On Beyond BACI: Sampling Designs that Might Reliably Detect Environmental Disturbances

A. J. Underwood

Ecological Applications, Vol. 4, No. 1. (Feb., 1994), pp. 3-15.

Stable URL:
http://links.jstor.org/sici?sici=1051-0761%28199402%294%3A1%3C3%3AOBBSDT%3E2.0.CO%3B2-A

Ecological Applications is currently published by The Ecological Society of America.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/esa.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Fri Oct 13 16:12:42 2006

Ecological Applications, 4(1), 1994, pp. 3–15
© 1994 by the Ecological Society of America

ON BEYOND BACI: SAMPLING DESIGNS THAT MIGHT RELIABLY DETECT ENVIRONMENTAL DISTURBANCES[1]

A. J. UNDERWOOD
Institute of Marine Ecology, Zoology Building A 08, University of Sydney, New South Wales 2006, Australia

Abstract. Much sampling to detect and quantify human environmental disturbances is flawed by a lack of appropriate replication. BACI (Before–After-Control–Impact) designs have only a single control location, and any conclusions from them are illogical. Asymmetrical designs using one putatively impacted and several control locations can reliably detect a variety of environmental impacts, including those that do not affect long-run mean abundances, but do alter temporal variance. When abundances of populations in different locations show temporal interaction, the asymmetrical designs allow tests for impact that are not possible in BACI designs. Asymmetrical designs are also extendable to sample at hierarchical spatial and temporal scales.

The power of tests using asymmetrical designs is great for non-interactive sets of abundances, but greatest for pulse (short-term) responses to disturbances, large alterations of temporal variance, or combinations of sustained, press responses in mean abundance coupled with altered temporal heterogeneity. Power in temporally interactive sets of data is generally poor.

Alternatives to pre-disturbance sampling, including generalized assessment of spatial and temporal variances and experimental impacts, may provide better guidance for detection of human disturbances.

Key words: asymmetrical analysis; environmental impact; multiple controls; power of sampling; sampling design; spatial and temporal interaction; statistical power.

INTRODUCTION

There are increasing needs for reliable detection of environmental disturbances due to human activities. There are also needs for ecological research to become more concerned with problems of anthropogenic influence on natural systems at spatial and temporal scales of relevance to the organisms and habitats affected (e.g., Peters 1991). Unfortunately, there are often problems in the use of appropriately valid experimental and sampling designs and replication for detection of unnatural disturbances to biological variables and for identifying a causal relation between an observed effect and the putative anthropogenic cause.

Here, I briefly review some of the problems of some of the most widely used procedures. Then, I discuss sampling designs that are appropriate to detect a variety of environmental perturbations. Finally, I consider some of the ecological research programs (sensu Lakatos 1974) that are needed for the future development of practices of environmental sampling and assessment.

Throughout, I have chosen to consider populations as the appropriate units for investigation. There is a lack of coherence as to the definition of some more compound units, such as communities (Underwood 1986), and a lack of predictive power of many sub-organismal variables (Underwood and Peterson 1988). In principle, however, any univariate measurements are analyzable in the ways suggested here. There may be multivariate analogues of these procedures, but they have not yet appeared in the relevant literature. There are procedures for simpler situations (Warwick 1986, Clarke and Green 1988, Warwick and Clarke 1991), but they cannot handle the temporal patterns of change described here.

There are many practical problems of detection of human influences on abundances of populations, but two are paramount in designing sampling programs. First is the large temporal variance of many populations, so that their abundances are very "noisy." Second, many populations show a marked lack of concordance in their temporal trajectories from one place to another. This results in considerable statistical interaction between changes in mean abundance from time to time and differences from place to place.

Sampling must therefore be sufficient to identify unusual patterns of change in a very interactive and very variable measurement. This is not a property of some of the most widely used procedures (see examples in Underwood 1991).

PROBLEMS WITH CURRENT SAMPLING DESIGNS

There are several problems with current, widely used environmental sampling designs (see also reviews in Underwood 1991, 1992). The principle on which they are based is Green's (1979) BACI (Before–After-Con-

[1] Manuscript received 16 June 1992; revised 28 February 1993; accepted 9 March 1993; final version received 9 April 1993.

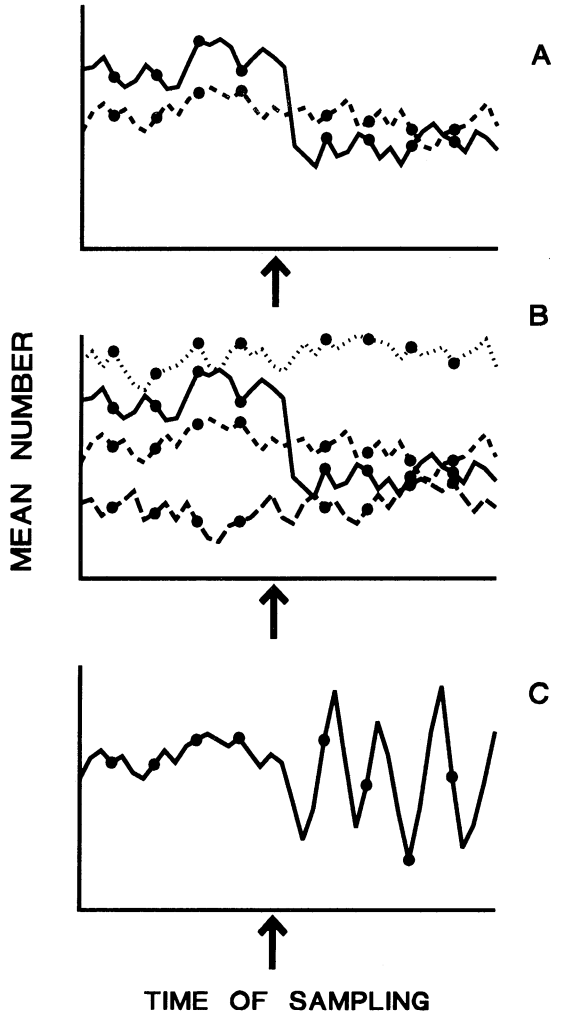

FIG. 1. Sampling to detect enviornmental impacts. (A) BACI design; replicated samples are taken at several times in a single Control (●– –●) and in the potentially Impacted location (●——●) Before and After a planned disturbance (at the time indicated by the arrow). (B) Sampling three Control locations to provide spatial replication. (C) An impact that has no effect on long-run mean abundance, but causes greater temporal variation.

trol–Impact) design. In this, a sample is taken Before and another sample taken After a possible impact, in each of the putatively disturbed (hereafter called the "Impact") and an undisturbed "Control" location. If there is an environmental disturbance that affects a population, it would appear as a statistical interaction between the difference in mean abundance of the sampled populations in the control and potentially impacted locations before the disturbance, and that difference after the disturbance.

As pointed out by Hurlbert (1984), this design is confounded. Any difference from before to after the potential disturbance may occur between two times of sampling, but may not be related to (caused by) the human activity. Thus, the design was extended by Bernstein and Zalinski (1983) and by Stewart-Oaten et al. (1986) to have several replicated times of sampling (Fig. 1A).

Stewart-Oaten et al. (1986) discussed the need for appropriate temporal replication, preferably in a non-regular frequency of sampling to avoid coincidences with natural cycles. They also considered, in detail, how to analyze the data, using t tests on the differences between two locations before and after the potential impact. All details of the rationale and procedure are available in their paper.

While this solved some of the problems of lack of temporal replication, it does not solve problems caused by the lack of spatial replication. The comparison between a single Impact and a single Control location is still confounded by any other cause of different time courses of abundances in the two places that is not due to the identified human activity. It is not unusual for populations to have different temporal trajectories in different locations, and temporal interaction among places is not at all uncommon.

Stewart-Oaten et al. (1986) discussed some aspects of such interaction. They concluded that if there were a trend or temporal change in the difference between the two locations before the putative impact, the variable being sampled would not be usable for assessment of the impact. Obviously, if the difference in mean abundance between two locations already varied from time to time before an impact, with some consistent temporal pattern of change, it would be expected to differ after an impact. Therefore there is no defined null hypothesis for any test so that an impact could be detected. Because many populations can be expected to interact in their patterns of abundance from time to time and location to location, this is a serious restriction on the usefulness of the proposed sampling design.

The problem of confounding (or "pseudoreplication," Hurlbert 1984) caused by comparing abundances in two locations, one potentially disturbed and the other a control, should be overcome by having several replicated disturbed and several control locations. This is, after all, the solution to the same problem in any routine field experiment (Hurlbert 1984, Underwood 1986). Of course, there will rarely be replicated planned developments causing the possible disturbance to a population. There is, however, no reason not to have replicated Control locations.

Thus, a randomly chosen set of locations is needed with the appropriate features of physical characteristics, mix of species, abundance of the target species, etc., as previously dictated the choice of a single control location. Note that there is no need to attempt to choose places with identical characteristics or abundances of the population. Not only is this impractical, it is unnecessary. The set of locations chosen to serve as con-

trols must simply represent the range of habitats like the one that might be disturbed (the Impact location). This is not a conceptually more difficult chore than the choice of two similarly representative locations, one Impact and one Control.

Obviously, the control locations (as with a single control in BACI designs) must be a representative sample of places of the same general habitat as that in which the impact is expected. So, if an outfall is to be placed on a rocky headland with steep cliffs and fast local currents, controls must be placed in similar locations. It can (and has) been argued that selecting such controls may be very difficult or impossible. Indeed, it may. It is, however, equally difficult to select a single control location. Where there cannot be any control (because, for example, the potentially impacted area is very large or is unique), there would be no comparative study comparable to a BACI or an asymmetrical, multi-control design. This is not a new problem for the designs considered here.

There may, nevertheless, be cases where multiple controls are not available and BACI procedures are all that can be available. Often, replication of control locations is not a more difficult task than choosing one control. Clearheadedness about the appropriately relevant locations that might be chosen is an identically valuable commodity for both types of design.

Now, assume that there is a sample of locations, representing a population of locations in which the monitored species can be found. The variance among mean abundances of a species from location to location represents the variance among any set of such locations (i.e., the set being sampled is a sample of locations from a population of many such locations). Locations represent a random factor in the sampling design (Snedecor and Cochran 1967, Winer 1971, Underwood 1981). The only constraint on the set of locations chosen for sampling is that it must (obviously!) include the location that might be disturbed by a planned human activity (the Impact location).

As explained in Underwood (1992), an environmental impact affecting the abundance of a sampled population in the impact location must cause the temporal pattern of abundance in that location to differ from the range of patterns in the set of control locations. There will probably be differences from location to location in the patterns of abundance from time to time. Nevertheless, to have an effect, the disturbance must cause more change in the impact location than occurs at the control locations.

An impact must also be detectable as a different pattern of statistical interaction from Before to After it starts, between the Impact and Control locations than occurs among the control locations. This is illustrated in Fig. 1B and described in full in Underwood (1992). Asymmetrical analyses of variance of sets of data from such sampling will be described below. Use of several control locations and asymmetrical analyses also allows a solution to Stewart-Oaten et al.'s (1986) problem of not being able to detect impacts in populations that have spatial and temporal interactions in their abundances. This will also be summarized below (and see Underwood 1992).

There are other problems with the BACI designs recommended by Bernstein and Zalinski (1983) and Stewart-Oaten et al. (1986). For their methods of analysis of the data, it is necessary for both locations (Control and Impact) to be sampled at the same time. This is not always possible because of logistic constraints, weather, time taken to sample, etc. Use of asymmetrical analyses of variance of the type recommended here can partially overcome this problem, but this is not considered further here. Full details are available in Underwood (1991, 1992).

Also, as illustrated in Underwood (1992), the temporally replicated BACI design assumes that the spatial scale of a putative environmental impact is known before it occurs. For example, a proposed outfall putting warm water into an estuary may have only a local effect on the surrounding few hundred square metres of mud flat. To detect this, appropriate control locations would be areas of a few hundred square metres elsewhere in the estuary. If, however, the estimated scale of impact is wrong and the outfall causes a change in abundance of some population over the whole estuary, such sampling will not detect it, because all the controls would also be affected.

To cover this possibility, control locations must also be examined in other estuaries along the coast, independent of any possible effects of the warm water from this outfall. Now, the impact would be detectable as a difference (from before to after the outfall's construction) in the abundance in the disturbed as opposed to the control locations. Such an effect will not be matched by any change in the controls. If the smaller scale were correctly predicted, the impact would be detected among the locations in the single estuary where the outfall is constructed. If this is wrong, there will be a different change in the controls in that estuary from that occurring in control locations elsewhere.

Because the spatial scale of such a disturbance might not be well defined in advance, sampling at two scales (estuaries and sites within estuaries) would be necessary. The Bernstein and Zalinski (1983) and Stewart-Oaten et al. (1986) design cannot be extended to analyze such data. In contrast, the asymmetrical analyses of variance reviewed here can be modified for this. Complete examples are provided in Underwood (1992), where I also discussed other advantages for the demonstration of causality by sampling several spatial scales.

As a final problem, there is a whole class of environmental disturbances that cannot at all be detected by a BACI design. These are disturbances that do not affect long-run mean abundances of a population, but,

TABLE 1. Analyses of variance in sampling designs to detect environmental impacts.

a) BACI design; data are collected in two locations (Impact and Control) at t randomly chosen times Before and After a planned disturbance. n replicate samples are taken at each time in each location. Relevant F ratios are calculated from expected values of Mean Squares as in Underwood (1981). B and C are fixed factors; Times are a random factor, nested in either Before or After.

Source of variation	Degrees of freedom	F ratio vs.	df for F ratio
Before vs. After = B	1	$T(B)$	
Control vs. Impact = C	1	$T(B)$	
$B \times C$	1	$T(B) \times C$*	1, $2(t-1)$
Times (Before or After) = $T(B)$	$2(t-1)$	Residual	
$T(B) \times C$	$2(t-1)$	Residual*	$2(t-1)$, $4t(n-1)$
Residual	$4t(n-1)$		

b) Similar design, but there is a total of l locations sampled; locations are a random factor, otherwise details are as above. There is no formal test for comparing Before versus After. This is irrelevant because an impact must cause an interaction ($B \times L$ or $T(B) \times L$); see text for details.

Source of variation	Degrees of freedom	F ratio vs.	df for F ratio
Before vs. After = B	1	No test	
Among location = L	$l-1$	$T(B) \times L$	
$B \times L$	$l-1$	$T(B) \times L$	$(l-1)$, $2(t-1)(l-1)$
Times (Before of After) = $T(B)$	$2(t-1)$	$T(B) \times L$	
$T(B) \times L$	$2(t-1)(l-1)$	Residual	
Residual	$2lt(n-1)$		$2(t-1)(l-1)$, $2lt(n-1)$

* This test is identical to the t test recommended by Stewart-Oaten et al. (1986).

instead, alter the temporal pattern of variance of abundance (Fig. 1C). Analysis of appropriate sampling for these is not considered in detail in this review, having been described in full in a previous paper (Underwood 1991).

The final part of this discussion is a consideration of the ecological research programs necessary to replace specific regimes of sampling to detect particular environmental impacts. These are required to solve the usual problem of a lack of sufficient time to sample before a possible impact and the lack of statistical power in many sampling designs used to detect impacts.

Asymmetrical Sampling Design to Detect Environmental Impact

First, consider the BACI design advocated by Bernstein and Zalinski (1983) and Stewart-Oaten et al. (1986). The analysis of variance of this design is summarized in Table 1a. As described in Underwood (1991), the F ratio indicated in Table 1a is the t test recommended by these authors. Also in Table 1b is the same analysis, extended to compare abundances in more than two locations.

In Table 2, the asymmetrical analysis is described, using a modelled set of data illustrated in Fig. 2A. Now, the useful contrasts of the Impacted vs. Control locations and its interactions with time can be extracted from the variation among all locations and its interaction with time.

An environmental impact should now be evident, in the simplest case, as an interaction between the difference between the mean abundance in the Impacted location and that in the Control locations Before compared to After the disturbance began (i.e., Table 2: $B \times I$; see Fig. 3 and Table 3 for examples). Alternatively, if the impact is not sustained, nor sufficient to alter the mean abundance in the impacted location over all times of sampling after the disturbance, it should be detected in the pattern of statistical interaction between the times of sampling and the contrast of the Impacted and Control locations (i.e., Table 2: T(Aft) $\times$ I). This is explained in full in Underwood (1992) and illustrated below.

Thus, a difference is sought between the time-course in the putatively impacted location and that in the controls. Such a difference would indicate an unusual event affecting mean abundance of the population in the single disturbed location, at the time the disturbance began, compared with what occurred elsewhere in undisturbed controls. The impact will either be detected as a different pattern of interaction among the times of sampling or at the larger time scale of Before to After the disturbance.

The patterns of significance in such analyses, under different types of responses to disturbance are illustrated below.

Patterns in Analyses to Detect Environmental Impacts

It is informative to consider disturbances of two types—pulse and press (Bender et al. 1984). The former are not sustained; the disturbance is removed after a period, although effects may be longer term. The latter are sustained disturbances. In environmental distur-

TABLE 2. Asymmetrical analysis of variance of model data from a single Impact (I) and three Control locations, sampled at six times Before and six times After a disturbance (that causes no impact, see Fig. 2a). [a] and [b] represent repartitioned sources of variation to allow analysis of environmental impacts as specific interactions with time periods ($B \times I$ or T(Aft) $\times$ I; see *Asymmetrical sampling design* . . . for more details).

Source of variation	Degrees of freedom	Mean square	F ratio	F ratio vs.
Before vs. After = B	1	331.5		
Among locations = L	3	25114.4		
[a]Impact vs. Controls = I	1	3762.8		
[a]Among Controls = C	2	35790.2		
Times (B) = T (B)	10	542.0		
$B \times L$	3	375.0		
[a]$B \times I$	1	454.0	1.51	Resid.
[a]$B \times C$	2	335.5	1.12	Resid.
T (B) $\times$ L	30	465.3		
[a]Times (Before) $\times$ L	15	462.2		
[b]T (Bef) $\times$ I	5	515.6		
[b]T (Bef) $\times$ C	10	435.9		
[a]Times (After) $\times$ L	15	468.2		
[b]T (After) $\times$ I	5	497.3		
[b]T (After) $\times$ C	10	453.6	1.51	Resid.
Residual	192	300.0		

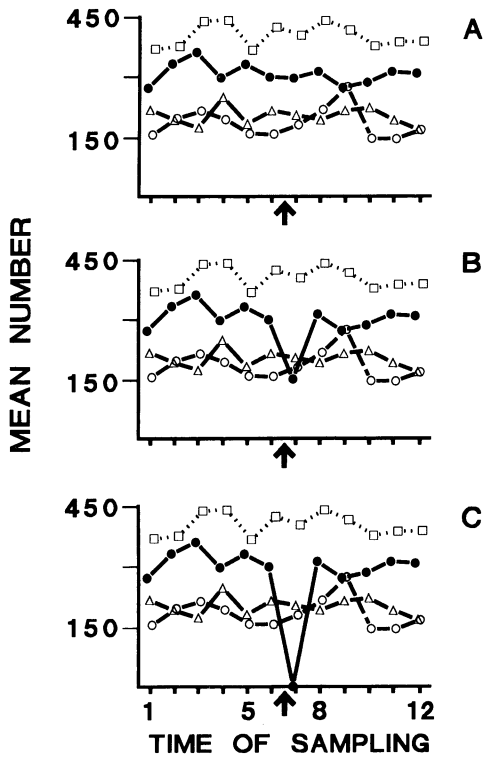

FIG. 2. Simulated environmental disturbances in one location (●——●), with three Controls, all sampled six times Before and After the disturbance (at the time indicated by the arrow). These data are analyzed in Table 3. (A) No effect of the disturbance; (B) a pulse reduction to 0.5 of the original mean; (C) a pulse reduction to 0.

bances, the construction of, say, a marina will cause pulse disturbances while it is being built. For example, there will be chemical and sedimentary changes to the habitat while building is done. This is a pulse disturbance—it ends when the marina is built. There is, however, also a press disturbance after the marina is finished, involving the altered water-flow around its walls, any hydrocarbon pollutants or antifouling leachates from the boats, etc. Either or both types of disturbance may have consequences for nearby populations of organisms.

There is some blurring of distinction between the two categories because they are not independent of the time scales of life histories of the organisms being affected (Underwood 1991), but they require different approaches, mechanisms, and interpretations (Bender et al. 1984).

Pulse effects

So, what happens in a location subject to a pulse disturbance that reduces the mean abundance of a population, which recovers very rapidly once the disturbance is removed? This situation is illustrated in Fig. 2B and C for two different magnitudes of response. In each case, only the population in the impact location was reduced.

The analysis of such data is modelled in Table 3, assuming that residual variances (i.e., among samples at each time and location) were not altered by the disturbance. To simulate the impact, the mean number of animals at the seventh time of sampling (the first after the impact) was reduced by the requisite proportion.

The effect on the analysis of the data is obvious in Table 3. There is now a significant interaction between the difference between Impact and Control locations and the times of sampling After the disturbance (T(Aft) $\times$ I in Table 3a and b). This is easily interpretable; the pulse caused a much more radical change from one time to another in the impacted location than in any of the controls. Thus, compared to the situation before the impact, there is now greater (and significant) interaction than before (T(Bef) $\times$ I in Table 3a and b was not significant), and there is no corresponding change among controls (T(Aft) $\times$ C is not significant). As explained in Underwood (1992), this test assumes no Type II error in pooling (Winer 1971, Underwood 1981) the non-significant variance component T(Aft) $\times$ C for testing for impact as T(Aft) $\times$ I.

Impacts affecting temporal variance

The second sort of impact is one that does not alter the mean abundance, but causes greater oscillations in

TABLE 3. Asymmetrical analyses to detect environmental impact; four locations are sampled, each with five random, independent replicates, at each of six times Before and six times After a putative impact starts in one location ("Impact"); there are three Control locations; "Locations" represents a random factor, and every Location is sampled at the same time;

Source of variation	df	(a) MS	(a) F	(b) MS	(b) F	(c) MS	(c) F	(d) MS	(d) F
Before vs. After = B	1	526.2		901.7		341.7		4212.3	
Among locations = L	3								
†Impact vs. Controls = I	1	2108.2		980.6		3593.9		1881.2	
†Among Controls = C	2	35790.2		35790.2		35790.2		35790.2	
$T(B)$	10	635.0		908.1		581.6		552.5	
B × L	3								
†$B \times I$	1	1125.6		2340.2		492.2		12625.6	26.55**
†$B \times C$	2	335.5		335.5		335.5		335.5	0.75
$T(B) \times L$	30								
†T(Bef) × L	15								
†T(Bef) × I	5	515.6		515.6		515.6		515.6	
†T(Bef) × C	10	435.9		435.9		435.9		435.9	
†T(Aft) × L	15								
†T(Aft) × I	5	1041.4	3.47**	2666.0	8.89**	1785.0	5.95**	435.3	1.45
†T(Aft) × C	10	453.6	1.51	453.6	1.51	453.6	1.51	453.6	1.51
Residual	192	300.0		300.0		300.0		300.0	

† Repartitioned sources of variation, Impacts can be detected by tests of B × I and/or T(Aft) × I (see *Patterns in analysis to detect . . . : Pulse effects* for details). Simulated disturbances are: (a), (b) pulses causing a brief reduction in mean abundance to 0.5 and 0, respectively; (c) temporal variance increased (standard deviation × 5) without altering mean abundance; (d), (e) press disturbance decreasing mean abundance to 0.5 or 0.2, respectively; (f) combination of altered temporal variance (standard deviation × 5) and press reducing mean to 0.5; (g) as (f), standard deviation × 0.5, press to 0.5.
* $P < .05$, ** $P < .01$.

numbers from time to time (Underwood 1991; Fig. 1C). In the first case, temporal variance was increased. This was modelled by calculating the deviation of the mean abundance in the Impact location at each time of sampling from the mean abundance in that location over all times of sampling After the disturbance. The deviations were then multiplied by the desired amount to alter the temporal standard deviation in that location. Results are illustrated in Fig. 3A and B and analyzed in Table 3c. Again, this has the effect of causing a significant interaction in the difference between the mean abundance in the impact and control locations and the time of sampling (T(Aft) × I in Table 3c). This is caused, again, by the greater fluctuations in that location than occurred before the disturbance or than occurred in the control locations after the disturbance.

Impacts on temporal variance can also cause decreased variation from time to time in the disturbed location, as shown in Fig. 3B. Here, there will again be a difference in the pattern of temporal interaction in the analyses, but the variance associated with such interactions will be smaller than in the control location and than occurred before the disturbance. This can still be identified in appropriate tests, which are described in Underwood (1992) and not dealt with in this summary.

Press disturbances

Press disturbances cause a different pattern in the data. If the press simply causes a sustained increase or decrease in mean abundance (as in Fig. 3C and D) it also causes an increase in the variation due to the interaction between the mean abundances in Impact and Control locations from Before to After the disturbance (Table 3d and e). This is, of course, the change detectable by the standard BACI design (Green 1979, Bernstein and Zalinski 1983, Stewart-Oaten et al. 1986, Underwood 1992). In the situations simulated here, the press disturbances were detectable as a significant pattern in the data ($B \times I$ was significant in Table 3d and e). Note that $B \times C$ was not significant and has been ignored in the analysis (see Underwood 1992 for details).

Combinations of effects

It is possible for a sustained series of disturbances to cause both a continuing impact, affecting numbers in the population and a simultaneous change in temporal variance. For example, an outfall pipe may cause overall reductions in survival of a population of fish, reducing mean abundance. The discharges from the outfall may, however, be intermittent and only coincide occasionally with episodes of recruitment. If recruits arrive when the outfall is operating, abundance will be pushed to very small numbers. If, however, recruits arrive when the outfall is inactive, the numbers entering the population will cause its abundance to rise dramatically. Thus, around the reduced mean abundance there will be greater temporal variance than in control locations. Several examples are illustrated in Fig. 4.

As can be seen in Table 3, both effects were detected as significant interactions. The altered mean abundance after the disturbance caused an increased inter-

"Times" represents a random factor nested in each of Before or After.

(e)		(f)		(g)	
MS	F	MS	F	MS	F
9687.5		4277.2		4228.5	
10149.1		1953.2		1899.1	
35790.2		35790.2		35790.2	
560.8		580.9		531.2	
29370.3	63.65**	12825.0	11.40**	12675.3	21.14**
335.5	0.75	335.5	0.75	335.5	0.75
515.6		515.6		515.6	
435.9		435.9		435.9	
407.2	1.36	1734.2	5.78**	683.3	2.28*
453.6	1.51	453.6	1.51	453.6	1.51
300.0		300.0		300.0	

action between the difference from Impacted to Control locations Before, compared with After, the disturbance ($B \times I$ is significant in Table 3f and g). In addition, the altered temporal variation caused interactions between the impacted and control locations in their temporal changes after the disturbance (T(Aft) $\times$ I is significant in Table 3f and g).

Thus, the combination of different types of impact is detected as a combination of the patterns of interactions they cause. The analyses in Table 3f and g demonstrate combinations of the patterns of significance of press effects (Table 3d and e) and altered temporal variation (Table 3c).

This not only indicates the usefulness of this design of sampling, but also shows that considerable information can be gained from the analysis about the nature of the biological responses to the disturbance. This would be of great benefit for predictions about effects of future proposed developments, prevention of their deleterious biological effects, or remedial action to reconstruct damaged ecosystems.

Temporally interactive populations

As discussed fully by Stewart-Oaten et al. (1986), in standard BACI procedures, data cannot be used for populations in which, before a disturbance, there is already an interaction among locations in the differences in abundance from time to time. The present designs can sometimes solve this problem, provided that impacts are large (because the tests are generally not likely to be powerful).

The logic of this procedure is described in full in Underwood (1992). If there exists, before the disturbance, a pattern of interaction among the locations, it will be detected as one or other, or both, of the F ratio tests on these interactions. Thus, in Table 3, T(Bef) $\times$ I, T(Bef) $\times$ C, or both, will be significant, depending on whether there is a different temporal pattern in the impact location from that in the controls, or whether there is, instead, or additionally, such a difference among the controls.

An environmental impact will result in a changed magnitude of interaction between the impact and control locations that is not matched in the controls. The interaction among times between the Control and Impact locations must alter relative to what occurred before the disturbance (T(Aft) $\times$ I must differ from T(Bef) $\times$ I in Table 3). Furthermore, there must be a difference in such interaction than now occurs among the controls (T(Aft) $\times$ I must be significantly larger than T(Aft) $\times$

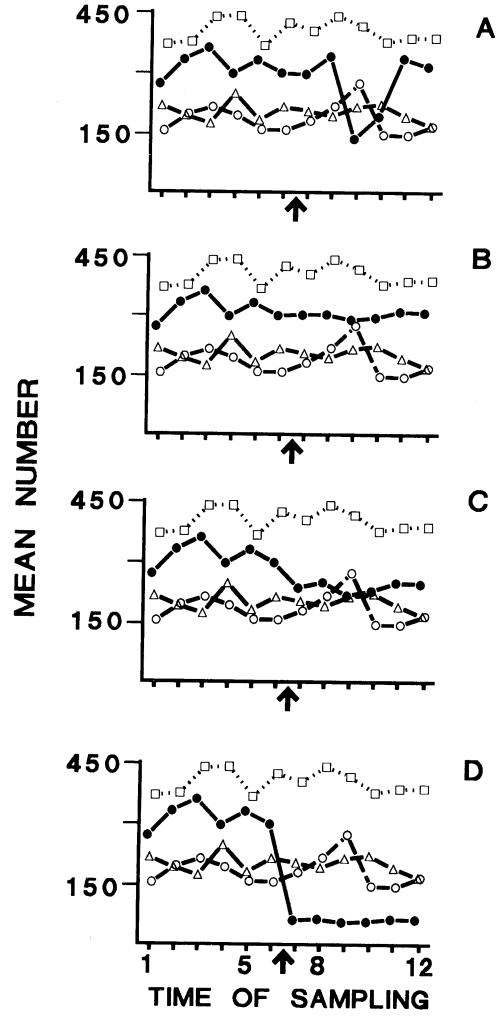

FIG. 3. Simulated environmental disturbances in one location (●——●), with three Controls, all sampled six times Before and After the disturbance (at the time indicated by the arrow). These data are analyzed in Table 3. (A, B) The impact is an alteration of temporal variance after the disturbance; temporal standard deviation $\times$ 5 in (A) and $\times$ 0.5 in (B). (C, D) A press reduction of abundance to 0.8 (C) and 0.2 (D) of the original mean.

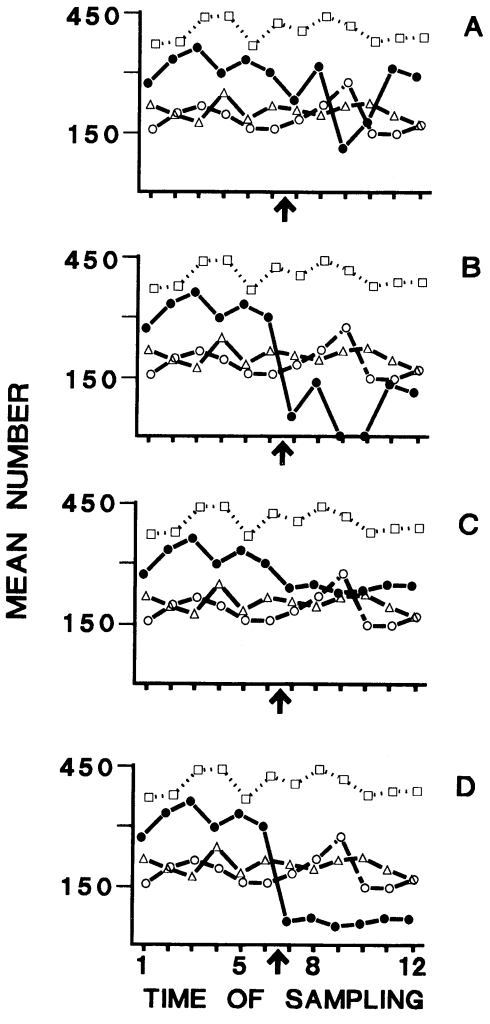

FIG. 4. Simulated environmental disturbances in one location (●——●), with three Controls, all sampled six times Before and After the disturbance (at the time indicated by the arrow). These data are analyzed in Table 3. In each case, disturbance causes a press response and, simultaneously, alters temporal variation as in Fig. 3. (A) temporal SD × 5, press to 0.8; (B) SD × 5, press to 0.2; (C) SD × 0.5, press to 0.8; (D) SD × 0.5, press to 0.2.

C in Table 3). Finally, in order to be sure that such patterns are not part of a more general environmental change, coincident with the disturbance, there must be similar amounts of interaction among controls After compared with Before the disturbance (T(Aft) × C and T(Bef) × C should be similar). Otherwise, the pattern identified as significant for the impact location is also present in the controls and cannot be identified as being due to the disturbance.

These tests are illustrated in Table 4 for the analyses in Table 3. Some of the statistical tests are two-tailed because a change in temporal interaction can result in a smaller or a larger mean square in an analysis of variance. Interactions among times of sampling After the disturbance can therefore be larger (if temporal variance increases) or smaller (if it decreases) relative to what occurred Before the disturbance (see also Underwood 1991).

The data in the simulations would have been identified as interactive if, for example, their residual variances had been much smaller. Many of the interactions before the disturbance would then have been significant. Under these circumstances, impacts would have been detected only for the most extreme conditions because the tests have few degrees of freedom and are not particularly powerful. Here, only a pulse to zero (a temporary extinction) would be unambiguously identifiable as an impact (only (b) in Table 4 has the appropriate pattern of significant results). This is, however, still a considerable advance on the lack of *any* test for such populations in the BACI procedure.

There is one other possible interactive structure in abundances of populations. That is an interaction over the longer time scale, from Before to After the disturbance, rather than among times before or among times after. If some general environmental change is happening while the disturbance occurs, the control locations may have different temporal patterns at a later time (i.e., coincidentally after the disturbance). There will then be an interaction among the locations in their differences after compared to before the disturbance (B × C will be significant in Table 3).

If this occurs, the only way the disturbance could have an effect is to cause a different pattern in the abundance in the impacted location. Thus, the interaction from before to after in the difference between the impact and control locations must differ from that among controls (and B × I will differ from B × C in analyses in Table 3). The F ratio test to detect this is shown in Table 4. Even though there are minimal degrees of freedom, in the simulated data there was virtually no longer term interaction among control locations (B × C in Table 3). The tests for press effects of disturbance that caused a major before-to-after interaction in the impacted location were still significant (Table 4d–g). Thus, even where there are few locations, if short-term interactions (among times of sampling) are absent, even non-parallel changes in abundance among locations do not prevent the detection of an environmental impact.

Power of tests

It is not realistic in this overview to consider the general properties of power of the tests (see Winer [1971] and Cohen [1977] for reviews of the principles). Nevertheless, it is useful to examine the particular sets of disturbances simulated here.

For each simulated disturbance described, I calculated the power of statistical tests to detect the effects. In each case I examined the altered variance (mean square, MS) associated with sources of variation af-

TABLE 4. *F* ratios for detecting environmental impacts when there are temporal interactions as described in full in Underwood (1992). Some tests are two-tailed because either mean square can be greater in response to different disturbances. MS and abbreviations are in analyses (a) to (g) in Table 3.

Test	df	Simulated disturbance (a)	(b)	(c)	(d)	(e)	(f)	(g)
1-tailed *F* ratios in analysis of variance in Table 3								
$B \times I$ vs. $B \times C$	1, 2	3.35	6.97	1.47	37.62*	87.54*	38.23*	37.78*
2-tailed *F* ratios of MS from analyses in Table 3								
T(Aft) $\times I$ vs. T(Aft) $\times C$	5, 10†	2.30	5.88**	3.93*	1.04	1.11	3.82*	1.51
T(Aft) $\times I$ vs. T(Bef) $\times I$	5, 5	2.02	5.17*	3.46	1.18	1.27	3.36	1.33
T(Aft) $\times C$ vs. T(Bef) $\times C$	10, 10	1.04	1.04	1.04	1.04	1.04	1.04	1.04

* $P < .05$, ** $P < .01$.
† Or 10, 5, depending on which MS is larger in the two-tailed test.

fected by the disturbance (as in Table 3). I then calculated the power of each *F* ratio test used (one-tailed tests in Table 3 and two-tailed tests in Table 4) to detect (as significant) that amount of variation. For example, a pulse to half the previous abundance caused the MS for the interaction between differences among times of sampling After the disturbance and the average mean difference between the Impact and Control locations (T(Aft) $\times I$ in Tables 2 and 3a) to change from 497.3 to 1041.4. The power of the *F* ratio to detect the latter (increased) MS is 0.95 (Table 5).

I calculated the power of all relevant tests on all simulated sets of data to gain some insight into types of environmental disturbance that were more likely to be detectable. Obviously, this is *not* an estimate of the power of the tests in real sets of data.

Power was calculated using the method for random sources of variation (variance components or Type II models) as described by Winer (1971). All tests were done with probability of Type I error (i.e., α) at .05, so power was calculated with this probability. I also calculated power with $\alpha = .10$, i.e., relaxing the stringency of tests and making detection of environmental impacts far less conservative. This was based on the premise that choice of $\alpha = .05$ is conventional, but not particularly well founded on any principle of the relative costs of Type I as opposed to Type II errors.

Considering first the tests for impacts in populations that do not have temporal interactions when undisturbed, the one-tailed *F* ratios described earlier (as in Table 3) were powerful for any large change to temporal variance or a very large pulse. Power was >0.90 in

TABLE 5. Power of tests used to detect environmental impacts. Calculations are from simulations as discussed in the text. Disturbances are (1) pulses causing a brief reduction (as specified) in mean abundance for the first time of sampling after the disturbance; (2) fluctuating, causing changes to temporal variance without altering the mean (multiplying temporal SD by the amount specified); (3) press disturbances causing sustained reduction (as specified) in mean abundance; (4) combinations of (2) and (3), a press disturbance also causing altered SD of temporal differences. Power is given if ≥0.50, for α (probability of Type I error) = .05 and .10.

Simulated disturbance	Analysis in Table 3	Data in Fig.	df	1-tailed *F* ratios: $B \times I$ / $T(B) \times I$ 1, 10		T(Aft) $\times I$ / Residual 5, 192		$B \times I$ / $B \times C$ 1, 2		2-tailed *F* ratios (Table 4): T(Aft) $\times I$ / T(Aft) $\times C$ 5, 10		T(Aft) $\times I$ / T(Bef) $\times I$ 5, 5	
			α	.05	.10	.05	.10	.05	.10	.05	.10	.05	.10
(1) Pulse to:													
0.5	(a)	2B		···	···	0.8	0.9	···	···	···	···	···	···
0	(b)	2C		···	···	0.95	0.95	···	···	0.7	0.8	0.5	0.6
(2) Temporal SD ×													
5	(c)	3A		···	···	0.95	0.95	···	···	0.5	0.6	···	0.5
2				···	···	···	···	···	···	···	···	···	···
(3) Press to:													
0.5	(d)	3C		0.6	0.7	···	···	0.5	0.6	···	···	···	···
0.2	(e)	3D		0.7	0.8	···	···	0.6	0.7	···	···	···	···
(4) Temporal SD ×/Press to:													
5/0.5	(f)			0.6	0.7	0.95	0.95	0.5	0.6	0.5	0.6	···	···
5/0.2		4B		0.8	0.9	0.9	0.9	0.6	0.7	···	···	···	···
2/0.5	(g)			0.6	0.7	0.7	0.8	0.5	0.6	···	···	···	···
2/0.2				0.8	0.9	0.7	0.8	0.6	0.7	···	···	···	···
0.5/0.5				0.6	0.7	···	···	0.5	0.6	···	···	···	···
0.5/0.2		4D		0.8	0.9	···	···	0.6	0.7	···	···	···	···

tests for such conditions (T(Aft) × I vs. Residual in Table 4). For smaller pulses (to 0.5), power remained at 0.80, for $\alpha = .05$.

Press effects were less likely to be detected. The power of the tests was in the range 0.6 to 0.7, even for an 80% (to 0.2) reduction of mean abundance ($B \times I/T(B) \times I$ in Table 5).

Press effects in combination with changes of temporal variance were generally more likely to be detected; power of the tests (T(Aft) × I/Residual in Table 5) was mostly somewhat larger for these disturbances (Table 5:(4)).

If the Residual variance had been smaller, making some of these tests more powerful, the interactions among times of sampling Before the disturbance would probably also be found to be significant. The populations would then be considered temporally interactive and the two-tailed F ratios in Table 4 would be relevant identifiers of impacts. The power of these was generally not great; they all had power in the range 0.5 to 0.8. Note that power of 0.5 implies that an impact, even of the magnitude of a sustained reduction to a fraction of the normal population's abundance, would only be detectable in one out of every two assessments.

As discussed below (see *Discussion: Improving designs to detect impacts*), the lack of power of programs of environmental sampling is a major problem. In contrast to the gloom, however, highly interactive populations are, by definition, characterized by large natural fluctuations from time to time and natural differences from place to place. Thus, to persist at all, they must probably be resilient and able to recover from nonanthropogenic disturbances, natural fluctuations in recruitments and mortality, etc.

In such populations, for human impact to be biologically meaningful, it must be very large, or it will not move the abundance out of its natural range of fluctuations. Therefore, in interactive populations, the only effects of human disturbances that are going to matter are very large ones. These sampling programs may well turn out to be powerful enough to detect such effects in real situations, but more work is needed to optimize the designs to maximize their power and efficiency.

Discussion

Scales of sampling

The sampling designs considered above are necessary for detecting environmental impacts because they include temporal and spatial replication to take into account natural variation. In many cases the spatial or temporal scale of the possible effects of a disturbance are unknowable before the disturbance happens. The choice of scales for sampling to detect the disturbance is then very difficult, if not impossible. Under these circumstances, hierarchical sampling of different spatial scales (Green and Hobson 1970, Underwood 1992) is appropriate and the asymmetrical designs can be extended to cover more than one spatial scale. I provided examples of such sampling in an earlier paper (Underwood 1992).

When time scales are uncertain, these designs are also extendible. Underwood (1991) illustrated their use for two different time scales to detect pulse and press disturbances that affect temporal variance, but not mean abundance. The use of hierarchical temporal sampling should probably be much more widespread in ecology (including environmental sampling) than is currently the case.

It is often assumed that differences seen from time to time in a monitoring program can be interpreted, for example, as indicating seasonal trends in mean abundance. Without knowledge of any fluctuations at shorter intervals, this is not a logically valid conclusion. The samples are confounded or pseudoreplicated (Hurlbert 1984) because there is only one set of replicates at one time in each season. This is illustrated in Fig. 5, where a single set of replicate samples, once in each of four seasons, reveals an apparent seasonal cycle of abundance. In fact, there is no cycle, but a slight general increase of abundance through time (Fig. 5C). If sampling were at two time scales, in this case three short intervals during each season, the lack of seasonal patterns of mean abundance would be revealed. The two time scales are then seasonal (as originally required) and shorter periods in each season. The alternative, of course, would be to demonstrate consistent seasonal patterns over several years (cycles). This would not only take longer, but would also be a different study. The time scale would be several years, which is presumably appropriate for some tests of hypotheses. It is, however, irrelevant for hypotheses about a particular (and appropriately defined) period, such as 1 yr.

Whatever the case, sampling at more than one temporal or spatial scale will be more expensive and time consuming, but may be mandatory. Cost–benefit procedures for optimizing the allocation of resources to each scale of sampling can be used where there are hierarchically structured spatial and/or temporal scales of sampling. Such procedures are well known (e.g., Cox 1958, Cochran 1963, Snedecor and Cochran 1967, Underwood 1981), and examples of their use are available in ecological field studies (e.g., Kennelly and Underwood 1985).

Improving designs to detect impacts

The power of environmental sampling to detect impacts is often poor, and there are serious deficiencies of capability of the designs discussed here to detect even quite large environmental disturbances (Table 5). Two conclusions are evident from this.

First, there must be more research on sampling procedures to optimize the designs and attempt to increase their power to detect unusual changes. There has been recent discussion on some of the related issues in relevant contexts (Andrew and Mapstone 1987, Clarke and Green 1988, Eberhardt and Thomas 1991). Some work is needed to determine the relative efficiency of increasing the number of times of sampling, number of control locations, and the number of replicate samples in each location at each time. There will obviously be different gains in precision for different types of effects. Note that only an increase in the number of Control locations will have any dramatic effect on detection of press disturbances in temporally interactive systems (Tables 4 and 5). A further consideration is rapid procedures for determining when temporally independent samples can be taken. If a population is sampled too frequently, the data from time to time are not independent, leading to major problems in any statistical analyses (e.g., Cochran 1947, Eisenhart 1947, Green 1979, Underwood 1981, Stewart-Oaten et al. 1986). Much more needs to be known about detection of non-independence and the frequencies with which temporally independent samples could be taken of various types of organisms with different life histories (see also Frank 1981, Connell and Sousa 1983, Keough and Butler 1983).

This is not as pessimistic as it seems. These designs are quite powerful for some types of disturbance, even with few control locations, times of sampling, and numbers of replicates (three, six, and five here). In the arbitrary case modelled, a number of disturbances would have been examined with power in excess of 0.90 (Table 5). This was also the case for real data in an extensive monitoring exercise as part of the CSIRO Jervis Bay Marine Environmental Projects (CSIRO 1991), where power to detect realistic sizes of disturbance was 0.90 or better for a number of populations monitored (A. J. Underwood, *unpublished data*).

Nevertheless, long time courses, particularly before a planned disturbance, are not the norm in environmental investigations. Logistical and financial constraints will continue to conspire to prevent adequate sampling unless the necessity for appropriately powerful sampling becomes a legal requirement of impact assessment. Alternative approaches are urgently needed, as discussed below.

Estimating spatial and temporal variances: removing the need for "Before" data

To overcome the problem of lack of time and resources for powerful sampling Before many impacts, research is needed to determine the rates of temporal change and the magnitudes of spatial differences for various populations. This could be done for key habitats in which repeated disturbances are planned. For example, there have been numerous disturbances of similar sorts in many habitats (such as those due to forestry, salmon fishing, agriculture, foreshore developments in mangrove forests, mussel and oyster farming, building marinas, boat ramps, outfalls from different types of industries, cooling-water outlets, etc.).

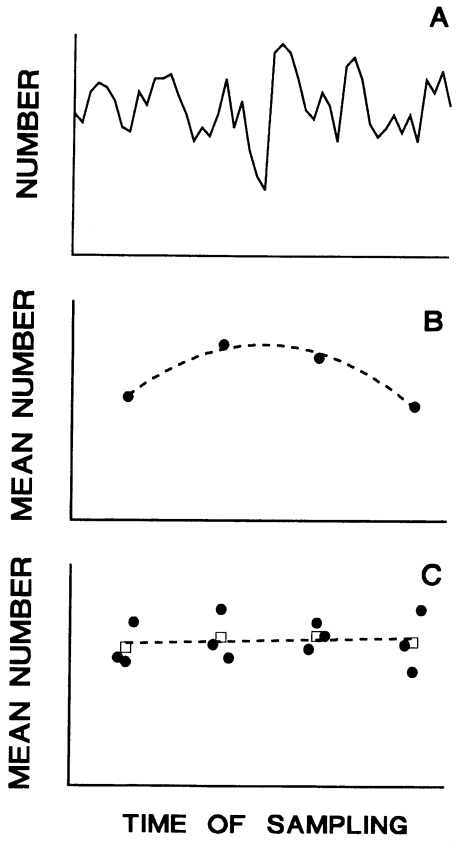

FIG. 5. Consequences of sampling without hierarchical temporal replication. (A) The real abundance, through time, of the sampled population; (B) samples, however well replicated, once in each season show an apparent seasonal trend; (C) sampling in each season is replicated at three times and there is (correctly) no cyclic trend in seasonal means.

Because these will recur repeatedly, it makes sense to choose some suitable species and to monitor now in a range of undisturbed, randomly chosen replicated habitats. If these are randomly chosen and the times of sampling are randomly assigned, the variance among times, locations, and their interactions are objective measures of the variances from a population of such locations and times. This is the essence of random factors in multifactorial analyses of variance (e.g., Scheffé 1959, Winer 1971, Underwood 1981, 1989). Thus, variances estimated in such a sampling program would serve as "before" data for any other set of locations at some subsequent time. They could be used to contrast against data collected "after" for a set of

locations chosen as controls and including one location in which a disturbance is occurring. This would obviate the need for specifically acquired "before" data and could be a very powerful set, based on numerous times of sampling and many locations (see also Underwood 1989).

Such a research program (sensu Lakatos 1974) needs to be implemented for those habitats and populations where there is reasonable expectation that a regime of sampling in some period of time will be representative of a subsequent period and for which there is an expectation of repeated, planned environmental assaults. Unless the latter is realistic, there is no point in collecting the information.

All of this begs the questions of which species (see, for example, Paine 1974, Underwood and Peterson 1988) and how to pay for the research. The former requires its own research. The latter requires a shift in emphasis of payment from developers. If this all works, the cost of individual Environmental Impact Statements (EISs) would be smaller, because of the reduced need for location- and time-specific "before" data. Also, such sampling would better protect developers from (and the prohibitive legal costs of defending) an accusation of causing environmental damage when none has occurred. While sampling continues to be in only two locations, many instances must arise where differences between the control and the putatively impacted location are not due, as claimed, to the identified human disturbance. Both should be political selling points for a change of strategy.

Experimental disturbances

The final new research program needed will be experimental impacts. As argued elsewhere (Underwood 1989), these should be of two kinds. One experiment entails the use of existing human disturbances. For example, Hilborn and Walters (1981) have made a convincing case for using existing Impacts as experiments. Far more can be done to evaluate what actually happened in places where human influences have occurred. This can be particularly powerful when repeated instances of similar disturbances have occurred (and can be used as replicates to contrast against a set of undisturbed Controls). Examination of such events will provide two types of information.

First, we would learn much more about the magnitudes (and directions) and temporal scales of changes caused by particular sorts of human disturbances. This would not only allow greater predictive accuracy for future planned development than is currently available from EIS procedures, but would also provide much of the information needed to optimize future environmental sampling in relation to future developments. Knowledge of the actual sizes of effects of disturbances would go a long way towards solving problems in the calculation of power in sampling designs (e.g., Cohen 1977, Peterman 1990).

Second, such studies will also provide evaluations of the worth of the predictions made before the development occurred. Thus, treating the development as an experiment will provide information to determine whether the prior EISs were realistic in their predictions. Or, we could discover whether they erred because of failure of understanding the biology of the system, inadequacy of the data gathered before the disturbance, incompetence on the part of the ecologists concerned, or a complete failure of the EIS process itself. Such environmental auditing can be useful (Buckley 1991), but needs to be more widespread and to make use of existing and previous disturbances as experiments (Hilborn and Walters 1981).

The other kind of experiment is a deliberate attempt to simulate the magnitudes of disturbances as they affect populations. Small-scale examples of these have proven very useful for estimating the effects of anthropogenic disturbances (e.g., McGuinness 1990) and are the bread and butter of much ecological research (e.g., Connell 1983, Schoener 1983, Hurlbert 1984, Hairston 1989, Peters 1991). What is needed is a shift of emphasis and perhaps of scale so that the experiments are recast to test explicit hypotheses about environmental disturbances. Some discussion of protocols, problems, and interpretations is in a previous paper (Underwood 1989) so I will not consider it further here.

Conclusion

Improved procedures for detection and interpretation of environmental impacts are needed. These must be coupled more tightly with logic and the sort of professionalism that field ecologists claim to use in more academic field experiments (although there are still too many problems for complacency about this; Underwood 1981, 1986, Hurlbert 1984). For too long, a perceived lack of rigor in environmental monitoring as opposed to "academic" ecological experimentation has been correlated with an argument that there is a difference between applied and pure science. It is important to get the debate back onto the distinction between good and bad science, regardless of its "purity" or its "application." Otherwise, decisions about environmental disturbances will continue to be dominated by (often) more expensive, more stochastic processes of law instead of being responsive to the successes of modern ecology.

Acknowledgments

The preparation of this paper was funded by the Australian Research Council, the Institute of Marine Ecology and the Research Grant of the University of Sydney. I am grateful for extensive discussions with and help from M. G. Chapman, K. R. Clarke, P. G. Fairweather, M. P. Lincoln-Smith, K. A. McGuinness, and C. H. Peterson and for the comments of two anonymous referees. Gee Chapman and Anne Under-

wood helped considerably with the preparation of the manuscript (twice!).

Literature Cited

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

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

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

Buckley, R. 1991. Auditing the precision and accuracy of environmental impact predictions in Australia. Environmental Monitoring and Assessment **18**:1–24.

Clarke, K. R., and R. H. Green. 1988. Statistical design and analysis for a "biological effects" study. Marine Ecology Progress Series **46**:213–226.

Cochran, W. G. 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics **3**:22–38.

———. 1963. Sampling techniques. John Wiley & Sons, New York, New York, USA.

Cohen, J. 1977. Statistical power analysis for the behavioural sciences. Academic Press, New York, New York, USA.

Connell, J. H. 1983. On the prevalence and relative importance of interspecific competition: evidence from field experiments. American Naturalist **122**:661–696.

Connell, J. H., and W. P. Sousa. 1983. On the evidence needed to judge ecological stability or persistence. American Naturalist **121**:789–824.

Cox, G. 1958. The planning of experiments. John Wiley & Sons, New York, New York, USA.

CSIRO. 1991. Jervis Bay baseline studies, fourth progress report. Commonwealth Scientific and Industrial Research Organization, Division of Fisheries, Hobart, Tasmania, Australia.

Eberhardt, L. L., and J. M. Thomas. 1991. Designing environmental field studies. Ecological Monographs **61**:53–73.

Eisenhart, C. 1947. The assumptions underlying the analysis of variance. Biometrics **3**:1–21.

Frank, P. W. 1981. A condition for a sessile strategy. American Naturalist **118**:288–290.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. Wiley Interscience, Chichester, England.

Green, R. H., and K. D. Hobson. 1970. Spatial and temporal structure in a temperate intertidal community, with special emphasis on *Gemma gemma* (Pelecypoda: Mollusca). Ecology **51**:999–1011.

Hairston, N. G. 1989. Ecological experiments: purpose, design, and execution. Cambridge University Press, Cambridge, England.

Hilborn, R., and C. J. Walters. 1981. Pitfalls of environmental baseline and process studies. Environmental Impact Assessment Review **2**:265–278.

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

Kennelly, S. J., and A. J. Underwood. 1985. Sampling small invertebrates on natural hard substrata in a sublittoral kelp forest. Journal of Experimental Marine Biology and Ecology **89**:55–68.

Keough, M. J., and A. J. Butler. 1983. Temporal changes in species number in an assemblage of sessile marine invertebrates. Journal of Biogeography **10**:317–330.

Lakatos, I. 1974. Falsification and the methodology of scientific research programmes. Pages 91–96 *in* I. Lakatos and A. E. Musgrave, editors. Criticism and the growth of knowledge. Cambridge University Press, Cambridge, England.

McGuinness, K. A. 1990. Effects of oil spills on macroinvertebrates of saltmarshes and mangrove forests in Botany Bay, New South Wales. Journal of Experimental Marine Biology and Ecology **142**:121–135.

Paine, R. T. 1974. Intertidal community structure: experimental studies on the relationship between a dominant competitor and its principal predator. Oecologia (Berlin) **15**:93–120.

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

Peters, R. H. 1991. A critique for ecology. Cambridge University Press, Cambridge, England.

Scheffé, H. 1959. The analysis of variance. John Wiley & Sons, New York, New York, USA.

Schoener, T. W. 1983. Field experiments on interspecific competition. American Naturalist **122**:240–285.

Snedecor, G. W., and W. G. Cochran. 1967. Statistical methods. Sixth edition. University of Iowa Press, Ames, Iowa, USA.

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

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

———. 1986. The analysis of competition by field experiments. Pages 240–268 *in* J. Kikkawa and D. J. Anderson, editors. Community ecology: pattern and process. Blackwells, Melbourne, Australia.

———. 1989. The analysis of stress in natural populations. Biological Journal of the Linnaean Society **37**:51–78.

———. 1991. Beyond BACI: experimental designs for detecting human environmental impacts on temporal variations in natural populations. Australian Journal of Marine and Freshwater Research **42**:569–587.

———. 1992. Beyond BACI: The detection of environmental impacts on populations in the real, but variable, world. Journal of Experimental Marine Biology and Ecology **161**:145–178.

Underwood, A. J., and C. H. Peterson. 1988. Towards an ecological framework for investigating pollution. Marine Ecology Progress Series **46**:227–234.

Warwick, R. M. 1986. A new method for detecting pollution effects on marine macrobenthic communities. Marine Biology **92**:557–562.

Warwick, R. M., and K. R. Clarke. 1991. A comparison of some methods for analysing changes in benthic community structure. Journal of the Marine Biological Association of the United Kingdom **71**:225–244.

Winer, B. J. 1971. Statistical principles in experimental design. Second edition. McGraw-Hill Kogakusha, Tokyo, Japan.