INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONE: 9240-1
CABLE: CENTRATOM - TELEX 460392 - I

SMR.301/26

FIRST AUTUMN WORKSHOP ON MATHEMATICAL ECOLOGY

(31 October - 18 November 1988)

THE RELATION BETWEEN THE NUMBER OF PARASITES/HOST AND HOST AGE:

POPULATION DYNAMIC CAUSES AND MAXIMUM LIKELIHOOD ESTIMATION

A.P. DOBSON

University of Rochester

Rochester, USA

These are preliminary lecture notes, intended only for distribution to participants.

Missing or extra copies are available from the Workshop Secretariat.

Parasitology (1988), **96**, 197-210
With 5 figures in the text

The relation between the number of parasites/host and host age: population dynamic causes and maximum likelihood estimation

S. W. PACALA[1] *and* A. P. DOBSON[2]

[1] *Department of Ecology and Evolutionary Biology, The University of Connecticut, U-42 Storrs, Connecticut* 06268

[2] *Department of Biology, Princeton University, Princeton, New Jersey* 08544

(*Accepted* 7 *July* 1987)

SUMMARY

We examined dynamical factors that shape the distribution of the number of parasites/host in constant or temporally varying environments, and with or without host-age dependent variation in host susceptibility and parasite mortality. We predict properties of the parasite distribution in the absence of density-dependent factors such as density-dependent mortality or recruitment and parasite-induced host mortality. These properties provide a criterion for the detection of density dependence in temporally variable systems with host-age dependent interactions. We have then introduced methods to estimate and statistically evaluate the effects of host age or size on the distribution of parasites/host. The methods are based on a maximum likelihood protocol for linear and non-linear regression when data are negatively binomially distributed. We have illustrated the use of the theoretical results and statistical methods by re-analysing the data of Halvorsen & Andersen (1984) on cestode infections in Norwegian arctic charr and by analysing new data on nematode infections in Caribbean *Anolis* lizards.

INTRODUCTION

Many recent studies in the parasitological literature have focused on the causes and population dynamic consequences of the pattern of parasite distribution among hosts. One productive approach has been to assume that the number of parasites/host follows some probability distribution such as the negative binomial, and then to determine with mathematical models the population dynamic consequences of the distribution (May, 1977; Anderson, 1978, 1982; Anderson & May, 1978, 1979, 1985; May & Anderson, 1978, 1979; Dobson, 1985). This body of work has shown convincingly that the commonly observed aggregated distributions (variance to mean ratio greater than 1) enhance density-dependent population regulation of host–parasite communities (*cf.* Anderson, 1982) and reduce the level of interspecific competition among parasites (Dobson, 1985).

A second approach, pioneered by Crofton (1971*a*, *b*) and more recently by Anderson & Gordon (1982), has been to model directly the factors that cause the pattern of parasite dispersion within a host population. Anderson & Gordon (1982) modelled parasite invasion–mortality processes in cohorts of aging hosts. They treated analytically the case in which the parasite invasion rate (the probability that a given host is invaded by an additional parasite/unit time) and the parasite mortality rate are independent of host age or time. If hosts differ in their susceptibilities to infection, due for example to immunological, behavioural or micro-habitat differences among hosts, then (a) parasite distributions are aggregated; (b) both the mean number of parasites/host and the variance to mean ratio increase monotonically with host age. Anderson &

Gordon (1982) assumed that host susceptibility to infection follows a Poisson distribution. In this paper we refer to the relation between the mean number of parasites/host and host age as the age–intensity relation.

Anderson & Gordon (1982) also performed Monte Carlo simulations of cases with density-dependent parasite mortality and parasite-induced host mortality. Density-dependent parasite mortality reduces the variance to mean ratio so that it may be either a peaked or decreasing function of host age. Age–intensity relations, however, remain monotonically increasing. With parasite-induced host mortality, both the mean and the variance to mean ratio may be peaked functions of host age. These quantities may each increase to a maximum as age increases and then subsequently decrease. See Dietz (1982) for a special case in which the problem with parasite-induced host mortality is analytically tractable. Anderson & Gordon (1982) concluded that the shapes of the age–intensity relation and the relation between the variance to mean ratio and host age are indicative of the kind of density dependence present. For example, a peaked age–intensity relation is evidence of parasite-induced host mortality. They pointed out, however, that non-monotonically increasing relations could also be caused by time or host-age dependent changes in host susceptibility, the abundance of infective stages or parasite mortality.

Empirical studies have repeatedly documented host-age dependent changes in the distribution of the number of parasites/host (see, for example, Gordon & Rau (1982), Halvorsen & Andersen (1984), Kennedy (1984) or the many human examples reviewed by Anderson & May (1985), and animal examples reviewed by Anderson & Gordon (1982). One important and widely recognized difficulty in using theoretical results to draw inferences from these data is that appropriate statistical procedures have been lacking (see discussion by Anderson (1982) and Anderson & Gordon (1982)). For example, to estimate the age–intensity relation, hosts are typically grouped into age classes and a mean number of parasites is calculated for each age class. A function of host age is then fit to these data by eye or by least-squares (Halvorsen & Andersen, 1984; Lester, 1984; Anderson & May, 1985). There are several obvious problems with this protocol. First, in regressing the mean for each age class versus age, one effectively reduces the sample size from the number of hosts to the number of age classes. Second, means estimated from samples of different sizes are weighted equally. One, of course, has more statistical confidence in a mean estimated from a large sample than in a mean estimated from a small sample. Not uncommonly, samples for the oldest age classes of hosts are small, and it is precisely these means that determine if the age–intensity relation is peaked. A third related problem is that the sampling distribution of the means may change form as a function of host age and sample size. Moreover, if the distribution of parasites/host is aggregated, then the sampling distribution of the mean will be commonly asymmetric. Thus, deviations of the sample mean below the population mean may be more likely than deviations above the population mean.

This paper is divided into two sections. We first generalize the analytical treatment of Anderson & Gordon (1982) as follows: (a) parasite invasion and mortality rates are arbitrary functions of time and host age rather than constant; (b) the probability density governing inherent differences among hosts is arbitrary rather than Poisson. We derive a criterion for detecting the effects of density dependence from the age–intensity relation and the relation between age and the variance to mean ratio. Here, by density dependence we mean density-dependent parasite mortality, parasite-induced host mortality or parasite-induced changes in a host's susceptibility to invasion by additional parasites. This criterion is unaffected by the age and time dependencies

built into the model and so could be used to detect density dependence in temporally varying systems with age-dependent epidemiological parameters.

We then describe a method to estimate and statistically evaluate the effects of host age or size on the distribution of parasites/host. The negative binomial distribution has proven to be an excellent empirical descriptor of the number of parasites/host (see Anderson & May, 1985). This distribution has two parameters: M, the mean, and k, a parameter that governs the degree of aggregation. Aggregation decreases as k increases; as k approaches infinity, the distribution approaches a Poisson distribution. Now suppose that M and k are functions of host age or size; we describe a maximum likelihood estimator for the parameters of the M and k functions. We have implemented an iterative algorithm that solves the likelihood equations for up to three-parameter M functions and two-parameter k functions, and produces boot-strap confidence intervals for the parameter estimates. In intuitive terms, our maximum likelihood protocol is a method for linear or non-linear regression when the data are negative binomially distributed with constant or variable k. In contrast, standard regression methods assume that residuals are normally distributed with constant variance.

O. Halvorsen has been kind enough to send us his and K. Andersen's data on cestode (*Diphyllobothrium ditremum*) infections in Norwegian arctic char (*Salvinius alpinus*) (Halvorsen & Andersen, 1984). To illustrate the maximum likelihood approach we re-analyse these data and show that the apparent peak in the char age–intensity relation is likely to be a sampling artifact. Nonetheless, k increases significantly with age. This, in light of our theoretical results, implies the presence of some form of density dependence. We also analyse new data on nematode (*Thelondros cubensis*) infections in Caribbean lizards (*Anolis bimaculatus*), to illustrate a case in which the hosts are classified by size rather than age.

THE MODEL

Let $q_n(t,a)$ be the probability that a host of age a at time t has n parasites. Also, let $\Delta a\lambda(t,a)$ be the probability that a host of age a is infected by an additional parasite during the time interval from t to $t+\Delta a$ (and hence between age a and $a+\Delta a$). Finally, suppose that $\Delta an\mu(t,a)$ is the probability that a death occurs in the population of n parasites infecting a host of age a within the time interval from t to $t+\Delta a$. Thus:

$$q_n(t+\Delta a, a+\Delta) = \Delta a\lambda(t,a)\,q_{n-1}(t,a)+\Delta a(n+1)\,q_{n+1}(t,a)\,\mu(t,a) + q_n(t,a)\,[1-\Delta a\lambda(t,a)-\Delta a\mu(t,a)].$$

Subtracting $q_n(t,a)$ from both sides, dividing both sides by Δa, and passing to the limit as $\Delta a \to 0$ yields

$$\frac{\partial q_n(t,a)}{\partial t}+\frac{\partial q_n(t,a)}{\partial a} = \lambda(t,a)\,q_{n-1}(t,a)+(n+1)\,\mu(t,a)\,q_{n+1}(t,a) - \lambda(t,a)\,q_n(t,a)-n\mu(t,a)\,q_n(t,a). \quad (1)$$

A solution of equation (1) is $$q_n(t,a) = e^{-\phi}\phi^n/n! \quad (2)$$

where $$\phi = \int_0^a \lambda(t-a+x,x)\,e^{-\int_x^a \mu(t-a+y,\,y)\,dy}\,dx.$$

Thus, the probability that there are n parasites in a host of age a at time t is given by a Poisson distribution with mean ϕ. Of course, equation (2) may be shown to be a

solution of equation (1) by direct substitution. In the special case of constant μ and λ studied by Anderson & Gordon (1982), ϕ reduces to $\lambda/\mu(1-e^{-\mu a})$.

Now suppose that the invasion rate $\lambda(t,a)$ is the product of a non-negative random variable Z and a non-negative function of host age and time $c(t,a)$. The random variable Z accounts for differences among hosts in their susceptibilities to infection. Each host is assigned a value of Z at birth and retains this value throughout its life. A submodel of infection that leads to the relation $\lambda(t,a) = Zc(t,a)$ follows: suppose that $\Delta a\Theta\omega$ is the probability that a host encounters an infective stage during the small time interval t to $t+\Delta a$, where Θ is the number of infective stages in the host's home range and $\Delta a\omega$ is the probability that the host encounters any one infective stage. Further, let ϵ be the probability that the host is infected given an encounter. Then: $\lambda(t,a) = \Theta\omega\epsilon$. There are at least three qualitatively different possible relations between $Zc(t,a)$ and $\Theta\omega\epsilon$. (a) Suppose that the mean density and spatial distribution of infective stages is constant in time, but that the spatial distribution of infective stages is patchy (some host home ranges contain more infective stages than others). If ϵ and ω vary temporally and with host age but do not vary among hosts of the same cohort, then $Z = \Theta$ and $\epsilon\omega = c(a,t)$. (b) Suppose that host behaviour varies, such that ω is independent of age and time but differs among hosts. This could occur, for example, due to between-host variation in prey preference. If Θ and ϵ vary temporally and with host age but do not vary among hosts of the same cohort then $Z = \omega$ and $\Theta\epsilon = c(a,t)$. (c) Suppose that some hosts are better able to resist infection, given an encounter, than others, but that ϵ is independent of age and time. This could occur if nutritional differences among hosts affected their ability to mount an immune response. If Θ and ω vary temporally and with host age but do not vary among hosts of the same cohort then $Z = \epsilon$ and $\omega\Theta = c(a,t)$.

We emphasize that the above interpretations do not represent an exhaustive list. Rather they are but three of many that will lead to the relation $\lambda(t,a) = Zc(t,a)$. With the invasion rate expressed as the product $Zc(a,t)$, the solution to equation (2) becomes:

$$q_n(t,a) = e^{-Zg(t,a)}[Zg(t,a)]^n/n!;\ g(t,a) = \int_0^a c(t-a+x,x)\, e^{-\int_x^a \mu(t-a+y,y)\,dy}\,dx. \qquad (3)$$

The probability that a randomly chosen host of age a at time t will have n parasites is

$$\int_0^\infty b(z)\, q_n(t,a \mid Z=z)\,dz, \qquad (4)$$

where $b(z)$ is the probability density of Z. If Z is a discrete random variable, then the integral in equation (4) is replaced by a discrete sum. In the case considered by Anderson & Gordon (1982), $b(z)$ is Poisson. Let $\pi(r)$ be the moment generating function of $b(z)$. The probability generating function of $q_n(t,a \mid Z=z)$ is $\nu(s)\, e^{zg(t,a)(s-1)}$, and so the probability generating function of equation (4) is $\pi(g(t,a)(s-1))$. We obtain the mean, $M(t,a)$ and variance of equation (4) directly from $\pi(g(t,a)(s-1))$ as

$$M(t,a) = \bar{B}g(t,a)$$

$$\text{Variance} = \bar{B}g(t,a) + \sigma^2 g(t,a)^2, \qquad (5)$$

where $\bar{B}$ and σ^2 are, respectively, the mean and variance of Z. Thus, the age–intensity relation at time t is given by equation (5) and the variance to mean ratio as a function of age and time ($VM(t,a)$) is

$$VM(t,a) = 1 + (\sigma^2/\bar{B})g(t,a). \qquad (6)$$

Together equations (5) and (6) provide the following result (Result 1). In the absence of density-dependent parasite mortality, parasite-induced host mortality and parasite-induced changes in a host's probability of further infection, the age–intensity relation is given by equation (5) and the variance to mean ratio is given by equation (6). Equations (5) and (6) are valid in constant or temporally variable environments, with or without host–age dependent changes in susceptibility or parasite mortality. Note that parasite distributions are always aggregated if the host population is heterogeneous ($VM(t,a) > 1$ if $\sigma^2 > 0$) and that *the degree of aggregation of the parasite distribution, $VM(t,a)$, is a linear increasing function of the mean number of parasites/host* (Fig. 1).

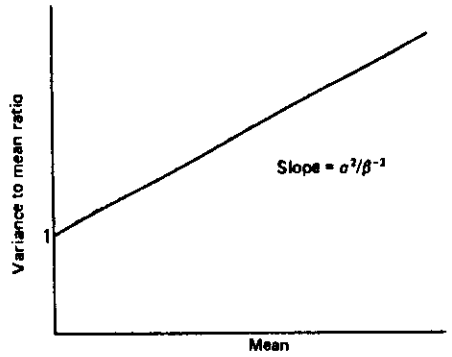

Fig. 1. The relation between the mean parasite burden and the variance to mean ratio predicted by Result 1.

The simulations of Anderson & Gordon (1982) show clearly that Result 1 is, in general, not valid in systems with density-dependent parasite mortality or parasite-induced host mortality. With density-dependent parasite mortality, $VM(t,a)$ often decreases as $M(t,a)$ increases and $VM(t,a)$ may be less than 1 (see Fig. 4 in Anderson & Gordon, 1982). We conjecture that the same would be true if a host's susceptibility were a function of its parasite burden (due for example to an immune response (Anderson & May (1985)). If host mortality increases with parasite burden then the relation between $VM(t,a)$ and $M(t,a)$ is, in general, not monotone (see Fig. 9 in Anderson & Gordon, 1982).

Result 1 may thus provide a criterion for the detection of density dependence that is valid even in temporally variable systems with host-age dependent parasite invasion and mortality rates. If $VM(t,a)$ is not a linearly increasing function of $M(t,a)$ then this is evidence of density dependence. It is important to realize that the converse is not necessarily true; a linear relation might be expected in some systems with density-dependent parasite mortality, parasite-induced host mortality or a parasite-induced immunological response that alters the probability of further infection. Density-dependent parasite fecundity, parasite-induced changes in host fecundity, and host density-dependent host fecundity or survivorship will affect the abundances of hosts and/or parasites, but will not cause the relation between $M(t,a)$ and $VM(t,a)$ to be non-linear.

We caution that Result 1 is dependent on the assumption that the probability density $b(z)$ does not vary with host age or time. The density $b(z)$ would be a function of host age, for example, if hosts become immunologically or behaviourally less heterogeneous as they age (Anderson & Gordon, 1982).

In practice, it may be difficult to detect directly non-linearly in the relation between the mean and variance to mean ratio. Suppose, however, that the distribution of

parasites/host in each cohort is negative binomial. Again, the negative binomial has proven to be an excellent empirical model in many studies. Because the variance to mean ratio of a negative binomial is $1+M/k$, Result 1 is obtained only if k is constant. Thus, in the negative binomial case, Result 1 can be restated simply as the parameter k is constant, $k=\bar{\beta}^2/\sigma^2$, and the age–intensity relation is given by equation (5). One could obtain evidence of density dependence from negative binomially distributed field data by determining whether or not k varies significantly from cohort to cohort. As before, this criterion is valid even in temporally variable systems with age-dependent epidemiological parameters. Incidentally, the distribution given in equation (4) is negative binomial for each value of a and t if Z is gamma distributed. See Adjei, Barnes & Lester (1986) for a method of estimating parasite-induced host mortality that is based on truncated negative binomial distributions.

Anderson & Gordon (1982) rightly pointed out that peaked age–intensity relations could be caused by age- or time-dependent parasite invasion or mortality rates as well as by parasite-induced host mortality. For example, it is straightforward to show from equation (5) that either of the following factors may cause peaks in an age–intensity curve: (a) parasite mortality increases with host age or (b) the invasion rate of parasites decreases with host age (e.g. decreasing susceptibility with increasing age). Moreover, suppose that the age–intensity relation is estimated using data from a longitudinal study in which one cohort is repeatedly sampled through time. A peak in the age–intensity relation from a longitudinal study could be caused by (a) parasite invasion rates that decrease over time (say due to a seasonal decrease in the abundance of infective stages) or (b) parasite mortality rates that increase over time. There is one interesting remaining case. Suppose that parasite invasion and mortality rates vary temporally but are unaffected by host age. Further suppose that a cross-sectional study is done; several host cohorts are sampled at a single point in time. Then, the invasion and mortality rates become $\lambda(t)$ and $\mu(t)$ and the age–intensity relation in equation (5) becomes

$$M(a)=\bar{B}\int_0^a \lambda(t-a+x)\,e^{-\int_x^a \mu(t-a+y)\,dy}\,dx. \tag{7}$$

Here, t should be taken as a constant because the study is cross-sectional. If $x'=x-a$ and $y'=y-a$, then

$$M(a)=\bar{B}\int_{-a}^{0} \lambda(t+x')\,e^{-\int_x^a \mu(t+y')\,dy'}\,dx'$$

and

$$\frac{dM}{da}=\bar{B}\lambda(t-a)\,e^{-\int_{-a}^{a}\mu(t+y')\,dy'}.$$

Because $\lambda(t-a)\geqslant 0$ and $\mu(t+y')\geqslant 0$, $dM/da\geqslant 0$. Thus, the mean number of parasites is a non-decreasing function of age. The implication of this result is that a peaked age-intensity relation is evidence of parasite-induced host mortality in a temporally variable system if the epidemiological parameters are independent of host age, and if the study is cross sectional.

Maximum likelihood estimation

Suppose that we collect a sample of N hosts and let X_i be the number of parasites in the ith host and further, suppose that the vector $\mathbf{A}_i$ is a series of measured attributes of the ith host excluding X_i. In the context of this study, $\mathbf{A}_i$ would include the age and/or time of collection of the ith host, but other host-specific variables such as habitat

descriptors or host size or reproductive condition could be included as well. We assume that the probability that a host with vector $\mathbf{A}_i$ has X parasites is given by a negative binomial distribution with parameters $M(\mathbf{A}_i)$ and $k(\mathbf{A}_i)$. Thus, the parameters of negative binomial are functions of the host-specific variables. Our goal is to estimate the parameters of the $M(\mathbf{A})$ and $k(\mathbf{A})$ functions from the $\mathbf{A}_i$ and the X_i.

With these assumptions, the log likelihood function for the data set $(X_i,\mathbf{A}_i)$, $i=1,\ldots N$ is

$$L=\sum_{i=1}^{N}\Big[k(\mathbf{A}_i)\log(k(\mathbf{A}_i))-k(\mathbf{A}_i)\log(M(\mathbf{A}_i)+k(\mathbf{A}_i))+X_i\log(M(\mathbf{A}_i)) -X_i\log(M(\mathbf{A}_i)+k(\mathbf{A}_i))+\sum_{j=1}^{X_i}\{\log(k(\mathbf{A}_i)+X_i-j\}-\log(X_i!)\Big]. \tag{8}$$

Let the r parameters of the $M(\mathbf{A})$ function be V_q, $q=1,2,\ldots,r$; and let the p parameters of the $k(\mathbf{A})$ function be W_h, $h=1,2,\ldots,p$. Values of the V_q and W_h that maximize L are maximum likelihood estimates. Our maximization protocol is described in the Appendix.

To estimate confidence limits for the parameters we employ a boot-strap procedure. One hundred data sets of size N are first obtained by randomly selecting values of $(\mathbf{A}_i,x_i)$ with replacement from the data. Parameter estimates are then obtained for each of the random data sets. The 96% confidence limits for each parameter are given by the third largest and third smallest estimate obtained.

Cestodes in arctic charr

We now analyse two sets of empirical data to illustrate the use of the above statistical and theoretical results. Halvorsen & Andersen (1984) obtained 5165 cestodes (*D. ditremum*) from 590 arctic charr that ranged in age from 2 to 13 years. The dark circles in Fig. 2A give the mean number of cestodes/host for each age class of host and the number above each mean gives the host sample size for the age class. Note that the age–intensity relation appears peaked; the mean number of parasites/host increases with age up to age 8 and then decreases. Halvorsen & Andersen (1984) also calculated values of the negative binomial parameter k for age classes 3–8. These estimates increase monotonically with age (Fig. 2B, dark circles), and so the distributions appear to become decreasingly aggregated in older fish. Finally, estimates of the variance to mean ratio first increase and then decrease with age (Halvorsen & Andersen, 1984). The authors conclude from the apparent peak in the age–intensity relation and the theoretical results of Anderson & Gordon (1982) that heavily infected fish suffer greater mortality than lightly infected fish. They speculate that this increased mortality is due to selective predation on heavily infected fish by black-throated divers.

We obtained maximum likelihood estimates from Halvorsen & Andersen's data for a three-parameter M function and a two-parameter k function. The mean function is

$$M(a)=V_1\left[\frac{1-e^{-V_3(a-V_2)}}{V_3}\right]. \tag{9}$$

This function is the age–intensity relation predicted in the absence of parasite-induced host mortality, if λ and μ are constant for each host and if hosts are not susceptible to infection until a threshold age, V_2 (see equation (5) and Anderson & Gordon, 1982). Apparently, plerocerids of the genus *Diphyllobothrium* are not found in 0- and 1-year-old arctic charr (Halvorsen & Andersen, 1984). The parameter V_2 is the mortality

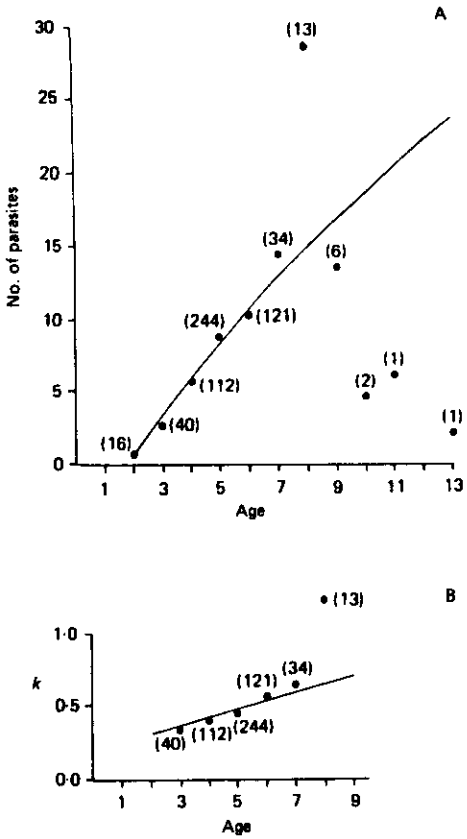

Fig. 2. (A) Age–intensity relation for cestodes in arctic charr (Halvorsen & Andersen, 1984). (●), Data means for each age class; the value in parenthesis above each data mean gives the host sample size. (—) Maximum likelihood fit of equation (9) to the data. (B) Relation between negative binomial parameter k and charr age. (●), Values calculated separately for each age class by Halvorsen & Andersen (1984). (—) Maximum likelihood best fit of equation (10).

rate of parasites and V_1 is the mean value of the parasite invasion rate for the host population. The k function is a linear function of host age

$$k(a) = W_1 + W_2 a. \tag{10}$$

Parameter estimates and 96% confidence limits are given in Table 1; predicted values of $M(a)$ and $k(a)$ are given by the lines in Fig. 2A and B respectively. Note that the predicted values of the mean and k closely match values calculated separately for each cohort, provided that the sample size of the age class is large (age classes 3–7). Deviations are larger for age classes with small sample sizes (classes 8–13). The line in Fig. 2B overshoots the separately calculated values for cohorts 3–5 and undershoots values for cohorts 6–8 primarily because the data in cohort 2 are not aggregated (and thus have infinite k).

Table 1. *Parameter estimates from a maximum likelihood fit of a three-parameter mean function and a two-parameter* k *function to Halvorsen & Andersen's* (1984) *cestode–arctic charr data*

Parameters	Estimate	Lower 96% confidence limit	Upper 96% confidence limit
V_1	2·889	1·593	4·276
V_2	0·062	−0·186	0·347
V_3	1·758	1·589	1·854
W_1	0·199	0·082	0·393
W_2	0·054	0·015	0·078

Halvorsen & Andersen (1984) fitted the model equation (9) directly to the *means* of the data from each age class (presumably by least squares). Their fit asymptotes at approximately 13 parasites/host rather than 46·6 as in Fig. 2A. The reason for the lower asymptote obtained by Halvorsen & Andersen (1984) is that their procedure weights equally the means derived from one or two individuals in age classes 10–13 and the means derived from over 100 individuals in age classes 4–6. Thus, the low means in age classes 10–13 are able to pull down their asymptote. In contrast, our procedure accounts for the differences among the sample sizes.

The maximum likelihood estimates indicate that the apparent peak in the age–intensity relation is likely to be an artifact of the small sample sizes in age classes 8–13. To see this, note first that the parameter V_2 is not significantly different from zero (Table 1). The parameter V_2 determines whether or not the function in equation (9) has a positive or negative second derivative. If V_2 is zero then the function in equation (9) is linear. Thus, we cannot conclude that the slope of $M(a)$ decreases with age, let alone becomes negative.

To demonstrate directly that the deviations between the sample and predicted means could be due to sampling artifacts we simulated the study as follows. Consider the cohort of hosts with age a_i and sample size N_i. We generated N_i negative binomial random numbers with the mean given by equation (9) evaluated at a_i, k given by equation (10) evaluated at a_i and using the estimated parameter values in Table 1. We then calculated the mean of these N_i numbers, repeated the entire procedure 200 times, and omitted the 5 highest and lowest sample means to obtain 95% sampling limits on the predicted mean. Of course, these limits could also be calculated directly from the negative binomial distribution. The upper and lower 95% limits for each age class are given by the lines in Fig. 3. Notice that *all* data means (the dark circles) fall within the 95% envelope. Because none of the deviations from the fitted model are statistically significant, we cannot conclude that the age–intensity relation is peaked. The marked widening of the 95% envelope as age increases is largely due to the small sample sizes of age classes 8–13. Finally, it is possible to show with similar methods, that the apparent peak in the variance to mean ratio reported by Halvorsen & Andersen (1984) is also likely to be a sampling artifact.

Unlike the age–intensity relation, the parameter estimates of the k function do provide evidence of density-dependent parasite mortality, parasite-induced host mortality or parasite-induced changes in host susceptibility. Specifically, the slope of the k function, W_2, is significantly positive (see the confidence limits in Table 1) and so the parasite distribution becomes decreasingly aggregated as host age increases. Because k is not constant, Result 1 is violated. We thus agree with Halvorsen and Andersen that there are indications of density dependence in their cestode–charr data.

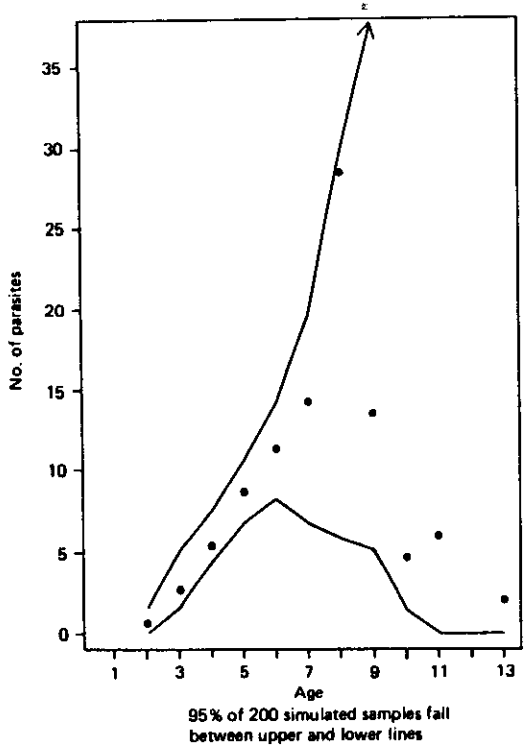

Fig. 3. Upper and lower 95% sampling limits on the mean for the negative binomial distribution with mean function of equation (9), k function of equation (10), parameter values in Table 1 and sample sizes in Fig. 2. Note that all data means (●), fall within these limits.

Nematodes in anoles

We collected 100 *Anolis bimaculatus* lizards on St Eustatius, Neth. Ant. in June 1985, and determined the number of *T. cubensis* nematodes in each. We cannot age these hosts and so, instead, we classify them by size. Size in *A. bimaculatus* is a reasonable indicator of age; fitted growth models for this species have been reported by Pacala & Roughgarden (1985). *A. bimaculatus* is sexually dimorphic. Males reach approximately 90 mm in snout–vent length and females reach approximately 65 mm. A total of 50 of the hosts were collected in relatively xeric coastal forest and 50 were collected in mesic montane (500 m elevation) forest. We used the maximum likelihood estimator to fit a linear mean function of length and constant k to these data. Four separate fits were obtained: montane males, montane females, coastal males and coastal females. The results, together with the data points are given in Fig. 4. Note that montane animals were more heavily infected than coastal animals.

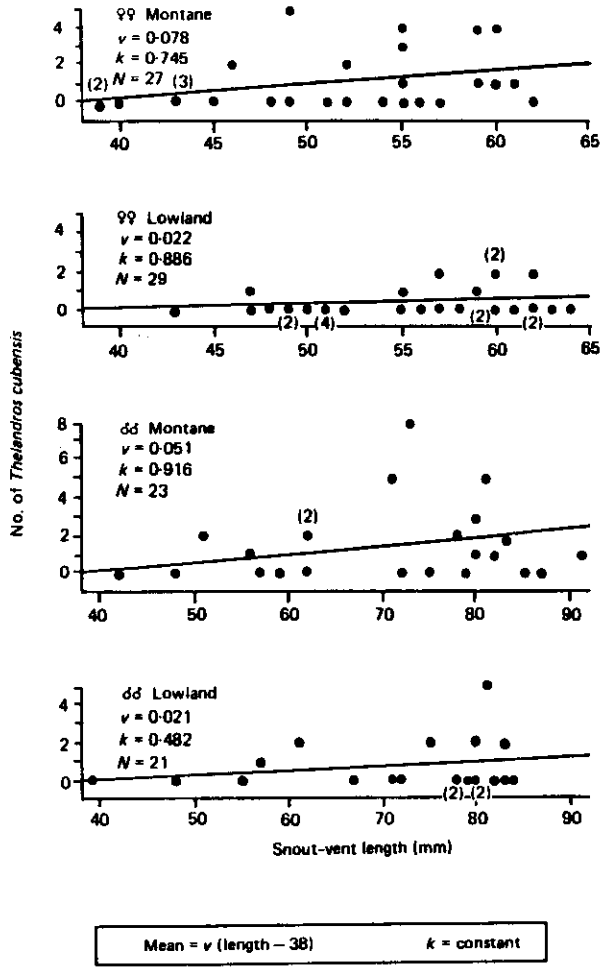

Fig. 4. Host size–intensity relations for nematode infections in Caribbean anoles.

DISCUSSION

The characterization of population dynamic processes from epidemiological patterns is at best a difficult endeavour because several different processes may lead to the same pattern. Here, we derive criteria to detect density dependence in the presence of confounding factors that are common in natural systems. These include temporal variation in the density of infective stages and time and host age-dependent variation in parasite mortality, probability of infection given an encounter, and host behaviour. We show that a peak in the age intensity relation from a cross-sectional study is

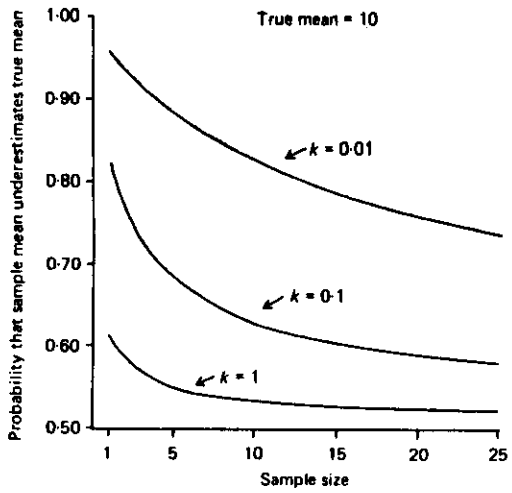

Fig. 5. Probability that the sample mean from a negative binomially distributed population will underestimate the true mean as a function of sample size. In all cases, $M = 10$.

evidence of parasite-induced host mortality whether or not parasite invasion and mortality rates vary in time. Such a peak may also be caused by increases in susceptibility or decreases in parasite mortality with host age. Moreover, we suggest a criterion for the detection of density dependence (Result 1) that is valid under quite general assumptions about the nature of host age and time dependencies. If the parasite distributions for each age class of host follows a negative binomial distribution, then this criterion is that k is constant in the absence of density dependence. The cestode–charr data are well described by the negative binomial (Halvorsen & Andersen, 1984). We caution that estimates of k could vary spuriously among age classes if the negative binomial distribution were fit to significantly non-negative binomially distributed data (see Pielou (1977) page 130). One other implication of Result 1, is that it lends support to epidemiological models that assume a constant k despite changes in parasite and host abundances (see Anderson & May, 1985).

The maximum likelihood statistical methods should prove useful regardless of the applicability of the theoretical results. In addition to estimating changes in the distribution of parasite burden with host age, the methods could be applied to any regression problem in which the data were negatively binomially distributed for all values of the independent variables. Methods similar to those that produce Fig. 4 could also be used to assess whether a seemingly abnormal value of any statistic calculated for an age class is a sampling artifact.

Fig. 4 shows that the peak in the charr age–intensity relation is not statistically significant and we conjecture that artifactual peaks are common in empirical studies. This is because the sample mean parasite burden for an age class will underestimate the true mean with high probability if the sample size is small and the parasite distribution is sufficiently aggregated. Fig. 5 shows for negative binomially distributed data, the probability that the sample mean will underestimate the true mean as a function of sample size. The three curves in the figure correspond to three different values of k. To

calculate these probabilities, we use the property that the sum of N negative binomial random variables, each with parameters M and k, is itself negative binomially distributed with parameters NM and Nk. The probability that the sample mean is an underestimate is always greater than 0·5 in Fig. 5 simply because the negative binomial distribution is not symmetric about the mean. Note that the probability of obtaining an underestimate increases as the sample size decreases and as the distribution becomes increasingly aggregated (as k decreases). Because of host mortality, the sample size for a host cohort typically decreases as host age increases, and so sample means for the oldest host cohorts are more likely to be underestimates than the other means. Thus, one expects artifactual peaks to be common in age–intensity relations. The important point here is that intuition honed by the study of symmetric distributions may be misleading when faced by asymmetrically distributed data. This underscores the need for statistical methods that are based on an aggregated distribution, such as the negative binomial.

APPENDIX

To maximize L in equation (8) with respect to the parameters $V_q, q = 1, 2, \ldots, r$ and $W_h, h = 1, 2, \ldots, p$ we must solve the system of $r+p$ equations:

$$\frac{\partial L}{\partial V_q} = 0 = \sum_{i=1}^{N} \frac{\partial M(A_i)}{\partial V_q}\left[\frac{X_i}{M(A_i)} - \frac{k(A_i)+X_i}{M(A_i)+k(A_i)}\right]; \quad q = 1, 2, \ldots, r \qquad \text{(A 1)}$$

$$\frac{\partial L}{\partial W_h} = 0 = \sum_{i=1}^{N} \frac{\partial k(A_i)}{\partial W_h}\Big|_{X_i} [1 + \log(k(A_i)) - \log(M(A_i)+k(A_i)) - \frac{k(A_i)+X_i}{M(A_i)+k(A_i)} + \sum_{j=1} 1/(k(A_i)+X_i-j)]; \quad h = 1, 2, \ldots, p,$$

for the roots: $\hat{V}_q$; $\hat{W}_h$, $q = 1, 2, \ldots, r$; $h = 1, 2, \ldots, p$.

We have implemented an iterative algorithm that will solve the system in equation (A 1) for $M(A)$ functions with up to three parameters and $k(A)$ functions with up to two parameters. The inputs for this programme are the data, the functional forms of $k(A)$ and $M(A)$, the first and second derivatives of these functions with respect to their parameters, and initial guesses of the parameter values. Given appropriate initial guesses, the programme converges to a local maximum of equation (8).

The programme's numerical root-finder is a modified Gauss–Newton algorithm. The advantage of the Gauss–Newton method is that it converges rapidly. We have written two versions of the programme, one in PL1 on an IBM mainframe, and the other in Pascal on a micro-computer. The micro-computer version is not prohibitively slow even with data sets including more than 500 hosts and 5000 parasites. The disadvantage of the Gauss–Newton method is that it is prone to oscillatory instabilities, and our implementation is certainly no exception. The algorithm is especially sensitive to the choice of initial values for parameters of the $k(A)$ function, as small changes in k result in large changes in the shape of the negative binomial. We have thus developed a hybrid procedure whereby a value for one parameter of the $k(A)$ function is chosen and held constant while the other $k(A)$ parameter and the parameters of $M(A)$ are solved by Gauss–Newton. By successively increasing or decreasing the value held constant and resolving for the remaining parameters each time by Gauss–Newton, the algorithm eventually converges to the roots of equation (A 1). Even so, preliminary data analysis is required to determine appropriate initial guesses for the parameters.

We have also implemented a simple hill-climbing (simplex) algorithm to solve the

system in equation (A 1). This programme will never exhibit oscillatory instabilities but is much slower than the Gauss–Newton routine. Nonetheless, the hill-climbing programme is useful to obtain initial guesses for the Gauss–Newton algorithm. Copies of the mainframe versions of all programmes are from S. Pacala and copies of the microcomputer versions are available from A. Dobson.

We thank J. Roughgarden for his field work, encouragement and friendship and the National Science Foundation (#BSR-83-16949) for funding the study.

REFERENCES

Adjei, E. L., Barnes, A. & Lester, R. J. G. (1986). A method for estimating possible parasite-related host mortality illustrated using data from *Calliterarhynchus gracilis* (Cestoda: Trypanorhyncha) in lizard fish (*Saurida*) spp. *Parasitology* **92**, 227–43.

Anderson, R. M. (1978). The regulation of host population growth by parasitic species. *Parasitology* **76**, 119–57.

Anderson, R. M. (1982). Epidemiology of infectious disease agents. In *Modern Parasitology* (ed. F. E. G. Cox), pp. 204–51. Oxford: Blackwell Scientific Publishers.

Anderson, R. M. & Gordon, D. M. (1982). Processes influencing the distribution of parasite numbers within host populations with special emphasis on parasite-induced host mortalities. *Parasitology* **85**, 373–98.

Anderson, R. M. & May, R. M. (1978). Regulation and stability of host–parasite population interactions. I. Regulatory processes. *Journal of Animal Ecology* **47**, 219–48.

Anderson, R. M. & May, R. M. (1979). Population biology of infectious diseases. I. *Nature, London* **280**, 361–7.

Anderson, R. M. & May, R. M. (1985). Helminth infections of humans: mathematical models, population dynamics, and control. In *Advances in Parasitology*, vol. 24, pp. 1–101. London: Academic Press.

Crofton, H. D. (1971*a*). A quantitative approach to parasitism. *Parasitology* **62**, 179–94.

Crofton, H. D. (1971*b*). A model of host–parasite relationships. *Parasitology* **63**, 343–64.

Dietz, K. (1982). Overall population patterns in the transmission cycle of infectious disease agents. In *Population Biology of Infectious Diseases* (ed. R. M. Anderson and R. M. May), pp. 87–102. Berlin: Springer-Verlag.

Dobson, A. P. (1985). The population dynamics of competition between parasites. *Parasitology* **91**, 317–47.

Gordon, D. M. & Rau, M. E. (1982). Possible evidence for mortality induced by the parasite *Apatemon gracilis* in a population of brook sticklebacks *Culaea inconstans*. *Parasitology* **84**, 41–7.

Halvorsen, O. & Andersen, K. (1984). The ecological interaction between arctic charr, *Salvelinus alpinus* (L.), and the plerocercoid stage of *Diphyllobothrium ditremum*. *Journal of Fish Biology* **25**, 305–16.

Kennedy, C. R. (1984). The use of frequency distributions in an attempt to detect host mortality induced by infections of diplostomatid metacercariae. *Parasitology* **89**, 209–20.

Lester, R. J. G. (1984). A review of methods for estimating mortality due to parasites in wild fish populations. *Helgolander Meeresunters* **37**, 53–64.

May, R. M. (1977). Dynamical aspects of host–parasite associations: Crofton's model revisited. *Parasitology* **75**, 259–76.

May, R. M. & Anderson, R. M. (1978). Regulation and stability of host–parasite population interactions. II. Destabilizing processes. *Journal of Animal Ecology* **47**, 249–68.

May, R. M. & Anderson, R. M. (1979). Population biology of infectious diseases. II. *Nature, London* **280**, 455–61.

Pacala, S. W. & Roughgarden, J. D. (1985). Population experiments with the *Anolis* lizards of St Maarten and St Eustatius. *Ecology* **66**, 129–51.

Pielou, E. C. (1977). *Mathematical Ecology*. New York: John Wiley.

Printed in Great Britain