Ecology (1975) **56**: pp. 244–247

ESTIMATING TIME-VARYING SURVIVAL OF ARTHROPOD LIFE STAGES FROM POPULATION DENSITY[1]

WILLIAM G. RUESINK
Illinois Natural History Survey, Urbana, Illinois 61801

Abstract. A new technique for calculating survival rates decomposes the density versus time curve for each life stage to obtain the rate individuals are entering and the rate they are leaving that stage. The survival rate is then estimated by a comparison of the rate they leave one stage with the rate they enter the next. The analysis of a hypothetical set of data indicates that the technique can handle relatively imprecise data when at least seven nonzero population estimates are available for each life stage per generation.

Key words: Arthropoda; insecta; mortality analysis; population dynamics; survival analysis.

INTRODUCTION

Many people are now developing mathematical models for the population dynamics of particular insect species. These models generally subdivide the population into components equivalent to life stages or aggregates of life stages. Each component in turn includes a submodel for survival rate, which in many cases appears as a simple constant for lack of better information.

One of the principles of modeling is to include as part of the model only those phenomena needed to capture the essence of the real system. Thus, if survival is constant for all available data, then our model should not consider the factors responsible for mortality. When survival varies with time in a recognizable pattern and the magnitude is predictable as a function of time, then it is best to model it simply as a function of time. To identify and include in the model the causes of mortality would violate the principle of simplicity. Reaching the best decision requires that the modeler have adequate information on survival in the natural ecosystem.

Most of the methods available to field ecologists for estimating age-specific survival of arthropod life stages from population density are not capable of detecting changes in survival with time. R. O. Barr (*pers. comm.*) presents a method for calculating time-varying survival from population densities if, among other things, the instantaneous birth rate is known. But when the birth rate is unknown, none of the existing techniques can be used. The purpose of this paper is to (1) present the theory of a new method for estimating time-varying survival rates, (2) explain how to use the method, and (3) investigate its sensitivity to sampling error via the analysis of a hypothetical population. Applications of the technique to certain insect species will be described in the entomological literature.

[1] Manuscript received October 18, 1973; accepted June 5, 1974.

MATHEMATICAL FORMULATION

This method assumes that an individual is much more likely to die during or near the time of molt than at any other time. This condition is represented by the solid line in Fig. 1; a close approximation is the step function indicated by the broken line. The mathematical derivations of the equations in this paper are based on the situation represented by the step function.

The essence of the method is to decompose the density vs. time curve for each life stage into two parts: the rate individuals are entering and the rate they are leaving. This is easily done if one knows the amount of time that an individual spends in each stage. Since all mortality is assumed to occur between stages, the survival between two successive stages can be determined by a comparison of the rate that individuals enter a given stage with the rate they leave the preceding stage. Although conventional time may be used, in most cases it is better to use physiological time (i.e., degree-days above the developmental threshold).

Let $N_j(t)$ = Population density of stage j at time t,
T_j = Developmental time of stage j,
$Q_j(t)$ = Rate individuals are entering stage j at time t,
$R_j(t)$ = Rate individuals are leaving stage j at time t,
$C_j(t)$ = Total number of individuals that have entered stage j up to time t,
$D_j(t)$ = Total number of individuals that have left stage j up to time t, and
S_{jj+1} = Survival between stages j and $j+1$.

Given $N_j(t)$ and T_j for a complete generation, then $C_j(t)$ and $D_j(t)$ can be computed from the equations:

$$C_j(t) = N_j(t) + C_j(t - T_j) \tag{1}$$

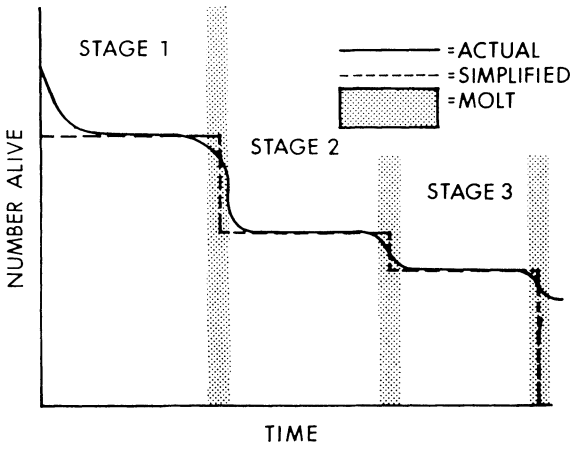

FIG. 1. The solid line (——) = the actual number surviving from an initially uniform-aged cohort; broken lines (– – –) = the simplified condition used to derive the equations.

$$D_j(t) = C_j(t - T_j) \tag{2}$$

where

$C_j(t) = 0$ for t earlier than the first nonzero $N_j(t)$. And the survival between stages j and $j+1$ is found from

$$S_{jj+1}(t) = \frac{Q_{j+1}(t)}{R_j(t)} = \frac{d[C_{j+1}(t)]/dt}{d[D_j(t)]/dt}. \tag{3}$$

COMPUTATIONAL PROCEDURE

These equations cannot easily be used to compute survival rates from field data. One reason is that N_j is available, not as a continuous function, but only as a series of estimates (with sampling errors) of N_j at discrete points in time. Even if a curve is drawn through the sampling points, it is difficult to write a concise mathematical expression for that curve. In other words, one cannot easily obtain an equation for N_j and hence cannot differentiate to obtain Q_j and R_j.

One way to handle the problem is by using graphical methods. This involves drawing a continuous line through the data points, either with straight line segments or with a smooth curve. From this continuous line one can construct $C_j(t)$ using equation (1), and then $D_j(t)$ by displacing $C_j(t)$ to the right a distance equal to the duration of the life stage. When these curves are constructed for consecutive life stages, the survival rate at any point in time can be found from the ratio of the slopes of the appropriate curves (equation (3)). In practice one may find that inaccuracies induced by sampling errors are sufficient to cause $C_j(t)$ and $D_j(t)$ to fluctuate rather than increase monotonically. In this case they should be smoothed by some curve-fitting technique prior to computing the derivatives; otherwise the computed survival will behave very erratically.

For persons who have access to a computer, the tedious work of drawing these curves can be avoided. A FORTRAN program is available which performs all of the necessary calculations, including the interpolation between sample points and the curve smoothing, and prints the time-varying survivals for up to 10 life stages (a listing of the program can be obtained from the author). For comparison, the printout includes an estimate of the mean generation survival as calculated by the area under the curve technique of Southwood (1966). Since digital computers operate much better with difference equations than with differential equations, the program uses

$$S_{jj+1}(t) = \frac{C_{j+1}(t+n) - C_{j+1}(t-n)}{D_j(t+n) - D_j(t-n)} \tag{4}$$

in place of equation (3). When the estimated densities contain large sampling errors, n must be made relatively large to obtain meaningful survival estimates. Increasing n has the effect of smoothing the $S_{jj+1}(t)$ curve.

EXAMPLE USING HYPOTHETICAL DATA

Consider a hypothetical case where $T_1 = 10$, $T_2 = 8$ and $Q_1(t)$ and $S_{12}(t)$ are as shown in Fig.

TABLE 1. Some parameter values for the hypothetical population derived from the Q_1 and S_{12} curves given in Fig. 2

t	N_1	C_1	D_1	R_1	S_{12}	Q_2	C_2	D_2	N_2
10	1	1	0	.00	.200	.00	.0	.0	.0
20	20	21	1	.75	.200	.15	.2	.0	.2
30	44	65	21	3.25	.200	.65	4.2	.6	3.6
40	50	115	65	5.00	.200	1.00	13.0	5.6	7.4
50	50	165	115	5.00	.240	1.20	23.3	15.0	8.3
60	50	215	165	5.00	.373	1.87	38.6	25.8	12.8
70	42	257	215	4.70	.507	2.38	60.4	42.5	17.9
80	32	289	257	3.70	.640	2.37	84.4	65.2	19.2
90	22	311	289	2.70	.773	2.09	106.9	89.1	17.8
100	12	323	311	1.70	.907	1.54	125.3	111.0	14.3
110	2	325	323	.70	1.000	.70	136.8	128.2	8.6
120	0	325	325	.00	1.000	.00	139.2	138.0	1.2
130	0	325	325	.00	1.000	.00	139.2	139.2	.0

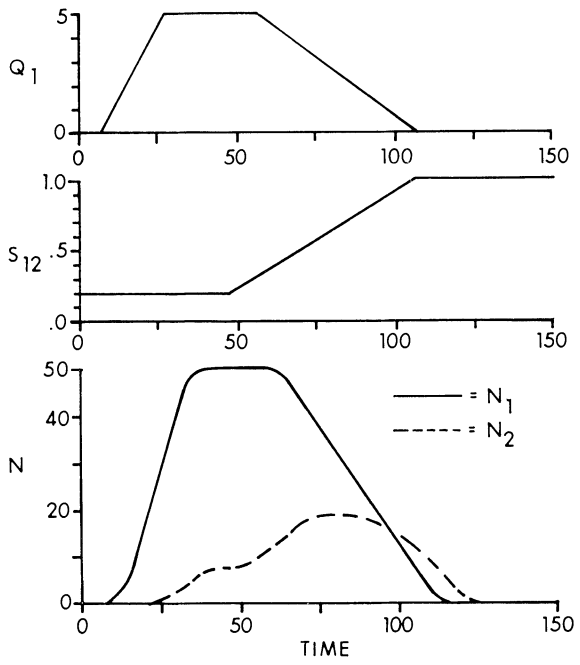

FIG. 2. Hypothetical population data used to evaluate the method. The population densities (N_1 and N_2) of stages 1 and 2 were computed from (1) the assumed input rate (Q_1) into stage 1, and (2) the assumed survival (S_{12}) at the instant of molt from stage 1 to stage 2.

2 and Table 1. These conditions, plus the assumption that all mortality occurs at molt, were used to generate the values for $N_1(t)$ and $N_2(t)$ (Fig. 2 and Table 1). Hence if N_1 and N_2 are as shown, and if $T_1 = 10$ and $T_2 = 8$, then $S_{12}(t)$ must be as it is shown. The generation survival (i.e., the number of individuals reaching stage 2 over the number reaching stage 1 during the entire generation) under these conditions is $139.2/325 = 0.428$.

Applying equations (1–3) to the mathematical description of $N_1(t)$ and $N_2(t)$ does, in fact, faithfully reproduce the $S_{12}(t)$ curve given in Fig. 2. Hence, we must determine how the accuracy of the estimate of $S_{12}(t)$ is affected by the nature of the available estimates of $N_1(t)$ and $N_2(t)$.

First, consider the consequence of infrequent sampling. Assume that on each sampling date the values of N_1 and N_2 are known exactly (i.e., no sampling error). Table 2(b) presents the calculated percent survival on selected days for a 2-, 5-, 10-, and 20-day sampling interval when n in equation (4) is set at 2.5. Only the 20-day interval shows a significant departure of the calculated value from the true value.

Next, consider the consequence of sampling error. Table 2(c) shows the calculated percent survival on a 5-day sample interval with 5%, 10%, 20%, and 40% sampling error. Ten percent error means that the number of samples taken on a given date is such that the standard error is 10% of the mean. Here, only the 40% error resulted in significant departures from the true value.

TABLE 2. Comparison of the actual percent survival to the estimated value for selected sampling conditions

Sampling interval (days)	Percent error	Mean generation survival[a]	Computed survival on day 20	40	60	80	100
(a)	Exact hypothetical values						
0	0	43	20	20	37	64	91
(b)	With no sampling error						
2	0	43	20	20	36	63	89
5	0	43	20	20	36	62	87
10	0	43	20	21	36	61	84
20	0	42	—	27	38	55	71
(c)	When the sample interval is 5 days						
5	5	43	20	20	35	65	93
5	10	42	20	21	34	67	99
5	20	42	21	21	32	73	117
5	40	41	21	22	28	87	187
(d)	When the sample interval is 20 days						
20	5	39	—	25	34	50	66
20	10	36	—	23	31	46	61
20	20	31	—	21	25	31	53
20	40	21	—	16	15	22	37

[a] Computed by Southwood's area-under-the-curve technique.

Table 2(d) presents the results of combining the worst cases: a 20-day interval with a sampling error of up to 40%. As expected, the results deviate considerably from the true value (Table 2(a)). Two points should be made, however: (1) the mean generation survival is also poorly estimated; and (2) in spite of the apparent inadequacy of a 40% sampling error and a 20-day sample interval, the true nature of the generation survival was accurately revealed: it is lowest early in the season and increases considerably later on.

Finally, consider the consequence of calculating

TABLE 3. Consequence of errors in estimation of the T_j's with a 5-day sampling interval and a 5% sampling error

T_1	T_2	Mean generation survival[a]	Computed survival on day 20	40	60	80	100
9	7.2	43	17	20	32	57	98
9	8.0	38	16	18	29	52	81
9	8.8	35	16	17	27	47	81
10	7.2	47	21	22	38	71	113
10	8.0	43	20	20	35	65	93
10	8.8	39	20	19	32	58	92
11	7.2	52	26	25	40	74	111
11	8.0	47	26	22	37	68	92
11	8.8	43	26	20	34	61	91

[a] Computed by Southwood's area-under-the-curve technique.

survival rates when the estimated developmental times are in error. Table 3 shows the results of a $\pm 10\%$ error in estimating T_1 and T_2, and usually developmental times are known more accurately than that. Mean generation survival is overestimated in proportion to the overestimate of T_1; it is underestimated in proportion to the overestimate of T_2.

Daily survival estimates respond in a more complex manner. Very early in the generation daily survival is overestimated in proportion to T_1 and is almost independent of T_2. By midgeneration the response follows the same pattern as for mean generation survival. Late in the generation daily survival is underestimated in proportion to the overestimate of T_2 and independent of T_1, except that when T_1 is underestimated, then daily survival is underestimated proportionally.

Discussion

In many situations the assumption that mortality occurs mostly at and near molts comes close to reality. Certainly an individual is more susceptible to predation and the hazards of inclement weather at that time. Examples of species which should meet this condition are (1) all immature stages of the cereal leaf beetle, *Oulema melanopus* (L.); (2) larval and pupal stages of the alfalfa weevil, *Hypera postica* (Gyllenhal); and, (3) larval and pupal stages of the northern corn rootworm, *Diabrotica longicornis* (Say). On the other hand, some life stages are of very long duration, e.g., the egg of the northern corn rootworm and, in the southern United States, the egg of the alfalfa weevil. These would usually encounter significant mortality during the central portion of the stadium. Hence, the technique presented in this paper should be applied judiciously and only to populations that can be expected to have a low mortality during the central portion of consecutive life stages. In general, the condition will be satisfied for any non-diapausing immature stage in an environment favorable for its rapid development.

The example just analyzed assumed a constant temperature and therefore a constant number of degree-days per day. For laboratory populations this is normally the actual case, but for field populations the only reasonable approach is to ignore days entirely and to base everything on degree-days above the developmental threshold. This is particularly important when the day-to-day temperatures vary considerably and not all life stages have the same developmental threshold.

Sampling intervals in Table 2 are 2, 5, 10, and 20 days. In this example the duration of the first stage was 10 days and that of the second stage was 8 days. So the 2- and 5-day intervals represent sampling more often than once per stage duration, whereas the 20-day interval represents sampling less often.

A more important consideration is how many data points are obtained during each life stage per generation: the more data points, the better. Figure 2 shows that the 2-day interval results in about 50 data points, but the 20-day interval results in only 5. The results presented in Table 2 plus other analyses not presented here suggest that 7 data points evenly spaced in the nonzero density area approaches the minimum desirable situation; as the sampling frequency increases, so does the allowable sampling error of the density estimates. A rough rule of thumb is that the sampling error (as a percent of the mean) should not exceed twice the number of non-zero data points. For example, if there will be 10 sampling dates on which the density is nonzero, then on each date enough samples should be taken to insure that the standard error is within 20% of the mean.

Acknowledgments

I thank my colleagues P. W. Price, R. G. Helgesen, and M. Kogan, who suggested useful changes in the manuscript. This work was supported in part by NSF Grant No. GB-34718 "The Principles, Strategies, and Tactics of Pest Population Regulation in Major Crop Ecosystems."

Literature Cited

Southwood, T. R. E. 1966. Ecological methods: With particular reference to the study of insect populations. Methuen, London. 391 p.