COMMENT

Limnol. Oceanogr., 44(7), 1999, 1840–1841
© 1999, by the American Society of Limnology and Oceanography, Inc.

How important are rare species in aquatic community ecology and bioassessment? A comment on the conclusions of Cao et al.

Cao et al. (1998) showed that if rare species are eliminated from a freshwater macroinvertebrate data set, then the differences in species richness (the criterion for bioassessment in this case) between sites will be changed to a varying extent depending on the species abundance patterns at each site. Diverse sites with many rare species will lose a greater percentage of their fauna than will impoverished (probably disturbed) sites with few rare species, if species are eliminated at the same level of rarity, e.g., all species that occur only once. If species richness is the sole criterion for bioassessment, it seems most unlikely that any species would be eliminated from a data set. However, let us accept this much of the argument.

The authors then claim that these differential effects of species reduction will affect other methods of bioassessment because these ''are strongly dependent on species richness.'' In particular, they focus on the similarity measures used in multivariate analyses of benthic community composition, even though they have avoided these techniques in their own analysis. They criticize the widely used practice of eliminating rare species, claiming that the justification invoked (that rare species seldom affect multivariate analysis) is false. They assert that similarity measures ''usually strongly underweight rare species and overweight abundant ones'' and that the justification for elimination of rare species thus becomes circular. This assertion is wrong. First, in most multivariate analyses, abundance data are transformed to logarithms (or double square roots) for the express purpose of removing the dominance of the particularly abundant species and of increasing the emphasis in the analysis on the species with moderate or low abundances. Only if raw abundances span a narrow range, e.g., 0–100, is transformation not necessary (Gauch 1982). Second, if the assertion is correct, then there should be major differences in the outcomes of multivariate analyses, e.g., ordination analyses, based on abundance data compared with those based simply on the presence or absence of species, i.e., binary data in which each species in a sample is on an equal footing. This is not the case. In my own work (Marchant 1990; Marchant et al. 1995), ordination of binary data on benthic invertebrates from stream surveys over single or multiple catchments reveals biological and environmental gradients that are indistinguishable from those derived from ordinations based on quantitative abundance data. This is not a surprising result, considering Gauch's (1982) conclusion that the bulk of the information from diverse communities lies in the qualitative differences in species composition between sites.

Moreover, removal of rare species from a quantitative data set (e.g., Marchant 1990; Marchant et al. 1994), selection of a subset of species (the Plecoptera, Ephemeroptera, and Trichoptera; Marchant et al. 1995), or reduction in taxonomic resolution (from species to family, Marchant et al. 1995) results in ordinations that also differ little from those based on fine-grained data. In other words, these data sets contain much redundant information, particularly that resulting from the presence of rare species. This is the common experience of those who have conducted multivariate analyses on freshwater benthic communities distributed across clear environmental gradients such as those sampled by Cao et al. (Furse et al. 1984; Bunn et al. 1986; Wright et al. 1988; Bowman and Bailey 1997). When community differences are more subtle, e.g., Barton 1996, redundancy is less, but it is usually still observable.

Those working with marine macrobenthos have reached identical conclusions (Warwick 1993; Clarke and Warwick 1994). Indeed, Clarke and Warwick (1994) summed up the general situation by stating that ''sample relationships can often be well summarized in a 2-dimensional ordination, which is reduced from a very much higher-dimensional species space. This implies that many species must be interchangeable in the way they characterise the samples.'' In contrast, Cao et al. conclude that exclusion of rare species leads ''to an unacceptable loss of ecological information'' and suggest that ''the effects of deleting rare species on multivariate analysis need to be further examined.'' This amounts to claiming that no redundancy exists in these types of data. I think the weight of evidence from an extensive literature simply contradicts them.

Finally, Cao et al. criticize the low sample size of 100–300 individuals employed by the rapid bioassessment protocols (RPB) of the U.S. Environment Protection Agency (EPA), claiming these are ''too small to effectively use the information macroinvertebrate communities can provide.'' How do they know? They provide no analysis of the EPA data and merely assert that 100 individuals would not provide adequate coverage of the one undisturbed site (site 1) they sampled. Similar RBP assessments of 49 undisturbed river sites in the Australian state of Victoria (Marchant et al. 1997; sample sizes of 200–300 individuals per site) allowed a detailed multivariate classification of sites to be undertaken and enabled a predictive model of invertebrate composition to be developed. Predictive ability was confirmed by testing the model on independent data, i.e., sites not included in the model. It seems to me the onus is on Cao et al. to demonstrate why this is not an effective use of such data.

R. Marchant

Museum of Victoria
Abbotsford, Victoria, 3051 Australia

References

BARTON, D. R. 1996. The use of Percent Model Affinity to assess the effects of agriculture on benthic invertebrate communities in headwater streams of southern Ontario, Canada. Freshwater Biol. **36:** 397–410.

BOWMAN, M. F., AND R. C. BAILEY. 1997. Does taxonomic resolution affect the multivariate description of the structure of freshwater benthic macroinvertebrate communities? Can. J. Fish. Aquat. Sci. **54:** 1802–1807.

BUNN, S. E., D. H. EDWARD, AND N. R. LONERAGAN. 1986. Spatial and temporal variation in the macroinvertebrate fauna of streams of the northern jarrah forest, Western Australia: Community structure. Freshwater Biol. **16:** 67–91.

CAO, Y., D. D. WILLIAMS, AND N. E. WILLIAMS. 1998. How important are rare species in aquatic community ecology and bioassessment? Limnol. Oceanogr. **43:** 1403–1409.

CLARKE, K. R., AND R. M. WARWICK. 1994. Change in marine communities: An approach to statistical analysis and interpretation. Natural Environment Research Council.

FURSE, M. T., D. MOSS, J. F. WRIGHT, AND P. D. ARMITAGE. 1984. The influence of seasonal and taxonomic factors on the ordination and classification of running-water sites in Great Britain and on the prediction of their macroinvertebrate communities. Freshwater Biol. **14:** 257–280.

GAUCH, H. G. 1982. Multivariate analysis in community ecology. Cambridge Univ. Press.

MARCHANT, R. 1990. Robustness of classification and ordination techniques applied to macroinvertebrate communities from the La Trobe River, Victoria. Aust. J. Mar. Freshwater Res. **41:** 493–504.

———, L. A. BARMUTA, AND B. C. CHESSMAN. 1994. Preliminary study of the ordination and classification of macroinvertebrate communities from running waters in Victoria, Australia. Aust. J. Mar. Freshwater Res. **45:** 945–962.

———, ———, AND ———. 1995. Influence of sample quantification and taxonomic resolution on the ordination of macroinvertebrate communities from running waters in Victoria, Australia. Mar. Freshwater Res. **46:** 501–506.

———, A. HIRST, R. H. NORRIS, R. B. UTCHER, L. METZELING, AND D. TILLER. 1997. Classification and prediction of macroinvertebrate assemblages from running waters in Victoria, Australia. J. North Am. Benthol. Soc. **16:** 664–681.

WARWICK, R. M. 1993. Environmental impact studies on marine communities: Pragmatical considerations. Aust. J. Ecol. **18:** 63–80.

WRIGHT, J. F., P. D. ARMITAGE, M. T. F. URSE, AND D. MOSS. 1988. A new approach to the biological surveillance of river quality using macroinvertebrates. Int. Verein. Limnol. Theor. Angew. Verh. **23:** 1548–1552.

Received: 27 January 1999
Accepted: 3 May 1999
Amended: 10 June 1999

Limnol. Oceanogr., 44(7), 1999, 1841–1842
© 1999, by the American Society of Limnology and Oceanography, Inc.

Rare species are important in bioassessment (Reply to the comment by Marchant)

Marchant (1999) has raised some points with respect to our conclusions (Cao et al. 1998) about the importance of rare species in community analysis and bioassessment that are worthy of further discussion. Whereas we indicated that rare species could be important for distinguishing among sites along an impairment gradient, Marchant argues that the weight of evidence suggests otherwise—that ample redundancy in species composition and abundance allows the elimination of rare species from many multivariate analyses without affecting conclusions significantly. Our study warned that environmental practitioners should be alert to situations in which rare species could make a significant difference in bioassessment interpretation. Marchant also questions some of our conclusions. Our detailed responses follow.

First, Marchant disagrees with Cao et al.'s criticism regarding the circular relationship between underweighting rare species in similarity measures and the effect of rare species on multivariate techniques. It is true that log or other forms of data transformations can increase the weight of less abundant and rare species. Cao et al. (1997) demonstrated how the transformation reduced the bias of overweighting abundant species inherent in most similarity measures (including Euclidean Distance, Bray–Curtis Index, and Percentage Similarity), but they also showed that the bias could not be eliminated by a log transformation, which complicated some others (e.g., Canberra Metric). Marchant suggests that the agreement between the solutions from binary (presence/absence) and quantitative data supports his argument. However, other explanations exist. For example, whether and how the solutions from the two types of data sources are different will depend on how strong the pattern is. If a pattern is obvious, both types of data are sufficient to detect it, and even the presence of just a few dominant/common species may be enough. We agree with Marchant that the agreement indicates that most of the information is contained in species composition rather than in relative/absolute species abundance. However, less abundant and rare species typically dominate the taxonomic makeup of natural systems, and the effectiveness of binary data supports the importance of rare species.

Marchant further suggests that emphasizing rare species is equivalent to denying the redundancy in community data. This argument is not logical. Whereas data sets do contain redundant information, Marchant does not demonstrate that it is the rare species information that is necessarily the redundant part for bioassessment. Defining community types or a major environmental gradient is different from detecting impacts (*see below*). Indeed, it can be strongly argued that, in contrast, it is those abundant species that respond to the