Impact Assessment Using the Before-After-Control-Impact (BACI) Model: Concerns and Comments

Eric P. Smith

Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

and

University Center for Environmental and Hazardous Materials Studies, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

and David R. Orvos[1] and John Cairns, Jr.

University Center for Environmental and Hazardous Materials Studies, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Smith, E. P., D. R. Orvos, and J. Cairns, Jr. 1993. Impact assessment using the before-after-control-impact (BACI) model: concerns and comments. Can. J. Fish. Aquat. Sci. 50: 627–637.

The effect of a change in an ecosystem can often be assessed through the use of a statistical model that incorporates the change. A sensible approach for assessing the effects of an industrial or power plant on the aquatic environment is to sample the environment both before and after the plant starts operation and test for a change in some biologically relevant parameter. To improve sensitivity, samples may be taken at a control site as well as at sites receiving the plant effluent. While this provides a powerful means for assessing effects, the implementation of the design is important and subsequent analysis of the collected data depends on proper implementation. Problems such as trends in the measurements, failure to meet the assumptions of the model, irregular sampling, confounding factors, and changes in the habitat can influence results, as we illustrate using a long-term impact assessment of a power plant on fish populations. In long-term studies, it may be difficult to separate effects due to the plant from effects due to other sources. Sound design requires both a good statistical model and an understanding of the underlying biological processes (what to measure) and careful planning (how to measure it well).

Les effets d'un changement survenant dans un écosystème peuvent souvent être évalués par l'utilisation d'un modèle statistique qui englobe ce changement. Une façon logique d'évaluer les effets d'une usine ou d'une installation de production d'énergie consiste à prélever des échantillons dans le milieu avant et après le début des opérations et à réaliser un test portant sur la variation d'un paramètre biologique pertinent. La sensibilité peut être accrue en prélevant des échantillons dans un site témoin et dans des sites récepteurs des effluents de l'usine. Cette façon de procéder est un très bon moyen d'évaluer les effets, mais la mise en oeuvre d'un tel protocole doit être faite de façon adéquate car elle influe sur les résultats de l'analyse des données recueillies. Nous illustrons les effets de certains problèmes, tels les tendances des mesures, l'impossibilité de se conformer aux hypothèses du modèle, l'irrégularité de l'échantillonnage, les facteurs à effets confondus et la modification de l'habitat, sur les résultats de par une évaluation des incidences à long terme d'une usine de production d'énergie sur les populations de poisson. Il peut s'avérer difficile, au cours des études à long terme, de distinguer entre les effets de l'usine et ceux d'autres sources. Une bonne conception suppose l'utilisation d'un modèle statistique approprié, la connaissance des processus biologiques sous-jacents (à mesurer) et une planification méticuleuse (comment mesurer adéquatement).

Received December 11, 1991
Accepted September 14, 1992
(JB330)

Reçu le 11 décembre 1991
Accepté le 14 septembre 1992

Measuring and assessing the impact of industrial operations on the environment are of great importance because recent changes in the legal and socioeconomic environment make determination of impact crucial to regulatory compliance and assessment of potential risk. Whenever a new industrial plant begins operation, the effect of the operation upon the ecosystem must be assessed. However, the design and analysis of ecological monitoring studies are not simple tasks unless the effects of the suspected impact are great and well defined. A number of articles have discussed the design of ecological impact assessments (Green 1979, 1989; Hurlbert 1984; Peterman 1990; Eberhardt and Thomas 1991).

The simplest design for impact assessment was presented in Green (1979). In this approach, potential ecological impacts are assessed by collecting data in a control and impact zone both before and after a potential impact begins. The impact could then be assessed using a two-factor analysis of variance (or multivariate extension for multivariate observations). Hurlbert (1984) has criticized ecological impact assessment methods (using as an example the above design) as not being sound from a statistical design view, but did not suggest alternatives. The problem that Hurlbert identified was one of lack of replication and the inability to randomize the impact to the sites. Stewart-Oaten et al. (1986), in a rejoinder to Hurlbert, suggested a

[1]Present address: Environmental Research Laboratory, SmithKline Beecham Corporation, 709 Swedeland Road, King of Prussia, PA 19406, USA.

TABLE 1. Data for the standard BACI analysis.

Treatment	Time	Control	Impact	Difference
Before	1	y_{b1c}	y_{b1I}	d_{b1}
	2	y_{b2c}	y_{b2I}	d_{12}
	·	·	·	·
	·	·	·	·
	·	·	·	·
	n	y_{bnc}	y_{bnI}	d_{bn}
After	$n + 1$	y_{a1c}	y_{b1I}	d_{a1}
	$n + 2$	y_{a2c}	y_{b2I}	d_{a2}
	·	·	·	·
	·	·	·	·
	·	·	·	·
	$n + m$	y_{amc}	y_{bmI}	d_{am}

design involving data on pairs of stations, one viewed as control and the other in the impact zone, collected both before and after the onset of the potential impact. This design differs from the simple two-factor design in that the control and impact sites are paired. The design is quite similar to the one used by Thomas et al. (1978), and both researchers used the design to assess effects of nuclear power plants.

This paper discusses the before-after-control-impact (BACI) design and describes some concerns that were realized during attempts to apply the design to assess the impact of a nuclear power plant on a river community. Some of the problems encountered that make the general task of assessing potential impact effects on organisms difficult are described. Recently, Stewart-Oaten et al. (1992) have discussed some potential problems with impact assessment. While their approach dealt primarily with the theoretical aspects of design and the resulting model, our approach focuses on a case study and specific problems encountered in assessments using fish in a river system. Some of the problems are specific to the BACI design; however, many are common to the more general problem of ecological impact assessment and even long-term ecological studies.

The BACI Approach

The BACI design (as described in Stewart-Oaten et al. 1986) requires data on two sites, corresponding to a control site and an impact site. Data are collected a number of times before the impact begins as well as after. Thus, there are two treatments: before-after, which is of primary interest, and control-impact, which is of secondary interest. Data for this analysis are summarized in Table 1. The sites are viewed as pairs and as forming a block in time. Data as described in Table 1 suggest analysis using a split-plot model (Milliken and Johnson 1984). The whole plots are the times of sampling and the split plots are the locations within the times. The model for the data is given by

$$y_{ijk} = \mu + \alpha_i + t_{ij} + \beta_k + (\alpha\beta)_{ik} + \epsilon_{ijk}$$

where μ is the overall mean, α_i is the effect associated with before-after ($i = b,a$), t_{ij} is the factor associated with time of sampling (block) ($j = 1,2, \ldots, n$ for $i = b$ and $j = 1,2, \ldots, m$ for $i = a$), β_k is the effect associated with control-impact ($k = c,I$), $(\alpha\beta)_{ik}$ is the interaction term, and ϵ_{ijk} is the error term. The analysis of variance table for this design is given in Table 2. Three tests are associated with this analysis, corresponding to two main effects (before-after and control-impact) and the interaction. The BACI test (when applied as a t-test, not a Mann–Whitney test) is equivalent to the test for an interaction effect (specifically, the BA × CI interaction). The above parametric approach is similar to a model used by Thomas et al. (1978) in the analysis of impacts at nuclear power plants.

TABLE 2. ANOVA table for BACI analysis.

Source	Sum of squares	Degrees of freedom	Mean square	F statistic
Before-after	SSBA	1	MSBA	MSBA/MSE1
Error 1	SSE1	$n + m - 2$	MSE1	
Control-impact	SSCI	1	MSCI	MSCI/MSE2
BA × CI	SSBACI	1	MSBACI	MSBACI/MSE2
Error 2	SSE2	$n + m - 2$	MSE2	
Total	SST	$2n + 2m - 1$		

Stewart-Oaten et al. (1986) suggested an alternative analysis that is analogous to the use of a paired t-test instead of a randomized block analysis when there are two treatment levels (e.g. see Kleinbaum et al. 1988, p. 392). Differences are formed between the control site and the impact site. These differences are then analyzed for a before-after treatment effect. Because the differences now appear as two samples of data (Table 1), the approach to the analysis is similar to that used for standard two-sample tests. The recommended approach is to use a two-sample t-test or a nonparametric test, such as the Wilcoxon rank sum or Mann–Whitney test, to assess change. If the two-sided t-test is used, the p-value is the same as that of the F-test for interaction (BA × CI effect in Table 2). The nonparametric approach is recommended to reduce the effects of extreme data values and failures in the assumptions that occur in most ecological data.

One potential problem with blocking designs is interaction between the blocking factor and the treatment factors, which is commonly referred to as nonadditivity. The recommended procedure also suggests transforming the data prior to differencing to achieve additivity. In the BACI analysis, additivity means that the differences in the control site are roughly parallel to those in the impact site, i.e. the shape of the plot of one site against time is similar to the plot of the other site except for a shift. The test is computed by testing the correlation between two series generated from the before period: one calculated as the sum of the measured values at the two sites at each sample time and the other calculated as the difference of these values.

Additivity is usually violated in one of two ways. First, the sites may differ in scale. A plot of the control site abundance versus time may resemble the impact site, but with more (or less) variability. Thus, one site may have both a larger mean and a larger variance. The second violation of additivity occurs when there is a trend in the observations in the before period for one of the sites or different trends in the observations at the different sites. These types of trends would suggest that the differences in the before period are not roughly constant but change with time. The first violation of additivity (scale) can often be removed by using a transformation. The second violation (trend) may not be removed using a transformation (especially the one that removes nonadditivity) and may lead to misleading significance (see fig. 3 in Stewart-Oaten et al. 1986). Stewart-Oaten et al. (1986) stated "Thus no attempt should be made to choose a transformation to correct this change. Rather, such correction should be avoided, the trend should be tested for, and the species not used in this analysis if the trend is present." If the same trend is present in control and impact sites, this would not violate the assumption of additivity, as this trend situation would imply parallel lines and con-

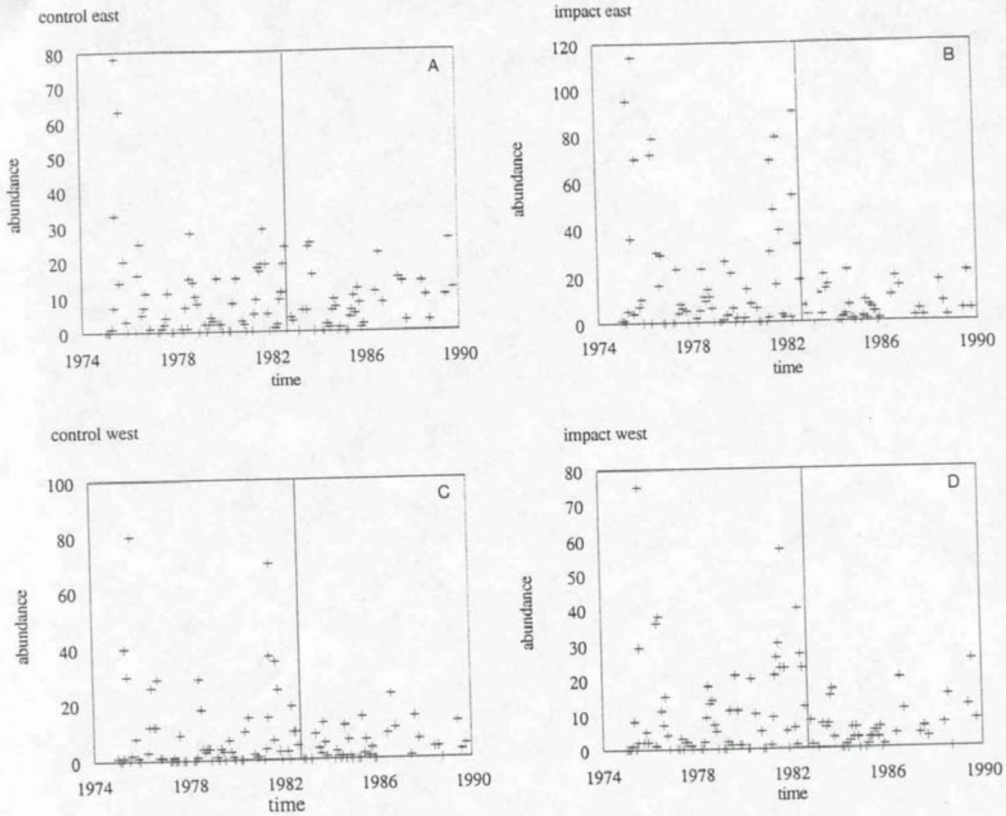

FIG. 1. Plot of the abundance of taxon 14 for (A) control east, (B) impact east, (C) control west, and (D) impact west.

stant differences between the lines. However, the presence of trend would have to be incorporated into the model to make accurate inferences.

To examine the utility of the BACI model, an actual case study is used in which data were collected both below and above a plant that discharged cooling water into a second-order river. Data were collected both before and after the plant began operation. The data set included information on a number of chemicals, phytoplankton, and benthic and fish taxa. In the fish study, data were collected using electrofishing at both impact and control sides of the river on 19 specific taxa; all remaining fish were classified into an unidentified group. Data were collected at two locations (east and west sides of the river) for almost 7 yr prior to plant operation and for 7 yr afterward (the start of the plant was in September 1982). Data from the two sampling sites at each location (east and west side of the river) initially were pooled to create one sample site at each location as suggested by Stewart-Oaten and his colleagues. Data for one of the 20 fish taxa are shown in Fig. 1 for each side of the river to illustrate some of the concerns. The data analyzed are the total catch from the electrofishing. Roughly equal effort (time and space) was expended for each collection, so the data were not adjusted for effort. The focus of this discussion is the analysis of the fish at this site and whether the plant is responsible for any change in fish abundance.

Table 3 presents the results of applying the BACI model to the fish species data. The additivity test based on Pearson's correlation coefficient was carried out on five transformations (none, square root, log, inverse square root, and inverse). A constant was added to the abundance prior to transformation (0.5 for the square root and 1.0 for the log, inverse square root, and inverse transformations). The transformation that gave the most nonsignificant p-value for additivity was used. After transformation, the Wilcoxon test was used on the differences to assess effect. Table 3 (the first column of tests) indicates that there are significant effects for five of the fish taxa. Thus, the preliminary conclusion is that the plant has a significant impact on the fish community ($\alpha = 0.05$). It may be the opinion of some readers that the testing should be multivariate or adjusted in some manner for the number of tests that are run, as a large

TABLE 3. Results of applying the BACI model to the 20 fish taxa. ***Significant at the 0.05% level (Wilcoxon test). Transformation used to correct for nonadditivity is given in parentheses following the p-value (S = square root, L = log, IS = inverse square root, I = inverse). Samples sizes are 70 for the before period and 41 for the after period. + + + indicates that none of the transformations was nonsignificant in the additivity test.

Taxon number	p-value using all the data	Sample sizes, excluding joint zeros	p-value excluding joint zeros
1	0.0037 (I)***	49, 21	0.0095 (I)***
2	0.7231 (IS)	62, 36	0.5383 (IS)
3	0.1493 (IS)	48, 26	0.1520 (IS)
4	0.0574 (IS)	70, 41	0.0574 (IS)
5	0.0051 (IS)***	63, 34	0.0014 (L)***
6	0.0367 (L)***	67, 40	0.0442 (S)***
7	0.5311 (I)	40, 9	0.0047 (I)***
8	0.4043 (I)	37, 28	0.2255 (I)
9	0.1244 (IS)	58, 37	0.0464 (L)***
10	0.0003 (IS)***	61, 37	0.0003 (S)***
11	0.9976 (S)	12, 11	0.9755 (L)
12	0.5971 (I)	29, 16	0.5693 (I)
13	0.2149 (I)	41, 28	0.3856 (I)
14	0.0201 (IS)***	62, 38	0.0197 (L)***
15	0.7716 (I)	26, 13	0.4839 (I)
16	+ + +	24, 28	0.0356 (I)***
17	+ + +	32, 16	0.8783 (I)
18	0.9659 (I)	32, 20	0.7349 (I)
19	0.7048 (IS)	69, 41	0.6969 (IS)
20 (unidentified)	0.1360 (L)	57, 41	0.0109 (IS)***

number of taxa are analyzed and the species abundances may not be independent. If a simple Bonferroni correction is used, significance would result when the p-value is less than 0.0025. Only one species (10) would be significant using this approach. One drawback to this approach is that the importance of the species is not accounted for. Species 14, for example, is a popular game fish. Further, if the number of taxa is large, the sensitivity of the inference may be compromised. The Bonferroni method is a conservative approach and, as such, results in industry being favored over the environment. Our preference is to consider the taxa on a more individual basis.

Concerns and Comments

While the BACI analysis represents an improvement over many other analyses that are or have been used, questions arise about its performance and the ability to assign cause to change in long-term biological impact assessments. During the course of analyzing the data from our case history, several concerns became apparent.

Zeros in the Data Set

The first difficulty encountered with the data and the model arose because of zeros in the data set. Zeros correspond to either an absence of the fish in the area at the time of sampling or an absence due to low sampling intensity. For example, the ability to catch fish using electrofishing at this particular site is different in winter than in summer. Furthermore, the winter months tend to have lower abundances, as the young fish of many species appear in spring. This may result in more zero data in winter than in summer. The difficulties in analyzing data with zeros are similar to the difficulties encountered when analyzing censored data (e.g. in chemical measurements). The estimates of parameters are affected, tests tend to have reduced power, and models do not adequately describe the data (see Helsel 1990 for a review of censoring and Lambert 1992 for a method for analyzing data with zeros). Figure 2 shows the differences (control-impact) for species 14; many of the differences are zero. The differences that are zero come from cases that were jointly zero and also cases where the abundances were small and of the same value (see also Fig. 1). Differences were quite large and negative during two intervals in the before data. This reflects sampling times where the abundance was large in the impact area but small in the control area. These large differences do not occur in the after period, however, and suggest an impact effect. Why the differences are so large is also of interest.

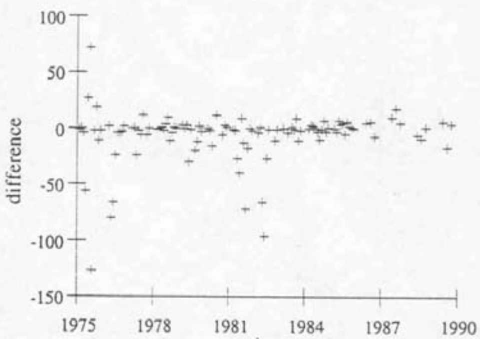

FIG. 2. Plot of the difference (control-impact) between control sites and impact sites for taxon 14.

One simple approach to dealing with zeros is to drop the sampling times in which zeros occur in both the control and impact sites. The results of applying the BACI model to fish data without joint zeros are given in the last column in Table 3. There are several more species that show significant changes, including the unidentified category. The increase in the number of significant species and a review of respective scatterplots indicate that the effect of the zeros is to increase the variance and make the means closer to zero. Some species have a large number of joint zeros (species 7, 11, and 15–18 in Table 3). Models to account for zeros are available for simple situations but present computational difficulties for models such as the split-plot model (Lambert 1992).

Additivity

A second concern with transformations is the relationship between the transformation to achieve additivity and the use of a nonparametric procedure to test for a plant effect. As biological field data tend to have nonnormal qualities, nonparametric methods are useful. Two concerns arise in their use. First, a philosophical problem is that the nonparametric test is applied because the researcher believes that a parametric procedure will not be valid; however, the test for additivity and choice of transformation depend on a parametric test. Second, the test for additivity, being based on a parametric analysis, will be sensitive to violations of some of the assumptions. What is the effect of using the nonparametric test after a transformation is chosen based on a parametric procedure when there are outliers or trends in the data? What are some other approaches to adjust for this problem and are they effective? Transformations such as the inverse and inverse square root may deemphasize the large values and overemphasize the small values (especially the zero values). What are the effects of zero values with these and other transformations on the nonparametric test? It is crucial that these questions receive further research.

A further philosophical concern with the BACI analysis is the test for additivity. As indicated in Table 3, all the fish species data required transformation, except for two for which additivity could not be achieved using any of the four transformations. Many transformations are either the inverse or inverse square root transformations. The test used in Table 3 is for only one type of violation of additivity: scale differences between the control and impact areas. In fact, the test is equivalent to testing equality of variance for the data from the control and impact areas in the before period (see Bradley and Blackwood 1989). Other forms of nonadditivity, in particular, that of trends in the differences in time, should be investigated in a BACI analysis (as noted by Stewart-Oaten et al. 1986). Trends may occur in the differences or in the individual locations.

Sampson and Guttorp (1991) noted that, if a transformation is made, then this transformation also affects the hypothesis being tested. The hypothesis of no impact effect in the original scale is not the same as the hypothesis of no impact effect in a transformed scale. For example, a test of no impact in the original scale corresponds to a test on the mean and variance in the inverse scale (x^{-1}), not a test on the means in the inverse scale. Furthermore, the BACI test of no interaction in the inverse scale may not mean that there is no interaction (impact) in the original scale. Thus, interpretation of the data may be confounded by the transformation. Sampson and Guttorp (1991) presented an approach for testing hypotheses about impacts in the original scale using data in a transformed scale.

Some of these difficulties are illustrated in Table 4. Several transformations are applied and additivity and a test for impact conducted using both parametric and nonparametric methods. This produced two interesting results. First, for both the combined data set and the river east side data set, the largest p-values for additivity tests are the inverse square root transformation and the inverse transformation, respectively, regardless of whether a parametric or nonparametric method is used. However, the parametric and nonparametric tests for impact give different results. In both cases, the nonparametric method results in significance, while the parametric test is not significant. The second concern occurs with the test of additivity for the river west side. The largest p-value using a parametric test is for the square root transformation. However, this transformation is significant (at the 0.05 level) using a nonparametric method. The nonparametric approach suggests using a log transformation.

Data transformation is also useful for normalizing the data and making the assumptions for parametric analysis more tenable. However, transformations do not work well with all distributions. Data used here are count data, specifically the number of fish of a particular species. When a large number of observations have the same value (in this case, zero), the transformation may not work as anticipated and will not remove the effect of the cluster of observations with the same value. Stewart-Oaten et al. (1992) further noted that transforming to achieve additivity may not always succeed and that such a failure may indicate that the treatment effect changes with another factor, for example, season.

Assumption of Independence

Another difficulty with the BACI analysis is the assumption of independence. While some evidence shows that the Mann–Whitney test is less sensitive to the independence assumption (Hirsch et al. 1982), it is still affected. Furthermore, the test of additivity may be sensitive to temporal dependence (serial autocorrelation). Temporal dependence could cause an excessive number of rejections. Seasonality may also be a problem because fish populations undergo natural fluctuations over a year. Differences in fish abundance in December may have quite different statistical properties than differences in June when the number of young fish increases. The BACI model is most effective if the variance in the difference does not change over sampling time. This is probably not true for fish in temperate rivers because the variance in the summer difference may be substantially higher than the variance in the winter months. Methods for analyzing impact data with seasonal effects are available (Thomas et al. 1978; Hirsch et al. 1982; van Belle and Hughes 1984).

Sampling Frequency

The BACI model works best if the timing of sampling is the same in the before and after periods. For example, if, during the before period the frequency of sampling is high, with samples taken on a monthly basis, but in the after period, sampling occurs on a monthly basis for some years and only in the summer months thereafter, there may be fewer small differences in the after period than in the before period (as winter months are excluded). For the power plant data, samples were taken approximately once each month for most of the study. However, in the last 3 yr, sites were sampled only in June, August, and October (Fig. 1). This unbalancing of data may cause variances

TABLE 4. p-values for additivity and the BACI test for a variety of transformations of data on taxon 14. Additivity is tested using a parametric (Pearson) and nonparametric (Kendall) correlation. The test for impact is carried out using a two-sample t-test (equal variances) and the Wilcoxon two-sample test. The analyses are done for each sampling site and for the combined sites. *Optimal transformation.

Transformation	Combined	East	West	Test
				Additivity
None	0.0001	0.0002	0.7927	Pearson
	0.0001	0.0002	0.0135	Kendall
	0.0024	0.0008	0.5366	t-test
	0.0030	0.0008	0.3843	Wilcoxon
				Additivity
Square root	0.0001	0.0001	0.8613*	Pearson
	0.0003	0.0001	0.0454	Kendall
	0.0011	0.0002	0.6867	t-test
	0.0031	0.0015	0.4723	Wilcoxon
				Additivity
Log	0.5243	0.0232	0.8108	Pearson
	0.3524	0.0139	0.7875*	Kendall
	0.0677	0.0024	0.9314	t-test
	0.0098	0.0035	0.9490	Wilcoxon
				Additivity
Inverse square root	0.9454*	0.3918	0.7728	Pearson
	0.8837*	0.4683	0.1702	Kendall
	0.1687	0.1021	0.3743	t-test
	0.0265	0.0089	0.6447	Wilcoxon
				Additivity
Inverse	0.7172	0.5812*	0.5862	Pearson
	0.3714	0.9935*	0.0751	Kendall
	0.3742	0.4196	0.3280	t-test
	0.0434	0.0149	0.5497	Wilcoxon

in the after period to be larger than in the before period, thus affecting the tests and possibly resulting in detection of differences that are not there.

It is difficult to sample rivers on a common day each month because events such as floods and foul weather can severely affect the performance of sampling. Electrofishing is commonly done at the same depth; however, if the river is flooded, then the only locations with an appropriate depth may be on a site that was previously not underwater. The effects of minor inconsistencies in sampling frequency should have minor effects on the analysis. However, it is best to have the after sampling as similar as possible to the before sampling.

Pooling of Sample Sites

The question of how to analyze data collected at multiple stations (locations within the sites) in impact and control sites is difficult to answer. The simplest approach is to pool data from the locations. However, there are a number of concerns with this approach. Pooling data from the locations is valid if the locations within the sites do not significantly alter the abundance of fish. However, two sites in a river may be quite different due to differences in the soils, depth of the river at the site, or the habitats. Site variability can be very important if the control sites are different from the impact sites or if the impact affects sites differentially. Site differences can also increase statistical variability and make the test less sensitive. In our case history, the plant is located near a bend in the river, and the habitat below the plant is different from that above the plant, especially for the site located on the outside of the bend (in our example, the west site). The bend may cause the habitat at this outside site below the plant to be different from that on the similar site above the plant as well as the site directly across the river. Using the fish data, the analysis can be applied to each site pair. Alternatively, a site variable can be added to the model, although this would make the analysis more complicated. It would be appropriate to use the bank to pair the sites; however, if the outer site below the plant is quite different in terms of habitat, then it may be that neither of the control sites provides an adequate control for this particular site. This problem cannot be resolved by analytical methods.

Temporal Blocking

The BACI is a sensitive method for analyzing data. Because differences are used in the analysis, sensitivity to events occurring in both areas (e.g. seasonality, flooding effects) is diminished. However, in planning BACI studies, the design is important (see Eberhardt and Thomas 1991). The design treats the two sampling sites (above and below the plant) as a temporal block. The design is most powerful if the temporal blocks are true blocks and reduce variation. Variation is reduced by blocking if the sites within a block are similar (and remain similar over time). If the control site is notably different from the impact site, then the sites are likely to respond differently to natural changes in environmental conditions and, as a result, confound conclusions. For example, habitat may change due to natural fluctuations in river flow over long periods. If these river flow patterns are different in the control site as compared with the impact site, then the habitat change in the control site is likely to be different from that in the impact site. Hence, fish abundance would be expected to change differentially. If the flows tend to be higher in the before period than in the after period, then before-after is confounded with flow. This confounding

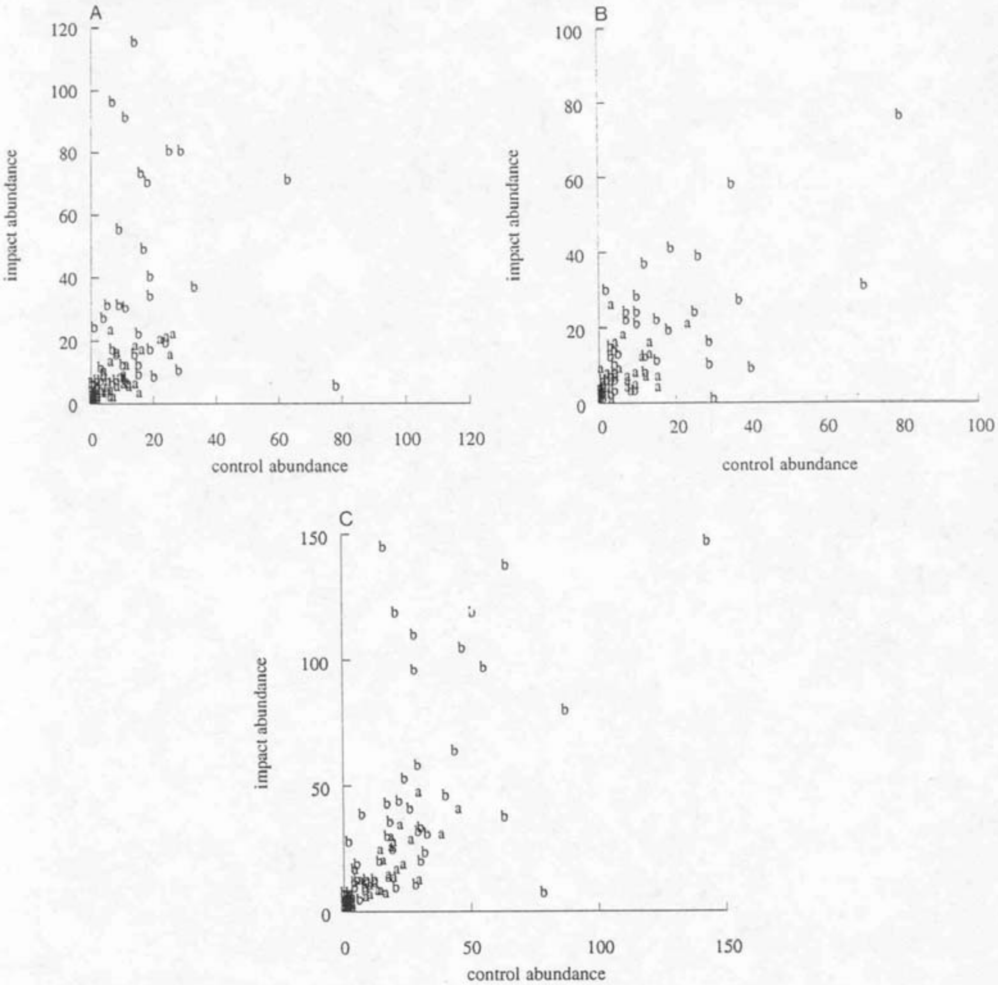

FIG. 3. Plot of the control site abundance versus the impact site abundance for taxon 14 for (A) east sites, (B) west sites, and (C) combined sites. The period is denoted b for before and a for after.

could be removed by adjusting the data for flow and then applying a BACI analysis. An indication of a flow effect could appear as a trend or some other pattern in the before data. While some of the effects of these factors can be removed, it requires planning (e.g. collect flow data from the start) and knowledge of the factor. Without planning and knowledge, the design may become flawed.

While it cannot be expected that dissimilar sites will change in an ecologically similar fashion, there is no guarantee that ecologically similar sites will remain ecologically similar over time (McCune and Allen 1985). Microhabitat differences in the sites may affect food chain dynamics and species abundances. Differential competition and predation may result in quite different species distributions.

A simple assessment of the blocking is obtained by graphing before data for the control site versus the impact site. In Fig. 3, the data for species 14 are plotted for both before (using the label b) and after (a). If the blocking is effective (i.e. greatly reduces variation), the data from the before period would be expected to fall near a straight line with a slope of 1 (but the intercept need not be zero). All three plots suggest that the blocking is not effective. Large deviations in the slope from 1

TABLE 5. p-values for test for trend in time for data from the before plant operation period using Kendall's tau statistic. The sign of the correlation is given in parentheses for the significant ($p < 0.05$) correlations. Cases with joint zeros in the control and impact were deleted.

Species	n	Control west	Control east	Impact west	Impact east
1	49	0.4903	0.8043	0.0004 (+)	0.7411
2	62	0.0637	0.0269 (+)	0.3110	0.2280
3	48	0.0542	0.6314	0.0646	0.5201
4	61	0.2377	0.5692	0.0097 (−)	0.0487 (−)
5	70	0.1752	0.0012 (−)	0.0464 (−)	0.0036 (−)
6	63	0.3875	C.9516	0.0007	0.8089
7	67	0.4920	0.4536	0.2618	0.0894
8	40	0.0031 (−)	0.4570	0.2143	0.0039 (−)
9	37	0.2215	0.1963	0.9214	0.6400
10	58	0.0717	0.0085 (+)	0.0000 (+)	0.0570
11	12	0.1111	0.3497	0.3364	0.9353
12	29	0.0823	0.6769	0.1979	0.4163
13	41	0.0006 (−)	0.2256	0.9130	0.0027 (−)
14	62	0.4733	0.4275	0.0478 (+)	0.9708
15	26	0.0901	0.0017 (−)	0.9090	0.0214 (−)
16	24	0.2931	0.5902	0.7089	0.1235
17	31	0.2586	0.1714	0.4811	0.0449 (−)
18	32	0.0016 (−)	0.0079 (−)	0.4586	0.6162
19	69	0.0003 (+)	0.0246 (+)	0.0000 (+)	0.0012 (+)
20	57	0.0004 (+)	0.0678	0.0174 (+)	0.0018 (+)

and low correlations suggest inadequate blocking. Figure 3A, in particular, indicates possible problems in that the high densities for impact are not matched with high densities in the control. Although the after data are plotted on the graphs, these values are difficult to observe because almost all of the values are small. Thus, there is a reduction of abundances not only for the impact area but also for the control area. The small differences noted in Fig. 2 for the after period are due to a reduction in both areas. Because of the lack of matching, it is difficult to ascertain if the effect is due to the operation of the plant or some other factor.

Confounding Factors

It is quite a leap of faith to assume that there are no other factors that influence the ecology of the system. The argument in favor of the BACI design is that, if the factor influences both the control and impact sites equally, the effect is removed by differencing. However, our experience with long-term studies suggests that factors affect sites differentially over time. The effects of these factors needs to be investigated. In the case of our example, the building of the plant and economic changes that result from plant construction may be an important factor. It may be quite difficult to disentangle the effect of construction from the plant operation, as these factors are confounded in time. In our example, a large number of confounding variables further suggests that the plant may not cause substantial ecological change. These include effects of a sewage treatment plant, addition of antifouling chemicals, differences in flow for different years, changes in the sampling plan, and changes in habitat. A bacterial insecticide (*Bacillus thuringiensis*) was applied in shallow areas starting in 1986 to control blackflies. The insecticide may affect macroinvertebrates and, hence, the food chain. Other indirect effects are also possible. The construction and operation of the plant require a large number of personnel that, in itself, could also have affected results. Increased fishing intensity in the impact area may have resulted in some significant changes. Data on these factors are needed to assess their influence. Then, the BACI analysis could be rerun, adjusting for the variables. Without this information, the results of the statistical analysis may be indefensible in a court of law. Even with the information, the analysis may be weakened.

Further Statistical Analysis of BACI Data

Because of the problems suggested in the above discussion, several additional analyses were conducted.

(1) A test of trend in the control and impact areas was conducted. As mentioned above, differences in trend in the control and impact area before plant startup in 1982 would be a violation in the assumption of additivity.

(2) An analysis of a reduced data set was conducted. Data were reduced by eliminating the first year of sampling (1975), which was a year of preliminary results. Also, because the final sampling was done only in June, August, and October, data from other months were deleted. This provided identical sampling times for both the before and after data (a more balanced data set).

(3) Separate analyses for the east and west sides of the river were conducted. The bend in the river would suggest that the sides are different and, hence, that pooling would not be appropriate.

(4) To examine variability in the before data, the BACI model was used on only the before data. Because it is important that the control be similar to the impact site prior to plant operation, we compared the control and impact sites using the BACI model with data from the period 1975–79 as the preimpact data and 1979–82 as the postimpact data. The month where the division of data was assigned was chosen to be similar to that of the full data set with the actual impact. One other split of the data was used to ascertain the influence of the split.

(5) Finally, an analysis of the presence and absence pattern of fish species in the samples was conducted. Species that are absent in a large number of samples are not useful indicators

TABLE 6. p-values for tests of trend in the difference in control and impact using Kendall's tau and the data for the before plant operation period. Cases with zeros in both the control and impact sites were deleted. ***Significant at the 5% level.

Species	Combined	East	West
1	0.0101***	0.8576	0.0008***
2	0.8837	0.7467	0.4544
3	0.7819	0.7356	1.00
4	0.0502	0.2161	0.0923
5	0.0072***	0.0268***	0.0150***
6	0.0063***	0.8751	0.0128***
7	0.1772	0.5728	0.4119
8	0.2084	0.0155***	0.1910
9	0.3466	0.6138	0.3227
10	0.0954	0.1898	0.0063***
11	1.000	0.5630	0.1013
12	0.3061	0.4788	0.0360***
13	0.0783	0.8182	0.0013***
14	0.3009	0.6006	0.0227***
15	0.1320	0.9100	0.3468
16	0.7999	0.4935	0.1844
17	0.9869	0.8913	0.4054
18	0.0111***	0.0992	0.0181***
19	0.0038***	0.0352***	0.0304***
20	0.1064	0.0214***	0.7979

of change in abundance; however, they may suggest changes in species richness.

Results of Additional Analyses

Tables 5 and 6 present results of the tests for trends in data collected prior to plant operation based on Kendall's tau measure of correlation. Results in Tables 5 and 6 suggest that the conclusions drawn from the previous data analysis need to be questioned. Table 5 indicates that a number of species show trends. Table 6 indicates that the differences between control and inpact abundance also show trends for a number of species. Many of the species have a positive trend, while a few exhibit negative trends. A positive trend in our case history indicates either large negative differences that become smaller or positive differences that become larger. A negative trend indicates negative differences becoming larger or positive differences becoming smaller. *If the species with significant trends are eliminated from the BACI analyses, as suggested by Stewart Oaten et al. (1986), then most of the significant differences identified in the initial analysis disappear* (Table 3).

Because imbalancing of sampling data may have resulted in false significances, the BACI analysis was rerun using only the months June, August, and October. These results are presented in Table 7 and, in general, are less significant than the previous analyses. For the west side, only two taxa (6 and 7) are significant ($\alpha = 0.05$). The p-values for these taxa are slightly less than 0.05. For the east side, three species are also significant. These are taxa 4, 14, and the unidentified species. The results for taxa 4 and 14 indicate highly significant differences. The combined set of locations indicated four species (4, 6, 14, and 20) were affected. Part of the lack of significance for species in this data may be due to the reduction in sample size. Given the smaller sample sizes in this study, the results, therefore, suggest a change in abundance due to the plant for several species, especially 4 and 14.

To assess the validity of the BACI analysis for this study, an analysis of the before data was done. Here the data from the

TABLE 7. p-values for the BACI test using the data for the months of June, August, and October. The year 1975 was dropped. The transformation was the same as the one used for the full data (U = untransformed, S = square root, L = log, IS = inverse square root, I = inverse). The sample size is 17 in the before period and 20 in the after period. ***Significant at the 5% level; + + + indicates that no transformation was successful at removing nonadditivity and — indicates that data were zero.

Species	Combined	East	West
1	0.1799 (I)	+ + +	+ + +
2	0.8430 (I)	0.5321 (IS)	0.4929 (I)
3	0.2228 (IS)	0.5222 (I)	0.1435 (I)
4	0.0010 (S)***	0.0028 (S)***	0.0908 (U)
5	0.8192 (IS)	0.5422 (L)	0.3294 (S)
6	0.0027 (I)***	0.1900 (I)	0.0154 (I)***
7	0.0908 (S)	0.1849 (L)	0.0293 (I)***
8	0.5938 (I)	0.5026 (I)	0.2005 (I)
9	0.4461 (I)	0.2861 (I)	0.1244 (I)
10	0.2060 (L)	0.1702 (IS)	0.9878 (L)
11	—	—	—
12	0.1435 (I)	+ + +	0.4020 (S)
13	0.7605 (I)	0.1702 (L)	0.8311 (I)
14	0.0031 (L)***	0.0017 (L)***	0.4929 (U)
15	0.4833 (I)	+ + +	0.1314 (I)
16	0.2931 (I)	0.8789 (I)	0.4553 (U)
17	0.6808 (I)	0.4193 (U)	0.4461 (I)
18	0.2287 (I)	+ + +	0.3219 (I)
19	0.4929 (IS)	0.8549 (L)	0.5123 (IS)
20	0.0109 (IS)***	0.0355 (U)***	0.0530 (S)

after period were deleted and the data from the before period were split into two groups. The data before September 1979 were treated as the before group and the data from September 1979 through August 1982 formed the after group. The results are given inTable 8 and provide some interesting observations. By chance alone, one would expect significance one time in 20 tests. The higher the proportion of significant results above this expected level, the less our confidence is in the BACI method as applied to our data. For the combined sites, 4 of 20 species were significant, 2 of 18 were significant for the east site, and 3 of 15 were significant for the west site. Although the species differ, the number that are significant is similar to that of the previous analysis described in Table 5, and the number of effects detected is greater than the number expected due to chance. There could be several explanations for these additional significances. The test may be detecting changes caused by plant construction (not operation), the presence of a sewage treatment plant (started in 1978), or other unknown effects. Alternatively, the significances may reflect the trends in the data at the sites that were detected in Tables 5 and 6.

Some of the significances associated with the BACI tests may be related to the proportion of zeros in the data. For example, species 8 was present between 24 and 32% of the time in the before period but decreased to almost zero in the after period for all sites except the east impact site where it occurred 15% of the time. Six other species (1, 11, 12, 15, 17, and 18) were absent approximately 70% or more of the time.

Conclusions and Comments

Although a number of fish showed significant change according to the BACI model, there are a number of reasons, both statistical and biological, for doubting whether these changes should be attributed to the plant. A number of species abun-

TABLE 8. p-values for the BACI test (Wilcoxon) using the data from the before period only, split into two groups at September 1979. The sample size is 43 for the first square root (ALL = transformation had no effect on the test. U = untransformed. S = square root. L = log. IS = inverse square root. I = inverse). ***Significant at the 5% level; + + + indicates that none of the transformations was successful at removing nonadditivity and — indicates that no test was performed, as most of the data were zeros.

Species	Combined	East	West
1	0.2299 (I)	0.7680 (I)	0.0615 (I)
2	0.1946 (I)	0.1112 (U)	0.6294 (L)
3	0.8801 (I)	0.4508 (I)	0.3883 (I)
4	0.9519 (I)	0.8187 (IS)	0.4841 (IS)
5	0.2230 (IS)	—	0.1616 (IS)
6	0.0036 (IS)***	0.5031 (I)	0.0052 (IS)***
7	0.6294 (S)	0.6208 (IS)	0.9231 (IS)
8	0.0923 (I)	0.0465 (I)***	0.9471 (I)
9	0.2299 (I)	0.8253 (I)	—
10	0.3883 (IS)	0.8944 (I)	0.0161 (S)***
11	0.6950 (ALL)	1.000 (ALL)	—
12	0.4765 (I)	+ + +	—
13	0.2467 (I)	0.8801 (IS)	0.0294 (I)***
14	0.4508 (IS)	0.9856 (I)	0.3405 (I)
15	0.1427 (I)	0.2966 (IS)	0.7860 (I)
16	0.0492 (I)***	0.3623 (I)	0.3051 (S)
17	0.6166 (I)	0.7721 (I)	—
18	0.1477 (I)	0.1427 (IS)	—
19	0.0177 (L)***	0.0741 (IS)	0.1315 (IS)
20	0.0047 (U)***	0.0107 (S)***	0.4051 (I)

dances indicated trends over time in the before period, suggesting interactions that were not included in the model. The sampling program was changed following the initiation of plant operation. Also, the analysis was made more difficult by the large number of zero catches in the data set and the need to transform the data to more closely meet the assumptions of the analysis. Finally, a large number of other factors should be eliminated as possible causative factors before a decision on plant impact is made. Given these other possibilities, we cannot conclusively determine whether the plant has a substantial biological impact on fish populations. It is clear that the results of this analysis would not support a decision against the power company in a court of law.

Our experience with fish data reinforces the commonly held view that fish populations vary considerably in natural settings. While the BACI analysis is intended to be robust for many effects altering natural variance, it is advantageous to have control sites as similar to impact sites as possible. In the case history examined, this is made difficult because the habitat may be changing and this change may cause unwanted interaction if it differs between the control and impact areas. The interaction is suggested by trends in time and suggests that the habitat change in the control site is different from that in the impact site.

The above case study illustrates some of the difficulties in ecological impact assessment. Although the view in the above analyses is rather pessimistic, there are a number of lessons that can aid future studies. Some of these are described below; the papers by Eberhardt and Thomas and their colleagues provide some excellent advice on the design of assessment programs (Thomas et al. 1978; Eberhardt and Thomas 1991; and references therein).

The BACI model is one of the best models for impact assessment, but its success depends on a sound sampling design. Some of the problems encountered (such as the sampling frequency) can be controlled by the agency carrying out the investigation. Control of the sampling program must be maintained throughout the time frame of the study. Major changes should be avoided or at least carefully considered in view of the needs of the proposed analysis. Supplemental information, such as from toxicity tests or on the age structure of the fish population, is relatively inexpensive to collect and can be quite useful for assessing effects and assigning causes (Pratt et al. 1988; Pontasch et al. 1989). Prior information about habitats is needed for selecting sites in the control and impact areas. It is useful to try to anticipate some of the changes that might be caused by other factors (such as construction) and to try to address these factors in the study design. Finally, the role of statistical significance (as opposed to biological significance) should not be overemphasized because statistical significance (with a fixed Type I error rate) is tied to sample size. As the number of sampling times increases (before and after), the amount of change required to produce a significant impact generally decreases. The important game fish, taxa 14, which was one of the taxa showing significant change, had a change in the difference between control and impact of between 1 and 4 fish per catch (i.e. the median difference between control and impact was roughly −1 in the before period and roughly 1 in the after period). Whether this change is important biologically and to what extent the change can be attributed to the plant are difficult questions. A measure of what is biologically significant in assessing impacts and making environmental decisions is important. However, the question of assessing change in field situations is a difficult one that requires sound knowledge not only of the statistical difficulties but also of the biological and political ones. More integration of these components would be beneficial. Sound design requires not only a good statistical model but also an understanding of the underlying biological processes (what to measure) and careful planning (how to measure it well). All of these components are required to achieve defensible impact assessments.

References

BRADLEY, E. L., AND L. G. BLACKWOOD. 1989. Comparing paired data: a simultaneous test for means and variances. Am. Stat. 43: 234–235.

EBERHARDT, L. L., AND J. M. THOMAS. 1991. Designing environmental field studies. Ecol. Monogr. 61: 185–199.

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

1989. Power analysis and practical strategies for environmental monitoring. Environ. Res. 50: 195–205.

HELSEL, D. R. 1990. Less than obvious: statistical treatment of data below the detection limit. Environ. Sci. Technol. 24: 1767–1774.

HIRSCH, R. M., J. R. SLACK, AND R. A. SMITH. 1982. Techniques of trend analysis for monthly water quality data. Water Resour. Res. 18: 107–121.

HURLBERT, S. J. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54: 187–211.

KLEINBAUM, D. G., L. L. KUPPER, AND K. E. MULLER. 1988. Applied regression analysis and other multivariable methods. 2nd ed. PWS-Kent Publishing Co., Boston, MA.

LAMBERT, D. 1992. Zero-inflated Poisson regression, with application to defects in manufacturing. Technometrics 34: 1–14.

MCCUNE, B., AND T. F. H. ALLEN. 1985. Will similar forests develop on similar sites? Can. J. Bot. 63: 367–376.

MILLIKEN, G., AND D. JOHNSON. 1984. The analysis of messy data. Vol. 1: Designed experiments. Lifetime Learning Publications, Belmont, CA.

PETERMAN, R. M. 1990. Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47: 2–15.

PONTASCH, K. W., B. R. NIEDERLEHNER, AND J. CAIRNS, JR. 1989. Comparisons of single-species, microcosm, and field responses to a complex effluent. Environ. Toxicol. Chem. 8: 521–532.

PRATT, J. R., N. J. BOWERS, B. R. NIEDERLEHNER, AND J. CAIRNS, JR. 1988. Effects of atrazine on freshwater microbial communities. Arch. Environ. Contam. Toxicol. 17: 449–457.
SAMPSON, P. D., AND P. GUTTORP. 1991. Power transformations and tests of environmental impact and interaction effects. Am. Stat. 45: 83–89.
STEWART-OATEN, A., J. R. BENCE, AND C. W. OSENBERG. 1992. Assessing the effects of unreplicated perturbations: no simple solutions. Ecology 73: 1396–1404.
STEWART-OATEN, A., W. W. MURDOCH, AND K. R. PARKER. 1986. Environmental impact assessment: pseudoreplication in time? Ecology 67: 929–940.
THOMAS, J. M., J. A. MAHAFFEY, K. L. GORE, AND D. G. WATSON. 1978. Statistical methods used to assess biological impact at nuclear power plants. J. Environ. Manage. 7: 269–290.
VAN BELLE, G., AND J. HUGHES. 1984. Nonparametric tests for trend in water quality. Water Resour. Res. 20: 127–136.

The following papers were not cited but are relevant:

Bernstein, B.B., and J. Zalinski. 1983. An optimal sampling design and power tests for environmental biologists. J. Environm. Managem. 16: 35-43.
Underwood, A.J. 1992. Beyond BACI: The detection of environmental impacts on populations in the real, but variable, world. J. Exp. Mar. Biol. Ecol. 161: 145-178.
Underwood, A.J. 1993. The Mechanics of spatially replicated sampling programmes to detect environmental impact in a variable world. Austr. J. Ecol. 18: 99-116.