* 17:41–52.
- MYERS, R. H. 1990. *Classical and modern regression with applications*. 2<sup>nd</sup> edition. PWS-Kent, Boston, Massachusetts.
- ONBERDORFF, T., D. PONT, B. HUGUENY, AND D. CHESSEL. 2001. A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology* 46:399–415.
- OSTERMILLER, J. D., AND C. P. HAWKINS. 2004. Effects of sampling error on bioassessments of stream ecosystems: applications to RIVPACS-type models. *Journal of the North American Benthological Society* 23:363–382.
- PARSONS, M., AND R. H. NORRIS. 1996. The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. *Freshwater Biology* 36:419–434.
- RAMSEY, F. L., AND D. W. SCHAFER. 1997. *The statistical sleuth: a course in methods of data analysis*. Duxbury Press, Belmont, California.
- REYNOLDS, T. B., R. H. NORRIS, V. H. RESH, K. E. DAY, AND D. M. ROSENBERG. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16:833–852.
- SIMPSON, J. C., AND R. H. NORRIS. 2000. Biological assessment of river quality: development of AUSRIVAS models and outputs. Pages 125–142 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). *Assessing the biological quality of fresh waters: RIVPACS and other techniques*. Freshwater Biological Association, Ambleside, UK.
- STODDARD, J. L., D. P. LARSEN, C. P. HAWKINS, R. K. JOHNSON, AND R. H. NORRIS. 2006. Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications* 16:1267–1276.
- USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2006. *Wadeable streams assessment: the states assess the nation's streams*. EPA 841/B/06/002. Office of Water, US Environmental Protection Agency, Washington, DC.
- VAN SICKLE, J., C. P. HAWKINS, D. P. LARSEN, AND A. T. HERLIHY. 2005. A null model for the expected macroinvertebrate assemblage in streams. *Journal of the North American Benthological Society* 24:178–191.
- VAN SICKLE, J., D. D. HUFF, AND C. P. HAWKINS. 2006. Selecting discriminant function models for predicting the expected richness of aquatic macroinvertebrates. *Freshwater Biology* 51:359–372.
- VAN SICKLE, J., D. P. LARSEN, AND C. P. HAWKINS. 2007. Exclusion of rare taxa affects performance of the O/E index in bioassessments. *Journal of the North American Benthological Society* 26:319–331.
- WRIGHT, J. F. 1995. Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology* 20:181–197.
- WRIGHT, J. F. 2000. An introduction to RIVPACS. Pages 1–24 in J. F. Wright, D. W. Sutcliffe, and M. T. Furse (editors). *Assessing the biological quality of fresh waters: RIVPACS and other techniques*. Freshwater Biological Association, Ambleside, UK.
- WRIGHT, J. F., M. T. FURSE, AND P. D. ARMITAGE. 1993. RIVPACS: a technique for evaluating the biological water quality of rivers in the UK. *European Water Pollution Control* 3:15–25.

Received: 30 August 2007

Accepted: 21 December 2007