ON SOME STATISTICAL PROBLEMS IN MARINE BENTHIC ECOLOGY

Ib Svane

*The Royal Swedish Academy of Sciences. Kristineberg Marine Research Station. Kristineberg 2130. S-450 34 Fiskebäckskil. Sweden. *Present address: SARDI Aquatic Sciences. PO Box 120, Henley Beach. SA 5022. Australia*

ABSTRACT

Statistical procedures in field and experimental ecology frequently cause problems to many researchers in the process of design of sampling programs, performance of field experiments, test of hypotheses, and their subsequent interpretation. It is my aim to discuss problems of observation and the logic components of research programs. I suggest ways of describing spatial and temporal patterns, emphasising some problems in statistical sampling methods and highlighting the importance of power analyses.

INTRODUCTION

The Ryoan-Ji Buddhist Temple of the Peaceful Dragon in Kyoto, is the site of one of the most famous rock-and-sand Seki-Taki gardens of Japan. Confusion exists regarding the origins of the garden but it most likely originates back to 1470 AD, created by an anonymous designer according to a principle of Zen, that insight is gained by meditation.

The astonishing feature of the garden is its simplicity. It consists of a bed of white and grey gravel with 15 stones, grouped irregularly in 5 moss-covered islands, apparently in no particular order. The garden is 10 by 30 meters, bounded on three sides by a low parapet wall and can only be viewed from the open side adjoining the temple. From this position, the viewer cannot see all 15 stones at the same time since any perspective allows only the view of 12 to 14. To comprehend the full design of the garden, and get a full count of the number of stones, the viewer has to change the perspective, remembering previous perspectives, integrating all the information to reveal its true nature. The 15th invisible stone appears only to the enlightened eye.

This situation is usually what meets ecologists when describing and interpreting patterns of nature. A single perspective cannot provide a full understanding and much will be hidden from our view. To reveal patterns by integrated observations, ecologist use statistics, but many new students of ecology frequently ask the question: "Why use statistics?" An answer to this question can be found in the analogy of the observation of the stone pattern of the Ryoan-Ji Garden: we can only know the characteristics of the true population of stones by integrating our estimates, based on sampling. In nature, all biological variables have a distribution and comparisons of distributions require statistics.

The purpose of this paper is to highlight the most common statistical problems I have encountered in the discussions with the TMMP participants and in the numerous papers and research proposals I have reviewed. I will do this by presenting a " statistical lift-off" by discussing important aspects of the sequences of events in ecological research, leading from observation to hypothesis testing, to interpretation and further to formulation of new hypotheses.

PROBLEMS OF OBSERVATION

The fundamental aim of ecological research is to explain patterns observed in nature.

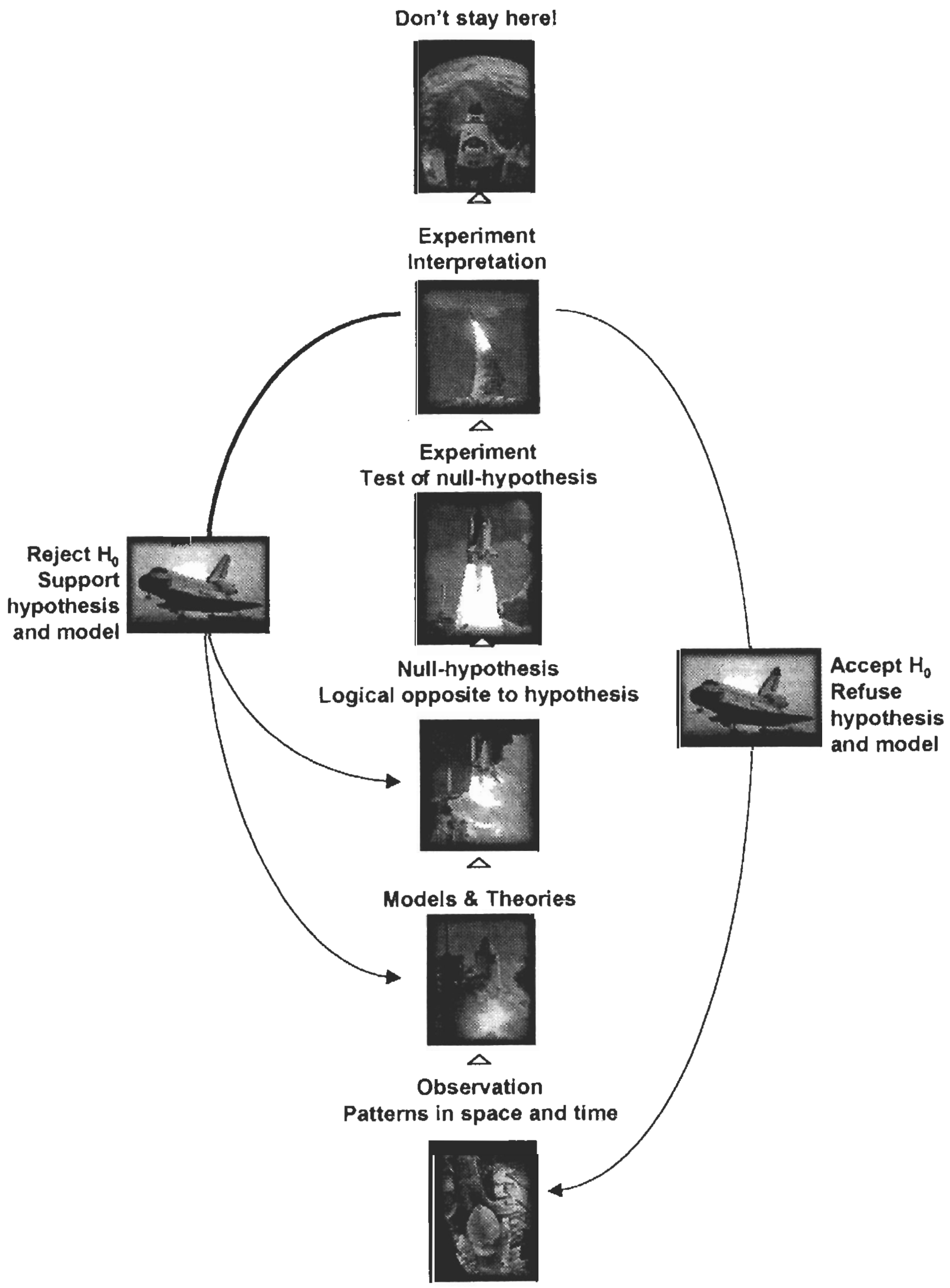


Fig. 1. Statistical lift-off describing the logical events in an ecological investigation starting at observation of patterns in nature, formulation of models and theories, formulating of a null hypothesis, experimental tests, interpretations, and new initatives based on rejection or acceptance of null-hypotheses (modified from Underwood, 1997).

One observation is that there is always some change or variation. No living object is perpetually constant, because metabolic processes are the fundamental characteristics of all life. Populations, assemblages, and ecosystems must change accordingly. Nevertheless, we do observe some constancy allowing us to do research on a particular object at different sites and at different times. On many occasions, therefore, the problem for an observer is the problem of scales of observation. This involves two major components namely variation in space (variation within and between units of observations = functional and structural scales) and variation in time (temporal scales). Irrespective of scale, the first task for the observer is to explain why a particular observation is made and why others are ignored. The second task is to demonstrate that the observation represents a natural phenomenon. The third task is to formulate an explanation, model or theory leading to an hypothesis. The fourth task is to test such an hypothesis by experiment. In many cases investigations end here. However, researchers have an obligation not only to interpret the analyses, but also to further improve models based on rejection or retainment of the null hypothesis. The generalised scheme in Fig. 1 shows the "lift-off" describing the logical steps in an ecological investigation (see Underwood, 1997).

A sound experimental design and a subsequent statistical test is a necessity in ecological research. Recent developments of statistical methods have greatly increased our understanding, thus making it possible to avoid the many pitfalls, notably problems concerning so-called "pseudoreplication" (Hulbert, 1984; Stewart-Oaten *et al.*, 1986; Underwood, 1991). Most ecological research programs, however, are constrained by financial realities and reductions in ambition levels have to be made. This usually involves decisions about scales of observation, since it is logistically impossible to deal with all scales and all sources of variation. Consequently, choices have to be made. It is thus necessary for the scientists to clearly state the objectives and the statistical limitations of a proposed research program, for example by applying an appropriate power analysis (Andrew, 1987; Cohen, 1988; Fairweather, 1991; Searcy-Bernal, 1994).

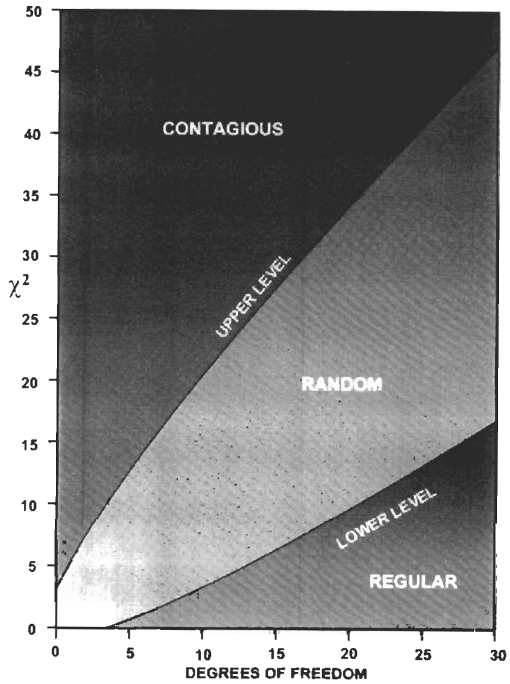
MODELS AND HYPOTHESES OF SPATIAL AND TEMPORAL PATTERNS IN BENTHIC ECOLOGY

Null hypotheses

When making a prediction about observed patterns in nature, ecologists formulate a hypothesis or a hypothetical prediction and subsequently design experimental tests by creating the conditions required by the hypothesis. To design an experiment that without any doubt proves the hypothesis is usually not possible, because this requires that any possible observation is available, or it requires the assumption that what happens in all possible circumstances not actually observed can be inferred from the cases actually available in the experiment (Underwood, 1997). It is, however, not necessary, and in many cases not possible, to prove the true nature of observed patterns. Instead, we can use the " falsificationist procedure" (Popper, 1968), attempting to falsify our hypothesis by constructing the logical null hypothesis, that is the opposite statement of our hypothesis. This is more simple because only one observation is necessary to disprove a hypothesis.

Spatial patterns

Marine benthic organisms are not usually randomly distributed. It is therefore of ecological importance to understand causes of variation in distribution. Causes of non-random distribution patterns are many but the most important ones are related to the



The 5% significance levels of χ^2. If χ^2 values are between significance levels, agreement with Poisson series is accepted at 95% probability level (P > 0.05)

χ^2 test (variance to mean ratio for agreement with a Poisson Series)

$$\chi^2 = \frac{s^2(n-1)}{x}$$

$$I = \frac{\text{sample variance}}{\text{theoretical variance}} = \frac{s^2}{\bar{x}}$$

$$\frac{s^2}{\bar{x}} < 1 \Rightarrow \text{regular}$$

$$\frac{s^2}{\bar{x}} \cong 1 \Rightarrow \text{random}$$

$$\frac{s^2}{\bar{x}} > 1 \Rightarrow \text{contagious}$$

Fig. 2. Test of the null hypothesis of spatial randomness using Chi-square test for the variance to mean ratio for agreement with a Poisson Series or Index of Dispersion. Redrawn from Elliott (1977).

nature of the substratum i.e. orientation (light), inclination (sedimentation), topography (small-scale heterogeneity) and structure (excavatability). These factors interact with already established assemblages or patches, attracting or repelling larval settlement and recruits either by biological, chemical or physical effects (Pawlik, 1992). To test the null hypothesis of spatial randomness is therefore important. The test is two-tailed; that is if the null hypothesis is rejected, two possible alternatives are available, either clumped or regular. If the null hypothesis is accepted, indicating a random distribution, then no further description is necessary (Andrew, 1987). If the null hypothesis is rejected, then various statistical methods and indices of different complexity are available to describe the observed patterns. The most simple and commonly used is the variance/mean ratio (Fig. 2) but also goodness-of-fit tests (χ^2 and G-tests or the non-parametric Kolmogorov-Smirnov test) can be used by comparing the estimated observed frequency of distribution with an expected distribution (see Andrew, 1987; Elliott, 1977; Fowler & Cohen, 1990). To estimate the significance of distance between organisms or patches of organisms, the so-called "Nearest Neighbour Index" can be used to test for randomness (Clark & Evans, 1954).

Temporal patterns

In the study of temporal patterns of structural variables, two statistical methods are applicable, namely univariate and multivariate analyses, where the latter are novel statistical analyses suited for multi-species assemblages (Clarke, 1993). Multivariate analyses have proved useful in detecting environmental changes in situations where the variance in single measures is too great for univariate analyses. There is, however, a problem with multivariate analyses because they cannot deal with any complexity beyond very simple nested or hierarchical sampling designs (equivalent to a two-factor experimental design with no interactions) (Underwood, 1996). Nevertheless, by using multivariate analyses an opportunity exists to reduce the cost of sample analysis, because this can be done with less taxonomic resolution (Gray et al., 1990; Warwick, 1988).

When sampling through time, small-scale temporal and spatial variations have the potential to confound large-scale comparisons (Morrisey et al., 1992; Underwood, 1991). When planning a monitoring program, the question usually asked is: "How often are the stations going to be sampled?" Many programs are designed so that samples are taken at fixed stations at regular intervals. These programs are confounded by the problems of pseudoreplication (Hulbert, 1984). Firstly, fixed stations may not adequately estimate spatial variation and replicates, if not randomised, are in many cases dependent (or at least not demonstrated to be independent). With regular temporal sampling, successive time series of samples may also be dependent (correlated with each other) because the observer returns to the same site, and cyclic differences may be overlooked (Green, 1993; Underwood, 1991). The design of a program must ensure that sampling is random, representative, but also random in time if time is a factor under investigation. A program of sampling at random intervals is therefore to be recommended and all stations should accordingly be sampled simultaneously, which may be impractical. A solution to the problem is to use a sampling schedule incorporating nested, replicated sampling times at different temporal scales. The contribution of each scale to the overall variation among samples can then be estimated (Morrisey et al., 1992). The choice of time scale is dependent on what kind of resolution is required to obtain the set goal. The nested hierarchical design for temporal variance is recommended and discussed in detail by Underwood (1991). The principle is suitable for detection of environmental impact, and is also applicable in experimental designs where variation in time is considered.

SAMPLING PROBLEMS

When deciding on a sampling method the first problem encountered is usually to determine the number and the size of quadrats or cores (=sampling units) to be used. With respect to number of sampling units, it is desirable that any sampling program is balanced (the number of replicates must be consistent for all factors). Analysis of unbalanced data is possible (but statistically less reliable) and most available computer programs reject unbalanced data or re-balance the data by randomly omitting values prior analysis.

Both the number and size of sampling units are related to the desired precision of the measured variables and are fundamentally related to the characteristics of the assemblages studied. By increasing the size of a sampling unit, precision will initially increase, but eventually decreases as the size of the sampling unit exceeds the mean distance between patches. In other words, unit size must be related to number and size of the sampled organisms: sampling elephants on the Serengeti Plains requires a large sampling unit, while sampling ants probably requires a relatively smaller unit! Because organisms differ considerably in size and form, a choice of unit size (and shape) will always be a compromise. Pringel (1984) reviewed the problem of sampling units and found in a study of macrophyte biomass that a large number of small sampling units (0.25 m^2) was more efficient than a smaller number of larger units (4.0 m^2).

In already established assemblages, estimation of a species-area curve (or individual/patch area) may solve the problem of size of sampling unit (Weinberg, 1978). As a rule of thumb one can say that

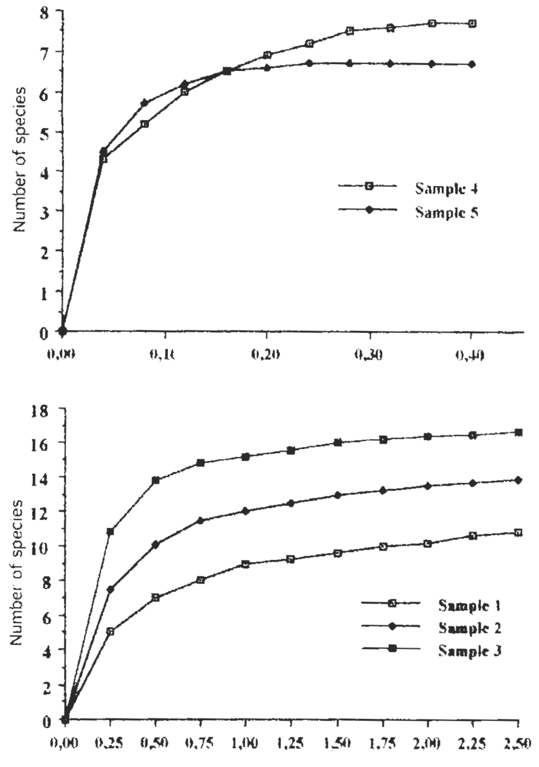


Fig. 3. Species-area curve for a Mediterranean rock epifauna in dark caves (above) and open assemblages (below). Redrawn from Weinberg (1978).

an acceptable area is sampled if the number of species/individuals/patches does not increase more than 10% when the sampled area is doubled (Fig. 3). Hawkins & Hartnoll (1980) used 0.25 m^2 quadrats and found that the structure of the assemblage and the size range of the individual organisms will have a considerable effect on the minimal area sampled. Hawkins & Hartnoll (1980) found for a rocky intertidal site, when using sampling units of 0.25 m^2, that an increment of 5% in species number was reached when a total area of 1.25 m^2 was sampled.

The minimum number of sampling units (replicates) for a statistical analysis is three (simple random sampling), but with such a low number an accurate estimate of the true population mean in natural marine benthic assemblages may not be obtained. Significant differences among means may only reflect differences in the relative accuracies. This is a consequence of heterogeneity. However, heterogeneity differs among assemblages and by the degree of development. If transect sampling (simple random sampling) is used the minimum number is also three for statistical comparison, because using transects rather than quadrats just increases the area sampled. Transects may allow a stratified design for comparisons with nested analyses of variance that nest strata within sites. The minimum number of replicates (sampling units) necessary is determined by homogeneity of the variance of the mean, which is a prerequisite for performing an analysis of variance. It is thus recommended that a pilot study be performed to determine the efficient number of replicates required. A simple approach is to determine precision by calculating:

$$p = \frac{SE}{\bar{x}} \quad \text{thus}$$

$$n = \left[\frac{s}{p\bar{x}}\right]^2$$; (p = precision, n = number of replicates, s = standard deviation) (Andrew 1987).

POWER ANALYSIS

In an experiment, the null hypothesis (H_0) is usually tested against an alternative hypothesis (H_a). We may reject H_0, and suggest treatment effects, or accept H_0, suggesting no treatment effects. Because our data are based on randomised sampling, a certain probability of making an erroneous conclusion is always present. Such a risk is found in all statistical tests. For a given number of samples and a sample size with a certain variance and error, the probability that an effect will be detected can be calculated. This probability is called " statistical power". To know the statistical power of our test is important because to make erroneous conclusions may have financial consequences both for our project (wasted working hours) and maybe also for our employer or the community (strategic

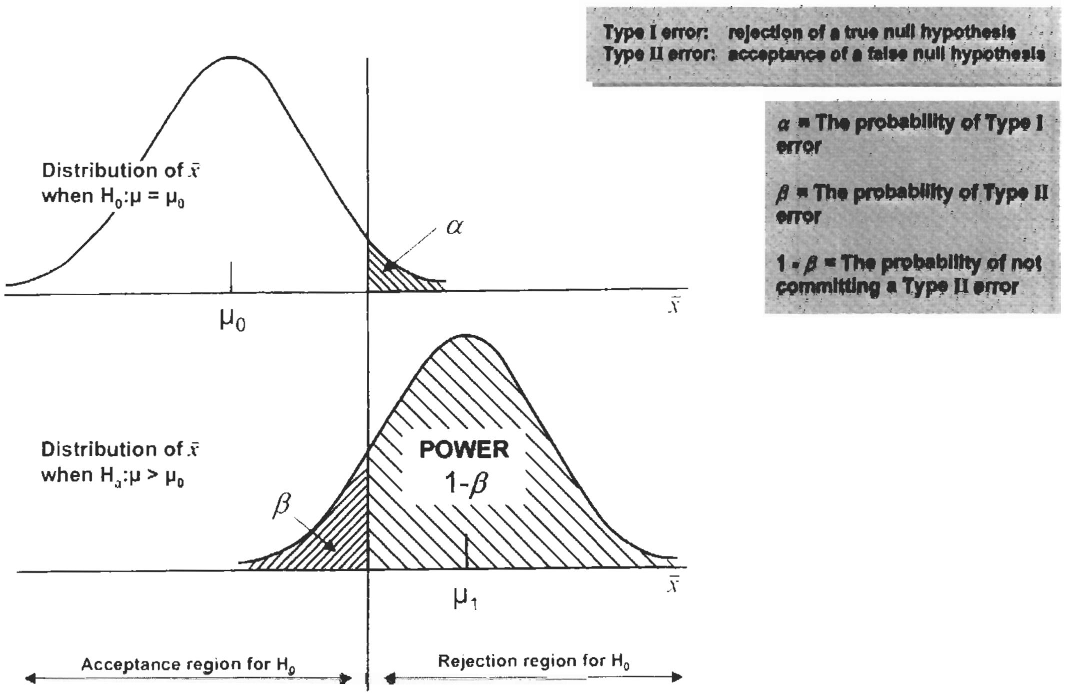


Fig. 4. Statistical errors, hypothesis testing and power in a one-tailed test (modified from Searcy-Bernal, 1994).

decisions).

In a statistical analysis we may make two kinds of errors, namely type I error: rejection of the true null hypothesis, or type II error: acceptance of the false null hypothesis. The probability of committing a type I error is called α and type II error is called β. In most cases we are not concerned with β but only α, which is chosen by convention to be 5%. This indicates that we always have 5% probability of rejecting a true null hypothesis ($\mu = \mu_0$), and making the conclusion that our measured μ does not belong to the population with a mean of μ_0 ($\mu \neq \mu_0$) (Fig. 4, above). If we accept H_0 then the probability of committing a type II error is β. Accordingly, the probability of not committing a type II error is 1-β (Fig. 4, below). Statistical power is defined as 1-β, expressing the probability of not committing a type II error (Fig. 4).

The power of a statistical test is a function of α, the number of treatments (k), the number of replicates (n), the population standard deviation (σ) and the effect size (ES). Effect size is a measure of the overall effect of the treatment on the response variable and describes how large the difference is between the null hypothesis and the alternative hypothesis. The β-value is inversely related to the effect size and proportional to the variability in the data set. β can only be indirectly controlled by the researcher.

In order to measure the power of a statistical test it is necessary to use tabulated values of effect size which can be found in Cohen (1988). According to Searcy-Bernal (1994), for a balanced (equal number of replicates) one-way ANOVA, the method to determine power is to calculate the standardised effect size index (f), given by:

$$f = \frac{\sigma_\mu}{\sigma} = \sqrt{\frac{\sum_{i}^{k}(\mu_i - \mu)^2}{k\sigma^2}}$$

where μ_i is the population mean for each treatment, μ is the grand population mean.

By consulting power tables provided by Cohen (1988) and Searcy-Bernal (1994) (one-way ANOVA), power values for effect size index (f) and n replicates can be found. Another method to determine power of ANOVA can be found in Underwood (1981), Underwood (1997) and Fairweather (1991). According to Cohen (1988), a power of 0.8 is the minimum desirable value in order to accept that the test has an adequate power.

ACKNOWLEDGEMENTS

I am grateful to Dr. Jørgen Hylleberg and DANIDA for financial support allowing me to participate in the Tropical Marine Mollusc Programme (TMMP). I am furthermore grateful to the participants of the TMMP who through their many research activities, papers and discussions generate a fruitful environment for achivement and progress in marine science and to Dr. Richard Emlet for linguistic corrections and valuable suggestions.

REFERENCES

Andrew, N.L., 1987. Sampling and the description of spatial pattern in marine ecology. - Oceanogr. Mar. Biol. Ann. Rev. **25**: 39-90.

Clark, P.J. & F.C. Evans, 1954. Distance to nearest neighbour as a measure of spatial relationship in populations. - Ecology. **35**: 23-30.

Clarke, K.R., 1993. Non-parametric multivariate analyses of change in community structure. - Aust. J. Ecol. **18**: 117-143.

Cohen, J., 1988. Statistical Power Analysis for Behavioural Sciences. L. Erlmbaum Associates, Hillsdale, NJ.

Elliott, J.M., 1977. Some methods for the statistical analysis of samples of benthic invertebrates. - Freshw. Biol. Ass. Sci. Publ. No. 25, 159 pp.

Fairweather, P.G., 1991. Statistical power and design requirements for environmental monitoring. - Aust. J. Mar. Freshwater Res. **42**: 555-567.

Fowler, J. & L. Cohen., 1990. Practical statistics for field biology. John Wiley & Sons, Chichester, England.

Gray, J.S., K.R. Clarke, R.M. Warwick & G. Hobbs, 1990. Detection of the initial effects of pollution on marine benthos: an example from the Ekofisk and Eldfisk oilfields, North Sea. - Mar. Ecol. Prog. Ser. **66**: 285-299.

Green, R.H., 1993. Application of repeated measures designs in environmental impact and monitoring studies. - Aust. J. Ecol. **18**: 81-98.

Hawkins, S.J. & R.G. Hartnoll, 1980. A study of the small-scale relationship between species number and area on a rocky shore. - Estaur. Coast. Mar. Sci. **10**: 201-214.

Hulbert, S.J., 1984. Pseudosampling and the design of ecological field experiments. - Ecological Monographs. **54**: 187-211.

Morrisey, D.J., A.J. Underwood, L. Howitt & J.S. Stark, 1992. Temporal variation in soft-sediment benthos. - J. Exp. Mar. Biol. Ecol. **164**: 233-245.

Pawlik, J.R., 1992. Chemical ecology of the settlement of benthic marine invertebrates. - Oceanogr. Mar. Biol. Ann. Rev. **30**: 273-335.

Popper, K.R., 1968. The logic of scientific discovery. Hutchinson, London.

Pringel, J.D., 1984. Efficiency estimates for various quadrat sizes used in benthic sampling. - Can. J. Fish. Aquat. Sci. **41**: 1485-1489.

Searcy-Bernal, R., 1994. Statistical power and aquaculture research. - Aquaculture. **127**: 371-388.

Stewart-Oaten, A., W.M. Murdoch & K.R. Parker, 1986. Environmental impact assessment: 'pseudoreplication' in time? - Ecology. **67**: 929-940.

Underwood, A.J., 1981. Techniques of analysis of variance in experimental marine biology and ecology. - Oceanogr. Mar. Biol. Ann. Rev. **19**: 513-605.

Underwood, A.J., 1991. Beyond BACI:

Experimental designs for detecting human environmental impacts on temporal variations in natural populations. - Aust. J. Mar. Freshwater Res. **42**: 569-587.

Underwood, A.J., 1996. Detection, interpretation, prediction and management of environmental disturbances: some roles for experimental marine ecology. - J. Exp. Mar. Biol. Ecol. **200**: 1-27.

Underwood, A.J., 1997. Experiments in Ecology. Cambridge University Press, 499pp.

Warwick, R.M., 1988. The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. - Mar. Poll. Bull. **19**: 259-268.

Weinberg, S., 1978. The minimal area problem in invertebrate communities of Mediterranean rocky substrata. - Mar. Biol. **49**: 33-40.