**Abstract.**–Aerial surveys designed to detect trends in the abundance of harbor porpoise *Phocoena phocoena* were conducted each autumn, 1986 through 1990. The number of porpoise seen per kilometer of survey effort was used as an index of abundance. Based on these surveys, an analysis of covariance was used to model porpoise abundance. Year was treated as a covariate, and factors which affected sighting rates were included as categorical variables. No significant changes were seen in the abundance of porpoise over the five survey years. Monte Carlo simulations were performed to determine the power of the ANCOVA to detect trends in abundance. We conclude that the ability to detect trends is poor if traditional levels of statistical significance ($\alpha = 0.05$) are used. A larger $\alpha$-error may be appropriate in the management context of this species and increases the power to detect trends. Additional survey years similarly improve the power to detect trends. Based on the results of the simulations, we suggest that power should be defined to include only the detection of the correct trend when two-tailed tests are employed.

Manuscript accepted 25 March 1991.
Fishery Bulletin, U.S. 89:367–377 (1991).

# Detecting Trends in Harbor Porpoise Abundance from Aerial Surveys Using Analysis of Covariance

**Karin A. Forney**
La Jolla Laboratory, Southwest Fisheries Science Center
National Marine Fisheries Service, NOAA, P.O. Box 271, La Jolla, California 92038

**Doyle A. Hanan**
California Department of Fish and Game, c/o Southwest Fisheries Science Center
La Jolla Laboratory, National Marine Fisheries Service, NOAA
P.O. Box 271, La Jolla, California 92038

**Jay Barlow**
La Jolla Laboratory, Southwest Fisheries Science Center
National Marine Fisheries Service, NOAA, P.O. Box 271, La Jolla, California 92038

Harbor porpoise *Phocoena phocoena* are caught incidentally during halibut fishing with gillnets along the central California coast (Diamond and Hanan 1986; Hanan et al. 1986, 1987; Barlow 1987; Barlow and Hanan 1990). To assess the potential impact of this fishery mortality, ship and aerial surveys have been used to estimate the abundance of harbor porpoise along the coast of California, Oregon, and Washington (Barlow 1988, Barlow et al. 1988). These authors showed that although aircraft can be used to survey a large area very quickly, abundance estimates from aerial surveys must be multiplied by a very large and uncertain correction factor to account for the majority of animals that will be underwater at any given instant. For this reason, ship surveys were concluded to be preferable for estimating absolute porpoise abundance.

The requirements are, however, less stringent if the only goal is to detect trends in the abundance of porpoise over time, rather than determining absolute abundance. The ability of aircraft to cover great distances relatively quickly and inexpensively makes them a logical platform for such surveys. If the fraction of animals detected from the air does not change over time, the correction factor becomes irrelevant, and indices of relative abundance can be used in place of absolute abundance measures.

We describe a series of five aerial surveys for harbor porpoise conducted in central California during autumn of 1986, 1987, 1988, 1989, and 1990. These surveys were designed specifically to detect changes in porpoise abundance. We used twin-engine aircraft to fly predetermined transect lines which zigzagged up the coast between Point Conception and the mouth of the Russian River (Fig. 1). Line transect methods were used with one observer on each side of the aircraft and a belly-port observer. A fourth person recorded information pertaining to sightings of porpoises and sighting conditions. Each year within the survey period, the transect lines were repeated 3–7 times, depending on weather conditions.

The number of porpoise seen per kilometer of search effort was used as a measure of relative abundance. A stepwise analysis of covariance procedure (ANCOVA) with year as the covariate was used to identify the best model describing porpoise seen

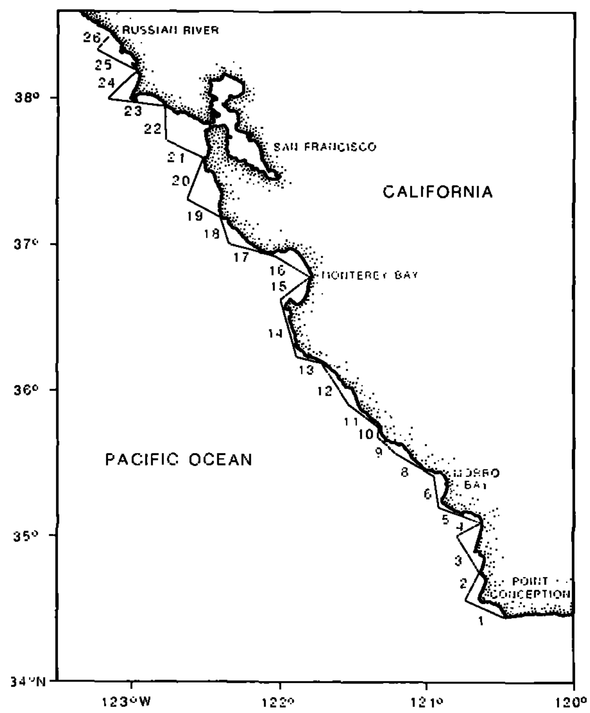


**Figure 1**
Flight transects for aerial survey of harbor porpoise in central California, 1986–90. Transect 7 was combined with transect 8 after 1986 and is not shown.

per kilometer. Standard ANCOVA F-ratio tests were applied to determine whether a significant trend with year is present. Monte Carlo methods were used to determine the power of this test to detect known trends in abundance.

## Methods

### Field methods

The surveys were established to monitor changes in abundance within the range of porpoise/gillnet fishery interactions (Point Conception to Russian River, California). Surveys were started only when there was a good likelihood of completing at least half of the survey (Point Conception to Monterey, or Monterey to Russian River) under good weather conditions. They were halted when viewing conditions deteriorated below acceptable levels (sea state higher than Beaufort 4 or 5, excessive dark cloud cover, rain, or fog).

A series of predetermined locations marking the beginning and end of each transect was entered into the aircraft's LORAN C navigational receiver to give the pilot a course to follow. The transects zigzagged in a generally northward progression between shore and roughly the 50-fathom (91-m) contour (Fig. 1). Sightings at the endpoint of a transect were very rare, and duplication at the start of the next transect did not occur. The transect lengths ranged from 5.2 to 44.8 km and averaged 24.8 km. The aircraft maintained an altitude of approximately 213 m and speeds of 90–100 knots (167–185 km/hour). To reduce sun glare, surveys were conducted only from south to north.

The surveys were flown in a twin-engine, high-wing, seven-passenger aircraft with the rear seat removed (Partenavia P68). Two observers sat behind the pilot and copilot seats and looked out the side windows; a third observer (belly observer) lay on the floor on his/her stomach just behind the right-side observer's seat and surveyed the water below the airplane through a 25 x 30-cm rectangular viewing port. Starting in 1988, the side windows were fitted with plexiglas bubble-type windows, allowing the side observers to see from the horizon to directly under the plane. This created an overlap with the belly observer's field of view; however, this did not result in double counting because the observers were in constant communication and discussed all possible sighting duplicates as they occurred.

The data recorder sat in the copilot position and recorded flight information, including location (latitude and longitude), time, weather (% cloud cover, Beaufort sea state, and sun position), viewing conditions, and porpoise sighting information. The data recorder entered weather and viewing conditions at the start of each transect and whenever conditions changed. Each observer subjectively evaluated viewing conditions as excellent, good, poor or "off effort," depending on estimated viewing depth into the water, sun glare, and sea state. To simplify the recording procedure and enhance accuracy of the data, a lap-top computer connected to the LORAN C navigational receiver replaced the hand-written flight log during the 1988–90 surveys.

The pilot, recorder, and observers communicated through headsets and voice-activated microphones. All communication was recorded on a central tape recorder. Additionally, each observer used a hand-held tape recorder for storage of individual sighting information. The two side observers used hand-held inclinometers to measure declination angles in degrees to the animals sighted. Due to space limitations, the belly observer could not use an inclinometer and estimated angles using marks applied to the viewing port. The observers changed positions approximately every 1–1.5 hours and between flights.

The observers actively searched (were "on effort") from start to finish of a transect, except when circling or when they declared themselves "off effort" because of poor sighting conditions. The pilot circled on porpoise sightings if there was any question about species identification or number of porpoise. Additional sightings made while circling were recorded as "off effort" sightings and were not included in the analyses.

During the first survey year (1986), observers reported all marine mammals sighted. However, the large number of California sea lion sightings took a disproportionate amount of time, so only harbor porpoise were recorded in 1987–90. Following the surveys, the data in the flight log or computer were checked for accuracy and, if needed, compared with the tape recordings. The data were transferred into microcomputer databases for summary and analysis.

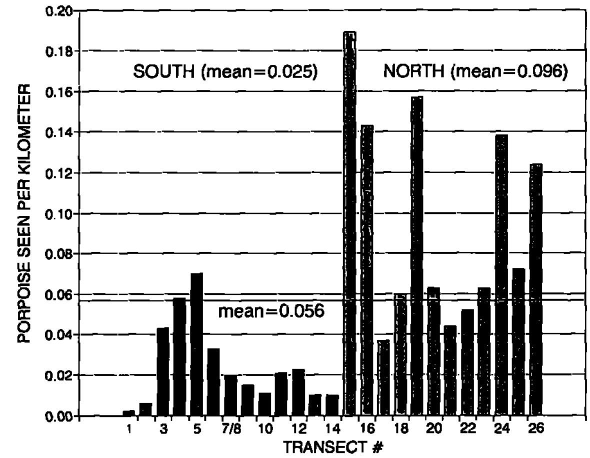


**Figure 2**

Porpoise seen per kilometer in transects 1 through 26 for 1986–90 surveys (including Beaufort sea states 0–3 and clear skies only). For the analysis, transects were divided into two areas at Point Pinos (between transects 14 and 15): south (low abundance) and north (high abundance).

## Analytical methods

Individual flight segments during which all sighting conditions were constant were combined to measure porpoise per kilometer in relation to each of the sighting variables. These variables included Beaufort sea state, cloud cover, viewing condition, individual observers, and an *a posteriori* geographic subdivision chosen on the basis of apparent porpoise abundance: south (low abundance) and north (high abundance) (Fig. 2). This subdivision was created to correct for slight interannual differences in survey effort for high- and low-density areas, caused by bad weather.

Cloud cover was recorded as a percentage and later coded into the categories "clear" (0–24%) and "cloudy" (25–100%). Sighting efficiency and sample sizes decreased dramatically when Beaufort sea state was higher than 3, so only segments with Beaufort 0–3 were used. Beaufort 0 was combined with Beaufort 1 because there was very little survey effort at Beaufort 0.

The data were fitted to an analysis of covariance (ANCOVA) model of the form:

$$P = \mu + \alpha_1 + \alpha_2 + \ldots + \delta(y - \bar{y}) + \varepsilon \qquad (1)$$

where P represents the log-transformed ($\log_e$) value of porpoise per kilometer, $\mu$ is the mean value of P, the $\alpha$ represent qualitative factors influencing observed porpoise abundance, $\delta$ represents the coefficient for the covariate year (y), $\bar{y}$ is the mean value of y, and $\varepsilon$ is a random error term. Such an additive model for logarithmic values is equivalent to a model describing multiplicative effects on the untransformed number of porpoise seen. This was deemed appropriate because sighting conditions affect the fraction of porpoise seen but not the absolute density of porpoise present. Because of the logarithmic transformation, a linear increase or decrease in the covariate would be interpreted as an exponential increase or decrease in porpoise abundance. The constant 0.001 was added to each value before transformation to avoid trying to take the logarithm of zero. This logarithmic transformation also made the data more nearly normal (Fig. 3).

It was not possible to include all potential variables in the model selection procedure, because this would have caused overstratification of the data. Individual observer effects were excluded because not all observers collected data each year, resulting in a large number of missing cell values. Viewing condition was also excluded because it is somewhat redundant with sea state and cloud cover and it is more subjective. Previous nonparametric tests of individual observer effects and viewing conditions with three years of data (Forney et al. 1989) yielded no significant differences in observed numbers of porpoise per kilometer.

In the ANCOVA, the data were weighted by the number of kilometers flown to correct for variability due to unequal sample sizes. A stepwise selection procedure with the SAS procedure GLM (Joyner 1985) was used to determine the best model for the observed data. At each step, all appropriate parameters and interaction effects were tested individually. The most significant parameter was added to the model, based on a criterion level of $\alpha = 0.05$. Each included variable was retested for significance at each subsequent step of the procedure.

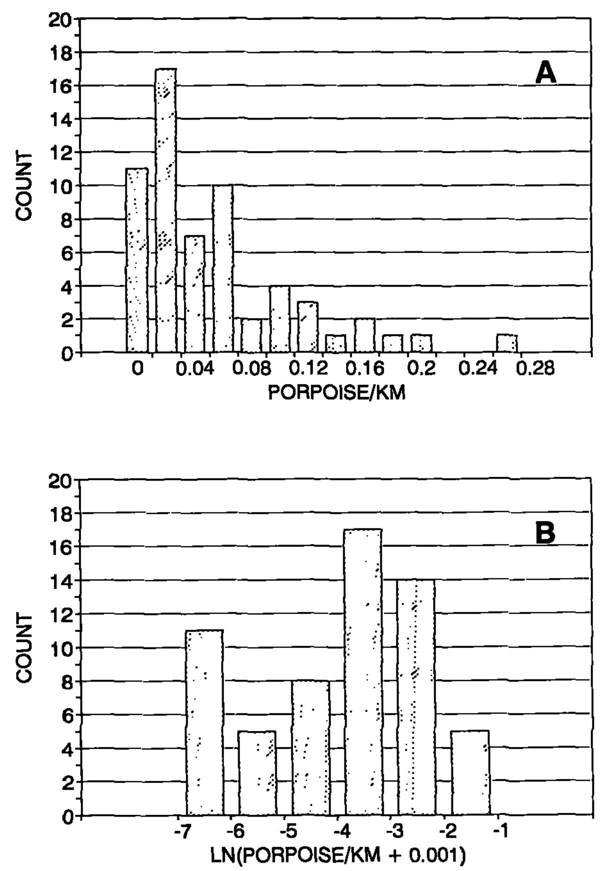


**Figure 3**
Distribution of observed porpoise/km values (A), and log-transformed porpoise/km values (B) for five years of aerial survey data. The transformation was ln(x + 0.001), where x is the observed number of porpoise/km.

**Table 1**
Average number of harbor porpoise seen per kilometer of search effort for 1986–90 aerial surveys. Numbers of porpoise seen per number of kilometers surveyed are given in parentheses. Categories are defined in the text. Overall mean number of porpoise per kilometer is 0.047 (796/16948).

| Area | Cloud cover | Beaufort 0,1 | Beaufort 2 | Beaufort 3 |
|---|---|---|---|---|
| South | Clear | 0.036 (75/2075) | 0.024 (54/2248) | 0.013 (24/1918) |
| | Cloudy | 0.030 (31/1026) | 0.015 (15/974) | 0.006 (3/518) |
| North | Clear | 0.163 (170/1042) | 0.106 (229/2165) | 0.046 (84/1814) |
| | Cloudy | 0.058 (30/521) | 0.039 (59/1530) | 0.020 (22/1117) |

## Simulation methods

Once the best model had been selected (see Results), Monte Carlo simulations were performed to determine the power of the ANCOVA to correctly detect a given trend in porpoise abundance. The analysis of power was divided into two main steps: (1) Simulations without a trend, to determine whether the procedure can create and correctly analyze simulated data; and (2) simulations with trends, to estimate how often the procedure correctly identifies a known trend in harbor porpoise abundance. Annual changes of ±5% and ±10% were tested over periods of five, six, eight, and ten years.

The random data sets were generated using the parameters and error structure obtained for the actual data from the best model (see Results). First, the expected logarithmic value of porpoise per kilometer for each combination of conditions was calculated from the fitted parameters. A random error term for each expected value was then drawn from a normal distribution with a mean of zero and standard error from the ANCOVA results of the best model. To allow weighted analysis of the simulated data, this error term was weighted inversely, i.e., multiplied times $\sqrt{(1/w)}$, where w is the number of kilometers flown under the given conditions. A set of 60 values for w, one for each of the 60 simulated porpoise-per-kilometer values, was obtained for each year by randomly selecting the actual numbers of kilometers flown from one of the five survey years. Complete yearly sets were chosen rather than individual values to avoid unlikely combinations of kilometers flown.

A yearly trend was incorporated into the simulation data by multiplying the calculated value of porpoise per kilometer times a factor representing the desired exponential change in porpoise abundance. To make the simulated data more like potential real data, all values were rounded to yield only integer values of porpoise over the given number of kilometers flown. In addition, to prevent unfeasible values of porpoise per kilometer, a new error term was drawn if the original one resulted in a value which was negative or greater than 0.4 porpoise per kilometer. The highest value observed in 1986–90 was 0.24 porpoise per kilometer; multiplying this value times the maximum simulated increasing trend yields an upper limit of approximately 0.4 porpoise per kilometer. Less than 5% of all error terms were redrawn in the simulations.

**Table 2**

Stepwise model selection procedure for 1986–90 aerial survey data. Parameters marked with an asterisk indicate variables included in the model at each step. P = ln(porpoise/km + 0.001); $\mu$ = mean value of P; BF = Beaufort sea state; AR = area; CL = cloud cover; YR = year; an × between letters indicates an interaction effect.

| STEP<br>Base model | 1<br>$P = \mu$ | 2<br>$P = \mu + AR$ | 3<br>$P = \mu + AR + CL$ | 4<br>$P = \mu + AR + CL + BF$ |
|---|---|---|---|---|
| P-values for tested variables | BF: 0.0587 | BF: 0.0079 | *BF: 0.0016 | YR: 0.8535 |
| | *AR: 0.0001 | *CL: 0.0021 | YR: 0.5865 | BF×AR: 0.8491 |
| | CL: 0.0256 | YR: 0.9628 | CL×AR: 0.3399 | BF×CL: 0.8378 |
| | YR: 0.9258 | | | CL×AR: 0.3875 |

## Results

### Survey results

A total of 16,948km of survey effort during 1986–90 resulted in 431 sightings, representing a total of 796 harbor porpoise. The overall mean number of porpoise per kilometer was 0.047. The average values of porpoise per kilometer over the five years are listed in Table 1 for different sighting conditions and areas. Mean group size was 1.85 porpoise, with a range of 1–10 and median 1.

### The model

The following model provided the best fit to the observed logarithmic estimates of porpoise per kilometer:

$$P = \mu + \alpha_{Bi} + \alpha_{Cj} + \alpha_{Ak} + \varepsilon_{i,j,k} \quad (2)$$

where $P = \log_e$ of [(porpoise/km) + 0.001],
$\mu$ = mean value of P,
$\alpha_{Bi}$ = effect of Beaufort sea state i on P,
$\alpha_{Cj}$ = effect of cloud cover j on P,
$\alpha_{Ak}$ = effect of area k on P,
$\varepsilon_{i,j,k}$ = normally distributed error with a mean of zero.

The model selection procedure is outlined in Table 2. None of the included variables lost significance and subsequently had to be dropped after inclusion of other variables. The yearly trend was not significant, so the model is essentially reduced to an analysis of variance (ANOVA) model. The results of the complete ANCOVA model testing for a yearly trend in the 1986–90 harbor porpoise data are shown in Table 3. The effects of area, cloud cover, and Beaufort sea state were significant ($P<0.0001$, $P<0.0006$, and $P<0.0021$, respectively), the yearly trend was not ($P = 0.8535$). None of the interaction effects were significant. The parameter estimates for the models with and without year are displayed in Table 4.

**Table 3**

Results of the weighted analysis of covariance.

| Source | df | Sum of squares | Mean square | F value | Prob. >F |
|---|---|---|---|---|---|
| Model | 5 | 15397 | 3079 | 10.46 | 0.0001 |
| Area | 1 | 9494 | 9494 | 32.24 | 0.0001 |
| Cloud cover | 1 | 3927 | 3927 | 13.34 | 0.0006 |
| Beaufort | 2 | 4095 | 2047 | 6.95 | 0.0021 |
| Year | 1 | 10 | 10 | 0.03 | 0.8535 |
| Error | 54 | 15900 | 294 | | |

**Table 4**

Parameter estimates from (A) the ANCOVA testing year, and (B) the 'best' model (ANOVA) chosen for the simulations. Standard errors are given in parentheses.

| Parameter | | (A) | | (B) | |
|---|---|---|---|---|---|
| $\mu$ | Mean | −2.1496 | (0.3288) | −2.1454 | (0.3251) |
| $\alpha_{B1}$ | Beaufort 0&1 | 0.0000 | — | 0.0000 | — |
| $\alpha_{B2}$ | Beaufort 2 | −0.0838 | (0.3304) | −0.0883 | (0.3266) |
| $\alpha_{B3}$ | Beaufort 3 | −1.1344 | (0.3516) | −1.1257 | (0.3453) |
| $\alpha_{A1}$ | Area 1 (South) | −1.5338 | (0.2701) | −1.5360 | (0.2674) |
| $\alpha_{A2}$ | Area 2 (North) | 0.0000 | — | 0.0000 | — |
| $\alpha_{C1}$ | Clear skies | 0.0000 | — | 0.0000 | — |
| $\alpha_{C2}$ | Cloudy skies | −1.0377 | (0.2841) | −1.0454 | (0.2787) |
| $\delta$ | Year | +0.0182 | (0.0982) | — | — |

### Analysis of power to detect trends

**No trend simulations ($\delta = 0$)** To determine the reliability of the simulation procedure to model trends in porpoise abundance, 500 simulated data sets with no yearly trend were created for five, six, eight, and ten survey years, using parameter set (B) in Table 4. The simulated data sets were analyzed using the full

**Table 5**
Actual $\alpha$-errors and fractions of positive and negative covariate coefficient estimates ($\delta$) for ANCOVA of 500 random data sets for five, six, eight, and ten simulated survey years.

| No. of years | % annual change | Actual $\alpha$-errors for $\alpha = 0.05$ | $\alpha = 0.10$ | $\alpha = 0.20$ | Fractions of $\pm\ \delta$ values |
|---|---|---|---|---|---|
| 5 | none | 0.05 | 0.10 | 0.20 | 0.48/0.52 |
| 6 | none | 0.06 | 0.11 | 0.21 | 0.54/0.46 |
| 8 | none | 0.04 | 0.10 | 0.20 | 0.49/0.51 |
| 10 | none | 0.07 | 0.12 | 0.23 | 0.53/0.47 |

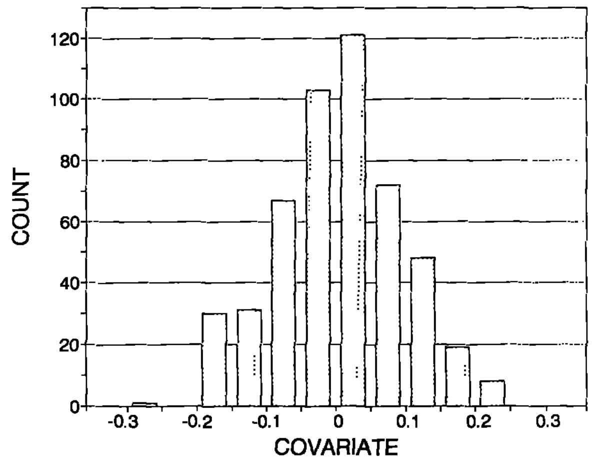


**Figure 4**
Distribution of covariate estimates ($\delta$) representing yearly change in abundance of harbor porpoise (from ANCOVA) for 500 simulations of five survey years with no annual trend in abundance.

ANCOVA model with a null hypothesis of no trend in abundance. The $\alpha$-error (Type I error) is the fraction of simulations which falsely detected a trend.

For all four simulations, the resulting $\alpha$-errors were close to the theoretical ones (Table 5). The average root mean-square-error term obtained for these data sets (16.63) was also close to the error for the actual data (17.01). The estimates of the covariate for year ($\delta$) in the simulated data were approximately normally distributed with a zero mean, as expected (Fig. 4). This confirms that the simulation procedures do not introduce substantial bias into the data or error structure.

***Simulations with trends ($\delta \neq 0$)*** To analyze the power of this procedure to detect given trends, random data sets spanning five, six, eight, and ten years were created with artificial changes in abundance of ±5% and ±10% per year. All other parameters were taken from Table 4, set (B), as above. For each combination of survey years and trend, 500 data sets were created and analyzed with the ANCOVA procedure.

In each simulation, a fraction of the analyses did not detect a trend: this represents the $\beta$-error (Type II error). A much smaller fraction detected a trend in the opposite direction of the true trend. The latter presents a special case (dilemma), and we have termed this type of error $\gamma$-error (Type III, cf. Carmer 1976). Figure 5 graphically illustrates $\alpha$, $\beta$, and $\gamma$ for a situation where an increasing trend is occurring and being tested against the null hypothesis in a two-tailed test (in a one-tailed test, $\gamma$ is zero). The three types of errors are interdependent: as $\alpha$ increases (i.e., the bars in Figure 5 move closer to zero), $\beta$ decreases, and $\gamma$ increases.

Power has been defined as the probability of correctly rejecting the null hypothesis when it is false, which numerically is $1-\beta$ (Rotenberry and Wiens 1985, Peterman 1990ab). However, this definition does not address the error associated with accepting a false alternate hypothesis ($\gamma$). In the case of trend analysis, this is the probability of rejecting the null hypothesis (no trend) in favor of a trend in the wrong direction. We therefore suggest that power be defined more precisely to include only the probability of detecting the correct alternate hypothesis, which numerically is $1-(\beta+\gamma)$.

Using this definition, the power to correctly detect trends in harbor porpoise abundance is displayed in Table 6 for six different levels of $\alpha$. The values listed under $\alpha = 1.0$ correspond to the fraction of the time that the sign of the covariate is correct, regardless of significance level. At this $\alpha$-level, the $\beta$-error is zero, because the null hypothesis of no trend is always rejected in favor of either an increasing or a decreasing trend. Both power and $\gamma$-errors are maximized when $\alpha = 1.0$ (see Discussion below).

At $\alpha = 0.05$, the ability to detect trends in abundance of harbor porpoise is poor (0.07–0.79) for all tested trends and numbers of survey years. This is below the level of power = 0.80 which has been suggested as a minimum standard (Skalski and McKenzie 1982, Peterman and Bradford 1987). Raising $\alpha$-levels improves the ability to detect trends, but also increases the chance of detecting a trend in the opposite direction of the true trend ($\gamma$-error). When $\alpha = 0.05$, $\gamma$-errors are less than 0.01 for the levels of change tested. In contrast, at $\alpha = 0.20$, $\gamma$-errors range from 0 to 0.05, and with $\alpha = 1.0$, $\gamma$-errors are between 0 and 0.33. Both $\beta$ and $\gamma$-errors are reduced with larger trends and more survey years.

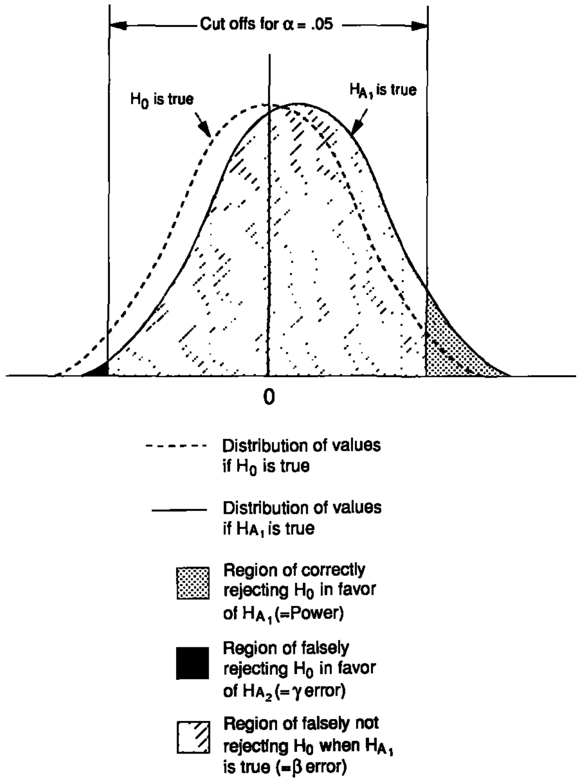


**Figure 5**

Graphic illustration of the errors associated with a two-tailed test, such as trend analysis. Solid line represents the distribution of coefficients for a hypothetical increasing trend. Dashed line represents the null distribution of coefficients (when there is no trend). Shaded areas represent the three error types, $\alpha$, $\beta$, and $\gamma$ (see text). $H_{A1}$ represents an increasing trend, $H_{A2}$ represents a decreasing trend, and $H_0$ represents no trend.

**Table 6**

Estimated power of the analysis to detect a given trend in harbor porpoise abundance correctly, based on 500 random data sets for each combination of change and number of survey years. Power is defined as $1-(\beta+\gamma)$.

| No. of years | % annual change | Power associated with given $\alpha$-error $\alpha$ = 0.05 | 0.10 | 0.20 | 0.30 | 0.40 | 1.0 |
|---|---|---|---|---|---|---|---|
| 5 | −10 | 0.17 | 0.24 | 0.36 | 0.43 | 0.52 | 0.82 |
| 6 | −10 | 0.23 | 0.35 | 0.52 | 0.60 | 0.68 | 0.90 |
| 8 | −10 | 0.52 | 0.63 | 0.74 | 0.82 | 0.86 | 0.98 |
| 10 | −10 | 0.79 | 0.87 | 0.92 | 0.95 | 0.97 | 1.00 |
| 5 | +10 | 0.11 | 0.18 | 0.29 | 0.39 | 0.47 | 0.78 |
| 6 | +10 | 0.21 | 0.32 | 0.45 | 0.54 | 0.63 | 0.89 |
| 8 | +10 | 0.46 | 0.56 | 0.70 | 0.80 | 0.85 | 0.97 |
| 10 | +10 | 0.74 | 0.82 | 0.88 | 0.93 | 0.96 | 1.00 |
| 5 | −5 | 0.08 | 0.14 | 0.22 | 0.28 | 0.35 | 0.70 |
| 6 | −5 | 0.10 | 0.16 | 0.29 | 0.38 | 0.44 | 0.72 |
| 8 | −5 | 0.17 | 0.27 | 0.42 | 0.50 | 0.59 | 0.84 |
| 10 | −5 | 0.27 | 0.37 | 0.55 | 0.65 | 0.71 | 0.91 |
| 5 | +5 | 0.07 | 0.13 | 0.20 | 0.27 | 0.34 | 0.67 |
| 6 | +5 | 0.10 | 0.16 | 0.25 | 0.34 | 0.40 | 0.73 |
| 8 | +5 | 0.15 | 0.23 | 0.35 | 0.45 | 0.53 | 0.83 |
| 10 | +5 | 0.25 | 0.34 | 0.48 | 0.57 | 0.66 | 0.89 |

## Discussion

The number of years necessary to detect trends in harbor porpoise abundance with the techniques described above will depend on two things: the rate of change to be detected, and the degree of certainty desired. A 5% annual change will be more difficult to detect than a 10% change over the same time period. If a large change must occur before the trend is detected, such methods may be of limited use in the management of populations, and more powerful techniques may be required.

If one does not need the ability to determine both increases and decreases, but merely wishes to determine whether a population is declining (objective 3, Peterman and Bradford 1987), one-tailed statistical tests can be used and will increase statistical power. Alternatively, if one is willing to accept a larger probability of inferring a trend when none is actually present, the power to detect trends can be improved by raising the level of $\alpha$ used to determine statistical significance.

It has been suggested that appropriate levels for $\alpha$ and $\beta$ should be determined based on the relative costs of committing each type of error (Toft and Shea 1983, Rotenberry and Wiens 1985, Hayes 1987, Peterman 1990b). If the cost of failing to detect a change in abundance is high relative to the cost of falsely detecting a trend for a stable population, then the traditional $\alpha$-level of 0.05 may be inappropriate. In such cases it may be preferable to minimize $\beta$-errors by increasing $\alpha$. For example, in the context of ecological monitoring, Hinds (1984) suggests that $\alpha$ should be made equal to $\beta$. However, it is important to remember that increasing $\alpha$ when power is low also raises $\gamma$ from virtually zero to potentially large levels. Rather than equalizing $\alpha$ and $\beta$, a tradeoff must be made between all three types of error. The magnitude of these errors can be estimated using simulations.

When $\alpha$ is raised to 0.10, ten years of data are sufficient to yield power greater than 0.80 and a $\gamma$-error of virtually zero when a 10% annual change is occurring. However, this corresponds to a very large total change in abundance (236% increase or 61% decrease). A smaller, but still substantial, change of 5% per year (total 155% increase or 37% decrease) would have a very low chance of being detected at this level of $\alpha$. If small changes are to be detected, then $\alpha$ may have to be set higher.

The most extreme form of raising $\alpha$-levels is accomplished by considering only the sign of the estimate for the covariate coefficient in the ANCOVA, thus setting $\alpha = 1$. In this case, the direction, rather than the presence, of a trend is tested for. This approach maximizes power, and may be an alternative for situations where power cannot be improved through other means (i.e., increasing the number of surveys conducted). For harbor porpoise trend estimation, the roughly equal fractions of positive and negative covariate coefficient estimates, $\delta$ (Table 5) indicate that such an analysis is not biased towards detecting either trend direction.

With $\alpha = 1.0$, power to detect the correct trend in harbor porpoise abundance ranges from 0.67 to 1.00 for 5–10 survey years and ±5% and ±10% annual population changes. Power of 0.80 or higher is achieved with $\alpha = 1.0$ after 5–6 survey years for a 10% annual change, or after 8 survey years for a 5% annual change. However, since the cost of low power in this case is a $\gamma$-error, power should be higher than when $\alpha$ is set at the traditional level of 0.05. In this case, eight survey years may provide high enough power to detect an annual 10% change, whereas even 10 years may not yield sufficient power to detect the smaller 5% annual change.

The magnitude of the $\gamma$-error when $\alpha = 1.0$ can be demonstrated with Figures 6 and 7. The three curves in these figures represent the distribution of covariate coefficients, $\delta$, for 500 simulated data sets with annual changes of (A) −10%, (B) 0%, and (C) +10%. The $\gamma$-error is represented by the area under curves A and C which lies on the incorrect side of zero. If this area is small or equal to zero, as when 10 annual surveys are conducted (Fig. 6), then the analysis will have a high probability of detecting the direction of a trend correctly. However, if the area is large, as when only five annual surveys are conducted (Fig. 7), then the procedure will not be able to detect the direction of trends accurately. The large degree of overlap between the three curves in Figure 7 also reflects the low power to detect trends. The dotted line marks the location of the covariate coefficient estimate ($\delta$) for the 1986–90 survey data. It is apparent that the estimate could reasonably come from any of the three distributions.

Setting $\alpha = 1.0$ is valid only if the costs of interpreting a nonexistent trend in a stable population are small in relation to the costs of failing to detect an existing trend. This may be the case if one needs to determine whether an existing level of take from a commercially exploited population is sustainable. The cost of not detecting a decreasing trend could be extinction of the population and the permanent loss of a resource. On the other hand, eliminating or reducing exploitation on a stable population which is incorrectly thought to be decreasing would cause smaller, short-term costs. In the case of marine mammals in the United States, existing laws mandate that all species be maintained at sustainable levels, so extinction represents an unacceptably high cost.

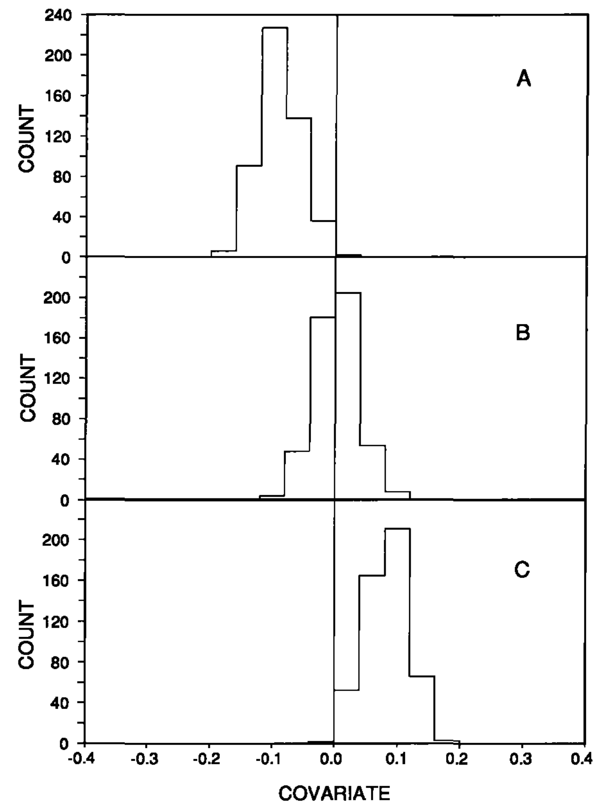


**Figure 6**
Distribution of covariate estimates ($\delta$) representing yearly change in abundance (from ANCOVA) for 500 simulations each of (A) 10% annual decrease, (B) no change, and (C) 10% annual increase in abundance over 10 survey years. Shaded area under curves A and C which lies on the incorrect side of zero represents the $\gamma$-error when $\alpha = 1.0$; here, it is essentially zero.

Several assumptions of the above procedures must be discussed. The most critical assumption is that the five years of data collected during 1986–90 characterize the level of variability expected in a longer time series. In addition, the results of the simulations are only accurate if the ANCOVA model is appropriate. The results indicate that the chosen model fits the data well ($P<0.0001$).

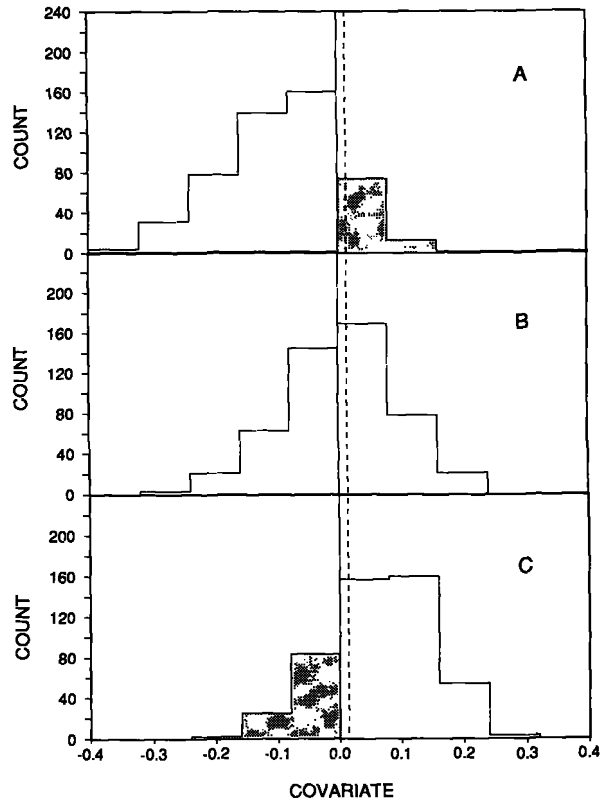


**Figure 7**
Distribution of covariate estimates ($\hat{\delta}$) representing yearly change in abundance (from ANCOVA) for 500 simulations each of (A) 10% annual decrease, (B) no change, and (C) 10% annual increase in abundance over five survey years. Shaded area under curves A and C which lies on the incorrect side of zero represents the γ-error when $\alpha = 1.0$. Dashed line marks location along the x-axis of the covariate estimate for actual 1986–90 harbor porpoise data.

The choice of adding the constant 0.001 in the log-transform may at first seem a bit odd, but in fact would be the same as the more familiar transformation $\ln(x + 1)$ if relative abundance had been defined as porpoise per thousand kilometers, rather than porpoise per kilometer. Several other constants were tested to determine if the choice of transformation might influence the analysis. The stepwise procedure yielded the same model in each case. The value 0.001 was chosen because it yielded the most normal distribution of porpoise per kilometer values (Fig. 3B).

This approach to trend analysis also assumes that the fraction of animals visible from the air does not change over time. The probability of detection can be influenced by many factors, particularly sighting conditions, porpoise behavior and group sizes, and observer differences. In our analysis, we controlled for sighting conditions by eliminating poor conditions and stratifying by the remaining ones. Changes in observers between years prevented tests of observer differences. However, based on previous tests with three years of data, they are not believed to be significant (Forney et al. 1989).

Harbor porpoise behavior, including frequencies of active versus inactive behaviors and mean group sizes, has been shown to vary by area and season (Calambokidis et al. 1990, Taylor and Dawson 1984, Sekiguchi 1987). To control for these potential differences, the surveys followed the same transect lines during the same season (autumn) each year. Nevertheless, group sizes in 1989 were significantly different than those in 1987 and 1988 (Kolmogorov-Smirnov test of cumulative distributions, $P = 0.02$ for both tests). The difference appears to be due to a larger percentage of groups containing three or more animals.

If group size affects harbor porpoise sightability, a substantial change in group size distribution could bias the trend analysis, either obscuring a present trend or creating a false one. To test for this possibility, the ANCOVA was repeated excluding the data for 1989. The overall results were similar, with the same final model, similar parameters, and no significant yearly trend ($P = 0.98$). We conclude that this slight difference in group sizes is not likely to have affected our analysis.

## Conclusion

The use of simulations allows researchers to estimate appropriate error levels for the analysis of surveys of animal populations. The ANCOVA model we used suggests that no trend in harbor porpoise abundance occurred between 1986 and 1990. However, our simulations show that the power of this model to detect trends using conventional α-levels of 0.05 or 0.10 is poor. Therefore, it is more correct to say that we could not reject the null hypothesis of no trend due to insufficient power.

Power can be increased by raising the acceptable level of α. If only the sign of the coefficient for the covariate year is used to determine the direction of a trend, regardless of significance level, then the ANCOVA has a high probability of detecting trends correctly, particularly with eight or more annual surveys. However, at higher α levels, the probability of detecting a change in the wrong direction (γ-error) increases.

When making decisions, there are distinct trade-offs between the error types which must be evaluated. In trend analysis, power should be defined as $1 - (\beta + \gamma)$ to include only detection of a trend in the correct direction. If the cost of making an $\alpha$-error, i.e., falsely concluding that a stable population is increasing or decreasing, is low, $\alpha$ can be increased to increase power. However, attention must be paid to both $\beta$ and $\gamma$-errors as power increases. If $\gamma$ is relatively large, then power should be greater than the previously suggested value of 0.80.

Additional surveys improve the power to detect trends and reduce $\gamma$-errors. Furthermore, if future research can identify and record additional factors affecting observed abundances, such as productivity of the area surveyed (Smith et al. 1986) or ocean temperature patterns (Reilly 1990), this may reduce the variability in the model and increase power.

Future research is planned to continue surveys and search for alternative methods of analyzing these data. The traditional approach to making statistical inference regarding trends has been hypothesis testing with a null hypothesis of no change. As seen in this paper, this is a complicated approach. One must first decide what levels of $\alpha$, $\beta$, and (now) $\gamma$ one is willing to tolerate. The range of these errors is dependent on many factors, including the level of change to be detected, and the number of years surveyed. Once inference is made, it cannot be presented to others without reference to this bewildering array of decision criteria.

Bayesian statistics (Iversen 1984, Press 1989) may offer an alternative approach to statistical inference, circumventing many of the complications discussed above. Bayesian methods would allow the calculation of the probability distribution of possible trends given the observed data. From this distribution it would be possible to directly calculate the probability that the population is increasing or decreasing. Such methods may be of more value than statistical test results which are highly dependent on the chosen error levels.

## Acknowledgments

Support, personnel, and flight time were provided by J. Lecky, Southwest Regional Office (NMFS), the Southwest Fisheries Science Center (NMFS), and the California Department of Fish and Game (CDFG). We thank R. Holt for assisting in planning the surveys. The observers and data recorders were M. Beeson, D. DeMaster, S. Diamond, T. Gallegos, A. Hudoff, L. Jones, E. Konno, S. Kruse, M. Larson, J. Lecky, and J. Scholl (also J.B., K.F., and D.H.).

The aircraft were chartered from Aspen Helicopters, Inc. and contracted from CDFG. The pilots were B. Cheney, B. Cole, L. Heitz, K. Lankard, E. Schutze, J. Turner, R. Riediger, and R. Van Benthuysen. The computer data acquisition software was written by J. Cubbage. We thank D. DeMaster, T. Gerrodette, S. Kruse, C. Lennert, B. Taylor, P. Wade, and two anonymous reviewers for their careful reviews of this manuscript. C. Lennert also provided valuable advice regarding statistical aspects of this paper.

## Citations

**Barlow, J.**
1987 An assessment of the status of harbor porpoise populations in central California. Admin. Rep. LJ-87-06, Southwest Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, P.O. Box 271, La Jolla, CA 92038, 34 p.
1988 Harbor porpoise (*Phocoena phocoena*) abundance estimation in California, Oregon and Washington: I. Ship surveys. Fish. Bull., U.S. 86:417–432.

**Barlow, J., and D.A. Hanan**
1990 An assessment of the status of harbor porpoise populations in California. Working Paper SC/42/SM6 presented to the IWC Scientific Committee, June 1990 (unpubl.). Southwest Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, P.O. Box 271, La Jolla, CA 92038, 31 p.

**Barlow, J., C. Oliver, T.D. Jackson, and B.L. Taylor**
1988 Harbor porpoise (*Phocoena phocoena*) abundance estimation in California, Oregon and Washington: II. Aerial surveys. Fish. Bull., U.S. 86:433–444.

**Calambokidis, J., C. Ewald, G.H. Steiger, S.M. Cooper, I.D. Szczepaniak, and M.A. Webber**
1990 Harbor porpoise studies in the Gulf of the Farallones. Final report for contract CX 8000-8-0001 to Gulf of Farallones Natl. Mar. Sanctuary, NOAA, San Francisco, by Cascadia Research Collective, Olympia, WA 98501, 57 p.

**Carmer, S.G.**
1976 Optimal significance levels for application of the least significant difference in crop performance trials. Crop Sci. 16:95–99.

**Diamond, S.L., and D.A. Hanan**
1986 An estimate of harbor porpoise mortality in California set net fisheries, April 1, 1983 through March 31, 1984. Admin. Rep. SWR-86-15, Southwest Region, Natl. Mar. Fish. Serv., NOAA, 300 S. Ferry Street, Terminal Island, CA 90731, 40 p.

**Forney, K.A., D.A. Hanan, and J. Barlow**
1989 Detecting trends in harbor porpoise abundance using aerial surveys: A preliminary report based on three years. Admin. Rep. LJ-89-20, Southwest Fish. Sci. Cent., Natl. Mar. Fish. Serv., NOAA, P.O. Box 271, La Jolla, CA 92038, 18 p.

**Hanan, D.A., S.L. Diamond, and J.P. Scholl**
1986 An estimate of harbor porpoise mortality in California set net fisheries, April 1, 1984 through March 31, 1985. Admin. Rep. SWR-86-16, Southwest Region, Natl. Mar. Fish. Serv., NOAA, 300 S. Ferry Street, Terminal Island, CA 90731, 38 p.
1987 An estimate of harbor porpoise mortality in California set net fisheries, April 1, 1985 through March 31, 1986. Admin. Rep. SWR-87-5, Southwest Region, Natl. Mar. Fish. Serv., NOAA, 300 S. Ferry Street, Terminal Island, CA 90731, 9 p.

**Hayes, J.P.**
1987 The positive approach to negative results in toxicology studies. Ecotoxicol. Environ. Saf. 14:73–77.

**Hinds, W.T.**
1984 Towards monitoring of long-term trends in terrestrial ecosystems. Environ. Conserv. 11:11–18.

**Iversen, G.R.**
1984 Bayesian statistical inference. Sage University paper on quantitative applications in the social sciences, 07-043. Sage Publ., Beverly Hills, CA.

**Joyner, S.P. (editor)**
1985 SAS/STAT guide for personal computers, Version 6 edition. SAS Inst. Inc., Box 8000, Cary, NC 27511-8000.

**Peterman, R.M.**
1990a Statistical power analysis can improve fisheries research and management. Can. J. Fish. Aquat. Sci. 47:2–15.
1990b Comment on forest decline and acidic deposition: The importance of reporting statistical power. Ecology 71: 2024–2027.

**Peterman, R.M., and M.J. Bradford**
1987 Statistical power of trends in fish abundance. Can. J. Fish. Aquat. Sci. 44:1879–1889.

**Press, S.J.**
1989 Bayesian statistics: Principles, models and applications. John Wiley & Sons, NY, 237 p.

**Reilly, S.B.**
1990 Seasonal changes in distribution and habitat differences among dolphins in the eastern tropical Pacific. Mar. Ecol. Prog. Ser. 66:1–11.

**Rotenberry, J.T., and J.A. Wiens**
1985 Statistical power analysis and community-wide patterns. Am. Nat. 125:164–168.

**Sekiguchi, K.**
1987 Occurrence and behavior of harbor porpoises (*Phocoena phocoena*) at Pajaro Dunes, Monterey Bay, California. M.S. thesis, San Jose State Univ., San Jose, CA, 36 p.

**Skalski, J.R., and D.H. McKenzie**
1982 A design for aquatic monitoring programs. J. Environ. Manage. 14:237–251.

**Smith, R.C., P. Dustan, D. Au, K.S. Baker, and E.A. Dunlap**
1986 Distribution of cetaceans and sea-surface chlorophyll concentrations in the California Current. Mar. Biol. 91:385–402.

**Taylor, B., and P.K. Dawson**
1984 Seasonal changes in density and behavior of harbor porpoise (*Phocoena phocoena*) affecting census methodology in Glacier Bay National Park, Alaska. Rep. Int. Whaling Comm. 34:479–483.

**Toft, C.A., and P.J. Shea**
1983 Detecting community-wide patterns: Estimating power strengthens statistical inference. Am. Nat. 122:618–625.