DETECTING ECOLOGICAL IMPACTS

Concepts and Applications in Coastal Habitats

Edited by

Russell J. Schmitt
Department of Biological Sciences
University of California
Santa Barbara, California

Craig W. Osenberg
Department of Zoology
University of Florida
Gainesville, Florida

Academic Press
San Diego New York Boston
London Sydney Tokyo Toronto

CHAPTER 6

DETECTION OF ENVIRONMENTAL IMPACTS

Natural Variability, Effect Size, and Power Analysis[1]

Craig W. Osenberg, Russell J. Schmitt, Sally J. Holbrook, Khalil E. Abu-Saba, and A. Russell Flegal

A principal challenge posed in field assessments of environmental impacts is to isolate the effect of interest from noise introduced by natural spatial and temporal variability. If the size of an impact from a human disturbance is small relative to natural variability, it will be difficult to detect with any degree of confidence. Therefore, it is critical to consider statistical power in planning and interpreting environmental impact assessment studies (Green 1989, Fairweather 1991, Faith et al. 1991, Osenberg et al. 1992a, Mapstone, Chapter 5; see also Peterman 1990, Cooper and Barmuta 1993). Consideration of power can also guide the selection of environmental parameters and sampling intensity. These are important design criteria because time and financial constraints typically limit the number of parameters that can be measured and the number of samples that can be collected.

Calculation of statistical power, which is the probability of rejecting the null hypothesis of "no effect" when it is false and a specified alternative is true, requires specification of the number of replicates as well as the ratio between the size of the "true" effect and the variability among the replicates (Cohen 1977). Because there are many assessment designs, each of which makes different assumptions about the meaning of "effect", "variability" and "replicate" (Green 1979, Stewart-Oaten et al. 1986, Eberhardt and Thomas 1991, Underwood 1991, Chapter 9, Osenberg and Schmitt, Chapter 1), the general assessment design must be specified before power can be discussed unambiguously. In assessing the environmental impacts of a particular anthropogenic activity, we typically

[1]This chapter was previously published in a modified form in Ecological Applications 4:16–30.
© 1994 by the Ecological Society of America.

require a design that explicitly deals with the lack of spatial replication and randomization (e.g., nuclear power plants are not replicated and placed at random sites along the U.S. coastline: Stewart-Oaten et al. 1986). The Before-After-Control-Impact Paired Series design (BACIPS: Stewart-Oaten et al. 1986, Schroeter et al. 1993, Stewart-Oaten, Chapter 7; see also Campbell and Stanley 1966, Eberhardt 1976, Skalski and McKenzie 1982, Bernstein and Zalinksi 1983, Carpenter et al. 1989) meets this criterion, and is the focus of our analyses and discussion.

In its simplest formulation, BACIPS requires simultaneous (*P*aired) sampling several times *B*efore and *A*fter the perturbation at a *C*ontrol and an *I*mpact site. The measure of interest is the difference (hereafter referred to as "delta", Δ) in a parameter value (in its raw or transformed state) between the Control and Impact sites as assessed on each sampling date (e.g., $\Delta_{Pi} = \log(C_{Pi}) - \log(I_{Pi})$, where C_{Pi} and I_{Pi} are estimates of the parameter at the Control and Impact sites on the i^{th} date of the Period P: i.e., Before or After). The average delta in the Before period is an estimate of the average spatial variation between the two sites, which provides an estimate of the expected delta that should exist in the After period in the absence of an environmental impact (i.e., the null hypothesis). The difference between the average Before and After deltas ($\Delta_{B\cdot} - \Delta_{A\cdot}$) provides an estimate of the magnitude of the environmental impact. Confidence in this estimate is determined by the variation in deltas (among sampling dates within a period, S_Δ), as well as the number of sampling dates (i.e., replicates) in each of the Before and After periods ($n_B + n_A = n$). For the purposes of this study, we define

$$\text{Effect Size} = \Delta_{B\cdot} - \Delta_{A\cdot} \quad (1)$$

$$\text{Variability} = S_\Delta = [\Sigma (\Delta_{Pi} - \Delta_{P\cdot})^2]^{1/2}/(n_P - 1)^{1/2} \quad (2)$$

$$\text{Standardized Effect Size} = |\Delta_{B\cdot} - \Delta_{A\cdot}|/(2S_\Delta) \quad (3)$$

We assume for convenience that variability (S_Δ), as well as sample size (n_P), are equal in the Before and After periods (but see Stewart-Oaten et al. 1992), and we double the standard deviation of deltas (S_Δ) in the denominator of Equation 3 based on the assumption that the resulting test will be two-tailed (Gill 1978).

Note that the standardized effect size (Equation 3), which consists of the components defined by Equations 1 and 2, expresses the effect size in standard deviation units and enters directly into conventional calculations of power (Cohen 1977). Note also that our terminology differs from that used by some authors (e.g., Cohen 1977): we use "effect size" to refer to the absolute magnitude of the effect and "standardized effect size" to refer to the standardized measure (which is often simply labelled "effect size").

Unlike other designs, the variability of interest, S_Δ, is not a simple measure of within-site sampling variability. Rather, it is a measure of the actual temporal variation in deltas, as well as within-site sampling error (which contributes to error in estimating the actual delta on any date). Figure 6.1 illustrates how this variability of deltas can be altered without any change in the average temporal

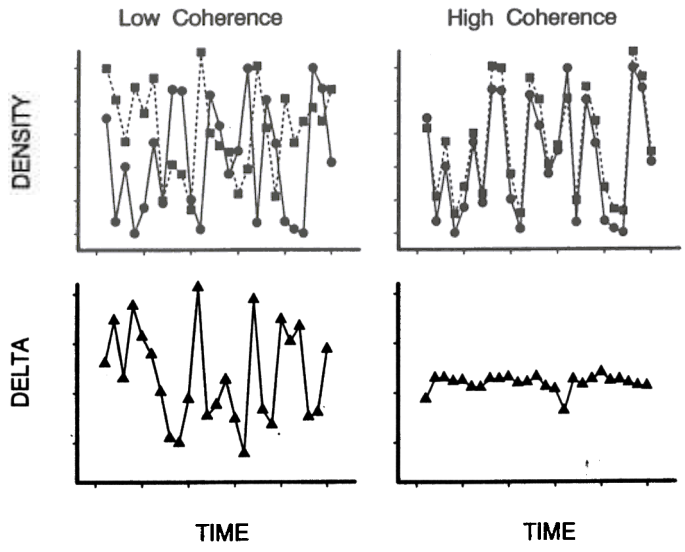

Figure 6.1. Patterns of spatial and temporal variation in population densities that lead to high and low variation in deltas. Simulated data (top panels) are from two pairs of sites. In both panels temporal variation in density (at a site) and the average difference between the sites are similar. The panels differ in the degree to which the estimated densities at the paired sites track one another through time. On the left, poor tracking (i.e., low coherence: Magnuson et al. 1990) leads to a low correlation between densities at the two sites ($r = -0.25$), while on the right, good tracking (i.e., high coherence) leads to a stronger correlation in densities ($r = 0.98$). The bottom graphs show the resulting differences in density (deltas). Low temporal coherence in densities (or any other parameter of interest) leads to high variability in deltas (i.e., relatively low power), while high coherence leads to low variability in deltas (i.e., relatively high power).

variability of a parameter (e.g., density), or in the amount of within-site sampling error. The critical feature in determining the variability among deltas is the extent to which estimates of parameters at the two sites track one another though time; Magnuson et al. (1990) refer to this as temporal coherence.

To aid in the planning of a BACIPS study, it would be helpful to find previous BACIPS studies conducted in a comparable situation (e.g., similar perturbation in a similar environment) and review the results for variability and effect size. This would permit estimation of the number of sampling dates needed to achieve a given level of power (e.g., Bernstein and Zalinski 1983) or a given amount of confidence in estimates of the effect size (e.g., Bence et al., Chapter 8, Stewart-Oaten, Chapter 2). For example, parameters with large standardized effect size (i.e., relatively large effect size and small variability) will yield more powerful assessments with fewer sampling events than parameters with low standardized effect size. Obtaining an adequate number of sampling events in the Before

period is crucial in a BACIPS assessment since once the perturbation begins, it is no longer possible to obtain additional Before samples. Unfortunately, there are few existing BACIPS studies that permit this type of analysis.

In the absence of this information, other data could be used to guide the design of BACIPS studies. Two types of non-BACIPS studies are more common and can offer insight. The first are long-term studies that document natural spatial and temporal variability, and therefore can provide estimates of S_Δ (Equation 2). The second are "After-only" (or Control-Impact) studies that assess impacts using a postimpact survey of sites that vary in proximity to the perturbation. After-only studies are a common type of field assessment approach, but they confound effects of the perturbation with natural spatial variability. Still, After-only studies can suggest the size of effects that might occur in response to a particular perturbation (Equation 1).

In this chapter, we illustrate how information from long-term studies and After-only studies can be combined to help plan BACIPS studies. We show how this information can be used to guide the selection of parameters and determine sampling schedules given constraints of time and funding. Our presentation consists of four analytical steps: (i) estimation of temporal variability of deltas using results from a long-term study; (ii) estimation of the likely magnitude of impacts using results from an After-only study; (iii) determination of the number of sampling dates required to detect the estimated impact given the background variability (at a specified level of power); and (iv) exploration of serial correlation, using the long-term data set, to assess the time necessary to achieve the required number of independent sampling dates. We contrast results for chemical–physical (e.g., chemical concentrations, sediment characteristics), individual-based biological (e.g., body size, growth), and population-based biological (e.g., density) parameters, and conclude there is a critical need to increase the use of individual-based parameters in field studies of environmental impacts.

Methods

Background

To help guide the planning of a BACIPS study of a particular planned intervention, it would be best to examine results of several preexisting BACIPS studies that examined impacts on many parameters in response to the same intervention in identical environments. Of course, such studies do not (and cannot) exist, but the congruence between this ideal and the realized match serves as a guide to the potential accuracy of the general guidelines that emerge.

The first step in this process is to define the intervention. To illustrate our approach, we focus on the nearshore discharge of an aqueous waste called produced water. Produced water is a complex wastewater generated from the production of oil and contains a variety of petroleum hydrocarbons, heavy metals,

and other potential pollutants (Middleditch 1984, Higashi et al. 1992). Although concerns have been raised about possible environmental effects of produced water in marine environments (Neff 1987, Neff et al. 1987, Osenberg et al. 1992b, Raimondi and Schmitt 1992), there have been no field assessments with sufficient Before data to allow separation of impacts from other sources of spatial and temporal variability (Carney 1987; also see Underwood 1991, Osenberg and Schmitt, Chapter 1).

We explore results from a long-term study of natural spatial and temporal variability and an After-only study to substitute for the absence of existing BACIPS studies. The two studies were both conducted in nearshore habitats along the coast of Santa Barbara County in southern California. The benthic environments are both dominated by soft-bottom habitats, and the studies used many of the same methods and quantified many of the same parameters. In each study, parameters had been selected based upon their perceived relevance to the impacts of produced water (e.g., Boesch and Rabalais 1987). (Because the long-term study is actually part of the "Before" sampling of a BACIPS study of produced water impacts, even the parameters examined in this study were selected with respect to produced water discharge.) However, these parameters, which include chemical, physical and biological characteristics (Table 6.1), are commonly measured in field assessments of other impacts in marine environments. We next review the two studies, the methods that were used and the parameters that were measured.

Natural Variability Assessed from a Long-Term Study

The two sites that comprise the long-term study are located approximately 1.6 km apart offshore of Gaviota, CA (ca. 34°27'29"N, 120°12'43'W) at a water depth of approximately 27 m. Various biological and chemical–physical parameters (Table 6.1) were sampled at the sites for periods ranging from 1.5 to just over 3 years beginning in February, 1988. For a given sampling date, a single value was obtained for each parameter at each site, and a delta was calculated as the difference between the log-transformed values:

$$\Delta_i = \log(X_{1i}) - \log(X_{2i}), \tag{4}$$

where X_{1i} and X_{2i} are the values of parameter X at each of the two sites (1 and 2) on the i^{th} date. Original parameter values were log-transformed to better satisfy assumptions of additivity required by BACIPS (Stewart-Oaten et al. 1986) and to facilitate comparison of deltas for parameters measured in different units (the transformed deltas are dimensionless). For each parameter, variability was quantified as the standard deviation of the deltas (S_Δ) calculated over all available sampling dates (Equation 2).

Population-Based Parameters. Densities of infaunal organisms were estimated approximately eight times per year. On each sampling date, 12 cores

Table 6.1. List of the Types of Parameters Used to Explore Natural Temporal Variability in Deltas (from the Long-Term Study) and to Obtain Estimates of Effect Size from an Existing Perturbation (from the After-Only Study)

Parameter type	Source: Long-term study (*Variability*)	Source: After-only study (*Effect size*)
Chemical–Physical		
Water temperature (No. depths)	2	2
Seston characteristics	3	0
Sediment quality	2	2
Sediment elements	11	9
Water column elements	12	8
Individual-based: Field collections		
Urchin size and condition	5	0
Cumacean body size	2	0
Individual-based: Transplants		
Mussel performance	(10)	12
Abalone performance	0	4
Population-based (No. taxa)		
Band transects	6	0
Infaunal cores	11	10
Quadrats	1	0
Emergence traps	4	0
Re-entry traps	3	0

Note: For each parameter type, we give the number of parameters quantified at each site (e.g., for infaunal density, 11 taxonomic groups yielded sufficient data for analysis in the long-term study). Details on parameters are given in the methods section. The ten estimates of variability for mussel performance, in parentheses, were collected as part of the After-only study but analyzed in the same manner as data from the long-term study.

(each 78 cm^2 × 10 cm deep) were collected. Samples were preserved in 10% buffered formalin and sieved through a 0.5 mm mesh sieve. Organisms were identified and counted from at least 4 of these cores per site per sampling date. Because this community is extremely speciose, with many species represented by only a few organisms or by zero counts on particular dates, and because zeros can cause difficulties in BACIPS analyses (Stewart-Oaten et al. 1986), infaunal organisms were grouped into broad taxonomic units, such as families and classes (see discussions on aggregation in Herman and Heip 1988, Warwick 1988, Frost et al. 1992, Carney, Chapter 15).

Numbers of infaunal organisms that migrated from the sediments into the overlying water (i.e., demersal zooplankton) were estimated using two emergence funnel traps (each covering a bottom area of 0.23 m^2) and three reentry traps (each 0.05 m^2 in area), which were deployed at both sites approximately eight times per year (for more detail on trap designs and function, see Alldredge and King 1980, Stretch 1983). Traps were set out for a 24-hr period. Following retrieval, contents were preserved, sieved through a 0.5 mm mesh sieve, and organisms were identified and counted as with the infaunal cores (Table 6.1).

Densities of larger epifaunal and demersal organisms (e.g., fish, sea stars, tube anemones) were estimated visually along band transects by divers. Two band transects (each 40 m by 1 m) were established along the 27 m isobath at both sites on each sampling date, and all large organisms within the transect were counted. Most were identified to species, although we grouped many of them into larger taxonomic units for these analyses. Due to their greater maximum density, white sea urchins (*Lytechinus anamesus*) were counted in 5 nonpermanent quadrats, each 1 m^2 in area, at both sites on all dates. Densities of urchins and other epifaunal and demersal organisms were estimated 8–12 times per year.

Individual-Based Parameters. The size (length of metasome) of two cumacean species was measured from samples obtained from the emergence traps. Other individual-based parameters (Table 6.1), were calculated from samples of the white sea urchin, *Lytechinus anamesus*: test diameter, gonad mass, somatic tissue mass, and gonadal-somatic index. In addition, the condition of urchins was estimated by calculating an adjusted mean for each site and date based on ANCOVA using each collection as a group, log(test diameter) as the covariate, and log(total tissue mass) as the response parameter. Urchins were sampled for these analyses 11 times during the study. As part of the After-only study, we also obtained estimates of variability for several other individual-based parameters derived from study of the mussel, *Mytilus californianus* (see below: *Combining Results on Effect Size and Natural Variability*).

Chemical–Physical Parameters. Chemical and physical parameters were examined that were thought to be indicative of the future plume's chemistry (e.g., elevated levels of certain heavy metals) or of the discharge's physical effects (e.g., altered sediment traits due to scouring of substrate or altered sedimentation rates and temperature due to local oceanographic effects) (Table 6.1). Seston flux was estimated by particulate accumulation in two sediment traps (5.1 cm diameter) that were filled with a mixture of seawater, formalin, and salt; the dense preservative remained in the sediment traps during the deployment and had an initial salinity of approximately 65 ppt and a formalin concentration of 5%. Sediment traps were deployed approximately 3 m above the sediments and retrieved by divers after 3–7 days. Traps were deployed approximately 8 times per year. Prior to analysis, large invertebrates were removed (aided by a dissecting microscope), following which the dry mass and ash free dry mass (AFDM)

of the particles were determined. Sedimentation rate was calculated as the mass of material (on a dry mass or AFDM basis) per cm^2d^{-1}. The percentage organic matter in the seston was estimated as the ratio of AFDM to dry mass.

Sediment grain size and percent organic matter were characterized from two sediment cores (20.3 cm^2/core, 5 cm deep) collected from both sites approximately 8 times per year. Sediment organic matter (SOM) was estimated based on combustion (for 4 hr at 450°C) of subsamples from one core. The fine sediment fraction (percent) was estimated from the other core as the percent (by dry mass) of the sample that passed through a 0.063 mm mesh sieve.

Water temperature was recorded approximately monthly at 3 m depth intervals. Here we use data for the 6 m and 21 m depths.

Surficial sediments (approximately the top 1 cm) were collected 4 times per year for analyses of trace and bulk elements. Three samples were collected at each site in acid-cleaned polyethylene containers by divers using trace metal clean sampling techniques. Any overlying water was decanted and samples were frozen. Sediments were later thawed and extractions performed by leaching 2 g sediment in 20 ml of 0.5 N HCl for 24 hr. The leachate was then filtered through a 0.45 μm mesh teflon filter using procedures reported previously (Oakden et al. 1984). This extraction is considered to be relatively selective for the biologically available concentrations of many metals, such as Pb, Cu, and Ag (Luoma et al. 1991). Leachates were analyzed for bulk elements (Al, Ca, Fe, Mg, Mn, P) and trace elements (Ba and Zn) by inductively coupled plasma-atomic emission spectrometry (ICP-AES). Other trace element (Cr, Cd, and Pb) concentrations were determined by graphite furnace atomic absorption spectrometry (GFAAS). Environment Canada reference sediments (BCSS-1, MESS-1, PACS-1) were analyzed concurrently to quantify the extraction efficiency for each element. All analyses were normalized to sediment dry weight.

Unfiltered water samples were collected two times per year from each site at two depths (surface and 21 m). The samples were extracted using the ammonium 1-pyrrolidinedithiocarbamate/diethylammonium diethyldithiocarbamate (APDC/DDC) extraction method described by Bruland et al. (1985). Trace element concentrations (Ag, Cd, Co, Cu, Fe, Ni, Pb, Zn) were measured by GFAAS. Procedural blanks were measured in each sample set. Each set of samples was analyzed in duplicate after a series of intercalibrations with Environment Canada reference seawater (CASS-1). These analyses were conducted concurrently with analyses of sea water from San Francisco Bay, and details of the procedural blanks and intercalibrations are provided in a report on those data (Flegal et al. 1991).

Effect Size Estimated from an After-Only Study

The After-only study was conducted at a produced water outfall located near Carpinteria, CA (34°23'10"N, 119°30'31"W) that was the subject of recent investigations of potential environmental impacts (Higashi et al. 1992, Krause et al. 1992, Osenberg et al. 1992b, Raimondi and Schmitt 1992, Raimondi and Reed, Chapter 10). The Carpinteria sites are approximately 50 km from the Gaviota sites. Although the two locations (Carpinteria and Gaviota) are both open coast, soft-bottom environments in the Santa Barbara Channel and have many species in common, the bottom depths sampled differed between the Carpinteria (11 m) and Gaviota (27 m) sites.

An intensive spatial survey of infauna was conducted along the 11 m isobath at the Carpinteria study area in 1990, approximately 12 years after produced water was first discharged at this location (Osenberg et al. 1992b). In a single survey, 20 sites were sampled along a spatial gradient from 2 to 1000 m upcoast (West) and downcoast (East) of the diffusers. Infaunal densities were estimated at each site by collecting eight cores (78 cm^2/core to a depth of 10 cm). These were processed as described for the long-term study, and a mean density was calculated for each taxon at each of the 20 Carpinteria sites.

All chemical–physical parameters examined as part of the long-term study at Gaviota were also estimated at the Carpinteria sites, except those related to seston quality and deposition and several elements. Methods were identical to those used at Gaviota (described above).

Individual-based biological data were obtained by transplanting individuals of known size and/or age to several of the sites. Mussels (*Mytilus californianus* and *M. edulis*) were transplanted to six sites to determine if proximity to the outfall influenced their individual growth and condition (Osenberg et al. 1992b). Forty individuals from a uniform size distribution (range 20–60 mm shell length) of a mussel species were put into a bag of 1.25 mm oyster netting, and one bag of *M. californianus* and one of *M. edulis* were attached to buoy lines approximately 3 m above the sediments. Mussels were retrieved and frozen after 3–4 months in the field. Final shell length, initial shell length, dry gonadal tissue mass, and somatic tissue mass were then measured for each mussel. Site-specific estimates of average gonadal condition (gonad mass at a given size), somatic condition, total condition, and gonadal-somatic index were obtained by running analyses of covariance (ANCOVA) for each parameter for each mussel species using log(final shell length) as the covariate. Average shell growth and tissue production were estimated using log(initial shell length) as the covariate. Adjusted means were obtained for each parameter at each of the 6 sites.

Abalone larvae were raised in the lab and transplanted in small flow-through cages to 6–8 sites located 5–1000 m from the diffuser (Raimondi and Schmitt 1992). Three measures of per capita settlement and metamorphosis were derived from transplants that lasted about four days: (i) the proportion of late-stage larvae that successfully settled in the field, (ii) the proportion of late-stage larvae that successfully metamorphosed in the field and (iii) the proportion of early-stage larvae that subsequently settled in the lab after addition of a chemical inducer (for details, see Raimondi and Schmitt 1992). An additional measure of individual performance was obtained from a short-term transplant: the proportion of early-stage larvae still swimming after 6 hr in the field.

To obtain estimates of the magnitude of impacts due to produced water, we calculated means (e.g., of density or performance) for three distance categories: Near (sites < 25 m of the diffuser), Far (25–200 m), and Control (> 200 m). We then calculated a near-field and far-field effect size as the difference between log(Mean Near or Mean Far) and log(Mean Control). This is equivalent to the impact size (expressed in log units) of a BACIPS study (Equation 1) assuming no natural spatial variation between the sites (i.e., $E(\Delta_B) = 0$). While this assumption cannot be tested without Before data (and is certainly false), available evidence suggests that natural spatial gradients are small relative to the impacts of produced water (Osenberg et al. 1992b, Raimondi and Schmitt 1992).

Combining Results on Effect Size and Natural Variability

For parameters that were common to both the After-only study and the long-term study, the standardized effect size was calculated as the ratio between the absolute value of the effect size, which was obtained from the After-only study, and twice the standard deviation of deltas, which was obtained from the long-term study: Equation 3. In some cases, however, the same parameters were not measured in both studies, and other steps were required before proceeding with the power analyses.

For example, there were four chemical–physical parameters that provided estimates of effect size but not variability. All four parameters were elemental concentrations (i.e., Cu in sediments and Co, Ag, and Pb in the water column), so we used the average standard deviation for other elements (in either the sediments or water column) in the calculation of the standardized effect size.

Conversely, there were chemical–physical and population-based parameters that provided estimates of variability but not effect size (i.e., parameters estimated from sediment traps, band transects, emergence traps, reentry traps, and quadrats in addition to several elemental concentrations: Table 6.1). For these parameters we calculated standardized effect sizes using the average effect size for similar parameters that were measured as part of the After-only study.

Estimating standardized effect sizes for individual-based parameters posed a more difficult analytical problem because the individual-based data from the long-term study were derived from field collections of organisms, whereas the transplants conducted in the After-only study used organisms of known size, or cohorts of known number and age. Therefore, the transplants removed several sources of potential variability present in estimates from the long-term study. Because mussels had been transplanted during four different periods (spread over a total of 14 months), we were able to obtain estimates of variability for the mussel parameters. The standard deviation of differences between log-transformed parameters measured at the 1000 m and 100 m sites was calculated for ten of the mussel parameters over the four periods. Because the 100 m site is probably influenced by the discharge of produced water (Osenberg et al. 1992b, Raimondi and Schmitt 1992), this approach will overestimate S_Δ if there is temporal variation in the effects of produced water.

Standardized effect size was then calculated as explained above using these new estimates of variability for all mussel parameters except tissue production (for which we had only one survey and therefore could not estimate S_Δ). The mean standard deviation of deltas for the mussel parameters was used to estimate the standardized effect sizes for mussel tissue production and abalone performance parameters, which lacked estimates of S_Δ. The standardized effect sizes for the individual-based parameters derived from the long-term study were calculated using the mean effect sizes based on the mussel and abalone transplants.

For each parameter, we estimated the sample size (total number of sampling dates in the Before and After periods) needed to have an 80% chance of detecting ($\alpha = 0.05$) an impact characterized by the parameter's standardized effect size. All power analyses were based on two-tailed *t*-tests as provided in Gill (1978). The number of sampling dates in the Before and After periods was assumed to be equal.

Serial Correlation

The power analyses yield the number of independent sampling events (i.e., dates) needed for a given level of power (e.g., 80%). The time scale over which those samples must be collected will depend on the amount of serial correlation in the time series of deltas for each parameter (Stewart-Oaten et al. 1986). Serial correlation can be directly incorporated into the analyses of BACIPS data (Stewart-Oaten et al. 1992, Stewart-Oaten, Chapter 7), but power is greatest when serial correlation is absent. Therefore, we tried to determine the most intensive sampling schedule that would avoid substantial amounts of serial correlation. By doing so, we could roughly translate the number of independent sampling events into an estimate of the minimum amount of time required by the BACIPS study.

Because rigorous analyses of serial correlation require long time series of data, and because the approach we outline here is imprecise to begin with (i.e., extrapolating from two different studies to the design of a future one), we used a simpler approach to provide a general guide to sampling frequency. For each parameter sampled as part of the long-term study, we examined the correlation between the delta measured on one sampling date (Δ_i) and the delta measured on the next date on which sampling for that parameter was conducted (Δ_{i+1}). Only parameters with data from 8 or more dates were included in the analyses.

Results

Natural Variability Assessed from a Long-Term Study

Data from the long-term study revealed that the variation in deltas (i.e., in the difference in parameter values between sites) was lowest for chemical-

physical parameters, intermediate for individual-based parameters, and greatest for population-based parameters (Figure 6.2). Most (28 of 30) of the chemical-physical parameters exhibited less variation in deltas than did the least variable population-based parameter. Almost all of the population-based parameters (24 of 25) were more variable than the most variable of the 7 individual-based parameters. Within a parameter group, no systematic differences were apparent among data collected using different techniques (e.g., densities based on infaunal cores versus band transects, or water column elements versus sediment elements), and there were no apparent trends among the population-based parameters related to the level of taxonomic aggregation (see Frost et al. 1992). All else being equal, these data suggest that chemical–physical parameters will provide more reliable indicators of environmental impacts than population-based parameters due to their smaller variability.

Effect Size Estimated from an After-Only Study

The After-only study provided estimates of effect sizes, which varied with proximity of the sampled sites to the produced water diffuser. In general, sizes of

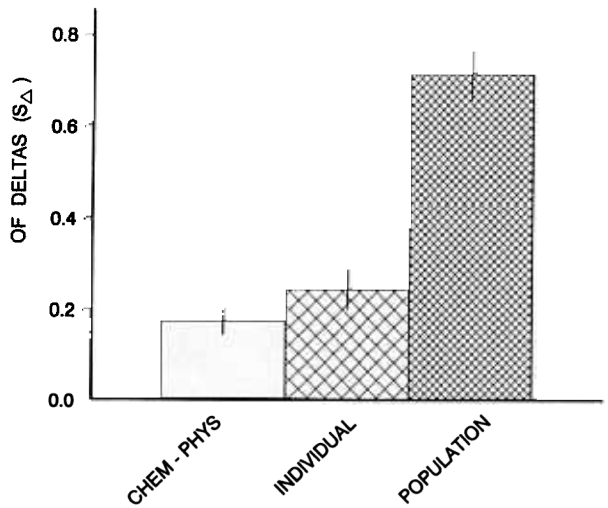

Figure 6.2. Temporal variability in estimates of the deltas (S_Δ) for chemical-physical, individual-based, and population-based parameters. Data were derived from the long-term study. For each parameter on each sampling date, a delta was estimated based on the difference between the log-transformed means at two sites [e.g., Log(mean density at Site 1 on date *i*) − Log(mean density at Site 2 on date *i*]. Shown are the standard deviations of deltas (mean ± SE) for parameters in each of the three groups. Means are based on 30, 7 and 25 different parameters for chemical-physical, individual, and population groups respectively. Here all individual-based data are derived from field collections.

effects were correlated (r = 0.62, n = 47) for sites near to and far from the diffuser (Figure 6.3), and the magnitudes of effects consistently were greatest nearer the diffuser. This pattern suggests that impacts diminished with distance away from the disturbance.

Both positive and negative changes in parameter values with distance from the diffuser were observed, and the sign depended on the particular parameter or parameter group examined. For example, concentrations of water column metals were higher nearer the diffuser, whereas most measures of individual performance were lower. Similarly, some taxa were more abundant closer to the diffuser, while others were less abundant. These two patterns in density probably reflect positive responses to organic enrichment (from oil constituents) and negative responses to toxicants present in produced water (e.g., Spies and DesMarais 1983, Osenberg et al. 1992b, Steichen 1994, see also Pearson and Rosenberg 1978, Ferris and Ferris 1979).

In evaluating power, the crucial factor is the absolute size of the change and not the sign (i.e., a positive or negative response). Although quite variable, the population-based parameters had absolute values of effect sizes that were about twice those for individual-based parameters, and four times larger than effect sizes for chemical-physical parameters (Figure 6.4). This pattern was similar for both Near and Far sites (r = 0.72, n = 47), although the overall magnitude of effects was lower at the Far sites (Figure 6.4). For simplicity, we focus on results from the Near sites in the following sections.

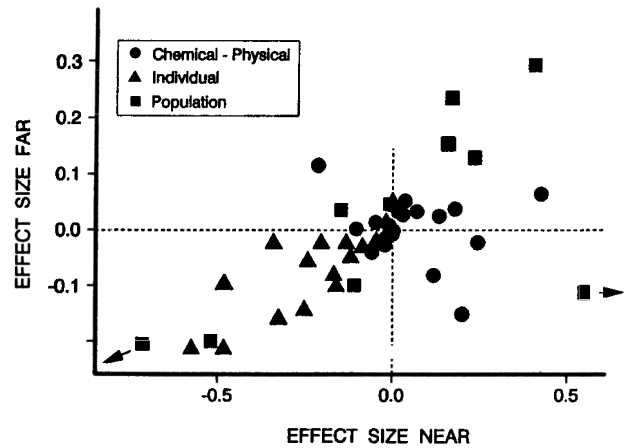

Figure 6.3. Effect sizes estimated from sites near and far from an operating produced water diffuser (an After-only study). Positive values indicate larger parameter values near (or far from) the diffuser relative to control sites, while negative values indicate the opposite. The two population-based parameters next to the arrows have effect sizes that are off the scale: (−0.92, −0.85) and (0.917, −0.13).

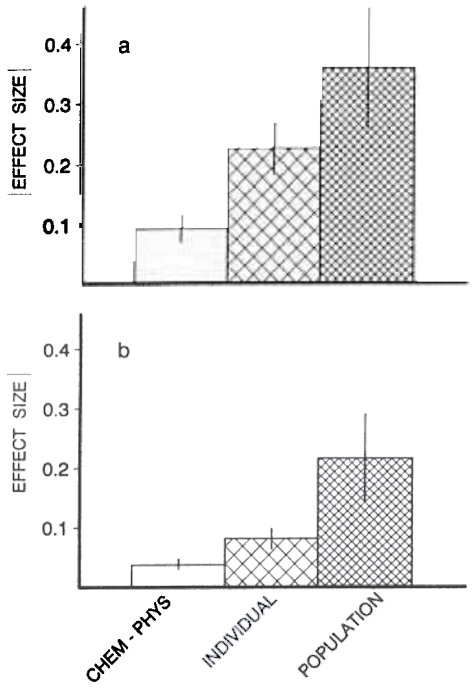

Figure 6.4. Absolute effect sizes (± SE) for chemical-physical, individual-based, and population-based parameters based on sites (a) near and (b) far from the diffuser. Sample sizes (number of parameters) were 21, 16 and 10 for the chemical-physical, individual, and population groups respectively.

Combining Results on Effect Size and Natural Variability

Estimates of natural variability in individual-based parameters were derived from field collections, whereas those for effect size were obtained from transplants. To make the estimates more comparable, we calculated variability of deltas for individual performance of mussels from four separate transplants in the After-only study. The results show that all ten indices of mussel performance were relatively invariable over time (Mean S_Δ = 0.080, S.E. = ± 0.20, Range 0.007 − 0.220). Indeed, most (70%) of these estimates of mussel performance were less variable than almost all (94%) of the parameters measured in the long-term study.

The results from the long-term study and the After-only study yielded the opposite conclusions about the power associated with different parameter groups. On one hand, the population-based (and individual-based) parameters should be the most powerful due to their larger average effect sizes (Figure 6.4),

whereas the chemical–physical (and individual-based) parameters should be more powerful due to their smaller average variability (Figure 6.2). Ultimately, the more powerful parameters will be those with the greatest standardized effect size (i.e., signal to noise ratio: Equation 3). Due to their relatively large effect sizes but low variability, individual-based parameters (particularly those derived from transplants) had larger standardized effect sizes than either the chemical-physical or population-based parameters (Figure 6.5). With respect to the individual-based parameters, the transplants yielded standardized effect sizes that were over three times larger than those derived from field collections.

The standardized effect sizes for both chemical-physical and population-based parameters were low and quite similar (Figure 6.5), due to the lower variability associated with chemical–physical parameters (Figure 6.2) and the greater effect sizes associated with population-based parameters (Figure 6.4). The standardized effect sizes for these two groups of parameters were 1/2 and 1/7 the magnitude of those for individual-based parameters derived from field collections and transplants respectively (Figure 6.5).

These results indicate that power to detect changes from exposure to produced water should be greatest for individual-based parameters derived from transplants, and next greatest for individual-based parameters obtained from field collections. For an equivalent number of estimates (i.e., sampling dates), power

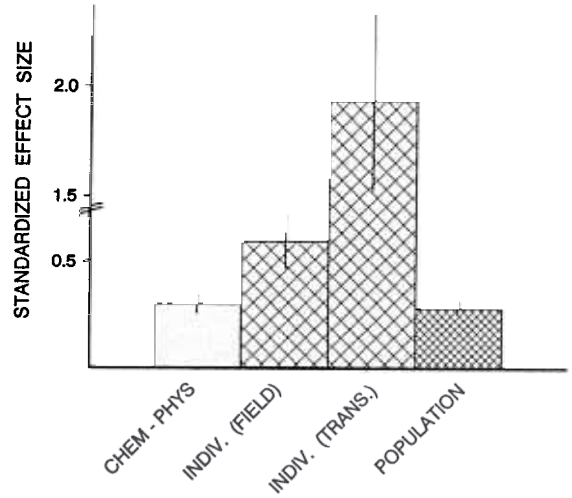

Figure 6.5. Standardized effect size (|Effect sizel / 2 × S_Δ) for each parameter group; the measure is the ratio of effect size to twice the standard deviation of delta. Shown are means (± SE), based on 34, 7, 16, and 26 parameters (from left to right). Individual-based parameters are divided into estimates derived from field collections and those derived from transplants of marked individuals or caged cohorts. Note the break in scale between 0.7 and 1.5.

should be considerably lower for chemical–physical and for population-based parameters. For example, based upon average standardized effect sizes (Figure 6.5) and a Type I error rate of 0.05, the numbers of independent sampling dates needed to achieve power of 80% are approximately four for individual-based parameters from transplants, 24 for individual-based parameters from field collections, 90 for chemical-physical parameters, and 95 for population-based parameters.

Most individual-based parameters required <20 (and typically < 10) sampling dates to achieve 80% power (Figure 6.6). Over half of the chemical–physical and population-based parameters required 100 or more sampling dates to reach 80% power (Figure 6.6). To provide an idea of how many parameters would have high power for a logistically reasonable number of surveys that would also permit model development and testing (Stewart-Oaten, Chapter 7), we determined the fraction of parameters in each group with a sufficiently large standardized effect size (> 0.52) to yield power of at least 80% with 30 sampling dates ($n_B = n_A = 15$). Using this guideline, 81% (13/16) of individual-based parameters from transplants and 43% (3/7) of those from field collections had power that exceeded 80%. By contrast, only 18% (6/34) of the chemical–physical and 4% (1/26) of the population-based parameters achieved this level of power from 30 surveys.

The preceding analyses were based on effect sizes estimated from sites near the produced water diffuser. Repeating the analyses using data from the Far sites yielded similar patterns, although as expected the overall power was much lower

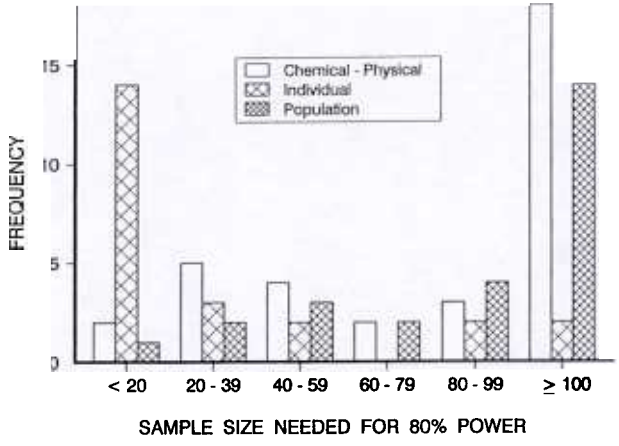

Figure 6.6. Frequency distribution of the sample size (number of independent sampling dates) for parameters in each group that is required for 80% power. Power analyses are based on standardized effect sizes (Figure 6.5).

or number of sampling dates needed for a given level of power was much higher. For example, the smaller effect sizes (estimated from sites far from the diffuser) resulted in more than half of the parameters in each of the three groups requiring >100 sampling dates to achieve 80% power. Only 26% of the individual-based parameters (all from transplants) required < 30 sampling dates, while none of the chemical-physical or population-based parameters achieved the same power with 30 dates.

Serial Correlation

Our analyses suggested that impacts on individual-based parameters are the most likely to be detected with a limited number of sampling dates. The analyses assumed that each sampling date provided an independent estimate of the true deltas (i.e., the underlying difference in parameter values between the Control and Impact sites). We examined patterns of serial correlation from the long-term study to gain insight into the frequency with which samples could be collected without grossly violating the assumption of temporal independence. This provided information on the time frame needed to collect series of independent samples.

There were no cases of significant ($P < 0.05$) negative serial correlation, and only 8% (4 of 50) of the parameters exhibited significant positive serial correlation (e.g., see Figure 6.7). Of the four parameters with positive serial correlation, two were chemical-physical parameters (seston sedimentation rate and seston percent organic matter), and two were population-based parameters (densities of sea pens and sea urchins: Figure 6.7c,d). None of the individual-based parameters exhibited significant serial correlation.

Serial correlation appeared to arise in the population-based parameters as a result of long-term trends in the deltas (Figures 6.7c,d). For example, the white sea urchin (*Lytechinus anamesus*) exhibited strong seasonal migrations, and was present during the winter and spring but absent during the summer and fall. The relative density at the two sites appeared to be set when urchins reappeared in winter; the ranking of the two sites were consistent within a year, but varied greatly among years (Figure 6.7c). This suggests that replicates should be collected only once per year, or a yearly average obtained from more frequent collections.

Density of sea pens (*Acanthoptilum* sp. and *Stylatula* sp.) exhibited an even longer term trend (Figure 6.7d). One site tended to have a greater density than the other site prior to October 1989, but the reverse was true for all samples collected after this date (Figure 6.7d). This could have arisen, for example, by a strong recruitment event in the Fall of 1989 at only one of the sites.

Despite these two examples, serial correlation was not a general problem for the various parameters estimated in our long-term study (e.g., Figure 6.7a,b). On average, the serial correlation for each of the three parameter groups was only

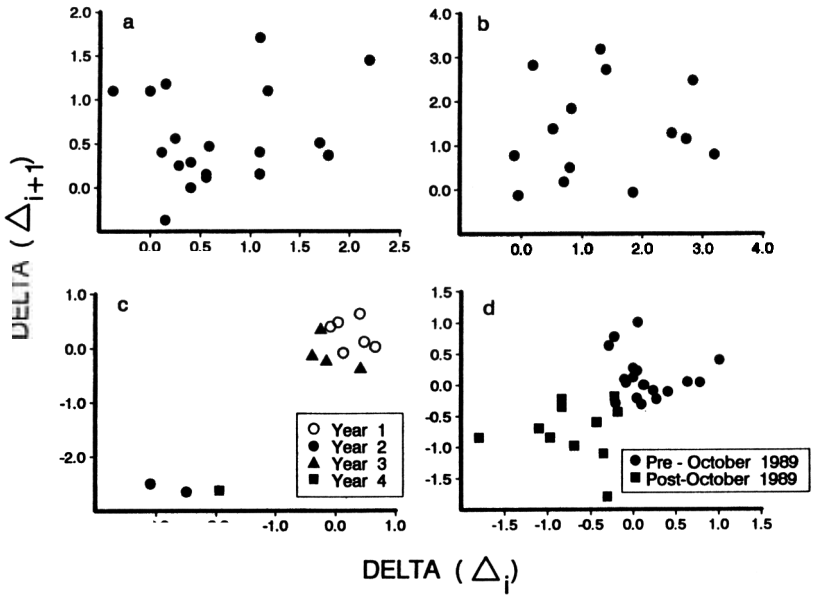

Figure 6.7. Patterns of serial correlation in deltas for four population-based parameters. These are the difference in density of: (a) cerianthid (burrowing) anemones (from band transect estimates); (b) copepods (from emergence traps); (c) white sea urchins (*Lytechinus anamesus*) (from quadrat samples); and (d) sea pen density (from band transects). There is significant serial correlation in (c) and (d), and data are separated into temporal groups to help distinguish the long-term patterns.

0.1–0.2 (Figure 6.8). Simulations suggest that serial correlation of this order introduce only small error into tests of impacts (Carpenter et al. 1989, Stewart-Oaten et al. 1992).

Based on these results, we assumed that sampling could occur every 60 days without yielding substantial amounts of serial correlation. Assuming that six samples are collected per year and the Before and After periods are of equal duration, the estimates of sample size (number of independent sampling events) can be translated into the number of years the assessment study must be conducted. Achieving 80% power would require 16 years for population-based parameters, 15 years for chemical–physical parameters, 4 years for individual-based parameters from field collections, and 1 year for individual-based parameters from transplants. To achieve 80% power for only a quarter of the parameters in each group, the required study duration is reduced to 11 years for population-based parameters, 7 years for chemical–physical parameters, 3 years for individual-based parameters from field collections, and <1 year for individual-based parameters from transplants.

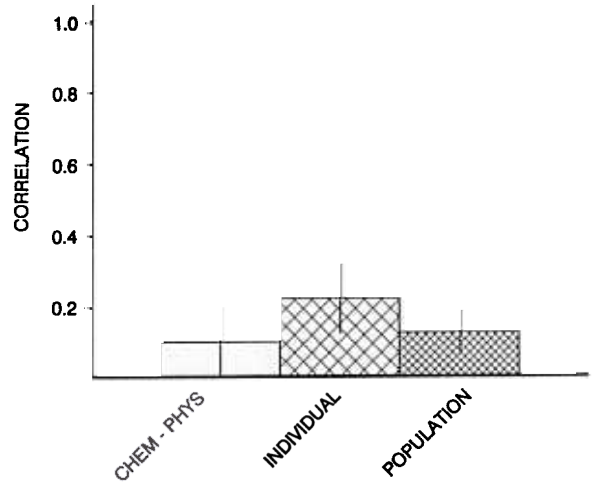

Figure 6.8. Degree of serial correlation in deltas for each parameter group. Shown are means (± SE), based on 18, 7, and 25 parameters for chemical-physical, individual and population groups respectively.

Discussion

Because relatively few well-designed studies of planned perturbations have been completed, there is a sparse empirical base to guide the design of future assessment programs (e.g., Carney 1987, Spies 1987, Underwood 1991, Stewart-Oaten, Chapter 7, Ambrose et al., Chapter 18). Recent discussions have highlighted general design considerations that should be incorporated in Before-After-Control-Impact approaches (e.g., Stewart-Oaten et al. 1986, 1992, Stewart-Oaten, Chapter 7, Underwood, Chapter 9), but these say little about specific considerations regarding sampling frequency and parameter selection. Often, a study must be planned in the absence of sufficient preliminary data to properly guide sampling decisions (Stewart-Oaten, Chapter 7). It is crucial to obtain good estimates of sampling variability and the size of impacts that might arise (or that are deemed ecologically important: Underwood and Peterson 1987, Yoccoz 1991), but this information typically is lacking. In the absence of a BACIPS (or analogous) study conducted previously on a similar perturbation in a similar habitat, it is vital that other existing data be used to guide specific design considerations.

Given limitation on time and funding, the selection of parameters and frequency of sampling are especially crucial features of the design process. One of the most acute constraints is the time available to collect data prior to the perturbation. In many situations, the Before period probably will be rather abbreviated for a variety of reasons beyond scientific control (Piltz, Chapter 16).

Therefore, parameter selection and sampling design should take into account the low numbers of temporal replicates that likely can be collected prior to the commencement of the disturbance (see Stewart-Oaten, Chapter 7 for discussion of model development based on these data). Key considerations in this regard are the likely variability in the parameter estimate (e.g., delta) and the probable magnitude of response to the disturbance, both of which influence statistical power to detect an effect. Constraints on the number of temporal replicates in the Before period are most likely to hamper detection of impacts on population density and chemical-physical characteristics, and least likely to affect detection of effects on individual performance. Unfortunately these results suggest that many field monitoring programs might be compromised because individual-based parameters rarely are examined (e.g., Carney 1987).

There are, however, compelling reasons to examine population and chemical-physical parameters despite the expected low power. First, chemical and physical properties describe the direct effect of many perturbations, and in many cases impacts could be ameliorated by subsequent intervention (e.g., source reduction, reduced discharge limits). Second, population attributes, such as density, reflect the ecological consequences of the disturbance, and are features of fundamental concern to resource managers and regulatory agencies. In addition, some species receive special regulatory consideration. Another reason is that, while the average power for population or chemical-physical parameters is low, some species or chemical-physical parameters will have greater power than others. The approach described here is equally useful in identifying promising candidates within a parameter group as it is in guiding allocation of effort among groups. Finally, the actual impacts of the new disturbance, of course, cannot be known *a priori*, and effects on populations and chemical-physical parameters certainly can be much larger (or variation much smaller) than anticipated based on extrapolations from other data sets.

It is useful to consider why the population and chemical-physical parameters had low and similar power, because low power arose for different reasons. Population parameters were highly responsive to produced water (i.e., larger impact), but exhibited much greater natural variability. In contrast, the chemical–physical parameters had much lower variability in deltas, but were not greatly altered by the discharge of produced water. It appears that these results generally will hold for other types of point source disturbances in the marine environment. Many chemical–physical parameters probably are influenced largely by large-scale oceanographic processes that similarly affect nearby sites. For example, certain chemical–physical attributes (e.g., sedimentation rate, water temperature, nutrient flux) are strongly associated with upwelling conditions, which is a region-wide phenomenon (e.g., Landry and Hickey 1989). In these situations, differences in these parameter values between Control and Impact sites (i.e., the deltas) will be similar through time (see also a related discussion in Magnuson et al. 1990, which discussed temporal coherence of chemical–physical and biological parameters in freshwater lakes).

The relatively small response of chemical–physical parameters we observed to the discharge of produced water is also consistent with recent analyses of the general effect of waste discharges on the distribution of trace elements in coastal waters. For example, massive discharges (10^9 L d^{-1}) of wastewaters in the Southern California Bight have had a negligible (< 1%) impact on concentration of cadmium in those waters (Sanudo-Wilhelmy and Flegal 1991). Similarly, Schmidt and Reimers (1991) found that, in the Santa Barbara Basin, the fraction of certain metals (Cd, Cu, Ni, Pb) from human sources that is deposited in sediments near municipal outfalls is quite small (< 1%) compared to the amount released. In both cases, natural inputs and physical mixing processes appeared to have reduced the contribution from human inputs to a small fraction of the background level. So for chemical–physical parameters, the large spatial scale of events that drive natural variation can lead to low variability in deltas, while other natural processes can greatly diminish the signal provided by anthropogenic perturbations.

Population density, by comparison, is known to be highly responsive to local conditions, and can exhibit considerably different temporal patterns among neighboring sites (e.g., Holbrook et al. 1990, Magnuson et al. 1990, Schmitt and Holbrook 1990). The high sensitivity to local conditions potentially can translate into strong local responses to natural phenomena (thus increasing S_Δ) as well as anthropogenic perturbations such as wastewater discharges (thus increasing effect size). Within-site sampling error also can contribute to the high variability as benthic populations are notoriously difficult to sample (Vezina 1988, Thrush et al., Chapter 4).

It is important to note that the variability reported here (e.g., Figure 6.2) is a measure of the variability (over time) in *estimates* of the *differences* between sites. This variability includes both the true temporal variation in deltas and variation due to sampling error within a site (which adds error to the estimation of delta on any date). The contribution of sampling error will be a function of spatial variability within a site and sampling intensity, and therefore will vary with the within-site sampling design. This suggests that the variation in deltas (S_Δ) for population-based parameters could be reduced by more intensive sampling on each date, rather than increasing the number of dates. However, partitioning of observed variation for the long-term data set revealed that the deltas for population-based parameters were more variable due both to sampling error (i.e., high within-site spatial variation) and site-specific temporal variability (i.e., high variation in the actual deltas through time) (Osenberg, personal observation); increasing the sampling intensity within a date would reduce the observed variation (S_Δ) by only about 50%. Therefore, even if sampling error were removed (e.g., through more exhaustive sampling), population-based parameters still would be more variable than the chemical-physical or individual-based parameters (see Figure 6.2).

Our estimates of S_Δ probably are typical because the within-site sampling design of our long-term study is similar to that used in many assessment studies

(see Thrush et al., Chapter 4). The costs and benefits of adjusting within-site sampling intensity to achieve greater power can be analyzed (e.g., the importance of within-site accuracy versus more sampling dates), although with limited resources, greater precision ultimately would be accomplished at the cost of fewer sampling dates (which is the unit of replication in a BACIPS design).

Difficulty in sampling populations or other parameters within sites not only can affect the variance of the estimate, it also might lead to overestimation of effect sizes from After-only studies (Figures 6.3, and 6.4). This would be true especially for a parameter that is not affected by the perturbation, and thus should have an effect size of 0. Our approach would overestimate this effect by confounding sampling error and any underlying spatial gradient as an effect of the perturbation. If so, the calculated number of surveys (sample size) needed for a given level of power would be underestimated. While this bias will exist for any parameter, our data suggest that, on average, it will be most acute for population-based parameters. Hence, limitations on detecting impacts at the population level may be even more difficult than our analyses suggest.

In contrast to population and chemical–physical parameters, individual-based parameters had relatively high power owing to relatively low levels of variability (Figure 6.2) and intermediate effect sizes (Figure 6.4). Although 80% power could be achieved for many of the parameters we examined with fewer than 10 sampling dates, it is unwise to reduce the sampling intensity below a level at which model development and testing can be performed (Stewart-Oaten, Chapter 7). Our data also indicate that variability in the deltas for individual-based parameters can be reduced by use of transplants (Figure 6.4), which results in increased power (Figure 6.5). This presumably occurs because, compared with estimates from field collections, transplants remove noise introduced by individual variation as well as variability between sites over time. For example, size-specific growth rates can be assessed accurately using marked individuals of known size; because size can influence growth and size-distributions can vary among sites (e.g., Osenberg et al. 1988), an analysis based on marked individuals is likely to be more powerful than one based on field collections.

It should be noted that several of the transplant-derived parameters for which we had relatively high power are closely related to population-based parameters, which had much lower power. For example, transplants of abalone larvae provided estimates of per capita settlement rates. In the field, natural rates of per capita settlement can be estimated from observed settlement rates and/or larval supply, both of which require estimation of density (e.g., Olson 1985; Keough 1986; Victor 1986; Raimondi 1990). Therefore, these field estimates would have considerable error for the same reasons that population parameters were highly variable. The use of transplants surmounted much of this problem by using cohorts of known size, thereby eliminating much of the variability that plagues the population parameters.

The observation that individual-based parameters may yield more powerful assessments is troubling given the rarity with which they are measured in field

assessments. Care must be taken to guard against only considering parameters that yield low probabilities of demonstrable results (e.g., chemical–physical and population attributes); inclusion of individual-based parameters could greatly increase the sensitivity of assessment studies (Carney 1987, Osenberg et al. 1992a; see also Jones et al. 1991). However, the need to investigate individual-based parameters goes far beyond power considerations; it is the individual-based (and demographic) parameters that provide the mechanisms that underlie changes at the population (and therefore community) level. Furthermore, these individual-based parameters provide an explicit connection with detailed laboratory studies that focus on individuals and mechanisms of toxicity. What is needed are more realistic studies of individual-based effects under field conditions combined with both mechanistic laboratory studies and field assessments of population-level consequences. Recent advances with individual-based models (DeAngelis and Gross 1992) provide an explicit framework for making these fundamental linkages among environmental chemistry, physiology and population ecology (e.g., Hallam et al. 1990). Such models provide a powerful, mechanistic approach to assessing impacts on natural populations and complement the traditional approach of monitoring environmental impacts.

Acknowledgments

The assistance of D. Canestro is gratefully acknowledged. Also assisting in the field or laboratory were S. Anderson, M. Carr, M. Edwards, D. Forcucci, D. Heilprin, T. Herrlinger, B. Hoffman, T. Kaltenberg, P. Krause, S. Lin, A. Martinez, M. Perez, P. Raimondi, D. Reed, D. Steichen, K. Sydel, and V. Vrendenburg. G. Scelfo performed trace metal analysis of seawater samples, and Bonnie Williamson provided technical and logistical assistance. P. Raimondi provided the data on abalone larvae and A. Stewart-Oaten provided helpful discussions. We also appreciate the critical comments of A. Stewart-Oaten, J. Levinton and an anonymous reviewer on an earlier draft. This research was funded by the Minerals Management Service, U.S. Department of Interior under MMS Agreement No. 14-35-001-3071, and by the UC Coastal Toxicology Program. The views and conclusions in this chapter are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

References

Alldredge, A. L., and J. M. King. 1980. Effects of moonlight on the temporal migration patterns of demersal zooplankton. Journal of Experimental Marine Biology and Ecology **44**:133–156.

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

Boesch, D. F., and N. N. Rabalais, editors. 1987. Long-term environmental effects of offshore oil and gas development. Elsevier Applied Science, New York, New York.

Bruland, K. W., K. H. Coale, and L. Mart. 1985. Analysis of seawater for dissolved cadmium, copper and lead: an intercomparison of voltametric and atomic absorption methods. Marine Chemistry **17**:285–300.

Campbell, D. T., and J. C. Stanley. 1966. Experimental and quasi-experimental designs for research. Rand McNally, Chicago, Illinois.

Carney, R. S. 1987. A review of study designs for the detection of long-term environmental effects of offshore petroleum activities. Pages 651–696 in D. F. Boesch and N. N. Rabalais, editors. Long-term environmental effects of offshore oil and gas development. Elsevier, New York, New York.

Carpenter, S. R., T. M. Frost, D. Heinsey, and T. K. Kratz. 1989. Randomized intervention analysis and the interpretation of whole-ecosystem experiments. Ecology **70**:1142–1152.

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

Cooper, S. D., and L. A. Barmuta. 1993. Field experiments in biomonitoring. Pages 399–441 in D. M. Rosenberg and V. H. Resh, editors. Freshwater biomonitoring and benthic macroinvertebrates. Chapman and Hall, New York, New York.

DeAngelis, D. L., and J. Gross, editors. 1992. Individual-based models and approaches in ecology: populations, communities, and ecosystems. Chapman and Hall, New York, New York.

Eberhardt, L. L. 1976. Quantitative ecology and impact assessment. Journal of Environmental Management **4**:27–70.

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

Fairweather, P. G. 1991. Statistical power and design requirements for environmental monitoring. Australian Journal of Marine and Freshwater Research **42**:555–567.

Faith, D. P., C. L. Humphrey, and P. L. Dostine. 1991. Statistical power and BACI designs in biological monitoring: comparative evaluation of measures of community dissimilarity based on benthic macroinvertebrate communities in Rockhole Mine Creek, Northern Territory, Australia. Australian Journal of Marine and Freshwater Research **42**:589–602.

Ferris, V. R., and J. M. Ferris. 1979. Thread worms (Nematoda). Pages 1–33 in C. W. J. Hart and S. L. H. Fuller, editors. Pollution ecology of estuarine invertebrates. Academic Press, New York, New York.

Flegal, A. R., G. J. Smith, G. A. Gill, S. Sanudo-Wilhelmy, and L. C. D. Anderson. 1991. Dissolved trace element cycles in the San Francisco Bay estuary. Marine Chemistry **36**:329–363.

Frost, T. M., S. R. Carpenter, and T. K. Kratz. 1992. Choosing ecological indicators: effects of taxonomic aggregation on sensitivity to stress and natural variability. Pages 215–227 in D. H. McKenzie, D. E. Hyatt and V. J. McDonald, editors. Ecological indicators. Volume 1. Elsevier Applied Science, New York, New York.

Gill, J. L. 1978. Design and analysis of experiments in the animal and medical sciences. Volume 1–3. Iowa State University Press, Ames, Iowa.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. Wiley and Sons, New York, New York.

Green, R. H. 1989. Power analysis and practical strategies for environmental monitoring. Environmental Research **50**:195–205.

Hallam, T. G., R. R. Lassiter, J. Li, and W. McKinney. 1990. Toxicant-induced mortality in models of *Daphnia* populations. Environmental Toxicology and Chemistry **9**:597–621.

Herman, P. M. J., and C. Heip. 1988. On the use of meiofauna in ecological monitoring: who needs taxonomy? Marine Pollution Bulletin **19**:665–668.

Higashi, R. M., G. N. Cherr, C. A. Bergens, T. W.-M. Fan, and D. Crosby. 1992. Toxicant isolation from a produced water source in the Santa Barbara Channel. Pages 223–233 in J. P. Ray and F. R. Englehardt, editors. Produced water: technological/environmental issues and solutions. Plenum Press, New York, New York.

Holbrook, S. J., R. J. Schmitt, and R. F. Ambrose. 1990. Biogenic habitat structure and characteristics of temperate reef fish assemblages. Australian Journal of Ecology **15**:489–503.

Jones, M., C. Folt, and S. Guarda. 1991. Characterizing individual, population and community effects of sublethal levels of aquatic toxicants: an experimental case study using *Daphnia*. Freshwater Biology **26**:35–44.

Keough, M. J. 1986. The distribution of the bryozoan *Bugula neritina* on seagrass blades: settlement growth and mortality. Ecology **67**:846–857.

Krause, P. R., C. W. Osenberg, and R. J. Schmitt. 1992. Effects of produced water on early life stages of a sea urchin: state-specific responses and related expression. Pages 431–444 in J. P. Ray and F. R. Englehardt, editors. Produced water: technological/environmental issues and solutions. Plenum Press, New York, New York.

Landry, M. R., and B. M. Hickey, editors. 1989. Coastal oceanography of Washington amd Oregon. Elsevier Science Publishers, Amsterdam, The Netherlands.

Luoma, S. N., D. J. Cain, C. Brown, and E. V. Axtman. 1991. Trace metals in clams (*Macoma balthica*) and sediments at the Palo Alto mudflat in South San Francisco Bay: April, 1990 - April, 1991. U.S. Geological Survey Report 91-460. 47 pp.

Magnuson, J. J., B. J. Benson, and T. K. Kratz. 1990. Temporal coherence in the limnology of a suite of lakes in Wisconsin, U.S.A. Freshwater Biology **23**:145–159.

Middleditch, B. S. 1984. Ecological effects of produced water effluents from offshore oil and gas production platforms. Ocean Management **9**:191–316.

Neff, J. M. 1987. Biological effects of drilling fluids, drill cuttings and produced waters. Pages 469–538 in D. F. Boesch and N. N. Rabalais, editors. Long-term environmental effects of offshore oil and gas development. Elsevier Applied Science, New York, New York.

Neff, J. M., N. N. Rabalais, and D. F. Boesch. 1987. Offshore oil and gas development activities potentially causing long-term environmental effects. Pages 149–173 in D. F. Boesch and N. N. Rabalais, editors. Long-term environmental effects of offshore oil and gas development. Elsevier Applied Science, New York, New York.

Oakden, J. M., J. S. Oliver, and A. R. Flegal. 1984. Behavioral responses of a phoxocephalid amphipod to organic enrichment and trace metals in sediments. Marine Ecology Progress Series **14**:253–257.

Olson, R. R. 1985. The consequences of short-distance larval dispersal in a sessile marine invertebrate. Ecology **66**:30–39.

Osenberg, C. W., E. E. Werner, G. G. Mittelbach, and D. J. Hall. 1988. Growth patterns in bluegill(*Lepomis macrochirus*) and pumpkinseed(*L. gibbosus*) sunfish: environmental variation and the importance of ontogenetic niche shifts. Canadian Journal of Fisheries and Aquatic Sciences **45**:17–26.

Osenberg, C. W., S. J. Holbrook, and R. J. Schmitt. 1992a. Implications for the design of environmental impact studies. Pages 75–89 in P. M. Griffman and S. E. Yoder, editors. Perspectives on the marine environment of southern California. USC Sea Grant Program, Los Angeles, California.

Osenberg, C. W., R. J. Schmitt, S. J. Holbrook, and D. Canestro. 1992b. Spatial scale of ecological effects associated with an open coast discharge of produced water. Pages 387–402 in J. P. Ray and F. R. Englehardt, editors. Produced water: technological/environmental issues and solutions. Plenum Press, New York, New York.

Pearson, T. H., and R. Rosenberg. 1978. Macrobenthic succession in relation to organic enrichment and pollution of the marine environment. Oceanography and Marine Biology Annual Review **16**:229–311.

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

Raimondi, P. T. 1990. Patterns, mechanisms, consequences of variability in settlement and recruitment of an intertidal barnacle. Ecological Monographs **60**:283–309.

Raimondi, P. T., and R. J. Schmitt. 1992. Effects of produced water on settlement of larvae: field tests using red abalone. Pages 415–430 in J. P. Ray and F. R. Englehardt, editors. Produced water: technological/environmental issues and solutions. Plenum Press, New York, New York.

Sanudo-Wilhemy, S., and A. R. Flegal. 1991. Trace element distributions in coastal waters along the U.S.-Mexican boundary: relative contributions of natural processes vs. anthropogenic inputs. Marine Chemistry **33**:371–392.

Schmidt, H., and C. E. Reimers. 1991. The recent history of trace metal accumulation in the Santa Barbara basin, Southern California borderland. Estuarine, Coastal and Shelf Science **33**:485–500.

Schmitt, R. J., and S. J. Holbrook. 1990. Constrasting effects of giant kelp on dynamics of surfperch populations. Oecologia **84**:419–429.

Schroeter, S. C., J. D. Dixon, J. Kastendiek, R. O. Smith, and J. R. Bence. 1993. Detecting the ecological effects of environmental impacts: a case study of kelp forest invertebrates. Ecological Applications **3**:331–350.

Skalski, J. R., and D. H. McKenzie. 1982. A design for aquatic monitoring systems. Journal of Environmental Management **14**:237–251.

Spies, R. B. 1987. The biological effects of petroleum hydrocarbons in the sea: assessments from the field and microcosms. Pages 411–467 in D. F. Boesch and N. N. Rabalais, editors. Long-term enviromental effects of offshore oil and gas development. Elsevier Applied Science, New York, New York.

Spies, R. B., and D. J. DesMarais. 1983. Natural isotope study of trophic enrichment of marine benthic communities by petroleum seepage. Marine Biology **73**:67–71.

Steichen, D. S. 1994. The response of benthic and demersal macrofauna to organic enrichment at a natural oil seep. M.A. Thesis. University of California, Santa Barbara, California.

Stetch, J. J. 1983. Habitat selection and vertical migration of sand-dwelling demersal gammarid amphipods. Ph.D. Dissertation. University of California, Santa Barbara, California.

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

Stewart-Oaten, A., J. R. Bence, and C. W. Osenberg. 1992. Assessing effects of unreplicated perturbations: no simple solutions. Ecology **73**:1396–1404.

Underwood, A. J. 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.

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

Vezina, A. F. 1988. Sampling variance and the design of quantitative surveys of the marine benthos. Marine Biology **97**:151–156.

Victor, V. C. 1986. Larval settlement and juvenile mortality in a recruitment-limited coral reef fish population. Ecological Monographs **56**:145–160.

Warwick, R. M. 1988. The level of taxonomic discrimination required to detect pollution effects on marine benthic communities. Marine Pollution Bulletin **19**:259–268.

Yoccoz, N. G. 1991. Use, overuse and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America **72(2)**:106–111.