

# RESAMPLING METHODS FOR COMPUTATION-INTENSIVE DATA ANALYSIS IN ECOLOGY AND EVOLUTION

*Philip H. Crowley*

Center for Evolutionary Ecology, T. H. Morgan School of Biological Sciences,  
University of Kentucky, Lexington, Kentucky 40506-0225, and NERC Centre for  
Population Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY  
United Kingdom

**KEYWORDS:** bootstrap, jackknife, Monte Carlo methods, permutation tests, randomization tests

## INTRODUCTION

The advent of fast, relatively inexpensive (thus, widely available) microcomputers is transforming the way we analyze data in ecological and evolutionary research. Even more profound, however, are the associated changes in questions asked, empirical methods used, studies conducted, and interpretations offered. Now that an array of computation-intensive statistical methods is newly available for general use, it seems particularly important to assess their advantages and limitations, to note how they are currently being used, and then to consider implications for the future.

I focus in this review on four related techniques known in the statistical and biological literature as *randomization* (or *permutation*) *tests*, *Monte Carlo methods*, *bootstrapping*, and the *jackknife*. I refer to them collectively as resampling methods, because each involves taking several-to-many samples from the original data set (randomization, bootstrap, jackknife) or from a stochastic process like the one believed to have generated the data set (Monte Carlo). Each of these methods is actually an extensive family of techniques and specific applications that cannot be thoroughly examined here; instead, I briefly characterize the focal methods and then survey the recent literature in ecology and evolution to identify the issues most frequently associated with these techniques. It emerges that resampling methods are well represented in

data analyses related to some of the most important issues and intense controversies currently in these fields of research.

The specific objectives of this paper are:

1. to acquaint a wider array of ecologists and evolutionary biologists with these useful techniques, which, at least until very recently, have been underemphasized or ignored in statistical training;
2. to document the association between certain research questions and one or more of these resampling methods;
3. to emphasize the role of the focal methods in expanding the range of feasible experimental designs and in shifting the conceptual basis of data analysis;
4. to compare and contrast resampling methods with more standard approaches, noting assumptions and other key features that bear on their appropriateness for particular applications; and
5. to highlight methods that need clarification and development, in the hope that these will soon be addressed by statisticians and biometricians.

The review is intentionally biased toward ecological studies, in accord with my own research experience and interests. Simpler, univariate analyses are emphasized in the interest of clarity and also because a review of computation-intensive multivariate methods in ecology is in progress (165). To respect page limits on contributions to this volume, I have restricted the number of examples cited and emphasized more recent papers most likely to contain additional citations of relevant work. I assume here that readers are familiar with rudimentary statistical concepts and basic methods.

The present review proceeds as follows: First, I describe briefly the four resampling techniques, including a relatively straightforward example of each from the literature of ecology and evolution. Next, I summarize results of a systematic literature search for applications, including a computer search of biological journals and edited volumes published during 1985–1990, and my own search by hand through all issues of two prominent ecological journals for the period 1985–1991. Resampling techniques are used to test for temporal trends in the use of these methods and for differences in frequency of use between ecological and evolutionary studies. Publications identified in the search are classified by topic and subtopic, from which are distilled seven major issues considered with example applications in more detail. Focus then shifts to the relation of resampling applications to classical and actively developing statistical methodology. Finally, I discuss advantages, disadvantages, and implications of these methods, highlight methodological questions that deserve attention, and close with some specific recommendations.

## BRIEF DESCRIPTIONS OF THE FOCAL METHODS

Lucid descriptions and examples of these methods and computer programs suitable for implementing them are available in recent books by Edgington (64), Noreen (186), and Manly (163).

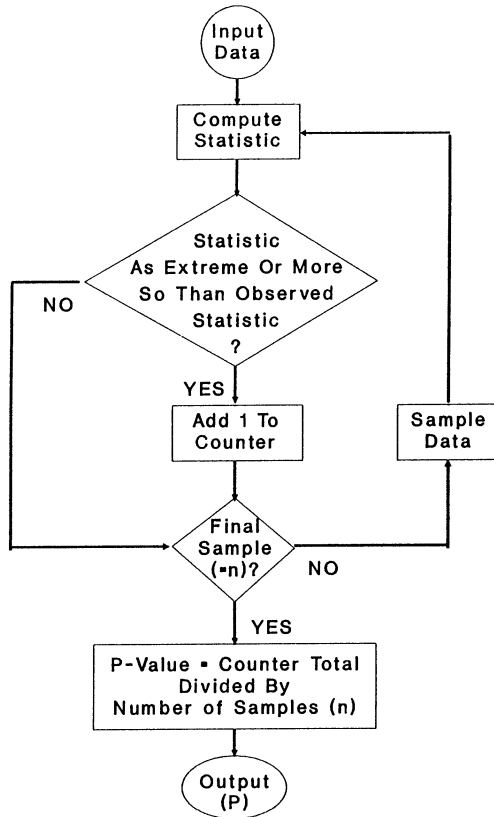
### *Randomization*

Referring to the use of randomization tests in analyzing data, R. A. Fisher once claimed that statistical "conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method" (75). In a randomization test, the chance of type 1 error under the null hypothesis (i.e. the  $p$ -value) is determined by repeated random assignment of the data to treatment levels. The  $p$ -value is simply the proportion of all data arrangements yielding test statistics at least as extreme in magnitude as the value resulting from the arrangement actually observed (see Figure 1).

When the null hypothesis is that the observed magnitude of the test statistic is (say) not larger than would be expected by chance (a one-tailed hypothesis), then the extreme values to be counted in calculating the  $p$ -value are those greater than or equal to the observed test-statistic value. When the null hypothesis is that the observed magnitude of the test statistic is not different from chance expectation (a two-tailed hypothesis), then separate counts are made of values greater than or equal to the observed and of values less than or equal to the observed; the lower of these counts is doubled and divided by the total number of data arrangements to obtain the two-tailed  $p$ -value (subject to the constraint that  $p < 1$ ).

Data resampling requires pooling all data from the treatment levels (i.e. experimentally established or "observed" groups) to be compared and then reassigning data randomly and without replacement to the treatment levels, keeping the number of observations per treatment level the same as in the original data. In some cases, all possible redistributions of data among treatment levels can be readily obtained, resulting in an "exact" randomization test. In other cases, generally when the potential number of different redistributions approaches or exceeds  $10^4$ – $10^5$ , some of these (often about  $10^3$ ) are sampled with replacement for the test, which is then known as a "sampled" randomization test.

Randomization tests are often based on standard test statistics (e.g.  $t$ ,  $F$ ); it is thus the method of resampling the data and of calculating  $p$  that are definitive, rather than the statistic used. But the potential to use special-purpose or ad hoc statistics is a particularly important advantage of the randomization approach (and of resampling more generally), since this may increase the statistical power to accept the relevant alternative hypothesis (64, 163).



*Figure 1* Flow diagram of hypothesis testing via randomization tests, Monte Carlo methods, and some types of bootstrapping techniques. Except for exact randomization when few samples are needed, this logic is usually implemented using a computer program that generates and processes a large number of samples  $n$  (typically  $n$  1000); only rudimentary programming skills are needed, unless the statistic is particularly complex or the data structure requires a sophisticated sampling algorithm. See Edgington (64), Noreen (186), and Manly (163) for example programs. This diagram is a slight modification of one on the cover (and Figure 1, p. 51) of E.S. Edgington's book (64).

The basic rationale for randomization methods is that under the null hypothesis of, for example, no difference between treatment-level means, any of the possible distributions of data among treatment levels is equally probable. This equiprobability is assumed to follow from (i) random sampling of populations being compared (contemporary applications generally avoid this assumption, but those that invoke it are known as permutation tests), (ii) random assignment of experimental units to treatment level, or (iii) for nonexperimental studies, simply taking the data to be "exchangeable" among levels in the absence of treatment effects (see 163, 250). However, random-

ization tests of differences among means are sensitive to differences in variances and other moments (e.g. see 15, 219, 242; contrary to assertions in 64, p. vi, and others), implying that none of the enumerated assumptions is strictly sufficient. In lieu of defensible alternatives, a sufficient assumption is that observed distributions of data are identical except for the features actually compared in the test. See Table 1.

For example, consider the study by Loreau (156) on temporal niche differentiation in carabid beetles. Loreau was interested in whether species shifted their periods of activity seasonally such that niche overlap or another index termed "mean competitive load" is reduced, as might be expected from competition theory. The data consisted of biweekly activity levels by species and habitat (correlated here with successional stage) over a four-year period. Subject to some constraints on the timing of peak activity and on the boundaries of the active period that were intended to preserve biological realism, the observed temporal distribution of activity within a species was seasonally shifted at random, and the two indices were calculated for each species-habitat combination. For each species and habitat, this procedure was repeated systematically (i.e. by exact randomization) when the total number of distinct reorderings was less than 4000, or randomly with replacement of reorderings (i.e. by sampled randomization) 2000 times when the total number of distinct reorderings exceeded 4000. The  $p$ -values were determined as in Figure 1.

Results differed among habitats, among constraints imposed, and among species subsets considered. In the beechwood and pinewood habitats,  $p$ -values for both response indices tended to approach or achieve statistical significance as increasingly severe and realistic constraints were imposed on the randomization process. In these cases, niche overlap and mean competitive load calculated from the original data were lower than 95% or more of the corresponding values generated from the random seasonal shifts in activity pattern. It is difficult to imagine how these hypotheses could have been tested with these data by standard statistical methods. This example and the continuing debate over the interpretation of such carabid data (53, 266, and their references) indicate some of the challenges that can arise in attempting to operationalize the null hypothesis and specify the most suitable randomization algorithm. Nevertheless, the overall pattern of statistical significance in the present study does suggest increasingly distinct niche differentiation from successional to "climax" beechwood forest, as would be expected according to competition theory and some other possible interpretations.

### *Monte Carlo*

In Monte Carlo methods, a particular random process (e.g. binomial coin flips or a complex stochastic simulation model) is assumed to underlie the observed data in determining confidence intervals or the expected response under the

**Table 1** Assumptions and restrictions on parametric, standard nonparametric, and resampling methods.

Assumptions/restrictions <sup>1</sup>	Standard parametric methods	Standard nonparametric		Resampling methods			
		Rank methods <sup>2</sup>	Categorical methods <sup>3</sup>	Randomization	Monte Carlo	Bootstrap	Jack-knife
GENERAL							
Statistically independent data	Yes <sup>4</sup>	Yes <sup>4</sup>	Yes	Yes <sup>4</sup>	No <sup>5</sup>	Yes	Yes <sup>6</sup>
Particular underlying distribution(s)	Yes <sup>7,8</sup>	No	Yes <sup>9</sup>	No	Yes <sup>10</sup>	No	Yes <sup>7,8,11</sup>
Empirical samples must be random	Yes	No	Yes	No	Yes	Yes	Yes
Relatively sensitive to outliers	Yes	No <sup>12</sup>	Yes	No <sup>12</sup>	Yes	No <sup>12</sup>	Yes <sup>13</sup>
Data ranked (or reduced to ranks)	No	Yes	No	No	No	No	No
Low values in data pose problems	No	No	Yes <sup>14</sup>	No	No	No	No
Ties in data pose problems	No	Yes <sup>15</sup>	Yes <sup>15</sup>	No	No	No	No
HYPOTHESIS TESTS FOR $\geq 2$ SAMPLES							
Identical underlying distributions	Yes <sup>7,16</sup>	Yes	No	Yes <sup>17</sup>	Yes	No <sup>18,19</sup>	Yes <sup>7,18</sup>
Relatively sensitive to sample size differences	Yes	Yes	No	Yes <sup>17</sup>	No	Yes <sup>18,19</sup>	No <sup>18</sup>

<sup>1</sup> This list is not exhaustive, nor are the table entries definitive. The table is intended to provide general comparisons as guidelines for statistical practitioners.<sup>2</sup> Mann-Whitney, Kruskal-Wallis, Friedman's ANOVA-by-ranks, Wilcoxon matched-pairs signed-ranks, etc.<sup>3</sup> Chi-square, G-tests, etc.

- <sup>4</sup> Using special techniques (e.g. time-series analysis), temporal or spatial autocorrelations may in some cases be taken into account without serious problems.
- <sup>5</sup> The underlying stochastic process (e.g. a Markov chain) may adequately represent non-independence in the data.
- <sup>6</sup> This assumption, in effect, is made and violated in all jackknife applications involving variance calculations, generally with unknown consequences (163).
- <sup>7</sup> For the simplest and most common applications, much statistical research indicates robustness to moderate departures from this assumption when the underlying distribution in question is the normal.
- <sup>8</sup> The Central Limit Theorem assures asymptotic agreement with this assumption for large sample sizes.
- <sup>9</sup> Generally chi-square.
- <sup>10</sup> Actually, the stochastic process that generates the data distribution is what must be known.
- <sup>11</sup> For confidence intervals and hypothesis testing only. In these cases, the jackknifed *statistic*, not the measured variable, is assumed to be normally distributed. In some applications, the jackknife is robust to underlying non-normality (e.g. 8), but not others (e.g. 179).
- <sup>12</sup> Ranks imply limits on effects of extreme values. In randomization tests and bootstrapping, outliers alter both the observed data and the distribution used for comparison (see 64, 163).
- <sup>13</sup> See Hinkley (112).
- <sup>14</sup> The standard textbook guidelines for chi-square have long been that no expected frequency should be  $< 1.0$ , and no more than 20% should be  $< 5.0$  (40). Expected frequencies in G-tests should be  $\geq 5.0$  (235). Generally, rows, columns, or both may need to be combined to satisfy these constraints. Results of recent statistical research on this issue are more equivocal (B. F. J. Manly, personal communication). Contingency-table tests based on small sample sizes can instead be conducted by Monte Carlo methods (e.g. see 153).
- <sup>15</sup> Ties in rank tests require approximated  $p$ -values, and protecting the chance of type I error in these cases can reduce power somewhat (64). In categorical tests, "ties" can sometimes arise as intermediates between categories, which generally must be omitted from the test.
- <sup>16</sup> Aspin-Welch  $t$ -tests are much more tolerant of unequal variances than are standard parametric tests (235, 242).
- <sup>17</sup> (219, 242). Assuming identical underlying distributions implies that values from different samples are exchangeable under the null hypothesis. Random assignment to treatment is insufficient to satisfy the assumption; this depends in part on the *responses* to treatments.
- <sup>18</sup> Hypothesis tests are not really standardized for these methods.
- <sup>19</sup> Bootstrapping samples separately and comparing the observed statistic with the resulting bootstrapped distribution should at least partly relax this assumption. See Manly (162) for an example of this approach.

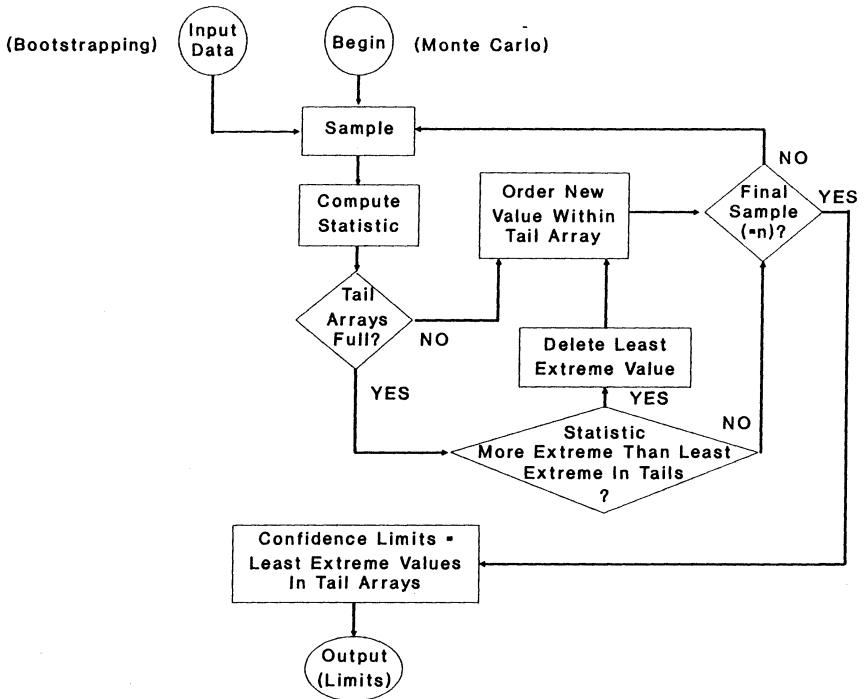
null hypothesis for hypothesis testing. This random process is then sampled repeatedly (e.g. many coin flips are simulated or the simulation model is run many times), with test statistics calculated in each case. For hypothesis testing (e.g. is a particular coin unequally likely to produce “heads” or “tails”? —or— do empirical observations differ significantly from the model’s predictions?), the  $p$ -value is found from this frequency distribution of test statistics exactly as for randomization (see Figure 1). In fact, randomization is generally considered to be the special case of Monte Carlo tests in which the relevant random process simply samples the distribution of test-statistic values associated with equiprobable rearrangements of data among treatment levels (163). Nevertheless, some differences in assumptions and restrictions arise in comparing more typical Monte Carlo tests (i.e. those in which the observed data are not used to implement the random process) and randomization tests (see Table 1).

Monte Carlo methods are often used to generate confidence intervals (whereas this is possible but uncommon and usually cumbersome with randomization—e.g. see 163, p. 18–20). Though not particularly difficult, this is procedurally more complex than hypothesis testing because it requires accumulating and maintaining ordered arrays of extreme values of the statistic (corresponding to the tails of the distribution) as these are generated (see Figure 2, which illustrates the “percentile method”; more “adventurous” methods, with clear advantages in some cases, are described e.g. in 67 and 69; for recent work on Monte Carlo methods see 12, 82, 85, and their references).

Consider a demographic study of the colonial gorgonian *Leptogorgia virgulata* using projection matrices and Monte Carlo methods to analyze time-varying population growth (95). Field measurements of recruitment, colony growth, and survival for five size classes over 24 months, supplemented by other fecundity data, were used to construct  $23\ 5 \times 5$  monthly projection matrices. Each entry represented the expected number of individuals in the row size class that arose by survival, growth, or reproduction from an individual in the column size class one month before. Multiplying this matrix by a column vector representing the numbers of individuals in each of the five size classes at the beginning of the month projected by the matrix yielded the numbers present in each size class a month later.

The less complex of two Monte Carlo applications in the paper concerns determinations of elasticity (i.e. proportional contributions by recruitment, growth, and survival rates to population growth rate) using the matrix techniques. The question of interest in the gorgonian study was whether the observed patterns in the data could be attributable simply to the general form of the matrices rather than their biological details. If so, then an arbitrary distribution of nonzero entries in the matrices should generate elasticity





*Figure 2* Flow diagram for constructing confidence intervals via Monte Carlo and bootstrapping according to the percentile method. This requires a computer program that uses "tail arrays" to collect and reorder the smallest and largest values of the statistics generated. If the confidence level of interest is  $100(1-\alpha)\%$ , where  $\alpha$  is the corresponding significance level, and  $n$  values are computed to estimate the confidence interval, then each tail array for two-tail limits will contain  $1 + \alpha n/2$  values (ignoring any fraction). Thus if  $n = 10000$  and  $\alpha = 0.05$ , then the tail arrays ultimately hold the 251 largest and the 251 smallest values of the statistic. Initially, the two arrays are filled by the first 502 values, such that the larger values are ordered in one tail and the smaller values are ordered in the other. Then each subsequent value smaller than the largest in the lower tail or larger than the smallest in the upper tail is ordered within the appropriate array, and the least extreme value is eliminated; intermediate values, insufficiently extreme for either tail, are not stored. After all  $n$  values have been calculated and ordered appropriately, the interval defined by the least extreme value in each of the two tails is the confidence interval. One-tailed confidence intervals are handled similarly. Suitable programs are provided by Noreen (186). With bootstrapping, substantial bias can result from this straightforward approach in some cases (see 67 and 69 for some ways of dealing with this potential problem)

patterns and vital rates statistically indistinguishable from those observed. One thousand random projection matrices were constructed with the same zero elements as in the data matrices, but with all nonzero elements drawn from a uniform distribution ranging from zero to one. (Note that if the nonzero elements had repeatedly been randomly scrambled, rather than drawn from a

particular statistical distribution, then this would have been classified as a randomization test.) Some of the observed pattern did indeed seem to be mimicked by the random matrices; but Monte Carlo tests showed that the gorgonian recruitment and population growth rates used as test statistics were significantly atypical of vital rates derived from the random matrices. So the tests helped separate general features of such matrices from the system-specific information contained in the data. As in the carabid randomization example above, it is difficult to see how any approach other than resampling could have been useful here.

### *The Jackknife*

The jackknife, like its all-purpose namesake, was intended to offer crude but effective assistance when a more precise tool is unavailable (252). It provides systematic methods of resampling the actual data using relatively few calculations that can often be done efficiently on a calculator. The direct results of these calculations are an "improved" (i.e. less sampling-biased) estimate of some sample parameter (e.g. mean, kurtosis, intrinsic rate of increase) and often of the approximate variance and confidence interval associated with the estimate. The confidence interval is sometimes used in hypothesis tests (occasionally the jackknifed data themselves, known as pseudovalues, are used; see 114 and 180 for reviews on the jackknife).

Though higher-order versions may occasionally be useful (e.g. 194, 195), the first-order jackknife is by far the most commonly used and proceeds as follows: Suppose that the parameter of interest  $K$  (e.g. the true standard error of the mean for the underlying normal distribution of means) is estimated appropriately over the whole sample of  $m$  observations as  $k$ . Pseudovalues  $\bar{\kappa}_i$  associated with each observation  $i$  are then obtained as  $\kappa_i = k - (m - 1)(k_{-i} - k)$ , where  $k_{-i}$  is just the standard parameter calculation with the  $i$ th observation deleted from the sample. The expression on the right-hand side of this equation is the sample parameter estimate minus a bias term, reflecting the deviation of the  $i$ -deleted estimate  $k_{-i}$  from the full sample estimate  $k$ . The mean of the pseudovalues  $\bar{\kappa}$  is then the jackknife estimate of  $K$ . The difference  $k - \bar{\kappa}$  measures the overall sampling bias of the original estimate  $k$  (bias can, for example, distort estimates of population density, particularly when the individuals are strongly clumped in space; see 67 for derivations of the above relationships).

Ignoring the correlations necessarily present among the pseudovalues, calculating their variance  $s^2$  in the usual way, and dividing by the number of observations then generates the variance  $s^2/m$  of the jackknife estimate  $\kappa$ . Now the assumption that such jackknife estimates are based on normally distributed error yields the parametric confidence interval for the estimate:  $k$

$\pm t^*s/m^{1/2}$ , where  $t^*$  is the appropriate two-tailed critical value of the  $t$ -distribution with  $m-1$  degrees of freedom.

Note that the variance estimate requires assuming that the correlations among pseudovalues are unimportant, but general conditions in which this might be valid have not been established (163). The normality assumption, though justifiable by the Central Limit Theorem for large sample sizes, is difficult to evaluate with the smaller-sample applications where it is more dubious. See Table 1.

In a study of interclutch intervals and reproductive success of feral pigeons nesting on a building at the University of Kansas, Johnson & Johnston (131) used the jackknife to test the relation between three selection parameters and four morphological features. Over 600 banded birds were included in the study; survival and reproductive activity were observed several times per week over a 17-month period and associated with measurements of body mass, tarsus length, bill length, and bill width. The three selection parameters of interest here were the standardized directional selection differential ( $i$ ), the slope of relative fitness regressed on the morphological trait; the standardized stabilizing selection differential ( $C$ ), for which a positive value indicates disruptive selection and a negative value implies stabilizing selection; and the standardized directional selection gradient ( $\beta$ ), resulting from multiple regression of relative fitness on the combined morphological variables. Though this was not explicitly stated, the parameters were presumably considered significantly nonzero when the two-tailed confidence interval failed to include zero, following the procedure for constructing the interval that is outlined above.

Results indicated a highly significant directional selection differential ( $i$ ) and gradient ( $\beta$ ) for female body mass, and a significant directional selection differential for female bill length, interpreted here as correlated selection. Thus fecundity selection related to interclutch interval apparently "targets" female body mass. In this example, for parameters of unknown statistical distribution calculated over the whole dataset, only the jackknife and bootstrap and their close kin could readily estimate the sampling variation required for hypothesis tests.

### *The Bootstrap*

Bootstrapping is a quite recent technique (66) that is still developing rapidly and attracting much attention in the statistical literature (e.g. 57, 68, 77). Like randomization and the jackknife, bootstrapping focuses on resampling the actual data to reveal some of the subtler patterns they imply (in fact, results obtained from the bootstrap are often closely approximated by those from the jackknife—66). Here, the basic notion is that the data themselves, viewed as a frequency distribution, represent the best available image of the frequency

distribution from which they were drawn. Thus the bootstrap metaphor refers to the sense in which the data are used in their own statistical analysis. To bootstrap a confidence interval for a statistic (e.g. mean, skewness, or species diversity) calculated from a single data set of  $m$  observations, for example, simply requires  $m$  random draws with replacement per sample from the original data, calculating the statistic, and repeating the process many times according to the scheme illustrated in Figure 2. Here again, this simply specifies a particular random process that technically represents a special case of the Monte Carlo method.

Bootstrapping can also be used in hypothesis testing (e.g. 49, 76, 100); for example, with data from each treatment level (or data set to be compared) sampled with replacement separately, tests can be formulated according to the extent of overlap between confidence intervals or by combining the bootstrapped samples to calculate a test statistic (as in Figure 1; see 163, p. 28, and 186, p. 80). This can still be considered a special case of the Monte Carlo method, but since separate random processes are used to generate the separate samples from which the comparisons are made, the approach is quite different from the usual Monte Carlo approach. Bootstrapping is distinct from randomization, which redistributes the original data set over treatment levels, and it contrasts with the parametric and less computation-intensive jackknife approach. By keeping the sampling process separate between the compared treatment levels, bootstrapping should be less dependent than most other statistical methods on similarity in underlying statistical distributions among treatment levels (B. F. J. Manly, personal communication; see Table 1).

In an extensive study of predator and parasitoid selection pressure on gall size of the goldenrod gall fly *Eurosta solidaginis*, Abrahamson et al (1) used bootstrapping to avoid problems with non-normality and correlated samples that arose in previous analyses. Selection intensities on gall size attributable to natural enemy attack were calculated as the difference between the mean gall diameter of the selected individuals and the population-mean gall diameter, divided by the population standard deviation. For each of 20 populations and two mortality sources (i.e. insects and birds), the observed number of linked observations (gall size in *mm*, survival from the relevant natural enemy as 0 or 1) were sampled with replacement from the original data, the selection intensity was calculated, and this process was repeated 1000 times to generate a two-tailed confidence interval, as in Figure 2. An observed selection intensity was considered significantly nonzero if its lower confidence limit was greater than zero. Two selection intensities were considered significantly different if their 95% confidence intervals did not overlap; this latter would be a very conservative approach to hypothesis testing, except that the  $p$ -values were not adjusted for the large number of comparisons implied.

The many significant selection intensities imposed by insects were all

positive, and the few significant selection intensities imposed by birds were negative. Taken together, natural enemies therefore seem to generate stabilizing selection on gall size, though the parasitoid effect predominated, resulting in some overall directional selection for increasing gall size as well.

## RECENT APPLICATIONS IN ECOLOGY AND EVOLUTION

### *Overview*

To determine how randomization, Monte Carlo, jackknife, and bootstrapping methods are currently being used in the literature, I conducted a search of a large computer database (BIOSIS Previews on-line database, 2100 Arch Street, Philadelphia, Pennsylvania 19103-1399 USA) for the publication years 1985–1990 and directly examined all issues of the journals *Ecology* and *Oecologia* for 1985–1991.

During the period of interest, approximately 9000 biological journals were being abstracted by BIOSIS, apparently including all major journals in ecology and evolution. Searching titles, abstracts, and key words yielded 391 references from 154 journals and 11 chapters from books, once the few obvious mistakes were eliminated by a direct scan of the abstracts. I relied on the BIOSIS classification scheme to draw appropriate distinctions between references classified as “ecology,” “evolution,” or both. It is possible that the Monte Carlo category is somewhat inflated relative to the others, since the term is sometimes used for a wider range of simulation methods than just the statistical techniques of interest here; but the direct examination of journals indicated that any such effect would be minor.

Some results of the computer search are presented in Table 2. All four resampling methods are well represented in the recent literature, with Monte Carlo methods overall about twice as frequent as bootstrapping, which in turn was almost twice as common as either randomization tests or the jackknife. The hypothesis that resampling methods are becoming more common in the literature was corroborated statistically, though the evidence to support this for any particular method was more equivocal (randomization, Monte Carlo, bootstrap) or clearly contradictory (jackknife) (tested by randomization; see Table 2 and Appendix 1). Each of the four methods (and all taken together) was used disproportionately in evolutionary studies relative to ecological studies (Monte Carlo tests; see Table 2, Appendix 1), as suggested by the observed proportions  $v$  of evolutionary studies ( $v = 0.144\text{--}0.360$ ) relative to the overall proportion in all papers published (0.098).

I scanned the methods and results sections and all figures and tables of the 1485 articles published in *Ecology* and the 2128 articles published in



Any of these	Ecology	44	28	44	51	62	77	$b = 1.407$	$v = 0.2285$
	Evolution	11	11	18	15	22	27	$r = 0.736$	$p = 0.0001***$
	Either	54	38	57	62	79	101	$p = 0.472^*$	
Total sample searched	Ecology	36,525	39,175	44,880	46,862	45,552	48,511		
	Evolution	4,364	5,004	5,716	5,464	5,523	6,315		
	Either	39,631	42,801	48,897	50,654	49,679	52,752		

<sup>1</sup> Data were obtained using the on-line BIOSIS database on December 6 and 16, 1991, searching for the "words" randomi(s)zation and permutation (combined here as randomization), Monte Carlo, bootstrap-, and both jack-knif- and jackknif-. After removing a few inappropriate references (e.g. those on jackknife clams and on jellyfish found at Monte Carlo), the remaining 391 abstracts were used to classify the studies for the BIOSIS contributions to Table 3. A total of 154 biological journals and 11 edited volumes were represented in the 1985-1990 data. The eleven journals with eight or more papers in the sample are *Atmospheric Environment* ( $n = 9$ ), *Biometrics* (11), *Canadian Journal of Fisheries and Aquatic Sciences* (18), *Ecology* (17), *Evolution* (18), *Genetics* (10), *Journal of Hydrology* (8), *Journal of Wildlife Management* (8), *Limnology and Oceanography* (8), *Oecologia* (10), and *Systematic Zoology* (8).

<sup>2</sup> Procedures are described in the Appendix. The hypotheses tested are that the focal methods considered in this review (i) are becoming more commonly used over the last several years and (ii) are used in differing frequency in ecology vs evolutionary studies. Results shown for Hypothesis 1 are based on frequencies normalized (i.e. divided) by the total sample searched in each year, indicating frequency per published paper. In each case,  $b$  is the slope of a least-squares regression line through the data,  $r$  is the product-moment correlation coefficient between normalized frequency and publication year, and  $p$  is the  $p$ -value of the randomization test. (Trends were in consistently better agreement with this hypothesis when frequencies were unnormalized, but this approach seemed less useful a priori.) For Hypothesis 2,  $v$  is the fraction of all exclusively ecology or evolution papers that are classified by BIOSIS as evolution. For both hypotheses, tests based on the "any of these" data were considered designed orthogonal comparisons, one-tailed for Hypothesis 1 and two-tailed for Hypothesis 2. Tests of the four individual methods were contingent on significance of the designed comparisons; in these cases, the Bonferroni adjustment (multiplying the observed  $p$ -value by the number of tests) was used. If  $p'$  is the  $p$ -value adjusted as appropriate for the number of tails and the number of tests, then significance levels are: \*\*\*,  $p' < 0.001$ ; \*\*,  $0.01 > p' \geq 0.001$ ; \*,  $0.05 > p' \geq 0.01$ ; otherwise  $p' \geq 0.05$ .

*Oecologia* from 1985–1991. (For 1985–1990 only, there were 1270 articles in *Ecology* and 1841 in *Oecologia*.) Regardless of the authors' original designation, I classified methods as Monte Carlo when a mathematical random process was executed repeatedly to generate an estimate of biological variation (e.g. a confidence interval) or to test a hypothesis, following Figure 1. Analyses specifically classified as binomial tests were excluded (except as explicitly noted), though all were equivalent to exact randomization tests (see below).

During 1985–1990, 87 or 6.9% of the *Ecology* papers and 53 or 2.9% of the *Oecologia* papers included one or more of the focal resampling techniques. (In *Ecology*, allowing for the 8 papers each using 2 of the methods, there were 31 randomization, 33 Monte Carlo, 15 jackknife, and 16 bootstrap. In *Oecologia*, with 3 papers each using 2 of the methods, there were 21 randomization, 22 Monte Carlo, 5 bootstrap, and 8 jackknife.) The BIOSIS search was thus relatively inefficient overall ( $17/87 = 19.5\%$  for *Ecology*, and  $10/53 = 18.9\%$  for *Oecologia*), and the efficiencies probably differed among methods. (This is not particularly surprising, since statistical methods may not often warrant mention in the title, abstract, or key words, though some may be more likely to be mentioned than others.) The overall number of ecology/evolution articles during 1985–1990 that used these resampling methods can be very roughly estimated as the total number of articles identified by the BIOSIS search divided by the mean of these two efficiencies expressed as a decimal fraction, which to the nearest integer equals 2036. Clearly, a thorough and comprehensive review of this and more recent material would be overwhelming, both for reviewers and readers.

Combining the full 1985–1991 direct-examination data with the 1985–1990 BIOSIS results and classifying the papers by content generated Table 3. Notice that some particularly controversial issues in ecology and evolution (e.g. null models, size-ratio theory, detecting density dependence, phylogeny) are well represented here, perhaps mainly to exploit the considerable flexibility of resampling methods in applications involving nonstandard models and test statistics. This flexibility can be a mixed blessing, however, as I note below. The poor representation of behavior and behavioral ecology in the table is probably artifactual, reflecting the separation of behavior from ecology and evolution within BIOSIS.

### *Some Active Areas of Application*

**NULL MODELS, COMPETITION, AND COMMUNITY STRUCTURE** Contemporary interest in competition as a mechanism underlying community structure led in the 1970s to the formulation of null (or neutral) models, with which statistical tests of predicted patterns could be conducted (35, 223, 229). Since



the notion of a probabilistic model that can generate the statistical distribution consistent with a null hypothesis is the essence of Monte Carlo hypothesis testing, Monte Carlo tests (e.g. 201, 223) and closely related exact randomization tests (e.g. 50 and 247, which referred to them using the broader term "binomial tests") were soon prominent in these analyses. The controversy that erupted between those formulating null models for this purpose (42–44, 229, 247) and those considerably less enthusiastic about this approach (41, 56, 91–93) provides a cautionary tale: the potential for differing null models, misunderstandings of methods, procedural errors, and alternative interpretations of similar results can be high with resampling methods (cf Monte Carlo tests in 43, 44, and 93; see 262). Nevertheless, the null model approach seems to have taken hold in the recent literature, partly via Monte Carlo methods (Table 3; see 202, 265).

There is much relatively untapped potential to use resampling tests for detecting community organization (e.g. guild structure: 124, 129, 265) and community similarity (118, 263, 264), and for testing whether environmental factors can account for community structure (29, 83). In assessing species diversity and the variation associated with these estimates, the jackknife has been used most often (e.g. 110, 194, 195), but bootstrap applications (e.g. 23) may become more common.

Much of the null model controversy has addressed the distribution of species abundance and presence/absence on islands (214, 244, and analogous situations considered in 265). A good overview of this issue and problems associated with choosing appropriate constraints on randomization is provided by Manly (163, p. 233 ff).

Numerous recent attempts to test for niche differences (34, 109, 162) and to measure overlap (2, 107) have used resampling methods, particularly randomization (125, 152, 265; see 202). In other cases, temporal niche shifts have been tested by resampling (61, 206, and the binomial or exact randomization test in 50). Simberloff & Boecklen's forensic analysis of Santa Rosalia (230) stimulated several resampling tests (96, 271, and the equivalent of exact randomization in 19) of the constant-size-ratio hypothesis from Hutchinson's original paper (120).

Considerable recent work in plant ecology has focused on competition from immediate neighbors (27, 136, 251) and related distortion of the population size distribution (133, 143, 227). The geometry of access to resources and thus of potential response to competition has also been characterized (136, 215). These plant neighborhood-competition and size-distribution citations involve the gamut of resampling methods considered in this review, both for hypothesis tests (27, 251) and to estimate confidence intervals for the Gini coefficient (an indicator of size inequality; 133, 227) or to calculate skewness by jackknifing (143).

**Table 3** Classification by research topic and statistical method of the relevant papers identified in the BIOSIS survey (1985–1990) and by direct examination of journals (*Ecology* and *Oecologia*, 1985–1991)<sup>1</sup>

TOPIC/Subtopic	Randomization	Monte Carlo	Bootstrap	Jackknife
<u>Competition (total)</u>	31	36	20	17
Null models	3	12	0	0
Niche differentiation, overlap & breadth	14	9	4	4
Size-ratio theory of niche displacement	1	3	0	1
Niche-shift dynamics & interaction intensities	3	1	1	0
Plant size hierarchies	1	0	7	3
Plant neighborhood competition	4	5	0	1
 <u>Community structure (total)</u>	14	16	7	7
Detecting organization	2	5	2	0
Diversity	2	2	3	6
Community similarity	7	2	2	1
 <u>Temporal variability and stability</u>	1	5	1	2
 <u>Detecting density dependence</u>	7	4	0	0
 <u>Spatial patterns and processes (total)</u>	15	33	3	9
Dispersion & spatial pattern	6	17	1	3
Dispersal & migration	5	7	1	4
Scale effects	4	3	0	1
 <u>Demography (total)</u>	5	47	10	8
Population size of density	2	19	6	7
Vital rates	2	9	4	7
Growth, size & age relationships	1	7	0	0
Stock-recruitment relations	0	5	0	0
 <u>Agricultural/fisheries</u>	0	8	0	2
 <u>Environmental factors (total)</u>	4	58	10	3
Absorption & scattering of light	0	16	1	0
Air-quality models & indicators	0	10	4	0
Aquatic environmental quality/toxicology	2	8	3	2
Lake & stream acidification	1	4	1	0
Surface, soil & groundwater	1	13	0	1
 <u>Behavior/behavioral ecology (total)</u>	10	16	2	2
Social organization	3	2	0	0
Foraging	6	7	2	1
 <u>Evolution/evolutionary ecology (total)</u>	61	44	49	19
Selection intensity & response	0	10	3	4
Genetic differentiation & correlation	1	2	2	2
Mutation rates	0	7	0	0
Morphometric comparisons	3	2	2	3
Phylogeny	5	1	33	9

<u>Statistical and modeling methods</u> (total)	41	63	43	44
Analysis of variance	3	2	0	1
Regression & correlation	5	8	0	5
Mantel's test	7	0	0	0
Discriminant function analysis	2	2	0	5
Nearest-neighbor analysis	1	6	0	0
Sensitivity, error & uncertainty	0	14	1	3
Power analysis	0	6	0	1
Confidence intervals & variance	1	4	31	16
Bias estimation & reduction	0	10	7	8
Grand total	145	330	139	114

<sup>1</sup> Direct search of the two ecological journals located all of the references on detecting density dependence and almost all of the competition studies. The statistical and modeling methods references were derived both from BIOSIS and from the direct search. References for the remaining topics were obtained almost entirely or entirely from the BIOSIS search. Data shown here were collected using a lengthier list of topics and subtopics, but those categories accumulating fewer than five citations are not shown; topic totals include the additional citations from subtopics not shown, and the grand total includes citations from topics not shown. Many of the papers are tallied in more than one category.

An underutilized randomization method of very wide potential application, particularly in community analyses, is Mantel's test (119, 166; see the description and examples in 163 and 165). This flexible technique tests for correlation between two (or more) square distance matrices. Typically, entries in one matrix express Euclidian distances (or some alternative measure) between (say) species in quantitative multivariate features (e.g. in diets), and the other matrix may represent a postulated pattern among species (e.g. zeroes and ones indicating membership or not in the same guild). By randomly reassigning rows and columns of one matrix to species, recalculating the correlation between corresponding off-diagonal matrix elements (where the test statistic is the sum of the multiplicative products of these corresponding matrix elements), and then repeating this sequence many times, the statistical tendency in the original data for the postulated pattern to match the distance pattern can readily be assessed (e.g. see 198).

**DETECTING DENSITY DEPENDENCE** Another controversy of long standing in the ecological literature concerns the role of density dependence in population dynamics. (See e.g. 11, 55, 267, and their references for evidence that the controversy continues unabated.) Two important milestones were the initiation of experimental field tests of density dependence (70) and the formulation of statistical methods to detect density dependence in temporal sequences of density data (26). Resampling methods have proven useful in both of these approaches, particularly the latter (48, 51, 54, 203, 204, 210, 258, 259; an application to analysis of a field test is in progress—D. M. Johnson, T. H. Martin, L. B. Crowder, P. H. Crowley, in preparation).

Though concerns have been expressed about the potential for detecting

density dependence in density sequences (84), recently developed methods, particularly the two tests based on randomization methods by Pollard et al (204) and Reddingius and den Boer (210), appear sufficiently powerful to be useful (48, 51, 259). Recent variations on this randomization theme have been used to evaluate bias in *k*-factor analysis (258), to extend the notion of detecting density dependence to the community level (48, 51), and to derive testable predictions about the direction of density changes (48, 51).

The Pollard et al (204) "randomization test" and the Reddingius and den Boer (210) "permutation test" both involve scrambling the order of the observed changes in log-transformed density (a measure of population growth over the time interval) for comparison with the observed sequence. Pollard et al (204) used the correlation coefficient between density at the start of each interval and the associated change in density during the interval as the test statistic (density dependence implies an inverse correlation; the randomization test avoids the problems inherent in the analogous parametric approach noted in 159 and 239). Reddingius and den Boer (210) used the log-range between the highest and lowest densities reached in the density sequence as the test statistic (density dependence implies a small log-range). Other test statistics may be more appropriate or powerful in particular cases (cf the "violation number" statistic in 48 and 51). It may often be helpful to use several different tests and test statistics on the same data set, since the test results are sometimes complementary (51, 259), though this may raise concerns about adequately protecting the chance of type 1 error over all tests.

**SPATIAL PATTERNS AND PROCESSES** Characterizing spatial patterns and processes is a major challenge in contemporary ecological research. A diverse array of resampling approaches has been used for this purpose. Descriptive methods include assessing the spatial distribution of sparsely sampled points and the spatial areas most closely associated with point locations (e.g. tree locations—236 and 136, respectively), and particularly spatial autocorrelation (149, 232). Hypothesis tests have been used to detect nonindependence of animal locations (237, 248); variations in territory size (249); differences in dispersion among size classes, species, and quadrats (105, an application of Mantel's test); differences in association of plant distribution and abundance with taxonomic composition vs vegetation structure (222, also via Mantel's test); and an association between spatial distribution and temporal dynamics (228). Monte Carlo methods have improved and extended the classic nearest neighbor analysis of Clark & Evans (39; see 33; 151; and 163, p. 21-23 and chapter 7).

In other cases, geographical limits of populations have been established (220), and the implications of spatial scale (4, 122, 215) and of environmental heterogeneity (4, 215, 222) have been addressed.

In considering insect dispersal processes, Monte Carlo simulation has been used to evaluate the need for a stochastic formulation to predict dispersal (253). Error associated with estimates of the diffusion coefficient (160, 188) or of the radius of patch detection (103) has been assessed primarily with the jackknife.

**ESTIMATING POPULATION SIZE AND VITAL RATES** Resampling methods, particularly Monte Carlo, are now in fairly common use to reduce bias and determine error associated with estimates of population density (111, 173, 181). The “smoothed bootstrap” (226) and randomization tests (171, 269) have been used to detect density changes, mainly in non-experimental studies.

One of those last randomization examples (i.e. 269) invoked an approach known as MRPP (multiresponse permutation procedures—176–178). MRPP, a special case of Mantel’s test (163, p. 209), is conceptually consistent with graphical representations of the data and readily extends to multivariate problems. With this method, predefined groups (e.g. sites, treatments) can be tested for differences using standard statistical distance measures, from which a test statistic is derived and then assessed by ordinary randomization procedures (Figure 1). Interestingly, standard *t* and *F* tests and common nonparametric tests are special cases of MRPP, though practitioners argue that nonstandard formulations are generally more appropriate (269).

Following the comparison of jackknife and bootstrap methods by Meyer et al (175), there has been much recent interest in measuring and testing for differences in demographic costs of predator defense (13, 212, 260) and of other environmental factors (88, 139), as measured by the per-capita increase rate of zooplankton. (See 95 for an assessment of temporal changes in the per-capita increase rate derived from dominant eigenvalues of a matrix model.) Resampling methods have also been applied to tests and error estimates for other vital rates (birth rate: 58; mortality: 153; relative growth rate: 37; transmission rate of an insect virus: 63; many different vital rates: 254), reproductive effort (86), and extinction rate (199, 200).

**ENVIRONMENTAL MODELING** As the need for reliable environmental predictions and monitoring has steadily increased, a broad range of relatively realistic, quantitative models has appeared in the basic and especially in the applied ecological literature. A focal issue in many of these studies is evaluating the model’s fit to data; for probabilistic models, Monte Carlo methods are often the best option and have commonly been used.

With regard to aquatic environments, resampling has been applied in toxicological models (20, 24, 221) and laboratory tests (205), time-series analysis of BOD data (197 via the “Bayesian bootstrap”), testing sensitivity of lakes to phosphorus loadings (28, 150), estimating an index of water quality

(97), and assessing the impact of acid deposition (62, 128, 132). Applications in soil and groundwater systems include those on soil hydraulic properties (117, 234), estimating runoff (10, 60, 99), and monitoring groundwater quality (155, 185, 240). In the atmosphere, resampling techniques have been used with models of carbon dioxide uptake and exchange (121, 138, 268), for impact assessment of radionuclide fallout (21, 211, 261), and especially in air-quality models and indicators (e.g. 25, 101, 108).

An important component of many climate and plant-growth models, both aquatic and terrestrial, is absorption and scattering of incident solar radiation. Monte Carlo applications are particularly common in these studies (e.g. 5, 38, 94).

**EVOLUTIONARY PROCESSES AND RATES** Resampling methods figure prominently in analyses of natural, sexual, and group selection. Examples include the introductory case studies of the jackknife and the bootstrap early in this review (1, 131), studies determining the magnitude of sexual selection (172, 187) and group selection (78, 102), and others concerned with various responses to selection (142, 182, 257).

Rates of evolution have been assessed and contrasted via resampling applications (78, 90, 140), as have mutation rates (79, 106, 189, 190) and evolutionary implications of genetic drift (189, 190, 217). In an analysis of taxon extinction rates, Raup & Sepkoski (209 and references therein) used randomization tests to identify significant periodicity of major extinction events in the geologic record (also 116; see 116 and 193 on speciation periodicity), but Quinn (207) argued that bootstrapping is more appropriate for this purpose (see 16 and the overview in 163, p. 192 ff).

All four resampling methods have been used to detect genetic differentiation between populations based on immunologic (225), electrophoretic (45, 59), and nucleotide-difference (216, 246) data. Discriminant function analysis, particularly with the help of Monte Carlo (225) or randomization (238) methods, can prove useful in such studies.

**PHYLOGENY** Phylogenetic analysis has evolved rapidly since the 1970s with widespread use both of molecular techniques and of computer simulation and data analysis. Pioneering simulation studies by Raup et al (208) demonstrated the possible importance of stochastic processes and potential biases in interpreting phylogenies. Some of the early work on null models emphasized biogeographic data (e.g. the binomial or exact randomization test of 247; also see 41 on the avoidable and unavoidable biases in such studies), and the usefulness of statistically contrasting proposed phylogenies against a null pattern is becoming more widely recognized (73; see the bootstrap approach of 89 and an exact Monte Carlo method in 233).

An important recent development was the formulation of techniques for establishing confidence intervals for monophyletic groups (jackknifing over taxa: 146; bootstrapping over characters: 72; see 196 for a comparative evaluation of these and related methods). Felsenstein's frequently used approach analyses character data contained in a species  $\times$  character matrix. Bootstrap samples of characters (or, strictly, of the columns of species-specific values for particular characters) are used to construct alternative phylogenetic trees; the percentage of these containing a monophyletic group present in the tree based on the original data then estimates the confidence that the group is indeed monophyletic. Inherent assumptions that characters were sampled randomly and evolved independently raise some concerns about the method's validity (72, 73, 224) but apparently have not deterred applications of this and related bootstrapping approaches (e.g. 126, 127, 270). Bootstrapping clearly predominates in recent resampling studies of phylogeny (Table 3), including both cladistic analyses (e.g. 47, 145, 147) and the phenetic studies emphasized above.

Another issue of current interest concerns whether particular patterns derived from phylogenetic data can be considered nonrandom. Here, randomization tests have been used to scramble character values among species to determine whether the tree derived from the original data required significantly fewer evolutionary step-changes than the trees derived from scrambled character values (nonrandomness was detected in 6 but not in 7; also see an analogous cladistic analysis in 192).

### *Statistical Methodology*

**RELEVANT TYPES OF ANALYSIS** There is much statistical and biometrical research in progress continuing the development of resampling methods (particularly the bootstrap). Moreover, to a greater extent than with other statistical approaches, each new application tends to extend the methodological possibilities because of the ad hoc nature of resampling analysis. Here, I note how resampling methods have been used to supplement or improve standard statistical methods and to stimulate or enhance new research initiatives as well.

Resampling methods avoid some of the more restrictive assumptions involved in standard regression and correlation analyses (e.g. see 64, p. 197), and there are now many published applications (e.g. regression: 30, 81, 170; correlation: 123, 249, 256). The useful generalized correlation methods known as Mantel's test and multiresponse permutation procedures (MRPP) have already been described and characterized in the literature summary above. Analysis of variance deserves special attention because of its central role in the design and analysis of experiments and because of restrictive assumptions

that can proliferate with complexity of the design (191; 64, p. 58 ff; recent applications include 130, 162, and 168). Help needed from statistical researchers on problems associated with ANOVA is noted below.

Many multivariate methods are currently being revitalized and extended via resampling (e.g. cluster analysis: 148, 184, 255; discriminant function analysis: 14, 29, 80; principal components analysis: 241; indirect gradient analysis: 144). It is primarily the methods readily applicable to single samples (Monte Carlo, bootstrap, jackknife) that are of interest in this context, though randomization can be useful for discriminant function analysis (163, 238). See Manly (163 and especially 165) for thorough review of multivariate resampling applications.

Many other standard issues and approaches in experimental design and data analysis have been addressed with resampling methods. Some of the more important of these are assessing errors associated with sampling (2, 103, 256) or direct measurement (3), estimating the power of hypothesis tests (32, 87, 130), determining and reducing bias (2, 173, 183), and determining the appropriate sample size (22, 154, 164).

In empirical studies, the need for methods of analyzing the ecological response to large-scale perturbations (31, 169) has led to some resampling applications associated with intervention analysis (Monte Carlo: 158) or randomized intervention analysis (randomization tests: 32). The approach here is usually based on paired systems, one experimental and one control; each is monitored extensively before and after the experimental system is manipulated, so that some of these observations can be assumed essentially independent (though autocorrelation is directly assessed). This general approach or a successor may prove valuable, particularly where replicated experiments are infeasible, but additional care should be taken to ensure that the null hypothesis is tested against an appropriate alternative (e.g. by transforming to reduce heteroscedasticity or other distributional differences that may confound the test; see 71 and 242).

In modeling studies, there is much current interest in incorporating age or size structure (36, 174) or explicitly representing individuals (52) within population models. Moreover, optimization models now more commonly include stochastic elements (e.g. see 161) or parameter uncertainties that complicate interpretation. In these and similar cases, resampling methods can prove particularly useful in characterizing the model's behavior and evaluating its consistency with empirical observations (e.g. age structure: 137; individual-based model: 157; optimization: 213).

**CLOSE RELATIONSHIPS WITH MORE STANDARD METHODS** The resampling techniques of interest here are all closely related to the more standard and



widely familiar statistical methods. By virtue of conceptual simplicity and the large number of nonparametric tests it has spawned, randomization can be considered fundamental to the standard methods (17, 46, 134). Transforming data to ranks is primarily a device to reduce data sets to a general form that permits construction of nonparametric significance tables, with entries at low sample sizes determined by randomization and at higher sample sizes by normal or chi-square approximations to the randomization results. The first three commonly used tests in Table 4 are examples of these; the sign test is also a kind of rank sum test. Fisher's exact test and the binomial (goodness-of-fit) test are directly calculated cases of exact randomization. The tests listed in the table are just a few of the more common nonparametric tests found in the *Ecology-Oecologia* sample.

Monte Carlo methods are generally used to derive statistical tables for tests based on data assumed to follow particular distributions, such as  $t$ ,  $F$ , and  $\chi^2$  tests. In the Monte Carlo tests of interest here, the actual statistical distribution may be unknown, so long as the relevant stochastic process can be simulated according to the scheme in Figures 1 and 2. In some cases among the *Ecology* and *Oecologia* articles, a smaller number of simulations was used to draw conclusions without a formal test (e.g. 104, 245) or were compared with observations using standard categorical tests (e.g. 55, 141) or parametric tests (e.g. 74, 98). Such hybrid approaches may often prove useful where the underlying assumptions can be met, but in several of these cases, the standard Monte Carlo test might have been more defensible and straightforward.

Applications of the jackknife involving hypothesis tests or determination of confidence limits rely on parametric critical values and significance tables (see above). Though the bootstrap is not inherently tied to parametric methods, one area of active development is known as the parametric bootstrap, in which the standard error of the mean is bootstrapped and then used in parametric analyses as with the jackknife (e.g. see 67, 186, 231). Of course, the

**Table 4** Percentages of papers published in *Oecologia* (1985–1991) featuring some common nonparametric tests, all of which are (or are equivalent to) randomization tests.

Test	Percentage
Mann-Whitney	11.3
Spearman rank correlation	7.1
Wilcoxon matched-pairs signed-ranks	4.7
Fisher's exact test	2.7
Sign test	1.3
Binomial test	1.2

asymptotic convergence of statistical sampling distributions on the normal distribution at sufficiently large sample sizes is implied by the Central Limit Theorem.

DISCUSSION

*Advantages and Disadvantages of These Methods*

An attempt to sort out the pros and cons of resampling methods relative to the more standard statistical techniques raises many issues of varying subtlety and complexity (Table 5). When their stringent assumptions are met, parametric procedures maximize power (i.e. the chance of rejecting a false null hypothesis in favor of a true alternative), for a specified type 1 error rate (i.e. chance of falsely rejecting a true null) (186). But rarely at small or moderate sample sizes can all of the assumptions be known or convincingly demonstrated to apply. The conservative approach is then to resort to standard nonparametric methods or resampling.

Nonparametric methods are generally slightly to considerably weaker than the stronger of parametric and resampling methods for several reasons. Essentially all nonparametric techniques in common use were necessarily designed for minimizing computation. In some cases, this resulted in inherently low power (e.g. the sign test—see 135). In others, a loss of power or inadequate protection of the type 1 error rate may be attributed to reducing data to ranks, approximations related to ties in rank tests, continuity corrections at low frequencies for categorical tests, or the possibility of inaccurate approximations in some tables at intermediate numbers of observations (64). Often, the

**Table 5** Key features of three categories of statistical methods

Feature	Standard Parametric Methods	Standard Nonparametric Methods	Resampling Methods
Statistical power	High (when assumptions met)	Moderate	High
Known by researchers	Very widely	Widely	Sometimes & increasing
Acceptance	Widespread	Widespread	Common & increasing
Standardization	Very high	High	Moderate
Flexibility	Low	Moderate	High
Assumptions (see Table 1)	Moderate-strong (robust to some departures)	Moderate	Weak-moderate
Population or sample	Population	Sample	Population (except randomization)
Time & effort cost	Moderate	Somewhat lower	Higher & decreasing
Conceptual complexity	High	Moderate	Low-moderate

available tables seriously constrain the analyses by incompleteness (e.g. Friedman's ANOVA-by-ranks), by providing only one-tailed or only two-tailed significance values (Fisher's exact test and chi-square, respectively), or by providing only a rough indication of the  $p$ -value's magnitude (most nonparametric tests) (see 64).

In contrast, randomization tests yield about the same significance level as parametric methods when the parametric assumptions are met (115, 218) but may have more power than parametric methods when data are from non-normal distributions (64, p. 94; 135). Less is known about circumstances in which Monte Carlo, bootstrap, and jackknife methods may be more powerful than standard parametric analysis (but see examples in 67).

An obvious current advantage of using standard parametric and nonparametric techniques is that they are widely known and accepted by editors and other researchers, though resampling methods are now clearly in common use as well. Yet bootstrapping may have been swept into the mainstream of ecological and particularly evolutionary research somewhat ahead of a full, balanced evaluation of its capabilities and shortcomings. Bootstrap confidence limits and hypothesis testing are not always reliable (e.g. see 67); some familiarity with current methodological advances and perhaps some ad-hoc checking could prove important, and use should be restricted to cases where randomization tests and parametric methods are inappropriate (163, 186).

Standardization and flexibility of statistical methods must trade off, to some extent. The psychological shift associated with the conceptually simple resampling approach, in which the data analyst necessarily controls and understands each step from hypothesis formation to designing a sufficiently powerful test statistic (and perhaps the test itself) to calculating an intuitively meaningful  $p$ -value, can be "liberating" (186). Otherwise infeasible experimental designs (e.g. those based on nonrandom sampling or requiring nonstandard response variables) become available with resampling methods. But this degree of versatility carries the cost that other defensible test statistics or procedures for conducting the test itself may lead to different conclusions (e.g. the survey of "null models" above)—or as a worst case even undermine the objectivity of the data analysis (see 9 and 113). It is thus important to consider a range of alternative test statistics and procedures and to justify the choices made, ideally before the data are analyzed. It should be clear that the appropriate alternative hypothesis would indeed be supported by rejecting the null; an instructive case is the randomization test of differences between means, in which the null hypothesis can be rejected for identical means but different variances (242).

The striking differences in assumptions underpinning the classical and resampling methods (Table 1; 186, p. 84-92) necessarily constrain the options to an extent that is often overlooked or ignored by editors, referees, and

researchers. Making dubious assumptions should obviously be avoided where possible, but this should be balanced against the tacit assumptions involved in using some of the more speculative resampling methods (particularly the bootstrap and jackknife).

Randomization tests (and by implication their derivative nonparametric tests) apply only to the samples themselves, not to some underlying population that may have been sampled. This obviates the need for random samples and for certain assumptions about the population of origin, but it also lengthens the interpretive extrapolation from the observed results to the general situation or population of interest. Some such extrapolation is almost always necessary in any case, and this problem is of greater theoretical than applied relevance (64).

Finally, resampling methods currently require on average more expenditure of time and effort per analysis than do classical methods, largely because of the necessary computer programming. In fact, the required programming is often quite straightforward, and programs for many common applications are widely available in the literature (e.g. 64, 163, 186) or as shareware. Soon, commercial mainframe and microcomputer packages will be available (186).

### *Implications for Interpreting and Communicating Results*

In hypothesis testing, the interpretation of results necessarily hinges on the  $p$ -value, or at least on its magnitude relative to the critical value. An attractive feature of resampling methods is that the direct calculation of the  $p$ -value obviates the discrete decision-theory distinction between significant and nonsignificant results, cleft sharply if arbitrarily at a knife-edge critical value. Instead, the  $p$ -value can simply be understood to measure the degree of consistency between the data and the null hypothesis, though the classical significance levels (0.05, 0.01, etc) retain their utility as benchmarks. Moreover, the directly calculated  $p$ -value may be much easier to communicate to nontechnical decision-makers; as noted in the introduction, in a randomization test of a difference between two means, the  $p$ -value is simply the proportion of random assignments of data to treatments that gives a difference between group averages at least as large as the difference obtained in the experiment (64, p. 10).

As with other statistical parameter estimates, an error estimate for the  $p$ -value is desirable, particularly where this error reflects only a moderate number of repetitions (e.g. 1000) for a resampling method (sampled randomization, Monte Carlo, or bootstrapping). In the latter case, the  $100(1-\alpha)\%$  confidence interval is well approximated by  $p \pm t_{1-\alpha}(p(1-p)/n)^{1/2}$  with an infinite number of degrees of freedom, where  $\alpha$  is the significance level,  $t_{1-\alpha}$  is the critical value of the  $t$  distribution for significance level  $\alpha$ , and  $n$  is the number of repetitions (e.g. see 186, p. 34). Note that these error bounds on

the  $p$ -value reflect variation derived from the intensity of resampling. More speculatively, it may be possible in some cases to obtain error bounds on the  $p$ -value associated with empirical sampling variation using the bootstrap or jackknife, but I am unaware that this has yet been attempted.

In research reports, error estimates via resampling generally require little fanfare—only the name of the method and the number of repetitions (if applicable). Hypothesis testing requires more information in the methods section, including specification and justification of the null and alternative hypotheses. Unless the approach is exotic or particularly central to the presentation, references to support resampling methods are unnecessary. The method, number of repetitions (if applicable), magnitude of the test statistic, and  $p$ -value (with confidence interval when appropriate) generally appear parenthetically with results of a hypothesis test.

### *Methodological Issues That Deserve Attention*

Much statistical research remains to be done to develop, improve, and evaluate these resampling techniques. Under the general theme of robustness, both parametric and resampling methods need to be further examined and compared in their sensitivity to non-normality, non-equivalence of distributions (e.g. unequal variances), and sample size. Also, what are the implications for randomization tests of nonrandom assignment to treatment levels and for other resampling and parametric tests of nonrandom sampling? Ideally, such studies should focus on features typical of small-to-moderate-sized samples and the ways such data are actually gathered, rather than exclusively on the characteristics of large statistical populations from which such samples may be drawn (64).

Until resampling methods became generally feasible relatively recently, the available statistical methods were sufficiently constraining that the formulation of null and alternative hypotheses has been relatively straightforward and unambiguous. Now that the horizon for these hypotheses has widened considerably, perhaps new and useful guidelines can be devised that will help practitioners to match hypotheses more effectively to tests and test statistics.

Many of the more complex descriptive and hypothesis-testing techniques traditionally based on relatively assumption-bound parametric methods can be effectively refitted as resampling methods (see above and 163, 165). Of particular importance in ecology and evolution are methods that relate directly to common experimental designs, like analysis of variance. ANOVA has traditionally been considered robust to departures from the standard parametric assumptions, though not all agree (e.g. 18, 243), and more complex variations (e.g. factorial ANOVA, ANCOVA, MANOVA) can be more vulnerable to violations of parametric assumptions (e.g. 191). In parallel with further studies of robustness, the development of resampling ANOVA and its variations

should continue (e.g. 8, jackknife ANOVA; 130, randomization ANOVA; 162, bootstrap ANOVA), and controversies like contrasting views on interaction terms in factorial ANOVA (see 64 vs 163) need to be resolved.

Further statistical research on Monte Carlo methods and especially on bootstrapping will continue to attract immediate interest and additional applications in ecology and evolution.

### *Some Specific Recommendations*

1. *Resampling methods should be part of basic statistical training in ecology and evolution.* At least until these methods are incorporated into mainframe statistics packages, this will require some computer-programming skills as well. In exchange, perhaps less emphasis can be placed on standard nonparametric methods.
2. *Parameter estimates should be accompanied by estimates of the associated variation.* Resampling methods make it possible for this principle to be very broadly (if not universally) applied.
3. *With small-to-moderate sample sizes, maintain a healthy skepticism about the appropriateness of parametric analysis.* Even failures to reject normality and equal variances as null hypotheses are rarely conclusive, since the power of tests to evaluate them is low at the relevant sample sizes. When the random process that generated the data is statistically uncharacterized, the conservative approach is to use defensible methods making the fewest strong and unverifiable assumptions.
4. *Transformations should be used to improve the equivalence of distributions in randomization tests in essentially the same way that these are used in parametric analyses.* This should help neutralize a potential problem with standard randomization methods that has often been unrecognized.
5. *Where equivalence of distributions is unlikely to hold or to be achieved by transformation, multisample hypothesis tests can be conducted by bootstrapping.* With this approach (termed "bootstrapped randomization" in 186), data for different treatment levels are bootstrapped independently before the test statistic is calculated. In our present state of ignorance, bootstrapping should not ordinarily be used where parametric or randomization methods apply.
6. *For confidence intervals and hypothesis testing, other more theoretically defensible methods should generally be used instead of the jackknife.* The jackknife is particularly useful for eliminating bias in parameter estimates (67), as a check or extension of the bootstrap (68), or in cases where the other heavily computation-intensive methods are not feasible (67).
7. *Where possible, researchers should attempt to ensure that their exper-*

*imental units constitute a random sample of some population of interest.* Though it may often be unclear how to accomplish this, the object is to retain defensible options for the statistical analysis. When the case can be made, it may be useful to explicitly identify the population that has been sampled randomly.

8. *In the absence of random sampling, hypotheses comparing two or more samples with equivalent distributions should be tested by randomization.* Randomization can also be used to construct confidence intervals, though these tend to be relatively conservative (64).
9. *Use randomization, Monte Carlo, or bootstrapping methods instead of standard nonparametric methods, particularly when maximizing power is essential.* Any nonparametric test can be replaced by a potentially more flexible and powerful but otherwise equivalent resampling test.
10. *Whenever possible and appropriate, use a large number of repetitions in resampling tests ( $\geq 20000$ ).* This is particularly important when it can influence the way the data are interpreted (e.g. when the  $p$ -value is near 0.05). For randomization,  $n = 1000$  and  $n = 5000$  are generally considered minimal for tests at the 5% and 1% significance levels, respectively (64, 65, 167).
11. *When using resampling methods, define null and alternative hypotheses with special care.* Justify these choices in the methods section of the research report.
12. *More attention should be paid to the applicability of assumptions underlying statistical analyses by researchers, editors, and referees.* With resampling methods becoming widely known and commonly used, standards for acceptably thorough and rigorous data analysis should continue to rise.

A summary of some common situations arising in data analysis and the most appropriate methods for dealing with them is presented in Table 6.

#### ACKNOWLEDGMENTS

It is a pleasure to thank some people who have aided the preparation of this review and development of the ideas herein: Peter Chesson, Dan Johnson, and Jerry Nagle for introducing me to randomization and bootstrapping; Eugene Edgington and Bryan Manly for encouragement and for breaking the ground so effectively; Ken Hudson for helping with the BIOSIS and journal searches of the literature; Bryan Manly, Brian McArdle, Cidambi Srinivasan, Clare Veltman, and participants in the 1992 course on experimentation in benthic ecology sponsored by the University of Umeå, Sweden, for stimulating discussions; Mick Crawley, Dan Johnson, John Lawton, Bryan Manly, Brian McArdle, and Clare Veltman for reading and commenting on the manuscript; Mick Crawley for letting me rummage at will through his journals; Brenda

**Table 6** Recommended methods.

Purpose	Sampling	Empirical sample size <sup>1</sup>	Distribution(s) (or underlying model)	Preferred method
Confidence intervals and single-sample hypothesis tests	Random	Large	Any	Parametric
	Random	Small-moderate	Normal	Parametric
	Random	Small-moderate	Known, but non-normal <sup>2</sup>	Monte Carlo
	Random	Small-moderate	Unknown <sup>2</sup>	Bootstrapping
	Non-random	Any	Any	Randomization <sup>3</sup>
Multisample hypothesis tests	Random	Any	Equivalent	Parametric
	Random	Any	Non-equivalent <sup>2</sup>	Bootstrapping
	Non-random	Any	Equivalent	Randomization
	Non-random	Any	Non-equivalent <sup>2</sup>	---

<sup>1</sup> In practice, large and moderate sample sizes are generally distinguished subjectively.

<sup>2</sup> Transformations of the data can at least sometimes reshape the distribution(s) adequately for parametric analysis (with random sampling) or randomization tests (with non-random sampling), though this may prove difficult to demonstrate convincingly.

<sup>3</sup> Awkward to implement, and tends to yield conservative confidence limits.

and Chauncey Curtz for new perspectives on null hypotheses; Production Editor Nancy Donham for grace under pressure; Marcel Dekker, Inc., for permission to use a slightly modified version of E.S. Edgington's diagram as Figure 1; John Lawton for hosting my sabbatical year at the Centre for Population Biology; and Lillie, Sarah, and Martin for being patient and understanding when the crunch came. I acknowledge funding from the National Science Foundation grant INT 9014938 and a Visiting Research Fellowship from the Royal Society of London.

#### APPENDIX: HOW THE DATA OF TABLE 2 WERE ANALYZED

I used randomization and a Monte Carlo method to test two a-priori hypotheses concerning the data summarized in Table 2. The rationale and procedure used in each case are briefly described here as examples of how this general approach can be implemented.

*Hypothesis 1: The focal methods considered in this review are becoming more commonly used over the last several years in ecology and evolution.*

The BIOSIS literature search over the six publication years 1985–1990 on topics in ecology and evolution (Table 2) provides the basis for a test. To assure adequate sample sizes for each separate method, references fitting either or both of these categories for a particular publication year were pooled, yielding publication frequencies for each of the six years to be tested for trend.

Hypothesis 1 predicts a positive trend in publication frequency over years, generating a one-tailed test against the null hypothesis of no trend. I chose the linear regression coefficient  $b$  or least-squares slope of publication frequency as a function of publication year to be the response measure. (It



can be demonstrated that using the correlation coefficient  $r$  for this purpose necessarily produces identical results; moreover, the much simpler metric  $\sum x_i y_i$ , where  $x_i$  and  $y_i$  are the coordinates of the  $i$ th data point, is equivalent to either of these and was the measure actually used in my tests—see 163, pp. 91–92.) Testing Hypothesis 1 then required calculating the observed  $b = b_0$  (or the simpler equivalent) and demonstrating that a value as large or larger is very unlikely to have arisen by chance alone.

Now if a trend-free null process generated the observed sequence, then there should be nothing special about the order in which the publication frequencies were observed; any reordering should produce a statistically equivalent sequence (163, p. 92). There are  $6! = 720$  different orderings of six numbers. Calculating a regression coefficient  $b$  (or equivalent) for each and comparing it with  $b_0$  indicated the proportion of these as large or larger than  $b_0$ . This proportion was taken to be the probability  $p$  that the observed sequence could have been generated by the same kind of trend-free process that produced the other 719 sequences. When  $p$  was smaller than the relevant significance level (generally 0.05), I rejected the null hypothesis in favor of the alternative (i.e. Hypothesis 1); otherwise I was unable to reject the null, and Hypothesis 1 was not supported.

This exemplifies systematic randomization, which determines exact  $p$ -values. My Pascal computer program (written in Turbo Pascal 6.0) to calculate  $b$ ,  $r$ , and  $p$  was just over 100 lines long, the majority of which were needed to generate the 720 reorderings of the data. The solution could instead have been found by sampled randomization using a much simpler algorithm (see 163), half as many program lines, and some additional run time (roughly 30 sec rather than a fraction of a second on a typical 386/387 microcomputer).

*Hypothesis 2: The focal methods considered in this review are used in differing frequency in ecological versus evolutionary studies.*

The data of Table 2 permit this hypothesis to be tested for each method. Only the data in the “ecology” and “evolution” rows of the table were used. First, the few papers within years classified as both ecological and evolutionary were removed from the observed frequencies, eliminating a source of positive correlation. The residuals were then summed along rows (i.e. over years), and the overall proportion  $v_0$  of evolutionary studies out of all ecological and evolutionary studies for the given method was calculated as the evolution row sum divided by the sum of the evolution row sum and the ecology row sum. (This is of course equivalent to using the overall proportion of ecological studies, which is simply  $1 - v_0$ .)

To determine whether  $v_0$  corroborated Hypothesis 2, I used a Monte Carlo method to generate a null distribution of  $v$ -values for comparison with  $v_0$ . If a given method were used just as frequently in both kinds of studies, then the observed frequencies in a particular year could have been generated as a sample from a binomial process based on the frequencies of all ecology and

all evolution studies in that year. I thus randomly generated the number of studies observed in each year, with the chance of any particular study's being "evolutionary" equalling the "evolutionary" proportion in the given year, and I tallied the distribution between evolution and ecology. When these distributions had been simulated for all six years of the sequence, the overall proportion  $v$  of evolutionary studies was found as above from the evolution and ecology row totals. Each of 20,000 such  $v$ -values was determined, and the proportion as extreme or more extreme than  $v_0$  became the  $p$ -value estimate. Though this  $p$ -value was derived from a random sample of  $v$ -values and was therefore inexact, the large number of iterations assured adequate precision. (See 163, pp. 32–36. The precision of this  $p$ -value could be assessed using the confidence-interval calculation described in the Discussion subsection entitled Implications for Interpreting and Communicating Results.)

Special provision must be made for two-tailed tests of hypotheses like Hypothesis 2. I noted whether  $v_0$  was larger or smaller than the overall expected value or mean of  $v$ , setting the computer program to calculate  $p$  from  $v$ -values equally far or farther from the mean. As is typical for two-tailed tests, this required that the significance level be halved to 0.025. Thus if 499 or fewer of the  $v$ -values were as extreme or more extreme than  $v_0$  (as in all cases in Table 2), then the null hypothesis was rejected and hypothesis 2 corroborated. My Pascal program to calculate  $v_0$  and  $p$  was just over 50 lines and was straightforward to write; the substantial number of iterations required about 10–60 sec to run on a 386/387 microcomputer, depending on the observed publication frequencies for the method of interest.

### Literature Cited

1. Abrahamson, W. G., Sattler, J. F., McCrea, K. D., Weis, A. E. 1989. Variation in selection pressures on the goldenrod gall fly and the competitive interactions of its natural enemies. *Oecologia* 79:15–22.
2. Abrams, P., Nyblade, C., Sheldon, S. 1986. Resource partitioning and competition for shells in a subtidal hermit crab species assemblage. *Oecologia* 69: 429–45.
3. Althawadi, A. M., Grace, J. 1986. Water use by the desert cucurbit *Citrullus colocynthis* (L.) Schrad. *Oecologia* 70:475–80.
4. Antonovics, J., Clay, K., Schmitt, J. 1987. The measurement of small-scale environmental heterogeneity using clonal transplants of *Anthoxanthum odoratum* and *Danthonia spicata*. *Oecologia* 71:601–7.
5. Antyufeev, V. S., Marshak, A. L. 1990. Inversion of Monte Carlo model for estimating vegetation canopy parameters. *Remote Sens. Environ.* 33:201–10.
6. Archie, J. W. 1989. A randomization test for phylogenetic information in systematic data. *Syst. Zool.* 38:239–52.
7. Archie, J. W. 1989. Phylogenies of plant families: A demonstration of phylogenetic randomness in DNA sequence data derived from proteins. *Evolution* 43:1796–800.
8. Arveson, J. N., Schmitz, T. H. 1970. Robust procedures for variance component problems using the jackknife. *Biometrics* 26:677–86.
9. Basu, D. 1980. Randomization analysis of experimental data: the Fisher randomization test. *J. Am. Stat. Assoc.* 75:575–82.
10. Benjamini, Y., Harpaz, Y. 1986. Observational rainfall runoff analysis for estimating effects of cloud seeding on

- water resources in northern Israel. *J. Hydrol.* 83:299-306
11. Berryman, A. A. 1991. Stabilization or regulation: what it all means. *Oecologia* 86:140-43
  12. Besag, J., Clifford, P. 1991. Sequential Monte Carlo *p*-values. *Biometrika* 78: 301-4
  13. Black, A. R., Dodson, S. I. 1990. Demographic costs of *Chaoborus*-induced phenotype plasticity in *Daphnia pulex*. *Oecologia* 83:117-22
  14. Block, W. M., Brennan, L. A., Gutierrez, R. J. 1991. Ecomorphological relationships of a guild of ground-foraging birds in northern California, USA. *Oecologia* 87:449-58
  15. Boik, R. J. 1987. The Fisher-Pitman permutation test: a non-robust alternative to the normal theory F-test when variances are heterogeneous. *Br. J. Math. Stat. Psychol.* 40:26-42
  16. Boyajian, G. E. 1991. Taxon age and selectivity of extinction. *Paleobiology* 17:49-57
  17. Bradley, J. V. 1968. *Distribution-Free Statistical Tests*. Englewood Cliffs, NJ: Prentice-Hall
  18. Bradley, J. V. 1978. Robustness? *Br. J. Math. Stat. Psychol.* 31:144-52
  19. Brandl, R., Topp, W. 1985. Size structure of *Pterostichus* spp. (Carabidae): aspects of competition. *Oikos* 44:234-38
  20. Breck, J. E., DeAngelis, D. L., Van Winkle, W., Christensen, S. W. 1988. Potential importance of spatial and temporal heterogeneity in pH, aluminum, and calcium in allowing survival of a fish population: a model demonstration. *Ecol. Modell.* 41:1-16
  21. Breshears, D. D., Kirchner, T. B., Otis, M. D., Whicker, F. W. 1989. Uncertainty in predictions of fallout radionuclides in foods and of subsequent ingestion. *Health Phys.* 57:943-53
  22. Briggs, J. M., Knapp, A. K. 1991. Estimating aboveground biomass in tallgrass prairie with the harvest method: determining proper sample size using jackknifing and Monte Carlo simulations. *Southwest Nat.* 36:1-6
  23. Brokaw, N. V. L., Scheiner, S. M. 1989. Species composition in gaps and structure of a tropical forest. *Ecology* 70:538-41
  24. Brueggemann, R., Trapp, S., Matthies, M. 1991. Behavior assessment of a volatile chemical in the Rhine River. *Environ. Toxicol. Chem.* 10:1097-104
  25. Brusasca, G., Tinarelli, G., Anfossi, D. 1989. Comparison between the results of a Monte Carlo atmospheric diffusion model and tracer experiments. *Atmos. Environ.* 23:1263-80
  26. Bulmer, M. G. 1975. The statistical analysis of density dependence. *Biometrics* 31:901-11
  27. Cain, M. L. 1990. Patterns of *Solidago altissima* ramet growth and mortality: the role of below-ground ramet connections. *Oecologia* 82:201-9
  28. Canale, R. P., Effler, S. W. 1989. Stochastic phosphorus model for Onondaga Lake, New York, USA. *Water Res.* 23:1009-16
  29. Capone, T. A., Kushlan, J. A. 1991. Fish community structure in dry-season stream pools. *Ecology* 72:983-92
  30. Carnes, B. A., Slade, N. A. 1988. The use of regression for detecting competition with multicollinear data. *Ecology* 69:1266-74
  31. Carpenter, S. R. 1990. Large-scale perturbations: opportunities for innovation. *Ecology* 71:2038-43
  32. Carpenter, S. R., Frost, T. M., Heisey, D., Kratz, T. K. 1989. Randomized intervention analysis and the interpretation of whole-ecosystem experiments. *Ecology* 70:1142-52
  33. Casas, J. 1990. Multidimensional host distribution and nonrandom parasitism: a case study and a stochastic model. *Ecology* 71:1893-903
  34. Case, T. J. 1990. Patterns of coexistence in sexual and asexual species of *Cnemidophorus* lizards. *Oecologia* 83: 220-27
  35. Caswell, H. L. 1976. Community structure: a neutral model analysis. *Ecol. Monogr.* 46:327-54
  36. Caswell, H. 1989. *Matrix Population Models*. Sunderland, Mass: Sinauer
  37. Chapin, F. S. III, Groves, R. H., Evans, L. T. 1989. Physiological determinants of growth rate in response to phosphorus supply in wild and cultivated *Hordeum* species. *Oecologia* 79:96-105
  38. Chuh, H. T., Tan, H. S. 1990. A multiconstituent and multilayer microwave backscatter model for a vegetative medium. *Remote Sens. Environ.* 31: 137-54
  39. Clark, P. J., Evans, F. C. 1954. Distance to nearest neighbor as a measure of spatial relationships in populations. *Ecology* 35:445-53
  40. Cochran, W. G. 1954. Some methods for strengthening the common  $\chi^2$  tests. *Biometrics* 10:417-51
  41. Colwell, R. K., Winkler, D. W. 1984. A null model for null models in biogeography. In *Ecological Communities: Conceptual Issues and the Evidence*,

- ed. D. R. Strong, D. Simberloff, L. G. Abele, A. B. Thistle, pp. 344–59. Princeton, NJ: Princeton Univ. Press
42. Connor, E. F., Simberloff, D. 1979. The assembly of species communities: Chance or competition? *Ecology* 60: 1132–40
43. Connor, E. F., Simberloff, D. 1984. Neutral models of species' co-occurrence patterns. See Ref. 41, pp. 316–31
44. Connor, E. F., Simberloff, D. 1984. Rejoinder. See Ref. 41, pp. 41–43
45. Corbin, K. W., Livezey, B. C., Humphrey, P. S. 1988. Genetic differentiation among steamer-ducks (Anatidae: Tachyeres): an electrophoretic analysis. *Condor* 90:773–81
46. Cotton, J. W. 1973. Even better than before. *Contemp. Psychol.* 18:168–69
47. Cracraft, J. 1991. Patterns of diversification within continental biotas: hierarchical congruence among the areas of endemism of Australian vertebrates. *Aust. Syst. Bot.* 4:211–27
48. Crowley, P. H. 1992. Density dependence, boundedness, and attraction: Detecting stability in stochastic systems. *Oecologia* 90:246–54
49. Crowley, P. H., Gillett, S., Lawton, J. H. 1988. Contests between larval damselflies: empirical steps toward a better ESS model. *Anim. Behav.* 36:1496–511
50. Crowley, P. H., Johnson, D. M. 1982. Habitat and seasonality as niche axes in an odonate community. *Ecology* 63: 1064–77
51. Crowley, P. H., Johnson, D. M. 1992. Variability and stability of a dragonfly assemblage. *Oecologia* 90:260–69
52. DeAngelis, D. L., Gross, L. J., eds. 0000. *Individual-Based Approaches and Models in Ecology*. New York: Chapman-Hall
53. den Boer, P. J. 1989. Comment on the article "On testing temporal niche differentiation in carabid beetles," by M. Loreau. *Oecologia* 81:97–98
54. den Boer, P. J. 1990. On the stabilization of animal numbers. Problems of testing. 3. What do we conclude from significant tests? *Oecologia* 83:38–46
55. den Boer, P. J. 1991. Seeing the trees for the wood: random walks or bounded fluctuations of population size? *Oecologia* 86:484–91
56. Diamond, J. M., Gilpin, M. E. 1982. Examination of the "null" model of Connor and Simberloff for species co-occurrences on islands. *Oecologia* 52: 64–74
57. DiCiccio, T. J., Romano, J. P. 1989. A review of bootstrap confidence intervals. *J. R. Stat. Soc. B* 50:338–54
58. Dorazio, R. M. 1986. Estimating population birth rates of zooplankton when rates of egg deposition and hatching are periodic. *Oecologia* 69:532–41
59. Douwes, P., Stille, B. 1988. Selective versus stochastic processes in the genetic differentiation of populations of the butterfly *Erebia embla* Thnbg. (Lepidoptera: Satyridae). *Hereditas* 109:37–44
60. Dressing, S. A., Spooner, J., Kreglow, J. M., Beasley, E. O., Westerman, P. W. 1987. Water and sediment sampler for plot and field studies. *J. Environ. Qual.* 16:59–64
61. DuBow, P. J. 1988. Waterfowl communities and seasonal environments: temporal variability in interspecific competition. *Ecology* 69:1439–53
62. Duff, K. E., Smol, J. P. 1991. Morphological descriptions and stratigraphic distributions of the chrysophycean stomatocysts from a recently acidified lake, Adirondack Park, New York, USA. *J. Paleolimnol.* 5:73–113
63. Dwyer, G. 1991. The roles of density, stage, and patchiness in the transmission of an insect virus. *Ecology* 72:559–74
64. Edgington, E. S. 1987. *Randomization Tests*. New York: Marcel Dekker. 2nd ed.
65. Edwards, D. 1985. Exact simulation-based inference: a survey, with additions. *J. Stat. Comput. Simul.* 22:307–26
66. Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Ann. Stat.* 7:1–26
67. Efron, B. 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*. Philadelphia: SIAM
68. Efron, B. 1992. Jackknife-after-bootstrap standard errors and influence functions. *J. Roy. Stat. Soc. B* 54:83–127
69. Efron, B., Tibshirani, R. 1986. Bootstrap methods for standard errors, confidence intervals, and other methods of statistical accuracy. *Stat. Sci.* 1:54–77
70. Eisenberg, R. M. 1966. The regulation of density in a natural population of the pond snail, *Lymnaea elodes*. *Ecology* 47:889–906
71. Faust, K., Romney, A. K. 1985. The effect of skewed distributions on matrix permutation tests. *Br. J. Math. Stat. Psychol.* 38:152–60
72. Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–91
73. Felsenstein, J. 1988. Phylogenies from

- molecular sequences: Inference and reliability. *Annu. Rev. Genet.* 22:521–65
74. Firbank, L. G., Watkinson, A. R. 1987. On the analysis of competition at the level of the individual plant. *Oecologia* 71:308–17
75. Fisher, N. I., Hall, P. 1990. On bootstrap hypothesis testing. *Aust. J. Stat.* 32: 177–90
76. Fisher, N. I., Hall, P. 1991. Bootstrap algorithms for small samples. *J. Stat. Plan. Inference* 27:157–69
77. Fisher, R. A. 1936. The coefficient of racial likeness and the future of craniometry. *J. R. Anthropol. Inst.* 66:57–63
78. Fix, A. G. 1984. Kin groups and trait groups: population structure and epidemic disease selection. *Am. J. Phys. Anthropol.* 65:201–12
79. Foley, P. 1987. Molecular clock rates as loci under stabilizing selection. *Proc. Natl. Acad. Sci. USA* 84:7996–8000
80. Forina, M., Armanino, C., Leardi, R., Drava, G. 1991. A class-modelling technique based on potential functions. *J. Chemometrics* 5:435–54
81. Gallagher, E. D., Gardner, G. B., Jumars, P. A. 1990. Competition among the pioneers in a seasonal soft-bottom benthic succession: field experiments and analysis of the Gilpin-Ayala competition model. *Oecologia* 83:427–42
82. Garthwaite, P. H., Buckland, S. T. 1992. Generating Monte Carlo confidence intervals by the Robbins-Monro process. *Appl. Stat.* 41:159–71
83. Gascon, C. 1991. Population- and community-level analyses of species occurrences of central Amazonian rainforest tadpoles. *Ecology* 72:1731–46
84. Gaston, K. J., Lawton, J. H. 1987. A test of statistical techniques for detecting density dependence in sequential censuses of animal populations. *Oecologia* 74:404–10
85. Gates, J. 1991. Exact Monte Carlo tests using several statistics. *J. Stat. Comput. Simul.* 38:211–18
86. Gebhardt, M., Ribi, G. 1987. Reproductive effort and growth in the proboscbranch snail, *Viviparus ater*. *Oecologia* 74:209–14
87. Gerrodette, T. 1991. Models for power of detecting trends—a reply to Link and Hatfield. *Ecology* 72:1889–92
88. Gilbert, J. J. 1990. Differential effects of *Anabaena affinis* on cladocerans and rotifers: mechanisms and implications. *Ecology* 71:1727–40
89. Gilinsky, N. L., Bamback, R. K. 1986. The evolutionary bootstrap: a new approach to the study of taxonomic diversity. *Paleobiology* 12:251–68
90. Gillespie, J. H. 1986. Variability of evolutionary rates of DNA. *Genetics* 113:1077–92
91. Gilpin, M. E., Diamond, J. M. 1982. Factors contributing to non-randomness in species co-occurrences on islands. *Oecologia* 52:75–84
92. Gilpin, M. E., Diamond, J. M. 1984. Are species co-occurrences on islands non-random, and are null hypotheses useful in community ecology? See Ref. 41, pp. 297–315
93. Gilpin, M. E., Diamond, J. M. 1984. Rejoinder. See Ref. 41, pp. 332–41
94. Gordon, H. R. 1991. Absorption and scattering estimates from irradiance measurements: Monte Carlo simulations. *Limnol. Oceanogr.* 36:769–77
95. Gotelli, N. J. 1991. Demographic models for *Leptogorgia virgulata*, a shallow-water gorgonian. *Ecology* 72: 457–67
96. Gotelli, N. J., Lewis, F. G. III, Young, C. M. 1987. Body-size differences in a colonizing amphipod-mollusc assemblage. *Oecologia* 72:104–8
97. G-Valdecasas, A., Baltanas, A. 1990. Jackknife and bootstrap estimation of biological index of water quality. *Water Res.* 24:1279–84
98. Haefner, J. W. 1988. Niche shifts in Greater Antillean *Anolis* communities: effects of niche metric and biological resolution on null model tests. *Oecologia* 77:107–17
99. Haith, D. A. 1986. Simulated regional variations in pesticide runoff. *J. Environ. Qual.* 15:5–8
100. Hall, P., Wilson, S. R. 1991. Two guidelines for bootstrap hypothesis testing. *Biometrics* 47:757–62
101. Hanna, S. R. 1988. Air quality model evaluation and uncertainty. *J. Air Pollut. Control Assoc.* 38:406–12
102. Harpending, H., Rogers, A. 1987. On Wright's mechanism for intergroup selection. *J. Theor. Biol.* 127:51–62
103. Harrison, S. 1989. Long-distance dispersal and colonization in the Bay Checkerspot butterfly, *Euphydryas editha bayensis*. *Ecology* 70:1236–43
104. Harvey, B. C., Stewart, A. J. 1991. Fish size and habitat depth relationships in headwater streams. *Oecologia* 87: 336–42
105. Harvey, L. E., Davis, F. W., Gale, N. 1988. The analysis of class dispersion patterns using matrix comparisons. *Ecology* 69:537–42
106. Hastings, A. 1987. Substitution rates

- under stabilizing selection. *Genetics* 116:479–86
107. Hayward, G. D., Garton, E. O. 1988. Resource partitioning among forest owls in the River of No Return Wilderness, Idaho. *Oecologia* 75:253–65
  108. Heidam, N. Z. 1987. Bootstrap estimates of factor model variability. *Atmos. Environ.* 21:1203–18
  109. Heller, K.-G., von Helversen, O. 1989. Resource partitioning of sonar frequency bands in rhinolophoid bats. *Oecologia* 80:178–86
  110. Heltsh, J. F., Forrester, N. E. 1985. Statistical evaluation of the jackknife estimate of diversity when using quadrat samples. *Ecology* 66:107–11
  111. Henle, K. 1989. Population ecology and life history of the diurnal skink *Morethia boulengeri* in arid Australia. *Oecologia* 78:521–32
  112. Hinkley, D. V. 1978. Improving the jackknife with special reference to correlation estimation. *Biometrika* 65:13–22
  113. Hinkley, D. V. 1980. Discussion of Basu's paper. *J. Am. Stat. Assoc.* 75: 582–84
  114. Hinkley, D. V. 1983. Jackknife methods. *Encyclopedia Stat. Sci.* 4:280–87
  115. Hoeffding, W. 1952. The large-sample power of tests based on permutations of observations. *Ann. Math. Stat.* 23: 169–92
  116. Holman, E. W. 1989. Some evolutionary correlates of higher taxa. *Paleobiology* 15:357–63
  117. Hopmans, J. W., Guttierrez-Rave, E. 1988. Calibration of a root water uptake model in spatially variable soils. *J. Hydrol.* 103:53–66
  118. Horvitz, C. C., Schemske, D. W. 1990. Spatiotemporal variation in insect mutualists of a neotropical herb. *Ecology* 71:1085–97
  119. Hubert, L. J., Schultz, J. 1976. Quadratic assignment as a general data analysis strategy. *Br. J. Math. Stat. Psychol.* 29:190–241
  120. Hutchinson, G. E. 1959. Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* 93: 145–59
  121. Ito, D., Oikawa, T. 1989. A simulation study on characteristics of light-interception and carbon dioxide assimilation in a densely planted mulberry field and their regional variation. *J. Sericul. Sci. Jpn.* 58:131–39
  122. Jackson, D. A., Harvey, H. H. 1989. Biogeographic associations in fish assemblages: local vs. regional processes. *Ecology* 70:1472–84
  123. Jackson, D. A., Somers, K. M. 1991. The spectre of "spurious" correlations. *Oecologia* 86:147–51
  124. Jaksic, F. M., Medel, R. G. 1990. Objective recognition of guilds: testing for statistically significant species clusters. *Oecologia* 82:87–92
  125. James, C. D. 1991. Temporal variation in diets and trophic partitioning by coexisting lizards (Ctenotus: Scincidae) in central Australia. *Oecologia* 85:553–61
  126. Jansen, R. K., Holsinger, K. E., Michaels, H. J., Palmer, J. D. 1991. Phylogenetic analysis of chloroplast DNA restriction site data at higher taxonomic levels: an example from the Asteraceae. *Evolution* 44:2089–105
  127. Jansen, R. K., Wallace, R. S., Kim, K.-J., Chambers, K. L. 1991. Systematic implications of chloroplast DNA variation in the subtribe Microseridinae (Asteraceae: Lactuceae). *Am. J. Bot.* 78:1015–27
  128. Jenkins, A., Whitehead, P. G., Musgrove, T. J., Cosby, B. J. 1990. A regional model of acidification in Wales, UK. *J. Hydrol.* 116:403–16
  129. Joern, A., Lawlor, L. R. 1981. Guild structure in grasshopper assemblages based on food and microhabitat resources. *Oikos* 37:93–104
  130. Johnson, D. M., Pierce, C. L., Martin, T. H., Watson, C. N., Bohanan, R. E., Crowley, P. H. 1987. Prey depletion by odonate larvae: combining evidence from multiple field experiments. *Ecology* 68:1459–65
  131. Johnson, S. G., Johnston, R. F. 1989. A multifactorial study of variation in interclutch interval and annual reproductive success in the feral pigeon, *Columba livia*. *Oecologia* 80:87–92
  132. Jones, M. L., Minns, C. K., Marmorek, D. R., Heltcher, K. J. 1991. Assessing the potential extent of damage to inland lakes in eastern Canada due to acidic deposition. IV. Uncertainty analysis of a regional model. *Can. J. Fish. Aquat. Sci.* 48:599–606
  133. Jurik, T. W. 1991. Population distributions of plant size and light environment of giant ragweed (*Ambrosia trifida* L.) at three densities. *Oecologia* 87:539–50
  134. Kempthorne, O. 1955. The randomization theory of experimental inference. *J. Am. Stat. Assoc.* 50:946–67
  135. Kempthorne, O., Doerfler, T. E. 1969. The behaviour of some significance tests under experimental randomization. *Biometrika* 56:231–48
  136. Kenkel, N. C., Hoskins, J. A., Hoskins,

- W. D. 1989. Edge effects in the use of area polygons to study competition. *Ecology* 70:272-74
137. Kimura, D. K. 1990. Approaches to age-structured separable sequential population analysis. *Can. J. Fish. Aquat. Sci.* 47:2364-74
  138. King, A. W., O'Neill, R. V., DeAngelis, D. L. 1989. Using ecosystem models to predict regional carbon dioxide exchange between the atmosphere and the terrestrial biosphere. *Global Biogeochem. Cycles* 3:337-62
  139. Kirk, K. L., Gilbert, J. J. 1990. Suspended clay and the population dynamics of planktonic rotifers and cladocerans. *Ecology* 71:1741-55
  140. Kitchell, J. A., Estabrook, G., MacLeod, N. 1987. Testing for equality of rates of evolution. *Paleobiology* 13:272-85
  141. Kjellsson, G. 1991. Seed fate in an ant-dispersed sedge, *Carex pilulifera* L.: recruitment and seedling survival in tests of models for spatial dispersion. *Oecologia* 88:435-43
  142. Knapp, S. J., Bridges, W. C. Jr., Yang, M.-H. 1989. Nonparametric confidence interval estimators for heritability and expected selection response. *Genetics* 121:891-98
  143. Knox, R. G., Peet, R. K., Christensen, N. L. 1989. Population dynamics in loblolly pine stands: changes in skewness and size inequality. *Ecology* 70:1153-66
  144. Knox, R. G., Peet, R. K. 1989. Bootstrapped ordination: a method for estimating sampling effects in indirect gradient analysis. *Vegetatio* 80:153-66
  145. Krantz, T. S., Theriot, E. C., Zimmer, E. A., Chapman, R. L. 1990. The Pleurastrophyceae and Micromonadophyceae: a cladistic analysis of nuclear ribosomal RNA sequence data. *J. Phycol.* 26:711-21
  146. Lanyon, S. M. 1985. Detecting internal inconsistencies in distance data. *Syst. Zool.* 34:397-403
  147. Lavin, M., Doyle, J. J. 1991. Tribal relationships of *Sphinctospermum* (Leguminosae): integration of traditional and chloroplast DNA data. *Syst. Bot.* 16:162-72
  148. Legendre, P., Dallot, S., Legendre, L. 1985. Succession of species within a community: chronological clustering with applications to marine and freshwater zooplankton. *Am. Nat.* 125:257-88
  149. Legendre, P., Fortin, M. J. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80:107-38
  150. Lesht, B. M., Fontaine, T. D. III, Dolan, D. M. 1991. Great Lakes USA-Canada total phosphorus model post audit and regionalized sensitivity analysis. *J. Great Lakes Res.* 17:3-17
  151. Lindberg, W. J., McElroy, D. A. 1991. Limitations in nearest neighbor tests of dispersion and association. *Fla. Sci.* 54:106-11
  152. Linton, L. R., Edgington, E. S., Davies, R. W. 1989. A view of niche overlap amenable to statistical analysis. *Can. J. Zool.* 67:55-60
  153. Loader, C., Damman, H. 1991. Nitrogen content of food plants and vulnerability of *Pieris rapae* to natural enemies. *Ecology* 72:1586-90
  154. Lobel, P. S., Anderson, D. M., Durand-Clement, M. 1988. Assessment of Ciguatera dinoflagellate populations: sample variability and algal substrate preference. *Biol. Bull.* 175:94-101
  155. Loganathan, G. V., Shrestha, S. P., Dillaha, T. A., Ross, B. B. 1989. Variable source area concept for identifying critical runoff-generating areas in a watershed. *Virginia Polytech. Inst. State Univ. Water Resour. Res. Center Bull.* I-XII:1-119
  156. Loreau, M. 1989. On testing temporal niche differentiation in carabid beetles. *Oecologia* 81:89-96
  157. Lumsden, C. J. 1988. Gene-culture coevolution: a test of the steady-state hypothesis for gene-culture translation. *J. Theor. Biol.* 130:391-406
  158. Madenjian, C. P., Jude, D. J., Tesar, F. J. 1986. Intervention analysis of power plant impact on fish populations. *Can. J. Fish. Aquat. Sci.* 43:819-29
  159. Maelzer, D. A. 1970. The regression of  $\log N_{n+1}$  on  $\log N_n$  as a test of density-dependence: an exercise with computer-constructed density-independent populations. *Ecology* 51:810-22
  160. Mallet, J. 1986. Dispersal and gene flow in a butterfly with home range behavior: *Heliconius erato* (Lepidoptera: Nymphalidae). *Oecologia* 68:210-17
  161. Mangel, M., Clark, C. W. 1988. *Dynamic Modeling in Behavioral Ecology*. Princeton, NJ: Princeton Univ. Press
  162. Manly, B. F. J. 1990. On the statistical analysis of niche overlap data. *Can. J. Zool.* 68:1420-22
  163. Manly, B. F. J. 1991. *Randomization and Monte Carlo Methods in Biology*. London: Chapman & Hall
  164. Manly, B. F. J. 1992. Bootstrapping to determine sample sizes in biological studies. *J. Exp. Mar. Biol. Ecol.* In press
  165. Manly, B. F. J. 1993. A review of

- computer intensive multivariate methods in ecology. In *Multivariate Environmental Statistics*, ed. G. F. Patil, C. R. Rao, N. P. Ross. Amsterdam: North Holland/Elsevier. In press
166. Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27: 209–20
  167. Marriott, F. H. C. 1979. Barnard's Monte Carlo tests: how many simulations? *Appl. Stat.* 28:75–77
  168. Martin, T. H., Wright, R. A., Crowder, L. B. 1989. Non-additive impact of blue crabs and spot on their prey assemblages. *Ecology* 70:1935–42
  169. Matson, P. A., Carpenter, S. R. 1990. Statistical analysis of ecological response to large-scale perturbations. *Ecology* 71:2037
  170. McArdle, B. H. 1988. The structural relationship: regression in biology. *Can. J. Zool.* 66:2329–39
  171. McKillup, S. C., Allen, P. G., Skewes, M. A. 1988. The natural decline of an introduced species following its initial increase in abundance: an explanation for *Ommatolius moreletii* in Australia. *Oecologia* 77:339–42
  172. McLain, D. K. 1986. Null models and the intensity of sexual selection. *Evol. Theory* 8:49–52
  173. Menkens, G. E. Jr., Anderson, S. H. 1988. Estimation of small-mammal population size. *Ecology* 69:1952–59
  174. Metz, J. A. J., Diekmann, O., eds. 1986. *The Dynamics of Physiologically Structured Populations*. Springer-Verlag Lecture Notes in Biomathematics 68: 475–94
  175. Meyer, J. S., Ingersoll, C. G., McDonald, L. L., Boyce, M. S. 1986. Estimating uncertainty in population growth rates: jackknife vs. bootstrap techniques. *Ecology* 67:1156–66
  176. Mielke, P. W. 1984. Meteorological applications of permutation techniques based on distance functions. In *Handbook of Statistics*, Vol. 4, ed. P. R. Krishnaiah, P. K. Sen, pp. 813–30. Amsterdam: North-Holland
  177. Mielke, P. W., Berry, K. J., Brier, G. W. 1981. Application of multiresponse permutation procedures for examining seasonal changes in monthly mean sea-level pressure patterns. *Monthly Weather Rev.* 109:120–26
  178. Mielke, P. W., Berry, K. J., Brockwell, P. J., Williams, J. S. 1981. A class of nonparametric tests based on multi-response permutation procedures. *Biometrika* 68:720–24
  179. Miller, R. 1968. Jackknifing variances. *Ann. Math. Stat.* 39:567–82
  180. Miller, R. G. 1974. The jackknife—a review. *Biometrika* 61:1–17
  181. Minta, S., Mangel, M. 1989. A simple population estimate based on simulation for capture-recapture and capture-resight data. *Ecology* 70:1738–51
  182. Mitchell-Olds, T., Shaw, R. G. 1987. Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution* 41:1149–61
  183. Mueller, L. D., Altenberg, L. 1985. Statistical inference on measures of niche overlap. *Ecology* 66:1204–10
  184. Nemec, A. F. L., Brinkhurst, R. O. 1988. Using the bootstrap to assess statistical significance in the cluster analysis of species abundance data. *Can. J. Fish. Aquat. Sci.* 45:965–70
  185. Newell, C. J., Hopkins, L. P., Bedient, P. B. 1990. A hydrogeologic database for ground-water modeling. *Ground Water* 28:703–14
  186. Noreen, E. W. 1989. *Computer Intensive Methods for Testing Hypotheses: An Introduction*. New York: Wiley
  187. Ocana, J., Ruiz de Villa, C., Ribo, G. 1991. Bias and efficiency in estimating sexual selection: Levene's Z index and some resampling alternatives. *Heredity* 67:95–102
  188. Odendaal, F. J., Turchin, P., Stermitz, F. R. 1989. Influence of host-plant density and male harassment on the distribution of female *Euphydryas anicia* (Nymphalidae). *Oecologia* 78: 283–88
  189. Ohta, T. 1988. Evolution by gene duplication and compensatory advantageous mutations. *Genetics* 120:841–48
  190. Ohta, T. 1988b. Further simulation studies on evolution by gene duplication. *Evolution* 42:375–86
  191. Olson, C. L. 1976. On choosing a test statistic in multivariate analysis of variance. *Psychol. Bull.* 83:579–86
  192. Page, R. D. M. 1988. Quantitative cladistic biogeography: constructing and comparing area cladograms. *Syst. Zool.* 37:254–70
  193. Page, R. D. M. 1991. Clocks, clades, and cospeciation: comparing rates of evolution and timing of cospeciation events in host-parasite assemblages. *Syst. Zool.* 40:188–98
  194. Palmer, M. W. 1990. The estimation of species richness by extrapolation. *Ecology* 71:1195–98
  195. Palmer, M. W. 1991. Estimating species richness: the second-order jackknife reconsidered. *Ecology* 72:1512–13
  196. Pamilo, P. 1990. Statistical tests of



- phenograms based on genetic distances. *Evolution* 44:689-97
197. Papadopoulos, A. S., Tiwari, R. C., Muha, M. J. 1991. Bootstrap procedures for time series analysis of BOD data. *Ecol. Model.* 55:57-66
198. Patterson, G. B. 1986. A statistical method for testing for dietary differences. *N. Zeal. J. Zool.* 13:113-15
199. Peltonen, A., Hanski, I. 1991. Patterns of island occupancy explained by colonization and extinction rates in shrews. *Ecology* 72:1698-708
200. Phillips, R. A., Neal, L., Johnson, D. M., Nagel, J. W., Laughlin, T. F., Lane, R. T. 1987. Simulating local extinction of trout populations. In *Tools for the Simulation Profession*, ed. R. Hawkins, K. Klukis, pp. 24-27. San Diego: Soc. Computer Simul.
201. Pimm, S. L. 1983. Appendix: Monte-Carlo analyses in ecology. In *Lizard Ecology: Studies on a Model Organism*, ed. R. B. Huey, E. R. Pianka, T. W. Schoener, pp. 290-96. Cambridge, Mass: Harvard Univ. Press
202. Pleasants, J. M. 1990. Null-model tests for competitive displacement: the fallacy of not focusing on the whole community. *Ecology* 71:1078-84
203. Poethke, H. J., Kirchberg, M. 1987. On the stabilizing effect of density-dependent mortality factors. *Oecologia* 74:156-58
204. Pollard, E., Lakhani, K. H., Rothery, P. 1987. The detection of density-dependence from a series of annual censuses. *Ecology* 68:2046-55
205. Pontasch, K. W., Smith, E. P., Cairns, J. Jr. 1989. Diversity indices, community comparison indices, and canonical discriminant analysis interpreting the results of multispecies toxicity tests. *Water Res.* 23:1229-38
206. Powell, G. L., Russell, A. P. 1985. Growth and sexual size dimorphism in Alberta populations of the eastern short-horned lizard, *Phrynosoma douglassi brevirostre*. *Can. J. Zool.* 63:139-54
207. Quinn, J. F. 1987. On the statistical detection of cycles in extinctions in the marine fossil record. *Paleobiology* 13:465-78
208. Raup, D. M., Gould, S. J., Schopf, T. J. M., Simberloff, D. S. 1973. Stochastic models of phylogeny and the evolution of diversity. *J. Geol.* 81:525-42
209. Raup, D. M., Sepkoski, J. J. 1988. Testing for periodicity of extinction. *Science* 241:94-96
210. Reddingius, J., den Boer, P. J. 1989. On the stabilization of animal numbers. Problems of testing. 1. Power estimates and estimation errors. *Oecologia* 78:1-8
211. Ribí, J., Tomczak, M. 1990. An impact assessment for the French nuclear weapon test sites in French Polynesia. *Mar. Pollut. Bull.* 21:536-42
212. Riessen, H. P., Sprules, W. G. 1990. Demographic costs of antipredator defenses in *Daphnia pulex*. *Ecology* 71:1536-46
213. Ritchie, M. E. 1990. Optimal foraging and fitness in Columbian ground squirrels. *Oecologia* 82:56-67
214. Roberts, A., Stone, L. 1990. Island-sharing by archipelago species. *Oecologia* 83:560-67
215. Robertson, G. P., Huston, M. A., Evans, F. C., Tiedje, J. M. 1988. Spatial variability in a successional plant community: patterns of nitrogen availability. *Ecology* 69:1517-24
216. Rogers, D. S. 1990. Genic evolution, historical biogeography, and systematic relationships among spiny pocket mice, subfamily Heteromyinae. *J. Mammal.* 71:668-85
217. Rogers, A. R., Eriksson, A. W. 1988. Statistical analysis of the migration component of genetic drift. *Am. J. Phys. Anthropol.* 77:451-58
218. Romano, J. P. 1989. Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Stat.* 17:141-59
219. Romano, J. P. 1990. On the behavior of randomization tests without a group invariance assumption. *J. Am. Stat. Assoc.* 85:686-92
220. Root, T. 1989. Energy constraints on avian distributions: a reply to Castro. *Ecology* 70:1183-85
221. Rose, K. A., McLean, R. I., Summers, J. K. 1989. Development and Monte Carlo analysis of an oyster bioaccumulation model applied to biomonitoring data. *Ecol. Model.* 45:111-32
222. Rotenberry, J. T. 1985. The role of habitat in avian community composition: physiognomy or floristics. *Oecologia* 67:213-17
223. Sale, P. F. 1974. Mechanisms of co-existence in a guild of territorial fishes at Heron Island. In *Proc. 2nd Int. Coral Reef Symp.* 1:193-206. Brisbane: Great Barrier Reef Comm.
224. Sanderson, M. J. 1989. Confidence limits on phylogenies: the bootstrap revisited. *Cladistics* 5:113-29
225. Schill, W. B., Dorazio, R. M. 1990. Immunological discrimination of Atlantic striped bass stocks. *Trans. Am. Fish. Soc.* 119:77-85
226. Schluter, D., Repasky, R. R. 1991.

- Worldwide limitation of finch densities by food and other factors. *Ecology* 72:1763-74
227. Schmitt, J., Eccleston, J., Ehrhardt, D. W. 1987. Density-dependent flowering phenology, outcrossing, and reproduction in *Impatiens capensis*. *Oecologia* 72:341-47
  228. Setzer, R. W. 1985. Spatio-temporal patterns of mortality in *Pemphigus populicaulis* and *P. populitransversus* on cottonwoods. *Oecologia* 67:310-21
  229. Simberloff, D. 1978. Using island biogeographic distributions to determine if colonization is stochastic. *Am. Nat.* 112:713-26
  230. Simberloff, D., Boecklen, W. 1981. Santa Rosalia reconsidered: size ratios and competition. *Evolution* 35:1206-28
  231. Simonoff, J. S. 1986. Jackknifing and bootstrapping goodness of fit statistics in sparse multinomials. *J. Am. Stat. Assoc.* 81:1005-11
  232. Slatkin, M., Arter, H. E. 1991. Spatial autocorrelation methods in population genetics. *Am. Nat.* 138:499-517
  233. Slowinski, J. B., Guyer, C. 1989. Testing the stochasticity of patterns of organismal diversity: an improved null model. *Am. Nat.* 134:907-21
  234. Smettem, K. R. J., Collis-George, N. 1985. Prediction of steady-state ponded infiltration distributions in a soil with vertical macropores. *J. Hydrol.* 79: 115-22
  235. Sokal, R. R., Rohlf, F. J. 1981. *Bio-metry: The Principles and Practice of Statistics in Biological Research*. New York: Freeman. 2nd ed.
  236. Solow, A. R. 1989. Bootstrapping sparsely sampled spatial point patterns. *Ecology* 70:379-82
  237. Solow, A. R. 1989. A randomization test for independence of animal locations. *Ecology* 70:1546-49
  238. Solow, A. R. 1990. A randomization test for misclassification probability in discriminant analysis. *Ecology* 71: 2379-82
  239. St. Amant, J. L. S. 1970. The detection of regulation in animal numbers. *Ecology* 51:823-28
  240. Starks, T. H., Flatman, G. T. 1991. RCRA ground-water monitoring decision procedures viewed as quality control schemes. *Environ. Monit. Assess.* 16:19-38
  241. Stauffer, D. F., Garton, E. O., Steinhorst, R. K. 1985. A comparison of principal components from real and random data. *Ecology* 66:1693-98
  242. Stewart-Oaten, A., Bence, J. R., Osenberg, C. W. 1992. Assessing effects of unreplicated perturbations: no simple solutions. *Ecology*. In press
  243. Still, A. W., White, A. P. 1981. The approximate randomization test as an alternative to the F test in analysis of variance. *Br. J. Math. Stat. Psychol.* 34:243-52
  244. Stone, L., Roberts, L. 1990. The checkerboard score and species distributions. *Oecologia* 85:74-79
  245. Stone, L., Roberts, L. 1991. Conditions for a species to gain advantage from the presence of competitors. *Ecology* 72:1964-72
  246. Strobeck, C. 1987. Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics* 117: 149-54
  247. Strong, D. R., Szyska, L. A., Simberloff, D. 1979. Tests of community-wide character displacement against null hypotheses. *Evolution* 33:897-913
  248. Swihart, R. K. and Slade, N. A. 1985. Testing for independence in animal movements. *Ecology* 66:1176-84
  249. Temeles, E. J. 1987. The relative importance of prey availability and intruder pressure in feeding territory size regulation by harriers, *Circus cyaneus*. *Oecologia* 74:286-97
  250. ter Braak, C. J. F. 1991. Permutation versus bootstrap significance tests in multiple regression and ANOVA. In *Bootstrapping and Related Techniques*, ed. K.-H. Jöckel. Berlin: Springer-Verlag. In press
  251. Thomas, S. C., Weiner, J. 1989. Including competitive asymmetry in measures of local interference in plant populations. *Oecologia* 80:349-55
  252. Tukey, J. W. 1958. Bias and confidence in not quite large samples (Abstract). *Ann. Math. Stat.* 29:614
  253. Turchin, P. 1987. The role of aggregation in the response of Mexican bean beetles to host-plant density. *Oecologia* 71:577-82
  254. van Straalen, N. M. 1985. Comparative demography of forest floor Collembola populations. *Oikos* 45:253-65
  255. Vassiliou, A., Ignatiades, L., Karydis, M. 1989. Clustering of transect phytoplankton collections with a quick randomization algorithm. *J. Exp. Mar. Biol. Ecol.* 130:135-46
  256. Venrick, E. L. 1990. Phytoplankton in an oligotrophic ocean: species structure and interannual variability. *Ecology* 71: 1547-63
  257. Verrier, E., Colleau, J. J., Foulley, J. L. 1990. Predicting cumulated response to directional selection in finite pan-

258. mictic populations. *Theor. Appl. Genet.* 79:833-40
259. Vickery, W. L. 1991. An evaluation of bias in *k*-factor analysis. *Oecologia* 85:413-18
260. Vickery, W. L., Nudds, T. D. 1991. Testing for density-dependent effects in sequential censuses. *Oecologia* 85: 419-23
261. Walls, M., Caswell, H., Ketola, M. 1991. Demographic costs of *Chao-borus*-induced defenses in *Daphnia pulex*: a sensitivity analysis. *Oecologia* 87:43-50
262. Whicker, F. W., Kirchner, T. B. 1987. PATHWAY: a dynamic food-chain model to predict radionuclide ingestion after fallout deposition. *Health Phys.* 52:717-38
263. Wilson, J. B. 1987. Methods for detecting non-randomness in species co-occurrences: a contribution. *Oecologia* 73:579-82
264. Wilson, J. B. 1989. A null model of guild proportionality, applied to stratification of a New Zealand temperate rain forest. *Oecologia* 80:263-67
265. Wilson, J. B., Hubbard, J. C. E., Rapson, G. L. 1988. A comparison of realized niche relations of species in New Zealand and Britain. *Oecologia* 76:106-10
266. Winemiller, K. O., Pianka, E. R. 1990. Organization in natural assemblages of desert lizards and tropical fishes. *Ecol. Monogr.* 60:27-55
267. Wolda, H. 1989. Comment on the article "On testing temporal niche differentiation in carabid beetles" by M. Loreau and the "Comment on the article of M. Loreau" by P. J. den Boer. *Oecologia* 81:99
268. Wolda, H. 1991. The usefulness of the equilibrium concept in population dynamics: A reply to Berryman. *Oecologia* 86:144-45
269. Yearsley, J. R., Lettenmaier, D. P. 1987. Model complexity and data worth an assessment of changes in the global carbon budget. *Ecol. Modell.* 39:201-26
270. Zimmerman, G. M., Goetz, H., Mielke, P. W. Jr. 1985. Use of an improved statistical method for group comparisons to study effects of prairie fire. *Ecology* 66:606-11
271. Zink, R. M., Dittmann, D. L., Rootes, W. L. 1991. Mitochondrial DNA variation and the phylogeny of *Zonotrichia*. *Auk* 108:578-84
272. Zwolfer, H., Brandl, R. 1989. Niches and size relationships in Coleoptera associated with Cardueae host plants: adaptations to resource gradients. *Oecologia* 78:60-68