ESTIMATION OF TEMPORAL VARIABILITY IN POPULATIONS

ALLAN STEWART-OATEN,[1,*] WILLIAM W. MURDOCH,[1,†] AND SANDRA J. WALDE[2]

[1]Department of Biological Sciences, University of California, Santa Barbara, California 93106;
[2]Department of Biology, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada

Abstract.—A common measure of the temporal variability of a population is the standard deviation of the logarithms of successive estimated population sizes, $\ln(D_t)$. This measure overestimates true temporal variability (the standard deviation of the logarithms of true population density, $\ln[\Delta_t]$) because it is contaminated by spatial variance (variability among samples taken on the same date). The random error in D_t causes an overestimation of temporal variance, both directly and also indirectly, by causing $\ln(D_t)$ to underestimate $\ln(\Delta_t)$. Both problems are more severe if spatial variance is large or the sample size, on a date, is small. We develop an alternative estimator, which uses an estimate of spatial variance to correct for both problems. To evaluate it, we sampled from simulated populations with a wide range of clumping. The results show that the standard estimate can be badly biased. The new estimator is much better and is quite accurate over a broad range of conditions. Our results suggest a reanalysis of some ecological studies that have estimated temporal variability to attack theoretically important questions. In particular, the apparently greater average temporal variability of terrestrial arthropods compared with terrestrial vertebrates could be an artifact caused by the fact that, typically, clumping is weaker and density estimates are more accurate in vertebrates.

Some populations appear to fluctuate in abundance a great deal, others very little. Much effort has been directed at accounting for such differences in temporal variability in population abundance. Taxonomic patterns have been postulated: Connell and Sousa (1983) found no evidence of differences in the temporal variability of terrestrial insect and vertebrate populations, while Schoener (1985) found lizards to have very low variability, and Ostfeld (1988) found a similar pattern for small rodents. Population variability has been found to be positively correlated with fecundity and high population growth rate (Spitzer et al. 1984; Spitzer and Leps 1988), predation (Fairweather 1988), and geographical range (Gaston and Lawton 1988*a*) and negatively associated with longevity (Connell and Sousa 1983; Williamson 1984), body size (Gaston 1988; Gaston and Lawton 1988*b*), and polyphagy (Watt 1965; Redfearn and Pimm 1988). The temporal variability of spider populations has been linked to the probability of extinction (Schoener and Spiller 1992) and that of carabid beetle populations to habitat stability (Hanski 1982). Temporal variability patterns are of interest to ecologists both practically (e.g., in relation to pest control and exploited or endangered populations) and theoretically (e.g., in relation to stability).

* E-mail: stewart@lifesci.ucsb.edu.
† E-mail: murdoch@lifesci.ucsb.edu.

Crucial to all of these problems is a measure of the temporal variability of the true population size. For virtually any real population, of course, we cannot know the true density, which we denote by Δ_t, but can only estimate it, typically by the average number found in a set of replicate samples taken at time t, which we denote by D_t. In studies such as those discussed above, estimates of temporal variability (most commonly the standard deviation of the log of successive population densities) have been obtained by replacing Δ_t with D_t.

The purposes of this article are to show that these measures of temporal variance are not appropriate because they are inflated by the effects of spatial variance (the variation of each D_t about its "target" Δ_t), to propose an alternative measure, and to demonstrate that much of the apparent difference in temporal variability among groups of organisms could be induced by errors of estimation.

We focus on the most popular measure, the standard deviation of the logs of estimated densities. This approach measures relative variation: it can be rewritten as a function of the proportions, $\Delta_t / \Sigma \Delta_t$. This may be of interest in its own right, since it has more ecological significance than other possibilities, such as variation among the densities themselves. It may also give comparisons among species that are independent of mean density although, as McArdle et al. (1990) point out, this requires variance $= \alpha(\text{mean})^2$, a version of Taylor's empirical "power law" (Taylor 1961).

CONFOUNDING OF TEMPORAL AND SPATIAL VARIANCE

Simple Examples

Current estimates of temporal variability conflate spatial variance with temporal variance and hence systematically overestimate temporal variability. Temporal variation in the *estimated* population mean consists of two components: temporal variation in the *true* population mean, which is the component we seek, and variation in the difference between this true mean and its estimate. The second source of variance arises because spatial variability leads to an error in the estimation of the true population density on each date (fig. 1). This error will increase as spatial variation increases or sample size decreases.

The point can be made by a simple and extreme example. Suppose that the mean number of insects per twig in a forest is precisely 11, but half of the twigs support two insects and half support 20. Suppose further that the density does not change over the entire study period and that, on each date, we sample a single randomly chosen twig. The sample will *never* correctly estimate the true population size, and abundance will appear to fluctuate from two to 20, yet this apparent temporal variability will be caused entirely by *spatial* variation in abundance.

The problem is also familiar from the one-way random effects ANOVA model (Snedecor and Cochran 1989, chap. 13): $X_{ti} = \mu + A_t + \epsilon_{ti}$. For example, X_{ti} might be the number of insects on the ith random twig from the tth random tree, μ the average density of insects over all trees, A_t the deviation of tree t from this average, and ϵ_{ti} the deviation of the ith twig from the tree t average. Suppose that we wish to estimate the variance among trees on a single date (i.e., the
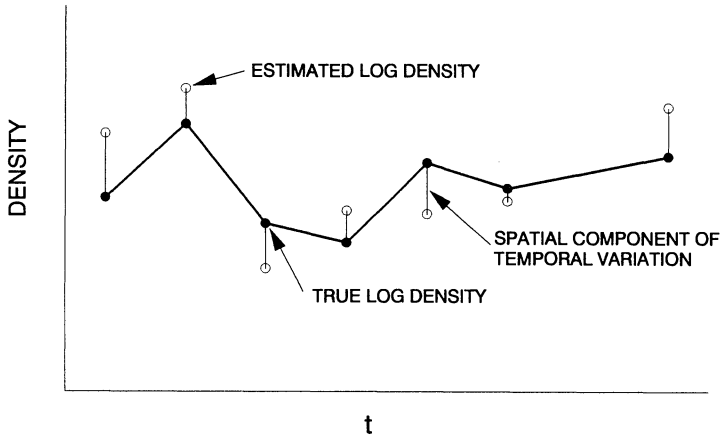
Fig. 1.—Illustration of how spatial variance inflates estimates temporal variability. True temporal variability is determined by the sequence of true population densities (indicated by solid circles). Spatial variance causes the estimates of population density (*open circles*) to be different from the true density. The errors (*vertical lines*) contribute to estimated temporal variability based on the estimated densities.

variance of A_t). We could use the sample variance of the tree averages, $X_{1\cdot}$, $X_{2\cdot}$, . . . (the dot in "$X_{t\cdot}$" signifies we are averaging over the twigs taken from tree t). But these are not the true mean densities, only an estimate of them: their variance, is partly attributable to variation between twigs within a tree. We therefore "correct" for this by subtracting an allowance for it; that is, we use (MS[between trees] − MS[within trees])/r:

$$s^2(A) = \Sigma(X_{t\cdot} - X_{\cdot\cdot})^2/(T - 1) - \Sigma_t s^2(X_t)/Tr, \qquad (1)$$

where $s^2(X_t) = \Sigma_i(X_{ti} - X_{t\cdot})^2/(r - 1)$, if there were T trees and r twigs per tree.

To return to temporal variability, suppose that t now stands for sampling date, we take twig samples from only one tree, and we are interested in variation across times rather than across trees. Current measures of temporal variability do not follow the ANOVA recipe. Instead they use the uncorrected variance among dates (i.e., the uncorrected variance of D_t or its transform). The error is illustrated in figure 1. The standard approach is based on the temporal variance of the successive density estimates (indicated by an open circle on the figure), but this variance is the sum of the temporal variance of the successive true densities (indicated by the solid circles in fig. 1) and the variance of the "errors" of estimation (indicated by the vertical lines connecting the estimates and the true densities). It is worth noting that other sources of variation in the estimate of mean density have the same effect (e.g., if densities on the sampled twigs are estimated rather than counted). We next develop a correction to subtract this spatial variance.

Adjusting for Spatial Variance

In this section, we address estimation of the variance of log(density), that is, $\ln(\Delta_t)$, using new random samples of twigs on each date (e.g., the sampling might

be destructive). The approach also works for any other transformation of density (provided approximations similar to those given below are accurate enough) or for no transformation at all, as in the ANOVA example. In some cases, it can also work under a different sampling regime (e.g., if the sampling is stratified or if an initial set of random twigs is chosen at the outset and revisited throughout the study). In these cases, such schemes can give better (smaller variance) estimates of temporal variability and can be informative about temporal variance on different spatial scales. However, they can give badly biased estimates if there is time × space interaction (e.g., if stratum 1 is more attractive than stratum 2 under some conditions but less attractive under others).

The simple examples illustrate the main problem of estimating the variance of $\ln(\Delta_t)$, but there are some additional complications. To give the details, we need to define some symbols. We use the example of sampling insects on a given tree, using the counts on randomly sampled twigs on each of T dates. We define

r_t = the number twigs sampled on date t;
D_{ti} = the number of insects on the ith sampled twig on date t;
$D_t.$ = the average of the D_{ti}'s over i (i.e., over twigs, within a date);   (2)
Δ_t = the true density per twig on date t;
Λ_t = $\ln(\Delta_t)$, the logarithm (base e) of the true density per twig on date t;

and

$\Lambda.$ = $\Sigma_t \Lambda_t / T$, the average of the Λ_t's (over sampling times, or dates).

Suppose we choose to measure temporal variability of the true population size by the variance of the logarithms of successive Δ_t's:

$$s_\Lambda^2 = \Sigma_t (\Lambda_t - \Lambda.)^2 / (T - 1).$$ (3)

(Strictly, this is the "sample" variance, because the divisor, $T - 1$, allows one degree of freedom for the estimation of the "true mean" of the Λ_t's by their average, $\Lambda.$. We will refer to this as the "variance" throughout.)

Since the Λ_t's cannot be observed, we must estimate them, using the D_{ti}'s. We discuss details of this estimation later. For the moment we define

$$L_t = \text{estimate of } \Lambda_t, \text{ based on } \{D_{ti}, i = 1, 2, \ldots, r_t\}.$$ (4)

So D is observed density, and L is estimated log density, based on D.

If, as in current measures of temporal variability, we simply replace the unobserved Λ_t's by their estimates, the L_t's, our estimate of true temporal variance, s_Λ^2, is

$$s_L^2 = \Sigma_t (L_t - L.)^2 / (T - 1),$$ (5)

where $L.$ is the average of $L_1, L_2, \ldots, L_t$.

But this estimate is contaminated by "within-tree" (i.e., spatial) variation. Thus, if the mean of L_t is

$$E\{L_t\} = \Lambda_t \tag{6}$$

and its (spatial) variance is

$$V\{L_t\} = \sigma_{Lt}^2, \tag{7}$$

then the mean of s_L^2, the estimated temporal variance, can be shown to be

$$E\{s_L^2\} = s_\Lambda^2 + \Sigma_t \sigma_{Lt}^2/T, \tag{8}$$

where the second term on the right-hand side is average spatial variance. We can therefore subtract an estimate of this contamination, as in the ANOVA analogy above.

Thus, suppose we can estimate the (spatial) variance of the estimate of $\ln(\Delta_t)$ by

$$s_{Lt}^2 = \text{estimate of } V\{L_t\} \text{ based on } \{D_{ti}, i = 1, 2, \ldots, r_t\}, \tag{9}$$

where s_{Lt}^2 is unbiased; that is,

$$E\{s_{Lt}^2\} = \sigma_{Lt}^2. \tag{10}$$

(Note the difference between s_L^2, the variance of the sample $L_1, \ldots, L_T$ [a variance between dates] and s_{Lt}^2, the estimated variance of L_t, due to spatial variation [i.e., between twigs within date t, of $D_{t1}, D_{t2}, \ldots$].)

Then we can adjust the estimate of s_Λ^2 to obtain

$$s_L^2(\text{adj}) = s_L^2 - \Sigma s_{Lt}^2/T, \tag{11}$$

where the second term is an estimate of the average spatial variance. Equations (8), (10), and (11) show that

$$E\{s_L^2(\text{adj})\} = s_\Lambda^2. \tag{12}$$

To summarize, we estimate the log of the population density, $\ln(\Delta_t) = \Lambda_t$, on each date, by L_t; we also estimate the variance of this estimate, $V\{L_t\}$, by s_{Lt}^2. We then estimate the across-dates variance of the Λ_t's, s_Λ^2, by the across-dates variance of the L_t's, s_L^2, minus the average of the estimated within-date (i.e., spatial) variances of the L_t's. The formula in equation (11) is a generalization of that in equation (1).

THE ESTIMATES OF LN(DENSITY) AND THEIR VARIANCES

We have not specified how the counts on twigs, $D_{t1}, D_{t2}, \ldots$, on a given date are to be converted into a value of L_t, the estimate of $\Lambda_t = \ln(\Delta_t)$ for that date. The obvious method is to average them, obtaining $D_{t\cdot}$, and let

$$L_t = \ln(D_{t\cdot}). \tag{13}$$

There are two problems with this method. One is that $D_{t.} = 0$ if all D_{ti}'s are zero (i.e., no insects at all are found on date t), and L_t is undefined. The other is that, even if this could be prevented (e.g., if somehow at least one twig was guaranteed to have at least one insect), and $D_{t.}$ was an unbiased estimator of Δ_t (i.e., $E\{D_{t.}\} = \Delta_t$), $L_t = \ln(D_{t.})$ would still be a biased estimator of Λ_t. This bias is caused by Jensen's inequality (Feller 1966, p. 152): $\ln(x)$ is a concave-down nonlinear function of x, and hence the mean of the log of a variable will be less than the log of the mean—for example, $E\{\ln(D_{t.})\} < \ln(E\{D_{t.}\}) = \ln(\Delta_t) = \Lambda_t$. This is most easily pictured by considering only two points, x_1 and x_2; the average of $\ln(x_1)$ and $\ln(x_2)$ is on the chord joining $\ln(x_1)$ and $\ln(x_2)$ on the graph of $\ln(x)$; this chord lies completely below the graph.

(If all twigs had at least one insect, an alternative estimator is $\Sigma_i \ln[D_{ti}]/r_t$, the average of the values of $\ln[\text{density}]$ for each twig. However, Jensen's inequality can be used to show that $E\{\Sigma_i \ln(D_{ti})/r_t\} < E\{\ln(D_{t.})\}$, so this estimator has even greater bias. We do not consider it further.)

Jensen's Inequality Adjustment

We next propose a correction for the underestimation of the natural logarithm of mean density, Λ, introduced by Jensen's inequality. First, we estimate the bias, as follows. Provided twigs are chosen randomly, $E\{D_{ti}\} = \Delta_t$. The deviation of twig i from the mean is

$$e_{ti} = D_{ti} - \Delta_t. \tag{14}$$

The observed average density on date t is

$$D_{t.} = \Delta_t + e_{t.}, \tag{15}$$

where $E\{e_{t.}\} = 0$. For the spatial (among-twigs) variance on date t, we write

$$\sigma_{Dt}^2 = V\{D_{ti}\} = V\{e_{ti}\}. \tag{16}$$

Assuming the twigs are independently sampled, the variance of $D_{t.}$ due to spatial variation is

$$V\{e_{t.}\} = \sigma_{D.}^2 = \sigma_{Dt}^2/r_t. \tag{17}$$

Thus, from equation (13), $L_t = \ln(\Delta_t + e_{t.})$. Taylor series expansion gives, approximately,

$$L_t = \ln(\Delta_t) + e_{t.}/\Delta_t - e_{t.}^2/2\Delta_t^2. \tag{18}$$

Since $E\{e_{t.}\} = 0$, $E\{e_{t.}^2\} = \sigma_{D.}^2$, and $E\{s_{D.}^2\} = \sigma_{D.}^2$, where

$$s_{D.}^2 = \Sigma(D_{ti} - D_{t.})^2/r_t(r_t - 1), \tag{19}$$

the estimated spatial variance of $D_{t.}$, we obtain, approximately,

$$E\{L_t + s_{D.}^2/2\Delta_t^2\} \approx \ln(\Delta_t). \tag{20}$$

This suggests approximating the bias by $s_{D.}^2/D_{t.}^2$ and using the estimator

$$L_{tJ} = \ln(D_{t.}) + s_{D.}^2/2D_{t.}^2. \tag{21}$$

as an improved (less biased) estimate of $\ln(\Delta_t)$. We use the notation L_{tJ} to remind us that this estimate corrects for Jensen's inequality.

Note also that, if $e_{t.}/\Delta_t$ is small, equation (18) gives $L_t \approx \ln(\Delta_t) + e_{t.}/\Delta_t$; thus $V\{L_t\} \approx \sigma_{D.}^2/\Delta_t^2$. This suggests estimating σ_{Lt}^2 (eq. [7]) by

$$s_{Lt}^2 = s_{D.}^2/D_{t.}^2 \tag{22}$$

(see eq. [19]), which can be used for the adjustment in equation (11).

The Problem of Zeros

The previous discussion is somewhat artificial, since it assumes that $D_{t.}$ cannot be zero. To facilitate the presentation, we only summarize here possible responses to the situation in which there are dates on which no organisms are found in samples; details are in the appendix. Such "zero dates," however, pose an awkward problem, and current solutions (mainly adding one to observations), and most potential solutions, involve arbitrary decisions that can affect conclusions. We distinguish two cases.

First, in the case of existing data sets, if $D_{t.} = 0$ for any date, t, then we cannot estimate $\ln(\Delta_t)$ by $\ln(D_{t.})$. Unfortunately, there seem to be no straightforward replacements. We make six suggestions in the appendix and assume for the rest of the discussion that an appropriate step has been taken to provide an estimate of $\ln(D_{t.})$ on dates on which no organisms have been found.

Second, in the case of future data sets, if "zero twigs" (i.e., $D_{ti} = 0$) are possible, and twigs are chosen randomly and independently, then zero dates ($D_{t.} = 0$) are also possible when fixed sample sizes are used. Unlimited sequential sampling—for example, sampling until two nonzero twigs have been found (Haldane 1945)—can guarantee unbiased nonzero estimates of Δ_t (assuming $\Delta_t > 0$) but may involve unacceptably large samples.

A compromise has been proposed (Stewart-Oaten 1995), based on a scheme of Kim and Nachlas (1984) and analysis by Kremers (1987). This is a modified sequential scheme in which an initial sample of r twigs is taken. If this contains fewer than m nonzero twigs, sampling continues until m nonzero twigs are found or, failing that, until the logistically feasible maximum number, R, have been taken. If $m > 1$, these samples provide an unbiased estimate of mean density, which requires an arbitrary adjustment away from zero only if no organisms are found in the R samples (appendix). This estimate replaces the average, $D_{t.}$, in the equations above: the two are often equal, but not always. The sequential samples also allow us to estimate the variance of this estimate. This estimated variance replaces the "s_D^2" of equation (19) for use in equations (22) and (11): the two are usually different, but not greatly. In this article, we avoid the zero dates problem by using this sequential scheme in the simulations presented next.

A COMPARISON OF STANDARD AND CORRECTED ESTIMATES

We now investigate how badly the standard approach overestimates temporal variability, as a function of spatial variability, and the improvements gained by removing spatial variation and adjusting for Jensen's inequality.

We compare four ways to estimate s_Λ^2, the temporal variance of $\ln(\Delta_t)$, depending on whether we correct for spatial variance (eq. [5] vs. eq. [11]) and on whether our estimate of L_t corrects for Jensen's inequality (eq. [13] vs. eq. [21]).

We present our results in terms of logs to base 10, as is usual, rather than to base e as in our discussion so far. Both s_Λ^2 and its estimates can be converted to base 10 by multiplying by $(\log_{10}e)^2$. We denote the converted value of s_Λ^2 by $s_{\Lambda 10}^2$; all estimates (s_L^2, s_{LJ}^2 [adj], etc.) are also converted, but we omit the subscript 10 for these.

The comparisons used successive true mean densities of $\Delta_t = 1, 2, 4,$ and 5 insects per twig, for $t = 1, 2, 3,$ and $4 = T$ dates. We chose these numbers because, with $\Lambda_{10t} = \log_{10}(\Delta_t)$, we get $s_{\Lambda 10}^2 = \Sigma(\Lambda_{10t} - \Lambda_{10\cdot})^2/(T - 1) = 0.1$. This value is very close to the mean and modal values observed in terrestrial vertebrate populations (see below).

The number of insects on a randomly chosen twig was assumed to be negative binomial, with mean Δ_t and variance $\Delta_t(1 + c\Delta_t)$. Clumping was defined by the parameter c. This is not the usual "clumping parameter," which is $k = 1/c$, but we prefer c because it increases as clumping increases, as seems more natural (the Poisson distribution has $c = 0$), and it more clearly separates highly clumped distributions, which seem to arise more often in the field and change more rapidly (e.g., distributions with $k = 0.2$ and $k = 0.1$ differ by far more than those with $k = 5$ and $k = 10$). We considered clumping values of $c = 0, 1, 2, \ldots, 10$, which seem to cover all but a few levels observed in the field.

To avoid dates with zero organisms, we used the sequential sampling scheme summarized above. For each date (a density and a clumping parameter), 10,000 density estimates were calculated, each based on a set of sequential samples. The minimum number of samples on a given date (the minimum number in a set) was $r = 5$, we required at least $m = 2$ positive samples, and the maximum was set at $R = 100$ samples. Further details are given at the end of the appendix. The average number of samples ranged from 5 to 9.7 (i.e., 9.7 was the average sample size over 10,000 "dates" with density = 1 and clumping = 10); the largest sample taken was 54, so it was never necessary to use the arbitrary adjustment for a zero date. We also report briefly on results for $r = 15$.

In the first set of simulations, clumping remained constant as mean density changed from date to date. Thus, for $c = 3$, an estimate of $s_{\Lambda 10}^2$ was based on at least r samples from the negative binomial distribution with mean one and clumping parameter $c = 3$ (so the variance was four), a second set of samples of at least r from the negative binomial distribution with mean two and clumping parameter three (so the variance was 14), and third and fourth sets of r or more, negative binomial with means four and five, clumping parameter three in both cases, and variances 52 and 80, respectively.

The four estimates are graphed against the clumping parameter values in figure 2, together with a flat line representing the true value of $s_{\Lambda 10}^2$, namely 0.10. The top curve on the graph is the standard uncorrected estimate. The graph answers the two questions raised at the start of this section, as follows.

First, spatial variability in abundance strongly inflates the standard measure of
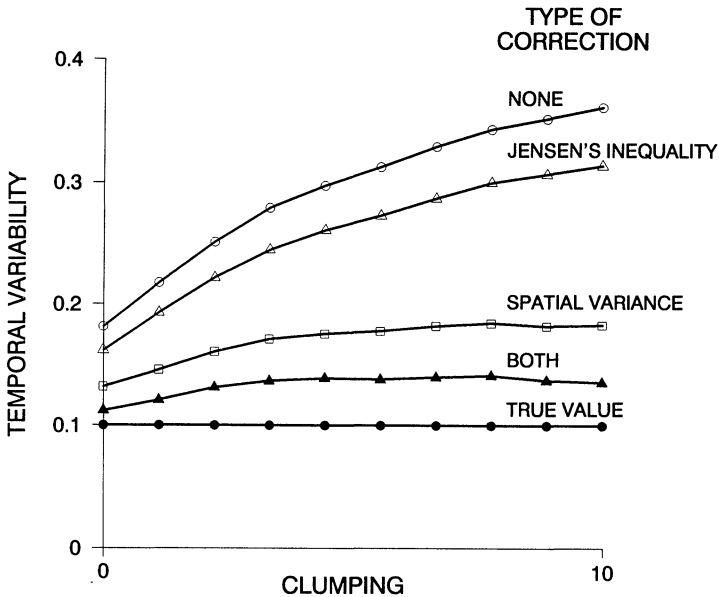
FIG. 2.—Results of simulations estimating temporal variability (variance of successive log densities) by four methods for a population that varied over a period of 4 yr at a density between one and five and was spatially distributed according to the negative binomial. The estimates were calculated for spatial clumping values ranging from zero (Poisson) to 10. The true variance, which is independent of spatial clumping, was 0.1. The top curve is the standard estimate uncorrected for either spatial variance or Jensen's inequality. The other curves show that the estimates come progressively closer to the mean as we correct for these two factors, the lowest curve resulting from our proposed estimator. For comparison, current mean or modal estimates are roughly 0.1 for terrestrial vertebrates and 0.35 for arthropods.

temporal variability, and the effect increases with clumping. For example, the *mean* value of the standard measure is three to four times the real value at clumping values often seen in nature ($5 < c < 10$; see below); actual values will often be higher still.

Second, our proposed estimator, $s_{LJ}^2(\text{adj})$, which adjusts for both spatial variation and Jensen's inequality, is a great improvement and overall does very well. Its mean value is very close to the real value (0.1) over a broad range of clumping and is never greater than 0.141, less than 50% greater than the true value. Even when we adjust only for spatial contamination, the mean of $s_L^2(\text{adj})$ is never greater than 0.185. The Jensen's inequality correction improves estimates, but not markedly: the reason seems to be that, while the unadjusted estimates underestimate the true value of Λ_t, they underestimate it by about the same amount for each density, Δ_t, so the spread in values, which is the focus of estimation of s_Λ^2, is not much affected.

Clumping commonly decreases with mean density in real populations (Murdoch and Stewart-Oaten 1989), so we calculated the four estimates for several situa-

tions in which clumping declined as mean density increased from one to five. We then examined the relationships between mean clumping and the four estimates of temporal variability. The results are very little changed from those presented above: the estimates were very close to those for the case with fixed clumping, though the standard estimate does slightly worse in all cases than before (by an average of 0.015) and the new estimate slightly better (by an average of 0.006).

Finally, we also looked at larger sample sizes (minimum number of samples for each density estimate, $r = 15$). Surprisingly, in this situation our estimator is more affected by clumping than when $r = 5$. For example, its mean estimate of the temporal variance (whose true value is 0.1) is 0.174 when $c = 10$. However it still does better than the other estimates, whose means are 0.190 (no adjustment for Jensen's inequality), 0.258 (no adjustment for spatial variation), and 0.275 (no adjustment for either) in this case. We suspect that the problem is related to the Taylor series approximation but have not yet been able to develop an acceptable improvement.

IMPLICATIONS FOR DIFFERENCES AMONG TAXA

The results in figure 2 raise the possibility that the published differences in estimated temporal variability among major taxonomic groups might be, in large part, artifactual. Terrestrial vertebrates have been recorded as having temporal variance values around 0.1: this is about the modal value for small mammal populations and about the mean for birds (equivalent to a standard deviation of logs about 0.3; see Ostfeld 1988, figs. 1 and 2). This is also, by design, the true value for the simulations presented in figure 2. By contrast, the modal value found to date for terrestrial arthropods is about 0.35 (equivalent to the published value for the standard deviation of logs of about 0.6) (Ostfeld 1988). It is possible, however, that terrestrial vertebrates and arthropod populations have much more similar temporal variabilities than these differences suggest, for two reasons, both related to differences in the accuracy of density estimates among taxa.

First, terrestrial insects have notoriously clumped spatial distributions. Clumping (c) values of 5–10 appear to be common, and even values around 20 have been observed (see, e.g., Atkinson and Shorrocks 1984; Reeve 1985; Hassell et al. 1987). Thus the standard approach, using common sample sizes, will often assign to insects values for temporal variance in the range of 0.3–0.4 (i.e., typically reported values) when the true temporal variance is actually only 0.1. We do not know of clumping estimates for terrestrial vertebrates, but such species, in contrast to arthropods, often have behavioral mechanisms, such as territoriality, that lead to even spatial distributions. Second, some terrestrial vertebrates such as breeding birds can be censused rather than sampled, in which case there is virtually no error in the estimate of temporal variability regardless of spatial clumping. Thus the observed values around 0.1 for many terrestrial vertebrates may be very close to their true value. Indeed, Xia and Boonstra (1992) have pointed out that the tendency for saturation of small mammal traps, nesting boxes in birds, and home ranges in other vertebrates (and one could add breeding territories in birds) induces *underestimation* of the temporal variability of the total

population whenever density moves into the range in which these effects occur. This type of problem does not arise in typical sampling methods for arthropods.

These points touch on a problem we have not addressed. Estimates of temporal variability are themselves variable and should have confidence intervals attached. We have not done this in our simulations partly because the estimation of the variance of temporal variance estimates is a complicated topic that would distract from our main concern, the bias arising from spatial variance. These intervals would often be wide, unless many dates were used, far more than the four of our simulations.

Still, the imprecision in temporal variance estimates is not a sufficient explanation for the consistent finding, across studies, of higher invertebrate than vertebrate variation. In some cases, of course, the explanation is likely to be that the finding represents reality: insect pests that defoliate large areas of forests and then virtually disappear for years at a time obviously fluctuate enormously in contrast to many vertebrate populations that are known to be quite stable. But we believe that the bias due to spatial variance, together with the other factors noted here, probably contributes substantially to many apparent differences.

<div style="text-align:center">DISCUSSION</div>

Many ecological questions require an estimate of the temporal variability of the abundance of organisms in a region. We have shown that one of the estimates currently used, the variance or standard deviation of successive estimated log densities, is not appropriate. It measures variability in the log of the estimated $(\ln[D_t])$ rather than the true $(\ln[\Delta_t])$ population density. On the average, this measure overestimates true temporal variability: it reflects not only the variation of $\ln(\Delta_t)$ over time but also sampling error—that is, the variation of $\ln(D_t)$ around $\ln(\Delta_t)$ due to spatial variation. The size of the overestimate increases if the spatial variance of the population increases or the sampling intensity decreases. Other standard estimates, such as the coefficient of variation, and D_{max}/D_{min}, seem likely to have a similar problem.

We propose the measures s_L^2 (adj) (eq. [11]), where s_L^2 is as in equation (5). For existing data sets with no zero dates, L_t is given by L_{tJ} in equation (21), and s_{Lt}^2 is given by equation (22). For future data sets, we propose the sequential scheme and the estimates in the appendix. These measures estimate s_Λ^2, the (temporal) sample variance of the natural logs of the true population densities, $\{\Delta_t\}$, over a set of observation dates, $\{t\}$. (For s_{L10}^2, using logs to base 10, multiply by $[\log_{10}e]^2$.) They take spatial variability into account by removing its estimated effects at two stages, and they sharply reduce the bias for most true densities and levels of clumping encountered in practice. We believe our new estimates are applicable to a wide range of ecological problems.

Data presented elsewhere (Murdoch et al. 1995) provide an example from red scale populations on grapefruit trees in a grove in California under biological control. The standard approach yielded an estimate of temporal variance (uncorrected $s_{L10}^2 = 0.176$) that was three times that obtained using the estimator in this article (s_{L10}^2 [adj] $= 0.06$). This was due to spatial variance, which remained large

even though sample size was large (a total of 32–64 twigs taken from 16 trees), and we used data from only one part of the tree (the outer region) and hence greatly reduced the variability in density among samples.

Our results suggest that existing analyses of temporal variability, and the conclusions drawn from them, need reappraisal. This is especially true for attempts to determine whether vertebrate populations, especially lizards, birds, and some mammals, are less variable through time than are arthropod populations, a question with broad implications for both population and community theory (Connell and Sousa 1983; Schoener 1985; Ostfeld 1988). Vertebrate populations may well be less variable on the average, but this result could also be, at least in part, an artifact of the statistic used.

Overestimation induced by spatial variance is not the only problem in the comparison of temporal variability among taxa. The zero dates problem, mentioned above, also confounds present comparisons. For example, Schoener (1986) added 0.1 to each estimate of spider density on Bahamian islands but not to estimates of lizard densities; adding a smaller (larger) amount would likely have increased (decreased) the estimate of spider variability. By contrast, J. H. Connell and W. P. Sousa (1983, personal communication) added one to all estimates of density whether or not zeroes existed, a procedure that is likely to have reduced estimated temporal variability more in low-density than in high-density populations.

McArdle et al. (1990) commented on these and several other issues. They note that estimated temporal variability tends to increase with the number of sampling periods, often because of autocorrelation (McArdle 1989). The true Λ_t's (and Δ_t's) are themselves random variables, forming part of a time series, the realization of a stochastic process. Values close in time are more likely to be similar in size; a small number of sampling periods implies a short duration for the entire study: all values are then close in time, so their variance, s_Λ^2, is likely to underestimate the variance of the process, σ_Λ^2. This offers yet another reason, in addition to sampling error (spatial variance) and trap saturation, for reviewing current evidence comparing the temporal variability of arthropods and mammals: the available arthropod estimates are based on many more years of data than the mammal estimates.

Over how long a time period should samples be taken, and what should be their frequency? Connell and Sousa (1983) among others have already pointed out the need for care in comparing organisms with dissimilar life histories, particularly generation times. Samples of long-lived organisms such as trees must be taken over long periods since the populations will change only slowly. Sampling of an insect population, with more or less continuous breeding and several overlapping generations per year, might be done over only a few years. These authors suggested that generation time should be the temporal unit of measurement. However, the issue is broader than differences in generation time and hinges instead upon the type of fluctuations we are interested in.

For example, freshwater zooplankton show substantial variations in density over most of the year on a frequency of weeks; they also show characteristic fluctuations in response to the seasons, and, in some situations, the peak spring density varies significantly on an annual basis. Each of these types of fluctuations arises from a different cause, and it may make sense to focus on only one for

comparative purposes. Analysis of the short-term fluctuations requires sampling every few days within a single season; analysis of spring peak dynamics might require only a few samples each year but many years of sampling.

Again, consider two annual insect species with, obviously, the same generation time. The winter moth fluctuates from year to year in English woodlands (Varley and Gradwell 1963). There does not appear to be any longer-term process at work. By contrast, the larch bud moth in Swiss forests cycles over several orders of magnitude with a period of almost a decade (Baltensweiler 1971), while another annual insect, the spruce bud worm, seems to have a period of several decades (Royama 1984).

One appealing measure of temporal variation in a stochastic process, $\{\Lambda_t\}$, is the total power: the mean (over "runs" of the process, which is assumed repeatable) of the average (over time) of $(\Lambda_t - \Lambda)^2$, where Λ is the mean of the average of the Λ_t's. In practice, we assume the time average is enough; that is, $\Lambda = \lim_{T \to \infty} \int_{-T}^{T} \Lambda_t \, dt/2T$ and power $= \lim_{T \to \infty} \int_{-T}^{T} (\Lambda_t - \Lambda)^2 dt/2T$. This is the integral of contributions due to fluctuations at different frequencies (Priestley 1981). Our zooplankton and insect examples above are to suggest total power will be poorly estimated if important frequencies are unrepresented; unimportant frequencies *can* be unrepresented without serious loss (e.g., the power for zooplankton cycles is concentrated at high frequencies; McCauley and Murdoch 1987), so frequent sampling can be compressed into a short period, while the power for trees is concentrated at low frequencies, so samples do not need to be frequent; and total power may *not* be what we want, if our interest is in the variation of particular kinds or from particular sources. For example, if our interest is in unpredictability rather than variation, we might wish to estimate the power after seasonal variation has been removed.

Several aspects of scale complicate the picture we have presented. The definition of a "population" is to some extent arbitrary (McArdle and Gaston 1993), and the amount of temporal variability will often depend on this definition. It has been noted (Murdoch et al. 1985, fig. 1) that the temporal variability of a metapopulation is likely to decrease as we sample from increasingly larger areas; Reeve (1988) provides theoretical examples. This implies we need to know something of the spatial dynamics of the populations we seek to assess or compare.

Variation in density and population will be equal if the same region is used for each survey and if it is large enough that no member of the population is ever outside it; for example, the population might be defined as all individuals within the sampled region. B. McArdle (personal communication) has pointed out that this restriction can be subtle. For example, if density is measured by the number of insects per twig, then the number of twigs in the region would need to be constant over sampling times. If it is not, then the population size might be estimated by (estimated number per twig) × (estimated number of twigs per tree) and estimated log(population) by estimated log(density) + estimated log(number of twigs), so that temporal variation would have two components, both subject to error, and to adjustments like those we have discussed.

McArdle et al. (1990) also suggest that estimates of temporal variability of different species need to be based on samples taken at "biologically equivalent" spatial scales: samples from an area a few meters square may suffice to indicate

the temporal variability of soil mite populations but not of breeding bird populations. We are not sure how to interpret this operationally, but it might be useful to relate the spatial scale of sampling to, say, the range of individual movement.

In contrast to these population-level questions, those focused on the behavior of individuals or on evolutionary questions will be poorly served by averaging across space in the way we have suggested. The experience of an individual forager is determined by spatial variation on the scale of its ambit, compounded by temporal variation at each point in space. In such cases, averaging done to answer population-level questions suppresses the relevant information.

APPENDIX
DEALING WITH ZERO DATES
EXISTING DATA SETS

The following discussion presents six possible responses to the existence in a data set of dates on which no organisms were found.

First, add a positive constant to all mean density estimates. This is the standard fix, but it has serious drawbacks. Large constants tend to linearize the values of the logs: $\ln(D_t. + c)$ approaches $\ln(c) + D_t./c$, so the aims and advantages of transforming to logs may be lost. Small constants have a large effect on estimated temporal variance because they have very negative logs. These effects will differ among populations, especially between large populations (large $D_t.$) and small ones: the choice of constant may be a major factor in a comparison.

This choice is little more than arbitrary. Formal transformation methods (see, e.g., Box and Cox 1964) do not work well in this case (Berry 1987). "Standard" constants, like 1, 1/2, 1/6 (Mosteller and Tukey 1977), or 3/8 (Anscombe 1948) have ad hoc defenses and are sometimes ambiguous: for example, "$c = 1$" could mean $\ln(D_t. + 1)$ or $\ln(D_t. + 1/r_t)$ (adding one to the total number found, before averaging) or something else (if $D_t.$ referred to an average density per cm^2, or per m^2, etc.). We would prefer to add the constant to the total, before averaging, partly to avoid linearizing and partly for ad hoc reasons like "pretend we saw half an extra animal."

It may be helpful to calculate temporal variability with a range of constants (e.g., from 0.1 to 1) to see whether the comparisons give about the same results.

Second, add the constant only when $D_t. = 0$. An argument for this option is that, while $D_t.$ might be a sensible estimate of Δ_t in other cases, we often know (for certain) that $\Delta_t > 0$ so $D_t. = 0$ is sure to be an underestimate. Again, a range of constants could be used to check consistency.

Third, calculate temporal variance on the basis of a running average of dates taken two at a time (or more, depending on the maximum number of adjacent dates with zeros).

This method has obvious drawbacks. It will reduce the estimate of temporal variability, especially that owing to short-term variability. The existence of only one zero, or series of zeros, will require smoothing all the data over the entire time scale.

Fourth, in some cases, it may be possible to assign a reasonable "prior" distribution to the possible values of Δ_t (or of Λ_t) and use Bayesian methods for the estimation (see Bishop et al. 1975, chap. 12). The prior distribution could include the possibility that $\Delta_t = 0$, but it will always lead to nonzero estimates provided it gives positive probability to some nonzero values of Δ_t that could lead (via sampling error) to zero values of $D_{t.}$.

The prior distribution would be arbitrary to some extent, but not necessarily any more arbitrary than the added constant in the first or second responses, and it may correspond better with common sense. One possibility is to assume that the true Δ_t's are independent and lognormally distributed and that the spatial variance of D_{t1}, D_{t2}, . . . is a function of Δ_t, with one or more parameters to be estimated. The parameters of the lognormal, including the temporal variance, could then be estimated directly from the D_{ti}'s (e.g., by maximum likelihood).

Fifth, when periods of true zero abundance are expected, the true variation of log(density) is infinite, so a logarithmic measure in this case has no meaning. If the zero period is known, it could be omitted: it may be of no biological interest, as when we are comparing species of insects whose adults appear only in the summer months.

Sixth, if there are genuine but unpredictable zeros, then the logarithmic measure is inappropriate. McArdle et al. (1990) discuss in detail the clearest alternative, the coefficient of variation (CV) of the raw $D_{t.}$'s. They point out that this may not estimate the "true" time series CV, $\sigma_\Delta/E\{\Delta\}$, as well as s_Δ^2 estimates its target, σ_Δ^2. If the Δ_t's are strongly skewed, as is common, then s_Δ^2 will usually underestimate σ_Δ^2, compensating by occasional gross overestimates. This problem may be worth studying in more detail, since it may not be as bad as it seems: $\Delta.$ will also usually underestimate $E\{\Delta\}$, and s_Δ^2 and $\Delta.$ are positively correlated.

In comparisons of species, it is possible to have $CV_1 > CV_2$ while $\sigma_{\Lambda 1}^2 < \sigma_{\Lambda 2}^2$, where the 1 and 2 refer to the species. If the distribution (over time) of the Δ_t's is lognormal, then the CV is $\sqrt{\exp(\sigma_\Lambda^2)} - 1$; if the distribution of the logs is uniform, the CV is $\sqrt{\sigma_\Lambda \sqrt{3} \coth(\sigma_\Lambda \sqrt{3})} - 1$. Thus if two species are both lognormal or both "log uniform," then the one with the larger CV will have the larger σ_Λ^2. But this may not be true if the distributions are different, and the situation is still messier if the CV depends on the mean of the logs as well as on the variance.

A third problem is that when the Δ_t's are not known and must be estimated, the coefficient of variation still conflates spatial and temporal variability. The effect of spatial variation on the CV may be complicated, because spatial variation affects both the numerator (estimated SD over time) and the denominator (estimated average over time).

Of course, an existing data set may not have any zeros. Even here, however, there are complications. If it is being compared to a data set that does have zeros, it seems reasonable that whatever "fix" is used on the latter should be applied to both. A deeper problem is that most of the properties of tests, estimators, and other methods in frequentist statistics depend not only on what was observed but also on what could have been observed. For example, to decide whether an estimator is unbiased, we need to know what its value would be for every possible sample, even if that sample did not arise in a particular use. Thus, unless it is actually impossible to get $D_{t.} = 0$, we cannot describe the estimate of Λ_t as "$\ln(D_{t.})$" since this is undefined at zero. The second option above offers the simplest way out since it does not affect a data set without zeros, though it adds an arbitrary element to comparisons with data sets containing zeros, since the choice of constant will affect one set but not the other.

FUTURE DATA SETS

The sequential sampling scheme used in the simulations presented in figure 2 is to choose (1) a minimum sample size, r; (2) a required number of nonzero twigs, m ($m \geq 2$); and (3) a maximum sample size, R. Take a sample of r twigs; if this contains fewer than m nonzero twigs, continue sampling until m nonzero twigs are found or until a total of R twigs has

been chosen. The estimate of mean density per twig is $D_{tS} = D_t.$, the average number per twig, unless the sampling is stopped in its "sequential" phase—that is, if $r < n =$ the number of twigs in the sample and m nonzero twigs are found; in this case, $D_{tS} = D_{t+} (m - 1)/(n - 1)$, where D_{t+} is the average number on the nonzero twigs. This estimate is unbiased; we modify it to avoid zeros by defining $D_{tS} = 1/2R$ if all R twigs are zeros.

The estimated variance of D_{tS}, σ_{DS}^2, is messier, though it is usually very similar to $s_D^2.$, in equation (19). It is given elsewhere (Stewart-Oaten 1995) and in more general form by Kim and Nachlas (1984) and Kremers (1987). We then estimate $\Lambda_t = \ln(\Delta_t)$ by $L_{tJ} = \ln(D_{tS}) + s_{DS}^2/2D_{tS}^2$, following equation (21), and the variance of L_{tJ} by $s_{Lt}^2 = s_{DS}^2/D_{tS}^2$, following equation (22).

For each combination of a mean density and a clumping parameter in the simulations presented in figure 2, we calculated $L_t = \ln(D_{tS})$, L_{tJ}, and s_{Lt}^2 for 10,000 samples and obtained the mean and variance of the L_t and L_{tJ} values, and the mean of the s_{Lt}^2 values. Taking these to be the true means and variances of these quantities, we then computed the means of the four estimates of s_Λ^2: $s_L^2 = \Sigma(L_t - L.)^2/(T - 1)$, $s_{LJ}^2 = \Sigma(L_{tJ} - L._J)^2/(T - 1)$, s_L^2 (adj) $= \Sigma(L_t - L.)^2 - \Sigma s_{Lt}^2/T$, and s_{LJ}^2 (adj) $= \Sigma(L_{tJ} - L._J)^2 - \Sigma s_{Lt}^2/T$, where the sums are over the densities (1, 2, 4, and 5). These means were then divided by $\ln(10)^2$ to convert them to base 10.

LITERATURE CITED

Anscombe, F. J. 1948. The transformation of Poisson, binomial and negative binomial data. Biometrika 35:246–254.

Atkinson, W. D., and B. Shorrocks. 1984. Aggregation of larval Diptera over discrete and ephemeral breeding sites: the implications for coexistence. American Naturalist 124:336–351.

Baltensweiler, W. 1971. The relevance of changes in the composition of larch bud moth populations for the dynamics of its numbers. Pages 208–219 in P. J. den Boer and G. R. Gradwell, eds. Dynamics of populations: proceedings of the Advanced Study Institute on "Dynamics of Numbers in Populations," Oosterbeek, The Netherlands, September 7–18, 1970. Centre for Agricultural Publishing and Documentation, Wageningen.

Berry, D. A. 1987. Logarithmic transformations in ANOVA. Biometrics 43:439–456.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. 1975. Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, Mass.

Box, G. E. P., and D. R. Cox. 1964. An analysis of transformations. Journal of the Royal Statistical Society B 26:211–252.

Connell, J. H., and W. P. Sousa. 1983. On the evidence needed to judge ecological stability or persistence. American Naturalist 121:789–824.

Fairweather, P. G. 1988. Predation can increase variability in the abundance of prey on seashores. Oikos 53:87–92.

Feller, W. 1966. An introduction to probability theory and its applications. Vol. 1. Wiley, New York.

Gaston, K. J. 1988. Patterns in the local and regional dynamics of moth populations. Oikos 53:49–57.

Gaston, K. J., and J. H. Lawton. 1988a. Patterns in the distribution and abundance of insect populations. Nature (London) 331:709–712.

———. 1988b. Patterns in body size, population dynamics, and regional distribution of bracken herbivores. American Naturalist 132:662–680.

Haldane, J. B. S. 1945. On a method of estimating frequencies. Biometrika 33:222–225.

Hanski, I. 1982. On patterns of temporal and spatial variation in animal populations. Annales Zoologici Fennici 19:21–37.

Hassell, M. P., T. R. E. Southwood, and P. M. Reader. 1987. The dynamics of the viburnum whitefly (Aleurotrachelus jelinekii): a case study of population regulation. Journal of Animal Ecology 56:283–300.

Kim, S., and J. A. Nachlas. 1984. Estimation in Bernoulli trials under a generalized sampling plan. Technometrics 26:379–387.

Kremers, W. K. 1987. An improved estimator of the mean for a sequential binomial sampling plan. Technometrics 29:109–112.

McArdle, B. H. 1989. Bird population densities. Nature (London) 338:628.

McArdle, B. H., and K. J. Gaston. 1993. The temporal variability of populations. Oikos 67:187–191.

McArdle, B. H., K. J. Gaston, and J. H. Lawton. 1990. Variation in the size of animal populations: patterns, problems and artifacts. Journal of Animal Ecology 59:439–454.

McCauley, E. F., and W. W. Murdoch. 1987. Cyclic and stable populations: plankton as paradigm. American Naturalist 129:97–121.

Mosteller, F., and J. W. Tukey. 1977. Data analysis and regression. Addison-Wesley, Reading, Mass.

Murdoch, W. W., and A. Stewart-Oaten. 1989. Aggregation by parasitoids and predators: effects on equilibrium and stability. American Naturalist 134:288–310.

Murdoch, W. W., J. Chesson, and P. L. Chesson. 1985. Biological control in theory and practice. American Naturalist 125:344–366.

Murdoch, W. W., R. F. Luck, S. L. Swarbrick, S. Walde, D. S. Yu, and J. D. Reeve. 1995. Regulation of an insect population under biological control: effects of ants on a refuge from parasitism. Ecology 76:206–217.

Ostfeld, R. S. 1988. Fluctuations and constancy in populations of small rodents. American Naturalist 131:445–452.

Priestley, M. B. 1981. Spectral analysis and time series. Academic Press, New York.

Redfearn, A., and P. L. Pimm. 1988. Population variability and polyphagy in herbivorous insect communities. Ecological Monographs 58:39–55.

Reeve, J. D. 1985. Biological control of the California red scale, *Aonidiella aurantii*, by the parasitoid *Aphytis melinus:* a test of ecological theory. Ph.D. diss. University of California, Santa Barbara.

———. 1988. Environmental variability, migration, and persistence in host-parasitoid systems. American Naturalist 132:810–836.

Royama, T. 1984. Population dynamics of the spruce budworm *Choristoneura fumiferana*. Ecological Monographs 54:429–462.

Schoener, T. W. 1985. Are lizard population sizes unusually constant through time? American Naturalist 126:633–641.

———. 1986. Patterns in terrestrial vertebrate versus arthropod communities: do systematic differences in regularity exist? Pages 556–586 *in* J. Diamond and T. J. Case, eds. Community ecology. Harper & Row, New York.

Schoener, T. W., and D. A. Spiller. 1992. Is extinction rate related to temporal variability in population size? an empirical answer for orb spiders. American Naturalist 139:1176–1207.

Snedecor, G. W., and W. G. Cochran. 1989. Statistical methods. Iowa State University Press, Ames.

Spitzer, K., and J. Leps. 1988. Determinants of temporal variation in moth abundance. Oikos 53: 31–36.

Spitzer, K., M. Rejmanek, and T. Soldan. 1984. The fecundity and long-term variability in abundance of noctuid moths (Lepidoptera, Noctuidae). Oecologia (Berlin) 62:91–93.

Stewart-Oaten, A. 1995. Sequential estimation of log(abundance). Biometrics (in press).

Taylor, L. R. 1961. Aggregation, variance and the mean. Nature (London) 189:732–735.

Varley, G. C., and G. R. Gradwell. 1963. The interpretation of insect population changes. Proceedings Ceylon Association of Advancement of Science 18:142–156.

Watt, K. E. F. 1965. Community stability and the strategy of biological control. Canadian Entomologist 97:887–895.

Williamson, M. H. 1984. The measurement of population variability. Ecological Entomology 9: 239–241.

Xia, X., and R. Boonstra. 1992. Measuring temporal variability of population density: a critique. American Naturalist 140:883–892.

Associate Editor: Robert D. Holt