

Influence of taxonomic resolution, biological attributes and data transformations on multivariate comparisons of rocky macrofaunal assemblages

Theresa Lasiak^{1,2,*}

¹Department of Zoology, University of Transkei, Private Bag X1 UNITRA, Umtata, Eastern Cape 5117, South Africa

²Present address: Centre for Research on Ecological Impacts of Coastal Cities, Marine Ecology Laboratories A11, University of Sydney, Science Road, New South Wales 2006, Australia

ABSTRACT: Multivariate comparisons of impacted and non-impacted biotic assemblages require decisions to be made about the taxonomic resolution to be used, biological attributes to be measured and whether the contributions of common, or combinations of common, intermediate and rare taxa should be emphasised. This study shows how these decisions affect comparisons of exploited and non-exploited rocky infratidal macrofaunal assemblages distributed across a biogeographic transition zone on the south-east coast of South Africa. Two-way crossed ANOSIM tests were used to test for differences in the structure of assemblages among regions of the coast and between exploited and non-exploited localities. In tests based on presence/absences, significant differences in assemblages among regions and between treatments were evident only at the species level. In tests based on densities and biomasses, regional differences were evident at the species, order, class and phylum level, and were irrespective of whether dominant or combinations of dominant, intermediate and rare taxa were emphasised. Significant differences in exploited and non-exploited assemblages were evident in all of the comparisons except those in which numerically dominant higher taxa were emphasised. These results do not suggest that the natural environmental gradient influences the fauna more by species replacements than by changes to the proportions of higher taxa, nor do they indicate that natural environmental variations and anthropogenic disturbances modify macrofaunal assemblages at different taxonomic levels. Further studies are needed to establish whether the taxonomic level, at which modification of assemblages occurs, varies with the type of disturbance (e.g. pollution vs exploitation, strong vs weak gradients) or biotic component (e.g. rocky vs soft-bottom assemblages) under consideration.

KEY WORDS: Rocky shores · Macrofauna · Exploitation · Biogeography · Taxonomic resolution · Data transformations · Biological attributes

Resale or republication not permitted without written consent of the publisher

INTRODUCTION

Comparative studies of impacted and non-impacted biotic assemblages require decisions about the level of taxonomic resolution to which organisms are identified, the biological attributes to be measured and the transformation of data prior to statistical analysis. The taxonomic level chosen is likely to depend on the objectives of the study (Ellis 1985), the time, funds and human resources available, as well as the extent to

which the biotic component of interest is known to be robust to taxonomic aggregation (Warwick 1993). Although identification of organisms to species used to be the norm (Olsgard et al. 1998), several authors have shown that polluted and non-impacted soft-bottom assemblages can be discriminated at higher taxonomic levels (Warwick 1988, Ferraro & Cole 1990, Gray et al. 1990, Somerfield & Clarke 1995, Olsgard et al. 1997). The response of some of these assemblages to pollution has, in fact, been found to be more clear-cut at

*Email: tlasiaak@hotmail.com

higher taxonomic levels (Warwick 1988, 1993). In large and/or heterogeneous areas, this may be because the natural environmental variables that confound differences due to contamination gradients influence the fauna by species replacement rather than by changes in the proportions of the major taxa (Warwick 1993). Olsgard & Somerfield (2000) have recently shown that the degree of concordance in the multivariate pattern between species and higher taxonomic levels varies with the level of pollution, being high in areas subjected to prolonged pollution and poor in pristine areas. Further studies are needed to establish whether other types of biota exhibit similar relationships, particularly assemblages separated by larger spatial scales than are typically studied along contamination gradients, and those subject to other forms of human perturbation.

The relative importance of taxa can be assessed in terms of their presence/absence, coded abundances, frequencies of occurrence, densities or biomass (Field et al. 1982). In environmental impact studies, densities are the most common, and more often than not the only, biological attribute measured. This is because they are easier and quicker to measure than biomass (Warwick 1993). Multivariate analyses of similarity measures derived from these different biological attributes are likely to indicate that different taxa are important in discriminating between sites.

In multivariate analyses based on quantitative data, transformations are used to weight the relative contribution of common, intermediate and rare taxa to the estimates of similarity between pairs of samples, with more and more down-weighting of common taxa occurring down the transformation sequence (Clarke & Green 1988). The transformation chosen determines the aspect of the assemblage that is emphasised in subsequent analyses and can affect the outcome of multivariate analyses, particularly when data are aggregated to higher taxonomic levels (Olsgard et al. 1997, 1998, Karakassis & Hatzilyanni 2000). If untransformed data are used, the focus is on patterns in common taxa only; however, if the fairly severe 4th-root transformation is used, as recommended by Field et al. (1982), the focus shifts to patterns within entire assemblages (Clarke & Warwick 1994).

The objective of this paper is to show how decisions about taxonomic resolution, the biological attribute that is measured, and use of raw versus transformed data, affect the outcome of multivariate comparisons of rocky infratidal macrofaunal assemblages. The assemblages chosen for study straddle a biogeographic transition zone on the south-east coast of South Africa and are also either subject to or protected from intense exploitation by subsistence fishers. Details of the study localities, sampling methods, data sets and results of multi-

variate analyses based on 4th-root transformed species' densities and biomasses are given in Lasiak (1998). The present study makes use of 3 data sets emanating from that paper. Additional characterisations of the assemblages were obtained by aggregating the species-level data to higher taxonomic levels, by using 3 different biological attributes (presence/absences, densities and biomasses) to assess the relative importance of these taxa, and by varying the contributions made by dominant, intermediate or rare taxa. The marked natural environmental gradient was expected to influence the distribution and abundance of the entire fauna but to do so primarily at the species level. This is because evolution results in the formation of groups of similar species each adapted to a relatively narrow range of environmental conditions (Warwick 1993). Exploitation was expected to modify the assemblages in an entirely different way because it is directed at one particular group of animals, the molluscs (Bigalke 1973, Lasiak 1992), and results in differences in abundance and biomass rather than species composition (Lasiak 1998).

METHODS

Owing to the limitations on the size of data set that PRIMER Version 4.0 can handle, a sub-set of each data matrix, consisting of the 160 most important of the 249 species and/or morphospecies recorded, was selected using the REDUCE routine. The estimates in the resultant data matrices were left in their raw form, or subjected to either a square-root or 4th-root transformation, prior to the computation of triangular similarity matrices based on the Bray-Curtis similarity measure. The species-level estimates were then aggregated into 29 orders, 17 classes and 10 phyla prior to the application of the analytical procedures. Taxonomic uncertainty precluded the aggregation of estimates to family and genus level. Two-way crossed analyses of similarities (ANOSIM) were used to test for significant differences in the structure of assemblages among coastal regions (i.e. south, central or north) and between exploited and non-exploited treatments (Clarke 1993). The specific hypotheses examined were that the quantitative differences in assemblages observed among coastal regions and between treatments were similar; (1) at all levels of taxonomic resolution, (2) irrespective of the severity of data transformation applied and (3) regardless of the biological attribute measured.

RESULTS

In the 2-way crossed ANOSIM tests based on similarity matrices derived from presence/absence data,

significant differences in the structure of assemblages between treatments and among regions were evident at the species level only (Table 1a). Pair-wise comparisons showed that the differences in the assemblages were significant in all 3 pairs of regions (Table 1b).

In the ANOSIM tests based on matrices derived from densities, significant regional differences in assem-

blages were evident regardless of the taxonomic level or data transformation used (Table 2a). Significant differences between exploited and non-exploited assemblages, however, were only apparent in the tests based on species-level data and on those containing transformed order, class and phylum densities. The R-statistic for these global tests generally increased, in value as the severity of the transformation increased and decreased as densities were aggregated to higher taxonomic levels. The value of R was always greater for the global test of differences among regions than for that between treatments (Table 2a). Pair-wise comparisons of the assemblages in the southern and northern regions were significant regardless of the taxonomic level and data transformation used (Table 2b). The comparisons of assemblages in the southern and central region, however, were only significant when analyses were based on species-level data and transformed densities of higher taxa. Significant differences in the structure of the assemblages in the central and northern regions were evident only in the tests based on species-level data and 4th-root transformed order densities.

In the ANOSIM tests based on matrices derived from biomasses, significant differences in the structure of assemblages were evident both among regions and

Table 1. Results of 2-way crossed ANOSIM tests based on similarity matrices derived from presence/absence data aggregated to various taxonomic levels. (a) Global tests on the effect of region and treatment; (b) pair-wise tests on the effect of region. R: ANOSIM statistic; p: probability level

(a) Taxonomic level	Source of variation			
	Regions		Treatments	
	R	p	R	p
Species	1.000	<0.001	0.655	0.003
Order	0.032	0.406	0.012	0.467
Class	-0.110	0.749	0.001	0.457
Phylum	-0.054	0.737	0.179	0.240

(b) South vs North	Source of variation				
	South vs Central		Central vs North		
R	p	R	p	R	p
1.000	0.01	1.000	0.033	1.000	0.025

Table 2. Results of 2-way crossed ANOSIMs based on similarity matrices derived from untransformed, square-root and 4th-root transformed densities of various taxonomic levels. (a) Global tests on the effect of region and treatment; (b) pair-wise tests on the effect of region. R: ANOSIM statistic; p: probability level

(a) Transform Taxonomic level	Source of variation											
	Differences among regions						Differences between treatments					
	None		Square root		4th root		None		Square root		4th root	
	R	p	R	p	R	p	R	p	R	p	R	p
Species	0.889	<0.001	0.957	<0.001	1.000	<0.001	0.559	0.027	0.822	0.007	0.872	0.007
Order	0.549	0.003	0.779	<0.001	0.838	<0.001	0.218	0.140	0.614	0.030	0.711	0.017
Class	0.582	0.002	0.728	<0.001	0.779	<0.001	0.228	0.140	0.512	0.040	0.539	0.003
Phylum	0.506	0.006	0.668	0.001	0.720	<0.001	0.269	0.103	0.502	0.033	0.707	0.001

(b) Transform	Taxon	Regions					
		South vs North		South vs Central		Central vs North	
		R	p	R	p	R	p
None	Species	0.911	0.01	1.000	0.033	0.761	0.025
	Order	0.665	0.02	0.798	0.067	0.236	0.325
	Class	0.599	0.02	0.798	0.067	0.427	0.125
	Phylum	0.498	0.02	0.731	0.067	0.188	0.325
Square root	Species	1.000	0.01	1.000	0.033	0.857	0.025
	Order	0.911	0.01	0.933	0.033	0.426	0.10
	Class	0.844	0.01	0.933	0.033	0.426	0.125
	Phylum	0.599	0.02	1.000	0.033	0.426	0.175
4th root	Species	1.000	0.01	1.000	0.033	1.000	0.025
	Order	0.888	0.01	1.000	0.033	0.618	0.025
	Class	0.866	0.01	1.000	0.033	0.426	0.15
	Phylum	0.755	0.01	1.000	0.033	0.330	0.175

Table 3. Results of 2-way crossed ANOSIMs based on similarity matrices derived from untransformed, square-root and 4th-root transformed biomasses of various taxonomic levels. (a) Global tests on the effect of region and treatment; (b) pair-wise tests on the effect of region. R: ANOSIM statistic; p: probability level

(a)		Source of variation											
Transform	Taxonomic level	Differences among regions						Differences between treatments					
		None		Square root		4th root		None		Square root		4th root	
		R	p	R	p	R	p	R	p	R	p	R	p
	Species	0.670	0.001	0.907	<0.001	0.932	<0.001	0.695	0.007	0.807	0.003	0.649	0.01
	Order	0.660	0.001	0.813	<0.001	0.846	<0.001	0.766	0.003	0.817	0.003	0.878	0.003
	Class	0.652	0.001	0.813	<0.001	0.880	<0.001	0.725	0.003	0.776	0.003	0.878	0.003
	Phylum	0.626	0.002	0.821	<0.001	0.872	0.001	0.705	0.003	0.735	0.003	0.735	0.003
(b)		Regions											
Transform	Taxon	South vs North		South vs Central		Central vs North							
		R	p	R	p	R	p	R	p				
None	Species	0.522	0.02	1.000	0.033	0.330	0.20						
	Order	0.477	0.02	1.000	0.033	0.521	0.10						
	Class	0.521	0.02	1.000	0.033	0.474	0.10						
	Phylum	0.454	0.05	1.000	0.033	0.474	0.10						
Square root	Species	0.922	0.01	1.000	0.033	0.665	0.075						
	Order	0.722	0.01	1.000	0.033	0.665	0.05						
	Class	0.722	0.01	1.000	0.033	0.665	0.05						
	Phylum	0.722	0.01	1.000	0.033	0.665	0.05						
4th root	Species	0.933	0.01	1.000	0.033	0.761	0.025						
	Order	0.755	0.01	1.000	0.033	0.714	0.025						
	Class	0.799	0.01	1.000	0.033	0.857	0.025						
	Phylum	0.844	0.01	1.000	0.10	0.665	0.075						

between treatments regardless of the taxonomic level or data transformation used (Table 3a). The ANOSIM statistic for both global tests (i.e. differences among regions and between treatments) generally increased in value as the severity of the transform increased, but showed no consistent response when biomasses were aggregated to higher taxonomic levels (Table 3a). All of the pair-wise comparisons of assemblages in the southern versus northern regions and southern versus central regions were significant. Differences between the assemblages in the central and northern regions, however, only became apparent when order, class and phylum biomasses were subjected to a square-root transformation and when species, order and class biomasses were 4th-root transformed (Table 3b).

DISCUSSION

Aggregation of species-level data to higher taxonomic levels alters the numerical structure of data matrices in 2 ways. Firstly, there is a marked decrease in the total number of entities contributing to estimates of dissimilarity. In the present study, the number of rows within the data matrices declined from 160 spe-

cies to 29 orders, 17 classes and 10 phyla. Secondly, because most samples contain representatives of higher taxonomic levels, there is an increase in the number of non-0 values in each column (Vanderklift et al. 1996). If presence/absence data are being used to characterise assemblages, there will be an increase in the number of ones in each column. When data sets based on this attribute are subjected to multivariate analyses, the emphasis is shifted towards patterns in intermediate and rare taxa rather than ubiquitous groups (Clarke & Warwick 1994). The failure of 2-way crossed ANOSIM tests to discriminate among regions and between treatments, when presence/absences are aggregated above species level, is probably due to the marked decline in the relative proportion of intermediate and rare versus common taxa.

In ANOSIM tests based on matrices derived from densities and biomasses, significant regional differences in the structure of assemblages were evident irrespective of whether: (1) the fauna was characterised at the species, order, class or phylum level; and (2) the emphasis was on common taxa only or on combinations of common, intermediate and rare taxa. The pair-wise comparisons of the southern and northern regions indicated that rare, intermediate and common

forms of all 4 taxa contributed to the differences between assemblages. Significant differences between the southern and central assemblages were evident in all of the comparisons except those emphasising the contribution of numerically dominant orders, classes and phyla. Differences in the central and northern assemblages were only apparent when the contributions of numerically dominant species in isolation and in combination with less dominant forms, numerically dominant, intermediate and rare orders combined, and biomass dominant, intermediate and rare taxa combined were emphasised. The greater number of significant differences detected in pair-wise comparisons involving the southern region probably reflects the fact that species which are restricted to or attain greater numbers or biomasses in this region make the greatest contribution to the dissimilarity in assemblages along this coast (Lasiak 1998).

Significant differences in exploited and non-exploited assemblages were evident in all of the comparisons except those emphasising numerically dominant orders, classes and phyla. The fact that biomass-dominant but not numerically dominant higher taxa contributed to the differences between assemblages probably reflects the difference in the size of individual organisms. It may also be because the group of taxa responsible for discriminating between the 2 treatments in tests based on densities differs from those in tests based on biomasses. This explanation seems plausible when one takes into account the fact that SIMPER analyses based on 4th-root transformed densities and biomasses have shown that different sets of species are important in discriminating between exploited and non-exploited treatments (Lasiak 1998). Only 9 of the 20 major discriminators were common to the SIMPER analyses based on densities and biomasses. These analyses also showed that the contribution of nontarget species, such as anemones, barnacles, amphipods, crabs, opisthobranchs, ophiuroids, echinoids and holothurians, to the dissimilarity between exploited and non-exploited assemblages was often greater than that of the species targeted by subsistence fishers.

The fact that the ANOSIM tests for differences among regions and between treatments were not adversely affected by taxonomic aggregation of quantitative data and that the value of the R-statistic generally decreased as taxonomic level increased concurs with the findings of Somerfield & Clarke (1995) and Olgard et al. (1997). Although these authors found that increasing the severity of data transformation had similar effects, in the present study, the value of the ANOSIM statistic and the number of significant pair-wise regional differences both increased when data were more strongly transformed. These discrepancies

may reflect differences in the relative proportions of taxa in rocky and soft-bottom assemblages or differences in the effects of different types of perturbations.

The ability of higher taxa to discriminate between assemblages has been attributed to functional coherence among species within higher taxa (Warwick 1993), redundancy of information in species-level data sets (Gray et al. 1988, Chapman 1998) and a hierarchical structure in biological response to stress (Ferraro & Cole 1990). The latter theory assumes that as stress increases, the adaptability of species is exceeded, in turn, by that of genera, families, orders, classes and phyla. The extent to which taxonomic aggregation affects multivariate analyses may also depend on the way in which species are distributed amongst the higher taxa (Somerfield & Clarke 1995). Here, 5 orders (Errantia, Sedentaria, Isopoda, Amphipoda and Archaeogastropoda) were considerably more speciose than the others and were major contributors to 3 of the classes and phyla. As these orders together accounted for 57% of the species and 65% of the individuals enumerated, it is not surprising that analyses based on higher taxa gave similar results to those based on species-level data.

This study does not indicate that gradients in natural environmental variables are likely to influence the fauna more by species replacement than by changes in the proportions of major taxa present (Warwick 1993). This may be because the rocky shore assemblages examined were spread over a larger spatial scale and were exposed to a much stronger natural environmental gradient than the soft-bottom assemblages hitherto examined (Warwick 1988, Ferraro & Cole 1990, Gray et al. 1990, Somerfield & Clarke 1995, Olgard et al. 1997, Olgard & Somerfield 2000). The results also do not support the notion that natural environmental variations and anthropogenic disturbances modify macrofaunal assemblages at different taxonomic levels (Warwick 1993). The taxonomic level at which modification of assemblages occurs may, of course, vary with the type of anthropogenic disturbance (e.g. pollution vs exploitation) and biotic component under consideration. Further studies on assemblages distributed along weak and strong natural environmental gradients and on other assemblages subject to these and other forms of anthropogenic disturbance are needed to test these hypotheses. We should also bear in mind that while the use of coarser levels of taxonomic resolution undoubtedly saves time, money and effort expended on impact assessments, it does not enhance our basic understanding of the responses of organisms to human perturbations. Researchers also need to decide whether assemblages should be characterised solely on the basis of dominant taxa or on some combination of dominant, intermediate and rare taxa.

Acknowledgements. This study was funded by the National Research Foundation, South Africa. Drs. Arthur Dye and Gee Chapman, and 3 anonymous referees are thanked for their comments on the previous draft.

LITERATURE CITED

- Bigalke EH (1973) The exploitation of shellfish by coastal tribesmen of Transkei. *Ann Cape Prov Mus Nat Hist* 9: 159–175
- Chapman MG (1998) Relationships between spatial patterns of benthic assemblages in a mangrove forest using different levels of taxonomic resolution. *Mar Ecol Prog Ser* 162: 71–78
- Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* 18:117–143
- Clarke KR, Green RH (1988) Statistical design and analysis for a 'biological effects' study. *Mar Ecol Prog Ser* 46: 213–226
- Clarke KR, Warwick RM (1994) Changes in marine communities: an approach to statistical analysis and interpretation. Plymouth Marine Laboratory, Plymouth
- Ellis D (1985) Taxonomic sufficiency in pollution assessment. *Mar Pollut Bull* 16:459
- Ferraro SP, Cole FA (1990) Taxonomic level and sample size sufficient for assessing pollution impacts on the Southern California Bight macrobenthos. *Mar Ecol Prog Ser* 67: 251–262
- Field JG, Clarke KR, Warwick RM (1982) A practical strategy for analysing multispecies distribution patterns. *Mar Ecol Prog Ser* 8:37–52
- Gray JS, Aschan M, Carr MR, Clarke KR, Pearson TH, Rosenberg R, Warwick RM (1988) Analysis of community attributes of the benthic macrofauna of the Frierfjord/Langesundfjord and in a mesocosm experiment. *Mar Ecol Prog Ser* 46:151–165
- Gray JS, Clarke KR, Warwick RM, Hobbs G (1990) Detection of initial effects of pollution on marine benthos: an example from Ekofisk and Eldfisk oilfields, North Sea. *Mar Ecol Prog Ser* 66:285–299
- Karakassis I, Hatziyanni E (2000) Benthic disturbances due to fish farming analyzed under different levels of taxonomic resolution. *Mar Ecol Prog Ser* 203:247–253
- Lasiak T (1992) Contemporary shellfish-gathering practices of indigenous coastal people in Transkei: implications for the interpretation of the archaeological record. *S Afr J Sci* 88: 19–28
- Lasiak T (1998) Multivariate comparisons of rocky infratidal macrofaunal assemblages from replicate exploited and non-exploited localities on the Transkei coast of South Africa. *Mar Ecol Prog Ser* 167:15–23
- Olsford F, Somerfield PJ (2000) Surrogates in marine benthic investigations— which taxonomic unit to target? *J Aquat Ecosyst Stress Recovery* 7:25–42
- Olsford F, Somerfield PJ, Carr MH (1997) Relationships between taxonomic resolution and data transformations in analyses of a macrobenthic community along an established pollution gradient. *Mar Ecol Prog Ser* 149:173–181
- Olsford F, Somerfield PJ, Carr MH (1998) Relationships between taxonomic resolution, macrobenthic community patterns and disturbance. *Mar Ecol Prog Ser* 172:25–36
- Somerfield PJ, Clarke KR (1995) Taxonomic levels, in marine community studies, revisited. *Mar Ecol Prog Ser* 127: 113–119
- Vanderklift MA, Ward TJ, Jacoby CA (1996) Effect of reducing taxonomic resolution on ordinations to detect pollution-induced gradients in macrobenthic infaunal assemblages. *Mar Ecol Prog Ser* 136:137–145
- Warwick RM (1988) Effects on community structure of a pollution gradient—summary. *Mar Ecol Prog Ser* 46:207–211
- Warwick RM (1993) Environmental impact studies on marine communities: pragmatical considerations. *Aust J Ecol* 18: 63–80

Editorial responsibility: Tony Underwood (Contributing Editor), Sydney, New South Wales, Australia

*Submitted: August 8, 2001; Accepted: November 19, 2002
Proofs received from author(s): February 25, 2003*