Transactions of the American Fisheries Society 121:123–131, 1992

Can We Determine the Significance of Key-Event Effects on a Recruitment Time Series?—A Power Study of Superposed Epoch Analysis [1]

MICHAEL H. PRAGER

*Southeast Fisheries Science Center, U.S. National Marine Fisheries Service
75 Virginia Beach Drive, Miami, Florida 33149, USA*

JOHN M. HOENIG

*Science Branch, Department of Fisheries and Oceans
Post Office Box 5667, St John's, Newfoundland A1C 5X1, Canada*

Abstract.—A persistent question in fishery research is whether extreme environmental events, such as climatic perturbations or discharges of toxic substances, influence recruitment. Superposed epoch analysis has been proposed as a statistical test to address such questions. In a superposed epoch analysis of recruitment, a test statistic is computed from the differences between recruitments in years with extreme environmental events ("key years") and recruitments in immediately surrounding years; the significance of the test statistic can be determined either parametrically or nonparametrically. Here we examine the power of two parametric and four nonparametric test statistics to detect, in a variety of simulated data sets analyzed by the superposed epoch method, associations between key events and unusual values of recruitment. The statistical significance of the nonparametric test statistics is determined by randomization, the significance of the parametric statistics by consulting tabled distributions. Under the simulated conditions, we observed essentially no loss of statistical power when conducting the superposed epoch analysis with a randomization test when the parametric approach was also appropriate. However, in situations when parametric testing was not appropriate, the randomization test was often much more powerful than a parametric test. We also evaluated the statistical power of superposed epoch analyses conducted with test statistics in which recruitment data from each key year were compared in a paired fashion to data from the surrounding years. For data with strong trend or a high degree of autocorrelation, such paired test statistics outperformed the corresponding unpaired statistics; otherwise, the unpaired statistics tended to be more powerful. In testing simulated conditions patterned after the population of chub mackerel *Scomber japonicus* off southern California, we estimated the power of the proposed randomization test as approximately 0.35 to 0.50.

An important, persistent question in fishery research is whether a specified type of external event—such as an El Niño–Southern Oscillation event (Sinclair et al. 1985) or a period of extreme water temperatures (Norton 1987)—influences a stock's recruitment or production. Hypotheses about such influences have sometimes been tested with parametric statistical methods such as regression or correlation analysis. However, the assumptions (independence, homogeneity, and normality of the errors) of parametric statistical tests are not always met, seriously affecting both the realized significance levels and the statistical power of the tests.

In general, testing such a relationship requires a statistical method designed for autocorrelated data. Many standard nonparametric tests, such as the Wilcoxon signed-rank test, assume independence of the errors and thus are inappropriate for time-series data (Hollander and Wolfe 1973). Time-series analysis in the sense of Box and Jenkins (1970) is often useful; however, the Box–Jenkins approach is not satisfactory when the dynamics of the time series are complicated and the length of the time series is short. Also, gaps in the time series complicate the analysis.

To address such questions, Prager and Hoenig (1989) proposed using superposed epoch analysis (Haurwitz and Brier 1981), an approach that compares recruitment in years containing specific, environmental events ("key events") to recruitment in immediately surrounding "background" years. (Such an analysis could also be applied to other response variables such as measures of species abundance along a transect.) To determine statistical significance, a superposed epoch analysis em-

[1] Contribution MIA-90/91-72 of the Miami Laboratory, Southeast Fisheries Science Center, U.S. National Marine Fisheries Service.

TABLE 1.—The six test statistics examined in this study.

Symbol	Type of statistic	Method of determining significance	Text equation
D	Difference of means	Randomization	(1)
T_t	Student's t	Tabled distribution	(2)
T_r	Student's t	Randomization	(2)
Q_t	Paired Student's t	Tabled distribution	(4)
Q_r	Paired Student's t	Randomization	(4)
W	Modified paired Student's t	Randomization	(6)

ploys a parametric or nonparametric test statistic; any number of statistics could be devised for a particular analysis. Most nonparametric test statistics used in such analyses are similar to parametric statistics, but to avoid the assumptions of parametric testing, their statistical significance is determined by randomization tests rather than by reference to tabled significance values. Thus the data may be nonindependent and the populations need not be normal.

Although superposed epoch analysis appears appropriate for recruitment studies, there remain two major questions about it. First, is it statistically powerful enough to be useful? Knowledge of statistical power (Cohen 1988) is particularly important in a management context (Peterman 1990). Second, how can one choose the test statistic that is likely to be the most powerful? Here we describe a series of Monte Carlo simulations addressing both questions. We also examine some consequences of conducting superposed epoch analyses with parametric test statistics in inappropriate situations—that is, when data have a component of time trend or autocorrelation.

The Hypothesis Test in Superposed Epoch Analysis

The goal of a superposed epoch analysis of recruitment data is to test whether there is an association between key events and high or low values in a time series of recruitment (Prager and Hoenig 1989). By definition, key events are observed in the explanatory variable. They need not be contemporaneous with corresponding recruitment responses; however, the presumed lag between cause and effect should be constant and should be determined a priori. (In the following treatment, we assume that there is no lag; however, the modification for a nonzero lag is straightforward.) In time series with annual observations, years incorporating key events are known as "key years." To define key years, it may sometimes be necessary to make a continuous explanatory variable (e.g., total rainfall) discrete by setting a threshold, either high or low. In other cases, key events are natural occurrences such as floods, earthquakes, or severe storms.

In superposed epoch analysis, the values of the response variable in key years are compared to those in "background years," i.e., in the b_1 years before and the b_2 years after a key event. The values of b_1 and b_2 are set according to the nature of the data. Here we use $b_1 = b_2 = 2$ years, but b_1 need not equal b_2. If key events are frequent, a key year τ_i may become a background year for a second key year τ_k, in which case the response variable for τ_i is considered to have a missing value whenever τ_i is used as a background year. However, if key events are infrequent and b_1 and b_2 are small, many years in the series will be neither key years nor background years.

When using superposed epoch analysis to test recruitment time series, we define the null hypothesis to be

- H_0: high (or low) values of the recruitment index are not associated with key events.

We define the alternative hypothesis to be

- H_a: high (or low) values of the recruitment index tend to be associated with key events.

Whether we look for increased or decreased recruitment to be associated with key events depends on our a priori beliefs about the interaction between environment and the population.

Test Statistics

Any number of statistics could be used within a superposed epoch analysis to test the null hypothesis given above. We examined six related statistics (Table 1). The simplest test statistic, D, is simply the difference between the mean recruitment in key years ($\bar{E}$) and the mean recruitment in background years ($\bar{B}$):

$$D = \bar{E} - \bar{B}. \qquad (1)$$

When D is used in a superposed epoch analysis, a randomization test (described later) is used to determine statistical significance.

The second and third test statistics, T_t and T_r, are computed as a standard two-sample Student's t-statistic:

$$T_t = T_r = \frac{(\bar{E} - \bar{B})}{S_T}. \tag{2}$$

S_T is the pooled estimate of the standard error, computed as in a standard Student's t-statistic:

$$S_T = \left[\frac{(SS_E + SS_B)(N_E + N_B)}{N_E N_B (N_E + N_B - 2)}\right]^{1/2}. \tag{3}$$

Here N_E and N_B are the numbers of observations of recruitment in key years and background years, respectively; SS_E is the sum of squared deviations of the recruitment values in key years from their mean; and SS_B is the sum of squared deviations of the recruitment values in background years from their mean. The distinction between T_t and T_r is that the statistical significance of T_t is obtained from tables of Student's t-distribution, whereas the significance of T_r is estimated by randomization. Use of T_t requires an assumption that variances are equal, which is appropriate for the simulations below but might be inappropriate for some data. Given enough key events to reliably estimate the variance of $\bar{E}$, one could use the Behrens–Fisher approximate t-statistic for unequal variances (Snedecor and Cochran 1980) instead of T_t. This consideration does not apply to the nonparametric statistic T_r; nonetheless, a similar nonparametric statistic might be based on the Behrens–Fisher approximation.

The fourth and fifth test statistics are each computed as a standard paired Student's t:

$$Q_t = Q_r = \frac{\sum_{i=1}^{N_E} (E_i - \bar{B}_i)}{N_E S_Q}; \tag{4}$$

E_i is the recruitment in key year i and $\bar{B}_i$ is the mean recruitment in the background years surrounding key year i. The estimated standard deviation S_Q is computed as for a standard paired Student's t-statistic:

$$S_Q = \left[\frac{1}{N_E - 1} \sum_{i=1}^{N_E} (E_i - \bar{B}_i)^2\right]^{1/2}. \tag{5}$$

The sixth test statistic, W (Prager and Hoenig 1989), is a modification of Q_r that accounts for the varying number of background years that may be associated with key years. Although the number of background years associated with a key year is normally $b_1 + b_2$, a key year whose epoch encompasses another key year, a missing value of recruitment, or either end of the time series usually has a smaller number of associated background years.

$$W = \frac{\bar{d}(N_B)^{1/2}}{S_W}, \tag{6}$$

$\bar{d}$ being the mean of all paired differences between recruitment in key-event years and recruitment in corresponding background years:

$$\bar{d} = \frac{1}{N_B} \sum_{i=1}^{N_E} \sum_{j=1}^{n_i} (E_i - B_{ij}). \tag{7}$$

The standard error S_w is computed as for a paired t-test:

$$S_W = \left[\frac{1}{N_B - 1} \sum_{i=1}^{N_E} \sum_{j=1}^{n_i} (E_i - B_{ij} - \bar{d})^2\right]^{1/2}; \tag{8}$$

n_i is the number of background years associated with the ith key event, and B_{ij} is the recruitment in the jth background year associated with key event i.

Randomization Tests

Under the null hypothesis, there is no association between the presence of a key event and the occurrence of unusually strong recruitment. Therefore, except for sampling variability, the mean recruitment in key years should be the same as the mean of any N_E observations of recruitment drawn randomly from the data. Under the null hypothesis, then, the expected value of a test statistic for a set of key years defined by an environmental variable is the same as its expected value with the key years drawn randomly from the time series. This equality can enable us to find the null distribution of any test statistic, either exactly or approximately.

To describe the exact null distribution of a test statistic, we compute the statistic for each possible set of locations of the N_E key events within the time series of N observations. There are $N!/[N_E! \cdot (N - N_E)!]$ such sets of locations, and the value of the test statistic must be computed for each. A test based on this method might be called an exact randomization test.

Because the computations for an exact randomization test can easily become overwhelming, we might prefer to conduct a Monte Carlo randomization test, in which we approximate the null distribution of the test statistic to a high degree of precision. In this test, we repeatedly select random positions for the N_E key events and compute the test statistic for each set of positions. The distribution formed by these statistics is the estimated null distribution of the particular statistic. For either an exact or a Monte Carlo randomization test, the critical value of the test statistic for a probability P_α of type I error is the value above which $100(1 - P_\alpha)\%$ of the computed values of the statistic occur. (This is for an upper one-tailed test.) If the test statistic computed from the true position of key events is greater than this critical value, the results are considered significant at the P_α level. In the Monte Carlo randomization method, the significance level is estimated as $(x + 1)/(v + 1)$, where x is the number of simulated test statistics greater than the actual test statistic and v is the number of simulations.

In the work described here, we used the Monte Carlo randomization method with 10,000 trials. This reduced the amount of computation needed for many of our simulations and allowed programming an algorithm that was quite general. Further explanation of randomization tests can be found in Sokal and Rohlf (1981), Edgington (1987), Noreen (1989), and Prager and Hoenig (1989).

Simulated Conditions for Power Analysis

Power tests of superposed epoch analysis are complicated by the unavailability, in most cases, of analytical estimates of power for randomization tests (Edgington 1987). We therefore used an additional, "outer" Monte Carlo simulation to estimate power. (This was distinct from the "inner" Monte Carlo simulation used in each randomization test.) In this approach, we generated many 35-year time series of simulated recruitment data, varying the simulated conditions to include a variety of realistic conditions. The conditions varied were (i) the underlying model for recruitment, (ii) the parameter values—trend and autocorrelation—of the underlying model, (iii) the increase δ in recruitment associated with a key event, and (iv) the number N_E of key events in the time series. Each combination of specific conditions for (i) through (iv) was termed a scenario. We generated 500 simulated data sets for each scenario and conducted a superposed epoch analysis of each data set, computing statistical power for each of the six statistics.

The simulated data sets were generated by three underlying recruitment models. Each of the three models was a special case of the following rather general model:

$$Y_\tau = \alpha Y_{\tau-1} + \beta\tau + \delta Z_\tau + e_\tau; \qquad (9)$$

Y_τ is recruitment in year τ; α is an autoregressive parameter; β is a trend parameter; δ is the mean increase in recruitment in key years; Z_τ is an indicator for key years,

$$Z_\tau = \begin{cases} 1 \text{ if } \tau \text{ is a key year} \\ 0 \text{ otherwise;} \end{cases}$$

and the random error term $e_\tau \sim N(0, 1)$.

The first model, with $\alpha = 0$, $\beta = 0$, and $\delta \geq 0$ in equation (9), simulated conditions in which a parametric two-sample t-test would be appropriate. Recruitment values were independently and identically distributed unit-normal random variables in all years except key years. In key years, which were randomly selected, recruitment was normally distributed with mean δ and unit variance. This first set of simulations was run to determine what penalty (if any), in terms of achieved power, arises from using the randomization procedure when the usually parametric t-test is appropriate. The scenarios used for this model were all combinations of $\delta \in \{0.0, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0\}$ and $N_E \in \{3, 5\}$.

The second model introduced linear trend but not autocorrelation into the time series: $\delta \geq 0$, $\beta > 0$, and $\alpha = 0$ in equation (9). Scenarios generated under this model comprised all combinations of $\delta \in \{0.0, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0\}$, $N_E \in \{3, 5\}$, and $\beta \in \{0.1, 0.2, 0.5\}$.

The third model was first-order autoregressive, with $\delta \geq 0$, $\alpha > 0$, and $\beta = 0$ in equation (9). Scenarios generated under this model comprised all combinations of $\delta \in \{0.0, 0.5, 1.0, 1.5, 2.0, 3.0, 4.0\}$, $N_E \in \{3, 5\}$, and $\alpha \in \{0.2, 0.35, 0.6, 0.9\}$.

For each scenario, we generated and analyzed 500 simulated data sets and conducted superposed epoch analyses with the six statistics (Table 1) to repeatedly test the null hypothesis of no significant association between key events and increased recruitment. For the models described, this was equivalent to testing the null hypothesis that $\delta = 0$. The number of times that a particular statistic led to rejection of the null hypothesis at $P_\alpha = 0.05$ was divided by 500 to provide the estimate ($\hat{\phi}$) of that statistic's power in that scenario. Because this proportion approximates a binomial random variable, a confidence band can be constructed about the estimate by using the large-sample normal approximation (e.g., Hollander and Wolfe 1973). With 500 realizations of each simulation, the 95% confidence interval is ± 0.045 when $\hat{\phi} = 0.5$; the confidence interval approaches zero as the estimated power $\hat{\phi}$ approaches 0 or 1. Because the confidence intervals were so small, they are not reported.

For completeness, we point out that the variance of our power estimator is slightly larger than that of a binomial random variable, because the outcome for each data set in the simulation is determined with some uncertainty. (This is beyond the variance due to random sampling from a binomial distribution.) The formula for the variance of our estimator is that used for two-stage sampling (Cochran 1977). However, as the number of realizations used to test a particular data set becomes large, the added uncertainty approaches zero. Because we used 10,000 realizations for each data set, the added variance is negligible and need not be considered further.

Results

The results summarize the power of superposed epoch analysis in a variety of situations and with a variety of test statistics. We observed two patterns that were expected. First, the power of each test statistic, under each set of simulated conditions, increased as the effect (δ) of a key event increased (Figures 1, 2). Second, each of the six test statistics was more powerful at detecting associations in data sets with five key events (Figures 1e–h and 2e–h) than in data sets with three key events (Figures 1a–d and 2a–d).

Used within a superposed epoch analysis, the most powerful statistics were D and W. In the simulated data sets with linear trend, D was the most powerful statistic except when the linear trend was very strong or extreme (Figure 1c, d, g, h). When the trend was very strong ($\beta = 0.2$ year^{-1}), D was nearly as powerful as the most powerful statistics; but when the trend was extreme, perhaps unrealistically so ($\beta = 0.5$ year^{-1}), D was substantially less powerful than the most powerful statistics. In the presence of extreme trend, W was the most powerful statistic; in other cases, it was nominally less powerful than D. In the data sets simulated with autocorrelation, D was most powerful except in scenarios with both five key events and very high autocorrelation; in these cases, W was superior, and W always performed well in the scenarios with autocorrelation.

The effects of using paired statistics to conduct the superposed epoch analysis varied with the data structure and the test statistic. When no autocorrelation or linear trend was present (Figure 1a), the paired statistics were less powerful than the unpaired statistics. In these cases, the two least powerful statistics were Q_r and Q_t, which are both based on the unmodified paired Student's t. However, under the same conditions the W-statistic, based on a modified paired Student's t, was similar in power to the most powerful statistics. For simulated data sets with strong trend or autocorrelation, the W-statistic was in some cases the most powerful choice for conducting a superposed epoch analysis (Figures 1d, h and 2h).

Discussion

The results demonstrate the relative strengths of these parametric and nonparametric test statistics when used in a superposed epoch analysis. As with many statistical procedures, the choice among test statistics cannot be made on the basis of power alone, but depends on the structure of the data. Test statistics based on parametric theory are usually most powerful when their assumptions are met, but many parametric (and some nonparametric) methods are inappropriate for data that are not independent. In a superposed epoch analysis of autocorrelated data, this consideration would dictate the use of a randomization test instead of a parametric test. Our results are in accordance with the theory.

We observed no loss of power when randomization tests were used in the scenarios without trend or autocorrelation, cases in which the parametric tests were also appropriate. In contrast, when the parametric statistics were used inappropriately, they were often much less powerful than the randomization tests. This emphasizes the danger of using any statistical procedure when its assumptions are not met. Both the power level and the realized significance level can be quite different from anticipated values, which can lead to inappropriate conclusions.

This result is consistent with the statistical literature, which finds that nonparametric statistics are usually quite efficient. Hollander and Wolfe (1973: 1) provided this summary of the situation: "Although at first glance most nonparametric procedures seem to sacrifice too much of the basic information in the samples, theoretical investigations have shown that this is not the case. More often than not, the nonparametric procedures are only slightly less efficient than their normal theory competitors when the underlying populations are normal (the home court of normal theory methods), and they can be mildly and wildly more efficient than these competitors when the underlying populations are not normal." Transforming the data to normality is frequently recommended;

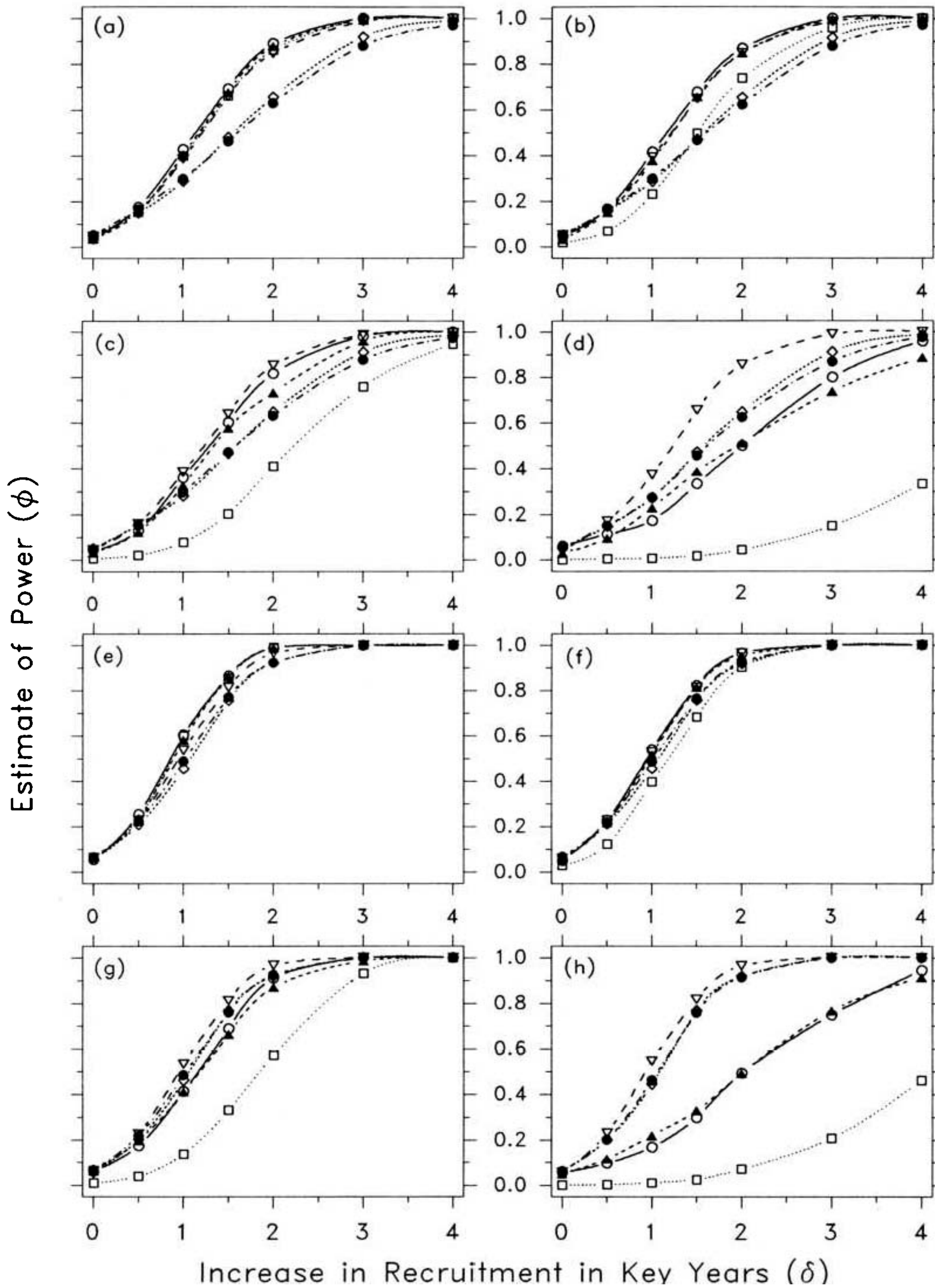

FIGURE 1.—Estimated power of six test statistics (Table 1) used in superposed epoch analyses with (a-d) three key years or (e-h) five key years in a 35-year simulated time series of recruitment. Recruitment in years without key events is normally distributed with $\mu = 0$, $\sigma^2 = 1$. Abscissa gives the expected increase in recruitment associated with a key event. The simulated data have no autocorrelation, but have the following levels of linear trend over time: (a) and (e) no trend; (b) and (f) moderate trend, 0.1 year^{-1}; (c) and (g) strong trend, 0.2 year^{-1}; (d) and (h) extreme trend, 0.5 year^{-1}. Symbols for statistics: ○ = D; ▲ = T_r; □ = T_t; ▽ = W; ◇ = Q_r; ● = Q_t.

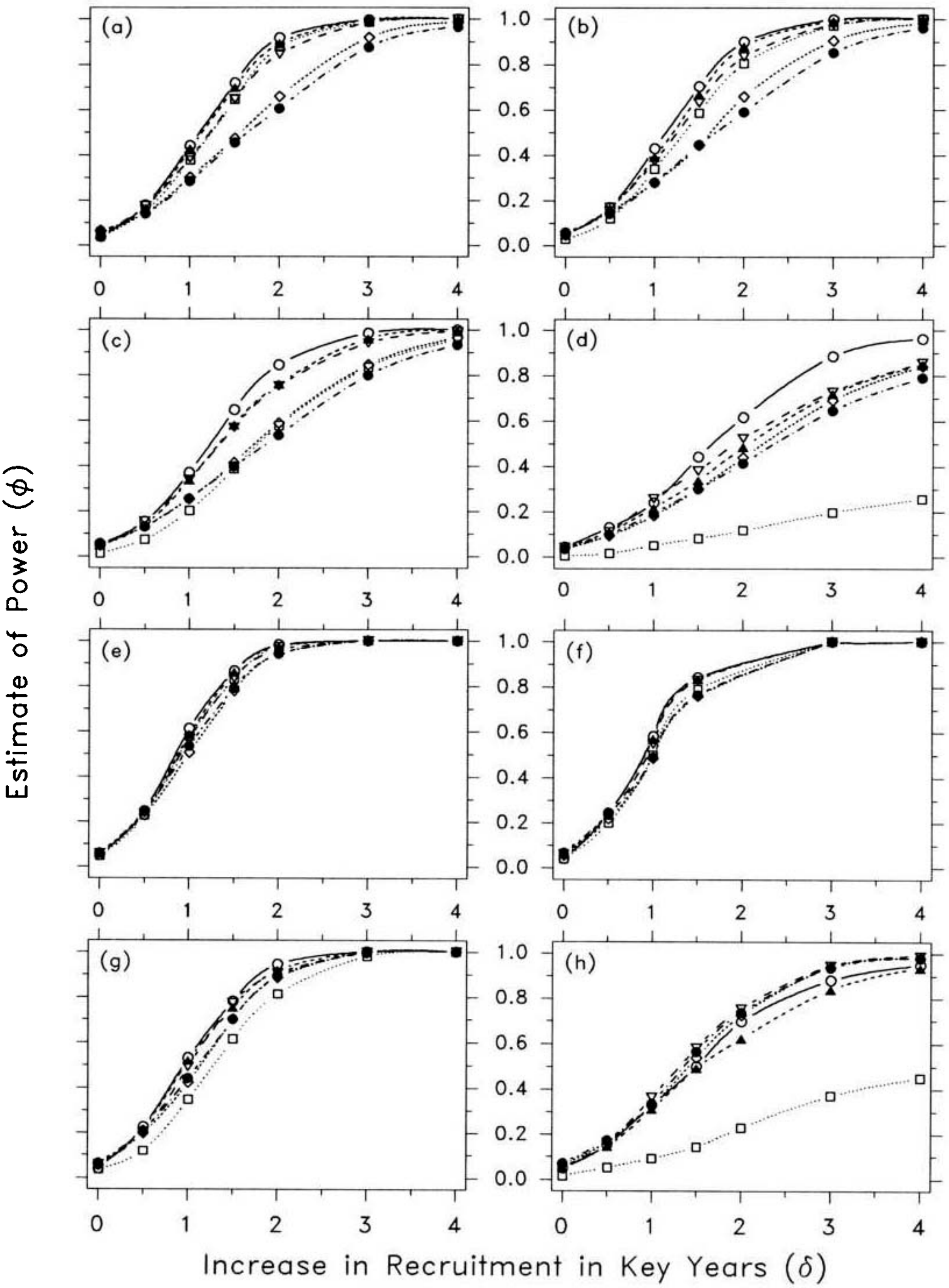

FIGURE 2.—Estimated power of six test statistics (Table 1) used in superposed epoch analyses with (a–d) three key years or (e–h) five key years in a 35-year simulated time series of recruitment. Recruitment data are generated according to the autoregressive model given in equation (9). Abscissa gives the expected increase in recruitment associated with a key event. The autoregressive parameter α took on the following levels: (a) and (e) α = 0.2; (b) and (f) α = 0.35; (c) and (g) α = 0.6; (d) and (h) α = 0.9. Symbols for statistics: ○ = D; ▲ = T_r; □ = T_t; ▽ = W; ◇ = Q_r; ● = Q_t.

however, this cannot always be accomplished and, moreover, this does not address nonindependence.

The comparison of paired and unpaired test statistics was also interesting. The paired test statistics were more powerful than the unpaired statistics when the data contained strong trends or strong autocorrelation. However, the paired statistics were often less powerful than the unpaired alternatives when the data contained only weak trends or weak autocorrelation. Despite these findings, tests provided by the paired statistic W were usually among the most powerful, suggesting that this statistic is the most widely useful of the six examined.

Although we simulated a limited number of scenarios, our results are not limited to those specific conditions. This is because tests based on the six statistics examined, like most statistical tests, are not affected by scaling the data. (This result was determined analytically in the simpler cases and confirmed by simulation.) Let the standard deviation of recruitment in non-key years (the error standard deviation) in equations (9), (10), and (11) be called σ. In our study, we used $\sigma = 1$, but the results are generalizable. For data with linear trend, the determining parameters are the two ratios $\delta{:}\sigma$ and $\beta{:}\sigma$, the ratios of the expected increase in recruitment and the linear trend, respectively, to the error standard deviation. Thus Figure 1c, which describes results with $\delta = \{0, 0.5, \ldots, 4\}$, $\beta = 2$, and $\sigma = 1$, is equally applicable to data with $\delta = \{0, 1.0, \ldots, 8\}$, $\beta = 4$, and $\sigma = 2$. For data with autocorrelation, the determining parameter is the ratio $\delta{:}\sigma$. The scale invariance means that the results presented in Figures 1 and 2 are applicable to a wide range of realistic conditions.

The most serious limitation in the generality of our results is that they apply only to 35-year time series with three or five key events and an epoch width of 5 years. However, induction can provide a rough indication of the power in other circumstances. For example, in a similar series with six key events, one would expect the power to be slightly higher than for one with five key events. For making precise estimates of power under a variety of conditions, the computer programs used here are available, with some refinements, in Hoenig et al. (1989).

Our power tests were conducted entirely within the framework of superposed epoch analysis; they did not include a comparison of superposed epoch analysis to other parametric or nonparametric procedures appropriate for autocorrelated data. It is likely that certain parametric procedures may be more powerful than superposed epoch analysis when all assumptions of the parametric test are met. The most widely accepted of these procedures, generalized least squares, requires estimating the variance–covariance matrix of the unknown random errors, which often cannot be done with a reasonable degree of precision (Kennedy 1979). An advantage of nonparametric superposed epoch analysis is that such estimation can be avoided. Another advantage is that, by examining only segments of the data (epochs) within which key events take place, the effects of random or systematic extraneous variation are lessened. This should increase the ability to detect an hypothesized effect, compared to a method that tries to test a univariate explanatory effect on all the data; however, we know of no formal study of this proposition. A third advantage of using a nonparametric superposed epoch analysis is that one need not know the true structure of the data, because the W-statistic appears to provide a powerful test in many, if not all, cases.

The present power study provides insight into our ability to detect the influence of key events on recruitment time series. As an additional example, we considered a scenario based on the logarithm of the first-year survival index for chub mackerel *Scomber japonicus* (Parrish and MacCall 1978; MacCall et al. 1985; Prager and MacCall 1988; Prager and Hoenig 1989). Sinclair et al. (1985) hypothesized that unusually high values of the chub mackerel survival index are associated with El Niño–Southern Oscillation events. In the time series of 40 years (1929–1968), 3 years have sea level anomalies of at least 1.5 SD above the mean. The index contains no significant linear trend, but is strongly autocorrelated. Ordinary least-squares estimates of the parameters of equation (9) are $\hat{\alpha} = 0.57$, $\hat{\delta} = 0.96$. Based on the power curves in Figure 2c, one would estimate the power to detect a significant association in such a data set at roughly 0.35, indicating a moderate level of statistical power. We believe this to be an underestimate, because the presence of adjacent key years in the chub mackerel data reduces the least-squares estimate of δ (and hence the estimate of power from Figure 2c), but not the ability of superposed epoch analysis to detect the effect from the actual data. We cannot estimate the amount of underestimation, but if the true value were $\delta = 1.25$, for example, the estimated power would be slightly more than 0.5. Based on this example and the other results presented here, we conclude that the statistical power of superposed epoch analysis should be sufficient in many cases to detect associations between environmental events and unusually high or low recruitment.

Acknowledgments

This study was begun while M.H.P. was at the Department of Oceanography of Old Dominion University, which provided partial support for the work. The Computer Center of Old Dominion University contributed central processing time (about 150 hours) on the IBM 3090/VF supercomputer. Partial support was also provided by the Department of Fisheries and Oceans, Canada; by the Miami Laboratory of the Southeast Fisheries Science Center, U.S. National Marine Fisheries Service; and by the Cooperative Institute for Marine and Atmospheric Studies through their Visiting Fellow program, funded through National Oceanic and Atmospheric Administration Cooperative Agreement NA85-WCH-06134 with the University of Miami. We thank Robert Crittenden, Michael Fogarty, Nicholas Payton, Pierre Pepin, Kenneth Pollock, Victor Restrepo, William Warren, Reg Watson, and an anonymous reviewer for suggestions. Remaining errors are the authors'.

References

Box, G. E. P., and G. M. Jenkins. 1970. Time series analysis—forecasting and control. Holden-Day, San Francisco.

Cochran, W. G. 1977. Sampling techniques, 3rd edition. Wiley, New York.

Cohen, J. 1988. Statistical power analysis for the behavioral sciences, 2nd edition. L. Erlbaum Associates, Hillsdale, New Jersey.

Edgington, E. S. 1987. Randomization tests, 2nd edition. Marcel Dekker, New York.

Haurwitz, M. W., and G. W. Brier. 1981. A critique of the superposed epoch analysis method: its application to solar-weather relations. Monthly Weather Review 109:2074–2079.

Hoenig, J. M., M. H. Prager, and N. Payton. 1989. Computer programs useful for testing hypotheses about recruitment time series. Canadian Technical Report of Fisheries and Aquatic Sciences 1713.

Hollander, M., and D. A. Wolfe. 1973. Nonparametric statistical methods. Wiley, New York.

Kennedy, P. 1979. A guide to econometrics. Massachusetts Institute of Technology Press, Cambridge.

MacCall, A. D., R. A. Klingbeil, and R. D. Methot. 1985. Recent increased abundance and potential productivity of Pacific mackerel (*Scomber japonicus*). California Cooperative Oceanic Fisheries Investigations Reports 26:119–129.

Noreen, E. W. 1989. Computer-intensive methods for testing hypotheses. Wiley, New York.

Norton, J. 1987. Ocean climate influences on groundfish recruitment in the California Current. Proceedings of the international rockfish symposium. University of Alaska, Alaska Sea Grant Report 87-2, Anchorage.

Parrish, R. H., and A. D. MacCall. 1978. Climatic variation and exploitation in the Pacific mackerel fishery. California Department of Fish and Game, Fish Bulletin 167.

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

Prager, M. H., and J. M. Hoenig. 1989. Superposed epoch analysis: a randomization test of environmental effects on recruitment with application to chub mackerel. Transactions of the American Fisheries Society 118:608–618.

Prager, M. H., and A. D. MacCall. 1988. Revised estimates of historical spawning biomass of the Pacific mackerel, *Scomber japonicus*. California Cooperative Oceanic Fisheries Investigations Reports 29: 81–90.

Sinclair, M., M. J. Tremblay, and P. Bernal. 1985. El Niño events and variability in a Pacific mackerel (*Scomber japonicus*) survival index: support for Hjort's second hypothesis. Canadian Journal of Fisheries and Aquatic Sciences 42:602–608.

Snedecor, G. W., and W. G. Cochran. 1980. Statistical methods, 7th edition. Iowa State University Press, Ames.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry, 2nd edition. Freeman, New York.

Received April 17, 1990
Accepted August 1, 1991