Transactions of the American Fisheries Society 118:608–618, 1989

Superposed Epoch Analysis: A Randomization Test of Environmental Effects on Recruitment with Application to Chub Mackerel

MICHAEL H. PRAGER

Department of Oceanography, Old Dominion University
Norfolk, Virginia 23529, USA

JOHN M. HOENIG[1]

Cooperative Institute for Marine and Atmospheric Studies
Rosenstiel School of Marine and Atmospheric Science
University of Miami, 4600 Rickenbacker Causeway
Miami, Florida 33149, USA

Abstract.—Superposed epoch analysis, a nonparametric technique, can be used to test the statistical significance of associations between extreme environmental events and recruitment success. A test statistic, similar to a paired t-statistic, is used to compare recruitment in years with extreme events to recruitment in the immediately surrounding years. Because statistical significance is determined by a randomization test, superposed epoch analysis does not rely on the usual assumptions (random sampling, normality, homogeneity of variance, independence of observations) of parametric testing. Thus, the method can be used when regression analysis, correlation, or a t-test would be inappropriate. As an example, we tested the association between elevated sea level (often associated with El Niño events) and high recruitment success of chub mackerel *Scomber japonicus* off the coast of southern California. The association was statistically significant ($P <$ 0.01) for the period preceding the collapse of the chub mackerel fishery in the late 1960s but not significant (P = 0.59) over the entire time series. This change may be due to statistical artifacts, a nonlinear relationship between sea level and recruitment, or biological causes. As with other statistical methods, a valid hypothesis test requires a priori formulation of the null hypothesis. Within this limitation, superposed epoch analysis is a useful method for conducting significance tests on autocorrelated time series, such as recruitment data.

It has often been noted (e.g., Parrish and MacCall 1978; Ulanowicz et al. 1982) that environmental events, such as variations in climate, seem to be associated with patterns in time series of recruitment data, and it is also possible that pollution incidents affect recruitment (Vaughan et al. 1986). However, few statistical methods have been developed for testing hypotheses about such associations. Although they have been used, parametric methods such as the t-test require assumptions not usually met by time-series data. The lack of appropriate methodology is surprising because a major emphasis of fisheries research has been to understand the causes of recruitment variability and, in doing so, to evaluate the role of environmental factors in recruitment dynamics.

In this paper, we use superposed epoch analysis (Haurwitz and Brier 1981) to test for a significant relationship between the occurrence of a specified type of event (a key event) and the strength of recruitment in corresponding years (the key response). We use the population of chub mackerel *Scomber japonicus* (locally called Pacific mackerel) in the northeast Pacific as an example for two reasons. First, a long time series of data is available, with a gap of 6 years near the end. Thus, the series forms a representative test case for application of the method. Second, recruitment data on this species were analyzed qualitatively by Sinclair et al. (1985), who suggested that El Niño–Southern Oscillation events, as indicated by high sea level, are associated with years of strong recruitment. It is therefore of interest to apply a new, quantitative procedure to the data set, which has been extended and revised (Table 1; Parrish and MacCall 1978; MacCall et al. 1985; Prager and MacCall 1988), to gain further understanding of the importance of climatic events to this species.

Development of Null Hypotheses and Test Statistics

The first step in using superposed epoch analysis (as with any statistical test) is to formulate

---

[1] Present address: Science Branch, Department of Fisheries and Oceans, Post Office Box 5667, St. John's, Newfoundland A1C 5X1, Canada.

TABLE 1.—Sea level anomalies[a] and chub mackerel survival and recruitment off the coast of southern California, 1929–1983. The first-year survival index and recruitment biomass are related to year of spawning, e.g., the table entry for 1929 recruitment occurred in 1930.

Spawning year	Sea level anomaly[a]	First-year survival index[b]		Recruitment biomass (kilotonnes)[c]	
		$\log_e$	Linear	$\log_e$	Linear
1929	0.175	−0.354	0.702	4.652	104.821
1930	0.902	−0.303	0.739	4.726	112.873
1931	0.453	−0.805	0.447	4.482	88.431
1932	0.451	−1.474	0.229	4.149	63.372
1933	−1.624**	−2.061	0.127	3.670	39.258
1934	−0.170	−2.559	0.077	3.107	22.346
1935	−0.238	−2.080	0.125	3.376	29.248
1936	0.311	−1.011	0.364	4.051	57.476
1937	0.217	−0.946	0.388	3.720	41.267
1938	−0.149	−0.399	0.671	4.161	64.149
1939	0.152	−0.634	0.530	3.872	48.039
1940	0.353	−0.307	0.736	3.851	47.019
1941	2.093**	0.349	1.417	4.372	79.229
1942	0.706	−0.423	0.655	3.591	36.256
1943	0.658	−0.916	0.400	3.448	31.439
1944	−0.087	−1.083	0.339	3.274	26.427
1945	−0.398	−1.831	0.160	2.174	8.795
1946	−0.092	−1.582	0.206	2.211	9.122
1947	−0.555	0.877	2.404	4.145	63.139
1948	−0.287	0.898	2.454	3.471	32.170
1949	−0.533	−1.866	0.155	1.626	5.083
1950	−0.694	−3.216	0.040	0.487	1.628
1951	1.197*	−2.941	0.053	0.546	1.725
1952	−0.050	0.521	1.683	3.223	25.107
1953	−0.409	2.306	10.029	3.877	48.297
1954	0.046	0.174	1.190	3.036	20.818
1955	−0.973*	0.125	1.133	3.725	41.463
1956	−0.421	−1.556	0.211	2.208	9.095
1957	1.491*	−1.015	0.362	2.542	12.702
1958	1.971**	0.909	2.483	3.729	41.646
1959	1.199*	0.737	2.089	3.461	31.833
1960	−0.547	0.714	2.042	4.067	58.363
1961	−0.875	0.051	1.053	3.659	38.827
1962	−0.831	−1.991	0.137	2.110	8.248
1963	0.053	−2.868	0.057	1.469	4.345
1964	−1.578**	−3.737	0.024	0.192	1.212
1965	0.406	−2.353	0.095	0.647	1.911
1966	0.102	−3.423	0.033	−1.446	0.235
1967	−0.322	−1.282	0.278	−0.727	0.484
1968	−0.609	0.461	1.586	0.747	2.111
1969	0.197				
1970	−1.102*				
1971	−0.679				
1972	0.952				
1973	−1.701**				
1974	−0.290			1.527	4.605
1975	−1.521**	1.390	4.014	0.136	1.146
1976	0.273	3.320	27.669	4.140	62.790
1977	−1.218*	1.377	3.962	3.708	40.762
1978	−0.151	0.976	2.653	4.740	114.395
1979	−0.921	−1.547	0.213	2.776	16.059
1980	−0.298	−0.320	0.726	4.432	84.104
1981	−0.417	−0.740	0.477	4.276	71.926
1982	1.506*	−2.779	0.062	2.183	8.871
1983	3.876**	−4.331	0.013	0.894	2.444

---

[a] Sea level anomalies are given as standard deviations from the grand mean for sea level during 1929–1983. Asterisks indicate high or low sea levels (≥1.0 SD* above or below the grand mean) and very high (≥2.0 SD** above the grand mean) or very low (≥1.5 SD** below the grand mean) sea levels.

[b] Recruitment biomass divided by spawning biomass of parent stock.

[c] Year-class biomass at age 1.

the null hypothesis and the test statistic. One can then determine the distribution of the test statistic when the null hypothesis is true. Parrish and MacCall (1978) defined a first-year survival (spawning success) index for the species. We call this index I and define it:

$$I = R'/P'; \tag{1}$$

R' is the recruitment biomass (the year-class biomass at age 1), and P' is the spawning biomass of the parent stock. Sinclair et al. (1985) suggested, after reviewing data from 1928 through 1965, that the survival index tends to be high in years of high sea level (Figure 1). In the following sections, we illustrate a statistical test for an association between years of elevated sea level, often associated with El Niño events, and high survival indices. We define the null hypothesis to be

- H_0: There is no association between elevated sea level and survival indices.

We define the alternative hypothesis to be

- H_a: Years of elevated sea level are associated with high survival indices.

The test is conducted by comparing the survival index for key-event years to indices for adjacent background years. Background years are defined for this application as the 2 years before and the 2 years after a year of elevated sea level.

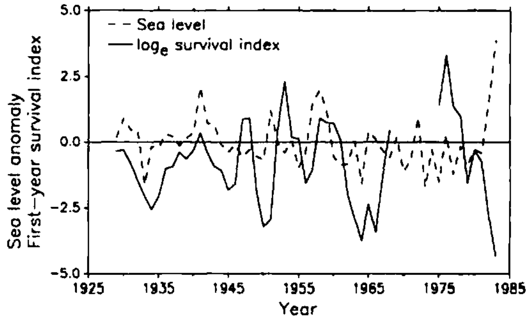

FIGURE 1.—Sea level anomalies (standard deviations from the long-term mean) and natural logarithm of first-year survival index of chub mackerel plotted as a function of year. Sea level is the mean of values from San Francisco, Los Angeles, and San Diego, California, in March–December of each year. The survival index is defined as the year-class biomass at age 1 divided by the spawning biomass. Spawning biomass incorporates weight at age, maturity at age, and number at age of the recruited (age-1 and older) population.

The Superposed Epoch

A comparison between background and key-event years can be pictured as a superposed epoch (Haurwitz and Brier 1981), an analytical tool used in meteorology and fluid dynamics. To construct the superposed epoch, we first define key-event years as years in which sea level was high. The survival indices for key-event years and background years are then arranged in a table with one row for each key event and five columns containing values of I for the 2 years before the key-event year, the key-event year, and the subsequent 2 years, in that order (Table 2). (This assumes that key events affect recruitment contemporaneously.) The five column means constitute the superposed epoch: when plotted (as in Figure 2), they provide a visual comparison of expected survival indices for key-event years and those for background years. The number and distribution of background years in each row, the designation of key-event years, and the temporal relationship of the key-response year in the superposed epoch to the key-event year in the independent variable (e.g., the same year, the following year, etc.) should be made a priori. The statistical tests are not sound if these choices are made by examining the response data.

Test Statistics

The usual parametric approach to comparing two means (e.g., the key-event versus background survival indices in the superposed epoch) would be the t-test. The assumptions for a valid t-test include random sampling, normality, homogeneity of variance, and independence of observations. (The same assumptions would apply to a parametric regression or correlation analysis.) Any of these assumptions might fail in the current circumstances, the last assumption in particular. (For the survival indices in Table 1, the lag-one autocorrelation coefficient is 0.593.) When the assumptions are not met, a t-statistic computed from the data in Table 2 is not expected to follow a Student's t-distribution under the null hypothesis. After performing a Monte Carlo study of a similar application, Haurwitz and Brier (1981) rejected the use of the parametric t-test because it provided incorrect significance levels. Thus, we will use other methods (described later) to determine the null distribution of the test statistic.

Study of a given null hypothesis may suggest several test statistics, and in our case at least three statistics appear potentially useful. The first statistic, D, is the difference between the mean survival index for key-event years ($\bar{E}$) and the mean index for background years ($\bar{B}$):

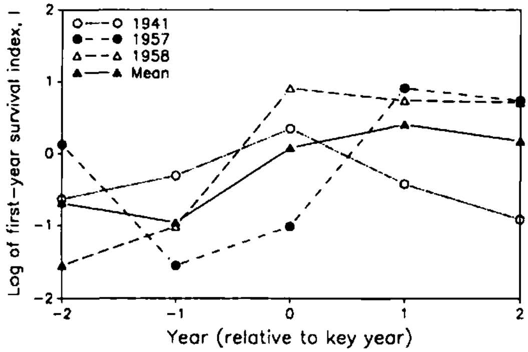

FIGURE 2.—Demonstration of superposed epoch analysis. Key events are the three years (1941, 1957, and 1958) with high or very high sea levels (at least 1.0 or 2.0 SD above the 1929–1984 mean, respectively). The response variable is the natural logarithm of the survival index for chub mackerel (Figure 1). The superposed epoch comprises the mean values of the survival index (column means in Table 2) in the years before, during, and after the key events. Log = $\log_e$.

$$D = \bar{E} - \bar{B}. \tag{2}$$

The second statistic, T, was used by Haurwitz and Brier (1981); we have made a minor correction for degrees of freedom. It is computed as one would compute an ordinary Student's t-statistic:

$$T = (\bar{E} - \bar{B})/S. \tag{3}$$

The pooled standard deviation estimate S is also computed as for a parametric t-test:

$$S = \left[\frac{(SS_E + SS_B)(N_E + N_B)}{N_E N_B (N_E + N_B - 2)}\right]^{1/2}; \tag{4}$$

N_E is the number of key events, N_B is the total number of survival indices for all background years, and SS_E is the sum of the squared deviations of survival indices for N_E key-event years from their mean. Therefore,

$$SS_E = \sum_{i=1}^{N_E} (E_i - \bar{E})^2; \tag{5}$$

E_i is the survival index for key-event year i. Similarly, SS_B is the sum of the squared deviations of N_B survival indices for background years from their grand mean.

The W-statistic, which is new, is analogous to Student's t-statistic computed for paired data. Each key-event year is compared with its own background years:

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

S_W is a measure of dispersion (defined shortly), and $\bar{d}$ is the mean of all paired differences between the survival index for key-event year i and the survival index for each corresponding background year (B_{ij}; $j \in [-2, -1, 1, 2]$). More precisely, we define $\bar{d}$ as

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

n_i is the number of background survival indices available for key event i. Generally, $n_i = 4$, but it is less than 4 when the key event occurs near a missing value of the response or near either end of the time series. In other words, missing values in recruitment are accommodated by eliminating them from computation of the test statistic and by adjusting n_i and N_B accordingly. The measure of dispersion S_W is computed as for a paired t-test:

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

How can we choose among these three statistics? We know of no simple method for determining which statistic provides the best test; it may be that none is uniformly most powerful or unbiased. By analogy to parametric testing, we suspected that the paired t-analog W would be more powerful than the unpaired t-analog T, and that T would be more powerful than the difference statistic D. Although we have not undertaken a de-

TABLE 2.—Construction of a superposed epoch from first-year survival indices[a] for chub mackerel. Index data are $\log_e$-transformed. Key-event years ($i = 0$) are those in which mean sea level off the California coast for March–December was at least 1.5 SD above the grand mean for 1929–1965. Background years ($i = -2, -1, 1$, and 2) are the two years immediately before and the two years immediately after a key-event year. The last row of the table constitutes the superposed epoch.

Key-event year or mean	$\log_e$ survival index for year i after key year				
	$i = -2$	$i = -1$	$i = 0$	$i = 1$	$i = 2$
1941	−0.634	−0.307	0.349	−0.423	−0.916
1957	0.125	−1.556	−1.015	0.909	0.737
1958	−1.556	−1.015	0.909	0.737	0.714
Mean	−0.688	−0.959	0.081	0.408	0.178

[a] Recruitment biomass divided by spawning biomass of parent stock.

tailed power study, a brief preliminary study supported the suspected ranking. We thus use the W'-statistic in this application.

Determining the Null Distribution of the Test Statistic

In a significance test, one decides whether the observed value of a test statistic is sufficiently extreme for rejection of the null hypothesis at the chosen significance level. To make this decision, one must know the null distribution of the test statistic (i.e., its distribution under the null hypothesis). For a parametric test, the null distribution is usually obtained from a table and is conditional on the assumptions of the test. For the nonparametric tests that we propose, null distributions are obtained not from a table, but from randomization tests.

Randomization Test

Assume that four key events occur during a time series. According to the null hypothesis, key events have no association with recruitment success; thus, any year might be called a key-event year. Using this insight, we can determine the null distribution of W' for the data set in question by computing W' for each possible set of positions of the four key years in the time series. The resulting frequency distribution of W' describes the exact null distribution of W' for these data. To test the significance of the observed W' (i.e., the value computed from the observed location of key events), we note where the observed value is located along this null distribution. If the observed W' is above the $100 \cdot (1 - P_\alpha)$th percentile, we reject the null hypothesis at P_α (for an upper one-tailed test).

The chub mackerel data (Table 1) include 49 years of observations of the first-year survival index. A full randomization test would require computation of 211,876 values of W' (the number of combinations of 49 items taken four at a time). If the number of key events were higher, the computations would be much increased. With five key events in 49 years, the number of computed W' values would be 1,906,884; with six events, almost 14 million. When computations for the full randomization method become excessive, one alternative is to compute the null distribution over a systematic subgroup of allocations (Edgington 1986, 1987). Another possibility is to determine the null distribution by Monte Carlo methods.

Monte Carlo Randomization

Monte Carlo simulation can be used to determine the null distribution of a test statistic to an arbitrary degree of precision. In this variant of the randomization method, we sample with replacement (rather than survey exhaustively) the sets of possible locations of key-event years. Repeatedly selecting random sets of positions for the key events (and thus the corresponding background years), we compute W' for each set. Significance level is computed as for the full randomization method. For example, an upper one-tailed test might be based on ν randomly generated values of W'. If x of the randomized values of W' were larger than the observed W', the estimated P of the observed W' would be $(x + 1)/(\nu + 1)$ (Edgington 1986).

Figure 3 illustrates an analysis (described later) of the chub mackerel data in which the observed W' (3.70) was on the right-hand tail of the null distribution. Of the 10,000 values of W' computed, 107 were larger than 3.70; the probability of a larger W' under the null hypothesis was 108/10,001 = 0.011. With P_α at 0.05, the null hypothesis would be rejected.

It is difficult to know a priori how many Monte Carlo trials are required for adequate precision in the results. Edgington (1987) has published formulae for the 99% probability interval that results from a given number of trials and a given P; these formulae can provide some guidance. An estimate of P achieved after a given number of trials can be obtained by the method of batch means (Law and Kelton 1982). A third technique, used here, is to plot the estimated P of the test statistic against the number of trials and to observe whether the results have stabilized within the specified number of trials (Figure 4).

With either full randomization or Monte Carlo methods, it may be necessary to restrict what constitutes a valid set of key events. Within each trial, the locations should be chosen without replacement (i.e., two key events should not have the same location in a single Monte Carlo or randomization trial). In addition, one may wish to discard sets in which key events occur closer together than some minimum spacing. The exact restrictions will depend on the null hypothesis being tested and the nature of the key events.

Reflected-Event Analysis: A More General Test

So far, we have assigned years to one of two categories: key-event years and non-key-event years. We may wish to consider a third category: anti-key-event years. For example, Sinclair et al. (1985) suggested that not only does the first-year survival index for chub mackerel tend to be high in years of high sea level, but the index also tends

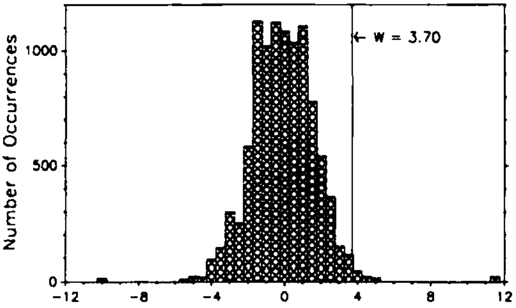

FIGURE 3.—Distribution of the W-statistic under the null hypothesis of no association between chub mackerel recruitment and sea level. The analysis was of survival index data of chub mackerel from 1929 to 1968 and corresponding sea level data (Table 1). Two key events were in 1941 and 1958. The distribution of the W-statistic was determined by Monte Carlo simulation. The vertical line indicates the observed value of the test statistic, which corresponds to $P = 0.011$.

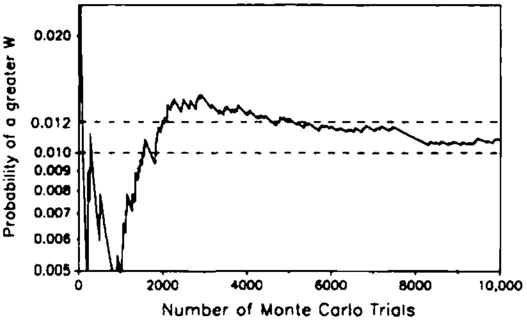

FIGURE 4.—Estimated significance level P of the observed value of the W-statistic as a function of the number of Monte Carlo trials. This illustration corresponds to the same analysis as Figure 3: a time series of chub mackerel recruitment from 1929 to 1968, with key events in 1941 and 1958. Because of the figure's scale, the estimated significance appears quite variable; however, the dashed horizontal lines illustrate that after about 5,000 trials, P remained between 0.010 and 0.012. The estimated P for this experiment was 0.011.

to be low in years of low sea level. For examination of this proposition, years of low sea level might be considered anti-key-event years. Consider the following pair of hypotheses:

- H_0: Survival index is unrelated to extremes in sea level;
- H_a: Survival tends to be high when sea level is high, survival tends to be low when sea level is low, or both.

We have devised reflected-event analysis, a new form of superposed epoch analysis, to address such pairs of hypotheses. The corresponding test statistic, R, is a modification of the W-statistic but is computed from the reflected paired differences d_{ij}^R between survival indices of a key-event year and its background years:

$$d_{ij}^R = \begin{cases} E_i - B_{ij}, & \text{if sea level is high in year } i; \\ B_{ij} - E_i, & \text{if sea level is low in year } i. \end{cases} \qquad (9)$$

The expected value of such paired differences is positive when the alternative hypothesis is true. The modified test statistic, R, is then

$$R = \frac{\bar{d}_R N_B^{1/2}}{S_R}; \qquad (10)$$

$\bar{d}_R$ is the mean of the N_B values of d_{ij}^R. The measure of variability, S_R, is analogous to S_W (equation 8) and is defined as

$$S_R = \left[\frac{1}{N_B - 1} \sum_{i=1}^{N_I} \sum_{j=1}^{n_i} (d_{ij}^R - \bar{d}_R)^2\right]^{1/2}. \qquad (11)$$

The null distribution of R can be found by the full randomization method or by the Monte Carlo randomization method, as with the W-statistic. However, significance tests of R are always one-tailed.

Application to Chub Mackerel

Using the techniques described above, we analyzed the relationship between chub mackerel recruitment and extremes of sea level. Our first set of analyses was performed with the natural logarithm of the survival index (equation 1) as the response variable. This index or its logarithm has been used in prior recruitment studies (Parrish and MacCall 1978; Sinclair et al. 1985) and can be derived from a multiplicative model of first-year survival. A second set of analyses was performed with the logarithm of recruitment biomass as the response variable. We performed this second set of analyses because recruitment rather than first-year survival is often the variable of ultimate concern. In each case, we used the logarithmic transform to give constant weight to proportional changes in the population.

Data

Recruitment and survival index data (Figure 1; Table 1) were taken from three sources. For 1925–1965, we used data from Prager and MacCall (1988). For 1965–1968, we used data from Parrish and MacCall (1978). Estimates from the latter source are within a few percent of the results of Prager and MacCall (1988) throughout the 1960s. For 1974–1983, we used data from MacCall et al. (1985), interpolated to reflect population sizes at May 1 rather than at July 1 of each year. To compute consistent biomass estimates for these final years, we used the weight and fecundity vectors of Prager and MacCall (1988). In all cases, we computed first-year survival with equation (1).

Sinclair et al. (1985) used sea level as an indicator of reduced southward transport by the California Current during an El Niño–Southern Oscillation event. Lacking a direct measure of transport, we have done the same. Although the low-frequency sea level record is undoubtedly associated with El Niño–Southern Oscillation events, we do not claim that our data are a measure of El Niño, which is defined in terms of conditions off Ecuador and Peru (Quinn et al. 1987). We know of no definitive survey of El Niño–Southern Oscillation effects in the northern hemisphere. (Quinn et al. 1978 surveyed large-scale events, but not over the entire period of interest.) Our analyses, although undoubtedly showing the influence of El Niño–Southern Oscillation events and the California Current, used sea level per se to define key events.

We used only sea levels from March through December of each year and thus included the month before, the month of, and the 7 months after the main spawning period for chub mackerel. The data are from Prager and MacCall (1987a, 1987b), who deseasonalized the data, estimated missing values, and standardized the series for each month. After computing the 9-month mean for each year, we standardized the resulting series to a mean of 0 and a standard deviation of 1. In the balance of this paper, the term sea level refers to this standardized 9-month annual mean.

Analyses

We used the W-statistic with Monte Carlo trials to test association of the responses (first-year survival and recruitment biomass) with two categories of years with extreme sea levels. The first category, which we call "very high sea level years," had sea levels at least 2.0 SD above the 1929–1984 mean. The second category, "high sea level years," had sea levels at least 1.0 SD above the mean. High and very high sea level years are indicated in Table 1.

In reflected-event analyses, we used the R-statistic to test the association of the two recruitment responses with high sea level years, very high sea level years, low sea level years, and very low sea level years. Low or very low sea level years, as indicated in Table 1, were those with sea levels at least 1.0 and 1.5 SD less than the mean, respectively. (Because no years had sea levels of -2.0 SD or less, we could not use a criterion symmetrical to that for very high sea level years.)

Because Sinclair et al. (1985) discussed the relationship between sea level and survival index for the years 1928–1965, we first applied our analyses to the data through 1965. We then applied the same statistical tests to the entire range of available data by adding data for 1966–1968 and 1975–1983.

Results

The association between very high sea level and high first-year survival index was significant (P = 0.011) when only the period from 1929 to 1968 was considered; however, when the full series was considered, the association was no longer significant (Table 3). This appears to be due to poor and declining survival during the extremely strong 1982–1983 El Niño.

Using reflected-event analysis, we found a significant association (P < 0.002) between very low sea level and low survival and between very high sea level and high survival. More formally, we were able to reject the null hypothesis and adopt the alternative hypothesis that survival tends to be high when sea level is high, survival tends to be low when sea level is low, or both. For the larger data set (1929–1983), P for this association drops to 0.105. When we tested the association between less extreme sea level conditions and survival, the results were not statistically significant (Table 3).

The second set of analyses (of recruitment rather than first-year survival) gave similar results (Table 3). The only notable difference was that P for the association between very high sea levels and the response was only 0.078.

Discussion

Population Considerations

The apparently very strong relationship between the chub mackerel survival index and sea level described by Sinclair et al. (1985) is absent from the most recent data (1975–1983). The estimated P for the relationship changed from 0.011 to 0.593 when the larger data set was considered. What conclusions can be drawn from these results? It is not very unusual for an apparently strong

TABLE 3.—Results of superposed epoch analyses and reflected-event analyses on sea-level-based key events as related to first-year survival and recruitment biomass (at age 1) of chub mackerel. Epoch analyses are based on the W-statistic, and reflected-event analyses are based on the R-statistic. Key events for reflected-event analyses are marked + for key-event years and − for anti-key-event years. In all cases, P was computed by a Monte Carlo randomization procedure.

Type of analysis and sea level criterion[a]	Number of key or anti-key events		P	
	Short time series	Long time series	Short time series (1929–1968)	Long time series (1929–1983)
	First-year survival			
Epoch analysis				
Very high sea level	2	3	0.011	0.593
High sea level	5	7	0.578	0.823
Reflected-event analysis				
Very high or very low sea level	+2, −2	+3, −3	0.002	0.105
High or low sea level	+5, −3	+7, −5	0.351	0.596
	Recruitment			
Epoch analysis				
Very high sea level	2	3	0.025	0.484
High sea level	5	7	0.717	0.910
Reflected-event analysis				
Very high or very low sea level	+2, −2	+3, −3	0.078	0.138
High or low sea level	−5, −3	+7, −5	0.643	0.665

[a] Very high sea levels were at least 2.0 SD above the mean for the time series. High and low levels were at least 1.0 SD above and below the mean, respectively. Very low levels were at least 1.5 SD below the mean.

but spurious statistical relationship to disappear when more data become available. The best information available would normally be the complete time series (up to and including 1983); from this point of view, our information does not support the hypothesis of a significant association between high sea level and high first-year survival.

The survival index and recruitment values are derived from cohort analysis. It is well known that the results of cohort analysis are least reliable for the most recent years (Gulland 1965; Murphy 1965). Thus, the most recent data and the data for the years just before the population collapse in the late 1960s are the least reliable. Coincidentally, the most recent data are also those for which the hypothesized relationship breaks down. Thus, one might suppose that the variability in the data masks the true association between recruitment and sea level, and therefore that the large decline in first-year survival from 1980 to 1983 (Figure 1) is an artifact of the cohort analysis. For two reasons, we do not believe this to be the case. First, MacCall et al. (1985) performed the virtual population analysis with several trial values of final fishing mortality F; all values yielded this same pattern of decline. This led MacCall et al. (1985) to comment that the 1982 and 1983 year-classes appeared to be unusually weak with respect to the exceptionally large spawning biomasses that produced them. Second, the raw catch data do not support high recruitment from the 1983 spawning. The numbers of age-1 fish caught in 1982–1984 can be used as a rough index of recruitment resulting from spawning in 1981–1983. The numbers are 34 × 10^6, 20 × 10^6, and 3 × 10^6 fish for each year, respectively. These numbers suggest a strong spawning biomass in 1983 (from the preceding large year-classes), but a poor resultant recruitment. This interpretation is further strengthened by the small catch of age-0 fish (prerecruits) in 1983: only 0.6 × 10^6 fish versus 16.2 × 10^6 in 1981 and 6.1 × 10^6 in 1982.

An interesting possibility is that the relationship between sea level and recruitment is real, although not linear or even monotonic. Indeed, one would expect that some range of sea levels (and related oceanic conditions) is optimal for recruitment and that recruitment does not continue to increase indefinitely with increasing sea level. The 1982–1983 El Niño was the strongest in our data series. The negative results may indicate an alternative relationship between sea level and recruitment such that extremely high sea levels have a negative effect on recruitment. We do not have strong evidence for such a relationship, but it is consistent with the observed patterns in the data.

A final explanation of our failure to detect a significant relationship in the time series may be density-dependent biological controls on recruitment. The chub mackerel population attained record biomasses in the late 1970s and early 1980s (MacCall et al. 1985). The Ricker recruitment function, which has been applied to this stock (Parrish and MacCall 1978; MacCall et al. 1985), can be rewritten as

$$\log_e(R'/P') = \log_e(\alpha) - \beta P'; \qquad (12)$$

α and β are constants, R' is recruitment biomass, and P' is spawning biomass. The left-hand side of equation (12) is the logarithm of the first-year survival index (equation 1). Thus, if we believe that the stock follows Ricker recruitment dynamics (perhaps confounded with other effects, such as those of sea level), equation (12), at many values of β, is consistent with the low first-year survival in 1983. The large spawning stock in 1983 implied a high degree of density-dependent reaction (by the Ricker model), i.e., low first-year survival and recruitment. This illustrates that there can be plausible biological reasons for departure from a monotonic relationship when a wide range of conditions (e.g., stock size) is considered.

Statistical Considerations

Randomization tests have been employed by statisticians for at least 50 years but have not been widely used in fishery science. Nonetheless, such tests have great potential for dealing with difficult statistical problems, such as hypothesis tests of autocorrelated data. A valuable feature of randomization tests is that they can estimate valid P-values from nonrandom samples. As Edgington (1987) stated: "Parametric statistical tables are applicable only to random samples, and the invalidity of application to nonrandom samples is widely recognized." Although the results of randomization tests apply only to the samples analyzed and cannot be statistically generalized to populations, generalizations (at least qualitative generalizations) may be possible on nonstatistical (e.g., biological) grounds. In many applications, the statistical restriction may be of little importance. It might suffice to say that an association is statistically significant for certain observed data or that it is not significant for other data.

The bootstrap, another sample-reuse technique, has been the subject of much work by Efron and others (Efron 1982; Efron and Tibshirani 1986). Although asymptotically equivalent in many cases, the bootstrap and the randomization test differ conceptually. Bootstrapping is used to make inferences about population parameters under the assumption of random sampling and is not valid for nonrandom samples. Randomization tests are used to investigate relationships among variables within a sample, which need not be random. In a randomization test, the sample itself is considered the population of interest.

Superposed epoch analysis is quite general and could be adapted to many other problems. We modified the test statistic described by Haurwitz and Brier (1981) into one that recognizes the paired-like nature of superposed epoch analysis. The new statistic (W) may prove more powerful; it was also easily adapted to the more general three-category problem (low or high sea level, or neither). Because superposed epoch analysis does not depend on tabled probability distributions, further modifications and generalizations are possible without extensive theoretical work.

The potential uses of superposed epoch analysis are many. Pearce and Phillips (1988) examined the possible effects of El Niño–Southern Oscillation events on recruitment of the western rock lobster off Australia. Hurricanes, rain, drought, and discrete releases of pollutants (e.g., chemical spills) are key events whose significance can be tested in relation to biological responses. These responses might include recruitment, growth, mortality, migratory timing (Mundy 1982), and other aspects of life history. In addition, epoch analysis may prove useful for studying spatial patterns, such as the association between a physical quantity and a biological response along a transect. For example, zooplankton and particulate organic matter may be more abundant in convergence areas associated with Langmuir cells than elsewhere (Sutcliffe et al. 1963). To test for such an association, a continuous record of temperature and abundance could be obtained along a transect, and the occurrence of sharp, temporary changes in temperature could be considered as encounters with convergence zones (key events).

Superposed epoch analysis is somewhat unusual because it tests an association between a response variable and an explanatory variable only at the extreme or extremes of the explanatory variable's range. This may prove of value when data are measured with error (A. D. MacCall, National Marine Fisheries Service, personal communication). In such a case, extreme values of both independent and response variables are most likely to be categorized correctly, but the categorization of less extreme values is less certain.

Standard time-series analyses, such as autoregressive integrated moving-average (ARIMA) models (Box and Jenkins 1976) could also be applied to a recruitment time series to answer questions about effects of key events and to extract additional information about the dynamics of the population. However, standard algorithms for such models require many consecutive observations (50 are often suggested); they also require special techniques for handling missing values (Cook and Campbell 1979). These requirements impede the application of ARIMA models to many fisheries problems. Superposed epoch analysis, on the other hand, can be used with smaller samples containing missing values in both the response and explanatory variables; it is also simpler to perform and is easily explained to nonspecialists. It provides a potentially valuable alternative to standard time-series analyses. To facilitate further use of superposed epoch analysis, we have made available the FORTRAN computer program used for the work described here (Hoenig et al. 1989).

A final important remark on statistical inference must be made. A major restriction on most statistical tests, including superposed epoch analysis, is that the hypotheses must be framed before inspection of the response data. When this restriction is ignored (e.g., when examination of the data results in the selection of a seemingly significant factor for analysis), significance levels can be exaggerated (Snedecor and Cochran 1980).

We used the chub mackerel data as an example even though the origin of the null hypothesis is obscure. Although we tested the hypothesis by using additional data (from the years 1975–1983), our significance levels may be exaggerated. In the context of hypothesis testing, this is of no importance because the overall results were not statistically significant.

The assumptions necessary for correct application of superposed epoch analysis have not been published. The most important requirement seems to be that hypotheses be formed before examination of the response data. A related assumption is that the structure of the epoch correctly reflects the relationship of the response to the independent variable. In particular, the response to the key event is assumed to be of known duration and short compared to the length of the time series.

Acknowledgments

We thank Michael Fogarty, William Fox, Jr., Dennis Heisey, Alec MacCall, Victor Restrepo, and the anonymous reviewers for helpful comments. Partial support for this study was provided through the Cooperative Institute for Marine and Atmospheric Studies by National Oceanic and Atmospheric Administration (NOAA) cooperative agreement NA85-WCH-06134. Partial support was also provided by a contract with the NOAA National Ocean Service, Ocean Assessments Division, to the National Marine Fisheries Service, Southwest Fisheries Center; by Old Dominion University; and by the Canada Department of Fisheries and Oceans.

References

Box, G. E. P., and G. M. Jenkins. 1976. Time series analysis: forecasting and control, revised edition. Holden-Day, San Francisco.

Cook, T. D., and D. T. Campbell. 1979. Quasi-experimentation: design & analysis issues for field settings. Houghton Mifflin, Boston.

Edgington, E. S. 1986. Randomization tests. Pages 530–538 *in* S. Kotz and N. Johnson, editors. Encyclopedia of statistical sciences, volume 7. Wiley, New York.

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

Efron, B. 1982. The jackknife, the bootstrap, and other resampling plans. Society for Industrial and Applied Mathematics, Philadelphia.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science 1:54–75.

Gulland, J. A. 1965. Estimation of mortality rates. Annex to Arctic Fisheries Working Group report. International Council for the Exploration of the Sea, C.M. 1965, Document 3, Copenhagen. (Reprinted. 1983. In D. M. Cushing, editor. Key papers on fish populations. IRL Press, Oxford, England.)

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. B. Payton. 1989. Computer programs for investigating the effects of environmental events on a time series of recruitment. Canadian Technical Report of Fisheries and Aquatic Sciences 1713.

Law, A. M., and W. D. Kelton. 1982. Simulation modeling and analysis. McGraw-Hill, New York.

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.

Mundy, P. R. 1982. Computation of migratory timing statistics for adult chinook salmon in the Yukon River, Alaska, and their relevance to fisheries management. North American Journal of Fisheries Management 4:359–370.

Murphy, G. I. 1965. A solution of the catch equation. Journal of the Fisheries Research Board of Canada 22:191–202.

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.

Pearce, A. F., and B. F. Phillips. 1988. ENSO events, the Leeuwin current, and larval recruitment of the western rock lobster. Journal du Conseil, Conseil International pour l'Exploration de la Mer 45:13–21.

Prager, M. H., and A. D. MacCall. 1987a. An environmental data base describing coastal southern California in the years 1920–1984. Part I: procedures and summaries (revised). National Marine Fisheries Service, Southwest Fisheries Center, Administration Report LJ-86-31, La Jolla, California.

Prager, M. H., and A. D. MacCall. 1987b. An environmental data base describing coastal southern California in the years 1920–1984. Part II: data tables (revised). National Marine Fisheries Service, Southwest Fisheries Center, Administrative Report LJ-86-32, La Jolla, California.

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.

Quinn, W. H., V. T. Neal, and S. E. Antunez de Mayolo. 1987. El Niño occurrences over the past four and a half centuries. Journal of Geophysical Research 92(C13):14449–14461.

Quinn, W. H., D. O. Zopf, K. S. Short, and R. T. Kuo Yang. 1978. Historical trends and statistics of the Southern Oscillation, El Niño, and Indonesian droughts. U.S. National Marine Fisheries Service Fishery Bulletin 76:663–678.

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.

Sutcliffe, W. H., E. R. Baylor, and D. W. Menzel. 1963. Sea surface chemistry and Langmuir circulation. Deep Sea Research 10:233–243.

Ulanowicz, R. E., M. L. Ali, A. Vivian, D. Heinle, W. A. Richkus, and J. K. Summers. 1982. Identifying climatic factors influencing commercial fish and shellfish landings in Maryland. U.S. National Marine Fisheries Service Fishery Bulletin 80:611–619.

Vaughan, D. S., J. V. Merriner, and W. E. Schaaf. 1986. Detectability of a reduction in a single year class of a fish population. Journal of the Elisha Mitchell Scientific Society 102:122–128.

Received November 29, 1988
Accepted September 18, 1989