

Ecological patterns in multivariate assemblages: information and interpretation of negative values in ANOSIM tests

M. G. Chapman*, A. J. Underwood

Centre for Research on Ecological Impacts of Coastal Cities and Institute of Marine Ecology, Marine Ecology Laboratories A11, University of Sydney, New South Wales 2006, Australia

ABSTRACT: Analysis of similarities (ANOSIM) has been widely used for testing hypotheses about spatial differences and temporal changes in assemblages and particularly for detecting environmental impacts. ANOSIM generates a value of R which is scaled to lie between -1 and $+1$, a value of zero representing the null hypothesis. Generally, R lies between zero and $+1$. Values much smaller than zero have been considered unlikely because they would indicate greater dissimilarity among replicate units within samples than occurs between samples. Nevertheless, in some habitats, frequent and large negative values of R are common. In this paper, assemblages that consistently gave negative R values when analysed using ANOSIM were examined to identify patterns of differences among replicates within and between samples to test the hypothesis that particular patterns of differences generated consistent negative R values. The hypothesised patterns were then tested by analysing simple assemblages generated by computer simulation and examining the frequencies of R values. In natural assemblages, negative R values were found when assemblages were very patchy so that replicates were variable, but each sample had similar amounts of variability among replicates. Large negative values of R were particularly common when either or both samples contained an outlier, or when the assemblage being sampled had 2 different states and the replicates had sampled each of these states. Negative values of R may therefore indicate the need for stratification of the sampling design, or problems of positive correlation between the different sets of samples. When negative values occur, they should not simply be ignored as anomalies. In fact, they identify important ecological information and identify issues about the design of sampling.

KEY WORDS: Assemblage · Multivariate · Sampling design · Stratification

INTRODUCTION

In recent years, analysis of similarities (ANOSIM) has been widely used for testing hypotheses about spatial differences and temporal changes in assemblages of organisms (Gray et al. 1988, Warwick 1988, Warwick et al. 1990, Chapman et al. 1995, Oliver & Beattie 1996 among others). These procedures were originally introduced for detecting environmental impacts. The logic, advantages and disadvantages of these procedures were reviewed in Clarke (1993). Briefly, ANOSIM is an analogue of a 1-factor analysis of variance based on multispecies data. The procedures use

the difference between average ranked values of Bray-Curtis measures of dissimilarity in abundances and types of organisms among replicates between samples (\bar{r}_B) and within samples (\bar{r}_W) to give a test statistic,

$$R = \frac{\bar{r}_B - \bar{r}_W}{n(n-1)/4},$$

where n is the total number of replicates summed for the 2 samples. R is scaled to lie between -1 and $+1$, a value of zero representing the null hypothesis of no differences among a set of samples (Clarke 1993). Therefore, there will be small negative or positive values near to zero when the null hypothesis is true. Generally, R lies between zero and $+1$, its magnitude depending on the amount by which samples differ. R

*E-mail: gee@bio.usyd.edu.au

substantially less than zero has been considered unlikely because it would indicate greater dissimilarity among individual replicates within any sample than exists among replicates taken from different samples (Clarke 1993).

In some studies, negative R values are, however, relatively common (Sommerfield & Clarke 1995, Olsgaard et al. 1997, Chapman & Underwood unpubl. data) and many can be substantially smaller than zero. Nevertheless, they have been generally ignored as irrelevant, anomalous or indicating 'incorrect labelling of samples' (Clarke & Warwick 1994). Because they indicate greater dissimilarity within than among samples, they may, however, contain important information about ecological patterns in the assemblages being examined or problems with the sampling design.

For example, non-independent sampling of patchy habitats from one time to another is likely to give large negative R values. Suppose assemblages are sampled in a set of replicate plots at the first time of sampling and then the same plots are sampled at a subsequent time. If there is much variability in the composition of assemblages among plots, the measures of dissimilarity (and therefore the average within-sample rank \bar{r}_W) will be relatively large at each time of sampling (illustrated by the lengths of the dotted lines in Fig. 1a,b).

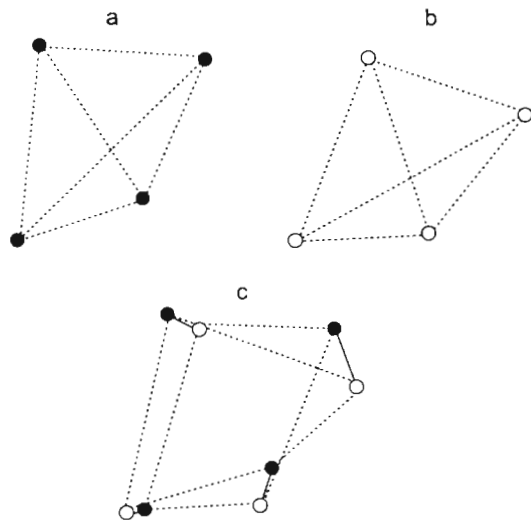


Fig. 1. Diagrams illustrating dissimilarities among replicate units in plots of non-metric multidimensional scaling of 2 samples. Distances between any 2 replicates indicate amount of dissimilarity between them. Dissimilarities among sample units have temporal non-independence because the same units were sampled at 2 times. (a) Illustrates large dissimilarity among units within the sample at Time 1; (b) illustrates large dissimilarity among units within the sample at Time 2; (c) illustrates that dissimilarity among sample units between Times 1 and 2 is composed of large measures similar to those within samples and very small measures (shown as solid lines) because the same units were sampled at each time

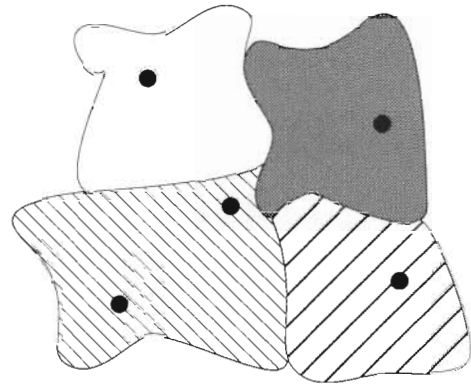


Fig. 2. There may be very large measures of dissimilarity among sample units within a sample when there are few units and the habitat being sampled is patchy with different assemblages in each patch. Comparison of such a sample with one from another area with similar scales of patchiness will give a small value of \bar{r}_B relative to \bar{r}_W because of the small measures of dissimilarity between those sample units which fall in the same types of habitat in each area

Further, suppose that the time between samples is not long enough for there to have been much change in the assemblages sampled. The measures of dissimilarity (and the average between-sample rank \bar{r}_B) will be smaller because they include comparisons between the non-independent samples (i.e. the same plots) which are very similar to each other (Fig. 1c). This will make \bar{r}_B smaller than \bar{r}_W and therefore R will be <0 .

For similar reasons, R may be very much smaller than zero if one compares samples in very patchy habitats using relatively few sampling units, so that only one (or a few) units fall into each type of habitat (Fig. 2). There will be very large differences among replicates within each set of samples because many such differences will be between units in 2 different types of patch. A second sample from a different area which has similar patchy habitats is also likely to have large dissimilarity among replicates and therefore large average \bar{r}_W . In this case, the average dissimilarity between samples and, therefore, the \bar{r}_B value will generally be smaller than the average \bar{r}_W value because there will be small values of dissimilarity between a sample unit from each of the 2 samples when both come from the same type of patch.

In some of our studies on intertidal rocky shores and boulder fields, large negative values of R are relatively common. In this paper, examples of spatial or temporal patterns of assemblages which frequently gave large negative R values when analysed using ANOSIM are illustrated to demonstrate the conditions under which negative R s occur and how to interpret them. Negative values near to zero are not considered because these are chance outcomes of comparing 2 samples from the

same assemblage. Most of the field-derived negative values discussed below were considerably smaller than zero. In some cases, many were also significantly small, i.e. in the extreme left-hand tail of the distribution of R from permutations of the data. It is important to note that the probability of getting large negative values of R in a comparison of 2 samples which are not different is 5%, as expected under the null hypothesis. In other words, this paper does not imply any problem with ANOSIM as a statistical test, but that negative values of the test statistic are important indicators of aspects of biological properties of the system sampled and of the sampling design.

The assemblages used as examples include algal and faunal assemblages in intertidal rock pools, algal assemblages on rocky shores and assemblages of chitons under boulders. The complete details of the studies of these assemblages are not important here. In addition, patterns in multispecies data sets were simulated on a computer and then tested using ANOSIM to identify patterns in the data that can give rise to negative values. Finally, the implications of negative R values in sampling designs are discussed.

METHODS

Analyses of spatial patterns. Intertidal rock pools of depths 5, 15 and 30 cm and diameters 15, 30 and 50 cm were constructed at a number of sites in the Cape Banks Scientific Marine Research Area in New South Wales, Australia (Underwood & Skilleter 1996). The development of the assemblage over a number of years was examined in several vertical strata (0–5, 5–15, 15–30 cm from the top of the pools) in each pool. The number of strata in each pool obviously depended on the depth of the pool. The percentage covers of algae and sessile animals were estimated and mobile animals counted in a number of vertical strata in each pool to test hypotheses about differences in assemblages in particular strata among pools of different depths. Details of the experimental design and sampling methods are given in Underwood & Skilleter (1996). Covers of encrusting and foliose algae and sessile animals were estimated from replicate quadrats in each stratum. Sessile animals that could be recognized as individuals (e.g. the tubeworm *Galeolaria caespitosa*) and mobile animals were counted and the numbers adjusted to give densities per unit area of each stratum (details in Underwood & Skilleter 1996). There were 75 taxa.

In this paper, we illustrate spatial patterns in the top stratum (5 cm deep from the rim of the pool) among pools of different depths, 5, 15 and 30 cm. Previous work has shown that assemblages did not differ significantly according to diameter of pools (Underwood &

Skilleter 1996). There were therefore 4 replicate pools (two 30 cm diameter and two 50 cm diameter) of each of the 3 depths in each of 24 sites (6 highshore, exposed sites; 6 lowshore, exposed sites; 6 highshore, sheltered sites and 6 lowshore, sheltered sites; Underwood & Skilleter 1996). These strata were compared across pools of each depth in each site separately at each of 3 times of sampling; October 1989, January 1990 and July 1990. The data collected at each time are known to be independent because of rates of turnover of organisms in these assemblages (Underwood unpubl. data).

Analyses of temporal patterns. As part of a study of spatial and temporal patterns of assemblages of chitons, gastropods and echinoderms under intertidal and shallow subtidal boulders on the coast of New South Wales, temporal changes in these assemblages were compared over periods of one or a few days, months and years to test hypotheses about the effects of disturbance on these animals (Chapman & Underwood 1996). These animals are very overdispersed, with large spatial variability from boulder to boulder. Analyses of differences in these assemblages from one time to another frequently gave negative R values, although the data were known to be independent from one time to another, because different boulders were sampled each time. Individual boulders had, however, been marked so that the relative sizes of R using non-independent or independent samples from one time to the next could be examined for this paper. Here, differences in the assemblage of chitons (16 taxa) from one day to the next are illustrated using data from the same 5 boulders (non-independent), as opposed to 5 different, randomly-selected boulders (independent) each day.

As part of a large-scale study of intertidal assemblages on wave-exposed shores in New South Wales, lowshore algal assemblages were compared within and among wave-exposed rocky shores along approximately 800 km of the coast, at 3-monthly intervals for 4 yr (details in Underwood & Chapman 1998). At each time of sampling, data were collected from 5 randomly-placed, replicate photoquadrats and the macroalgae identified to species or coarser taxonomic identity at each time (31 taxa). Spatial and temporal patterns of differences in these algal assemblages were examined using ANOSIM and analyses of variance. Temporal change was measured in each site over replicate periods of 3 mo and replicate periods of a year throughout the 4 yr period of the study. In some sites, analyses of changes from time to time in the assemblages gave large negative R values.

Computer simulations. Computer simulations were done to determine whether outliers in the data or unrecognized patterns of stratification within the sam-

Table 1. Details of computer simulations; for each sample in each test, the mean and variance (Var.) were used to generate abundances of each of 15 species in each replicate; the mean value of R and the proportion of negative R values were calculated from 1000 independent simulations (details in text); n is number of replicate units of the given condition in each sample

Test	Sample 1						Sample 2						Mean R	Proportion of negative R values
	Mean	Var.	n	Outliers		n	Mean	Var.	n	Outliers		n		
1—no outliers	100	45	4				100	45	4				0.01	0.52
2—no outliers	100	45	10				100	45	10				0.00	0.53
3—no outliers	100	45	4				100	45	10				-0.01	0.55
4—1 outlier in each sample	100	45	3	10	3	1	100	45	3	10	3	1	-0.13	0.94
5—1 outlier in each sample	100	45	3	700	300	1	100	45	3	700	300	1	-0.15	0.96
6—1 outlier in each sample	100	45	9	700	300	1	100	45	9	700	300	1	-0.02	0.70
7—1 outlier in each sample	10	45	9	700	300	1	10	45	3	700	300	1	-0.06	0.82
8—complete stratification	100	45	2	10	3	2	100	45	2	10	3	2	-0.14	1.00
9—2 outliers in each sample	100	45	8	700	300	2	100	45	8	700	300	2	-0.04	0.87
10—complete stratification	100	45	5	700	300	5	100	45	5	700	300	5	-0.05	0.99

Table 2. Numbers of pairwise comparisons that gave negative R values and, in brackets, number with $P > 0.90$ in comparisons of assemblages in the top stratum (5 cm) of rockpools 15 and 30 cm deep ($n = 4$ pools for each comparison); there are 6 possible comparisons for each set of environmental conditions (i.e. 1 comparison in each of the 6 sites)

Environmental conditions	Oct 1989	Jan 1990	Jul 1990
Low, sheltered	3	4 (2)	4 (1)
High, sheltered	4	4 (1)	3 (1)
Low, exposed	2	3	5 (2)
High, exposed	3 (2)	5	4

ples could actually give rise to large negative values of R . Simulations were not supposed to represent real sets of data (with overdispersed distributions, very different means and variances from one species to another, etc.), but to demonstrate that outliers can give rise to negative values of the test statistic. Simple assemblages were simulated using 2 samples, a range of means, variances and sample sizes, with or without outliers in either or both samples. Each replicate had the same 15 species and came from assemblages with the same structure (i.e. all species had the same means and variances for any sample taken). Abundances of each species were randomly generated from a normal distribution of given mean and variance. For each test, 2 samples of a chosen number of replicates were randomly generated and R was calculated. This was repeated for 1000 sets of data. It was predicted that, when the null hypothesis was true and there were no outliers in each sample, the R values would be symmetrical around and centred on zero. Further, to test whether our interpretations of the field data were correct, we predicted that the presence of outliers in the samples would increase the probability of negative R

values, particularly when the samples were stratified with equal numbers of replicates in each stratum. Details of tests of these hypotheses are given in Table 1.

RESULTS

Analyses of spatial patterns

In each of 24 sites at each of 3 times of sampling, R values were calculated for comparisons of assemblages in the top stratum of pools between pools 5 and 15 cm deep, between pools 5 and 30 cm deep and between pools 15 and 30 cm deep. For comparisons of the shallowest (5 cm) and deeper (15 or 30 cm) pools, there were relatively few negative values of R . Negative R values were, however, very common in comparisons of the 2 deeper sets of pools (15 vs 30 cm; Table 2) and several of these negative values were significantly large (with probability $P > 0.90$). Such frequent and large negative values of R have generally been considered to be unlikely in natural assemblages.

The results suggest that there was great variability in these assemblages among replicate 15 and 30 cm deep pools, but the amount of variability was similar among pools of different depths. Differences between the shallow pools and the 15 or 30 cm deep pools were more frequently greater than differences from pool to pool within each set.

Transforming the data to reduce the effects of abundant species relative to rare species has been shown to influence the significance of the results of ANOSIM, i.e. the value of the R statistic (Olsgaard et al. 1997). To test the hypothesis that transforming the data does not influence the prevalence of negative R values, χ^2 tests were used to compare the proportion of negative val-

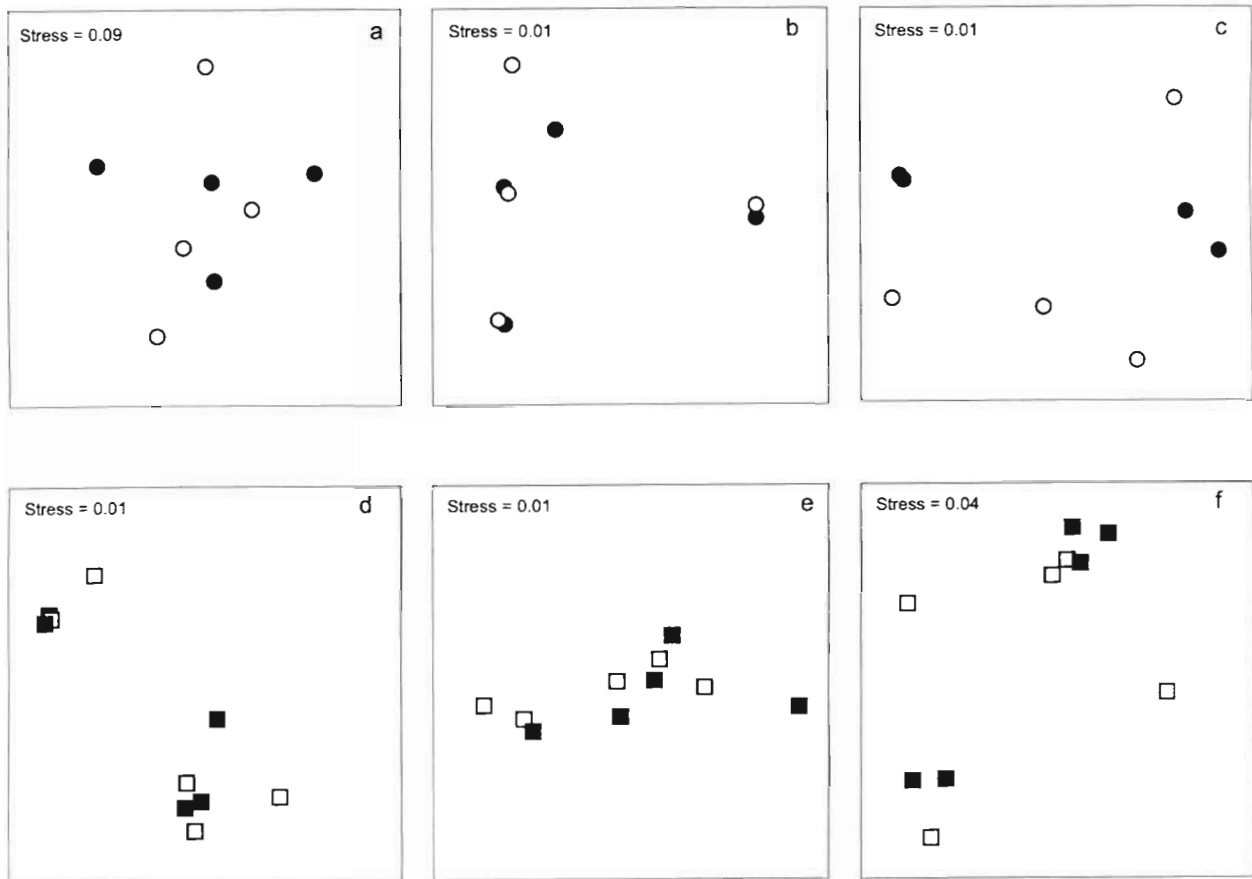


Fig. 3. nMDS plots of spatial patterns of assemblages in the top stratum of pools of different depths (filled and empty symbols; details in text) in 3 different sites (details of sites in Underwood & Skilleter 1996). (a) (○) 5 cm and (●) 30 cm deep pools in a low-shore sheltered site; (b) (○) 15 cm and (●) 30 cm deep pools in a low-shore sheltered site; (c) (○) 15 cm and (●) 30 cm deep pools in a mid-shore sheltered site. (d) Differences in assemblages of chitons under intertidal boulders in samples 1 d apart at Cape Banks (details of site in Chapman & Underwood 1996); (e, f) differences in a low-shore algal assemblage between (e) (□) March 1980 and (■) July 1980 at Flat Rock Island and (f) (□) January 1981 and (■) January 1982 at Cape Banks (details of sites in Chapman & Underwood 1998)

ues calculated using untransformed data and the proportion calculated from data transformed by double square root for all pairwise comparisons within each set of environmental conditions. All χ^2 values were non-significant ($\chi^2 = 1.78, 1.18, 0.15, 0.18$; $P > 0.10, P > 0.25, P > 0.50, P > 0.50$, respectively), indicating no effect of transforming the data on whether R values were likely or not to be negative. There was also little effect of transformation on the results of ANOSIM or the nMDS plots, probably because these assemblages do not vary greatly in abundances from taxon to taxon.

To illustrate the types of patterns that frequently gave large negative R values, nMDS plots of 3 sets of data are given in Fig. 3a,b,c. The simplest example is shown in Fig. 3a, for the comparison between 5 and 30 cm deep pools in a low-shore, sheltered site; $R = -0.25$, which was significant, $P = 0.94$. In this and other similar cases, there was very large spatial variability within

each of the samples and considerable overlap between the 2 samples. There were no clear outliers or subgroups in either sample. This example shows large variability in assemblages among replicate pools of each depth and relatively similar levels of variability between the 2 sets of pools. Therefore, for some replicates, between-sample comparisons give relatively small measures of dissimilarity, leading to $\bar{r}_B < \bar{r}_W$.

Although there were 75 taxa in the data set, relatively few were important in contributing to dissimilarity among replicate pools. Only the cover of coralline crust contributed more than 10% (12.7%) to average measures of dissimilarity among pools of different depths. In general, the same taxa (particularly covers of algae and the tube worm *Galeolaria caespitosa*) were important in contributing to variability among pools of each depth and for comparisons among pools of different depths (i.e. the percentage contribution to

the average measures of dissimilarity within each sample and between samples were relatively similar for each taxon). The only exception was the large limpet *Patella peroni* which was only really abundant in 1 pool in 1 sample. Therefore, the large measure of within-sample variability was caused by the covers/abundances of a small subset of species, with the same species being important in each sample (see also Underwood & Chapman 1998). The important taxa included foliose and encrusting algae, sessile animals and a few species of mobile gastropods.

As a second example, negative R was obtained when the pools in each sample separated into 2 distinct groups, but each sample contained both groups (illustrated for the comparison of the top stratum between 15 and 30 cm deep pools in a lowshore, sheltered site in Fig. 3b; $R = -0.25$; again this was significant, $P = 0.94$, coincidentally the same result for the data in Fig. 3a). One pool from each sample had zero cover of the tubeworm *Galeolaria caespitosa* and cover of the encrusting alga *Ralfsia verrucosa* of 38 and 45% (for 15 and 30 cm deep pools, respectively). These 2 pools plotted separately on the right-hand side of Fig. 3b. The other 6 pools all had cover of *G. caespitosa* between 27 and 70% and 5 had cover of *R. verrucosa* between 0 and 25%. Therefore, the pools stratified into 2 groups, largely based on the cover of 2 species, a tubeworm and an encrusting alga.

As a final example, an extreme value of negative R was obtained when the replicates within the samples separated into 2 subgroups, but the assemblages in the subgroups differed between the 2 samples (illustrated for comparisons of assemblages in the top strata of pools 15 or 30 cm deep in a midshore, sheltered site; Fig. 3c; $R = -0.33$; significant at $P = 1.00$). The assemblages in the 15 cm deep pools stratified into a single outlier and a rather variable subgroup of 3 pools (open symbols in Fig. 3c). This was largely due to the amount of cover of the encrusting alga *Hildenbrandia rubra*, which had 20% cover in the outlier and less than 4% cover in the other 3 pools. This species alone contributed 21% to average measures of dissimilarity among replicate 15 cm deep pools (SIMPER analysis; Clarke 1993). Other important species contributing to dissimilarity among replicates were the limpet *Cellana tramoserica* (13%), the snail *Nerita atramentosa* (11%) and the starfish *Patiriella exigua* (13%). Numbers of these animals were, however, very variable among pools, but there were no clear patterns of stratification of abundances.

The assemblages in the other sample (the 30 cm deep pools) stratified into 2 distinct subgroups (filled symbols in Fig. 3c), primarily due to a similar suite of species (*Ralfsia verrucosa*, 25% contribution of dissimilarity among replicates; *Hildenbrandia rubra*, 12%;

Cellana tramoserica, 12%; the snail *Austrocochlea porcata*, 11%; *Patiriella exigua*, 12%). The two 50 cm diameter pools had relatively large cover of *R. verrucosa*, large numbers of *C. tramoserica* and few *P. exigua*. The two 30 cm diameter pools had zero cover of *R. verrucosa*, fewer *C. tramoserica* and more *P. exigua*. Patterns of cover of *H. rubra* and numbers of *A. porcata* were simply variable among pools and no clear patterns were identified. Thus, in the case of the 30 cm deep pools, it appears that the assemblages were different between pools of different diameter, primarily due to 3 taxa. This was a chance finding. Such differences between diameters were not generally found in univariate analyses of many individual species and multivariate analyses of the whole assemblages (Underwood & Skilleter 1996), hence the use of pools of different diameter as replicates.

Analyses of temporal patterns

Assemblages of chitons under boulders were compared using non-independent samples (the same 5 boulders were sampled on each of 2 days) and independent samples (a different set of 5 boulders was sampled at each time). Each of these comparisons gave significantly large negative R values ($R = -0.20$, $P = 1.00$ and $R = -0.16$, $P = 0.98$, respectively). In the former case, as expected from non-independent sampling, there was large variability among boulders at each time of sampling, but assemblages on each boulder were the same or very similar from time to time. Hence, the \bar{r}_B measures includes very small measures of dissimilarity. Therefore, non-independence produced extreme negative R values. Such values may cause concern if there is a possibility that samples are not independent from one time of sampling to the next.

In the second comparison, samples were known to be independent (different boulders were sampled at each time). The large negative R value was due to the samples falling into 2 subgroups of assemblages and each subgroup being represented at each time of sampling (Fig. 3d; see also Fig. 3b). These subgroups were formed because the chitons were very overdispersed among boulders, with none or very few chitons under some boulders and many chitons under fewer boulders. Therefore the boulders with very few animals formed the larger subgroup and, at each time of sampling, a few boulders which had very large numbers of chitons were also present in the sample. These formed a smaller subgroup. The boulders always have such overdispersion of chitons and the proportion with chitons is fairly consistent from time to time. Therefore, it is typical that samples from 2 different times will have

similar proportions of boulders with very large or very small numbers of chitons.

Similar patterns were found in some comparisons of differences in lowshore algal assemblages over periods of 3 mo or 1 yr on exposed rocky shores (data transformed to $X^{1/4}$). In these cases, the negative values were not as large as described above and probability levels varied from approximately $P = 0.50$ to $P = 0.94$. Two temporal patterns that gave extreme negative R values are illustrated in Fig. 3e,f. Fig. 3e illustrates variability in the assemblages over a 3 mo period at Flat Rock Island (sites described in Chapman & Underwood 1998; $R = -0.14$, significant at $P = 0.94$). This is similar to the spatial comparisons shown in Fig. 3a. At each time of sampling, there was large variability from quadrat to quadrat and similar measures of dissimilarity were obtained at each time (average dissimilarity within samples was 16 and 18% for Times 1 and 2, respectively). This was mainly caused by only 2 taxa, cover of *Ulva lactuca* (percentage contribution to the average dissimilarity among quadrats was 47 and 51%, respectively) and a suite of species of *Sargassum* (41 and 46%, respectively). The quadrats were mostly covered with *Corallina* spp. (between 55 and 98% cover), but variability in cover of *Sargassum* spp. and *U. lactuca* was more important in contributing to variability within and among samples. These were very patchy assemblages, with a similar range of patchiness being sampled by random selection at each time of sampling.

The next example illustrates changes in the assemblage over a period of 1 yr in sites in the Cape Banks Scientific Marine Research Area (description of sites in Chapman & Underwood 1998). This comparison gave an R value of -0.10 , $P = 0.80$. The replicates at the first time of sampling were very variable (Fig. 3f, open symbols), again primarily due to very variable covers of *Corallina* spp., *Ulva lactuca* and *Sargassum* spp. from one quadrat to the next. The replicates from the second time of sampling (Fig. 3f, filled symbols) were stratified into 2 subgroups. The smaller group was almost entirely covered with *Sargassum* spp. (94 to 99%), with little cover of any other taxa; the larger group had very variable amounts of *U. lactuca* (11 to 39%), *Corallina* spp. (50 to 59%) and *Sargassum* spp. (5 to 30%).

Computer simulations

The above results indicated that stratification or outliers in the data would cause many negative R values. To test this, data were simulated as described in the 'Methods', with the null hypothesis true and with or without stratification in each or both samples. It was predicted that when data were simulated for stratified

assemblages, large values of negative R would be found.

In the first 3 tests, there were no outliers, but the number of replicates in each sample varied (Table 1). R was centred on and approximately symmetrical around zero, although there were more negative than positive values (Table 1).

In Tests 4 to 7, 1 replicate in each sample was obtained from a distribution with a different mean and variance, thereby creating an outlier amongst the replicates. In all cases, this generated many more negative values (between 70 and 96% of all R values were negative), although the range of R values was smaller. This was not affected by the relative magnitude of the mean and variance in the outlier relative to the other replicates (Tests 4 and 5). It was, however, affected by the number of replicates, with an outlier in samples with fewer replicates having a greater effect on the prevalence of negative R values (compare Tests 5, 6 and 7) than occurred when the outlier was in the larger sample.

Finally, increasing the proportion of outliers also increased the prevalence of negative R values, with all or nearly all values negative when the replicates in each sample were stratified into 2 equal subsamples representing different assemblages (compare Tests 4 and 5 with Test 8 or Tests 6 and 9 with Test 10 in Table 1).

DISCUSSION

R is a measure of variation between samples compared to variation within samples in dissimilarities of a suite of species using ranked differences among replicates (Clarke 1993). It is scaled to lie between -1 and $+1$, with values close to 0 representing the null hypothesis. Increasing positive values represent increasing differences among samples. Large or frequent negative values have been considered to be unlikely because they would represent greater variability within samples than among samples. It has been suggested that they are likely to represent errors in the data because of mislabelling of the samples (Clarke & Warwick 1994). This is undoubtedly true and therefore the quality of data giving large negative R values should be checked. Nevertheless, in many studies, large negative R values (i.e. towards or in the left-hand tail of the null distribution) appear to be common (here and Somerfield & Clarke 1995, Ols-gard et al. 1997, Chapman & Underwood unpubl. data), suggesting that they may be indicative of particular patterns of difference among samples. If so, they may also be informative about potential ecological processes creating differences. To investigate such

patterns, this study examined spatial and temporal patterns in data from a variety of intertidal habitats, which had previously been shown to give many large negative R values.

Spatial patterns were illustrated for comparisons of the assemblages in the top strata of rock pools across pools of different depths. Three general patterns gave rise to negative values. In some comparisons, they simply arose from very patchy assemblages among each set of pools (i.e. patchiness within samples), with a range of different taxa contributing to these differences. Some comparisons between pairs of replicates taken between samples were smaller than those within samples, but there were no strong patterns in the nMDS plots. Such examples are not particularly informative, except to illustrate that when very patchy assemblages are sampled with relatively few replicates ($n = 4$ in this case), negative values can be quite common when the null hypothesis is true.

The other 2 examples illustrated that stratification of the replicates into 2 subgroups also gave extreme negative R values (i.e. towards or in the left-hand tail of the null distribution). In the first case, each sample stratified into the same groups, largely due to the relative covers of 2 species (an encrusting alga and a tube worm). In the second case, each sample stratified into 2 subgroups, but these were not the same from one sample to another. Again, stratification was due to a few (although more than 2) components of the assemblage. Therefore, spatial comparisons of assemblages that give many or extreme values of R might usefully be examined for outliers or apparent stratification. Outliers could focus attention on the particular replicate that may (1) have errors in the quality of the data or (2) not be representative of the assemblage as a whole. When the data divide into 2 or more subgroups, the sampling design might be better if a stratified design were used. In either case, examining the causes for negative R values would be useful.

With respect to comparisons through time, non-independence is likely to give rise to negative R values, especially when there is little temporal change (due to large variability among replicates at each time, but small variability between times because the same samples are examined). The example of assemblages of chitons in boulder fields illustrated that similar results can be obtained from independent replicates when assemblages are very patchy, but the patchiness is similarly represented at each time of sampling. This is entirely analogous to problems of temporal (or spatial) non-independence in univariate analyses. For example, if there is marked positive correlation from time to time in some measure, variances within samples at each time are larger than the corresponding variances estimated among times in analysis

of variance (Underwood 1994, 1997). As a result, when the null hypothesis of no difference among times is false, F -ratios are smaller than predicted by the null hypothesis and there is excessive Type II error. In the multivariate case described here, R values are smaller than zero (the value predicted by the null hypothesis).

In this case, many boulders had very few chitons (usually only 1 or 2), a few boulders had very many (more than 30). When the same proportion of replicates at each time of sampling included the boulders with many chitons, extreme negative R values are obtained. Variability among boulders at each time of sampling was very large. Variability between times was smaller because assemblages on boulders with many chitons were similar at each time and assemblages on boulders with few chitons were similar at each time. Again, these data suggest that the samples might need to be stratified into boulders with few chitons and those with many chitons if changes through time are to be measured.

The final example was of algal assemblages on wave-exposed shores and, again, these temporal patterns mimicked the spatial patterns discussed above. In some cases, large negative values arose from comparisons of very patchy assemblages from one time to another, usually due to the cover of only a few species. In other cases, the replicates stratified into 2 groups, again due to only a few species.

Finally, the computer simulations illustrated that, for comparisons of 2 samples, outliers and stratification of the replicates increased the prevalence of negative R values when the null hypothesis is true. Negative values close to zero are unlikely to be more informative other than to support the null hypothesis. Frequent or large negative values may usefully focus attention on potential errors in the data (perhaps due to human error as suggested by Clarke & Warwick 1994). They may, however, also focus on how representative the replicates are of the assemblage by identifying outliers or the potential need to stratify the assemblage being examined. Further studies of patterns that give rise to negative values might, therefore, also be useful. Close examination of data giving rise to consistent negative values of R will reveal information that is crucial for being able to understand the nature of the assemblages being studied.

Acknowledgements. We thank the Australian Research Council, University of Sydney Small Grant Scheme, the Institute of Marine Ecology and Centre for Research on Ecological Impacts of Coastal Cities for supporting this research and the preparation of this manuscript. Many assistants helped with the field work. V Mathews assisted with the graphics. As is becoming routine, we thank Dr K. R. Clarke (Plymouth Marine Laboratories) for a massive amount of help.

LITERATURE CITED

- Chapman MG, Underwood AJ (1996) An experimental investigation of the effects of sampling on biota living under intertidal boulders. *J Exp Mar Biol Ecol* 207:103–126
- Chapman MG, Underwood AJ (1998) Inconsistency and variation in the development of intertidal algal assemblages. *J Exp Mar Biol Ecol* 224:265–289
- Chapman MG, Underwood AJ, Skilleter GA (1995) Variability at different spatial scales between a subtidal assemblage exposed to the discharge of sewage and two control assemblages. *J Exp Mar Biol Ecol* 189:103–122
- Clarke KR (1993) Non-parametric multivariate analyses of changes in community structure. *Aust J Ecol* 18:117–143
- Clarke KR, Warwick RM (1994) Changes in marine communities: an approach to statistical analysis and interpretation. Bourne Press Ltd, Plymouth
- Gray JS, Aschan M, Carr MR, Clarke KR, Green RH, Pearson TH, Rosenberg R, Warwick RM (1988) Analysis of community attributes of the benthic macrofauna of Frierfjord/Langesundfjord and in a mesocosm experiment. *Mar Ecol Prog Ser* 66:285–299
- Oliver I, Beattie AJ (1996) Invertebrate morphospecies as surrogates for species: a case study. *Conserv Biol* 10:99–109
- Olsgard F, Somerfield PJ, Carr MR (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
- Somerfield PJ, Clarke KR (1995) Taxonomic levels, in marine community studies, revisited. *Mar Ecol Prog Ser* 127:113–119
- Underwood AJ (1994) Things environmental scientists (and statisticians) need to know to receive (and give) better statistical advice. In: Fletcher DJ, Manly BJ (eds) *Statistics in ecology and environmental monitoring*. University of Otago Press, Dunedin, p 33–61
- Underwood AJ (1997) *Experiments in ecology: their logical design and interpretation using analysis of variance*. Cambridge University Press, Cambridge
- Underwood AJ, Chapman MG (1998) Variation in algal assemblages on wave-exposed shores in New South Wales. *Mar Freshw Res* 49:241–254
- Underwood AJ, Skilleter GA (1996) Effects of patch-size on the structure of assemblages in rock pools. *J Exp Mar Biol Ecol* 197:63–90
- Warwick RM (1988) Analysis of community attributes of the macrobenthos of Frierfjord/Langesundfjord at taxonomic levels higher than species. *Mar Ecol Prog Ser* 46:167–170
- Warwick RM, Clarke KR, Gee JM (1990) The effect of disturbance by soldier crabs, *Mictyris platycheles* H. Milne Edwards, on meiobenthic community structure. *J Exp Mar Biol Ecol* 135:19–33

Editorial responsibility: Otto Kinne (Editor), Oldendorf/Luhe, Germany

*Submitted: May 4, 1998; Accepted: December 7, 1998
Proofs received from author(s): April 7, 1999*