Aust. J. Mar. Freshwater Res., 1991, **42**, 569–87

Beyond BACI: Experimental Designs for Detecting Human Environmental Impacts on Temporal Variations in Natural Populations*

A. J. Underwood

Institute of Marine Ecology, Zoology Building A08,
University of Sydney, NSW 2006, Australia.

Abstract

Biological effects of environmental impacts are usually defined simplistically in terms of changes in the mean of some biological variable. Many types of impact do not necessarily change long-run mean abundances. Here, designs for detection of environmental impact are reviewed and some of their shortcomings noted. New sampling designs to detect impacts that cause changes in temporal variance in abundance of populations, rather than their means, are described. These designs are effective at distinguishing pulse and press episodes of disturbance and could be used for other variables of interest (size, reproductive state, rate of growth, number of species, etc.) for monitoring. The designs require sampling different time-scales before and after a proposed development that might cause impact. Cases are discussed in which there is a single control location. Inadequacies of this approach for detection of environmental impact are mentioned, with some discussion of the consequences for management of impacts that cause temporal change rather than alterations of the mean abundance of a population.

Introduction

The assessment of environmental disturbance by humans is a matter of increasing concern but remains an area of primitiveness in terms of the designs of programmes of monitoring and evaluation (Green 1979; Underwood 1989, 1991*a*). Much, if not all, of the focus of monitoring in most habitats is on detection of changes in the mean abundance, size, diversity of species or whatever variable is considered to be appropriate. Sampling to detect any change in means is often oversimplistic and based on poor logic. Even where a change can be identified that is correlated with whatever human activity (building, development, pollution, harvesting of resources, etc.), it is not often clear that the relationship between human activity and the perceived disruption of the environment is causal (Underwood and Peterson 1988).

The preoccupation with changes in averages is often made even more illogical by the notion that only decreases in numbers of organisms (or their sizes, or the richness of species) are considered as deleterious environmental impacts. For example, the environmental impact statement for the proposed development of yet another runway in Botany Bay (New South Wales) stated that the new sea-walls would be of benefit to some organisms because they would cause increased numbers of fish (Kinhill Engineers 1990). Apart from begging the question of whether the fish would actually be increased in number or would simply be aggregated around these man-made structures, the assumption that increases in numbers of fish are beneficial is not well-conceived (except from the point of view of fishermen). The fish presumably eat something else or are involved in other ecological interactions, potentially creating increases or decreases in abundances of numerous other species. An increase in numbers of fish (or, for that matter, organisms that cause much more emotional

---

*Based on a paper presented at the workshop 'Detecting Human Impacts on Aquatic Environments', University of Melbourne, November 1990.

0067-1940/91/050569$10.00

human reactions, such as kangaroos or koalas) is a measurable response to disturbances of the environment. A considerable body of opinion has led to ecological theory that there *should* be increased diversity or numbers of species in response to increased perturbation in habitats where disturbances are naturally rare or small (Connell 1978; Sousa 1979*a*, 1979*b*, 1980; McGuinness 1984). So, even if we are to be preoccupied with average conditions, any change in means that is correlatable to the start of some new human activity must be considered an environmental impact.

All in all, assessment of environmental changes should not be totally preoccupied with changes in averages of some variable. These are not the only types of environmental change that matter. In this paper, therefore, sampling designs are considered that should be useful for detecting environmental changes that do not themselves cause any differences to long-run averages of densities of populations. Instead, the results of these environmental perturbations are manifested in changes of temporal variance in, for example, the numbers of organisms in a population. Thus, human disturbances may cause changes in the rates or magnitudes of fluctuations in numbers of a population around their long-run averages. If sampling is too infrequent, which costs of monitoring programmes often cause, many types of temporal change may not be detected. Alternatively, the effect of environmental disturbance may be to alter the frequency or timing of important ecological events, such as recruitment of juveniles to populations, which is important in many coastal marine processes (Underwood and Denley 1984; Underwood and Fairweather 1989). As a result, the mean numbers of animals or plants over several months may not be altered, but the time-course of changes in numbers may be affected, with potential consequences for the probability of the population going extinct.

Before designs to detect such disturbances are examined, existing sampling and monitoring designs are briefly reviewed. The designs fall into several classes, some of which are popularly used and some of which are touted as being 'optimal' (Bernstein and Zalinski 1983). Most, if not all, have serious flaws in their logic or their use for statistical analysis (see also Underwood 1990). Some have serious limitations in terms of experimental design, including a lack of spatial controls, so that detection of environmental impact is not possible against a background of differential variability in numbers of organisms from site to site. Possible solutions to this major problem will, because of lack of space here, be suggested elsewhere (Underwood, unpublished data).

Finally, the consequences of detecting such environmental disturbances for the management or rehabilitation of disturbed habitats are briefly considered.

Sampling Designs for Detecting Environmental Impact

Several different types of sampling design have been used in monitoring and assessment of actual or potential environmental damage. These were reviewed most extensively by Green (1979). The major types are considered briefly here, with some commentary on their inadequacy, including points not made by Green (1979) and other more recent discussions.

Before/After Contrasts at a Single Site

The simplest design that could be used to detect a change in mean abundance of an organism before and after some putative environmental disturbance is a single sample taken in the site before and a single sample taken after the potential disturbance (Fig. 1*a*). This is widely used in response to certain obvious accidental incidents of potential impact, such as oil spills, where, fortuitously, some prior information was available. If subsequent sampling reveals differences, these are attributed to the event. Obviously, there may be no relationship between the observed event and the change in numbers of an organism. The change may have been due to any intrinsic cause, entirely coincidental to the observed human activity. There are no controls in time or space to demonstrate whether such changes are not widespread without there being an incident of pollution or any other human action in that site.

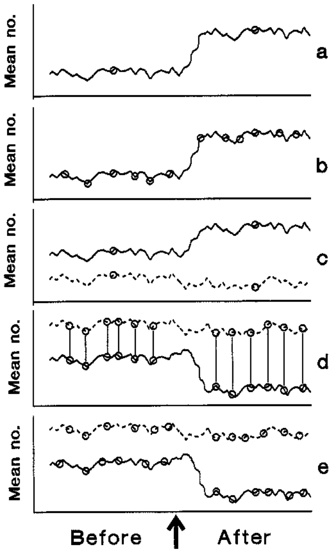

Fig. 1. Common sampling designs to detect environmental impacts, with circles indicating times of sampling: (*a*) a single sample in one location before and after an impact (at the time of the arrow); (*b*) random samples in one location before and after an impact; (*c*) BACI design as proposed by Green (1979), with a single sample before and after the impact in each of a control (dashed line) and the putatively impacted location (solid line); (*d*) BACI according to Bernstein and Zalinski (1983) and Stewart-Oaten *et al.* (1986); differences between mean abundances in the control and potentially impacted locations are calculated for random times of sampling before and after the disturbance begins (vertical lines indicate differences); (*e*) modified BACI design (*d*) to allow sampling at different times in each location (see text for further details).

These irrational forms of monitoring are sometimes even more badly conceived in that no quantitative data were collected before an obvious incident. Information is, however, gathered afterwards in an attempt to demonstrate that environmental change has occurred, but there is no prior information against which to assess it. This might even be termed the 'Australian Environmental Lobby Design' because it is so widely used as a means of provoking environmental confrontation. It seems to be a common response to politically inspired decisions to do something about, for example, an oil spill, any inadvertent release of chemicals, or other highly visible intrusion into a natural habitat (e.g. for oil spills, Baker *et al.* 1976; McGuinness 1988).

Repeated Before/After Sampling at a Single Site

One possible embellishment on the theme of sampling before and after a planned or accidental disturbance to a site would be to take repeated, randomly timed samples at the location prior and subsequent to the development or accident that might cause environmental change (Fig. 1*b*). Such a design is, again, only able to detect that a change in mean numbers of the sampled population has occurred coincidentally with the onset of the potential environmental disturbance. Data of this sort cannot indicate that there is some relationship between the change in numbers and the putative disturbance (Underwood and Peterson 1988; Underwood 1989, 1991*a*). The data are 'pseudoreplicated' (Hurlbert 1984); any change in numbers may be due to any number of causes (including that identified as a disturbance by man). Statistical analysis of such data is, however, quite straightforward (Table 1*a*). It is remarkable how often such sets of data are used (e.g. Bachelet 1986; Buchanan and Moore 1986; Dauvin and Ibanez 1986; Lopez-Jamar *et al.* 1986) or are even recommended (Lundalv *et al.* 1986) in environmental work.

Stewart-Oaten *et al.* (1986), in a different context, identified the great advantages of taking the samples at random intervals of time, rather than on some fixed schedule. Marine ecologists, by and large, usually wish to take samples on some regular basis, apparently because of an intrinsic belief that four seasons of the year, or some monthly schedule of changes, may be more usefully investigated in this way. For example, current environmental studies in Jervis Bay (New South Wales) are dominated by contractual obligations to take samples every 3 months (CSIRO 1989—a schedule so far justifiable only in terms of the schedule of payments made by CSIRO to their contractors). As Stewart-Oaten *et al.* (1986)

Table 1. Statistical analyses for the detection of environmental impact, using various sampling designs

In each case, analysis of variance is used to provide a standard framework for all designs; alternative analyses are possible for some of the simpler designs. In all cases, n replicate samples are taken at each time and site of sampling. In this and subsequent tables, expected values of Mean Squares and, thereby, F ratios were calculated according to Scheffe (1959), Winer (1971) and Underwood (1981)

a. Replicated Before/After sampling at a single location; samples are taken at t random times before and t times after the putative impact (see Fig. 1*b*)

Source of variation		Degrees of freedom	F ratio versus	Degrees of freedom
Before versus After	= B	1	T(B)	1, $2(t-1)$
Times (Before versus After)	= T(B)	$2(t-1)$		
Residual		$2t(n-1)$		
Total		$2tn-1$		

b. BACI: A single time of sampling at two locations, one Control and one potentially Impacted (see Fig. 1*c*)

Source of variation		Degrees of freedom	F ratio versus	Degrees of freedom
Before versus After	= B	1		
Locations: Control versus Impact	= L	1		
Interaction B × L		1	Residual	1, $4(n-1)$
Residual		$4(n-1)$		
Total		$4n-1$		

c. BACI: Replicated Before/After sampling at two locations, one Control and one potentially Impacted; samples are taken at t random times before and t times after the putative impact, but at the same times in each site (see Fig. 1*d*)

Source of variation		Degrees of freedom	F ratio versus	Degrees of freedom
Before versus After	= B	1		
Locations: Control versus Impact	= L	1		
Interaction B × L		1[A]	L × T(B)	1, $2(t-1)$
Times (Before versus After)	= T(B)	$2(t-1)$		
Interaction L × T(B)		$2(t-1)$[A]	Residual	$2(t-1)$, $4t(n-1)$
L × T(B) Before[B]		$t-1$	Residual	$t-1$, $4t(n-1)$
L × T(B) After[B]		$t-1$	Residual	$t-1$, $4t(n-1)$
Residual		$4t(n-1)$		
Total		$4tn-1$		

[A] This is the same test as the t test recommended by Stewart-Oaten *et al.* (1986).

[B] Repartitioned sources of variation; Impact in an interactive system can be examined by an F ratio of Mean Square C × T(B) After/Mean Square C × T(B) Before, with $(t-1)$ and $(t-1)$ degrees of freedom; this is a two-tailed test (see text).

Table 1 (*continued*)

d. BACI: Replicated Before/After sampling at two locations, one Control and one potentially Impacted; samples are taken at t random times before and t times after the putative impact, but at different times in each site (see Fig. 1*e*)

Source of variation		Degrees of freedom	F ratio versus	Degrees of freedom
Before versus After	= B	1		
Locations: Control versus Impact	= L	1		
Interaction B × L		1	T(B × L)	$1, 4(t-1)$
Times (B × L)	= T(B × L)	$4(t-1)$		
Residual		$4t(n-1)$		
Total		$4tn-1$		

suggested, taking samples at randomly placed intervals of time will tend to ensure that no cyclic differences unforeseen by the sampler will influence the magnitude of the difference before and after the onset of a potential environmental disturbance. Taking samples at regular, as opposed to randomly chosen, intervals means that temporal variance will not be estimated accurately and that the magnitude of an impact may be over- or underestimated. There are such serious consequences that regular temporal sampling should be avoided (see the more extensive discussions in Green 1979 and Stewart-Oaten *et al.* 1986).

BACI: Before/After and Control/Impact Sites

Green (1979) suggested that a so-called BACI design would generally be the most utilitarian for detecting environmental change. He managed to incorporate the necessity of a control site—one in which no human-induced change is planned, to contrast against the magnitude of change in the site in which a development is planned that may cause environmental damage. His design (Fig. 1*c*) involved a single sample taken in each of the two sites. One is the potentially disturbed site—usually called the 'Impact' site—even though no impact is yet known to be occurring. The second site is a similar area to serve as a control that can be sampled in identical fashion and independently of any change in the first site. Ideally, the two sites are sampled at the same time before and again after the developments.

The habit of calling one of these sites the 'Impact' site raises the potential confusion and bias caused by terminology in this field. Long before sampling is done to determine whether there is, in fact, an impact, one of the sites is already designated as being impacted. This is probably appropriate where human activity is inevitably going to affect biological variables and the only issue is the magnitude of the impact. In cases in which the existence of an impact is in doubt, the terminology is very biased.

As a result of having two sites, provided that the usual procedures are done to ensure that samples are representative, unbiased, independent and so forth (Snedecor and Cochran 1967; Dixon and Massey 1969; Green 1979; see Underwood 1981 for marine examples), the data are straightforward to analyse. The best test is a two-factor analysis of variance (provided its assumptions can be met), and the existence of some environmental impact would be shown by a significant statistical interaction between the two sources of variation (Control versus Impact sites and Before versus After times of sampling) (Table 1*b*). Such an interaction would indicate that the magnitude of difference from before to after in the mean numbers of the sampled population in the Control site is not the same as the difference between the means in the Impact site. This is a logical procedure for detecting differences in the amount of change in two populations from one time to another. It is not, however, rational to conclude that any interaction is *caused* by the human activity in one site that did not occur in the other, control site. There may be intrinsic changes due to any number of processes that cause changes in mean abundance in a population. It is not uncommon, particularly in marine examples, for populations in two locations to diverge or converge through time without there being an environmental impact caused by man at some particular time in one of the two locations. This is a common feature of natural populations (Underwood 1989). Thus, use of a single control site is not adequate to associate the presence of the statistical interaction with the putative impact of the activity in the impact site (see the related discussion in Bernstein and Zalinski 1983). Several control sites should be used (Underwood 1989). The design and analysis of data from monitoring one potentially impacted and several control sites will be described elsewhere (Underwood, unpublished data).

BACI: Repeated Before/After Sampling at Control and Impacted Sites

One of the most widely used designs is a modification of Green's (1979) BACI design and was foreshadowed by him as an extension of that method. It was first formally analysed by Bernstein and Zalinski (1983) and later discussed in detail by Stewart-Oaten *et al.* (1986). The design involves sampling the site that is planned to be affected by some development and a single control site. Each site is sampled several times, at random, prior to and then after the start of the potential disturbance. The two sites are sampled at the same times (i.e. times of sampling are orthogonal to the two sites), and there are similar (ideally, the same) numbers of samples taken before and after the planned development. As discussed in detail by Stewart-Oaten *et al.* (1986), times of sampling are random and thus form a random, nested source of variation in the data (Winer 1971; Underwood 1981). The times of sampling are nested in either the period before or the period after the potential impact starts.

The rationale behind the design is that the replicated sampling before the development gives an indication of the patterns of differences, over several periods of potential change of numbers of organisms, between the two sites. If the development in one site causes some change in the mean abundance of the sampled organisms, there will be a different magnitude of difference between the two sites after it starts from that prevailing before. By having several replicated times of sampling, it is possible to control for some random elements of difference between the two sites, in contrast to Green's (1979) design. It is no longer possible for the two sites to differ after the putative impact simply because, at a single time of sampling, they happen to have a different pattern of difference from that before. To be detected, an impact must cause a sustained pattern of difference.

Both Bernstein and Zalinski (1983) and Stewart-Oaten *et al.* (1986) considered that the best statistical analysis was to calculate the differences between the mean abundances in the two sites for each time of sampling and to analyse the difference from before to after in the means of these differences (Fig. 1*d*). Both sets of authors suggested a *t*-test, with appropriate elaborations to cover problems of nonindependence or nonhomogeneity of variances in the data. Stewart-Oaten *et al.* (1986) further warned that species of organisms that showed considerable interaction between the abundances in the two sites before the proposed development had started would be unsuitable for monitoring, because a pattern of difference from before to after would be expected even if there were no human impact.

The proposed statistical analysis using a *t*-test is, algebraically, identical to the *F*-ratio identified in the analysis of variance described in Table 1*c*. It is easy to demonstrate that the two tests are identical and have identical assumptions of homogeneity of variance and independence of samples (although this is not done here). To keep this design in the same context as others considered here, the framework of analysis of variance is used.

One possibility that could be explored for this design would be a solution to Stewart-Oaten *et al.*'s (1986) problem of significant Time × Site interaction before the putative impact. It is possible to test for impact even though such prior interactions occur. Consider the situation in which the difference in mean abundance of some monitored organism in the two sites is not constant from time to time. This will be detected as a significant interaction between time of sampling and site sampled before the possible disturbance begins. One test of the presence of impact is to partition the interaction (Site × Time) in the entire set of data (i.e. after the potential impact has started and data have been collected for several times from the two sites) into two components: that before and that after the possible impact (Table 1*d*). If an impact occurs, it must cause a change in the pattern of this interaction, which can be tested by an *F*-ratio of the two mean squares (Site × Time After versus Site × Time Before). This is a valid test, but it differs from the usual analysis-of-variance test because it is two-tailed. The magnitude of interactions between the sites may be greater or smaller as a result of an impact, thus the test is two-tailed in contrast to the one-tailed tests used in analysis of variance (Snedecor and Cochran 1967). The test is not problematic, but it will not be powerful (Cohen 1977; Underwood 1981; Bernstein and Zalinski 1983; Andrew and Mapstone 1987; Peterman 1990) unless numerous times are sampled before and after the potential impact, or unless there is very small temporal variance in the populations. The former is a relatively unlikely event, given that most environmental studies are of short duration—particularly before a planned development. There is obviously room to develop estimates of such interactions for monitorable species in a number of habitats so that the estimates could be used for particular instances of planned development without the need to collect samples over long periods of time separately for every new proposed disturbance of a given type of habitat.

Neither Bernstein and Zalinski (1983) nor Stewart-Oaten *et al.* (1986) considered the further problem that times of sampling must be identical for the two sites, which in some studies would not be possible because of the logistic support necessary for sampling, the time taken to complete samples in each site, or vagaries of weather that prevent the two sites being sampled together. This may not matter if the span of time over which the purportedly simultaneous sampling is completed is small relative to the rates of changes in the abundance of the population. This will, however, usually not be known whilst the study is in progress. As a result, a modification of the design should be considered that does not require the two sites to be sampled at the same times (Fig. 1*e*). This is presented in Table 1*d*. It involves no orthogonal, interactive contrast of Site × Time of sampling. Instead, Times are randomly sampled for each site separately; thus, Times of sampling are nested in the combination of Site and Before or After. This analysis is straightforward, but it must inevitably suffer from problems in which there are interactions between Site and Time but these cannot be examined at all.

Unfortunately, despite the assertions of authors such as Stewart-Oaten *et al.* (1986), the repeated-sampling BACI design *is still problematic in its interpretation*. Discovery of a difference between the two sites that is of different magnitude from the difference existing before the putative disturbance does not demonstrate that the human activity *caused* the impact. There is still no reason to expect two sites to have the same time-course of changes in mean abundance of a given species. What is needed is a set of control sites and then to demonstrate that the pattern of change in mean abundance of the population being sampled from before to after the onset of the human disturbance is greater in the Impact site than on average in the Control sites (Underwood 1989). This is possible using asymmetrical analysis of variance (Underwood 1978, 1984, 1986, 1991*b*; Underwood and Verstegen 1988) and appropriate sampling designs (Underwood, unpublished data). Nevertheless, if the temporally replicated design discussed by Bernstein and Zalinski (1983) and Stewart-Oaten *et al.* (1986) were expanded to include more than one control site, it would be a substantial improvement on any of the other designs considered above.

Sampling to Detect Changes in Temporal Variances

As indicated in the Introduction, whatever the potential problems for interpretation of the designs considered above, there remains the problem of how to sample for possible impacts that will not reveal themselves as differences in mean abundance of organisms after a planned or accidental human intervention in a site. Some types of these impacts are illustrated in Fig. 2 and considered in turn below.

Press and Pulse Disturbances

It is convenient to think of human disturbance as being in two classes—press and pulse (Bender *et al.* 1984). Pulse phenomena are short-term, acute episodes of disturbance that are then removed. An example is an oil spill, in which the incursion of oil is usually brief (although it may leave some longer-term aftermath in terms of chemical residues) and then removed. This is quite different from, for example, a daily discharge of sewage from a point source into a bay. The latter is an example of a press disturbance—one that is chronic and sustained. These are quite different in their potential effects on organisms or assemblages of species (Bender *et al.* 1984; Underwood 1989; see also the categories of impacts in McGuinness 1988). For consequences of the differences between the two, consult the cited reviews. There are difficulties with this dichotomy. Notably, if the time-scale of the disturbance is long or short relative to the time-scale or turnover time of the organisms, there is no really useful distinction between the two. For example, the Ranger Uranium Mine in the Northern Territory of Australia has a planned life of some 50 years. Thus, in terms of many ecological processes and many species with shorter life spans, it is a press phenomenon. Any impacts on local populations would be applied continuously throughout several generations. For organisms, including humans, with life cycles longer than 50 years, it is best considered as a pulse—it is only there for 50 years, or less than one generation. Despite such difficulties of definition, it is useful to categorize environmental disturbances into the two types. For example, prevention of pulse phenomena caused by accidents requires very stringent efforts, legislation and regulation. Rehabilitation or reduction of the effects may, however, be relatively easy, compared with press phenomena, because the cause of the impact is removed again quickly. In contrast, long-term press phenomena are much harder to detect until they have become of great magnitude (see Underwood 1989 for examples). As a result, monitoring programmes are less likely to detect them until it is too late. Management and rehabilitation of environments would then be difficult if not impossible. Many planned developments have both phenomena associated with them, each causing different responses in local populations. For example, building a marina involves pulse phenomena during the period of construction (e.g. movements of sediments, releases of chemicals from building materials), but these stop when, or shortly after, the construction is complete. The marina will, however, also cause long-term, possibly permanent, press changes in such variables as water flow and release of sewage or oil from boats. Any of these may cause press environmental responses. A sensible monitoring scheme would be able to detect both types of environmental impact, as considered below.

Nested Designs for Temporal Variance

Introduction

To detect the presence of some environmental impact that causes changes in temporal variance rather than simple effects on the average numbers of some organism, it is necessary to do sampling at different time-scales. This suggestion parallels that by Green (1979) in his excellent book, in which he recommended nested spatial sampling in monitored locations in case there was spatial heterogeneity in the variable being measured. The point here is that some types of environmental impact may cause temporal heterogeneity and thus require temporally hierarchical sampling designs. One method that should be able to detect and distinguish between pulse and press phenomena is to use two different time-scales, one longer than the other. This is illustrated, as an example, for a single potentially affected location in Fig. 2. Samples are taken before and after the onset of some putative environmental disturbance. In each case, data are collected from two periods, in each of which the population is sampled three times. At each time, there would be a series of replicated estimates of the population mean (to cover small-scale spatial variance in the location being sampled). The data thus form a nested (or hierarchical) series, with Before versus After a major source of variation, Periods nested randomly in each of these and Times of sampling nested randomly in each Period. Finally, replicate estimates at each time are fully nested inside all the other sources of variation. For the structure of hierarchical designs, consult Winer (1971) and Underwood (1981).

The advantages of this design are several. First, it may not actually cost any more than routine, fixed-interval monitoring, in terms of personnel, time, logistic costs, money, etc., because it could often be done by modifying the temporal scale of an existing or planned programme of monitoring. Often, monitoring is done routinely on some seasonal, or monthly, basis (or at some smaller time-scale) for no compelling reason except history, routine, fashion, or lack of imagination. Under such circumstances, the same total number of samples could still be taken but would be arranged differently to allow assessment of the patterns of difference in mean abundances of the monitored population at different time-

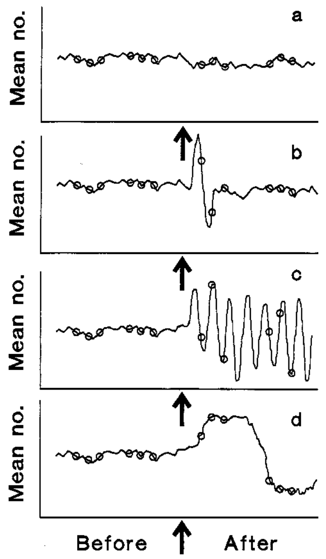

Fig. 2. Diagrams of mean abundance (or any other variable of interest) through time in one location subject to various environmental disturbances: (*a*) no environmental impact; (*b*) an impact causes pulse responses in the population; (*c*) an impact causes short-term, but continued, fluctuations in numbers around the previous mean; (*d*) an impact causes longer-term fluctuations. Circles indicate times of sampling (3 Times in each of 2 Periods Before and again After the impact starts as indicated by the arrows).

Table 2. Sample data for an imaginary population subjected to different types of disturbances
Four conditions are considered (as illustrated in Fig. 2): *a* is no impact, *b* is a pulse impact, *c* is a press disturbance causing increased rapid fluctuations, *d* is a press impact causing longer-term fluctuations (as in Figs 2*a*–2*d*). In each case, the location was sampled at three times (for example, 2 weeks apart) in each of two periods (for example, 6 months apart) before the disturbance and again after the beginning of the disturbance to the habitat. These data are analysed in Table 3

Condition	Before						After					
	Period 1, Time			Period 2, Time			Period 1, Time			Period 2, Time		
	1	2	3	1	2	3	1	2	3	1	2	3
a	100	120	110	115	103	114	120	111	104	109	118	103
b	100	120	110	115	103	114	65	155	111	109	118	103
c	100	120	110	115	103	114	65	155	111	67	148	119
d	100	120	110	115	103	114	140	150	135	65	80	73

Table 3. Analyses of a set of data as an example of detection of impacts in a single location

The location is subjected to disturbances causing different changes to temporal patterns in abundance of a population (see Fig. 2 and data in Table 2); the Residual variance is set at 100 and there are $n = 5$ samples at each time. Asterisks indicate a significant *F* ratio ($P < 0\cdot05$); *t* is a *t* test similar to that recommended by other authors (see text for explanation)

Source of variation		Condition:	*a*		*b*		*c*		*d*	
	Degrees of freedom	*F* ratio versus	Mean sqauare	*F* ratio	Mean square	*F* ratio	Mean square	*F* ratio	Mean square	*F*ratio
Before versus After = B	1	P(B)	203·75	0·96	200·42	0·99	203·75	0·99	350·4	0·02
Periods (Before or After) = P(B)	2	T(P(B))	212·08	0·40	202·08	0·07	205·42	0·04	18055·42	34·47*
Periods (Before)^{A}	1	T(P Bef)	203·33	0·36	203·33	0·36	203·33	0·36	203·33	0·36
Periods (After)^{A}	1	T(P Aft)	220·83	0·44	200·83	0·04	207·50	0·02	35907·50	73·78*
Time (Periods) = T(P(B))	8	Residual	532·08	2·66	2983·33	14·92*	5017·50	25·09*	523·75	2·62
Times (Periods Before)^{A} = T(P Bef)	4	Residual	560·83	2·80	560·83	2·80	560·83	2·80	560·83	2·80
Times (Periods After)^{A} = T(P Aft)	4	Residual	503·33	2·52	5405·83	27·03*	9474·17	47·37*	486·67	2·43
Times (Period 1 After)^{B} = T(P1 Aft)	2	Residual	521·67	2·61	10326·67	51·63*	10326·67	51·63*	491·67	2·46
Times (Period 2 After)^{B} = T(P2 Aft)	2	Residual	485·00	2·43	485·00	2·43	8621·67	43·11*	481·67	2·41
Residual	48		200·00	–	200·00	–	200·00	–	200·00	–
F ratio (two-tailed): P(Aft)/P(Bef)				1·09		0·99		1·02		176·59*
F ratio (two-tailed): T(P Aft)/T(P Bef)				0·90		9·64*		16·89*		0·87
F ratio (two-tailed): T(P1 Aft)/T(P2 Aft)				1·08		21·29*		1·20		1·02
t test (After – Before)				0·12		0·01		–0·03		0·20

[A,B] Repartitioned sources of variation.

scales. This would then cost no more than the usual pattern of sampling but would produce much more useful data.

A second advantage is that the information obtained will allow assessment of a whole range of types of environmental impacts (if they exist) rather than the routine one of determining that the long-run average of a population has somehow been altered.

The principle: one location only

To illustrate the procedures and their rationale, consider first the situation in which only one location is monitored before and after a proposed development that may cause environmental change affecting populations in that location. Obviously, this is not a satisfactory design (see earlier) and improvements are discussed below.

Where there is no impact (Fig. 2*a*), there will be no significant variation detected between Before and After in the patterns of differences among Periods and among Times within each period of sampling. In contrast, where there is a pulse disturbance, it temporarily affects a population by increasing the rate at which its numbers change around the long-run average that existed before the development. This is illustrated in Fig. 2*b*. Analyses of these two situations are presented in Table 3 (conditions *a* and *b*) for a hypothetical set of data (Table 2). Pulse disturbances would be detected, if the sampling were sufficiently intense to provide a powerful analysis, by a significant *F*-ratio for Times (After) versus Times (Before) and for Times (Period 1 After) versus Times (Period 2 After). Here, with only two Periods and three Times of sampling, the tests are significant, but in general they are not likely to be powerful. With more Periods and more Times, smaller changes would be detectable as significant. The Mean Squares and *F*-ratios in Table 3 serve to illustrate the sorts of changes that are to be expected under different types of impact. These tests are significant because the impact causes oscillation (creating great differences among means in the short term immediately after the onset of the disturbance), but these then disappear.

In contrast, press disturbances can cause differences either in the short-term rates of change (Times within Periods) or in the longer term (Between Periods), as illustrated in Figs 2*c* and 2*d*. Again, the analyses of these situations is straightforward (Table 3, conditions *c* and *d*). In the first case, there is, again, a significant *F*-ratio comparing Times within Periods After the onset of the disturbance with Times within Periods Before, but there is no significant difference between the patterns of difference among Times within each Period After. This is because the sustained disturbance, after it starts, does not disappear as with a pulse phenomenon. Where there are longer-term changes (Fig. 2*d* and Table 3, condition *d*), there are no differences detected among Times within the longer Periods, but there would now be differences between the Periods after the onset of the disturbance that were not present before. Thus, there would be a significant *F*-ratio for comparing Periods After and Periods Before the beginning of the development. In the case of some disturbance that caused short- and long-term fluctuations, both *F*-ratios (Times within Periods After and Periods After) would be significant.

In all cases, the analyses are able to detect significant patterns of difference after the start of the disturbance compared with before. Of course, the data must be sufficiently invariant and the number of replicate samples, times and periods sufficient to ensure suitably large power for any predetermined size of environmental impact that should be detected. What matters is that a whole series of different types of environmental impact could be detected by using a sampling regime of the sort described here. They could be distinguished by the different patterns of significance in the various statistical tests (Table 3). Cost–benefit procedures are ideal for determining the appropriate levels of replication of samples, times and periods for any given power (Snedecor and Cochran 1967; Underwood 1981; Kennelly and Underwood 1984, 1985).

Cost–benefit procedures would be particularly valuable because relatively small changes in the ways replication is allocated to different temporal scales will make potentially very large differences to the power of tests. For example, sampling two Periods before and two Periods after a putative impact, each with four Times, results in a quite different analysis from sampling with four Periods before and after, each with two times of sampling. Note that both schemes require the same amount of work; procedures for optimal allocation of effort to Periods and Times would clearly be of great value. Considerable thought and planning coupled with pilot studies and cost–benefit procedures are essential to design an optimal programme of sampling.

None of these impacts would be detected by the sort of simplistic analysis proposed as being optimal by Bernstein and Zalinski (1983) and Stewart-Oaten *et al.* (1986). To demonstrate this, a simple comparison was made with a theoretical control (i.e. undisturbed) location. To keep the modelled situation as simple as possible (see below for the complexity of spatial controls), consider that the undisturbed location has a constant mean abundance of the sampled organism of 200 individuals per sample unit. This is obviously unrealistic, but it serves to illustrate the point. The difference in mean abundance of the population between the control and disturbed locations was calculated for every time of sampling for the four sets of conditions in Table 2. These differences were then examined for any pattern of difference between the samples before the putative impact and those taken after. This is, in essence, the test proposed by Bernstein and Zalinski (1983) and Stewart-Oaten *et al.* (1986). In no case could these t tests detect any impact (Table 3, conditions *a*–*d*). These impacts could not be detected by the t tests because these tests are designed to detect only sustained, consistent differences. The disturbances considered here (Fig. 2 and Tables 2 and 3) produce a different type of effect, no less plausible biologically, but for which the traditional BACI design is completely inadequate. No one can tell in advance what

Table 4. Beyond BACI: Replicated Before/After sampling at two locations

One Control and one potentially Impacted location are sampled in p random periods before and p periods after the putative impact, but at the same times in each location; in each Period, t random times are sampled; at each Time, n random replicates are sampled

Source of variation		Degrees of freedom	F ratio versus	Degrees of freedom
Before versus After	= B	1		
Control versus Impact	= L	1		
Interaction B × L		1		
Periods (B)	= P(B)	$2(p-1)$	T(P(B))	$2(p-1),\ pt(n-1)$
Periods (Before)[A]	= P(Bef)	$p-1$	T(P(Bef))	$p-1,\ pt(n-1)$
Periods (After)[A]	= P(Aft)	$p-1$	T(P(Aft))	$p-1,\ pt(n-1)$
Interaction L × P(B)		$2(p-1)$	L × T(P(B))	$2(p-1),\ pt(n-1)$
L × P(Bef)[A]		$p-1$	L × T(P(Bef))	$p-1,\ pt(n-1)$
L × P(Aft)[A]		$p-1$	L × T(P(Aft))	$p-1,\ pt(n-1)$
Times (Periods(B))	= T(P(B))	$2p(t-1)$	Residual	$2p(t-1),\ 4pt(n-1)$
T(Periods(Before))[A]	= T(P(Bef))	$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
T(Periods(After))[A]	= T(P(Aft))	$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
T(Periods(Bef))Control[B]	= T(P(Bef))C	$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
T(Periods(Aft))Control[B]	= T(P(Aft))C	$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
T(Periods(Bef))Impact[B]	= T(P(Bef))I	$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
T(Periods(Aft))Impact[B]	= T(P(Aft))I	$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
L × T(P(B))		$2p(t-1)$	Residual	$2p(t-1),\ 4pt(n-1)$
L × T(P(Bef))[A]		$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
L × T(P(Aft))[A]		$p(t-1)$	Residual	$p(t-1),\ 4pt(n-1)$
Residual		$4pt(n-1)$		
Total		$4ptn-1$		

[A, B] Repartitioned sources of variation; the latter are repartitioned from either the Control (C) or Impact (I) Location only.

kinds of impacts are likely, so procedures like the ones described here are always preferable to the traditional designs.

Minimal Provision of Spatial Controls

The above analyses demonstrate the theoretical utility of arranging temporal sampling in hierarchical series, rather than treating monitoring as a series of equally spaced samples. The designs of sampling shown are not, however, sufficient to demonstrate that any pattern of difference at any temporal scale is related to the putative effect of human activity (as discussed earlier). At the very least, there should be some attempt to control for other sorts of changes that affect more than the single location being sampled and other changes that might occur in one location regardless of the activities of people. The former problem is considered here.

As used elsewhere, a single undeveloped and otherwise unchanged habitat could be sampled to serve as a control to contrast against the location that will potentially be affected by human disturbance. Each location should be sampled in the same pattern of longer and shorter time intervals as used earlier for a single location (Fig. 2). There is actually no need in these designs to take the samples at the same times for the two locations—Times and Periods of sampling are randomized and nested sources of variation in these analyses and could therefore be nested separately for each location. This has consequences for the power of the analyses and would require careful thought before changing the design to this alternative. The consequences are not considered here because of lack of space and because such nested sampling cannot deal with populations that show significant interactions between locations and time or period of sampling before the planned development begins.

The hypothetical sampling design described here is considered to involve collecting data at the same (fully orthogonal) times of sampling in two locations, one of which will serve as an undisturbed control. Both locations will have fluctuations in densities of the sampled organisms that are unrelated to the planned development. Only the noncontrol site will also have any effect of the development (if there is one). Each location could be sampled, as before, for two longer Periods before and again after the development starts. Again, consider that, within each Period, data are taken at three times (i.e. Times nested in each Period) and that five replicate samples are taken at each Time in each of the two locations.

Under such a sampling regime, the data could be analysed as in Table 4. If the temporal trajectories of the mean abundances of the monitored populations in the two locations are consistent before the onset of the putative impact, there will be no statistical interaction between locations and time nor between locations and period (Table 4; L × T(P(Bef)) and L × P(Bef) will not be significant sources of variation). Under these circumstances, any environmental impacts would be detected because they must cause interaction between locations and times or periods (depending on the nature of the impact) after the onset of the disturbance. This would make their detection very straightforward. Pulse impacts would be detected (in a similar manner to that described in Table 3, condition *b*) by significant *F*-ratios as summarized in Table 5. Press disturbances causing short- or long-term fluctuations larger than previously observed for the populations would be detected by the patterns of significant *F*-ratios described earlier (Table 3) and summarized in Table 5.

If, in contrast, the populations in the two locations do not fluctuate synchronously before any human disturbance, there will be significant interactions between locations and either times or periods of sampling before any putative impact occurs (in Table 4, L × T(P(Bef)) or L × P(Bef), or both, will be significant). Under these circumstances, the tests are not as straightforward as before. There may not be significant patterns of interaction after the onset of the putative disturbance (if it causes descreased statistical interaction between the two locations and the various temporal scales of sampling). Reliable tests involve additional two-tailed *F*-ratios to examine the magnitude of interaction before and after the onset of the potential disturbance. These are also summarized in Table 5. No modelled data have been

Table 5. Patterns of significance in analyses of various conditions of environmental impact

The design and abbreviations are as in Table 4. Conditions are: *b*, pulse; *c*, press causing short-term fluctuations; *d*, press causing long-term fluctuations (see Fig. 2 and Tables 2 and 3 for details). Noninteractive and interactive sets of patterns refer to situations in which there is no (noninteractive), or is a (interactive), significant interaction between the differences between the control and potentially impacted locations and times and/or periods of sampling (see text for further details). Asterisks indicate a significant *F* ratio, Xs indicate a non-significant *F* ratio and question marks indicate that the *F* ratio could be either, depending on the significance of other sources of variation in the data

F ratio	Noninteractive			Interactive			Comments
Condition:	*b*	*c*	*d*	*b*	*c*	*d*	
P(Bef)	?	?	?	?	?	?	
P(Aft)	?	?	*	?	?	*	Significant only when long-term fluctuations occur (see Table 3)
P(Aft)/P(Bef)[A]	X	X	*	X	X	*	If powerful enough, this test should detect impacts for condition *d* even when there are L×P(B) interactions
L×P(Bef)	X	X	X	*	*	*	
L×P(Aft)	X	X	*	*	*	*	Where L×P(B) is significant, this test will also be significant and thus cannot detect impact
L×P(Aft)/L×P(Bef)[A]	X	X	*	X	X	*	As above, only detects impact in noninteractive situations
L×T(Bef)	X	X	X	*	*	*	
L×T(Aft)	*	*	X	*	*	?	Where L×T(P(B)) is significant, this test cannot detect impact
L×T(Aft)/L×T(Bef)[A]	*	*	X	?	?	X	
		For Impact location only					
T(P(Bef))	?	?	?	?	?	?	May be significant if data are very variable from time to time, even where there is no impact
T(P(Aft))	*	*	?	*	*	?	As above
T(P(Aft))/T(P(Bef))[A]	*	*	X	*	*	X	If powerful enough, this test should detect impacts for conditions *b* and *c* even when there are L×T(P(B)) interactions
T(P1 After)/T(P2 After)[A,B]	*	X	X	*	X	X	If powerful enough, this test should detect impacts for condition *b* even when there are L×T(P(B)) interactions
T(P1 After)/T(P1 Before)[A,B]	*	*	X	*	*	X	If powerful enough, this test should detect impacts for conditions *b* and *c* even when there are L×T(P(B)) interactions
T(P2 After)/T(P2 Before)[A,B]	X	*	X	X	*	X	If powerful enough, this test should detect impacts for condition *c* even when there are L×T(P(B)) interactions

[A] Two-tailed *F* ratio. [B] Not shown in Table 4.

provided to demonstrate the effectiveness of these tests, but they are similar to those already illustrated in Table 3. Different types and magnitudes of environmental disturbances that do not cause any change in long-run averages in the affected location will be detectable, provided only that the data are sufficiently replicated at the various time-scales to provide powerful tests.

The Problem of Adequate Spatial Replication

As indicated earlier, sampling designs that use only a single control site to contrast against a single potentially impacted site are not sufficient. Any interpretation about the role of human disturbance is completely confounded with any natural, inherent cause of variability between the two locations in the time-course of mean abundances of the organisms. This provides a perfect let-out clause for developers or governments who wish to avoid responsibility for damage to natural environments. It also obfuscates issues of the nature and type of disturbance that actually cause stresses to biological systems (Underwood 1989).

Such sampling also serves to undermine scientific procedures. Almost any randomly chosen pair of locations will eventually reveal some pattern of difference in mean abundance of an organism through time without there being any novel, human-induced disturbance in the environment. As a result, current environmental sampling tends to be antidevelopment for unsound scientific reasons. All such differences will become identified as environmental impacts and attributed to the human activity. Legions of examples abound. For instance, in Botany Bay, New South Wales, it is often claimed by amateur environmentalists that all sorts of changes have occurred in mangrove swamps since an oil spill during the 1970s. Detailed replicated sampling and experimental tests have demonstrated few, if any, identifiable effects of the spill on the marine fauna (McGuinness 1988, 1990), although there are long-term consequences to trees.

The need for scientific, objective and quantified monitoring has not been challenged. The need for rational interpretations of the data is, however, very great and not possible without removing the confounding inherent in routine sampling of only two locations.

As previously argued (Underwood 1989), although there is no prospect (and no desirability) to have more than one impacted location, there is no reason not to have several randomly chosen, representative locations to act as controls. Sampling designs to detect environmental impacts in such sets of locations have been developed and will be described elsewhere (Underwood, unpublished data).

Conclusions

It is not difficult to imagine ecological processes that could be affected by environmental impacts resulting in the types of changes discussed here. Note that, in all cases considered, there was no actual difference between the mean numbers in the population before and after the impact occurred (Fig. 2). All the differences were manifested in short- or long-term changes in the rates of fluctuations around the mean. For example, building a jetty may cause long-term alterations to the patterns of recruitment of larvae to some population of fish. As a result, instead of there being fairly steady trickles of larvae into the monitored population, as was the case before the impact occurred, there are now periods of excessive larval input (causing a long-term increase in abundance until eventually this cohort dies out) and other periods of decline in abundance because the population is not receiving any larvae at all. Such disruption of larval supply would cause the sort of oscillations seen in Fig. 2*d*. In contrast, some developments may cause intermittent periods of increase in abundance of a monitored population. For example, material leaking from a marina may be nutrient-rich for several seasons. Then, a period occurs when the material released is toxic and the organisms are reduced in number. This is followed by restoration of the pattern of nutrient enrichment. This could occur, for example, where a development causes the release of sewage, which occasionally (say, after rainfall causing urban runoff) contains pesticides, heavy metals and other materials. This would cause the pattern of fluctuation illustrated in Fig. 2*c*. Other examples are easy to construct but are not possible so far to identify in the literature because this sort of sampling has not been done. Wherever some disturbance causes increased rates of change in the monitored population, there is an increased probability that the population will go extinct when any other stress arrives. This occurs because the mean abundance of the population, after the disturbance, wanders considerably closer to zero and is sometimes at very small numbers compared with the situation before the impact. Small populations of usually abundant organisms have very increased chances of extinction (Simberloff 1986).

Thus, the timing of sampling must be chosen very carefully with respect to the known biology of the organisms being monitored, the natural rates of change and temporal variances of the population before the putative impact begins and the scale and likely consequences of the sorts of impact that might be anticipated from a particular development. To make decisions about the design of sampling programmes so that they have appropriate time-scales is an urgent task for biologists. It is, however, a biological task—we should already know or be measuring natural rates of change in organisms as part of any other studies. It is surprising how little information there is about temporal and spatial variances in natural populations despite the many years of investigation (and the millions of dollars spent on sampling programmes). Where such data are available, the sorts of impacts that might matter to populations can be assessed before any development (or accident). Also, the magnitude of change in rates that might matter to the population (where 'matter' must be defined for each case; see Underwood and Peterson 1988) could also be determined in advance of any sampling. This would enable sampling frequencies, intensities and replication to be sufficient to provide powerful tests for such magnitudes of impact.

Determining an appropriate frequency of sampling is extremely important if pulse phenomena that do not last long are to be detected. Short-term transient fluctuations (as in Fig. 2*b*) cannot be detected unless samples happen to be taken during the period of oscillation. These are probably the least important environmental changes to detect because populations recover and there are no long-term consequences for the monitored species. There is, nevertheless, a great difference between human disturbance that causes no environmental change and one that causes short-term fluctuations. Such fluctuations may themselves cause much more extensive changes in other species that are not being monitored. Also, for some types of disturbance that are repeated in time and space (e.g. building marinas, jetties and outfall pipes, etc.) it will be important to determine whether there are rapid pulse responses by populations. If so, remedial action should be taken in subsequent developments to ensure that the pulse does not cause a local, but potentially long-term, extinction because the transient disturbance happens to be sufficiently larger than detected on previous occasions.

Throughout this examination of sampling designs, the mean abundance of a single species has been the focus of attention. There is no reason for other variables to be considered any differently (Underwood 1989). The designs used here can be used equally well for sizes, reproductive rates, physiological states—the whole gamut of types and scales of measures used in attempts to detect environmental disturbances (Underwood and Peterson 1988). There will, however, be differences in the degree to which different variables may conform to the underlying assumptions of analyses of variance (Winer 1971; Underwood 1981). There will also be potential problems because different variables have different frequency distributions and variances. Thus, allocation of replication at different levels of sampling will not be the same for different variables. The power of tests (i.e. their capacity to detect human disturbances) will be intrinsically different for different variables. Care will therefore be needed to choose appropriate variables in any specific case and to use suitably powerful designs where more than one variable is being monitored.

Also, here, to avoid excessive complexity in the introduction of the need for hierarchical temporal sampling, only two periods were suggested before and two after the potential impact. Of course, in practice, many more periods may be necessary or desirable. The designs illustrated above can deal with more periods of sampling and more contrasts. Extensions would simply be a matter of common sense applied to the appropriate tests for each possible type of impact. Obviously, the more data the better, in the sense that more varied and smaller impacts are likely to be detected with more data (a serious problem for much environmental analysis—see Bernstein *et al.* 1984). These designs could also, by obvious extension, be used where there are more than two relevant different time-scales. Only two were considered here, but any number of nested time-scales would be managed using identical logic to that presented here. Where there is little or no information about the sorts of natural rates of change in numbers in a population (or whatever variable has been chosen), a series of nested time-scales is probably essential during the phase of monitoring before a proposed potential impact. Extensions of nested designs to cover more levels of nesting are straightforward (Snedecor and Cochran 1967; Winer 1971; Underwood 1981). Note the earlier comments about optimal allocation of sample sizes to the various levels of nesting.

Supposing there are different types of environmental disturbance? What does it matter? First, it matters to the long-term conservation and diversity of species that any changes that might increase chances of extinction be identified early enough to do something about them. Sampling that can only detect long-run changes in means are not adequate. Secondly, when such disturbances have been identified, does it matter to know what sort of effects are caused? The answer to this is obvious. Any attempt to provide remedial action or restoration or reconstruction of habitat and the associated populations must be done on the basis of what is actually happening. If management of wastes, for example, is to be effective and wastes cause intermittent pulse responses in sampled populations, their management should focus on removing the extreme circumstances or time-courses that create the pulses. Alternatively, if the disposal of wastes causes chronic press phenomena, then alternative strategies to create pulses might well be better. Options for such practical problems are more contrastable if the nature of the effects that each may or does have could be properly assessed.

Alternatively, if some environmental disturbance causes increased temporal fluctuations of a long-term nature (as in Fig. 2*d*), management should depend on preventing further problems during periods when abundances are small. This is, of course, the basis of theoretical management of fisheries (although the laws of economic supply and demand will help because when resources get very scarce, they may not be worth harvesting and therefore a source of human disturbance is removed). This type of management would be more difficult when populations are fluctuating very rapidly (management cannot keep up).

Finally, the main reason for contemplating different time-scales in sampling during environmental monitoring programmes is that more thought will then be given to the whole problem of replication in time, which is conspicuously lacking in many ecological investigations. The assumption that a change from one season to another is actually seasonal cannot be tested without replicated times of sampling within each season. This simple point seems to have been lost in the routine aspects of monitoring programmes that have never had their time-courses justified by replicated sampling at smaller, nested scales to demonstrate the appropriateness of the time-scales chosen. More work needs to be done on these designs to investigate the appropriateness and robustness of the various *F*-ratios suggested and to determine their power under different field conditions of natural change. The fact that obvious environmental perturbations could be demonstrated using these procedures that could not have been discovered by routinely used designs does, however, demonstrate that a lot more effort is needed before complacency about environmental monitoring is allowed to become further entrenched.

Acknowledgments

The preparation of this paper was supported by funds from the Institute of Marine Ecology of the University of Sydney and from the Australian Research Council and was made possible, whilst on holiday, by the patience of my family and the provision of repeated sustenance by Mr D. B. and Dr K. E. McFadyen. Very useful discussions were held with several colleagues, notably Ms M. G. Chapman, Dr P. G. Fairweather and Mr M. Lincoln-Smith. Two anonymous referees worked hard to improve the manuscript, for which I thank them.

References

Andrew, N. L., and Mapstone, B. D. (1987). Sampling and the description of spatial pattern in marine ecology. *Annual Review of Oceanography and Marine Biology* **25**, 39–90.

Bachelet, G. (1986). Recruitment and year-to-year variability in a population of *Macoma balthica*. In 'Long-term Changes in Coastal Benthic Communities'. (Eds C. Heip, B. F. Keegan and J. R. Lewis.) pp. 233–58. (Junk: Dordrecht.)

Baker, J. M., Campodonico, I., Guzman, L., Texera, J. J., Texera, B., Venegas, C., and Sanhueza, A. (1976). An oil spill in the Straits of Magellan. In 'Marine Ecology and Oil Pollution'. (Ed. J. M. Baker.) pp. 441–72. (Applied Science Publishers: Barking.)

Bender, E. A., Case, T. J., and Gilpin, M. E. (1984). Perturbation experiments in community ecology: theory and practice. *Ecology* **65**, 1–13.

Bernstein, B. B., and Zalinski, J. (1983). An optimum sampling design and power tests for environmental biologists. *Journal of Environmental Management* **16**, 35–43.

Bernstein, B. B., Smith, R. W., and Thompson, B. E. (1984). 'Sampling Design and Replication for Benthic Monitoring.' (Southern Californian Coastal Water Research Report: Long Beach.)

Buchanan, J. B., and Moore, J. J. (1986). Long-term studies at a benthic station off the coast of Northumberland. In 'Long-term Changes in Coastal Benthic Communities'. (Eds C. Heip, B. F. Keegan and J. R. Lewis.) pp. 121–8. (Junk: Dordrecht.)

Cohen, J. (1977). 'Statistical Power Analysis for the Behavioural Sciences.' (Academic Press: New York.)

Connell, J. H. (1978). Diversity in tropical rain forests and coral reefs. *Science (Washington, DC)* **199**, 1302–10.

CSIRO (1989). 'Jervis Bay Baseline Studies, Second Progress Report. May 1989.' (CSIRO Division of Fisheries: Hobart.)

Dauvin, J. C., and Ibanez, F. (1986). Variations á long-terme (1977–1985) du peuplement des sables fins de la Pierre Noire (baie de Morlaix, Manche Occidentale): analyse statistique de l'évolution structurale. In 'Long-term Changes in Coastal Benthic Communities'. (Eds C. Heip, B. F. Keegan and J. R. Lewis.) pp. 171–86. (Junk: Dordrecht.)

Dixon, W. J., and Massey, F. J. (1969). 'Introduction to Statistical Analysis.' (McGraw-Hill Kogakusha: Tokyo.)

Green, R. H. (1979). 'Sampling Design and Statistical Methods for Environmental Biologists.' (Wiley: Chichester.)

Hurlbert, S. J. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**, 187–211.

Kennelly, S. J., and Underwood, A. J. (1984). Underwater microscopic sampling of a sublittoral kelp community. *Journal of Experimental Marine Biology and Ecology* **76**, 67–78.

Kennelly, S. J., and Underwood, A. J. (1985). Sampling of small invertebrates on natural hard substrata in a sublittoral kelp forest. *Journal of Experimental Marine Biology and Ecology* **89**, 55–67.

Kinhill Engineers Pty Ltd (1990). 'Proposed Third Runway, Sydney (Kingsford Smith) Airport: Draft Environmental Impact Statement.' (Federal Airports Corporation: Mascot, N.S.W.)

Lopez-Jamar, E., Gonzalez, G., and Mejuto, J. (1986). Temporal changes of community structure and biomass in two subtidal macroinfaunal assemblages in La Coruna Bay, NW Spain. In 'Long-term Changes in Coastal Benthic Communities'. (Eds C. Heip, B. F. Keegan and J. R. Lewis.) pp. 137–50. (Junk: Dordrecht.)

Lundalv, T., Larsson, C. S., and Axelsson, L. (1986). Long-term trends in algal-dominated rocky subtidal communities on the Swedish west coast—a transitional system? In 'Long-term Changes in Coastal Benthic Communities'. (Eds C. Heip, B. F. Keegan and J. R. Lewis.) pp. 81–96. (Junk: Dordrecht.)

McGuinness, K. A. (1984). Equations and explanations in the study of species-area curves. *Biological Reviews of the Cambridge Philosophical Society* **59**, 423–40.

McGuinness, K. A. (1988). 'The Ecology of Botany Bay and the Effects of Man's Activites: A Critical Synthesis.' (Institute of Marine Ecology, University of Sydney: Sydney.)

McGuinness, K. A. (1990). Effects of oil spills on macro-invertebrates of salt marshes and mangrove forests in Botany Bay, New South Wales, Australia. *Journal of Experimental Marine Biology and Ecology* **142**, 121–36.

Peterman, R. M. (1990). Statistical power analysis can improve fisheries research and management. *Canadian Journal of Fisheries and Aquatic Sciences* **47**, 2–15.

Scheffe, H. (1959). 'The Analysis of Variance.' (Wiley: New York.)

Simberloff, D. (1986). The proximate causes of extinction. In 'Patterns and Processes in the History of Life. Dahlem Konferenzen'. (Eds D. M. Raup and D. Jablonski.) pp. 259–76. (Springer-Verlag: Berlin.)

Snedecor, G. W., and Cochran, W. G. (1967). 'Statistical Methods.' 6th edn. (University of Iowa Press: Ames.)

Sousa, W. (1979*a*). Disturbance in marine intertidal boulder fields: the nonequilibrium maintenance of species diversity. *Ecology* **60**, 1225–39.

Sousa, W. P. (1979*b*). Experimental investigations of disturbance and ecological succession in a rocky intertidal algal community. *Ecological Monographs* **49**, 227–54.

Sousa, W. P. (1980). The responses of a community to disturbance—the importance of successional age and species life histories. *Oecologia (Berlin)* **45**, 72–81.

Stewart-Oaten, A., Murdoch, W. M., and Parker, K. R. (1986). Environmental impact assessment: 'pseudoreplication' in time? *Ecology* **67**, 929–40.

Underwood, A. J. (1978). An experimental evaluation of competition between three species of intertidal prosobranch gastropods. *Oecologia (Berlin)* **33**, 185–208.

Underwood, A. J. (1981). Techniques of analysis of variance in experimental marine biology and ecology. *Annual Review of Oceanography and Marine Biology* **19**, 513–605.

Underwood, A. J. (1984). Vertical and seasonal patterns in competition for microalgae between intertidal gastropods. *Oecologia (Berlin)* **64**, 211–22.

Underwood, A. J. (1986). The analysis of competition by field experiments. In 'Community Ecology: Pattern and Process'. (Eds J. Kikkawa and D. J. Anderson.) pp. 240–68. (Blackwells: Melbourne.)

Underwood, A. J. (1989). The analysis of stress in natural populations. *Biological Journal of the Linnean Society* **37**, 51–78.

Underwood, A. J. (1990). Experiments in ecology and management: their logics, functions and interpretations. *Australian Journal of Ecology* **15**, 365–89.

Underwood, A. J. (1991*a*). 'Biological monitoring for human impact: how little it can achieve.' In 'Proceedings of the 29th Congress of the Australian Society of Limnology, Jabiru, NT, 1990'. (Ed. R. V. Hyne.) pp. 105–23. (Office of the Supervising Scientist Alligator Rivers Region, Australian Government Publishing Service: Canberra.)

Underwood, A. J. (1991*b*). Competition in marine plant/animal interactions. In 'Plant/Animal Interactions in the Marine Benthos'. (Eds S. J. Hawkins, G. A. Williams, D. M. John and J. H. Price.) (Oxford University Press: Oxford.) (in press).

Underwood, A. J., and Denley, E. J. (1984). Paradigms, explanations and generalizations in models for the structure of intertidal communities on rocky shores. In 'Ecological Communities: Conceptual Issues and the Evidence'. (Eds D. R. Strong, D. Simberloff, L. G. Abele and A. Thistle.) pp. 151–80. (Princeton University Press: Princeton.)

Underwood, A. J., and Fairweather, P. G. (1989). Supply-side ecology and benthic marine assemblages. *Trends in Ecology and Evolution* **4**, 16–20.

Underwood, A. J., and Peterson, C. H. (1988). Towards an ecological framework for investigating pollution. *Marine Ecology Progress Series* **46**, 227–34.

Underwood, A. J., and Verstegen, P. H. (1988). Experiments on the association between the intertidal amphipod *Hyale media* Dana and the limpet *Cellana tramoserica* (Sowerby). *Journal of Experimental Marine Biology and Ecology* **119**, 83–98.

Winer, B. J. (1971). 'Statistical Principles in Experimental Design.' 2nd edn. (McGraw-Hill Kogakusha: Tokyo.)

Manuscript received 5 February 1991, revised and accepted 6 May 1991