

**Randomized Intervention Analysis and the Interpretation of  
Whole-Ecosystem Experiments**



Stephen R. Carpenter; Thomas M. Frost; Dennis Heisey; Timothy K. Kratz

*Ecology*, Vol. 70, No. 4 (Aug., 1989), 1142-1152.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9658%28198908%2970%3A4%3C1142%3ARIAATI%3E2.0.CO%3B2-5>

*Ecology* is currently published by The Ecological Society of America.

---

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

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

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

---

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

## RANDOMIZED INTERVENTION ANALYSIS AND THE INTERPRETATION OF WHOLE-ECOSYSTEM EXPERIMENTS<sup>1</sup>

STEPHEN R. CARPENTER, THOMAS M. FROST, DENNIS HEISEY, AND  
TIMOTHY K. KRATZ

*Center for Limnology, University of Wisconsin, Madison, Wisconsin 53706 USA*

**Abstract.** Randomized intervention analysis (RIA) is used to detect changes in a manipulated ecosystem relative to an undisturbed reference system. It requires paired time series of data from both ecosystems before and after manipulation. RIA is not affected by non-normal errors in data. Monte Carlo simulation indicated that, even when serial autocorrelation was substantial, the true  $P$  value (i.e., from nonautocorrelated data) was  $<.05$  when the  $P$  value from autocorrelated data was  $<.01$ . We applied RIA to data from 12 lakes (3 manipulated and 9 reference ecosystems) over 3 yr. RIA consistently indicated changes after major manipulations and only rarely indicated changes in ecosystems that were not manipulated. Less than 3% of the data sets we analyzed had equivocal results because of serial autocorrelation. RIA appears to be a reliable method for determining whether a nonrandom change has occurred in a manipulated ecosystem. Ecological arguments must be combined with statistical evidence to determine whether the changes demonstrated by RIA can be attributed to a specific ecosystem manipulation.

**Key words:** *ecosystem; experiment; intervention analysis; lake; manipulation; randomization tests; replication; statistics.*

### INTRODUCTION

Whole-ecosystem experimentation has made vital contributions to ecology (Hasler 1964, Likens 1985). Ecosystem-level experiments have been especially valuable when results of smaller scale experiments have contradictory implications for ecosystem phenomena (Schindler 1988). At the community level, whole-system manipulations of large, mobile predators have revealed responses at spatial scales much larger than experimental enclosures (Carpenter et al. 1987, Schindler 1987). Ecosystem experiments are properly scaled for many important management issues and so are especially useful in applied ecology (Magnuson et al. 1984a, National Research Council 1986, Frost et al. 1988, Kitchell et al. 1988).

Two analytical challenges arise in the interpretation of ecosystem experiments: (1) did the manipulated ecosystem change following the manipulation? and (2) did the manipulation cause the change? (Frost et al. 1988). Both questions can be answered by randomized, replicated experimental designs (Hurlbert 1984, Frost et al. 1988). However, replication is rarely possible in whole-ecosystem experiments because of limited funding, limited access to experimental ecosystems, and public health, political or environmental difficulties with the necessary manipulations. Moreover, high inter-annual and interecosystem variability dictates that large numbers of replicates are needed in many cases to achieve adequate statistical power. For example, in experiments on primary production in lake ecosys-

tems, about five replicate lakes per treatment are necessary to detect even large manipulation effects (Carpenter 1989). Many community properties are even more variable than primary production (Carpenter and Kitchell 1987, Schindler 1987, 1988) and would require even more replicates to detect effects. Problems of time-treatment interaction may further increase the numbers of replicates needed (Walters et al. 1988). In many cases, there will simply not be enough ecosystems to achieve adequate replication. For example, two prominent experimental lake reserves in North America, Canada's Experimental Lakes Area and the University of Notre Dame Environmental Research Center, contain only 46 and 27 lakes, respectively. Given the large number of important questions that require whole-ecosystem experiments (National Research Council 1986, Schindler 1987, 1988), these critical studies will rarely, if ever, be adequately replicated.

As an alternative to replication, whole-ecosystem experiments can be used selectively to determine the response potential of ecosystems to powerful manipulations which produce massive, unequivocal responses in one system at one time (Kitchell et al. 1988). In these unreplicated experiments, experimental and reference ecosystems are observed before and after a manipulation is applied to the experimental system. Replicated small-scale experiments, process studies, and models can then be used to determine the mechanisms that underlie ecosystem responses (Frost et al. 1988, Kitchell et al. 1988). This approach has been used in successful ecosystem experiments for nearly 40 yr (Hurlbert 1984, Likens 1985, Schindler 1988). Generally, ecosystem experiments are designed to evoke

<sup>1</sup>Manuscript received 25 March 1988; revised and accepted 8 August 1988.

massive responses in target variables. However, questions often arise about less pronounced but potentially important consequences of major manipulations. These questions are particularly critical in the search for early indicators of change, those subtle responses of ecosystems that forecast much larger changes to come (Schindler 1987, 1988). Novel statistical approaches to the detection of change in experimental ecosystems are needed to resolve such questions. Here we introduce and test Randomized Intervention Analysis for determining whether change has occurred following experimental manipulation of ecosystems.

#### RANDOMIZED INTERVENTION ANALYSIS

Antecedents of randomized intervention analysis (RIA) are described by Box and Tiao (1975), Edgington (1980), and Stewart-Oaten et al. (1986). Box and Tiao (1975) present tests to detect nonrandom change in a series of observations made before and after manipulation of a single system. Stewart-Oaten et al. (1986) described before-after-control-impact (BACI) analysis, in which experimental (impact) and reference (control) ecosystems are compared before and after the treatment of the experimental system. These approaches use parametric statistics, and the authors discussed many concerns about problems of independence and distribution of errors. Mohr (1988) has summarized and compared these approaches to intervention experiments.

RIA uses randomization to derive an error distribution from the data itself, so that non-normality does not affect the test results (Edgington 1980). This feature is a distinct advantage, because temporal trends and lagged responses commonly cause non-normal error distributions in ecosystem experiments. Also, RIA is not affected by heterogeneous variances (unlike some alternatives, such as the *t* test). While RIA is affected by autocorrelations in the data, our results show that lack of independence among sequential observations will not cause equivocal results in many ecosystem applications of RIA.

RIA begins with a series of parallel observations of experimental and reference ecosystems, paired in time, spanning periods before and after a manipulation (Fig. 1). A time series of intersystem differences is then calculated, and from these are calculated mean values for the premanipulation and postmanipulation differences,  $\bar{D}(\text{PRE})$  and  $\bar{D}(\text{POST})$ , respectively. The absolute value of the difference between  $\bar{D}(\text{PRE})$  and  $\bar{D}(\text{POST})$  is the test statistic. Its distribution is estimated by random permutations of the sequence of intersystem differences.

The null hypothesis assumes Nature assigns intersystem differences to pre- or posttreatment at random, hence the term randomization. More precisely, the null model states that all possible permutations of the data have an equal probability of being observed. Since we can compute  $|\bar{D}(\text{PRE}) - \bar{D}(\text{POST})|$  for each

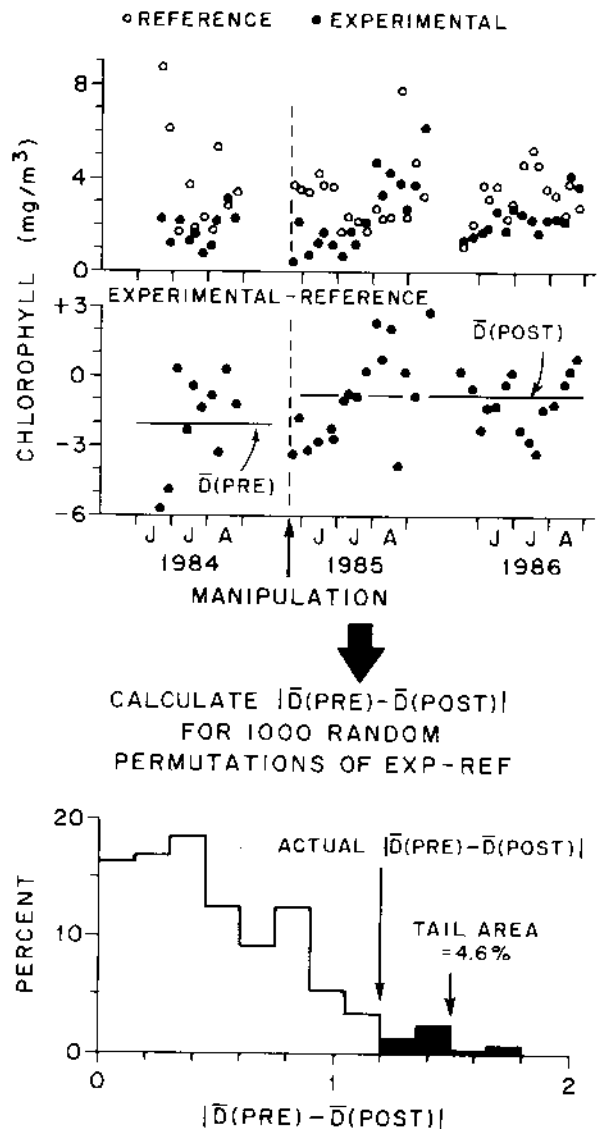


FIG. 1. Calculations for Randomized Intervention Analysis. From the paired data from both experimental and reference ecosystems before and after manipulation, intersystem differences (EXP - REF) are calculated. Mean intersystem differences before and after manipulation,  $\bar{D}(\text{PRE})$  and  $\bar{D}(\text{POST})$ , respectively, are then calculated. Random permutation of the intersystem differences yields the distribution of the test statistic,  $|\bar{D}(\text{PRE}) - \bar{D}(\text{POST})|$ . Data in example from lakes Peter (experimental) and Paul (reference).

of these possible permutations, the distribution of the test statistic can, in theory, be determined exactly. Exact determination is practical if the number of permutations is relatively small (Green 1977). If the number of permutations is large, exact *P* values may be computationally infeasible, in which case *P* values can be approximated by Monte Carlo methods. To do this, differences are randomly assigned to times before or after the manipulation, regardless of their position in the actual sequence. Many such permutations are generated, and the resulting frequency distribution of

$\{\bar{D}(\text{PRE}) - \bar{D}(\text{POST})\}$  is examined. The proportion of values of  $|\bar{D}(\text{PRE}) - \bar{D}(\text{POST})|$  that exceeds the observed value is the approximate  $P$  value. The  $P$  value indicates the probability of a test statistic as or more extreme than that observed, under the null hypothesis. If this  $P$  value is low, then one concludes that a non-random change in the interecosystem difference occurred following the manipulation. RIA does not indicate when the change occurred, though this information may be evident from the time series of observations.

In a general sense, RIA tests the null hypothesis that the same changes or trends occurred in the experimental and reference ecosystems following the manipulation. It is critical to note that rejection of this null hypothesis does not demonstrate that the manipulation caused the change (Box and Tiao 1976, Hurlbert 1984). Like any intervention analysis, RIA can only establish whether or not a change occurred (Box and Tiao 1976). To establish that the manipulation caused the response, one must show beyond reasonable doubt that no alternative causes could produce the observed change (Hurlbert 1984, Walters et al. 1988). In unreplicated ecosystem experiments, determination of causality will rest on ecological rather than statistical arguments (Hurlbert 1984, Frost et al. 1988). RIA indicates whether a change has occurred, and therefore, whether or not there is any need to consider its cause.

The performance of RIA in ecosystem experiments must be explored systematically. Given the high temporal variability of certain ecosystem parameters (Carpenter and Kitchell 1987, Kratz et al. 1987, Schindler 1987, 1988), it is conceivable that RIA could detect a high frequency of changes in unmanipulated ecosystems or, conversely, fail to detect even massive changes in manipulated systems. Here we examine the performance of RIA on parallel data from two whole-lake experiments and a set of nonmanipulated, long-term study lakes, all located within a 30 km radius in the Northern Highlands Lake District of Wisconsin. We test for (1) the ability of RIA to detect apparent change in experimentally manipulated ecosystems and (2) the frequency with which RIA detects change in lakes that have not been manipulated. The two whole-lake experiments are the Cascading Trophic Interactions project in which two lakes were subjected to food web manipulations and a third was maintained as a reference system (Carpenter et al. 1987), and the Little Rock Lake Experimental Acidification Project in which a lake was separated by a curtain with one basin acidified and the other maintained as a reference system (Brezonik et al. 1986, Watras and Frost 1989). Seven additional, nonmanipulated lakes are elements of the North Temperate Lakes Site of the Long-Term Ecological Research (LTER) Program (Magnuson et al. 1984b, Kratz et al. 1986). The food web of one LTER lake, Sparkling Lake, appears to be changing dramatically following the inadvertent introduction of an ex-

otic planktivore (Magnuson and Beckel 1985, McLain and Magnuson 1988).

## METHODS

### *Statistical procedures*

We determined the effects of sample size and autocorrelation on RIA by Monte Carlo simulation. In simulations to determine the effects of sample size,  $N$ , the premanipulation sample consisted of  $N/2$  normally distributed random numbers with mean  $\bar{X}$  and standard deviation  $s$ . The postmanipulation sample consisted of  $N/2$  normally distributed random numbers with mean  $\bar{X} + ms$ , where  $m$  is the specified manipulation effect.

Simulations to determine the effects of autocorrelation followed a similar procedure except that the sequences of random numbers were autocorrelated. In this paper, autocorrelation pertains to the time series of interecosystem differences (Fig. 1). Both autoregressive and moving average models were used.

In the autoregressive model,

$$Y(t) = C \cdot Y(t-1) + Z(t)$$

where  $Y(t)$  is the random variate (with mean zero) at time  $t$ ,  $C$  is the autoregression coefficient, and  $Z(t)$  is a random effect from a normal distribution with mean zero and unit variance. The variance of  $Y$  is

$$s^2 = 1/(1 - C^2)$$

and the autocorrelations are  $r_1 = C$ ,  $r_2 = C^2$ ,  $r_3 = C^3$ , etc., where the subscript denotes the lag (Chatfield 1980). We performed calculations for specified values of  $r_1$ .

In the moving average model,

$$Y(t) = Z(t) - C \cdot Z(t-1)$$

where  $C$  scales the moving average term, and other terms correspond to those in the autoregressive model. The variance of  $Y$  is

$$s^2 = 1 + C^2$$

and only the lag 1 autocorrelation is nonzero (Chatfield 1980)

$$r_1 = -C/(1 + C^2).$$

We performed calculations for specified values of  $r_1$ . In solving the quadratic equation for  $C$  given  $r_1$ , we used the root between  $-1$  and  $+1$ .

In simulated experiments with autoregressive or moving average models, a total of 40 autocorrelated  $Y$ s were generated. The first 20 represented the premanipulation data. A manipulation effect equal to a specified multiple of the standard deviation was added to the second set of 20  $Y$ s. These sample sizes were similar to those in our 3-yr data sets. We analyzed only positive autocorrelations. No cases of negative autocorrelation occurred in our data. While positive autocorrelations are anticonservative (i.e., lead to over-

TABLE 1. Manipulated and reference lakes studied. Code for projects: CTI = Cascading Trophic Interactions; LRLEAP = Little Rock Lake Experimental Acidification Program; LTER = Long-Term Ecological Research.

Lake	Project	Comment
Manipulated lakes:		
Peter	CTI	Piscivores removed, planktivores added
Little Rock Acid	LRLEAP	Acidified
Tuesday	CTI	Piscivores added, planktivores removed
Reference lakes:		
Allequash	LTER	
Big Muskellunge	LTER	
Crystal	LTER	
Crystal Bog	LTER	
Paul	CTI	
Little Rock Reference	LRLEAP	
Sparkling	LTER	Exotic planktivore (smelt) invading; for further information, see McLain and Magnuson (1988)
Trout	LTER	
Trout Bog	LTER	

estimation of  $P$ ), negative autocorrelations are conservative (i.e., lead to underestimation of  $P$ ) and are therefore of less concern.

Because our limnological data included replicate reference ecosystems, we could compare each manipulated lake with the set of reference lakes by a conventional two-sample  $t$  test. For this analysis, we calculated the response of each lake as the postmanipulation mean minus the premanipulation mean. These tests, involving 4–9 reference lakes, have very low statistical power because of no replication in one group (Snedecor and Cochran 1967). With 10 lakes, for example, the balanced design (5 experimental and 5 reference systems) has a  $t$  statistic 2.4 times larger than that of the design with 1 experimental and 9 reference systems, all other terms being equal (Snedecor and Cochran 1967).

The goal of this paper is to examine the performance of RIA for the typical application to one experimental and one reference ecosystem. In most applications, multisystem comparisons such as those we present will not be possible and control of experimentwise error rates will not be an issue. Therefore, we discuss significance at the nominal level of 5%, rather than at a lower level adjusted for simultaneous comparisons.

#### Limnological procedures

We computed RIAs for surface pH, light extinction coefficient of the epilimnion, epilimnetic chlorophyll, and densities of three zooplankton taxa (*Keratella cochlearis*, all *Daphnia* species combined, and all calanoid copepods combined). Here we emphasize three variables (pH, chlorophyll concentration, and *Daphnia* density) which represent the diversity of patterns we observed. The ability of RIA to detect changes directly related to manipulation was examined using pH, which exhibits modest temporal variability and was affected directly by the acidification of the experimental basin of Little Rock Lake (Brezonik et al. 1986). Chlorophyll concentration is a variable that is affected indirectly by both acidification (Schindler et al. 1985) and fish

manipulation (Carpenter et al. 1987). It exhibits greater temporal variability than pH, but is much less variable than the concentrations of individual plankton taxa. Light extinction, like chlorophyll, is affected indirectly by manipulations but is even more variable. *Daphnia* density is also affected indirectly by acidification and fish manipulation, and is among the most variable limnological parameters. To check for effects of taxonomic resolution, we also examined a species-level variable (*Keratella cochlearis*) and a suprageneric variable (total calanoid copepods).

Overall, we analyzed data from 12 lakes for 1984–1986 (Table 1). Three lakes were subjected to whole-system manipulations in the last week of May 1985. Little Rock Lake was divided by a neoprene curtain in 1984 and acidified by the addition of concentrated  $H_2SO_4$  on 29 May 1985. Acidification continued to maintain pH 5.6 during the ice-free periods of 1985 and 1986. Peter and Tuesday lakes underwent reciprocal fish manipulations during 23–31 May 1985. Ninety percent of the piscivore biomass of Peter Lake was transferred to Tuesday Lake, and 90% of the planktivore biomass of Tuesday Lake was transferred to Peter Lake (Carpenter et al. 1987). For some parameters, data were not available for every interlake comparison.

The two basins of Little Rock Lake and the LTER lakes (Allequash, Big Muskellunge, Crystal, Crystal Bog, Sparkling, Trout, and Trout Bog) were sampled and analyzed with the same procedures, except for zooplankton densities. Each basin or lake was sampled at a central, deep water station every 2 wk during the ice-free season and  $\approx$  every 5 wk throughout the rest of the year. Both basins of Little Rock Lake were sampled on the same day and in alternate weeks to the LTER lakes. Sample order for the LTER lakes was random within a week. Reported pH values are for samples not equilibrated with the atmosphere (Kratz et al. 1987). Epilimnetic light extinction was calculated from irradiance measurements made at 1-m intervals through-

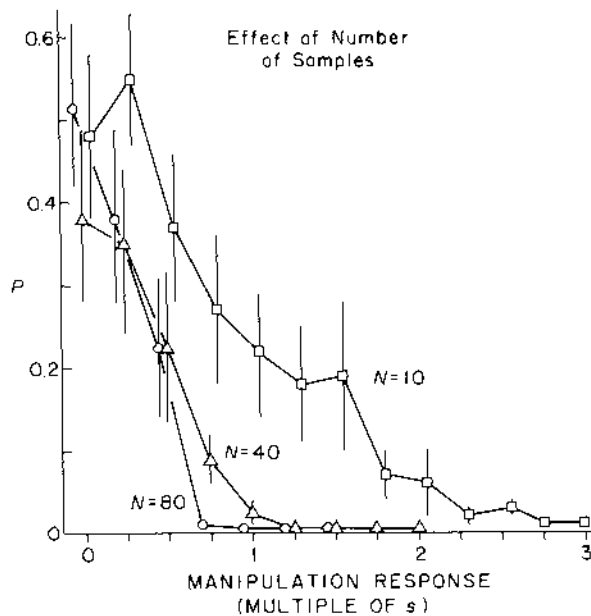


FIG. 2. Effects of sample size ( $N$ ) on RIA.  $P$  value from RIA is plotted vs. manipulation response as a multiple of the standard deviation( $s$ ). Curves are shown for series of 10, 40, and 80 samples. In all cases the manipulation was applied at the midpoint of the time series. Error bars are 95% confidence intervals based on 36 simulated experiments.

out the epilimnion with a flat PAR sensor and a reference deck cell. Epilimnetic chlorophyll was calculated as the mean of 2–6 samples taken with a peristaltic pump at different depths above the thermocline. Samples were filtered in-line, extracted with acetone-DMSO, and chlorophyll, corrected for pheopigments, was determined spectrophotometrically. In Little Rock Lake, zooplankton were sampled from three depths using a Schindler-Patalas trap (Frost and Montz 1988). Densities are reported as a simple average or as a volume-weighted average determined by adjusting for the lake volume represented by each depth stratum. In the LTER lakes, zooplankton samples were collected with a Schindler-Patalas trap but were pooled prior to counting, with individual samples weighted to represent the volume of their depth stratum.

Lakes involved in the Cascading Trophic Interactions project (Paul, Peter, and Tuesday lakes) were sampled weekly from mid-May through mid-September each year (Carpenter et al. 1987). Paul, Peter, and Tuesday lakes were sampled on Monday, Tuesday, and Wednesday of each week, respectively, at central deep water stations. Surface pH was determined after stirring samples for at least 10 min to equilibrate them with the atmosphere. The epilimnetic light extinction coefficient was calculated from irradiance measurements at 6–10 depths obtained with a spherical quantum sensor and reference deck cell. Epilimnetic chlorophyll was calculated as the mean of 3 or 4 samples taken by van Dorn bottle from different depths above

the thermocline. Samples were filtered, extracted in methanol, and chlorophyll concentration, corrected for pheopigments, was determined fluorometrically. Zooplankton densities were determined by enumerating animals collected in vertical net hauls. Filtering efficiencies of the net were determined by comparing vertical hauls with profiles obtained by Schindler-Patalas traps (Carpenter et al. 1987).

## RESULTS

### Simulations

When only 10 samples were obtained in simulated experiments, manipulation effects of 2 standard deviations or less were not generally detected by RIA at the 5% level (Fig. 2). In cases with 40 samples, comparable to our data sets, manipulation effects of 1 standard deviation or more were consistently detected by RIA. Doubling the sample size to 80 produced a small increase in sensitivity.

Serial autocorrelation in the time series of interecosystem differences causes RIA to underestimate the true  $P$  value (Fig. 3). We present results for  $r_1$  values

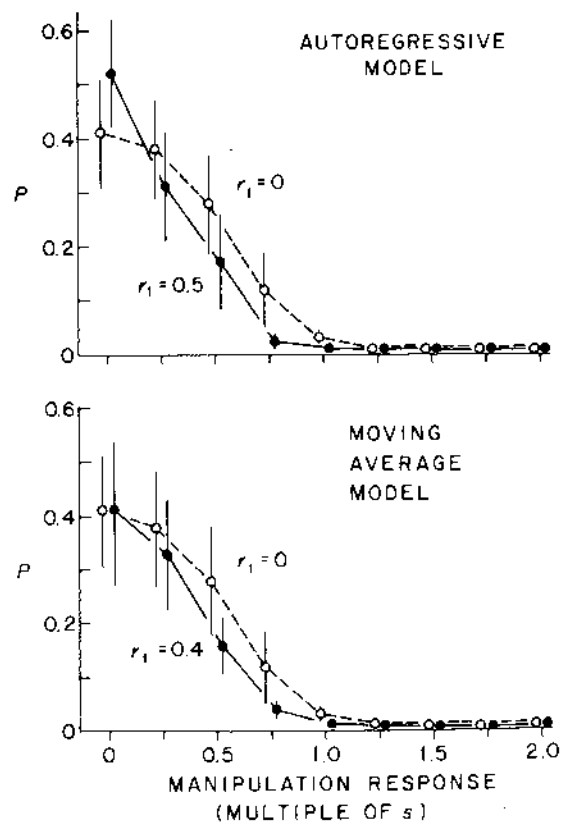


FIG. 3. Effects of autocorrelation on RIA.  $P$  value from RIA is plotted vs. manipulation response as a multiple of the standard deviation. Curves are shown for independent and autocorrelated time series calculated with both autoregressive (upper panel) and moving average (lower panel) models. Error bars are 95% confidence intervals based on 36 simulated experiments.

TABLE 2. *P* values from RIA of surface pH from nine reference lakes and the acidified half of Little Rock Lake. Underline denotes *P* < .05.

Lake	1	2	3	4	5	6	7	8	9	10
1. Allequash	...									
2. Big Muskellunge	0.513	...								
3. Crystal	0.495	0.332	...							
4. Crystal Bog	0.270	0.719	0.586	...						
5. Paul	0.945	0.774	0.065	0.715	...					
6. Little Rock acidified	<u>0.002</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	<u>0.002</u>	...				
7. Little Rock reference	<u>0.733</u>	<u>0.589</u>	<u>0.346</u>	<u>0.139</u>	<u>0.621</u>	<u>0.000</u>	...			
8. Sparkling	0.867	0.734	0.267	0.523	0.733	<u>0.000</u>	0.665	...		
9. Trout	0.628	1.000	0.807	0.461	0.906	<u>0.000</u>	0.624	0.390	...	
10. Trout Bog	0.497	0.105	0.112	<u>0.000</u>	0.417	<u>0.000</u>	0.074	<u>0.045</u>	0.433	...

near the maximum observed in our data sets. In all of the autocorrelated simulations examined, the true *P* value was < .05 if the *P* value from RIA was less than .01. Therefore, as a conservative rule of thumb, the *P* value from RIA should be < .01 to reject the null hypothesis when the time series is autocorrelated. If the time series is autocorrelated and *P* from RIA is between .01 and .05, then the results are equivocal.

#### Limnological data

RIA of surface pH data consistently indicated changes after the acidification of Little Rock Lake (Table 2). Of the 36 comparisons involving reference lakes, only two (5.6%) were significant at the 5% level. Both significant comparisons of reference lakes involved Trout Bog.

RIA of epilimnetic chlorophyll consistently indicated changes after fish manipulation in Tuesday Lake (Table 3), where piscivore addition caused fivefold reductions in algal biomass that were sustained through 1985–1987 (Carpenter et al. 1987, Carpenter and Kitchell 1988). The fish manipulation in Peter Lake caused 10-fold increases in algal biomass that persisted only ≈ 2 mo in 1985 (Carpenter et al. 1987). This transient response produced significant RIAs in comparisons with only two reference lakes, Paul and Little Rock. No response was evident for chlorophyll in Little

Rock Lake. Of the 21 comparisons involving reference lakes, only one (4.8%) was significant at the 5% level.

Because the mean is more sensitive to extreme observations than the median, we tested the effects of substituting medians for means in the calculation of  $\bar{D}(\text{PRE})$  and  $\bar{D}(\text{POST})$  for the chlorophyll data (Table 3). On the basis of significant results at the 5% level, medians and means gave almost identical results. As expected, *P* values were usually larger when medians were used (for example, see results for Tuesday Lake below). However, that was not always the case. Two (9.6%) of the comparisons involving reference lakes were significant at the 5% level when medians were used.

Most of the significant RIAs for *Daphnia* density involved manipulated lakes (Table 4), consistent with the high sensitivity of *Daphnia* to planktivorous fishes (Carpenter et al. 1987) and acidification (Schindler et al. 1985). The consistently significant RIAs involving Sparkling Lake are, perhaps, moot. An exotic planktivore (rainbow smelt, *Osmerus mordax*) invaded Sparkling Lake in 1982 (McLain and Magnuson 1988). Rapid expansion of the smelt population appears to be causing changes in *Daphnia* density comparable to those produced by deliberate fish manipulations in Peter and Tuesday lakes (McLain and Magnuson 1988).

Conventional *t* tests detected significant manipulation effects in Little Rock (pH), Peter (extinction coef-

TABLE 3. *P* values from RIA of epilimnetic chlorophyll concentration from three experimental lakes (\*) and seven reference lakes. Above diagonal: RIA based on means; below diagonal, RIA based on medians. Underline denotes *P* < .05.

Lake	1	2	3	4	5	6	7	8	9	10
1. Allequash	...	0.992	0.993	0.794	0.680	0.833	0.250	0.803	0.329	0.000
2. Big Muskellunge	0.233	...	0.215	0.875	0.601	0.261	0.260	0.353	<u>0.022</u>	<u>0.000</u>
3. Crystal	0.292	0.287	...	0.624	0.576	0.641	0.596	0.931	0.143	<u>0.000</u>
4. Little Rock acidified*	0.513	0.537	0.143	...	0.211	0.906	<u>0.006</u>	0.713	<u>0.002</u>	<u>0.000</u>
5. Little Rock reference	0.679	0.665	0.416	0.683	...	0.386	<u>0.037</u>	0.276	0.181	<u>0.000</u>
6. Paul	0.770	<u>0.030</u>	0.150	0.411	0.490	...	<u>0.032</u>	0.484	1.000	<u>0.000</u>
7. Peter*	0.692	0.279	0.734	<u>0.050</u>	0.200	0.401	...	0.360	0.993	<u>0.000</u>
8. Sparkling	0.387	0.214	0.237	0.335	0.709	0.564	0.614	...	0.512	<u>0.000</u>
9. Trout	0.416	<u>0.035</u>	0.298	<u>0.012</u>	0.143	0.615	0.696	0.709	...	<u>0.000</u>
10. Tuesday*	<u>0.006</u>	<u>0.000</u>	<u>0.000</u>	<u>0.004</u>	<u>0.046</u>	<u>0.000</u>	<u>0.000</u>	<u>0.000</u>	0.000	...

TABLE 4. *P* values from RIA of *Daphnia* density from three experimental lakes (\*) and five reference lakes. † Underline denotes *P* < .05.

Lake	1	2	3	4	5	6	7	8
1. Crystal	...	0.464	0.374					
2. Little Rock acidified*	0.331	...	<u>0.001</u>	<u>0.018</u>	0.065	<u>0.003</u>	<u>0.048</u>	0.618
3. Little Rock reference	0.352	<u>0.000</u>	...	<u>0.008</u>	0.128	<u>0.005</u>	0.865	0.246
4. Paul	<u>0.020</u>	<u>0.028</u>	0.207	...				
5. Peter*	0.900	0.063	0.596	<u>0.015</u>	...			
6. Sparkling	0.000	<u>0.008</u>	0.000	<u>0.009</u>	<u>0.000</u>	...		
7. Trout	<u>0.546</u>	<u>0.005</u>	0.785	<u>0.116</u>	<u>0.595</u>	<u>0.019</u>	...	
8. Tuesday*	0.114	0.517	0.280	<u>0.000</u>	<u>0.001</u>	<u>0.009</u>	0.413	...

† Little Rock Lake densities were computed by two methods: simple averaging (below diagonal) as in Paul, Peter, and Tuesday lakes, and hypsometrically-weighted averaging (above diagonal) as in Crystal, Sparkling, and Trout lakes.

ficient, total calanoid copepods), and Tuesday (chlorophyll concentration, total *Daphnia*) lakes (Table 5). However, because the power of this test is very low, lack of significance does not demonstrate lack of a manipulation effect. In two cases with significant *t* tests, all RIAs were significant (pH in Little Rock Lake, chlorophyll concentration in Tuesday Lake). In the other cases, RIA detected differences between the manipulated lake and only some of the reference lakes. In many of these cases, graphs of the time series suggested that any change in the manipulated lake was paralleled by changes in some reference lakes. For example, declining chlorophyll concentration in the acidified basin of Little Rock Lake and increasing chlorophyll in Peter Lake corresponded to similar changes in certain reference lakes. In a few cases, no response of the manipulated lake would have been expected. For example, Tuesday Lake is stained and its extinction coefficient

is very insensitive to changes in chlorophyll concentration (Elser 1987).

Overall, few significant RIAs occurred in comparisons of nonmanipulated lakes (14 of 108, or 12.9%, Table 5). Four of these involved differences in *Daphnia* density between Sparkling Lake and the other nonmanipulated lakes (Table 4), which was associated with the smelt invasion of Sparkling Lake (McLain and Magnuson 1988). Discounting these cases, only 9.6% (10/104) of the comparisons involving reference lakes produced significant RIAs.

The incidence of significant autocorrelations in time series with significant RIAs was about the same as in the complete set of time series (Table 5). Overall, 30.1% of the time series (66/219) were autocorrelated. Of the time series with significant RIAs 32.9% (23/70) were autocorrelated. Autocorrelation was especially frequent in the pH time series, and especially infrequent

TABLE 5. Summary of RIA results and incidence of significant autocorrelations at lag 1 ( $r_1$ ) for the interlake comparisons we examined.

Variable	Lake	Significant RIAs/total RIAs		Significant $r_1$ s*	
		Manipulated lakes	Nonmanip. lakes	Per test	Per sig. RIA
pH	*Little Rock	(9/9)	2/36	20/45	5/11
Extinction coefficient	Little Rock	(4/9)	2/21	4/45	0/10
	*Peter	(4/9)			
Chlorophyll concentration	Tuesday	(0/9)			
	Little Rock	(3/9)	1/21	9/45	2/14
	Peter	(4/9)			
<i>Keratella cochlearis</i>	*Tuesday	(9/9)			
	Little Rock	(1/7)	4/10	13/28	4/9
	Peter	(2/7)			
<i>Daphnia</i> spp.	Tuesday	(2/7)			
	Little Rock	(4/7)	5/10	9/28	5/15
	Peter	(3/7)			
Total calanoid copepods	*Tuesday	(3/7)			
	Little Rock	(5/7)	0/10	11/28	7/11
	*Peter	(4/7)			
	Tuesday	(2/7)			

\* Numbers of significant autocorrelations are tabulated per number of comparisons as well as per comparison with significant RIA.

† Lakes with significant changes by *t* test against all nonmanipulated lakes. No significant *t* tests occurred for nonmanipulated lakes.

in the time series for extinction coefficients and chlorophyll concentration. Autocorrelation was only rarely associated with marginal  $P$  values. Overall, only 2.7% (6/219) of time series had significant autocorrelations and  $P$  values between .01 and .05.

#### DISCUSSION

RIA appears to be a reliable means of detecting changes in time series from whole-ecosystem experiments. Differences following ecosystem manipulations were indicated regularly, while differences were not common in analyses of reference ecosystems. RIA was especially helpful when changes due to manipulation were similar in magnitude to the short-term variability of the time series and visual judgments were therefore difficult. The power of the test increases slowly with increasing sample size when time series contain more than  $\approx 40$  observations. Our analyses identified three areas for caution in the application of RIA: pairing of sample dates, autocorrelations in the time series of intersystem differences, and timing and duration of responses. In many cases, the latter two problems can be resolved by inspection of the time series of both ecosystems and their differences.

##### *Pairing sample dates*

In general, samples for RIA should be paired in time as closely as possible. When samples are not closely paired, possible effects of the pairing method on the results should be considered.

We found one example in which the method of pairing sample dates affected the results (Fig. 4). Little Rock Lake is sampled within 6 d of the previous week's LTER samples, and within 11 d of the subsequent week's LTER samples. Therefore, in Tables 1–3 we paired Little Rock samples with the previous week's LTER samples, their nearest neighbors in time. In rare cases, very different results arose when Little Rock sample were paired with the subsequent week's LTER samples. Extinction coefficients in Big Muskellunge Lake and the acidified basin of Little Rock Lake provide the most dramatic example:  $P = .008$  for pairing with previous (nearest) samples, and  $P = .753$  for pairing with subsequent samples (Fig. 4). Neither pairing method shows large changes, and the sensitivity of the results to the pairing methods suggests that these results are equivocal. Big Muskellunge Lake is much larger than the acidified basin of Little Rock Lake, and attendant differences in heat budget and phenology of the plankton may account for the sensitivity of these results to the time lag between samples.

##### *Autocorrelations*

Positive autocorrelation in the time series of intersystem differences causes RIA to underestimate the true  $P$  value (Fig. 3). A conservative correction for this problem is to require  $P < .01$  instead of the nominal value of .05 when significant autocorrelation exists.

This rule of thumb would lead to correct interpretations of even the most severely autocorrelated data sets in our sample. We observed no  $r_1$  values  $> 0.5$ . If  $r_1$  exceeds 0.5, effects on RIA can be estimated by Monte Carlo analysis of the autoregressive model described in the Methods.

Autocorrelation was not a severe obstacle to interpretation of our data. Overall, about one-third of the time series were autocorrelated, but in most of these cases the  $P$  values were so high or low that the results of RIA were unequivocal. Autocorrelation caused equivocal results in  $< 3\%$  of the cases we examined. These observations are consistent with those of Stewart-Oaten et al. (1986), who argued on theoretical grounds that autocorrelation may be only a minor problem in the analysis of time series of differences between experimental and reference ecosystems.

When autocorrelated time series cause equivocal results, intervention analyses that explicitly model the serial dependency may be an alternative to RIA (Box and Tiao 1976). However, our experiences with this method were generally unsatisfactory because of problems with nonstationarity and model identification in our relatively short time series.

##### *Timing and duration of responses*

RIA gives no information about the timing or nature of any change that it detects. Some changes detected by RIA are not consistent with the interpretation that the manipulation caused the change. For example, if a trend in the interecosystem difference existed before the manipulation, or if a change occurred in the reference system but not the manipulated system, then a significant RIA could occur even if the manipulation had no effect. In such cases, misinterpretations can be prevented by examining plots of the time series for each ecosystem and their difference.

A less obvious limitation of RIA is its inability to detect transient effects that span  $\approx 10$  or fewer sampling intervals. In unreplicated experiments, such brief responses will be difficult to evaluate by any statistical technique.

We found an informative example in which the RIA was significant even though no manipulation effect was evident. Chlorophyll concentrations in Big Muskellunge Lake and Trout Lake show weak but opposite trends which translated into a significant trend in the interlake difference ( $P = .022$ ) even though autocorrelation coefficients were small and nonsignificant (Fig. 5). However, the plots show that this low  $P$  value results from a gradual, noisy trend rather than an abrupt change following the date of manipulation. We analyzed random subsamples of the data to determine the sensitivity of the conclusions to particular data points (Fig. 6).  $P$  values rose and became much more variable when subsets of the data were analyzed. In contrast, the chlorophyll response of Tuesday Lake relative to its nearby reference system, Paul Lake, was much more pro-

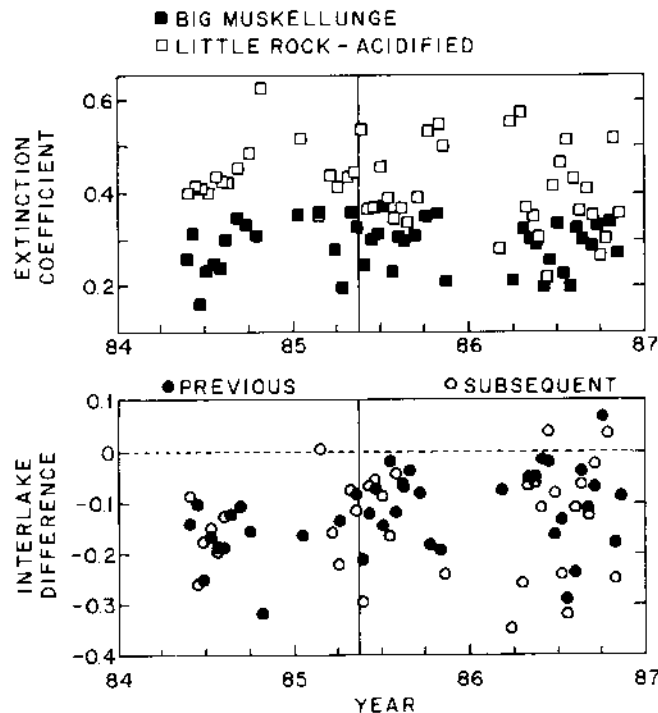


FIG. 4. Upper: extinction coefficient ( $m^{-1}$ ) in the acidified basin of Little Rock Lake and Big Muskellunge Lake vs. time, 1984–1986. Vertical line denotes the acidification of Little Rock Lake. Lower: interlake differences in extinction coefficient between Little Rock Lake and the previous week's sample in Big Muskellunge Lake, and between Little Rock Lake and the subsequent week's sample in Big Muskellunge Lake.

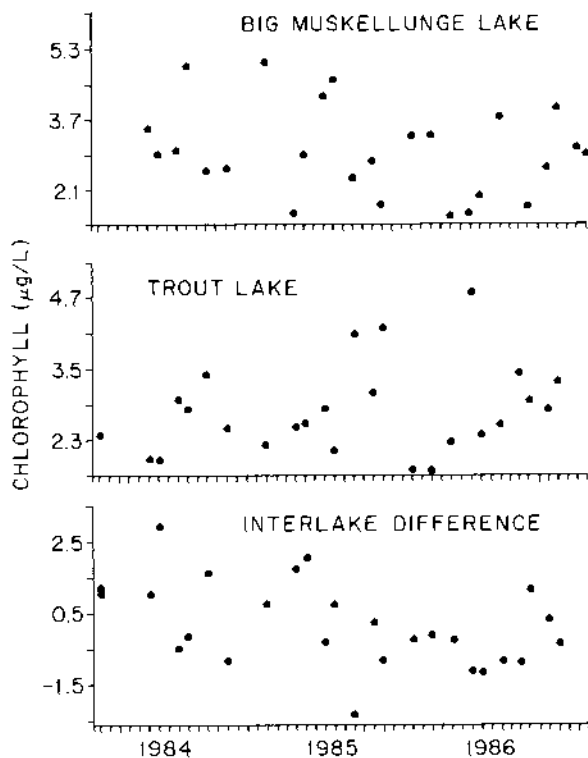


FIG. 5. Epilimnetic chlorophyll concentrations in Big Muskellunge and Trout lakes, and the interlake difference, vs. time, 1984–1986.

nounced (Carpenter et al. 1987).  $P$  values were very low for Tuesday and Paul lakes even when randomly chosen subsets of only 25% of the data were analyzed. We conclude that the changes detected by RIA in Big Muskellunge and Trout lakes result from a few influential data points, and that no changes are evident when these influential points are removed.

#### Ecological considerations

As with any statistical technique, it is critical to consider the ecological significance of RIA results. We found that graphs of both lakes' responses and the interlake difference vs. time were indispensable aids in the interpretation of  $P$  values from RIA.

Ecological criteria provide a further check on low  $P$  values from RIA of individual response variables. Ecosystem hypotheses usually predict concomitant changes in a series of variables. If an experiment is associated with changes in only one of these several variables, then it is unlikely that the postulated mechanism caused the change.

We recommend that RIA be used for initial analysis of data from ecosystem experiments with unreplicated treatments to determine whether or not a change has occurred. If no change is indicated, then RIA provides no evidence for an effect of the manipulation. Of course, statistically nonsignificant results may be biologically significant, if the number of observations is too low to

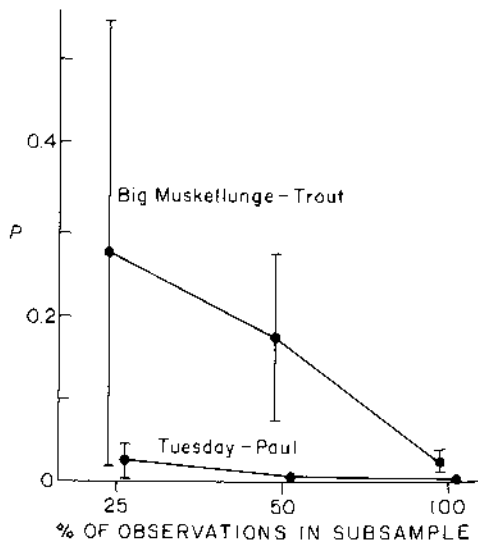


FIG. 6. Rarefaction of chlorophyll interlake differences between Big Muskellunge and Trout lakes and between Tuesday and Paul lakes. For subsamples of 25% and 50% of the data, error bars are 95% confidence limits based on 10 randomly chosen subsets. For 100% of the data, error bars are 95% confidence limits for binomial proportions.

detect the change. On the other hand, if RIA yields a low  $P$  value then a change has occurred which may have been caused by the manipulation. Time series plots for both experimental and reference ecosystems as well as the intersystem difference are valuable adjuncts to RIA and necessary for the ecological interpretations that must ultimately be used for evaluating unreplicated whole-ecosystem experiments.

#### ACKNOWLEDGMENTS

We thank J. Magnuson, J. Kitchell, and B. Benson for helpful discussions, and H. Caswell, A. Stewart-Oaten, and an anonymous referee for comments that improved the manuscript. S. R. Carpenter and T. M. Frost thank S. H. Hurlbert for extensive discussion and correspondence about RIA. Though we disagree about approaches to ecosystem experimentation, this manuscript has benefitted from his criticisms. This is a contribution from the University of Notre Dame Environmental Research Center and from Trout Lake Station, Center for Limnology, University of Wisconsin. Support was provided by N.S.F. grants DEB-80-12313, BSR 83-08918, BSR 85-14330, BSR 86-06271, and by cooperative agreements with the U.S. E.P.A. for the Little Rock Lake Experimental Acidification Project.

#### LITERATURE CITED

- Box, G. E. P., and G. C. Tiao. 1975. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association* **70**: 70-79.
- Brezonik, P. L., L. A. Baker, J. Eaton, T. Frost, P. Garrison, T. Kratz, J. Magnuson, J. Perry, W. Rose, B. Shepherd, W. Swenson, C. Watras, and K. Webster. 1986. Experimental acidification of Little Rock Lake, Wisconsin. *Water, Air, and Soil Pollution* **31**:115-121.
- Carpenter, S. R. 1989. Replication and treatment strength in whole-lake experiments. *Ecology* **70**:453-463.
- Carpenter, S. R., and J. F. Kitchell. 1987. The temporal scale of variance in lake productivity. *American Naturalist* **129**:417-433.
- Carpenter, S. R., and J. F. Kitchell. 1988. Consumer control of lake productivity. *BioScience* **38**:764-769.
- Carpenter, S. R., J. F. Kitchell, J. R. Hodgson, P. A. Cochran, J. J. Elser, M. M. Elser, D. M. Lodge, D. Kretchmer, X. He, and C. N. von Ende. 1987. Regulation of lake primary productivity by food web structure. *Ecology* **68**:1863-1876.
- Chatfield, C. 1980. *The analysis of time series*. Halsted, New York, New York, USA.
- Edgington, E. S. 1980. *Randomization tests*. Marcel Dekker, New York, New York, USA.
- Elser, J. J. 1987. Evaluation of size-related changes in chlorophyll-specific light extinction in some north temperate lakes. *Archiv für Hydrobiologie* **111**:171-182.
- Frost, T. M., D. L. DeAngelis, S. M. Bartell, D. J. Hall, and S. H. Hurlbert. 1988. Scale in the design and interpretation of aquatic community research. In S. R. Carpenter, editor. *Complex interaction in lake communities*. Springer-Verlag, New York, New York, USA.
- Frost, T. M., and P. K. Montz. 1988. Early zooplankton response to experimental acidification of Little Rock Lake, Wisconsin, USA. *Internationale Vereinigung für theoretische und angewandte Limnologie Verhandlungen* **23**:2279-2285.
- Green, B. F. 1977. A practical interactive program for randomization tests of location. *American Statistician* **31**:37-39.
- Hasler, A. D. 1964. Experimental limnology. *BioScience* **14**: 36-38.
- Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* **54**:187-211.
- Kitchell, J. F., S. M. Bartell, S. R. Carpenter, D. J. Hall, D. J. McQueen, W. E. Neill, D. Scavia, and E. E. Werner. 1988. Epistemology, experiments, and pragmatism. In S. R. Carpenter, editor. *Complex interactions in lake communities*. Springer-Verlag, New York, New York, USA.
- Kratz, T. K., J. J. Magnuson, C. J. Bowser, and T. M. Frost. 1986. Rationale for data collection and interpretation in the Northern Lakes Long-Term Ecological Research Program. Pages 22-33 in B. G. Isom, editor. *Rationale for sampling and interpretation of ecological data in the assessment of freshwater systems*. American Society For Testing and Materials, Philadelphia, Pennsylvania, USA.
- Kratz, T. K., R. B. Cook, C. J. Bowser, and P. L. Brezonik. 1987. Winter and spring pH depressions in northern Wisconsin lakes caused by increases in  $pCO_2$ . *Canadian Journal of Fisheries and Aquatic Sciences* **44**:1082-1088.
- Likens, G. E. 1985. An experimental approach for the study of ecosystems. *Journal of Ecology* **73**:381-396.
- Magnuson, J. J., J. P. Baker, and F. J. Rahel. 1984a. A critical assessment of effects of acidification on fisheries in North America. *Philosophical Transactions of the Royal Society of London, Series B* **305**:501-516.
- Magnuson, J. J., C. J. Bowser, and T. K. Kratz. 1984b. Long-term ecological research on north temperate lakes (LTER). *Internationale Vereinigung für theoretische und angewandte Limnologie Verhandlungen* **22**:533-535.
- McLain, A. S., and J. J. Magnuson. 1988. Analysis of recent declines in cisco (*Coregonus artedii*) populations in several northern Wisconsin lakes. *Finnish Fisheries Research, in press*.
- Mohr, L. B. 1988. *Impact analysis for program evaluation*. Dorsey, Chicago, Illinois, USA.
- National Research Council. 1986. *Global change in the geosphere-biosphere*. National Academy Press, Washington D.C., USA.
- Schindler, D. W. 1987. Detecting ecosystem response to anthropogenic stress. *Canadian Journal of Fisheries and Aquatic Sciences* **44** (supplement):6-25.

- . 1988. Experimental studies of chemical stressors on whole lake ecosystems. *Internationale Vereinigung für theoretische und angewandte Limnologie Verhandlungen* **23**: 11–41.
- Schindler, D. W., K. H. Mills, D. F. Malley, D. L. Findlay, J. A. Shearer, I. J. Davies, M. A. Turner, G. A. Linsley, and D. R. Cruikshank. 1985. Long-term ecosystem stress: the effects of years of experimental acidification on a small lake. *Science* **228**:1395–1401.
- Snedecor, G. W., and W. G. Cochran. 1967. *Statistical methods*. Iowa State University Press, Ames, Iowa, USA.
- Stewart-Oaten, A., W. Murdoch, and K. Parker. 1986. Environmental impact assessment: "Pseudoreplication" in time? *Ecology* **67**:929–940.
- Walters, C. J., J. Collie, and T. Webb. 1988. Experimental designs for estimating transient responses to management disturbances. *Canadian Journal of Fisheries and Aquatic Sciences* **45**:530–538.
- Watras, C., and T. M. Frost. 1988. Little Rock Lake: perspectives on an experimental ecosystem approach to seepage lake acidification. *Archives of Environmental Contamination and Toxicology* **17**:157–165.