

Detecting trends in raptor counts: power and Type I error rates of various statistical tests

*Jeff S. Hatfield, William R. Gould IV, Brett A. Hoover, Mark R. Fuller,
and Edward L. Lindquist*

Abstract We conducted simulations that estimated power and Type I error rates of statistical tests for detecting trends in raptor population count data collected from a single monitoring site. Results of the simulations were used to help analyze count data of bald eagles (*Haliaeetus leucocephalus*) from 7 national forests in Michigan, Minnesota, and Wisconsin during 1980–1989. Seven statistical tests were evaluated, including simple linear regression on the log scale and linear regression with a permutation test. Using 1,000 replications each, we simulated $n = 10$ and $n = 50$ years of count data and trends ranging from -5 to 5% change/year. We evaluated the tests at 3 critical levels ($\alpha = 0.01, 0.05,$ and 0.10) for both upper- and lower-tailed tests. Exponential count data were simulated by adding sampling error with a coefficient of variation of 40% from either a log-normal or autocorrelated log-normal distribution. Not surprisingly, tests performed with 50 years of data were much more powerful than tests with 10 years of data. Positive autocorrelation inflated α -levels upward from their nominal levels, making the tests less conservative and more likely to reject the null hypothesis of no trend. Of the tests studied, Cox and Stuart's test and Pollard's test clearly had lower power than the others. Surprisingly, the linear regression t -test, Collins' linear regression permutation test, and the nonparametric Lehmann's and Mann's tests all had similar power in our simulations. Analyses of the count data suggested that bald eagles had increasing trends on at least 2 of the 7 national forests during 1980–1989.

Key words autocorrelation, bald eagle, *Haliaeetus leucocephalus*, raptor, serial correlation, statistical power analysis, survey, trend analysis

A fundamental question in monitoring abundance of species is whether a species is increasing or decreasing in population size over an area during a given time. We expect a stable population size if the population is not changing. Stability, however, does not necessarily mean constancy, rather that the probability distribution of population sizes is constant or stationary over the period of interest.

Researchers have developed methods to test the hypothesis of stability by analyzing counts, signs, or other indices of population size collected systematically at monitoring sites over several years. Route-regression analysis, developed by Geissler and Noon (1981), uses analysis of covariance on the log scale to control for observer effects and uses bootstrapping to produce significance levels. Route-regression performs well as a statistical method under a variety of

During this research, all authors except Lindquist were with the National Biological Service, Patuxent Wildlife Research Center, 11510 American Holly Drive, Laurel, MD 20708, USA. Address for Edward L. Lindquist: U.S. Forest Service, Superior National Forest, P.O. Box 338, Duluth, MN 55801, USA. Current address for William R. Gould IV: Department of Economics, University Statistics Center, New Mexico State University, Las Cruces, NM 88003, USA. Current address for Brett A. Hoover: National Biological Service, Inventory and Monitoring, 12100 Beech Forest Road, Laurel, MD 20708, USA. Current address for Mark R. Fuller: National Biological Service, Raptor Research and Technical Assistance Center, 3948 Development Avenue, Boise, ID 83705, USA.

circumstances (Geissler 1984, Geissler and Link 1988) but requires numerous monitoring sites. It is widely used to estimate large-scale, continental, and state population trends for the North American Breeding Bird Survey (Droege 1990, Geissler and Sauer 1990).

In many cases, however, monitoring the relative abundance of a species occurs at only 1 or several small-scale sites over n sample periods of interest. If the number of sample periods is large, time-series analysis probably is most appropriate for analyzing population count data collected at a single site because it can adjust for autocorrelation among the counts (Jassby and Powell 1990, Barker and Sauer 1992). Positive autocorrelation (also called serial correlation) is found frequently in raptor and other biological count data because repeated measurements over time are collected from the same sampling areas and vital rates can be correlated from 1 period to the next (Pendleton 1989, Burgman et al. 1993). Autocorrelation generally arises because the area of interest is not surveyed at random each year. For economic or logistic reasons, the same site or route often is surveyed from 1 year to the next, even though the goal is to estimate trends for a larger area of interest. Although time-series analysis may be best for these data, it is rarely used for raptor count data because of insufficient data, violations of assumptions, and difficulty of the analysis.

Titus et al. (1989) reviewed methods to analyze raptor count data collected at several sites in the eastern United States and made recommendations to reduce the likelihood of incorrect conclusions concerning trend. Methods proposed for analysis of count data include linear regression of the natural logs of the counts versus time (Gerrodette 1987), linear regression with a permutation test (Collins 1990), a t -test and a permutation test based on first and last counts (Pollard et al. 1987, Link and Hoover 1991), and 3 nonparametric tests (Mann 1945, Lehmann 1975, Conover 1980).

Thus, at least 7 methods with a variety of statistical assumptions are available for resource managers to analyze count data. The relative performance of these trend tests, in terms of Type I and Type II error rates under realistic conditions, has not been widely investigated. A Type I error occurs if a trend is found in the sample data when in fact the true population count distribution had no trend over the time studied. A Type II error occurs if no trend is found when a trend actually existed in the true population counts. The power of a statistical trend test, defined as the probability of correctly concluding that a trend in the count data exists, is equal to 1 minus the Type II error rate.

Biologists recognize the importance of calculating the power of statistical tests to detect trends in data from ecological studies, yet rarely is it calculated (Temple and Wiens 1989, Peterman 1990). Gerrodette (1987) provided a power analysis for linear regression in his paper, but incorrect assumptions invalidated his results (Link and Hatfield 1990). Several other researchers published power or related analyses for biological surveys, but these analyses are not necessarily relevant to raptor studies (Eberhardt 1978, de la Mare 1984, Harris 1986, Kendall et al. 1992, Green and Young 1993, Taylor and Gerrodette 1993).

We report on simulations estimating power and Type I error rates of various trend tests for both independent and autocorrelated raptor count data and evaluate the merits of these tests for estimating trends in raptor populations. Furthermore, using the trend tests described above, we analyzed counts of bald eagles (*Haliaeetus leucocephalus*) from 7 national forests of the Eastern Region of the U.S. Department of Agriculture Forest Service (USFS) surveyed during 1980-1989 (Marita 1989). Data were collected systematically by the USFS using aerial surveys of bald eagle nests in Chequamegon and Nicolet (Wis.), Chippewa and Superior (Minn.), and Hiawatha, Ottawa, and Huron-Manistee (Mich.) National Forests. Two non-random surveys, each examining all known eagle territories, were conducted each year in each national forest: 1 verified occupancy of bald eagle nests in early spring (Fig. 1a), and 1 determined the success (productivity) of the same nests (Fig. 1b) in late spring or early summer (Fraser et al. 1983).

We chose 1980-1989 for our analyses because observer effort within each forest was thought to have been relatively constant during this period, following survey improvements stimulated by comments on survey inconsistencies by Mattson and Grewe (1976). Constant observer effort is important because Geissler and Noon's (1981) route-regression analysis found observer covariates to be a significant cause of bias in bird surveys (Geissler and Sauer 1990).

Four summary statistics (Table 1) are relevant to the trend tests applied to the bald eagle occupancy and productivity data shown in Fig. 1. The percent rate of increase for these data can be defined in several ways, of which we chose 2: as $r_1 = 100(e^{\text{slope}} - 1)$, where the slope is that obtained from a linear regression of the log counts versus time; and as a function of the $(n - 1)$ root of the ratio of the last and first counts, yielding $r_2 = 100([C_{89}/C_{80}]^{1/9} - 1)$, a statistic derived from the Pollard et al. (1987) t -test where C_{80} and C_{89} are the counts for 1980 and 1989, respec-

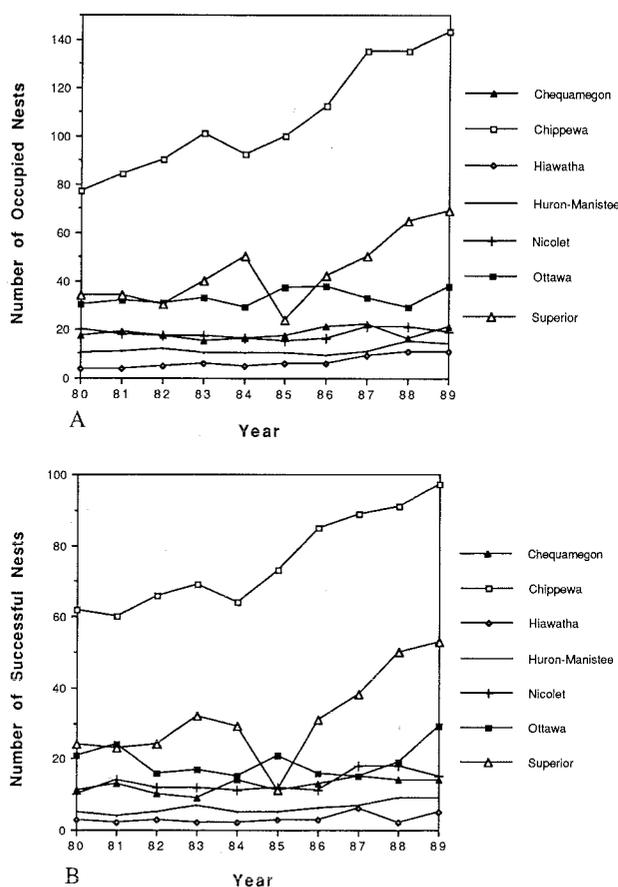


Fig. 1. Number of occupied (A) and successful (B) bald eagle nests in 7 national forests in Michigan, Minnesota, and Wisconsin, 1980–1989.

tively. Similarly, the coefficient of variation can be calculated in at least 2 ways: as $CV_1 = 100([e^{MSE} - 1]^{1/2})$, which assumes normality of the regression residuals of log counts versus time and where MSE is the mean square error of the regression line; and as $CV_2 = 100(SD/\bar{x})$, where SD is the standard deviation of the counts and the mean is the arithmetic average over the 10 counts.

Trend tests

Each of the statistics we consider tests the same null hypothesis of no trend in the counts ($H_0: r = 0$), where r is some measure of growth rate such as percent rate of increase, versus alternative hypotheses of either a positive or negative trend in the counts ($H_a: r > 0$ or $H_a: r < 0$, respectively). The tests programmed for our simulations are described below.

Simple linear regression on the natural log scale (Gerrodette 1987). The usual regression t -test with $n - 2$ degrees of freedom tests whether the slope differs from 0 of the line fitted to the natural log

of the counts versus time. This test assumes the regression residuals are independent and have a normal distribution. Eberhardt and Simmons (1992) also advocated linear regression on the log scale for studying trends in animal populations, but they did not provide a power analysis.

Simple linear regression with a permutation test (Collins 1990). This test uses simple linear regression on the natural log scale with a permutation procedure to test for significance (Collins 1990). A permutation test can be constructed for many statistics and assumes only that the sample counts are independent and identically distributed random variables. The permutation test is also known as a randomization test (Manly 1991), but we reserve this terminology for another process (see below). To construct a permutation test, one randomly reorders or permutes the data many times (e.g., 999 in our simulations) and compares the observed test statistic to the percentiles of the distribution of statistics generated by the permutation procedure to obtain the significance level for the observed statistic. Although we programmed our own permutation tests, note that Manly (1991) is the companion book to software that can perform permutation tests for linear regression and other types of analyses.

Pollard's test (Pollard et al. 1987) as modified by Link and Hoover (1991). This statistic is defined as:

$$\sqrt{\frac{n-2}{n-1}} \frac{\log C_{y_n} - \log C_{y_1}}{\sqrt{\frac{\sum (\log C_{y_i} - \log C_{y_{i-1}})^2}{n-1} - \frac{(\log C_{y_n} - \log C_{y_1})^2}{(n-1)}}}$$

where C_{y_i} = count for year i , $i = 1, 2, \dots, n$.

Assuming normality and independence of the differences between successive log counts, Link and Hoover (1991) showed that this statistic has a t -distribution with $n - 2$ degrees of freedom. Pollard et al. (1987) incorrectly concluded that a simpler version of this statistic had a t -distribution with $n - 1$ degrees of freedom. This t -test requires the years (y_i) to be spaced equally apart but the statistic can be modified if they are not.

Pollard's test (Pollard et al. 1987, Link and Hoover 1991) with a permutation test. This test uses a permutation procedure as discussed above and, like the nonparametric tests below, does not require any normality assumptions.

Mann's trend test (Mann 1945). This statistic is defined as:

$$\sum_{i < j} I(C_{y_i} < C_{y_j})$$

$i, j = 1, 2, \dots, n$, where I is the indicator function, taking the value 1 if the quantity within the parentheses is true, and 0 if false. This statistic totals the times each

Table 1. Percent rate of increase (r_1 , r_2) and coefficient of variation (CV_1 , CV_2) for the number of occupied and successful bald eagle nests in 7 national forests in Michigan, Minnesota, and Wisconsin, 1980–1989.

National forest	Occupied nests				Successful nests			
	r_1^a	CV_1^b	r_2^c	CV_2^d	r_1	CV_1	r_2	CV_2
Chequamegon	1.9	12.9	2.4	13.6	3.4	14.5	2.7	16.2
Chippewa	7.1	6.0	7.1	21.9	5.7	5.8	5.1	18.0
Hiawatha	12.4	13.5	11.9	39.8	6.0	38.3	5.8	44.2
Huron–Manistee	2.7	15.0	3.8	17.3	7.7	16.8	6.7	28.2
Nicolet	0.9	12.4	–0.6	12.0	4.5	16.8	4.6	21.6
Ottawa	1.5	10.2	2.7	10.7	0.7	23.6	3.7	23.6
Superior	8.1	25.7	8.2	33.6	8.7	40.9	9.2	40.4

^a $r_1 = 100(e^{\text{slope}} - 1)$, where slope is obtained from a linear regression of log counts versus time.

^b $CV_1 = 100((e^{\text{MSE}} - 1)^{1/2})$, which assumes normality of the regression residuals of log counts versus time and where MSE is the mean square error of the regression line.

^c $r_2 = 100((C_{89}/C_{80})^{1/9} - 1)$, a statistic derived from the Pollard et al. (1987) t -test where C_{80} and C_{89} are the counts for 1980 and 1989, respectively.

^d $CV_2 = 100(\text{SD}/\bar{x})$, where SD is the standard deviation of the counts and the mean is their arithmetic average over the 10 counts.

later count is greater than each earlier count. It is large for increasing trends and small for decreasing trends.

Lehmann's trend test (Lehmann 1975). Similar to Mann's trend test, Lehmann's (1975) statistic is defined as:

$$\sum_{i < j} (j - i) I(C_{y_i} < C_{y_j}).$$

This statistic weights the pairwise comparisons relative to the number of sample periods between them.

Cox and Stuart's trend test (Conover 1980). This statistic is defined as:

$$\sum_{i=1}^m I(C_{y_i} < C_{y_{i+c}}),$$

where $c = n/2$ and $m = c$ if n is even and $c = (n + 1)/2$ and $m = c - 1$ when n is odd. The middle count is not used when n is odd. Ties between counts are dropped from the data set, and m and c are adjusted accordingly.

Simulation methods

Our simulations were performed using the GAUSS programming language (Edlefsen and Jones 1986). Population growth and decline were assumed to be exponential with a rate of increase or decrease specified by $r = 0, 1, 2, 3, 4,$ or 5% change/year. These values were chosen because they were thought to be similar to actual yearly population change that USFS managers would be interested in detecting.

We assumed a log-normal sampling error distribution for the population counts with a coefficient of variation (CV) of 40%. This variability was supported by literature (Titus et al. 1989) and was verified in this study as a maximum (conservative) estimate for the variability in bald eagle surveys. Variability was added to the simulated counts by incorporating log-normally distributed sampling errors in the counts using the exponential model $C_i = C_1(1+r)^{i-1}$, $i = 1, 2, \dots, n$.

Two models were used for the sampling error distribution. The first assumed independent errors from year to year while the second assumed positively autocorrelated errors. Researchers usually assume independence of year-to-year counts or regression residuals when population trend statistics are computed for an individual area (e.g., Titus and Fuller 1990, Wilson 1991), but this assumption is expected to make the tests less conservative (more likely to reject H_0). This is because positive autocorrelation is expected to inflate the α -levels of the statistical tests since the sampling variance can be seriously underestimated (Neter and Wasserman 1974, Barker and Sauer 1992). The positive autocorrelation was added to the errors in our simulations by assuming a Toeplitz format of the autocorrelation matrix (i.e., counts i years apart have an autocorrelation of 0.5^{i-1}). This is a relatively high level of year to year serial correlation.

We simulated $n = 10$ and $n = 50$ years of count data. The α -levels of the nonparametric tests, which have discrete probability distributions, were adjusted to the nominal α -level using a randomization test

(Hogg and Craig 1978:255). Three nominal α -levels ($\alpha = 0.01, 0.05, \text{ and } 0.10$) were tested with 1-tailed tests for each statistic. We performed 1,000 replicates of each simulation for each of the combinations of years, α -level, and growth rate. For tests requiring a permutation procedure (Collins' linear regression and Pollard's test), 999 random permutations of the data were performed for each simulation to make it easier to calculate the percentiles for the permutation tests in each simulation. For statistics in the simulations that required the log of the count be calculated, 0.5 was added to each count before the test so that 0 counts could be included. Note that this con-

stant was not added to the USFS eagle count data; because 0 counts never occurred in those data sets, the natural log of the count could always be computed.

The estimated Type I error rate of a test statistic for a given combination of number of years monitored, error distribution, and nominal α -level is the proportion of rejections of H_0 out of 1,000 simulations when the expected growth rate is $r = 0$. This estimate has a standard error associated with it from which confidence levels can be attached to the estimated error rate. This standard error is estimated with the usual binomial variance formula (Link and Hatfield 1990). The estimated power of a test statistic for a particular combi-

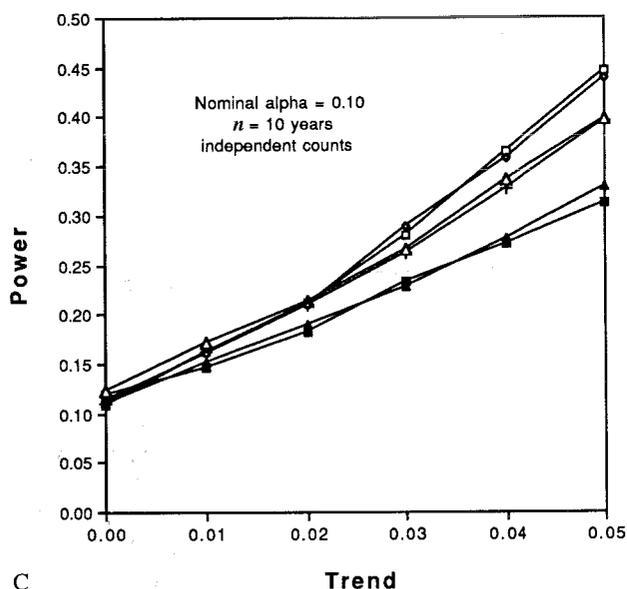
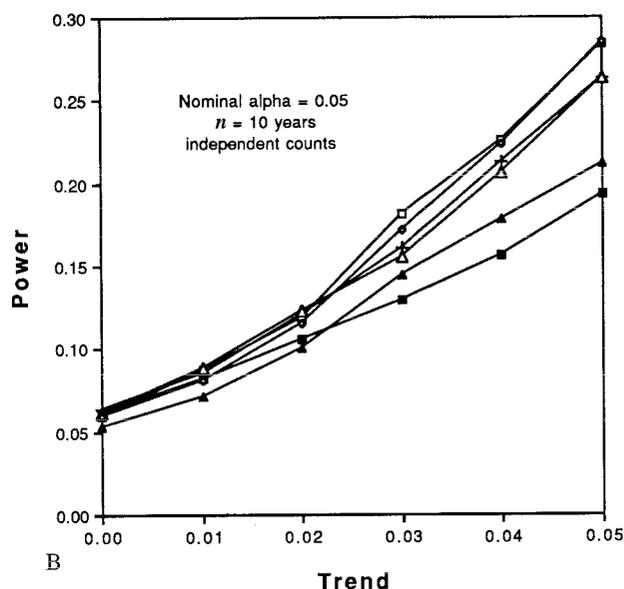
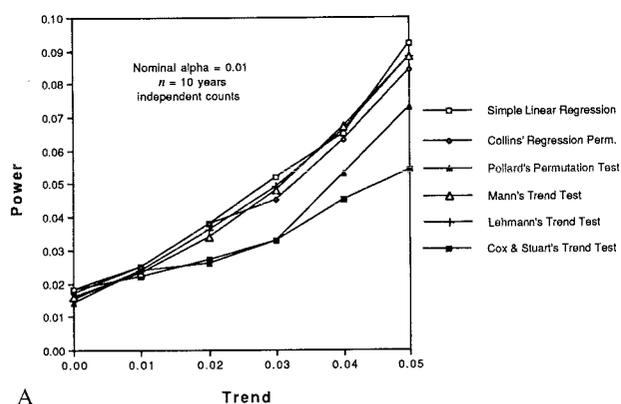
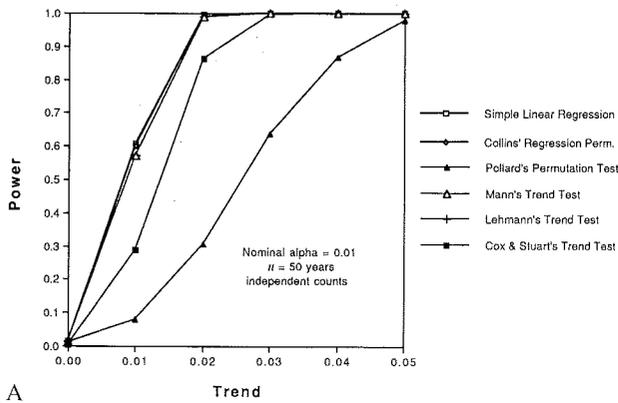
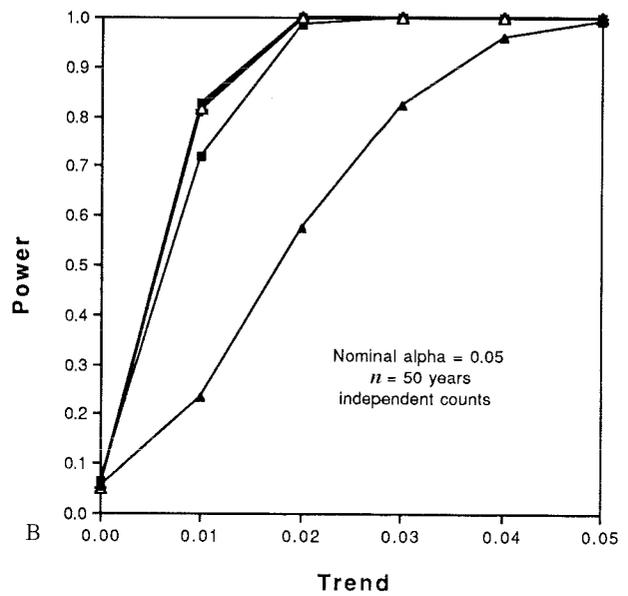


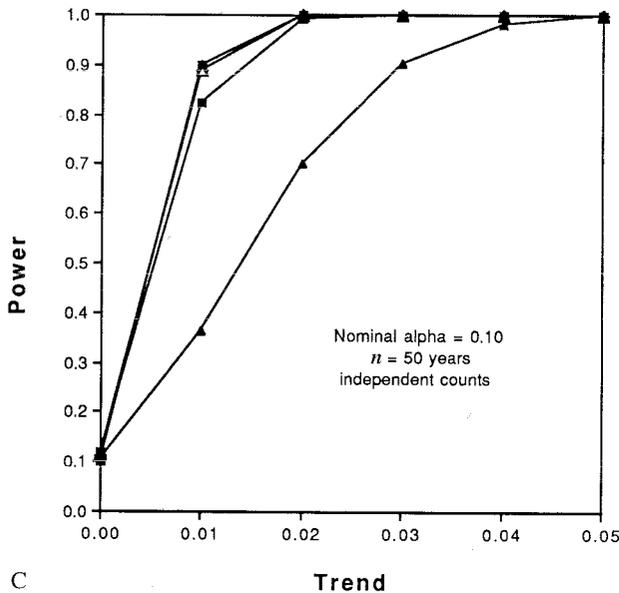
Fig. 2. Estimated power and Type I error rates of 6 statistical tests used for detecting a trend in 10 years of independent, exponentially increasing, log-normally distributed sample counts, with a coefficient of variation of 40%, and $\alpha = 0.01$ (A), $\alpha = 0.05$ (B), and $\alpha = 0.10$ (C). The Type I error rate is the power at a trend of 0, although it is only approximately the nominal level because of sampling error introduced by the simulations.



A



B



C

Fig. 3. Estimated power and Type I error rates of 6 statistical tests used for detecting a trend in 50 years of independent, exponentially increasing, log-normally distributed sample counts, with a coefficient of variation of 40%, and $\alpha = 0.01$ (A), $\alpha = 0.05$ (B), and $\alpha = 0.10$ (C).

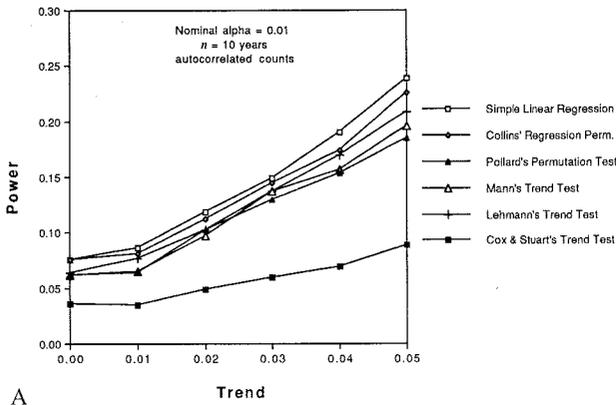
nation is the proportion of rejections out of 1,000 replicates for a given expected percent growth rate ($r \neq 0$).

Comparing the power curves of each test allows the manager to decide which test is preferred for analyzing raptor data; in our case, the USFS bald eagle count data. In addition, comparison of the tests can be facilitated by computing each test statistic's significance level for the actual bald eagle data to verify concordance of conclusions among the tests. For the bald eagle data, however, we never can know the "truth" (i.e., actual population trend), only estimates of it. The simulation results, therefore, are the best way to compare the relative performance of the

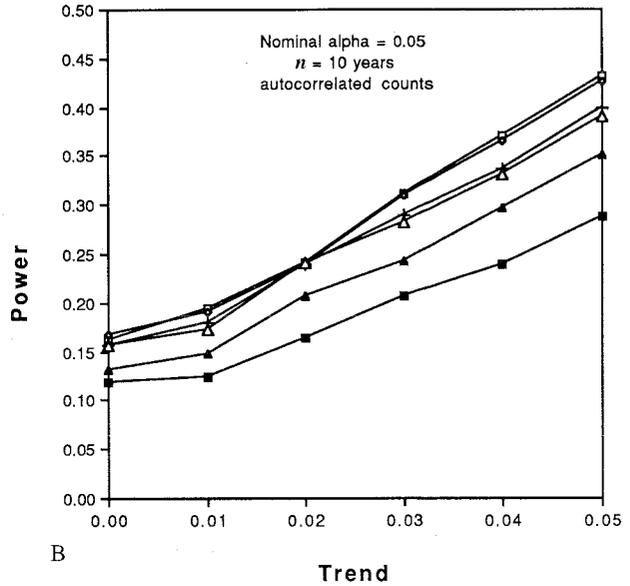
trend tests because we define the true values of the population parameters.

Results and discussion

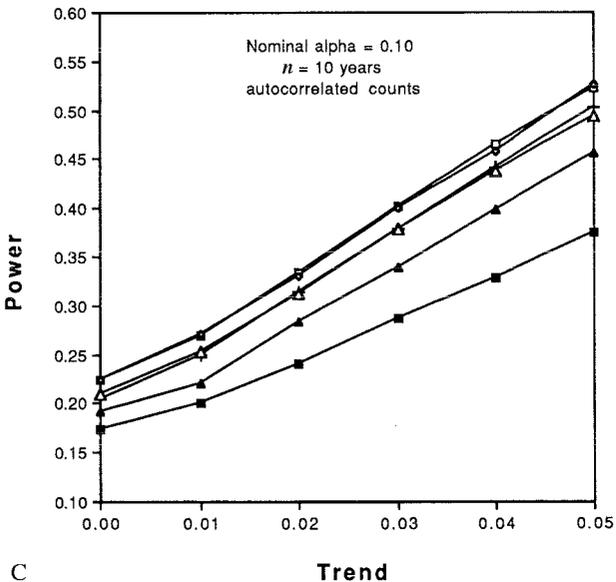
For the USFS eagle nest occupancy data (Fig. 1a), the rate of increase r_1 of the counts on the 7 national forests ranged from 0.9% to 12.4% with a mean of 4.9%; r_2 ranged from -0.6% to 11.9% with a mean of 5.1% (Table 1). For the success data (Fig. 1b), r_1 ranged from 0.7% to 8.7% with a mean of 5.3%, and r_2 ranged from 2.7% to 9.2% with a mean of 5.4% (Table 1). Based on these results, the range of r values used in the simulations was appropriate for evaluating the



A



B



C

Fig. 4. Estimated power and Type I error rates of 6 statistical tests used for detecting a trend in 10 years of autocorrelated, exponentially increasing, log-normally distributed sample counts, with a coefficient of variation of 40%, and $\alpha = 0.01$ (A), $\alpha = 0.05$ (B), and $\alpha = 0.10$ (C).

power of the trend tests applied to these USFS eagle data.

The estimated coefficients of variation reported in Table 1 ranged from 5.8% to 44.2%. Because we used a coefficient of variation of 40% in our power simulations, these simulations represent a conservative or minimum estimate of power of the trend tests for bald eagle count data because most of the USFS bald eagle data had coefficients of variation less than 40%. For any of the trends with a coefficient of variation less than 40%, the power of the tests will actually be greater for detecting trends in the bald eagle data.

The power curves for our simulations are graphed in Figs. 2-5 for the upper-tailed tests (for positive or

increasing trend): The lower-tailed tests (for negative or decreasing trend) had power curves nearly identical to Figs. 2-5 and therefore are not shown here. We did not graph the power curves of the Pollard's t -test. Since this t -test is based on the differences between log counts in successive years and these differences are assumed to be independent, the way we generated our data in these simulations violated one of the assumptions of this test. In our study, counts were assumed to be independent, but differences between counts were not, which generated negative autocorrelation between the data points for Pollard's test. Consequently, the test almost never rejected the null hypothesis for our simulated data and all the

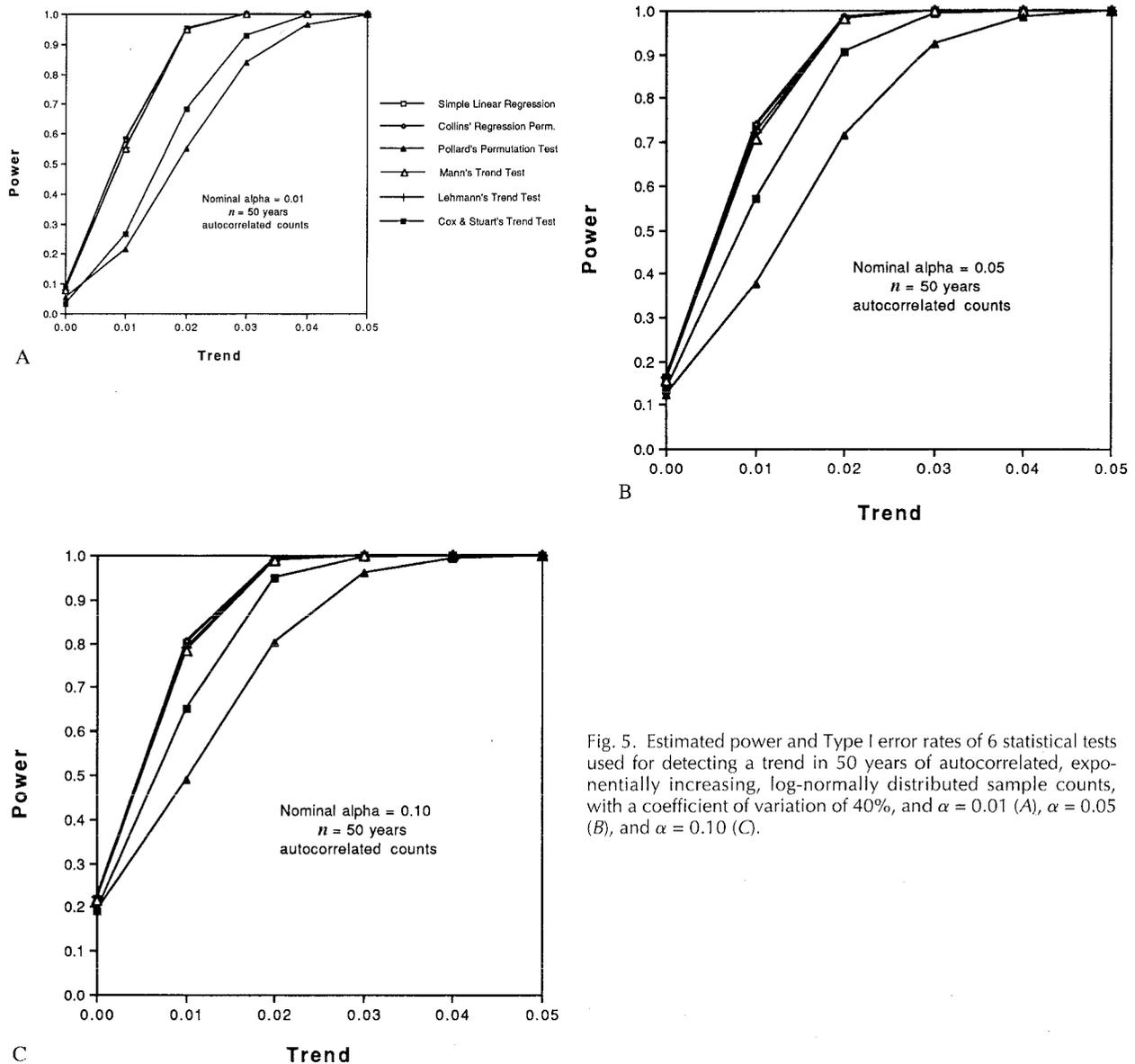


Fig. 5. Estimated power and Type I error rates of 6 statistical tests used for detecting a trend in 50 years of autocorrelated, exponentially increasing, log-normally distributed sample counts, with a coefficient of variation of 40%, and $\alpha = 0.01$ (A), $\alpha = 0.05$ (B), and $\alpha = 0.10$ (C).

power curves were essentially 0. This test still might be valid to use with real data, however, if assumptions are valid for that data. The permutation procedure used with Pollard's test corrects the problem and is not affected by the dependencies in the successive differences between log counts, allowing the test to be evaluated properly.

As expected, tests performed with 50 years of data (Figs. 3 and 5) were much more powerful than those with 10 years of data (Figs. 2 and 4). Of the 6 tests graphed, clearly Cox and Stuart's test and Pollard's permutation test had lower power than the others. Surprisingly, the linear regression *t*-test, Collins' linear regression permutation test, Lehmann's test, and Mann's test all had similar power. Linear regression

should yield the most powerful test because it is the likelihood-ratio test under the assumptions of our simulations (Neter and Wasserman 1974). Pollard's test is the likelihood-ratio test under the assumption of multiplicative errors in the counts (Pollard et al. 1987).

The effect of positive autocorrelation, as expected, was to shift the α -levels upward from nominal α -levels. The shift was not great: the nominal 0.01 level tests had an average actual α -level of 0.07 (Figs. 4a and 5a); the nominal 0.05 level tests had an average actual α -level of 0.15 (Figs. 4b and 5b); and the nominal 0.10 level tests had an average actual α -level of 0.20 (Figs. 4c and 5c). Inflation of the nominal α -levels is important when testing autocorrelated data. With this relatively high positive autocorrelation, the null hypothesis

Table 2. *P*-values of upper-tailed trend tests for increasing trends in the number of occupied bald eagle nests in 7 national forests in Michigan, Minnesota, and Wisconsin, 1980–1989.

National forest	Simple linear regression log scale	Collins' linear regression permutation	Pollard's trend test	Pollard's permutation test	Mann's trend test	Lehmann's trend test	Cox & Stuart's trend test
Chequamegon	NS ^a	NS	NS	NS	NS	NS	0.10
Chippewa	0.005	0.005	0.025	0.005	0.005	0.005	0.05
Hiawatha	0.005	0.005	0.05	0.005	0.005	0.005	0.05
Huron–Manistee	0.10	0.10	NS	0.025	NS	NS	NS
Nicolet	NS	NS	NS	NS	NS	NS	NS
Ottawa	NS	NS	NS	0.10	NS	NS	NS
Superior	0.025	0.025	NS	0.025	0.01	0.01	NS

^a NS, not significant ($P > 0.10$). Other *P*-values are upper bounds.

should not be rejected unless the chosen test is significant at $\alpha < 0.01$ so the true α -level is < 0.10 . However, given that an actual trend exists in the population, the tests are more likely to reject the null hypothesis in favor of the alternative hypothesis than if the counts did not show positive autocorrelation. Although it is not appropriate to compare the power of tests with different α -levels, the positive autocorrelation appears to make the tests more powerful than their independent counterparts at the expense of a higher Type I error rate. If we set the nominal Type I error rate lower than we usually would (e.g., $\alpha = 0.01$), we can mitigate autocorrelation in the counts and therefore achieve a reasonable actual Type I error rate (e.g., $\alpha < 0.10$).

Survey and management implications

Which test is best to use? Under our simulation assumptions, the linear regression *t*-test should be the most powerful test. Under other assumptions, Pol-

lard's *t*-test or a nonparametric test could be more powerful. Because the Collins' linear regression permutation test, Lehmann's test, and Mann's test do not assume normality of the counts, their regression residuals, or the differences between successive log counts, and because these tests were nearly as powerful as the linear regression *t*-test in our simulations, 1 of these tests would be preferable for analyzing count data in which the normality assumptions were violated. Because α -levels shift upward from their nominal values when positive autocorrelation exists, we should choose a low nominal α -level to maintain a reasonably small probability of Type I error. We do not recommend concluding statistical significance for count data unless the test statistic is significant at $\alpha < 0.01$.

When the 7 trend tests were applied to the bald eagle nest occupancy and success data, none of the lower-tailed tests found a decreasing trend in any of the counts, implying that populations were either stable or increasing (i.e., $P > 0.10$ for all lower-tailed tests). In general, we concluded that regres-

Table 3. *P*-values of upper-tailed trend tests for increasing trends in the number of successful bald eagle nests in 7 national forests in Michigan, Minnesota, and Wisconsin, 1980–1989.

National forest	Simple linear regression log scale	Collins' linear regression permutation	Pollard's trend test	Pollard's permutation test	Mann's trend test	Lehmann's trend test	Cox & Stuart's trend test
Chequamegon	0.05	0.05	NS ^a	NS	0.10	0.025	NS
Chippewa	0.005	0.005	0.05	0.005	0.005	0.005	0.05
Hiawatha	0.10	0.10	NS	NS	NS	NS	NS
Huron–Manistee	0.005	0.01	NS	0.025	0.025	0.005	0.10
Nicolet	0.025	0.025	NS	0.05	NS	0.05	NS
Ottawa	NS	NS	NS	NS	NS	NS	NS
Superior	0.05	0.05	NS	0.10	0.025	0.01	NS

^a NS, not significant ($P > 0.10$). Other *P*-values are upper bounds.

sion residuals for both upper- and lower-tailed tests were not different from normally distributed, but power might be a problem with just 10 years of data.

Most of the 7 upper-tailed trend tests of the actual bald eagle data agreed fairly well in their conclusions on trends. They tended to conclude there were increasing trends ($P < 0.01$) for the number of occupied bald eagle nests in Chippewa and Hiawatha National Forests (Table 2) and for the number of successful bald eagle nests in Chippewa and Huron-Manistee National Forests (Table 3). Similarities and differences among tests are useful to point out in Tables 2 and 3, but we reiterate that any conclusions from these comparisons really only apply to these eagle data and are not necessarily generalizable to other data sets.

Collins' linear regression permutation test was almost always as powerful for these eagle data as the usual linear regression t -test. Since the permutation test does not need to assume normality of regression residuals, the permutation test would be better to use if linear regression was the chosen trend analysis. Pollard's t -test and permutation test were generally not as powerful as other tests, similar to what was found in the simulations. However, Pollard's permutation test was always more powerful for these data than Pollard's t -test and may be due to the assumption of independence of counts versus independence of differences between successive counts, as discussed above. Although not recommended in this case due to lower power, if 1 of Pollard's tests were chosen for the trend analysis of these data, the permutation test would be better because it was more powerful.

Of the 3 nonparametric tests, Lehmann's test was always the most powerful test and thus would be the best choice for analyzing these data if a nonparametric test was chosen for the trend analysis. Cox and Stuart's test was the least powerful of all 7 tests, except for occupied nests in Chequamegon National Forest, for which it was the most powerful test. Note that, for the Superior National Forest (SNF), Lehmann's test was more powerful than every other test and found increasing trends ($P < 0.01$) on this forest. Lehmann's test was the most powerful test for SNF because a change in survey design on this forest caused the counts in 1985 to be outliers. The Boundary Waters Canoe Area Wilderness (contained within SNF) was not surveyed in 1985 and, in later years, was surveyed from ground and canoe rather than by air. This biased the counts

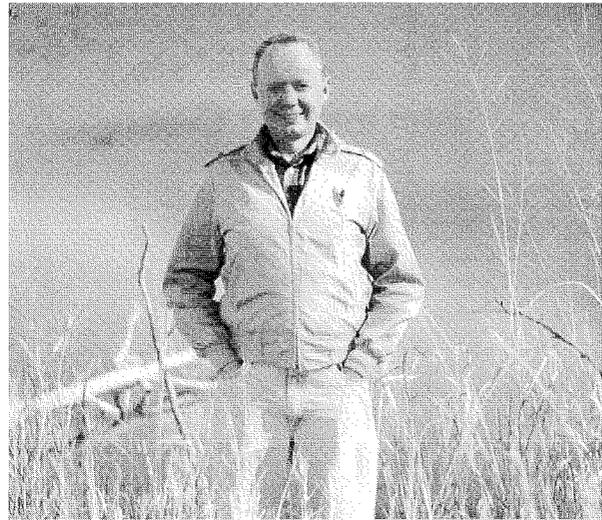
downward for SNF for 1985-1988 (Fuller et al. 1995). Thus, the trends in bald eagle counts on SNF were probably real but best evaluated with a nonparametric test.

Acknowledgments. Helpful comments were provided to earlier versions of this manuscript in reviews by J. Bart, R. M. Erwin, C. J. Henny, W. A. Link, J. S. Ramsey, and 1 anonymous reviewer.

Literature cited

- BARKER, R. J., AND J. R. SAUER. 1992. Modelling population change from time series data. Pages 182-194 in D. R. McCullough and R. H. Barrett, eds. *Wildlife 2001: populations*. Elsevier Sci. Publ., London, U.K.
- BURGMAN, M. A., S. FERSON, AND H. R. AKÇAKAYA. 1993. Risk assessment in conservation biology. Chapman and Hall, London, U.K. 314pp.
- COLLINS, B. T. 1990. Using rerandomizing tests in route-regression analysis of avian population trends. Pages 63-70 in J. R. Sauer and S. Droege, eds. *Survey designs and statistical methods for the estimation of avian population trends*. U.S. Fish and Wildl. Serv. Biol. Rep. 90(1).
- CONOVER, W. J. 1980. *Practical nonparametric statistics*. John Wiley and Sons, Inc., New York, N.Y. 493pp.
- DE LA MARE, W. K. 1984. On the power of catch per unit effort series to detect declines in whale stocks. *Rep. of the Int. Whaling Comm.* 34:655-661.
- DROEGE, S. 1990. The North American breeding bird survey. Pages 1-4 in J. R. Sauer and S. Droege, eds. *Survey designs and statistical methods for the estimation of avian population trends*. U.S. Fish and Wildl. Serv. Biol. Rep. 90(1).
- EBERHARDT, L. L. 1978. Appraising variability in population studies. *J. Wildl. Manage.* 42:207-238.
- EBERHARDT, L. L., AND M. A. SIMMONS. 1992. Assessing rates of increase from trend data. *J. Wildl. Manage.* 56:603-610.
- EDLEFSEN, L. E., AND S. D. JONES. 1986. GAUSS programming language manual. Aptech Systems, Inc., Kent, Wash. 720pp.
- FRAZER, J., L. D. FRENZEL, J. E. MATHISEN, F. MARTIN, AND M. E. SHOUGH. 1983. Scheduling bald eagle reproduction surveys. *Wildl. Soc. Bull.* 11:13-16.
- FULLER, M. R., J. S. HATFIELD, AND E. L. LINDQUIST. 1995. Assessment of ground-based counts of nestling bald eagles in northeastern Minnesota. *Wildl. Soc. Bull.* 23:169-174.
- GEISSLER, P. H. 1984. Estimation of animal population trends and annual indices from a survey of call-counts or other indications. Pages 472-477 in 1984 proceedings of the section on survey research methods. Am. Stat. Assoc., Washington, D.C.
- GEISSLER, P. H., AND W. A. LINK. 1988. Bias of animal population trend estimates. Pages 755-759 in E. J. Wegman, D. T. Gantz, and J. J. Miller, eds. *Computing science and statistics: proceedings of the 20th symposium on the interface*. Am. Stat. Assoc., Alexandria, Va.
- GEISSLER, P. H., AND B. R. NOON. 1981. Estimates of avian population trends from the North American breeding bird survey. Pages 42-51 in C. J. Ralph and J. M. Scott, eds. *Estimating numbers of terrestrial birds*. *Stud. in Avian Biol.* No. 6, Cooper Ornithol. Soc., Los Angeles, Calif.
- GEISSLER, P. H., AND J. R. SAUER. 1990. Topics in route-regression analysis. Pages 54-57 in J. R. Sauer and S. Droege, eds. *Survey designs and statistical methods for the estimation of*

- avian population trends. U.S. Fish and Wildl. Serv. Biol. Rep. 90(1).
- GERRODETTE, T. 1987. A power analysis for detecting trends. *Ecology* 68:1364-1375.
- GREEN, R. H., AND R. C. YOUNG. 1993. Sampling to detect rare species. *Ecol. Appl.* 3:351-356.
- HARRIS, R. B. 1986. Reliability of trend lines obtained from variable counts. *J. Wildl. Manage.* 50:165-171.
- HOGG, R. V., AND A. T. CRAIG. 1978. Introduction to mathematical statistics. Fourth ed. Macmillan Publ. Co., Inc., New York, N.Y. 438pp.
- JASSBY, A. D., AND T. M. POWELL. 1990. Detecting changes in ecological time series. *Ecology* 71:2044-2052.
- KENDALL, K. C., L. H. METZGAR, D. A. PATTERSON, AND B. M. STEELE. 1992. Power of sign surveys to monitor population trends. *Ecol. Appl.* 2:422-430.
- LEHMANN, E. L. 1975. Nonparametrics: statistical methods based on ranks. Holden-Day, Inc., San Francisco, Calif. 457pp.
- LINK, W. A., AND J. S. HATFIELD. 1990. Power calculations and model selection for trend analysis: a comment. *Ecology* 71:1217-1220.
- LINK, W. A., AND B. A. HOOVER. 1991. The detection of density-dependence from a series of annual censuses: a comment. *Ecology* 72:1514-1515.
- MANLY, B. F. J. 1991. Randomization and Monte Carlo methods in biology. Chapman and Hall, New York, N.Y. 281pp.
- MANN, H. B. 1945. Nonparametric tests against trend. *Econometrika* 13:245-259.
- MARITA, F. J. 1989. The 27th annual report of raptor nesting success in the national forests of the eastern region. U.S. For. Serv. Biol. Rep. 38pp.
- MATTSSON, J. P., AND A. H. GREWE, JR. 1976. Bald eagle nesting in the Superior National Forest. U.S. For. Serv. Res. Note NC-198. 2pp.
- NETER, J., AND W. WASSERMAN. 1974. Applied linear statistical models. Richard D. Irwin Publ. Co., Inc., Homewood, Ill. 842pp.
- PENDLETON, G. W. 1989. Statistical considerations in designing raptor surveys. *Natl. Wildl. Fed. Sci. and Tech. Ser.* 13:275-280.
- PETERMAN, R. M. 1990. The importance of reporting statistical power: the forest decline and acidic deposition example. *Ecology* 71:2024-2027.
- POLLARD, E., K. H. LAKHANI, AND P. ROTHERY. 1987. The detection of density-dependence from a series of annual censuses. *Ecology* 68:2046-2055.
- TAYLOR, B. L., AND T. GERRODETTE. 1993. The uses of statistical power in conservation biology: the vaquita and northern spotted owl. *Conserv. Biol.* 7:489-500.
- TEMPLE, S. A., AND J. A. WIENS. 1989. Bird populations and environmental changes: Can birds be bio-indicators? *Am. Birds* 43:260-270.
- TITUS, K., AND M. R. FULLER. 1990. Recent trends in counts of migrant hawks from northeastern North America. *J. Wildl. Manage.* 54:463-470.
- TITUS, K., M. R. FULLER, AND J. L. RUOS. 1989. Considerations for monitoring raptor population trends based on counts of migrants. Pages 19-32 in B. U. Meyburg and R. D. Chancellor, eds. *Raptors in the modern world*. Int. Coun. Bird Preservation, World Working Group on Birds of Prey and Owls, Berlin, Germany.
- WILSON, U. W. 1991. Responses of three seabird species to El Niño events and other warm episodes on the Washington coast, 1979-1990. *Condor* 93:853-858.



Jeff S. Hatfield (photo) is a Biostatistician with the Migratory Bird Research Branch (MBRB) at Patuxent Wildlife Research Center (PWRC) in Laurel, Maryland. He received his B.A. and M.S. in Zoology from Miami University and his M.A.S. in Applied Statistics and Ph.D. in Biostatistics from Ohio State University. Jeff has worked at PWRC since 1988, specializing in population and community ecology and the design and analysis of wildlife and ecological studies. He spends most of his time at the computer but occasionally gets out of the office to do field work in Hawaii, Maryland, and Minnesota. **William (Bill) R. Gould IV** is an Assistant Professor in the Department of Economics, University Statistics Center, at New Mexico State University. He received his degrees from Clemson University (B.S. in Math Science) and North Carolina State University (M. Stat. in Statistics, Ph.D. in Biomathematics). Bill has worked on a temporary basis for PWRC, the National Biological Service (NBS), and the Montana Department of Fish, Wildlife and Parks. His primary research interests involve the estimation and modeling of fisheries and wildlife population dynamics. **Brett A. Hoover** is a Computer Specialist with the NBS Inventory and Monitoring Group, based at PWRC. He received his B.S. in Wildlife Science with a minor in Statistics at Pennsylvania State University. Currently Brett's primary responsibilities include establishing telecommunication services (internet, bulletin board, etc.) for dissemination of monitoring data and protocols. He also has an interest in analysis of radiotelemetry data. **Mark R. Fuller** is Director of the NBS Raptor Research and Technical Assistance Center in Boise, Idaho. He attended Colorado State University (B.S.), Central Washington University (M.S.), and University of Minnesota (UM; Ph.D.), and worked 2 years at UM. Mark spent the next 15 years in the MBRB at PWRC before moving to Boise. His work includes population dynamics and bird-habitat associations, emphasizing raptors. His interests include basic raptor biology, migration, and conservation. **Edward (Ed) L. Lindquist** is a Biological Scientist at the Superior National Forest (SNF) in Duluth, Minnesota. He has been forest biologist at SNF, the Ashley Forest in Utah, and the Sawtooth National Recreation Area in Idaho. Ed attended the University of Montana (B.S.) and various other universities. He has been a member of The Wildlife Society since 1974, was president of the Minnesota Chapter from 1983-1985, and was on the North Central Section Executive Board in 1984. Ed's work includes habitat analysis and management, development of geographic information systems, endangered species management and reintroduction, and work in temperate and tropical forestry.

Associate Editor: Tacha

