

JEMBE 01810

## Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world

A. J. Underwood

*Institute of Marine Ecology, University of Sydney, New South Wales, Australia*

(Received 9 August 1991; revision received 3 April 1992; accepted 10 April 1992)

**Abstract:** BACI (Before/After and Control/Impact) sampling is widely used in investigations of environmental impacts on mean abundance of a population. The principle is that an anthropogenic disturbance in the “impact” location will cause a different pattern of change from before to after it starts compared with natural change in the control location. This can be detectable efficiently as a statistical interaction in an analysis of variance of the data. Usually, samples are taken at replicated, random intervals of time before and after the putative impact starts; this ensures that chance temporal fluctuations in either location do not confound the detection of an impact. These designs are, however, insufficient because *any* location-specific temporal difference that occurs between the two locations will be interpreted as an impact even if it has nothing to do with the human disturbance. Alternatively, abundance in the single control location may change in the same direction, cancelling the effects of an impact. Here, asymmetrical designs are developed that compare the temporal change in a potentially impacted location with those in a randomly-selected set of control locations. An impact must cause a different temporal change in the disturbed location from what would be expected in similar locations. This can be detected for short-term (pulse) or long-term (press) impacts by different patterns of significance in the temporal interactions between times of sampling and locations. From these novel designs, tests are derived that demonstrate whether an unusual pattern of temporal change in abundance of organisms is specific to the supposedly impacted location and correlated with the onset of the disturbance. Examples are presented of how to use these designs to detect impacts at different spatial scales. Other aspects of their use are discussed.

**Key words:** Asymmetrical design; BACI sampling; Environmental impact

### INTRODUCTION

A large general class of problems for ecologists to solve is the development of rapid, reliable and cost-effective procedures for detecting human effects on natural populations. There have been several discussions of appropriate experimental designs for this activity in some environments (Green, 1979). In a previous review (Underwood, 1991a), a sequence of possible sampling designs that has been recommended in the past was examined, exposing their flaws. As a result, a new, albeit more complicated, design was proposed to cover problems of inadequate sampling to detect impacts that affect temporal variances in populations. The need for adequate spatial replication was also

Correspondence address: A. J. Underwood, Institute of Marine Ecology, Zoology Building A 08, University of Sydney, New South Wales 2006, Australia.

stressed, but the problems of detecting impacts in a spatially heterogeneous habitat were not solved.

The major problem to be overcome in assessing environmental impact is that there is usually only one (therefore unreplicated) potentially impacted site (e.g. Stewart-Oaten et al., 1986). It is, of course, quite unreasonable to demand that every time a sewage outfall, a nuclear power plant, a marina, etc., is built, that there should be two or more of them. Presumably as a result of there being only one putatively impacted site, it appears to have become widespread that it is appropriate to only have one "control" site to contrast against the potentially impacted one. This is a false way to proceed (Underwood, 1989, 1991a). Obviously, two arbitrarily chosen sites may very well differ in their changes through time, regardless of whether there has been some actual human influence (or impact) in only one site. Alternatively, a real impact (e.g. causing a decrease in the numbers of a population in one site) may not be detected because a similar decrease happens to occur by chance in the single "control" site.

Here, the problems of spatial replication are considered, using the much more realistic device of having several, randomly-chosen sites to serve as "controls". New designs are proposed, using several spatially replicated sampling sites. Two variations of such designs are suggested, depending on the scale of spatial effects that might result from a proposed development or impact.

The other major issue addressed is the need for replicated sampling in time, preferably before the putative impact occurs (e.g. before a proposed development takes place) and after it has happened. This approach has been suggested, for example, by Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986) in their Before/After, Control/Impact (or BACI) design. They have, however, ignored the fact that there also needs to be spatial replication. An appropriate combination of replicated sampling in time and replicated sampling at appropriate spatial scales is absolutely mandatory before any attempt to determine potential impact is likely to succeed.

Practical experience suggests that most natural populations show fluctuations from time to time that are not parallel from place to place. As a result, there is considerable interaction between space and time in the data from any sampling design. It is therefore necessary to have methods for estimating potential change in the magnitude of such variations, in addition to measuring any potential or actual changes in the mean numbers of the target species. As demonstrated earlier (Underwood, 1991a), an impact may, in fact, be on the size of variance in time rather than on the absolute abundance of some species. Here, designs are proposed to detect impacts that affect spatial differences, temporal variances, or their interactions. Designs to determine impacts that have no effect on long-run mean abundances were described elsewhere (Underwood, 1991a).

Other aspects of the design of monitoring programmes are not considered here. There are serious difficulties with the implementation of monitoring as the sole, or even as the major tool for detection of impacts. Some of the considerations about choice of species have been discussed elsewhere (Underwood & Peterson, 1988). Other prob-

lems with the design or interpretation of monitoring programmes are considered in Underwood (1989, 1991b). What follows is an attempt to demonstrate the usefulness of a novel design for sampling that involves several spatial scales and that will determine the effect of some human disturbance for which it is known when a possible impact will start. This is the basis for many assessments of environmental impact (although there are exceptions where impacts must be assessed without any data being available before a disturbance occurs; Green, 1979).

Finally, examples of the sorts of data that are routinely available on natural populations are analysed. These indicate some of the problems associated with the determination of impact, from the point of view of the design of a sampling study. The procedures used are those of traditional experimental design. The general issues have been well summarised in ecological papers (see particularly Andrew & Mapstone (1987), Clarke & Green (1988), Eberhardt & Thomas (1991), Green (1979), Hurlbert (1984), Underwood (1981, 1986, 1988)). Throughout, univariate analyses of abundance of a single species are used as the indicator of impact. Undoubtedly, multivariate analogues could be developed for some aspects of the designs proposed here. The purpose of these developments is not only to demonstrate a more rational procedure for detection of impacts, if they occur, but also to be more sure about attribution of causality. Thus, these designs improve the ability of environmentalists to demonstrate that a change in a population in one place is associated with a particular human activity. Such demonstrations are not usually the case in this field. The hope is that such scientific underpinning for or against the relationship between a human activity and changed numbers of a species will take environmental decision-making out of the random processes known as legal procedures.

## THE PROBLEMS

### SINGULARITY OF THE IMPACTED SITE

The major problem confronting any attempt to determine potential effects of some human activity on a target population is that there is usually only one impacted site. Thus, there is one nuclear power plant, one marina, one housing development, one sewage outfall (although there are, occasionally and fortuitously, exceptions to this). Where there are exceptions, it should be possible to improve the designs discussed here to take into account, wherever practicable, orthogonal contrasts of control and impacted sites. Under those circumstances, the task of identifying environmental impacts will be considerably easier.

For reasons that are completely illogical, it has become routine to consider contrasts between the potentially impacted site and those of a similar, hopefully very similar, site chosen some distance away to represent a piece of the world which could not possibly be impacted. This leads directly to the notion of contrasting one potentially impacted and one control site with all of the concomitant problems of spatial confounding (more recently termed "pseudoreplication", Hurlbert, 1984).

## VARIANCE IN SPACE

Most natural populations oscillate in ways that are not concordant from one place to another. Thus, abundances of most species will fluctuate from time to time independently in any two sites (one potentially impacted and one control). Any attempt to determine that a difference from one site to another is due to the potential environmental impact when it starts is doomed. All that a study of such a system could do is to demonstrate that there are differences in temporal patterns between the two sites and not that the difference is due to the potential impact (see particularly Hurlbert, 1984).

Populations are known to vary from place to place for all sorts of reasons and it is normal for the mean number of some organism to be quite different in one part of the habitat from the mean number in another. It is therefore impossible to identify the cause of the difference unless potential causes due to one site being geographically, historically and in many other ways different from the other can be falsified. This is the basis of previous reviews of spatial confounding called "pseudoreplication" by Hurlbert (1984) and discussed by Andrew & Mapstone (1987) and Underwood (1981, 1986).

## VARIANCE IN TIME

It is also generally the case that the numbers of organisms in populations vary from time to time. There is no reason at all to suppose that the population of some species is going to stay constant, unless there are no, or are only absolutely compensating, processes of birth or death or immigration or emigration (e.g. Krebs, 1978). As a result, the arrival of some development or potential impact is not always going to be the cause of the next change in numbers. They may be changing because of some other process operating coincidentally from the time of the start of the impact. This point was well made by Stewart-Oaten et al. (1986). As shown below, they suggested that this problem could be overcome by replicated sampling at random intervals before the potential impact starts and then again after the impact. As will be discussed later, this is not, on its own, a satisfactory solution to the problem.

## INTERACTIONS BETWEEN SPACE AND TIME

The worst feature of natural biological populations is that the time courses of abundances of a population are rarely the same from one place to another. There has been considerable discussion about this in the literature. Nevertheless, in practice, most studies have demonstrated considerable change from time to time and difference from place to place and have demonstrated differences in the temporal changes from one place to another. The problem for assessment of environmental impact is that there do not seem to be widely used, or widely known, procedures for analysis of differences in populations when their natural numbers interact in complex ways, at least none that is widely cited in the literature on environmental impact. A particularly powerful analytical tool to identify such interactive processes is the collection of procedures known

as analysis of variance (Underwood, 1981); these will form the basis for the designs outlined here.

### APPARENT SOLUTIONS TO THE PROBLEM

#### GENERAL CONSIDERATIONS

Green (1979), in his excellent book, proposed a sampling design that would handle the arrival of some specifically human imposed perturbation to a natural population. He proposed that the habitat to be impacted or developed should be sampled once before and once after the proposal was implemented. In addition, he proposed that a control site, of suitable size and area should also be sampled once before and once after the development proceeded. He also suggested that this design could be extended to include more than one time or place. As has been pointed out before (Hurlbert, 1984; Bernstein & Zalinski, 1983; Stewart-Oaten et al., 1986; Underwood, 1991a), this design has serious flaws (illustrated in Fig. 1A). First, there may be many differences between the putatively impacted site and the control site which have nothing to do with the development. More importantly, there may be differences in the time courses of the populations in the two sites that also have nothing to do with the development (Fig. 1A). Therefore, it is not possible using the simplest sampling design advocated by Green (1979) to do other than to demonstrate that there is a difference between the two sites

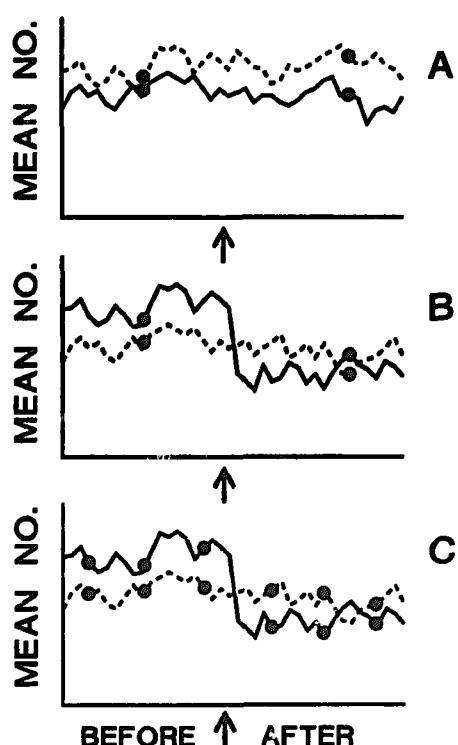


Fig. 1. Typical sampling for assessment of environmental impact on abundance of an organism. Solid line is the putatively impacted location; dashed line is a single control location. Circles represent times of sampling; the arrow indicates the beginning of the impact. (A, B) BACI design with a single sample before and after the putative impact. In A, if variability of abundance were small, a difference between the control and impacted locations would be detected, even though no impact occurred. In B, a real impact would not be detected because of chance variations in mean abundance. C is the BACI design of Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986) with replicated times of sampling in each location.

(if one exists), or that a difference has occurred between the time before impact began and the time after (which might, perhaps, be some function of the impact). At best, this design can demonstrate an interaction between the difference between control and impacted sites before and after the impact began. The presence of the interaction would be an indication that the difference between the control and the potentially impacted sites before the development is not the same as the difference between these two sites afterwards. As a result, the interpretation proposed by Green (1979) is that this interaction is brought about by the arrival of an actual impact into the impacted site.

This interpretation is completely confounded, although Green's (1979) introduction of rigour and logic into this aspect of environmental science is still a great advance on previous efforts. All that can be deduced is that there is indeed a difference between the two sites after the impact and that difference is not the same as the difference before the impact. This cannot possibly be singularly interpreted to mean that the impact caused the difference (Hurlbert, 1984; Bernstein & Zalinski, 1983; Stewart-Oaten et al., 1986; Underwood, 1991a).

Further, if there is a response to human disturbance in one area, it might be masked by a simultaneous fluctuation in the opposite direction in the single control site. In this case, the impact would never be detected (Fig. 1B). Thus, differences or an interaction between the two sites are not necessarily indicative of an impact (although such evidence is widely used to suggest that there is one). Lack of differences or interaction between the two sites, sampled at one time, is not evidence of the lack of an impact.

#### BACI DESIGN

Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986) discussed, at length, the deficiencies of this design, coupled with a critique of the problems already noted by Hurlbert (1984). As a result, they proposed a design, now termed the BACI-design. In this procedure, the control (C) and potentially impacted (I) sites are sampled before (B) and after (A) the development occurs. This before-after-control-impact (BACI) design therefore covers Green's (1979) proposed design. The new twist is that Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986) proposed that there should be several, i.e. temporally replicated, independent assessments of the abundance of the population before the development starts and again afterwards (Fig. 1C). They advocated taking samples in the two locations at replicated times, but at the same time in each location. Their test procedure consists of calculating the difference between the control and the potentially impacted site at each time of sampling and then completing a *t*-test (or alternative) on the set to replicate such differences before versus after the potential impact. Other analytic procedures may be equally, or more, useful, for example a three-factor analysis of variance (discussed in detail in Underwood, 1991a). This design has, undoubtedly, resolved one of the problems of that suggested by Green (1979). There is a proper temporal resolution that, in theory, allows interpretation of the differences from before to after as being something more sustained than simple random noise in time between the two sites (Fig. 1, C compared to A).

## PROBLEMS WITH THE BACI DESIGN

This has not, however, resolved any of the real difficulties. There still remains the problem that the two different sites may have different time-courses in the numbers of the target species, without regard to the arrival of the potential impact itself. Thus, it is quite reasonable to anticipate an interaction in this sort of analysis but not to be sure that it is due to a human impact. For example, there may be some general trend through time that causes the two sites to diverge in the abundance of the species of interest, in such a manner that would have occurred whether or not the development had ever taken place. This was, however, discussed by Stewart-Oaten et al. (1986, Fig. 3b, p. 934). They suggested that when differences between the two sites in the time courses of the populations already existed, the species concerned should not be used for the assessment of environmental impact. This is self-defeating. The majority of species about which appropriate data have been accumulated suggest that such interactions are a normal part of the operation of abundances of most populations. Numerous examples can be found in the literature. Of those studies with proper temporal replication, a very large proportion demonstrate these sorts of patterns. It is for this reason that some (but not enough: Connell, 1983; Underwood & Denley, 1984) experimental studies are multiply replicated in space and repeatedly sampled through time. As a corollary, of course, studies that have not been repeated in time and space cannot identify interactions, even though they probably exist.

Bernstein & Zalinski's (1983) and Stewart-Oaten et al.'s (1986) design does, however, have the advantage that it is possible, provided that sufficient temporal sampling is done, to estimate the temporal variance in the population in each of the sites before and after the proposed impact occurs. This has recently been demonstrated to be useful when the impact is not on the mean abundance but on the temporal variance itself, as demonstrable using the protocols in Underwood (1991a). This will not be discussed further here.

This BACI design, however well intentioned, is not sufficient to demonstrate the existence of an impact that might unambiguously be associated with some human activity thought to cause it.

## A POSSIBLE SOLUTION

### MULTIPLE CONTROLS

The first major improvement is the use of multiple control sites. Ideally, of course, there would be replicated control and replicated impact sites, but this is an ideal only for statistical purposes. It is not ideal, for example, to build nuclear power plants simply in order that there should be more of them in randomly-chosen places! Inevitably, therefore, most environmental impact assessment will be done using only one experimental, or potentially impacted site. There is, however, no reason why there should be only one control site.

There should be a series of sites, randomly chosen out of a set of possible sites that have similar features to those of the one where the development is being proposed. The only constraint on random choice of sites is that the one planned to be impacted must be included in the sample. This is not nearly as difficult as it seems; the sites do not have to be identical, in any sense of the word. They simply have to follow the normal requirements that they come from a population of apparently similar sites. This is the basis of all random sampling theory. Of course, the sites should be independently arranged so that there is no great spatial autocorrelation among them. For most marine habitats, sites can be sufficiently widely spaced that they are not correlated by processes of recruitment or disturbances.

The logic of the design is that an impact in one site should cause the mean abundance of animals there to change more than expected on average in undisturbed sites. Abundances in the control, undisturbed sites will continue to vary in time, independently of one another. On average, however, they can be expected to continue as before. Note the important difference from a BACI design. Here, the individual control sites may differ and may change significantly from one another. On average, however, the set of controls should continue to show average behaviour. Impacts are those disturbances that cause mean abundance in a site to change more than is found on average.

Formally, the hypothesis that there is going to be some impact of the proposed development is a statement that after the development has started, there will be some difference between the potentially impacted site from the average of that in the controls. This hypothesis is in its simplest (and most primitive) form and will be developed later. It is, however, instructive to examine this as opposed to the simple notion that a contrast between one control and a potentially impacted site is all that is required. The analysis is developed starting in Table I. What is required is one planned (a priori) orthogonal contrast (Scheffe, 1959; Winer, 1971). This contrast is between the potentially impacted (identified as I in Table I) and the control locations. Differences among the control locations are not of intrinsic interest, but a real environmental impact requires there to be greater difference between the impacted and control locations than there is generally among the controls.

Examination of Fig. 2A will illustrate the principle of the procedure. First, the abundances of some organism may show any pattern of difference among the locations before a putative impact. These will presumably continue whether or not an impact occurs. Thus, differences among locations are to be expected. Second, as discussed above, there is no reason to presume that the differences among locations are constant through time. Thus, there will be statistical interactions between locations and time of sampling. In this particular design (Fig. 2A), sampling is done twice, once before and once after a putative impact. It is reasonable, therefore, to expect that the differences among locations are not the same before and after. There should be an interaction, identified as  $B \times L$  in Table I.

If, however, there really is some effect on the population in the putatively impacted

TABLE I

Asymmetrical sampling design to detect environmental impact;  $l$  Locations are sampled, each with  $n$  random, independent replicates, once Before and once again After a putative impact starts in one Location ("Impact"); there are  $(l - 1)$  Control Locations; Locations represent a random factor.

Source of variation		Degrees of freedom	
Before vs. After	= B	1	
Among Locations	= L	$(l - 1)$	
Impact vs. Controls <sup>a</sup>	= I	1	
Among Controls <sup>a</sup>	= C	$(l - 2)$	
$B \times L$		$(l - 1)$	
$B \times I^{a,b}$		1	
$B \times C^a$		$(l - 2)$	<i>F</i> -ratio vs. Residual
Residual		$2l(n - 1)$	
Total		$2ln - 1$	

<sup>a</sup> Repartitioned sources of variation.

<sup>b</sup> Impact can be detected by the *F*-ratio Mean Square  $B \times I$ /Mean Square  $B \times C$ . If  $B \times C$  is not significant (the system is non-interactive), impact can be detected by the *F*-ratio Mean Square  $B \times I$ /Mean Square Residual (see text for details).

location (as in Fig. 2A), the difference between that location and the others before the impact occurs should not be of the same magnitude (or, for some locations, the same direction, as in Fig. 2A). In the example illustrated, mean abundance is reduced by the

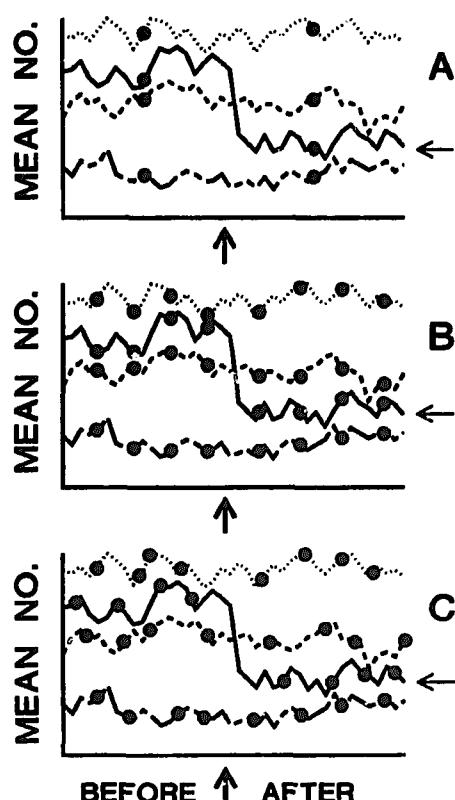


Fig. 2. Sampling for assessment of environmental impact with three control locations and a single impacted location (indicated by the arrow at the right). In A, a single time of sampling before and after the impact begins is used to illustrate the form of the analysis (as in Table I). In B, there are 4 times of sampling before and again after the impact starts (see Table II). In C, the locations are sampled at 4 different times before and again after the impact starts (see Table III).

impact, causing a larger negative difference from that in the location with the greatest abundance, a smaller positive difference from that in the location with the smallest abundance and a smaller (as opposed to larger) abundance than that in the other location (Fig. 2A).

In an analysis of such a situation (Table I), the impact would be detected as an interaction between time of sampling (Before versus After; B in Table I) and the difference between the putatively Impacted and Control locations (I in Table I). Where only fixed contrasts are involved, this source of variation can be readily extractable in asymmetrical analyses of variance (Winer, 1971; Underwood, 1978, 1984, 1986, 1991a; Underwood & Versteegen, 1988) and is identified as  $B \times I$  in Table I.

Here, however, there is a major difference in the nature of the analysis compared with routine asymmetrical analyses. The locations are in the sampling programme in two sets or groups. One is the set of control locations, which are randomly chosen and represent, truly, a random factor in the design. The putatively impacted location is the sole (i.e. unreplicated) member of the other group. Thus, the appropriate model for the analysis is a form of replicated repeated measures design (e.g. Winer, 1971; Hand & Taylor, 1987; Crowder & Hand, 1990), except that one group (the "group" of a single Impact location) is spatially unreplicated (although there are replicates within the location).

It is, of course, still possible that this interaction could be significant because of some differential fluctuations in the abundances of organisms in the different locations that have nothing to do with the putative impact. In this case, there should generally be interactions between the time of sampling (B in Table I) and the differences among the Control locations (C in Table I), regardless of what happens in the putatively impacted Location. This would be detectable as a significant  $B \times C$  interaction in Table I.

This consideration leads to the ecologically realistic proposition that an environmental impact on the abundance of a population in some location should be defined as an anthropogenic perturbation that causes more temporal change in the population than is usually the case in similar populations in other similar locations where no such disturbance occurs. Formally, the hypothesis is that the interactions in time between the putatively impacted and a set of control locations ( $B \times I$  in Table I) should be different from the naturally occurring interactions in time among the Control locations ( $B \times C$  in Table I).

If an anthropogenic disturbance does cause a small impact, such that the magnitude of change in the population is small, an ecologically realistic interpretation is that the fluctuation in the impacted population is within the bounds of what occurs naturally elsewhere. It is within the resilience of natural populations (e.g. Underwood, 1989) and therefore no cause for concern.

Using the general principles for calculating expected values of mean squares in an analysis of variance (e.g. Winer, 1971; Underwood, 1981), the appropriate tests can be constructed. The interaction from before to after the disturbance between the pu-

tatively impacted and the average of the control locations ( $B \times I$  in Table I) is what must be affected by an environmental disturbance. The appropriate test is an *F*-ratio of the Mean Square for  $B \times I$  divided by the Mean Square for  $B \times C$ .

If, however, it can be shown that there is no natural interaction among the Control locations from before to after the disturbance (i.e.  $B \times C$  is itself not significant when tested against the Residual Mean Square), an impact would be detected by the more powerful *F*-ratio of the Mean Square for  $B \times I$  divided by the Residual Mean Square. This latter test involves post-hoc elimination of a component of variation in the  $B \times I$  Mean Square (see Winer, 1971; Underwood, 1981). The details of this are not important here, but there are some considerations. Technically, this post-hoc removal of a component of variance involves making the assumption that the component is zero. This raises the possibility of Type II error (accepting that a value is zero when, in fact, the test is not capable of detecting it). This is the rationale behind post-hoc pooling procedures to reduce the probability of Type II errors (there are detailed discussions in Winer, 1971; Underwood, 1981). In this paper, the problem will be ignored on the grounds that Type II errors leading to inappropriate pooling will have the effect of increasing the probability of Type I errors in the tests for impacts. Thus, impacts may be detected when, in fact, they are not present. This seems a more desirable error than concluding that no impact is present when there is one — a likely risk in some of the tests described below which have few degrees of freedom and, probably, not very great power.

If the test for  $B \times I$  were significant, it would indicate the presence of an unnatural perturbation associated with the putatively impacted location. This is as close to a causal relationship between the disturbance and its measured impact (Underwood & Peterson, 1988) as could be expected in an experiment without replication of the experimental treatment.

If the test were not significant, it would not mean that the disturbance has had *no* effect. Rather, it indicates that whatever effects have occurred are too small to cause any untoward change in the population being examined. This is equivalent to a legal verdict of "not proven" rather than "innocent". What would be clear from such a result is, however, that there is no evidence of an impact and therefore no requirement for management or remedial action (assuming, of course, that the test was powerful enough to detect important changes; Green, 1979; Peterman, 1990; Cohen, 1977; Andrew & Mapstone, 1987; Underwood, 1981).

This design is only illustrative. It still suffers from the problem that there is no temporal replication and therefore chance differences, of no importance in the context of assessment of environmental impact, would be detected and interpreted as being indicative of impact (see the earlier discussion). It is not a useful analysis — see the above comments on temporal replication — but it demonstrates the appropriate asymmetrical contrasts (Winer, 1971; Underwood, 1978, 1984, 1986; Underwood & Verstegen, 1988).

## MULTIPLE CONTROLS, MULTIPLY SAMPLED BEFORE AND AFTER

An appropriate analysis would consist of appropriately replicated sampling at several times before the development and several times after, as in the Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986) procedure, but in the potentially impacted and in the replicated control locations. From these data, it is possible to ascertain whether there is an interaction between the difference between the impacted and control sites through time. This is much more likely to be attributable to the development itself than any simpler analysis. The procedure for analysis is illustrated in Table II. There is one orthogonal a priori contrast between the potentially impacted and the other locations

TABLE II

Asymmetrical sampling designs to detect environmental impact;  $l$  Locations are sampled, each with  $n$  random, independent replicates, at each of  $t$  Times Before (= Bef) and  $t$  Times After (= Aft) a putative impact starts in one Location ("Impact"); there are  $(l-1)$  Control Locations; Locations represent a random factor and every Location is sampled at the same time; Times represent a random factor nested in each of Before or After.

Source of variation	Degrees of freedom	
Before vs. After	= B	1
Among Times (Before or After)	= T(B)	$2(t-1)$
Among Locations	= L	$(l-1)$
Impact vs. Controls <sup>a</sup>	= I	1
Among Controls <sup>a</sup>	= C	$(l-2)$
$B \times L$		$(l-1)$
$B \times I^{a,d,e}$		1
$B \times C^{a,d}$		$(l-2)$
$T(B) \times L^a$		$2(t-1)(l-1)$
$T(Bef) \times L^a$		$(t-1)(l-1)$
$T(Bef) \times I^{a,c}$		$(t-1)$
$T(Bef) \times C^{a,c}$		$(t-1)(l-2)$
$T(Aft) \times L^a$		$(t-1)(l-1)$
$T(Aft) \times I^{a,b,c}$		$(t-1)$
$T(Aft) \times C^{a,b}$		$(t-1)(l-2)$
Residual		$2lt(n-1)$
Total		$2ltn - 1$

<sup>a</sup> Repartitioned sources of variation.

<sup>b</sup> If  $T(Aft) \times C$  is not significant, impact can be detected by:

(i)  $F = \text{Mean Square } T(Aft) \times I / \text{Mean Square Residual}$ ;

(ii) 2-tailed  $F = \text{Mean Square } T(Aft) \times I / \text{Mean Square } T(Bef) \times I$  is not significant.

<sup>c</sup> If  $T(Aft) \times C$  is significant, impact can be detected by (see text for details):

(i)  $F = \text{Mean Square } T(Aft) \times I / \text{Mean Square } T(Aft) \times C$  is significant;

(ii) 2-tailed  $F = \text{Mean Square } T(Aft) \times I / \text{Mean Square } T(Bef) \times I$  is significant;

(iii) 2-tailed  $F = \text{Mean Square } T(Aft) \times C / \text{Mean Square } T(Bef) \times C$  is not significant.

<sup>d</sup> If and only if there are no short-term temporal interactions in <sup>b,c</sup> and  $B \times C$  is not significant, impact can be detected by  $F = \text{Mean Square } B \times I / \text{Mean Square Residual}$ .

<sup>e</sup> If and only if there are no short-term temporal interactions in <sup>b,c</sup>, but  $B \times C$  is significant, impact can be detected by  $F = \text{Mean Square } B \times I / \text{Mean Square } B \times C$ .

(Fig. 2B). The focus of interest is, however, on the interaction between that contrast before and after the development begins.

The features of this design that warrant careful thought are the use of randomly determined sampling times prior and then subsequent to the development itself. Sampling is done in all locations at the same times, but these are randomly chosen. This ensures that the time courses of the populations in each of the sites are sampled in some way that will break up potential cyclic patterns that would exacerbate or minimise temporal variance. See the discussion in Stewart-Oaten et al. (1986) on this point. Although it is often considered that sampling for environmental impact should be done on some seasonal, or other humanly-driven determination of time course, there is no particularly compelling justification for this. Even though there may be strict seasonal cycles in the abundance of some species, unless they are absolutely concordant from year to year, sampling from one 3-monthly period to another may not represent the seasonal pattern at all (see examples in Stewart-Oaten et al., 1986). It is better, under these circumstances, to sample randomly through time, but not at too widely different time periods so that it is reasonable to determine temporal variance without having imposed different sorts of temporal patterns into the data (Stewart-Oaten et al., 1986). The best designs would incorporate several temporal scales of sampling (Underwood, 1991a), but these are not further discussed here.

This design now has several key features of importance. First, there is spatial replication. Second, there is temporal replication before and after the proposed development. Finally, there is the possibility of detecting impacts as an actual change in abundance in only the impacted location, which would appear as a difference between after and before in that location only. This would require interpretation of some multiple comparisons procedure having demonstrated an interaction between the impacted and other locations through time (Table II). More usually, however, there will already be such an interaction in the data before the development occurs. Therefore, this design also allows formal tests on the magnitude of interaction in the impacted versus the control locations. This is demonstrated in Table II.

In this case (Table II; Fig. 2B), sampling is done at several ( $t$ ) independent times at random intervals before and after the putative impact begins. Independence through time may be difficult to achieve because of inherent serial correlations in abundances of populations. All locations are sampled at these times (see below for what happens when this cannot be achieved); time of sampling and locations are fully orthogonal factors in the sampling design.

It is not important in this design that Mean Square estimates do not allow a formal test of differences from Before to After, nor between Impacted and Control locations. It is not important to determine whether or not there are consistent differences in abundances of the sampled organism among locations. An impact must appear as an interaction between the differences among locations before the impact starts and those differences prevailing after it begins.

Analysis for environmental disturbance then depends on the temporal scale of the

effects of the impact. In the simplest case, consider the situation when there is no short-term temporal interaction among control locations after the disturbance starts (i.e.  $T(Aft) \times C$  is not significant in Table II). There is no difference in the temporal pattern from one control location to another. A disturbance that affects the impacted location so that it differs from the controls to a different extent from time to time can then be detected using the  $F$ -ratios of  $T(Aft) \times I$  divided by the Residual, as indicated in Table II (footnote <sup>b</sup>). As noted before, this involves an assumption that there is no error in concluding that  $T(Aft) \times C$  is not significant.

Often, however, fluctuations in abundances from time to time vary significantly from location to location even when there is no human disturbance. Thus, in Table II, the interaction among control locations from time to time of sampling before and after the disturbance will be significant. This will occur when relatively large changes in numbers of populations occur out of phase in different places. A disturbance affecting the population in an unnatural manner must cause a changed pattern of temporal interaction after it begins. Thus, the impact must cause an altered pattern of differences between the mean abundance in the impacted and those in the control locations. This is obvious in Fig. 2B; consider at each time of sampling the differences in abundances between the impacted location and the location with the greatest abundance. These differences are altered by the impact.

So, there should be a difference between the interaction between times of sampling and the differences between impacted and control locations that occur after the impact starts ( $T(Aft) \times I$  in Table II) and this pattern of interaction among the control locations after the impact starts ( $T(Aft) \times C$  in Table II). Furthermore, the pattern of interaction ( $T(Aft) \times I$  in Table II) should no longer be the same as occurred before the impact started ( $T(Bef) \times I$  in Table II). Both propositions can be tested by  $F$ -ratios identified as footnote <sup>c</sup> in Table II.

Finally, any change in the interaction found after the disturbance between the impacted and control locations might be due to general changes coincident with the disturbance. To demonstrate that the change is associated with the disturbance and not part of a general change, there should be no change in the temporal interactions among control locations after ( $T(Aft) \times C$ ) compared with before ( $T(Bef) \times C$ ) the disturbance. This is testable as in Table II, footnote <sup>c</sup>.

Note that two of these tests ((ii) and (iii) in footnote <sup>c</sup> in Table II) are 2-tailed, unlike traditional 1-tailed  $F$ -ratios in analysis of variance. This is because interactions among times of sampling may be increased or decreased from before to after a disturbance starts. There is no a priori requirement that a particular one of the component Mean Squares will be larger than the other if they are not equal. The remaining test ((i) in footnote <sup>c</sup> of Table II) is 1-tailed; it can be derived directly as a 1-tailed test from the appropriate linear model of the analysis of variance (see, for example, Underwood (1981) for general procedures).

In biological terms, these tests detect whether a putative impact has caused some change to the population in one location making it vary from time to time differently

from the temporal pattern found on average in control locations ( $T(Aft) \times I$  will differ from  $T(Aft) \times C$ ). The second test determines whether the disturbance makes the population different from before in the time-course of its abundance ( $T(Aft) \times I$  will differ from  $T(Bef) \times I$ ).

There is also the possibility that an environmental impact occurs in a system which is not interacting temporally ( $T(Aft) \times C$  in Table II is not significant) and causes an impact of greater duration. Thus, after the disturbance, there is no increased interaction from time to time between the impacted and control locations.

Instead, the impact is a sustained effect on the population in one location after the impact starts (Fig. 2B), causing a larger difference in the abundance after versus before in the impacted location than would be found in the control locations. Where there is no shorter-term temporal interaction ( $T(Aft) \times C$ ,  $T(Aft) \times I$  are not significant in the tests described previously), the more sustained impact (affecting  $B \times I$  in Table II) can be tested.

First, consider what to do when there is no general interaction among control locations from before to after the disturbance (i.e.  $B \times C$  is not significant when tested against the Residual in Table II). Under these circumstances, a sustained interaction should cause the difference between impacted and control locations after the disturbance to differ from that before. Thus,  $B \times I$  will be significant as identified in footnote<sup>d</sup> in Table II.

If, however, there is a general change in the mean abundances of populations in all locations that is coincident with the beginning of some putative impact,  $B \times C$  would also be significant. Under these circumstances, an impact, to be detected, must cause a larger  $B \times I$  interaction than that measured as  $B \times C$ . This is realistic. An impact must make some pattern of temporal difference in the affected location that is unusual compared to those that naturally occur elsewhere. This can be tested by the  $F$ -ratio of the Mean Squares for  $B \times I$  and  $B \times C$  (in footnote<sup>e</sup> of Table II). This test for  $B \times I$  would not be very powerful, unless there were many control locations (making  $l$  large in Table II). This still agrees with common sense — it is easier to detect unnatural phenomena in a relatively invariant world.

#### WHAT IF SAMPLING CANNOT BE DONE AT THE SAME TIMES?

The great advantage of analysis of variance over the  $t$ -test advocated by Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986) is that it can be more decisive in situations when the locations cannot be sampled simultaneously. The effects of impacts can still be analysed (Table III), using different times of sampling in each location, provided they are randomly scattered. Only major impacts affecting the disturbed location in a sustained ("press") manner would be detectable. Effects of shorter duration, or that influence shorter-term temporal trajectories of mean abundance cannot be examined. This is because the interactions of locations with times of sampling ( $T(Aft) \times C$ ,  $T(Aft) \times I$ , etc., in Table II) are no longer analysable, because the times of sampling vary from location to location (technically are nested in locations and the

TABLE III

Asymmetrical sampling designs to detect environmental impact;  $l$  Locations are sampled, each with  $n$  random, independent replicates, at each of  $t$  Times Before and  $t$  Times After a putative impact starts in one Location ("Impact"); there are  $(l - 1)$  Control Locations; Locations represent a random factor and each Location is sampled at different times; Times represent a random factor nested in combinations of Location and Before or After.

Source of variation		Degrees of freedom	
Before vs. After	= B	1	
Among Locations	= L	$(l - 1)$	
Impact vs. Controls <sup>a</sup>	= I	1	
Among controls <sup>a</sup>	= C	$(l - 2)$	
$B \times L$		$(l - 1)$	
$B \times I^{a,b,c}$		1	
$B \times C^a$		$(l - 2)$	<i>F</i> -ratio vs. $T(B \times L)$
Among Times ( $B \times L$ )	= T( $B \times L$ )	$2l(t - 1)$	
Residual		$2lt(n - 1)$	
Total		$2ltn - 1$	

<sup>a</sup> Repartitioned sources of variation.

<sup>b</sup> If there is no interaction from before to after among Controls (i.e.  $B \times C$  is not significant), an impact can be detected by  $F = \text{Mean Square } B \times I / \text{Mean Square } T(B \times C)$ .

<sup>c</sup> If there is a significant interaction from before to after among Controls (i.e.  $B \times C$  is significant), an impact can be detected by  $F = \text{Mean Square } B \times I / \text{Mean Square } B \times C$ .

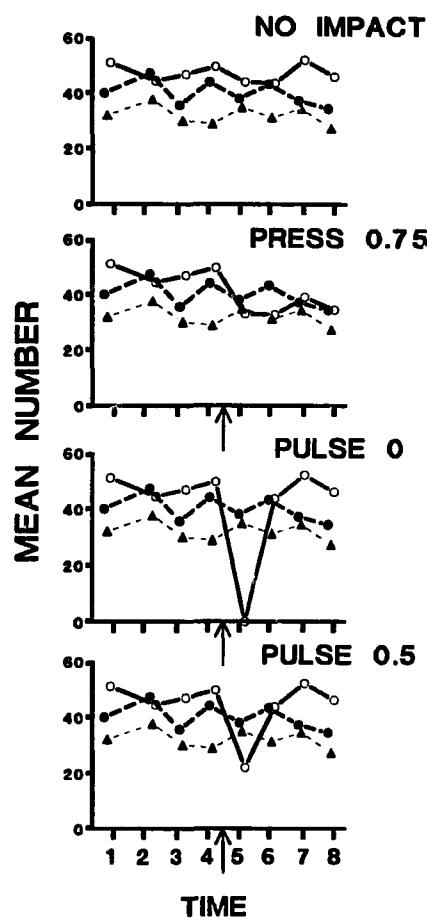
period before or after the impact). The procedures to detect sustained impacts (i.e. causing significant changes measured as  $B \times I$  in Table II) follow those described previously (see footnotes in Table III) and are not discussed further here. Under circumstances where sampling cannot possibly be at the same (orthogonal) times in all locations, these procedures are useful and a considerable advance on attempting to analyse the data using *t*-tests.

### APPLICATION OF THIS SAMPLING DESIGN

#### EXAMPLE OF ENVIRONMENTAL IMPACTS

Consider the situation where there is to be some potential impact, such as the development of a runway in an estuary, or the move of a Naval facility, or the construction of a tunnel under the sea floor in a harbour, or dredging of a main shipping channel, etc. Each may be considered as a potential impact on a very large scale — a bay-wide or an estuary-wide impact. Thus, in this design, the appropriate scale of replication is whole estuaries. Sampling is done on some random allocation of sample units within each estuary, independently at each time of sampling, for several samples before and several samples after a development starts, as in Fig. 3. Similar sampling is done at the same times in two control locations (other bays containing the population being sampled; Fig. 3, Table IV).

An example set of data has been analysed here in order to determine the effects of



**Fig. 3.** Modelled environmental impact in one affected and two control locations (see Table IV for data). Mean abundance in the disturbed location is shown by the solid line. Four conditions of impact are illustrated (no impact, a press to 0.75 of the original mean and pulses to 0 and 0.5 of the original mean). These data are analysed in Table V.

TABLE IV

Data used for simulated environmental impacts affecting a location; in each case, 3 locations (1 Impacted and 2 Controls) were sampled at the same 4 Times before and again 4 Times after an impact began; data are simulated means from  $n = 5$  replicate samples in each location at each time (see Fig. 3); four conditions are simulated in the Impacted location.

Location condition	Control 1	Control 2	Impacted			
			(i) No impact	(ii) "Press" impact to 0.75 of original mean	(iii) "Pulse" impact to 0	(iv) "Pulse" impact to 0.5 of original mean
<b>Time Before</b>						
1	40.0	32.0	51.0	51.0	51.0	51.0
2	47.2	37.6	44.4	44.4	44.4	44.4
3	35.6	30.0	46.8	46.8	46.8	46.8
4	44.0	28.8	49.6	49.6	49.6	49.6
<b>Time After</b>						
1	38.0	34.8	44.0	33.0	0.0	22.0
2	43.2	31.2	43.6	32.7	43.6	43.6
3	37.2	34.4	52.0	39.0	52.0	52.0
4	34.4	27.2	46.0	34.5	46.0	46.0

an impact. To simulate an environmental impact, data after the putative development started were altered as a theoretical exercise.

Two types of disturbance were simulated. First, a "press" disturbance (Bender et al., 1984) continuously removed one-quarter of the population in the impacted site. Thus, the population was "pressed" down to 0.75 of its natural abundance by a sustained disturbance (Fig. 3). This mimics the sort of environmental perturbation caused by sustained discharge of toxic chemicals, or sewage, or continued disruption to a population by harvesting (e.g. Underwood, 1989).

As a result, there was a significant interaction between the period before and after the impact began and the difference between the impacted and control locations ( $B \times I$  was significant in Table V(ii)). Note that this can reasonably be attributed to the disturbance in the impacted location because no such change in patterns of difference from place to place occurred in the control locations ( $B \times C$  was not significant in Table V(ii)). In this simple case, the impact was easily detected.

It is, of course, still possible that the change in abundance in the disturbed location was not due to the disturbance, but was coincidental. If there were more control locations, this would be less reasonable as an alternative explanation because such events would not have been found in a larger set of places and therefore would have been considered to be unlikely by chance. Even with two control locations, it is more realistic to attribute change to the defined disturbance, because two other locations showed no coincident change. This is a very different proposition from discovering the equivalent interaction when there is only one control location which may, itself, have a change of abundance causing an interaction that has nothing to do with the putative impact. The more locations sampled which show no interaction through time, the less likely it is that an interaction with the impacted location could be coincidental and not due to the disturbance. Such conclusions will always be subject to greater doubt than in controlled, replicated experiments (e.g. Hilborn & Walters, 1981; Underwood, 1989; Underwood & Peterson, 1988).

The second type of disturbance was an extreme "pulse" (Bender et al., 1984) which eliminated the species for a brief period from the disturbed location. Pulse disturbances are short-term and then removed. Environmental pulses are often acute, short-term and accidental, caused by such processes as an oil-spill, a flood or the short-term disturbance during the construction of a jetty. Responses are characterised by large and/or rapid changes in abundances. They may, of course, lead to long-term disappearance or even local extinction of a species in the affected location, even though the disturbance is short-term. Here, pulses removing all or half (Fig. 3) of the population were modelled. In passing, the distinction between press and pulse disturbances is not nearly as well-defined as the description by Bender et al. (1984). Some apparently pulse disturbances have longer-term influences; for example, an oil-spill may leave residues in the sediments which continue to act as a press disturbance. There are also cases in which the definition may differ according to the relative longevities of the different organisms in the habitat (Underwood, 1991a).

TABLE V

Simulations of environmental impact using data in Fig. 3; four conditions are analysed; in each case, 3 Locations (1 Impacted and 2 Controls) were sampled at the same 4 Times before (Bef) and again 4 times after (Aft) the putative impact began; there were  $n = 5$  replicate samples in each Location at each Time (see Table II for further details of the design and see text for details of the impacts); the variance among replicates in every sample (the Residual Mean Square) was set at 100; \* indicates  $p < 0.05$ ; \*\*,  $p < 0.01$ ; ns,  $p > 0.05$ ; only relevant  $F$ -ratios are indicated with superscripts identifying the tests as in footnotes to Table II.

Source of variation	Df	(i)		(ii)		(iii)		(iv)	
		No impact		“Press” impact to 0.75 of original mean		“Pulse” impact to 0 of original mean		“Pulse” impact to 0.5 of original mean	
		MS	F	MS	F	MS	F	MS	F
Before vs. After = B	1	191.88		1046.41		980.21		485.21	
T(B)	6	176.91		172.36		572.91		274.08	
Among Locations = L	2								
Impact vs. Controls	1	3445.07		877.60		966.40		2004.07	
Among Controls	1	1364.05		1364.05		1364.05		1364.05	
B × L	2								
B × I	1	100.60		d1.01 ns		951.27		280.27	
B × C	1	154.45		d1.54 ns		154.45		154.45	
L × T(B)	12								
T(Bef) × L	6								
T(Bef) × I	3	223.14		223.14		223.14		223.14	
T(Bef) × C	3	141.60		141.60		141.60		141.60	
T(Aft) × L	6								
T(Aft) × I	3	185.46		b1.85 ns		161.66		b21.65**	
T(Aft) × C	3	145.97		b1.46 ns		145.97		b1.46 ns	
Residual	96	100.00		100.00		100.00		100.00	

In both modelled cases, the impact caused a major increase in the interaction between times of sampling and the difference between the impacted and control locations ( $T(Aft) \times I$  was significant in Table V(iii) and (iv)). There was no interaction among times of sampling and locations before the impact ( $T(Bef) \times I$  and  $T(Bef) \times C$  were not significant in Table V(iii) and (iv)). Nor was there any interaction among the control locations and time of sampling after the impact ( $T(Aft) \times C$  was also not significant in Table V(iii) and (iv)). Thus, there was clear evidence that something had happened in one location that was coincident with the onset of the putative disturbance. The changes seen there and then were not matched by any change in the controls and were clearly different from what occurred before the putative disturbance began. Where a pulse impact caused a temporary decrease in abundance in one location, this was readily identifiable and detectable because it caused a major temporal interaction in the data.

The complete statistical tests for the interaction caused by the impact are shown as 2-tailed *F*-ratios in Table V. In the case where the pulse disturbance caused a massive change (to zero) in the abundance of the population in the impacted location, the interaction between times of sampling after the impact and the difference between the impacted and control locations was significantly greater after the impact than before (the first 2-tailed test in Table V(iii)). There was no corresponding change in the control locations (second 2-tailed test in Table V(iii)). Both tests provided statistically significant evidence that something untoward occurred in the impacted location after the putative impact.

In the case of the smaller pulse disturbance (to half of the original abundance; Table V(iv)), the corresponding tests were not significant. This is largely because the power of the tests is very small. Power would be notably improved by increasing the number of locations or the number of times sampled, or both. Optimising procedures for allocation of sampling effort could and should be used to increase the sensitivity of these tests (Clarke & Green, 1988; Cochran & Cox, 1957; Cox, 1958; Green, 1979; Underwood, 1981, 1991a). Determination of appropriate sampling effort and the design of the necessary pilot studies is outside the scope of this discussion.

#### EXAMPLE WHERE THERE ARE TWO SPATIAL SCALES OF POTENTIAL IMPACT

Where the spatial scale of a potential disturbance cannot be predicted before it occurs, but its time-course is known, these designs can be extended. In this case, it is assumed that the impact will not necessarily affect an entire estuary or bay. Consider a development that might affect only a small area or a whole region. For example, building a jetty in one site in an enclosed bay may only affect the population of fish in the site directly disturbed by the jetty. There may be pulse effects while construction takes place and press effects after it is in place (e.g. it may cause a permanent change in local water-flow or arrival of larvae, etc.). Thus, impacts might be predicted on a local scale; other parts of the bay should not be disturbed.

Yet, it may be that the jetty is to be built in association with, for example, refuelling of boats and, as a result, there may be short-term (pulse) leaks of water-borne chemicals, or long-term, chronic (press) small leaks of fuel that gradually affect the whole bay. In this case, impacts on the whole bay should be considered.

Under such circumstances, two-stage spatial sampling (bays and sites within each bay) and the appropriate analyses are essential. This can be done by extension of the previous design. In addition to several control locations being sampled, several random sites are sampled in every location (including the potentially impacted one). In the potentially disturbed location, the set of sites sampled must, of course, include the one that might be disturbed. All sites, in all locations must, as before, be sampled at the same times; times of sampling must be orthogonal in all locations and sites.

In some instances, there will be no choice between the various options. There are no natural bays or estuaries or equivalents in the environments to be sampled. Thus, a point source of pollution along a coast does not have to be examined at two different spatial scales of replication (or rather three, including the randomized replicates within each sampling site). There are no locations, as such, naturally in the system. Nevertheless, it may be good practice to nest a number of spatial scales inside the locations to be sampled, simply to cover the possibility that the impact is wider or narrower than thought (by local environmentalists) or denied (by the developer) before the development actually takes place. Use of several spatial scales can obviate arguments after the development that one has simply analysed everything in the wrong places. Similar arguments might apply to time courses (Underwood, 1991a), but careful consideration of natural rates of change in the species of interest should resolve that sort of argument because biological information on these natural rates of change would, of course, be used in planning the time scale of sampling (at least in theory!).

### SIMULATED IMPACTS

The situation where there are two possible spatial scales is modelled here, using real data on abundances of a small snail, *Littorina unifasciata*, that is common on rocky shores in southern Australia and which has been sampled as part of various studies (Underwood, 1981; Underwood & Chapman, 1985, 1989, 1992). The details are unimportant, but snails were counted in samples of 10 quadrats ( $125 \text{ cm}^2$ ) on 8 randomly-chosen occasions between 1980 and 1986. The timing of samples was sufficiently spaced to be confident that populations in each location were being sampled independently. Four locations in the Cape Banks Scientific Marine Research Station, some 30 to 100 m apart, were sampled. In each location, 4 sites, some 3–10 m apart were sampled. The former are equivalent to “bays” in the previous discussion; the latter are arbitrarily defined, but randomly-chosen sites of the size that might be disturbed by a human activity. Thus, these data have realistic variances and temporal interactions of the sort expected for an abundant marine invertebrate. In simulated impacts, the first four times of sampling were considered as “before” and the impact was imposed from the fifth time of sampling (Figs. 4 and 5).

To examine analytical procedures to detect impacts, press and pulse impacts were simulated at each spatial scale. The analytical framework is shown in Table VI and the appropriate *F*-ratio tests to detect impacts at different spatial and temporal scales are described in Table VII. The results of simulations, using data on littorinids, are shown in Figs. 4 and 5 and Tables VIII and IX. Note that when there were no impacts, there were no interactions among times of sampling and either sites within each location or

TABLE VI

Asymmetrical sampling designs to detect environmental impact at one Site (the "Impacted Site",  $S_I$ ) at one Location ("Impact",  $I_m$ ); at each of  $l$  Locations (Impact and  $(l - 1)$  Controls),  $s$  Sites are sampled, each with  $n$  random, independent replicates, at each of the same  $t$  Times Before and the same  $t$  Times After the putative impact starts; in the potentially impacted location, there are  $(s - 1)$  Other Sites, which are Controls.

Source of variation		Degrees of freedom
Before vs. After	$= B$	1
Among Times within Bef. or Aft.	$= T(B)$	$2(t - 1)$
Among Locations	$= L$	$(l - 1)$
"Impact versus Controls	$= I$	1
"Among Controls	$= C$	$(l - 2)$
$B \times L$		$(l - 1)$
" $B \times I$		1
" $B \times C$		$(l - 2)$
Among Sites within Locations	$= S(L)$	$l(s - 1)$
$B \times S(L)$		$l(s - 1)$
" $B \times S(I_m)$		$(s - 1)$
" $B \times S_I(I_m)$ vs. $O(I_m)$		1
" $B \times O(I_m)$		$(s - 2)$
" $B \times S(C)$		$(l - 1)(s - 1)$
$T(B) \times L$		$2(l - 1)(t - 1)$
" $T(Bef) \times L$		$(l - 1)(t - 1)$
" $T(Bef) \times I$		$(t - 1)$
" $T(Bef) \times C$		$(l - 2)(t - 1)$
" $T(Aft) \times L$		$(l - 1)(t - 1)$
" $T(Aft) \times I$		$(t - 1)$
" $T(Aft) \times C$		$(l - 2)(t - 1)$
$T(B) \times S(L)$		$2l(t - 1)(s - 1)$
" $T(B) \times S(I)$		$2(t - 1)(s - 1)$
" $T(B) \times S_I(I_m)$ vs. $O(I_m)$		$2(t - 1)$
" $T(Bef) \times S_I(I_m)$ vs. $O(I_m)$		$(t - 1)$
" $T(Aft) \times S_I(I_m)$ vs. $O(I_m)$		$(t - 1)$
" $T(B) \times O(Im)$		$2(t - 1)(s - 2)$
" $T(Bef) \times O(Im)$		$(t - 1)(s - 2)$
" $T(Aft) \times O(Im)$		$(t - 1)(s - 2)$
" $T(B) \times S(C)$		$2(l - 1)(t - 1)(s - 1)$
" $T(Bef) \times S(C)$		$l(t - 1)(s - 1)$
" $T(Aft) \times S(C)$		$l(t - 1)(s - 1)$
Residual		$2lts(n - 1)$
Total		$2ltsn - 1$

<sup>a</sup> Repartitioned sources of variation; impacts can be detected by tests described in Table VII, depending on their temporal and spatial scales.

TABLE VII

Dichotomous key to sequence of appropriate statistical tests to detect environmental impacts at different temporal and spatial scales. MS = Mean Square from analysis in Table VI.

---

**1. Disturbance causes impact in one Site in the affected Location ( $S_1$  in Im in Table VI)**

**1a. Impact affects short-term temporal interactions**

**1a(1). There are no short-term temporal interactions among control sites:**

$F = \text{MS T(Aft)} \times \text{O(Im)}/\text{MS Residual}$  is NOT SIGNIFICANT

Impact detected as:  $F = \text{MS T(Aft)} \times S_1$  vs.  $\text{O(Im)}/\text{MS Residual}$  is SIGNIFICANT  
and 2-tailed:  $F = \text{MS T(Aft)} \times S_1$  vs.  $\text{O(Im)}/\text{MS T(Bef)} \times S_1$  vs  $\text{O(Im)}$  is SIGNIFICANT

$F = \text{MS T(Aft)} \times \text{O(I)}$  vs.  $\text{MS T(Bef)} \times \text{O(I)}$  is NOT SIGNIFICANT

$F = \text{MS T(Aft)} \times S(C)$  vs.  $\text{MS T(Bef)} \times S(C)$  is NOT SIGNIFICANT

**1a(2). There are short-term temporal interactions among control sites:**

$F = \text{MS T(Aft)} \times \text{O(Im)}/\text{MS Residual}$  is SIGNIFICANT

Impact detected as:  $F = \text{MS T(Aft)} \times S_1$  vs.  $\text{O(Im)}/\text{MS T(Aft)} \times \text{O(I)}$  is SIGNIFICANT  
and 2-tailed  $F$ -ratios as above under 1a(1).

**1b. Impact does not affect short-term temporal interactions; all tests in 1a(1) and 1a(2) are NOT SIGNIFICANT**

**1b(1). There are no short-term temporal interactions and no before/after interactions among control sites:**

$F = \text{MS B} \times \text{O(Im)}/\text{MS Residual}$  is NOT SIGNIFICANT

Impact detected as:  $F = \text{MS B} \times S_1$  vs.  $\text{O(Im)}/\text{MS Residual}$  is SIGNIFICANT

**1b(2). There are before/after interactions among control sites:**

$F = \text{MS B} \times \text{O(Im)}/\text{MS Residual}$  is SIGNIFICANT

Impact detected as:  $F = \text{MS B} \times \text{O(Im)}/\text{MS B} \times \text{O(I)}$  is SIGNIFICANT  
and 2-tailed:  $F = \text{MS B} \times \text{O(Im)}$  vs.  $\text{MS B} \times S(C)$  is NOT SIGNIFICANT

**2. Disturbance causes impact detectable only at the scale of a whole location (Im); all tests in 1 above are NOT SIGNIFICANT**

**2a. Impact affects short-term temporal interactions**

**2a(1). There are no short-term temporal interactions among control sites:**

$F = \text{MS T(Aft)} \times C/\text{MS Residual}$  is NOT SIGNIFICANT

Impact detected as:  $F = \text{MS T(Aft)} \times I/\text{MS Residual}$  is SIGNIFICANT  
and 2-tailed:  $F = \text{MS T(Aft)} \times I$  vs.  $\text{MS T(Bef)} \times I$  is SIGNIFICANT

$F = \text{MS T(Aft)} \times C$  vs.  $\text{MS T(Bef)} \times C$  is NOT SIGNIFICANT

**2a(2). There are short-term temporal interactions among control sites:**

$F = \text{MS T(Aft)} \times C/\text{MS Residual}$  is SIGNIFICANT

Impact detected as:  $F = \text{MS T(Aft)} \times I/\text{MS T(Aft)} \times C$  is SIGNIFICANT  
and 2-tailed  $F$ -ratios as above under 2a(1).

**2b. Impact does not affect short-term temporal interactions; all tests in 2a(1) and 2a(2) are NOT SIGNIFICANT**

**2b(1). There are no short-term temporal and no before/after interactions and no before/after interactions among control sites:**

$F = \text{MS B} \times C/\text{MS Residual}$  is NOT SIGNIFICANT

Impact detected as:  $F = \text{MS B} \times I/\text{MS Residual}$  is SIGNIFICANT

**2b(2). There are before/after interactions among control sites:**

$F = \text{MS B} \times C/\text{MS Residual}$  is SIGNIFICANT

Impact detected as:  $F = \text{MS B} \times I/\text{MS B} \times C$  is SIGNIFICANT

---

locations (analyses in Tables VIII and IX). Thus, the time-courses of populations were very similar in these data, even though there were differences among the mean abundances in each location and among sites in each location (Figs. 4 and 5).

#### SMALL-SCALE PULSE

The first situation simulated was a small-scale pulse, removing all organisms at one site ( $S_1$ ) in the affected location (Fig. 4). The populations immediately recovered. In this scale of impact, there should be effects on the temporal differences between the impacted site ( $S_1$ ) and the other sites in the impacted location. These are analysed in Table V(II)(ii). Thus, after the impact begins there should be:

- (i) significant interaction among times of sampling and the difference between the affected site and the mean of the control sites in the disturbed location. This was significant in Test 1a(1) in Table IX.
- (ii) an increase from before to after in the temporal interaction between the affected site and the mean of the control sites in the disturbed location (i.e. the detected effects are temporally related to the onset of disturbance). This was significant in the first 2-tailed  $F$ -ratio in Test 1a in Table IX.
- (iii) no increase from before to after the disturbance in the temporal interaction among control sites in the disturbed location (the second 2-tailed  $F$ -ratio in Test 1a in Table

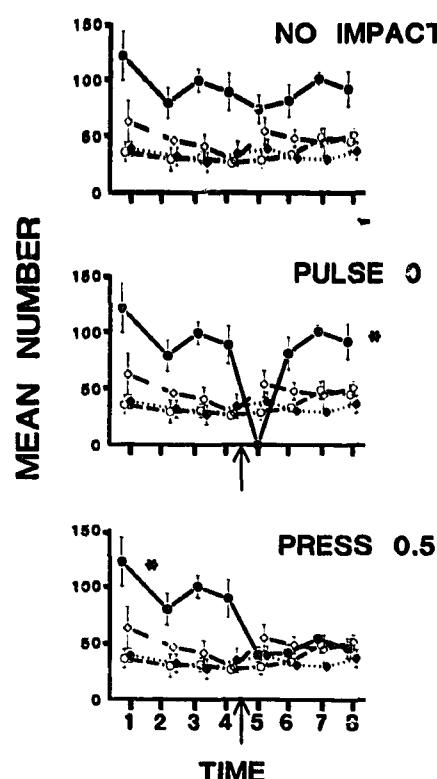


Fig. 4. Modelled environmental impacts affecting the abundance of the snail *Littorina unifasciata* in one site (indicated by the asterisk and solid line), with three control sites, in one location. Note that the means over all four sites in the disturbed location, when there was no impact, are the means of that location plotted in Fig. 5. A pulse to 0 and a press disturbance to 0.5 of the original mean abundance are illustrated. These data are analysed in Tables VIII and IX.

TABLE VIII

Analysis of variance of environmental impacts affecting mean abundance of *Littorina unifasciata* in one Site ( $S_1$ ) in the Impacted Location (Im) (Fig. 4); (ii) pulse to 0; (iii) press to 0.5 of original mean; or in an entire Location (Im) (Fig. 5); (iv) pulse to 0; (v) press to 0.5 of original mean; (i) is no impact; see text for details of data and simulated impacts.

Source of variation	df	(i)		(ii)		(iii)		(iv)		(v)	
		MS	MS	MS	MS	MS	MS	MS	MS	MS	MS
B	1	9494.0	12540.3	17082.3	18479.3	31949.8					
T(B)	6	1862.2	1932.4	1899.9	2015.4	1888.1					
L	3										
I	1	74264.2	88631.2	108982.5	115056.5	170852.7					
C	2	320756.9	320756.9	320756.9	320756.9	320756.9					
S(L)	12	96596.7	94338.1	92054.9	95404.1	93792.6					
B × L	3										
B × I	1	4063.3	1486.3	37.5	9.0	5943.0					
B × C	2	10552.7	10552.7	10552.7	10552.7	10552.7					
B × S(L)	12										
B × S(Im)	3										
B × S <sub>1</sub> vs. O(Im)	1	1695.0	8383.7	24457.5	4083.2	10182.9					
B × O(Im)	2	548.8	548.8	548.8	328.7	181.8					
B × S(C)	9	2415.9	2415.9	2415.9	2415.9	2415.9					
T(B) × L	18										
T(Bef) × L	9										
T(Bef) × I	3	1333.1	1333.1	1333.1	1333.1	1333.1					
T(Bef) × C	6	1310.1	1310.1	1310.1	1310.1	1310.1					
T(Aft) × L	9										
T(Aft) × I	3	745.9	3881.5	478.6	14578.8	381.8					
T(Aft) × C	6	2075.8	2075.8	2075.8	2075.8	2075.8					
T(B) × S(L)	72										
T(B) × S(Im)	18										
T(B) × S <sub>1</sub> vs O(Im)	6										
T(Bef) × S <sub>1</sub> vs O(Im)	3	698.0	698.0	698.0	698.0	698.0					
T(Aft) × S <sub>1</sub> vs O(Im)	3	1454.7	7921.9	137.4	2565.1	130.0					
T(B) × O(Im)	12										
T(Bef) × O(Im)	6	1245.6	1245.6	1245.6	1245.6	1245.6					
T(Aft) × O(Im)	6	1254.7	1254.7	1249.9	1168.7	78.2					
T(B) × S(C)	54										
T(Bef) × S(C)	27	1126.0	1126.0	1126.0	1126.0	1126.0					
T(Aft) × S(C)	27	1249.9	1249.9	1249.9	1249.9	1249.9					
Residual	512	1726.5	1720.4	1709.4	1710.5	1688.6					

IX(ii)), nor among sites in control locations (the final 2-tailed *F*-ratio in Test 1a in Table IX(ii)). Thus, the only change from before to after is in the impacted site in the affected location. Exactly this pattern was found.

#### SMALL-SCALE PRESS

The second simulation was that of a sustained (press) reduction to half the previous abundance, but in only one site in the affected location. At each time of sampling

TABLE IX

Results of analyses of simulated environmental impacts; *F*-ratios are as indicated in Table VII, analyses of variance are in Table VIII. Only relevant tests are presented (except under 1a as discussed in the text). For 1-tailed tests, ns denotes not significant;  $p > 0.05$ ; \*  $p < 0.05$ ; \*\*  $p < 0.01$ . For 2-tailed tests, ns denotes not significant,  $p > 0.01$ ; \*  $p < 0.10$ , to increase power because of small numbers of degrees of freedom in the tests; \*\*  $p < 0.01$ . (i) Control, no impact; (ii) pulse to 0, (iii) press to 0.5 of original mean in one Site ( $S_1$ ) in the Impacted Location (Im) (Fig. 4); (iv) pulse to 0, (v) press to 0.5 of original mean in an entire Location (Im) (Fig. 5).

Test (see Table VII)	(i) <i>F</i>	(ii) <i>F</i>	(iii) <i>F</i>	(iv) <i>F</i>	(v) <i>F</i>
1a. T(Aft) × O(Im)/Residual	0.73 ns	0.73 ns	0.73 ns	0.68 ns	0.05 ns
1a(1). T(Aft) × $S_1$ vs. O(Im)/Residual	0.84 ns	4.60*	0.08 ns	1.50 ns	0.08 ns
1a(2). T(Aft) × $S_1$ vs. O(Im)/T(Aft) × O(Im)					
2-tailed tests, 1a(1) and 1a(2)					
T(Aft) × $S_1$ vs. O(Im) vs. T(Bef) × $S_1$ vs. O(Im)		11.35*			
T(Aft) × O(Im) vs. T(Bef) × O(Im)		1.01 ns	(1.01 ns)		(15.93**)
T(Aft) × S(C) vs. T(Bef) × S(C)		1.11 ns	(1.11 ns)		(1.11 ns)
1b. B × O(Im)/Residual	0.32 ns		0.32 ns	0.19 ns	0.11 ns
1b(1). B × $S_1$ vs. O(Im)/Residual	0.98 ns		14.31**	2.39 ns	6.03*
1b(2). B × $S_1$ vs. O(Im)/B × O(Im)					
2-tailed test, 1b(2)					
B × O(Im) vs. B × S(C)			(1.27 ns)		(13.27**)
2a. T(Aft) × C/Residual	1.20 ns			1.21 ns	
2a(1). T(Aft) × I/Residual	0.43 ns			8.52**	
2a(2). T(Aft) × I/T(Aft) × C					
2-tailed tests, 2a(1) and 2a(2)					
T(Aft) × I/T(Bef) × I				10.94**	
T(Aft) × C/T(Bef) × C				1.58 ns	
2b. B × C/Residual	6.11*				
2b(1). B × I/Residual					
2b(2). B × I/B × C	0.38 ns				

after the impact began, approximately half of the organisms in each replicate count were removed (Fig. 4). The actual proportion removed was a random variate from a binomial distribution of mean 0.5; this simulated a form of randomized mortality over the site.

Here, after the impact starts, there should be a different pattern of temporal interaction between the impacted and the mean of the control sites in the affected location from that evident prior to the disturbance. Thus, the impact is identifiably affecting one site only. To corroborate this, the interaction from before to after the disturbance among control sites in the impacted location should continue to be similar to that among sites in control locations. Thus, there is not some general pattern of interaction elsewhere that matches that for the impacted site in the impacted location. Only the one site in the one location shows a different pattern after the impact starts. Similarly, there should be no interaction from before to after the disturbance among the sites in the other control locations, so that there is clear evidence that the sort of changes

happening in the impacted location are not indicative of more general changes that have nothing to do with the disturbance.

These were precisely the patterns found. There was a significant interaction from before to after the disturbance in the difference between the impacted and control sites ( $B \times S_I$  vs.  $O(Im)$  in Table IX(iii); Test 1b(1)). There was no such interaction among other sites in the impacted location ( $B \times O(I)$  was not significant in Table IX(iii); Test 1b), nor among sites in the other locations (Table VIII(iii);  $F$ -ratio for  $B \times S(C) = 1.41$ , with 9 and 512 df;  $P > 0.10$ ).

#### LARGE-SCALE PULSE

A pulse-impact affecting the whole location was applied for one time-period. This was an extreme, short-term catastrophe — all organisms were killed in all four sites in the impacted location (Fig. 5), but the population recovered by the subsequent time of sampling.

A large-scale pulse, affecting a whole location should cause a major change in the interaction between locations and times of sampling after the impact compared with before. This was detected as a significant interaction  $T(Aft) \times I$  in Table IX(iv), Test 2a(1). To demonstrate that this was only affecting the impacted location and was not a general change occurring coincidentally with the period of sampling, the difference between impacted and other locations should interact with time differently from the temporal interaction among control locations.

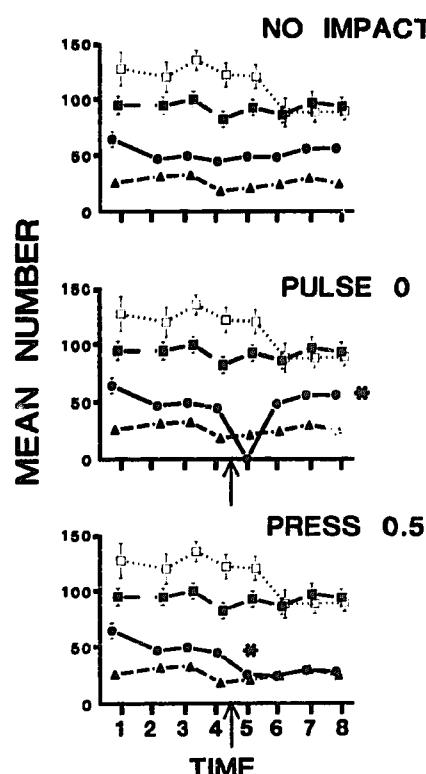


Fig. 5. Modelled environmental impacts affecting the abundance of the snail *Littorina unifasciata* (at a larger spatial scale than in Fig. 4) in one location (indicated by the asterisk and solid line), with three control locations. A pulse to 0 and a press disturbance to 0.5 of the original mean abundance are illustrated. These data are analysed in Tables VIII and IX.

Formally, two conditions should be met. There should be more temporal interaction in the difference between the impacted and control locations *after* the impact than *before* (i.e. any detected effects are temporally related to the onset of the disturbance). This was found to be significant in Table IX(iv), the first 2-tailed *F*-ratio in Test 2a. There should also be similar temporal interaction among control locations *after* and *before* the impact (i.e. no general change has occurred at the same time as the impact). This was again found; the second 2-tailed *F*-ratio in Table IX(iv), Test 2a was not significant.

#### LARGE-SCALE PRESS

The final, large-scale impact was a simulated press for the impacted location (Fig. 5). This was simulated by reducing each quadrat by a proportion, randomly chosen from a binomial distribution of mean 0.5 (as described above for the small-scale press).

The outcome of a large-scale, long-term change in abundances at all sites in one location will vary according to the nature of differences among sites and temporal patterns in these differences before the disturbance occurs. As a result, large-scale disturbances may be detected in tests for changes in small-scale interactions (i.e. among sites within a location). For example, if abundances at all sites in one location are all reduced, there should be a change in the temporal interactions among sites in the impacted location compared with those occurring before the impact. If all sites have decreased abundance after the impact starts, the interactions through time must be different from those occurring when the means were larger before. This is how the impact was detectable in the simulation (Table IX(v), Test 1b(1)). This was the same result described above for a small-scale press, the temporal interactions among control sites in the impacted location after the disturbance were much smaller than before ( $T(Aft) \times O(Im)$  and  $T(Bef) \times O(Im)$  in Table VIII(v); 2-tailed test in 1a(2) in Table IX(v)). This was due to the decreased abundances in these sites (Fig. 5); they did not vary as much from time to time.

The analysis does, however, reveal substantial change in all sites in the impacted location. This was not the case for the short-term pulse described previously. Control sites in the impacted location did not change when a small-scale pulse was simulated (see  $T(Aft) \times O(Im)$  and  $T(Bef) \times O(Im)$  in Table VIII(iii) and Test 1a in Table IX(iii)).

If a press impact increased the mean densities in the disturbed location, there would be generally greater temporal interaction among the sites there than before and than in control locations. The 2-tailed tests used above would still detect the change. If, however, there was little interaction among sites through time before an impact started, a press could not cause a decrease. Under these circumstances, a press should cause the more general interaction among locations from before to after the disturbance. There should be a greater interaction from before to after the impact started in the mean difference between the impacted and control locations. This would be detectable in Tests 2b in Table IX. Nevertheless, a variety of scenarios will lead to different patterns in the results of significance tests in Table IX.

## SOME CONSIDERATIONS FOR IMPROVEMENT OF THESE DESIGNS

The designs introduced above are demonstrably superior in logic and for interpretation than the more traditional, unreplicated BACI design of Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986). There are several features of them that require more development. First, it would be useful to determine, in advance, the power of the various tests, particularly the 2-tailed *F*-ratios used after the analyses of variance. Calculation of power requires some knowledge, in advance, of the magnitude of an environmental impact if it is going to occur (Clarke & Green, 1988; Cohen, 1977; Peterman, 1990; Underwood, 1981). This is not an easy task in this context because predictions must be made about a temporal interaction, rather than a relatively simple shift in mean abundance.

On the other hand, as argued by Peterman (1990), it should be straightforward to define, in advance, the magnitude of change in abundance of a species that should be considered to be biologically important in some habitat. This would then set the minimal change that should be detected and, thus, enable sampling to be designed with adequate power.

In addition to some statement about the anticipated magnitude of change in a population after an impact begins, calculation of power is also dependent on knowledge of the natural rates of variation in abundance of the population. This is clearly necessary, because it will influence the size of replication in the sampling programme (see below) and should be crucial in influencing the choice of species to monitor (Underwood, 1989; Underwood & Peterson, 1988). Species with very large natural temporal variation in abundance will only reveal environmental impacts if they are massive and long-term. On the other hand, species with very little temporal change may be useless for detection of impact because they are probably inert and resistant to most perturbations that affect their environments. That is why they show little natural temporal change. As a result, they are unlikely to respond to many types of human disturbance. Knowledge of the variance in time of monitored species is crucial for planning a powerful sampling programme.

Second, the analyses involve preliminary decisions being made and subsequent tests being used after components of variation have been eliminated, post-hoc, from the linear model for the analysis (as described in Table VII). As discussed earlier, these procedures are inherently uncertain because of potential Type II errors. Post-hoc pooling procedures recommended by Winer (1971) and Underwood (1981) could be applied. Had they been used here, most of the statistical conclusions would have been identical. There does, however, need to be research on appropriate procedures to protect against Type II errors when using the analyses recommended in Table VII.

Third, impacts that affect abundances should be properly considered as multiplicative, rather than additive. For example, in the cases used above, press impacts were described as decreasing a population to some proportion of its natural abundance. This

would be better analysed by transforming abundances to logarithms to create an additive, linear effect on the data (e.g. Green, 1979; Underwood, 1981; Winer, 1971).

Fourth, the analyses involve several forms of replication. Replicate control locations are necessary, as are replicated times of sampling. Allocation of effort to the various types and scales of replication will require careful thought and appropriate pilot studies. In many cases, there should be several scales of replication to ensure that human disturbances are not more widespread than claimed by those responsible. Thus, several sites in an estuary might be used as controls for assessment of the effects of an outfall (in the "impacted" site). If there are no detectable differences from the controls in the abundance, or its time-course, of the sampled species in the site of the outfall, it must be concluded that there is no impact. Yet, the outfall may have had a much more widespread effect, disturbing organisms in all of the sites. Hence, spatial replication should include sites elsewhere (for example, in other estuaries).

Replication in time also requires careful thought. Stewart-Oaten et al. (1986) discussed the problems of fluctuating populations. Some impacts may not influence the mean abundance of a population (Underwood, 1991a). Therefore, consideration should be given to sampling designs that include several hierarchical or clustered series of samples at different time-intervals. To be efficient, this sort of design requires detailed information about the natural time-course of the population being studied. On the other hand, such detailed knowledge is crucial for the *a priori* assessment of potential power of sampling (see above).

Fifth, the analyses depend on assumptions about the data being collected (Cochran, 1947; Eisenhart, 1947; summarised in Underwood, 1981). There are minor assumptions about normality of data and often resolvable issues about homogeneity of variances within samples. Analyses of variance are robust with respect to these problems, particularly in large, balanced designs (Underwood, 1981, for a review). There is, however, a major issue concerning independence of the data, which must be collected independently from time to time, place to place and sample to sample. In the context of environmental sampling, temporal independence (i.e. lack of serial correlation from one time of sampling to another) is probably the most difficult to achieve (Clarke & Green, 1988; Green, 1979; Stewart-Oaten et al., 1986). Again, achieving independent sampling through time requires knowledge of the natural rates of change, turnover and longevity of the organisms being sampled. There are also too many examples of non-independent spatial sampling (Green, 1979; Hurlbert, 1984; Underwood, 1981, 1986) for complacency about the basics of valid environmental assessment. There are, however, no new problems imposed by the designs considered here; independent sampling is necessary for all analyses of time-courses of populations in environmental work. The designs discussed here are no exception to this, but equally cause no greater problems.

All of these considerations lead to an inescapable general conclusion about the future of environmental monitoring. There should be long-term, experimental research programmes to determine the natural rates of change and their spatial interactions in species of interest for detecting environmental impact (see particularly Underwood,

1989). This should be coupled with experimental disturbances (where appropriate) and with post-hoc analyses of previous disturbances to determine what actually happens when human activities intrude into the real world (see also Hilborn & Walters, 1981; McGuinness, 1990). Such knowledge would allow very precise determination of power for future sampling. It would also determine the spatial scales of those types of disturbance that have already been seen on previous occasions. It would allow assessment of the effects of impacts on the size of populations, because the consequences of already existing perturbations could be investigated in terms of fecundity, recruitment, growth, longevity, etc. All in all, future environmental assessment should be based on much better fundamental understanding of natural populations.

## CONCLUSIONS

Sampling to detect potential environmental impact in very heterogeneous environments with markedly divergent time courses in the abundances of species of interest, requires rather complicated sampling designs. These are, however, tractable and efficient. They will, nevertheless, involve the expenditure of more money than has usually been the case. This is particularly apparent when assessment of much environmental impact consists simply of constructing lists of species and there is no quantitative component at all. It is worth, therefore, examining why the extra expenditure of money during the sampling exercises is warranted. First, without this increase in number of control locations, there is no logical or rational reason why any apparently detected impact should be attributed to the human disturbance of the apparently impacted location. This would not be accepted in normal and routine ecological and experimental analyses. The results would always be rejected by reputable journals (at least, should be rejected by reputable journals, although there are many examples to the contrary — see Hurlbert (1984) and Underwood (1986). The expenditure of extra effort is necessary and the lack of replicated control sites provides insufficient evidence for an impact being due to the development. Thus, such unreplicated sampling can always result in differences of opinion about what the results mean, leaving, as usual, the entire assessment of environmental impact to those random processes known as the legal system.

The other reason that the money should be expended is to protect developers from unwarranted claims that they have caused some impact, when there is no biological evidence to support this contention at all. This is probably an important political argument that will carry some weight with those who have to pay for assessment of potential environmental impact.

Too often, even where there is some realistic quantitative assessment of what is happening to target species in a habitat that might be affected by development, the results are totally ambiguous. Where there is only one control and one potentially impacted site, any data that show differences after the potential impact is presumed

to have started will be interpreted to mean that the development is responsible. This is unwarranted given the inadequacies of the sampling design, even in the improved BACI design proposed by Bernstein & Zalinski (1983) and Stewart-Oaten et al. (1986). It would seem reasonable to demand the extra expense of money in order to produce a result that is unarguable. This is more cost-effective than continuing to spend money first on assessment of environmental impact by biologists and then on lawyers to argue about what it might mean.

Finally, the procedures proposed here would need to be tested by examination of real environmental impacts. Opportunities to do this are available. Such assessments should be done to determine the validity, utility and power (Peterman, 1990) of these proposed procedures when attempting to deal with real variances in real populations.

However assessments of environmental impact are done, two important rules should be obeyed. First, at all times and places where monitoring for environmental impact is attempted, there must be clear statements about the aims of the exercise. Thus, the hypotheses of interest about potential impact must be clearly and unambiguously stated (see, for example, Green, 1979; Underwood, 1990). Only then will it be possible to determine whether the sampling design was appropriate for the job, without imposing confounding in space, time, or both into the data. Second, but much more difficult, is the requirement that eventually assessment of environmental impact will only really become scientific when impacts are themselves treated as experiments. This approach has been advocated several times (e.g. Hilborn & Walters, 1981). For a recent account and review of such possibilities see Underwood (1989). Environmental impact caused by human developments will occur whether or not biologists decide to take notice of them in some integrated and philosophically sound manner. Although increasing validity and worth and the ease of interpretation of monitoring data for assessment of impact is an important exercise, on its own it is insufficient. Much more progress is needed in creating situations where repeated types of impact and planned experimental disturbances that simulate impacts are used as long-term experimental tools for unravelling the effects of human disturbances to natural populations. This ought to be straightforward where there already exist replicated patches of disturbed habitats. For example, on many coastlines there are already several independently sited marinas, power stations or outfall pipes. These should be examined, with replicated controls, as an experimental environmental disturbance to determine, as in other field experiments, what actually happened to populations in disturbed areas. In the absence of such experimental procedures, the designs described here will provide better evidence of causal links between unusual patterns of abundance and the human disturbance in the affected location than is the case with more usual BACI designs.

#### ACKNOWLEDGEMENTS

The thoughts in this paper have been developed with financial support from the Australian Research Council, the CSIRO Jervis Bay Marine Environmental Projects

and the Institute of Marine Ecology of the University of Sydney. Discussion on these topics with M.G. Chapman, K.R. Clarke, P.A. Underwood, M.P. Lincoln-Smith, K.A. McGuinness and A.J. Butler have been extremely valuable. Ms. K.L. Astles greatly assisted with the analyses and modelling. The very helpful reviews of K.R. Clarke and R.H. Green prevented errors in the proposed procedures; thank you, but residual errors are mine.

## REFERENCES

- Andrew, N.L. & B.D. Mapstone, 1987. Sampling and the description of spatial pattern in marine ecology. *Annu. Rev. Oceanogr. Mar. Biol.*, Vol. 25, pp. 39–90.
- Bender, E.A., T.J. Case & M.E. Gilpin, 1984. Perturbation experiments in community ecology: theory and practice. *Ecology*, Vol. 65, pp. 1–13.
- Bernstein, B.B. & J. Zalinski, 1983. An optimum sampling design and power tests for environmental biologists. *J. Environ. Manage.*, Vol. 16, pp. 35–43.
- Clarke, K.R. & R.H. Green, 1988. Statistical design and analysis for a "biological effects" study. *Mar. Ecol. Progr. Ser.*, Vol. 46, pp. 213–226.
- Cochran, W.G., 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, Vol. 3, pp. 22–38.
- Cochran, W.G. & G. Cox, 1957. *Experimental Designs*. Wiley, New York, second edition 617 pp.
- Cohen, J., 1977. *Statistical Power Analysis for the Behavioural Sciences*. Academic Press, New York, 474 pp.
- Connell, J.H., 1983. On the prevalence and relative importance of interspecific competition: evidence from field experiments. *Am. Nat.*, Vol. 122, pp. 661–696.
- Cox, G., 1958. *The Planning of Experiments*. Wiley, New York, 308 pp.
- Crowder, M.J. & D.J. Hand, 1990. *Analysis of Repeated Measures*. Chapman and Hall, London, 253 pp.
- Eberhardt, L.L. & J.M. Thomas, 1991. Designing environmental field studies. *Ecol. Monogr.*, Vol. 61, pp. 53–73.
- Eisenhart, C., 1947. The assumptions underlying the analysis of variance. *Biometrics*, Vol. 3, pp. 1–21.
- Green, R.H., 1979. *Sampling Design and Statistical Methods for Environmental Biologists*. Wiley, Chichester, 257 pp.
- Hand, D.J. & C.C. Taylor, 1987. *Multivariate Analysis of Variance and Repeated Measures*. Chapman and Hall, London, 212 pp.
- Hilborn, R. & C.J. Walters, 1981. Pitfalls of environmental baseline and process studies. *Environ. Impact Assess. Rev.*, Vol. 2, pp. 265–278.
- Hurlbert, S.J., 1984. Pseudoreplication and the design of ecological field experiments. *Ecol. Monogr.*, Vol. 54, pp. 187–211.
- Krebs, C., 1978. *Ecology: The Experimental Analysis of Abundance and Distribution*. Harper & Row, New York, 678 pp.
- McGuinness, K.A., 1990. Effects of oil spills on macro-invertebrates of saltmarshes and mangrove forests in Botany Bay, New South Wales. *J. Exp. Mar. Biol. Ecol.*, Vol. 142, pp. 121–135.
- Peterman, R.M., 1990. Statistical power analysis can improve fisheries research and management. *Can. J. Fish. Aquat. Sci.*, Vol. 47, pp. 2–15.
- Scheffe, H., 1959. *The Analysis of Variance*. Wiley, New York, 477 pp.
- Stewart-Oaten, A., W.M. Murdoch & K.R. Parker, 1986. Environmental impact assessment: "pseudoreplication" in time? *Ecology*, Vol. 67, pp. 929–940.
- Underwood, A.J., 1978. An experimental evaluation of competition between three species of intertidal prosobranch gastropods. *Oecologia (Berlin)*, Vol. 33, pp. 185–208.

- Underwood, A.J., 1981. Techniques of analysis of variance in experimental marine biology and ecology. *Annu. Rev. Oceanogr. Mar. Biol.*, Vol. 19, pp. 513–605.
- Underwood, A.J., 1984. Vertical and seasonal patterns in competition for microalgae between intertidal gastropods. *Oecologia (Berlin)*, Vol. 64, pp. 211–222.
- Underwood, A.J., 1986. The analysis of competition by field experiments. In, *Community Ecology: Pattern and Process*, edited by J. Kikkawa & D.J. Anderson, Blackwells, Melbourne, pp. 240–268.
- Underwood, A.J., 1989. The analysis of stress in natural populations. *Biol. J. Linn. Soc.*, Vol. 37, pp. 51–78.
- Underwood, A.J., 1990. Experiments in ecology and management: their logics, functions and interpretations. *Aust. J. Ecol.*, Vol. 15, pp. 365–389.
- Underwood, A.J., 1991a. Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. *Aust. J. Mar. Freshwater Res.*, Vol. 42, pp. 569–587.
- Underwood, A.J., 1991b. 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*, edited by R.V. Hyne, Office of the Supervising Scientist, Alligator Rivers Region, Australian Government Printing Service, pp. 105–123.
- Underwood, A.J. & M.G. Chapman, 1985. Multifactorial analysis of directions of movement of animals. *J. Exp. Mar. Biol. Ecol.*, Vol. 91, pp. 17–43.
- Underwood, A.J. & M.G. Chapman, 1989. Experimental analyses of the influences of topography of the substratum on movements and density of an intertidal snail, *Littorina unifasciata*. *J. Exp. Mar. Biol. Ecol.*, Vol. 34, pp. 175–196.
- Underwood, A.J. & M.G. Chapman, 1992. Experiments on topographic influences on density and dispersion of *Littorina unifasciata* in New South Wales. *J. Moll. Stud.*, in press.
- Underwood, A.J. & E.J. Denley, 1984. Paradigms, explanations and generalizations in models for the structure of intertidal communities on rocky shores. In, *Ecological Communities: Conceptual Issues and the Evidence*, edited by D.R. Strong et al., Princeton University Press, New Jersey, pp. 151–180.
- Underwood, A.J. & C.H. Peterson, 1988. Towards an ecological framework for investigating pollution. *Mar. Ecol. Progr. Ser.*, Vol. 46, pp. 227–234.
- Underwood, A.J. & P.H. Verstegen, 1988. Experiments on the association between the intertidal amphipod *Hyale media* Dana and the limpet *Cellana tramoserica* (Sowerby). *J. Exp. Mar. Biol. Ecol.*, Vol. 119, pp. 83–98.
- Winer, B.J., 1971. *Statistical Principles in Experimental Design*. McGraw-Hill Kogakusha, Tokyo, second edition, 907 pp.