



International Association for Ecology

On the Stabilization of Animal Numbers. Problems of Testing. 1. Power Estimates and Estimation Errors

Author(s): J. Reddingius and P. J. den Boer

Source: *Oecologia*, Vol. 78, No. 1 (1989), pp. 1-8

Published by: [Springer](#) in cooperation with [International Association for Ecology](#)

Stable URL: <http://www.jstor.org/stable/4218824>

Accessed: 24/10/2014 04:13

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Springer and International Association for Ecology are collaborating with JSTOR to digitize, preserve and extend access to *Oecologia*.

<http://www.jstor.org>

Original papers

On the stabilization of animal numbers. Problems of testing*

1. Power estimates and estimation errors

J. Reddingius and P.J. den Boer

Biological Station LUW, Kampsweeg 27, 9418 PD Wijster, The Netherlands

Summary. It is tried to remove some misunderstandings about the “regulation” of animal numbers by density-dependent processes. Staying between limits is called “stabilization”, and only when this results from density-dependent processes it is considered “regulation”. We discuss two tests that may be used to detect the existence of regulation: the parametric tests of Bulmer and a nonparametric “permutation” test. The powers of these tests are compared. The first test of Bulmer and the permutation test do not differ very much in power, but the second test of Bulmer has hardly any power and therefore cannot be used in cases where densities were only estimated. The arguments from which Bulmer’s second test is derived are critically discussed and found not to be very convincing. We propose, rather than using Bulmer’s second test, to correct the test statistic of his first test for estimation error, and present a possible solution for this. In a following paper this method will be applied to some long series of population counts of univoltine insects, for, a basic assumption of all regulation-tests is, that the sequence of population counts is a realization of a piece of first-order Markov chain. This highly restricts the usefulness of regulation-tests. Some other recent attempts to construct such tests are discussed within the present context

Key words: Density dependence – Regulation – Stabilization

Since the classical paper of Howard and Fiske (1911) arguments pro and contra the hypothesis of the “regulation” of animal numbers by density-dependent processes have dominated the literature on population dynamics (see e.g. Pielou 1974). Even today this debate cannot yet be closed, not in the least because some conceptual misunderstandings have still to be removed.

For instance, Varley et al. (1973: 112) state: “Density-dependent mortality serves to regulate the population density and keeps it within limits”. But this statement is far less obvious than it may seem. To begin with, we need a definition of “regulate”. Let us assume this definition will be something like: “the operation of negative feedback processes that balance density around some stable norm or equilibrium value” (compare Varley et al. 1973: 19). Then it must be noted that:

* Communication No. 261 of the Biological Station, Wijster

Offprint requests to: P.J. den Boer

(a) Density-dependent mortality alone is neither necessary, nor sufficient for regulation to occur. See: Klomp (1966), Varley and Gradwell (1968), Den Boer (1986, 1987, 1988).
 (b) Regulation in the sense indicated is not equivalent to “keeping density within limits”, because density may remain between limits for appreciable lengths of time by mere stochastic fluctuations (Reddingius 1971; Reddingius and Den Boer 1970), and density might also remain between limits without the existence of a stable equilibrium value (e.g. limit cycles: Kolmogoroff 1936; May 1973).

Therefore, to keep the following intelligible, and to promote adequate testing, we have to make some distinctions. The often supposed (general) tendency for population density to stay for some time between relatively narrow limits will be called “stabilization” (Den Boer 1968). Stabilization may, but need not, result from the influence of “governing” (density-dependent) factors (Reddingius 1971). Only if this is the case it shall be called “regulation”. When developing adequate tests we have therefore to distinguish two questions:

(1) Did density stay between certain (narrow) limits? In this case it is not determined whether the restricted density range (range stability) is caused by governing factors (regulation), or results, for instance, from spreading of risk (Den Boer 1968; Reddingius and Den Boer 1970; Reddingius 1971).
 (2) Is the growth rate (coefficient of net reproduction) of the population significantly and negatively correlated with density? In this case the limits of density are not considered.

It has often been tried to develop tests that would answer the second question. Several authors, among whom Reddingius (1971) and Pielou (1974) proved, however, that the simple regression techniques generally used to show the occurrence of regulation are inadequate. After discussing the objections, Reddingius (1971, Ch. 10) extensively treated a valid parametric test based on the idea of governing (regulation) by density-dependent processes. Independently, Bulmer (1975) developed a similar test.

Reddingius (1971, Ch. 11) also considered a nonparametric test suitable for answering the second question. The test uses the rank of the log-range [$LR = \log(\text{highest density}) - \log(\text{lowest density})$; Reddingius and Den Boer 1970; Den Boer 1971] of the field data within the collection of log-ranges obtained from all possible permutations of the coefficients of net reproduction that were derived from the field data. The idea behind this test is, that in the case of stabilization of density the actual sequence of net repro-

duction values can be expected to have resulted in a better "range stability" (a lower log-range) than in a significant majority of the permuted series. This permutation test was not developed further by Reddingius because, exactly when the field data become interesting, i.e. cover a satisfactory number of generations, the test becomes unmanageable: the number of permutations of m values, $m!$, increases rapidly with m ; e.g. with $m=8$ we already have 40320 permutations, and with $m=10$ even 3628800. With the present rapid computers, however, it becomes feasible to take large samples from all possible permutations, by which it may be expected that the chance to accept the null hypothesis wrongly can be sufficiently reduced.

Note that, although "narrow limits" cannot be defined independently of the data, the permutation test helps us in answering the first question (the one concerning stabilization) by comparing the actual log-range to what might be expected given the variation in net reproduction as it happens to occur in these data.

It is argued above that question (1) is not equivalent to question (2). Nevertheless, "regulation" in any acceptable sense of the term will result in densities remaining between limits. Therefore, if the density does not stay between reasonable limits, regulation does not occur. The permutation test, then, may be used to investigate whether in a given case there is a sufficient amount of evidence for limitation to warrant an investigation for density-dependent processes. In practice, therefore, the permutation test may be used as a nonparametric alternative for Bulmer's test, with the restriction that we cannot exclude "range stability" over a considerable period from other sources than strict "regulation" (Den Boer 1981), e.g. from what may be called "stochastic boundedness" (Murdoch 1979).

We will therefore compare the powers of Bulmer's tests and the permutation test using a first-order autoregression model in which the value of a certain parameter determines whether or not the density, according to this model, is regulated. In a following paper the tests will be applied to some published population counts, from which we may gather an impression of the strength of the evidence supporting the regulation thesis.

Methods

Let N_1, N_2, \dots, N_n denote population densities, or sizes, of a given population of a univoltine species at equidistant times $t=1, 2, \dots, n$. Let $X_t = \ln N_t$, $t=1, 2, \dots, n$, and let $Y_t = X_{t+1} - X_t = \ln R_t$ for $t=1, 2, \dots, n-1$. We wish to test the hypothesis that Y_t is independent of X_t against the alternative that $E(Y_t | X_t = x)$ is a decreasing function of x . We may do this by means of either a parametric test or a nonparametric test.

A survey of parametric tests was given by Reddingius (1971). If $\mu = E(X_t)$ it may be postulated that

$$X_{t+1} - \mu = \beta(X_t - \mu) + Z_t, \quad t=1, 2, \dots, n-1$$

i.e.

$$Y_t = (1 - \beta)\mu - (1 - \beta)X_t + Z_t$$

where $\beta \leq 1$, and Z_1, Z_2, \dots, Z_{n-1} is a sequence of normally distributed, independent random variables with expectation zero and a constant variance σ^2 . This model has been discussed in detail by Reddingius (1971). It can be shown that the random process $(X_t, t=1, 2, \dots)$ is stationary in the weak

sense if, and only if, $|\beta| < 1$. If $\beta = 1$, it will be seen that Y_t does not depend on X_t at all, but is just a random number Z_t . So our testing problem may be put in parametric terms thus: The null hypothesis $H: \beta = 1$ has to be tested against the alternative $A: |\beta| < 1$. Of the tests considered, the best appeared to be the likelihood ratio test, which on theoretical grounds is expected to be fully efficient asymptotically anyway. However, no exact significant points were available for this test, so these as well as estimates of its power for various values of β were estimated by Monte Carlo simulations (Reddingius 1971). A few years later, Bulmer (1975) proposed a test statistic which is almost equivalent to the likelihood ratio statistic of Reddingius (1971), and for the critical values of which he was able to give accurate approximation formulae. We therefore consider the test based on this statistic as our best parametric test, and in the following we will indicate it as "Bulmer's first test". Bulmer envisaged that $0 < \beta < 1$; in our opinion, this restriction is not necessary for his first test; in fact, it is interesting to consider values of β such that $-1 < \beta \leq 0$ as well. Bulmer also discussed the case where population densities cannot be exactly measured, but have to be estimated from samples. In that case, we do not measure X , but rather X^* , with

$$X_t^* = X_t + E_t$$

where E_t is the estimation error. Again, we will assume that E_1, E_2, \dots, E_n are independent, and have expectation zero. Bulmer assumed the E 's to have a constant variance σ_d^2 . For this case he proposed a test which we will indicate as "Bulmer's second test". Let us define

$$U = \sum_{t=1}^{n-1} (X_{t+1} - X_t)^2$$

$$V = \sum_{t=1}^n (X_t - \bar{X})^2, \quad V^* = \sum_{t=1}^n (X_t^* - \bar{X}^*)^2$$

$$W = \sum_{t=1}^{n-2} (X_{t+2}^* - X_{t+1}^*)(X_t^* - \bar{X}^*)$$

where

$$\bar{X} = \left[\sum_{t=1}^n X_t \right] / n, \quad \bar{X}^* = \left[\sum_{t=1}^n X_t^* \right] / n$$

Bulmer's first test uses the statistic $R = V/U$. It may be seen that the null hypothesis must be rejected for small values of R . The null hypothesis is rejected at the α level of significance whenever $R < R_{L\alpha}$, with $R_{L\alpha} = 0.25 + (n-2)x_\alpha$, where x_α can be taken from Table 1 of Bulmer (1975); for example, for a 0.05 level test, $x_{0.05} = 0.0366$. Bulmer's second test is based on the statistic $R^* = W/V^*$. The null hypothesis is rejected at α level whenever $R^* < R_{L\alpha}^*$. Approximation formulae for $R_{L\alpha}^*$ are given by Bulmer (l.c.). For example, for $\alpha = 0.05$, we have approximately,

$$R_{L\alpha}^* = -13.7/n + 139/n^2 - 6113/n^3.$$

For the nonparametric test, we take the Y -values just as they happen to occur. The null hypothesis states that the order in which they appear is a random one, i.e. all permutations of the sequence of R -values would have been equally likely. The alternative hypothesis is that the Y -values occur in such an order that the resulting fluctuations in the X -sequences are restricted. As a measure for the amount

of fluctuation we use the logarithmic range, LR , as before. To illustrate the idea, suppose we have observed $X_1 = 7.196, X_2 = 7.613, X_3 = 6.385, X_4 = 6.342, X_5 = 6.409, X_6 = 7.196, X_7 = 7.618, X_8 = 7.064, X_9 = 6.391, X_{10} = 7.251$. Then $LR = 7.618 - 6.342 = 1.276$. The corresponding Y -values (e.g. $Y_1 = 7.613 - 7.196 = 0.417$, etc.) are $0.417, -1.228, -0.043, 0.066, 0.568, 0.642, -0.554, -0.673, 0.860$.

If they might have occurred in any order, we might as well have had $0.860, 0.417, 0.066, 0.568, 0.642, -0.554, -1.228, -0.043, -0.673$,

but then the resulting X -sequence would have been (e.g. $X_2 = 7.196 + 0.860 = 8.056$, etc.): $7.196, 8.056, 8.473, 8.539, 9.107, 9.749, 9.195, 7.967, 7.924, 7.251$ with $LR = 2.553$.

The question then is whether the observed LR (1.276 in this case) is significantly smaller than might be expected under the hypothesis of equally likely random permutations. As remarked in the introduction, working out of the permutation distribution of the LR 's is too formidable a task. Therefore we now propose the following solution: by means of pseudo-random numbers generated by a computer one generates a random sample of the possible permutations of the observed Y 's, compute the ensuing X -es, find their LR 's, and thus test the hypothesis that the originally observed LR is a random drawing from the same population as the sample of LR 's obtained by permutations. This is a special case of a Mann-Whitney test where one of the two samples to compare has size 1. Suppose we have k permutations, resulting in k LR 's. Let the observed LR be denoted by L_0 , and the first, second, ..., $k + 1$ st order statistic of the combined set of LR 's be $L_{(1)}, L_{(2)}, \dots, L_{(k+1)}$.

Suppose we wish a test of size α . Let the smallest integer that is not less than $(k + 1)\alpha$ be c . Then the null hypothesis will be rejected whenever

$$L_0 \leq L_{(c)}$$

Another way to formulate the same is this: let r be the rank of L_0 in the sequence of $L_{(i)}$'s; then, if $r/(k + 1) \leq \alpha$, the null hypothesis is rejected. Note that r is the number of LR 's in the sample of permutations that are not greater than L_0 .

Hence in order to perform this test, it is only necessary to count how many times an LR is obtained which is not greater than L_0 . Also, $r/(k + 1)$ is a P -value which may be mentioned as a result of the test. The generation of random permutations is discussed in Appendix 1.

Results and discussion

The power of the three tests considered was estimated for various values of β by performing the test a large number of times on random realizations of segments of the stochastic process ($X_t, t = 1, 2, \dots$), defined in the previous section. For a given value of μ (which is not relevant to our testing problem), an initial value X_1 was drawn at random from a normal distribution with expectation μ and a given standard deviation σ . Then for a given sequence of β -values $\beta_1, \beta_2, \dots, \beta_k$ (e.g. $\beta_i = -0.8, -0.6, \dots, 0.6, 0.8, 1.0$) we computed simultaneously, for a given n :

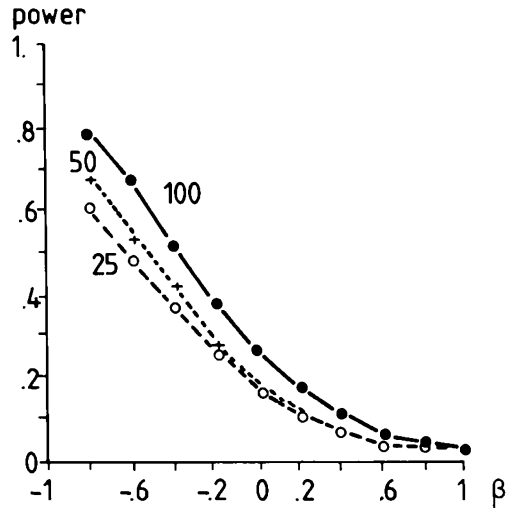


Fig. 1. Estimated power of the permutation test for tests using 25, 50, and 100 permutations respectively (10 generations). 500 Monte Carlo runs

Table 1. Estimated powers of Bulmer's first test for 10 generations. These estimates were based on 1000 Monte Carlo runs except the power for $\beta = 1$ (the null hypothesis), where 2500 runs were used to check the adequacy of the approximation formula for the critical value

value of β	Bulmer, 1975 1000 and 2500 runs	We, 1987 100 runs
0.50	0.19	0.21
0.75	0.10	0.08
0.90	0.06	0.06
1.00	0.044	0.04

$$X_{i,t+1} = \beta_i X_{i,t} + (1 - \beta_i)\mu + Z_t, \quad i = 1, 2, \dots, k$$

$$t = 1, 2, \dots, n - 1$$

where Z_t is a pseudo-random normal number with expectation zero and a given standard deviation σ ; in order to be able to compare the results without distortion by random estimation errors, the same Z_t was used for all i . The tests then were applied to all sequences $X_{i,1}, X_{i,2}, \dots, X_{i,n}$, using the 0.05 level of significance. This was repeated a large number of times, and it was counted, for all i , how often the null hypothesis $\beta = 1$ was rejected. The resulting relative frequencies are used as estimates of the power of the tests for $\beta_1, \beta_2, \dots, \beta_k$.

With the permutation test, a first question is: how many permutations to use? Fig. 1 shows estimated power functions for the permutation test using 25, 50, and 100 permutations. The power of the test clearly increases with increasing number of permutations. As the program is rather time-consuming, however, we left it at 100 permutations. For the same reason, we never used more than 500 Monte Carlo runs for estimating powers, and usually we used only 100 runs. We think this is warranted, because our aim is to compare powers, and not to estimate them accurately. Bulmer estimated the powers of his tests, for $\beta = 0.50, 0.75, 0.90$ using 1000 runs, and for $\beta = 1.00$ using 2500 runs. His values and ours (based on 100 runs), for the first test, are given in Table 1 for $n = 10$, and in Table 2 for $n = 25$. On

Table 2. Estimated powers of Bulmer's first test for 25 generations. See further legend of Table 1

value of β	Bulmer, 1975 1000 and 2500 runs	We, 1987 500 runs
0.50	0.74	0.66*
0.75	0.26	0.28
0.90	0.09	0.08
1.00	0.056	0.06

* For comparison between Bulmer's value and ours using a 2×2 table, $\chi^2 = 10.045$, 1 d.f.

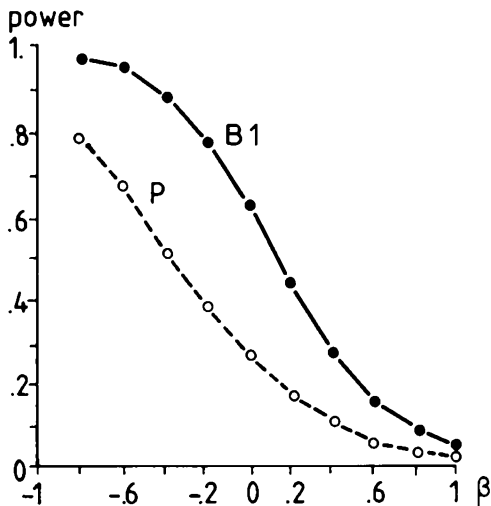


Fig. 2. Comparison of estimated powers of the permutation test (*P*) using 100 permutations, and Bulmer's first test (*B1*), for 10 generations. 500 Monte Carlo runs

the whole, the agreement is reasonable, but unfortunately, in one instance there is a significant difference between the two, see Table 2. We have no explanation for this, but feel no reason to be much alarmed either.

The estimated power functions of Bulmer's first test and the permutation test based on only 100 permutations, are compared in Fig. 2 ($n=10$), and in Fig. 3 ($n=25$). As was to be expected, Bulmer's test appears to be uniformly more powerful than the permutation test in the case that our preassumed model applies. It should be noted, however, that part of this difference between the two curves is due to the fact that the probability of rejecting the null hypothesis is somewhat smaller than 0.05 in the case of the permutation test, because the test statistic has a discrete distribution, and the requirement is that the probability of rejecting the null hypothesis be less than or equal to 0.05. In Fig. 3 all powers are overestimated due to random sampling errors.

With both Bulmer's first test and the permutation test, the powers increase with increasing n , the number of generations. This is shown in Fig. 4 for Bulmer's first test, and in Fig. 5 for the permutation test based on 100 permutations.

In order to study the power of Bulmer's second test, we used an approach different from that of Bulmer. Bulmer simply assumed that $\text{var}(E_i)$ has a constant value σ_a^2 , but in the comparison of estimated powers he then assumed this quantity to be zero. We have considered the case where

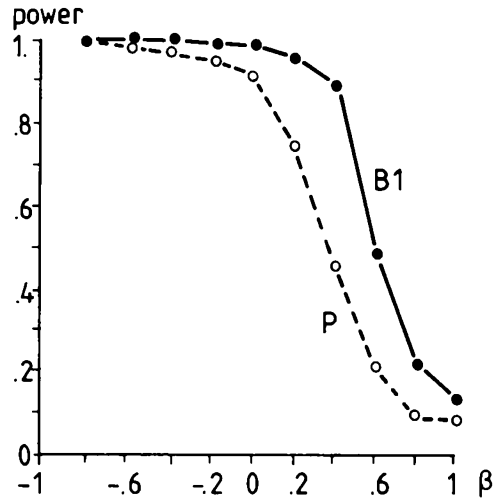


Fig. 3. Comparison of estimated powers of Bulmer's first test (*B1*) and the permutation test (*P*), for 25 generations and for estimated population densities with variances of estimates equal to expectations

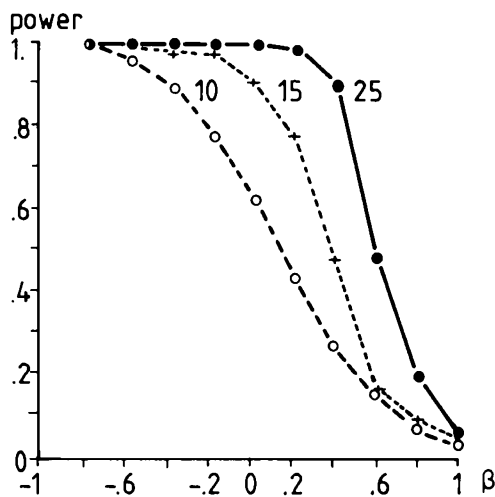


Fig. 4. Estimated power of Bulmer's first test for 10, 15, and 25 generations respectively. 100 Monte Carlo runs

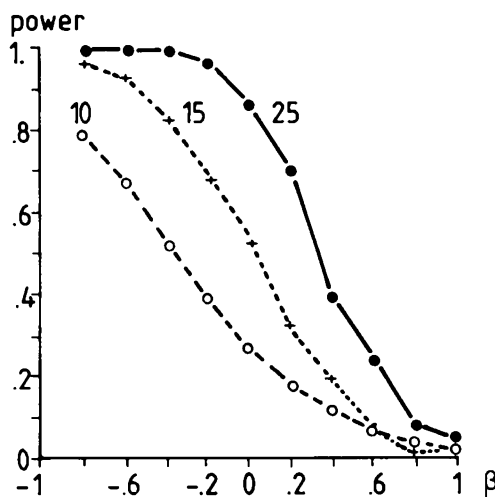


Fig. 5. Estimated power of the permutation test for 10, 15, and 25 generations respectively, using 100 permutations. 100 Monte Carlo runs

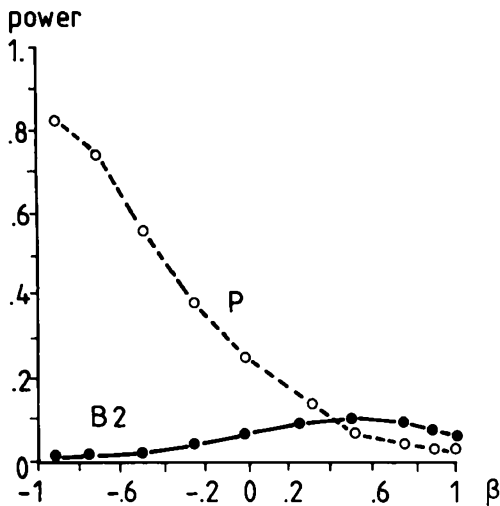


Fig. 6. Comparison of estimated powers of Bulmer's second test (B2) and the permutation test (P) using 100 permutations, for 10 generations, and for estimated population densities with variances of estimates equal to expectations

one has unbiased estimates N_t^* of N_t , so $E(N_t^* | N_t = m) = m$, and where the variance of N_t^* is equal to its expectation: $\text{var}(N_t^* | N_t = m) = m$. One would have this case if one were randomly sampling from a Poisson distribution. We have assumed that some law of large numbers would enable us to approximate the distribution of N_t^* for given N_t by the normal distribution, so we computed values of the N_t 's according to the assumed model (with, of course, $N_t = \exp(X_t)$), determined N_t^* 's by pseudo-random drawings from normal distributions with expectation N_t and variance N_t , and subjected the $X_t^* = \ln N_t^*$ to Bulmer's tests and to the permutation test, using, as before, the 0.05 level of significance. With 100 Monte Carlo runs, the null hypothesis was, in cases where $\beta = 1$, rejected 8 times with the permutation test, and 13 times with Bulmer's first test. With a significance level of 0.05 the expected number of rejections is, of course, 5. For both tests therefore, the number of rejections was too high, but the difference between 8 and 5 is not significant ($P > 0.12$) and the difference between 13 and 5 is ($P < 0.0003$). This illustrates a point that is already obvious for theoretical reasons (see Bulmer 1975): if population densities are estimated rather than measured, Bulmer's first test will reject the null hypothesis too often. With Bulmer's second test the null hypothesis was rejected 2 out of 100 cases, which does not differ significantly from 5 ($P > 0.12$). This agrees with Bulmer's own results; he estimated the probability of committing an error of the first kind for σ_d equal to zero times, to 0.5 times, and to 1.0 times σ , respectively, with $n = 10, 25, 50, 100$: at the 0.05 level, Bulmer's second test seems at the conservative side, but not significantly so.

Estimated powers of the permutation test and Bulmer's second test are compared in Fig. 6 ($n = 10$) and in Fig. 7 ($n = 25$). For values of β greater than 0.5 the difference between the two tests is small, but for smaller values of β the permutation test is much better. For negative values of β , Bulmer's second test appears to be worthless. This is doubtless the reason why Bulmer assumed negative values of β away, but of course, there are no biological reasons to do so. Moreover, the permutation test seems to be robust against the sampling errors in the X_t^* 's. We did another experiment with 500 Monte Carlo runs in order to get some-

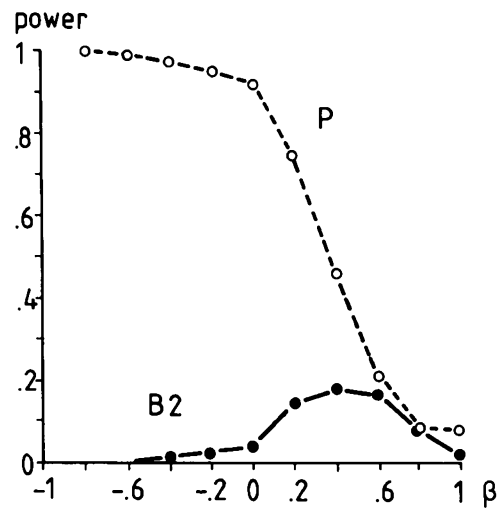


Fig. 7. Comparison of estimated powers of Bulmer's second test (B2) and the permutations test (P) using 100 permutations, for 25 generations, and for estimated population densities with variances of estimates equal to expectations

what more accurate results, using $n = 10$. The error of the first kind was committed 11 times, or 2.2% of the cases, with the permutation test, and 25 times, or 5.0% of the cases, with Bulmer's second test. Now, the number of rejections with the permutation test is significantly too small ($P = 0.003$), so it seems to be conservative rather than to overestimate power.

Discussion of the parametric tests

Bulmer (1975) pointed out that his first test is almost equivalent to using the first serial correlation coefficient of the X -es, r_1 , because his R is approximately $(1 - n^{-1})/[2(1 - r_1)]$, the difference being only in the weights given to the first and last observations. It is easily seen that therefore R is approximately an increasing function of r_1 . Now if we cannot observe X_t , but have to use $X_t^* = X_t + E_t$ instead, it turns out that r_1 and R will be decreased, hence the null hypothesis will be rejected too often. Concerning his second test, Bulmer (1975) states: (notation adapted to usage in the present paper) " R^* is nearly equivalent to $(n-2)(r_2 - r_1)/n$. The theoretical serial correlation with lag h is

$$\rho_h = [\sigma_d^2 / (\sigma_d^2 + \sigma^2)] \beta^h \quad h \geq 1$$

so that all the correlations are depressed by the same amount. It seems natural to base a test on a comparison between the first two serial correlation coefficients, and it is shown by Quenouille ([1957], Chapter 6) that such a test will be fully efficient in large samples".

This first statement: that all correlations are depressed by the same amount, is strictly true only if the variance of the estimation error, σ_d , is constant. It will therefore be approximately true if the variances of the estimates of population densities are proportional to these densities, as the logarithmic transformation is reported to take care of this type of heteroscedasticity (See e.g. Sokal and Rohlf 1969: 369-371). But is the difference between the first two serial correlation coefficients a "natural" statistic to use? From the theory given in Reddingius (1971) it may be derived that for a stationary sequence (X_t) , in fact

$$\begin{aligned}\text{cov}(X_t, X_{t+h}) &= \beta^h \sigma^2 / (1 - \beta^2) \\ \text{var}(X_t) &= \text{var}(X_{t+h}) = \sigma^2 / (1 - \beta^2)\end{aligned}$$

if $|\beta| < 1$. Then

$$\begin{aligned}\text{cov}(X_t^*, X_{t+h}^*) &= \text{cov}(X_t + E_t, X_{t+h} + E_{t+h}) \\ &= \text{cov}(X_t, X_{t+h})\end{aligned}$$

$$\begin{aligned}\text{var}(X_t^*) &= \sigma^2 / (1 - \beta^2) + \sigma_t^2, \\ \text{var}(X_{t+h}^*) &= \sigma^2 / (1 - \beta^2) + \sigma_{t+h}^2\end{aligned}$$

where we have written σ_t^2 for $\text{var}(E_t)$. The correlation coefficient with lag h between X_t^* and X_{t+h}^* is

$$\begin{aligned}\rho_h^* &= \frac{\beta^h \sigma^2 / (1 - \beta^2)}{\sqrt{[\sigma^2 / (1 - \beta^2) + \sigma_t^2][\sigma^2 / (1 - \beta^2) + \sigma_{t+h}^2]}} \\ &= \frac{\beta^h \sigma^2}{\sqrt{[\sigma^2 + \sigma_t^2(1 - \beta^2)][\sigma^2 + \sigma_{t+h}^2(1 - \beta^2)]}}\end{aligned}$$

If $\sigma_t^2 = \sigma_{t+h}^2 = \sigma_d^2$, this reduces to

$$\rho_h^* = \frac{\beta^h \sigma^2}{\sqrt{[\sigma^2 + (1 - \beta^2) \sigma_d^2]^2}} = \beta^h \frac{\sigma^2}{\sigma^2 + (1 - \beta^2) \sigma_d^2}$$

which differs from Bulmer's expression by a factor $(1 - \beta^2)$ in the denominator. We then have approximately, if $\sigma_t^2 \approx \sigma_{t+h}^2 \approx \sigma_d^2$:

$$\rho_2^* - \rho_1^* = (\beta^2 - \beta) \sigma^2 / [\sigma^2 + (1 - \beta^2) \sigma_d^2].$$

If $\beta = 1$, we have, writing $\text{var}(X_t) = \theta_t^2$:

$$\begin{aligned}\text{cov}(X_t, X_{t+h}) &= \text{var}(X_t) = \theta_t^2 \\ \text{var}(X_{t+h}) &= \text{var}(X_t) + h\sigma^2 = \theta_t^2 + h\sigma^2\end{aligned}$$

hence

$$\begin{aligned}\text{cov}(X_t^*, X_{t+h}^*) &= \theta_t^2 \\ \text{var}(X_t^*) &\approx \sigma_d^2 + \theta_t^2 \quad \text{and} \\ \text{var}(X_{t+h}^*) &\approx \sigma_d^2 + \theta_t^2 + h\sigma^2\end{aligned}$$

so that

$$\rho_h^* \approx \frac{\theta_t^2}{\sqrt{(\sigma_d^2 + \theta_t^2)(\sigma_d^2 + \theta_t^2 + h\sigma^2)}}$$

whence

$$\begin{aligned}\rho_2^* - \rho_1^* &= \frac{\theta_t^2}{\sqrt{(\sigma_d^2 + \theta_t^2)}} \\ &\cdot \left\{ \frac{1}{\sqrt{(\sigma_d^2 + \theta_t^2 + 2\sigma^2)}} - \frac{1}{\sqrt{(\sigma_d^2 + \theta_t^2 + \sigma^2)}} \right\}\end{aligned}$$

which is negative. Considering $\rho_2^* - \rho_1^*$ as a function of β , it will be seen that it decreases on the interval $-1 < \beta < 0.5$ and increases on the interval $0.5 \leq \beta < +1$, the graph of $\beta^2 - \beta$ being a parabola with zeroes at $\beta = 0$ and $\beta = 1$ and a minimum at $\beta = 0.5$. At $\beta = 1$, this function $\rho_2^* - \rho_1^*$ is discontinuous, because its value is negative, whereas $\lim_{\beta \rightarrow 1} (\rho_2^* - \rho_1^*) = 0$. It will be seen now why Bulmer considered only positive values of β : for negative values of β , his W statistic will not work. But even for positive values of β , the statistic does not seem "natural" to us at all. Its power for values of β close to zero is much smaller than its power in the neighbourhood of $\beta = 0.5$ (Figs. 6 and 7).

Rather than using Bulmer's second test, it seems to us

advisable to try to correct the test statistic for his first test, R , for estimation error. We may write

$$\begin{aligned}V^* &= \sum_{t=1}^n (X_t - \bar{X})^2 + \sum_{t=1}^n (E_t - \bar{E})^2 + 2 \sum_{t=1}^n (X_t - \bar{X})(E_t - \bar{E}) \\ U^* &= \sum_{t=1}^{n-1} (X_{t+1} - X_t)^2 + \sum_{t=1}^{n-1} (E_{t+1} - \bar{E})^2 + \sum_{t=1}^{n-1} (E_t - \bar{E})^2 \\ &\quad - 2 \sum_{t=1}^{n-1} (E_{t+1} - \bar{E})(E_t - \bar{E}) + 2 \sum_{t=1}^{n-1} (X_{t+1} - \bar{X})(E_{t+1} - \bar{E}) \\ &\quad + 2 \sum_{t=1}^{n-1} (X_{t+1} - \bar{X})(E_t - \bar{E}) - 2 \sum_{t=1}^{n-1} (X_t - \bar{X})(E_{t+1} - \bar{E}) \\ &\quad + 2 \sum_{t=1}^{n-1} (X_t - \bar{X})(E_t - \bar{E}).\end{aligned}$$

Considering $(E_t - \bar{E})^2$ as estimators of σ_d^2 , and all cross product terms as estimators of zero, we have, approximately,

$$R^{**} = V^*/U^* = (V + n\sigma_d^2)/[U + 2(n-1)\sigma_d^2]$$

and $R \approx (V^* - n\sigma_d^2)/[U^* - 2(n-1)\sigma_d^2]$.

It will be seen that for $n \geq 3$, R^{**} will tend to be smaller than R , as was pointed out by Bulmer (1975). Suppose now that the estimate of the population density at time t is based on a number of samples, and that each sample in itself would provide an estimate of the density. So we could have a number of estimates of X_t . The sample variance of these estimates will provide us with an estimate of σ_d^2 , which then could be used to compute an approximation of R , using the above formula. The corrected test statistic,

$$R^c = (V^* - nS_d^2)/[U^* - 2(n-1)S_d^2],$$

say, would include the stochastic term S_d^2 , and one should investigate whether the critical values given by Bulmer are still correct, i.e. whether the test based on R^c would have the proper level of significance. We will leave this to a mathematician more able than we are, and will in the next paper explore how much of a difference it will make in practice whether one uses R^c or R^{**} . (Compare the discussion in Reddingius (1971), Chapter 12 where, unfortunately, Table 12.1 contains calculation errors). Then, we will also apply this test as well as the permutation test to long series of published population data to get an impression of the evidence that actually supports the regulation hypothesis.

Some final remarks

A basic assumption underlying the tests considered is that the sequence of population densities as measured or estimated is a realization of a piece of first-order Markov chain. This assumption is not valid in cases where animals live longer than one generation, or where long-term trends in environmental factors such as weather or climate induce temporal dependencies in the data. In our opinion, a close connection of "stabilization" – as defined in the introduction – with the concepts of "regulation" and "density dependence" is valid only in a first-order Markov chain setting anyway. Where cyclicity or other long-term trends are considered, the idea of "density dependence" as it is usually defined is misleading. In such cases, as well as in cases where a population consists of animals in different age classes,

it is not at all clear which net reproduction is supposed to depend on which density, and there is a lot more to say about how a regulation process might work other than by some simple relationship between net reproduction and overall density. Note, that these criticisms also apply to the examples treated by Bulmer (1975).

The attempts of Slade (1977) and Vickery and Nudds (1984) to construct tests of density dependence suitable for dealing with trendy data therefore, in our opinion, are ill conceived if they are supposed to contribute to the questions whether or not stabilization or regulation occur (cf. Gaston and Lawton 1987). Pollard and Lakhani (1987) developed a similar kind of test, which is closely related, however, to the approach of the present authors. Next to being subjected to the above restrictions their test disconnects density dependence from regulation in the sense of "keeping density within limits" (introduction).

Also, the permutation test does not entirely escape from the above restrictions. Where long-term trends or cyclicity induce temporal dependencies in the data permutating the net reproduction values as found in the field is generally not realistic, because the time order of these values depends on processes that work across years. Nevertheless, such permutations can give us a first indication whether or not the series of census data under consideration is remarkably stable in the sense of question (1) of the introduction, without giving any hint about causes.

It seems to us useful to emphasize that no statistical test can show the existence of any biological phenomenon. Such a test can only tell us whether or not there is something in the data that is not easily explained away as being the result of pure chance. We got the impression that many ecologists expect a direct connection between some biological process, called "regulation", and a striking degree of stability of population numbers. Such a connection, however, is not as obvious as it appears on first thoughts, because a population number is the single result of many biological processes. Therefore, even if we have good reasons to suppose that one of these processes is responsible for a remarkable degree of stability in a sequence of population numbers, we have still to show (1), that this degree of stability indeed differs in a statistically significant degree from what might be expected in a random time series, and (2), that the process concerned is the most obvious cause of this stability. The latter means, that we should be able to compare in a sequence of population numbers the degree of stability with the biological process included to that with that process excluded. Den Boer (1986, 1987, 1988) tried to do this in simulation experiments in which all other processes are kept as these were found in the field population. As these other processes will not be completely independent of the supposed stabilizing process, real experiments in real populations might have been more conclusive, but hard to perform.

Appendix

The generation of random permutations and density ranges

Let $R(1), R(2), \dots, R(M)$ be the net reproduction values of a given field population in the natural order of successive generations, $L = 1, 2, \dots, M$. A random permutation of such a series of R -values can be obtained with following program section (Fortran 10):

```

DO 50 L = 1, M
  IF (L.EQ.M) GO TO 18
20  CALL RANDU (IX,IY,RNDGEN)
    KX = RNDGEN*(M + 1)
    IF (KX.LT.1) GO TO 20
    IF (NR(KX).EQ.0) GO TO 20
    RX(KX) = R(L)
    NR(KX) = 0
    GO TO 50
18  DO 50 JI = 1, M
    IF (NR(JI).NE.0) RX(JI) = R(M)
50  CONTINUE

```

in which RANDU generates pseudo-random numbers (RNDGEN), which are homogeneously distributed between 0 and 1. After being multiplied by $(M + 1)$ the next lower integer KX gives the new place of $R(L)$ in the permuted series according to $RX(KX) = R(L)$. A book-keeping of these new places is run in a parallel series of indexed integers (NR(..)) according to

```

K = 0
DO 110 KN = 1, M
  NR(KN) = K + 1
110 K = K + 1

```

in which the occupied places are marked "zero" with $NR(KX) = 0$ (above). When replacing the $R(L)$ -values the already occupied new places are passed by with $IF(NR(KX).EQ.0) GO TO 20$, after which another place is randomly chosen. The last value $R(M)$ gets the only open place left with $DO 50 JI = 1, M$ (above).

In a following program section the permuted series $RX(..)$ is used to create the series of densities belonging to it, starting from PS (\ln of the first density), after which the log-range $RLR = \ln(\text{highest density}) - \ln(\text{lowest density})$, RH , minus RL , of this series is established according to

```

PL = PS
RL = PL
RH = PL
DO 60 J = 1, M
  PL = PL + ALOG(RX(J))
  IF (PL.LT.RL) RL = PL
  IF (PL.GT.RH) RH = PL
60  CONTINUE
  RLR = RH - RL
  RANGE(N) = RLR

```

The above procedure is repeated 500 times, $N = 1, 2, \dots, 500$, after which the log-range values, $RANGE(..)$, of these 500 permutations are compared with the value FLR of the field series according to

```

LLR = 0
DO 70 N = 1, 500
  IF (RANGE(N).LE.FLR) LLR = LLR + 1
70  CONTINUE
  CLR = LLR
  CL = 500 + 1
  PLR = CLR/CL

```

in which PLR represents the chance that permuted series of the net reproduction values (RX) will give equal or smaller ranges between the highest and lowest density ($RANGE$) than the field series (FLR from the natural order

of R's). On the DEC-1091 of the Agricultural University Wageningen for series of 18–22 R-values this program (500 permutations) asked a CPU time of only about 2 s.

References

- Bulmer MG (1975) The statistical analysis of density dependence. *Biometrics* 31:901–911
- Den Boer PJ (1968) Spreading of risk and the stabilization of animal numbers. *Acta Biotheor* (Leiden) 18:165–194
- Den Boer PJ (1971) Stabilization of animal numbers and the heterogeneity of the environment: The problem of the persistence of sparse populations. In: Den Boer PJ, Gradwell GR (eds) *Dynamics of populations*. Pudoc, Wageningen pp 77–97
- Den Boer PJ (1981) On the survival of populations in a heterogeneous and variable environment. *Oecologia* 50:39–53
- Den Boer PJ (1986) Density dependence and the stabilization of animal numbers. 1. The winter moth. *Oecologia* 69:507–512
- Den Boer PJ (1987) Density dependence and the stabilization of animal numbers. 2. The pine looper. *Neth J Zool* 37:220–237
- Den Boer PJ (1988) Density dependence and the stabilization of animal numbers. 3. The winter moth reconsidered. *Oecologia* 75:161–168
- Gaston KJ, Lawton JH (1987) A test of statistical techniques for detecting density dependence in sequential censuses of animal populations. *Oecologia* 74:404–410
- Howard LO, Fiske WF (1911) The importation into the United States of the parasites of the gipsy moth and the brown-tail moth. *Bull US Bur Entom* 91:1–344
- Klomp H (1966) The dynamics of a field population of the pine looper *Bupalus piniarius* L (Lep. Geom) *Adv ecol Res* 3:207–305
- Kolmogorov AN (1936) Sulla teoria di Volterra della lotta per l'estistenza. English translation. In: Scudo FM, Ziegler JR (eds) *The golden age of theoretical ecology: 1923–1940*. Springer, Berlin Heidelberg New York pp 287–295
- May RM (1973) *Stability and complexity in model ecosystems*. Princeton Univ Press, Princeton
- Murdoch WW (1979) Predation and dynamics of prey populations. In: Halbach U, Jacobs J (eds) *Population ecology*. G Fischer, Stuttgart pp 295–310
- Pielou EC (1974) *Population and community ecology. Principles and methods*. Gordon & Breach, New York Paris London
- Pollard E, Lakhani KH (1987) The detection of density dependence from a series of annual censuses. *Ecology* 68:2046–2055
- Quenouille MH (1957) *The analysis of multiple time series*. Griffin, London
- Reddingius J (1971) Gambling for existence. A discussion of some theoretical problems in animal population ecology. *Acta Biotheor* 20 [Suppl] 1–208
- Reddingius J, den Boer PJ (1970) Simulation experiments illustrating stabilization of numbers by spreading of risk. *Oecologia* 5:240–284
- Slade NA (1977) Statistical detection of density dependence from a series of sequential censuses. *Ecology* 58:1094–1102
- Sokal RR, Rohlf FJ (1969) *Biometry. The principles and practice of statistics in biological research*. Freeman, San Francisco
- Varley GC, Gradwell GR (1968) Population models for the winter moth. In: Southwood TRE (ed) *Insect abundance*. Blackwell, Oxford Edinburgh pp 132–142
- Varley GC, Gradwell GR, Hassell MP (1973) *Insect population ecology, an analytical approach*. Blackwell, Oxford London
- Vickery WL, Nudds TD (1984) Detection of density-dependent effects in annual duck censuses. *Ecology* 65:96–104

Received June 10, 1988