Analysis of aggregated parasite distributions: a comparison of methods

K. WILSON,*† B. T. GRENFELL and D.J. SHAW‡
Zoology Department, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

Summary

1. Empirically, parasite distributions are often best described by the negative binomial distribution; some hosts have many parasites while most have just a few. Thus identifying heterogeneities in parasite burdens using conventional parametric methods is problematical. In an attempt to conform to the assumptions of parametric analyses, parasitologists and ecologists frequently log-transform their overdispersed data prior to analysis. In this paper, we compare this method of analysis with an alternative, generalized linear modelling (GLM), approach.

2. We compare the classical linear model using log-transformed data (Model 1) with two GLMs: one with Poisson errors and an empirical scale parameter (Model 2), and one in which negative binomial errors are explicitly defined (Model 3). We use simulated datasets and empirical data from a long-term study of parasitism in Soay Sheep on St Kilda to test the efficacies of these three statistical models.

3. We conclude that Model 1 is much more likely to produce type I errors than either of the two GLMs, and that it also tends to produce more type II errors. Model 3 is only marginally more successful than Model 2, indicating that the use of an empirical scale parameter is only slightly more likely to generate errors than using an explicitly defined negative binomial distribution. Thus, while we strongly recommend the use of GLMs over conventional parametric analyses, either GLM method will serve equally well.

Key-words: Aggregation, generalized linear modelling, GLIM, macroparasites, negative binomial, Splus

Functional Ecology (1996) **10**, 592–601

Introduction

Macroparasites, in common with many herbivorous insects (Southwood 1978), tend to be aggregated over their host populations; some hosts have high parasite burdens, while others have none or a few (e.g. Pennycuick 1971; Anderson 1974, 1978; Pacala & Dobson 1988). If parasites were randomly distributed over their host population, their frequency distribution would conform to the Poisson and the population variance in parasite load, σ^2, would be equal to its mean, μ. However, for most parasites, $\sigma^2 > \mu$, and thus there is greater variance in parasite loads than expected by chance. Theoretical studies suggest that the degree of aggregation may profoundly influence the stability of host–parasite interactions (e.g. Anderson & May 1985, 1991), and so it is important to establish the causes and extent of such patterns of dispersion. A

first step in this direction is to quantify the observed patterns accurately. Generalized linear modelling provides a readily accessible technique for doing this and is the subject of this paper.

In general, parasite distributions are empirically best described by the negative binomial distribution (Crofton 1971; Anderson 1974; 1978; Elliot 1977; Southwood 1978; Shaw & Dobson 1995), which is completely defined by its mean and the exponent k. The variance of a negative binomial distribution is described as follows:

$$\sigma^2 = \frac{\mu - \mu^2}{k}.$$

<div align="right">eqn 1</div>

Thus, the degree of aggregation declines as k increases until, as k approaches infinity (or in practice, above about 20), the distribution converges on the Poisson (Elliot 1977). In order to fit the negative binomial distribution, we need to estimate the exponent k. The most accurate estimate of k is obtained by maximum-likelihood methods (Elliott 1977), but a moment-estimate can be calculated by rearrangement of equation 1 and this is usually close to the maximum-likelihood estimate:

*To whom correspondence should be addressed.
†Present address: Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, UK.
‡Present address: Department of Parasitology, Faculty of Veterinary Medicine, University of Gent, Salisburylaan 133, 9820 Merelbeke, Belgium.

$$\hat{k} = \frac{\bar{x}^2}{(s^2 - \bar{x})}, \qquad \text{eqn 2}$$

where $\bar{x}$ is the sample mean, s^2 is the sample variance and $\hat{k}$ is the estimated value of k.

Traditionally, classical linear regression models (and analyses of variance) are used by ecologists and parasitologists to determine differences between mean parasite burdens. However, they have the disadvantage that they assume that the error distribution is normal (i.e. Gaussian). Because parasite distributions tend to be aggregated, this assumption is usually violated and so the analysis is often performed on logarithmically transformed data (Gregory & Woolhouse 1992; Fulford 1994). While the log-normal distribution provides a good fit to much parasite data, it fails when the parasites are highly aggregated, as this often results in the distribution of the transformed data being bimodal because of an excess of zeros (Wilson & Grenfell, in press). As a result, both type I and type II errors (as defined below) are likely when using this method.

Clearly, differences between mean parasite burdens would be assessed more accurately if one could specify a model that explicitly accounted for the negative binomial error distribution. Generalized linear modelling provides a readily accessible technique for doing this (for an gentle introduction into the ecological applications of generalized linear modelling, see Crawley 1993). GLMs are generalizations of classical linear models (which assume normally distributed errors) in which the *error structure* is explicitly defined by one of a series of distributions from the exponential family, and in which the expected values of the response variable are mapped on to the explanatory variables by a *link function*. For the negative binomial distribution (which is not in the exponential family), the GLM error structure is defined by equation 1 and the log or square-root link is usually used (Crawley 1993; Venables & Ripley 1994).

A commonly used alternative to explicitly defining a negative binomial error distribution is to assume a Poisson distribution and to adjust the *scale parameter* (or *dispersion parameter*) so that the ratio of the residual deviance and its degrees of freedom is approximately equal to unity (Aitkin *et al.* 1989; Crawley 1993; see below). Thus, instead of asssuming that the variance of the parasite distribution is *equal* to its mean ($\sigma^2 = \mu$, for the Poisson distribution), we assume that it is *proportional* to it (i.e $\sigma^2 = \phi\mu$, where ϕ is referred to as the scale or dispersion parameter).

This paper is in two parts. In the first part, we use randomly generated data from the negative binomial distribution to show that, in general, GLMs provide a major improvement upon conventional linear regression analyses of log-transformed data, especially when negative binomial errors are explicitly defined. In the second part, we use post-mortem worm counts and faecal egg counts from an unmanaged population of Soay Sheep on St Kilda to show that GLMs and traditional methods can give markedly different pictures as to the heterogeneities in parasite loads. We conclude by assessing the limitations of the GLM technique.

Materials and methods

GENERALIZED LINEAR MODELS

Most currently available statistical packages have GLM procedures. However, only a limited number of error structures (those in the exponential family) are generally available. For example, *GLIM* and *Genstat* allow only for normal, gamma, Poisson, binomial and inverse Gaussian errors. At present, no statistical packages explicitly provide for negative binomial errors. Thus, when data are overdispersed, as indicated by a ratio of residual deviance to residual degrees of freedom significantly greater than 1 for the minimal model (Crawley 1993), the model must be adjusted. This is usually achieved by multiplying the scaled deviances by an empirical scale parameter. Often this is calculated as simply the ratio of the scaled deviance to its degrees of freedom, but Aitkin *et al.* (1989) recommend that it be estimated as Pearson's χ^2 divided by the residual degrees of freedom. In either case, the parameter estimates are not affected but the standard errors are inflated (Crawley 1993) and, in a manner directly analogous to standard analysis of variance and regression procedures, the scaled deviances for terms in the model are compared using *F-tests* instead of χ^2-tests (Aitkin *et al.* 1989; Crawley 1993).

Although no statistical packages explicitly describe negative binomial errors, there are appropriate functions available in the public domain. For example, Crawley (1993) provides macros both for estimating k by maximum-likelihood methods and for using that k in a GLM. Here, we use the *Splus* statistical package (MathSoft, Inc. 1993; Seattle, WA) and the *glm.nb* function available by anonymous ftp from StatLib (*ftp lib.stat.cmu.edu*; see also Venables & Ripley 1994). This function also uses maximum-likelihood methods to estimate both the common k (referred to as *theta* in *glm.nb* output, see below) and the parameter values associated with the terms in the model.

SIMULATED DATASETS

In order to assess the utility of the GLM approach, we first compared it with traditional methods using a series of randomly generated datasets from the negative binomial distribution. The datasets comprised 20, 100 or 500 numbers from distributions with means ranging between 1 and 2000, and with k values ranging between 0·5 and 20. These ranges cover those of most parasite burdens and faecal egg counts. For each of a pair of datasets, with either identical means or

means differing from each other by 100% (1 *vs* 2, 5 *vs* 10, 10 *vs* 20, etc.), the statistical significance of the difference between the means was assessed over 100 trials using three models of increasing sophistication:

1. Model 1: a generalized linear model with Gaussian errors using $\log_{10}$-transformed data (equivalent to a classical linear regression or analysis of variance model).
2. Model 2: a generalized linear model with Poisson errors and adjusted scale parameter, using untransformed data.
3. Model 3: a generalized linear model with negative binomial errors, using untransformed data.

The explanatory variable of the three models comprised a single factor, which coded for each of the pair of distributions. The number of times that the different models detected significant differences between distributions was scored over 100 trials using *F*-tests (Model 1 and Model 2) or χ^2 tests (Model 3). Thus, by comparing the output of the three models it was possible to assess the probabilities of each performing:

1. Type I errors (i.e incorrectly *rejecting* the null hypothesis of no difference between the means).
2. Type II errors (i.e. incorrectly *accepting* the null hypothesis).

SOAY SHEEP PARASITE DATA

In the second part of the analysis we use the three methods to examine the factors influencing the post-mortem worm counts and faecal egg counts of an unmanaged population of Soay Sheep on the Scottish island group of St Kilda (see Gulland 1992, Gulland & Fox 1992). This population exhibits severe 'crashes' every 3–4 years, when up to 60% of the sheep die. This mortality is mainly due to the population overexploiting its winter food supply (Grenfell *et al.* 1992; Clutton-Brock *et al.*, in press), but parasites have been implicated in the differential survival of their hosts (Gulland 1992; Gulland *et al.*, 1993; K. Wilson *et al.*, unpublished observation).

In the first comparison, we use worm burdens of sheep that died during the population crash of 1991/92 (described in detail in K. Wilson *et al.* unpublished observations). Out of the 75 sheep that were autopsied at this time, 53 were lambs (< 12-months-old), 14 were yearlings (22–26-months-old) and eight were adults. Preliminary analysis indicated that a maximum of two age-classes were distinguishable: lambs (< 12-months-old) and 'adults' (> 12-months-old). Thus, analyses are presented in which the maximal model included two factors and their interaction: AGECLASS (two levels) and SEX (two levels).

In the second comparison, we examined the factors influencing the August faecal egg counts of sheep (K. Wilson *et al.* unpublished observations). Preliminary analyses established that sampling year (1988–93)

explained more variance in August faecal egg counts than sheep population density alone, and that the minimum number of definable sheep age-classes was four: lambs (4-months old), yearlings (16-months old), two-year-olds (28-months old) and adults (≥40-months old). Thus, analyses were performed in which the maximal model included 3 factors and their interactions: AGECLASS (four levels), SEX (two levels) and YEAR (six levels).

In these analyses, as well as testing the efficacies of the three linear models against each other, we also tested them against a simple non-linear maximum-likelihood model (see Shaw 1994 for details). The primary difference between this model and Model 3 is that, if appropriate, it can fit separate *k* values to each of the component distributions, rather than having to rely on a single common *k*.

Results

SIMULATED DATASETS

Comparisons are made between the three different models under five scenarios: (a) when the means and the *k* values of the two distributions are identical; (b) when the means are identical, but the *k* values differ; (c) when the means differ, but the *k* values are the same; (d) when the means and the *k* values both differ, and the *k* values and means are positively related (in other words, distributions with relatively larger means also have relatively larger *k* values, as appears to be true of most parasite distributions; Shaw & Dobson 1995; Grenfell *et al.* 1995; present study); and (e) when the means and *k* values both differ, and the *k* values and means are negatively related. Mistakes in the cases (a) and (b) are referred to as type I errors (i.e. incorrectly rejecting the null hypothesis of no difference between the means), and in the cases of (c), (d) and (e) are referred to as type II errors (i.e. incorrectly accepting the null hypothesis).

(a) Comparisons of distributions with the same means and the same k *values*

In this and the following series of comparisons, we determined the probability of each of the models producing type I errors. Thus, we assessed the number of times (out of 100 trials) that each of the models indicated a significant difference between two simulated distributions that had exactly the same mean and *k* value. All three models, over all combinations of sample sizes (*n*), means ($\bar{x}$) and *k* values had similar, and very low, probabilities of producing this kind of type I error (Fig. 1).

(b) Comparisons of distributions with the same means, but different k *values*

In this series of comparisons, the probability of
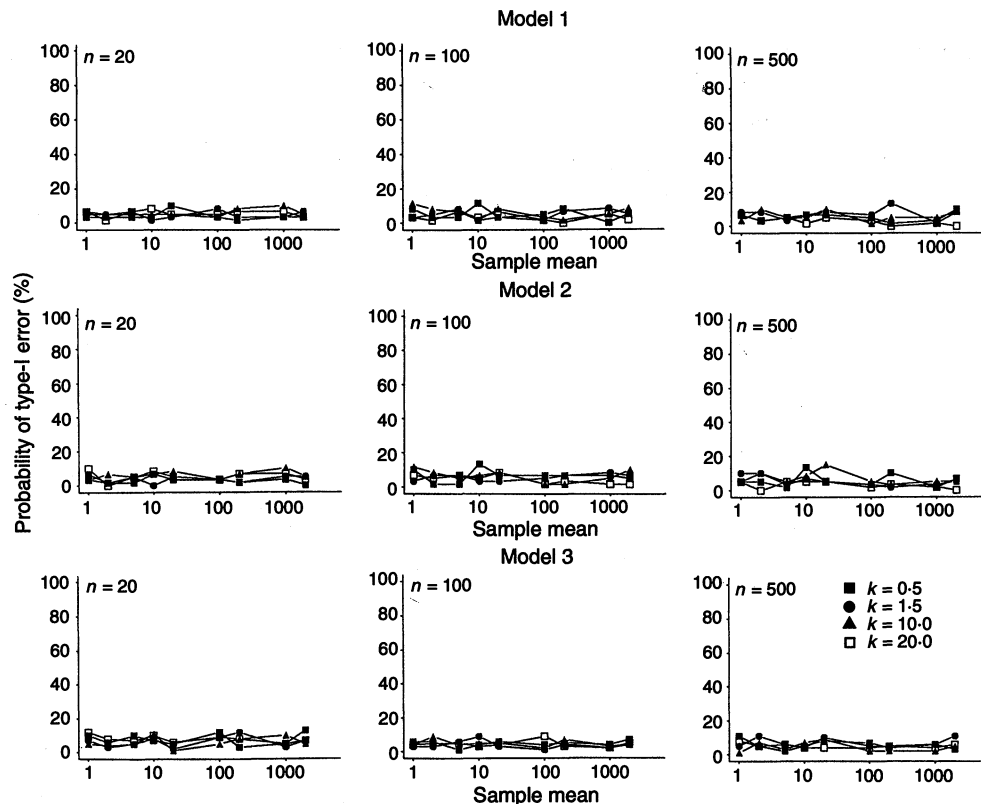
Fig. 1. Probability of producing type I errors when the component distributions have the same mean and the same k value. Each box refers to a series of simulations to assess the probability of three statistical models making type I errors and incorrectly rejecting the null hypothesis (see main text for a description of the three models and the simulations). The sample size for each of the component datasets is indicated in the top left-hand corner of each box and the k values are indicated by the different symbols, shown in the bottom right-hand box. All three models have a low probability of producing a type I error when the means and k values are identical.

Model 1 (the linear regression model using log-transformed data) producing type I errors ranged between between 0 and 100% and increased with sample size (n), sample mean ($\bar{x}$) and difference between the component k values. By comparison, both the GLMs (Model 2 and Model 3) produced many fewer type I errors over all values of n, $\bar{x}$ and k, and both models failed with a probability ranging between about 0 and 15% (Fig. 2).

(c) Comparisons of distributions with different means, but the same k values

The probability of producing a type II error (incorrectly accepting the null hypothesis) was first determined for two distributions that had different means, but the same k values (Fig. 3). For all three models, there was a significant probability of producing a type II error from such a comparison, except when sample sizes were very large ($n > 100$) or the k value was large ($k \geq 10$).

At low sample sizes ($n = 20$), the probability of Model 1 incorrectly accepting the null hypothesis increased as the distribution became increasingly aggregated (and k declined), such that for $k \ll 10$ between 50 and 80% of comparisons resulted in

type II errors over the range of sample means examined. As sample size increased (to $n = 100$), the probability of type II errors was approximately zero, except when the distributions were highly aggregated ($k < 1$). Thus, when comparisons are made between samples with different means, but similarly shaped distributions, linear models using log-transformed data are highly likely to produce type II errors, especially when sample sizes are small (< 100) or the distributions are highly overdispersed ($k \ll 10$).

Model 2 and Model 3 again tended to perform better than the linear model, though only marginally so. The probability of either of them producing type II errors was significant only when sample sizes were small ($n = 20$) and when the distributions were severely overdispersed ($k \ll 10$) or the sample mean was very low $\bar{x} < 5$). The GLM with negative binomial errors (Model 3) was slightly less likely to produce type II errors than the Poisson one with an empirical scale parameter (Model 2), or the classical linear model (Model 1). For example, for a comparison of two distributions with means of 100 and 200, k values of 1, and sample sizes of 20, the probability of type II error was 62%, 50% and 40% for models 1–3, respectively.
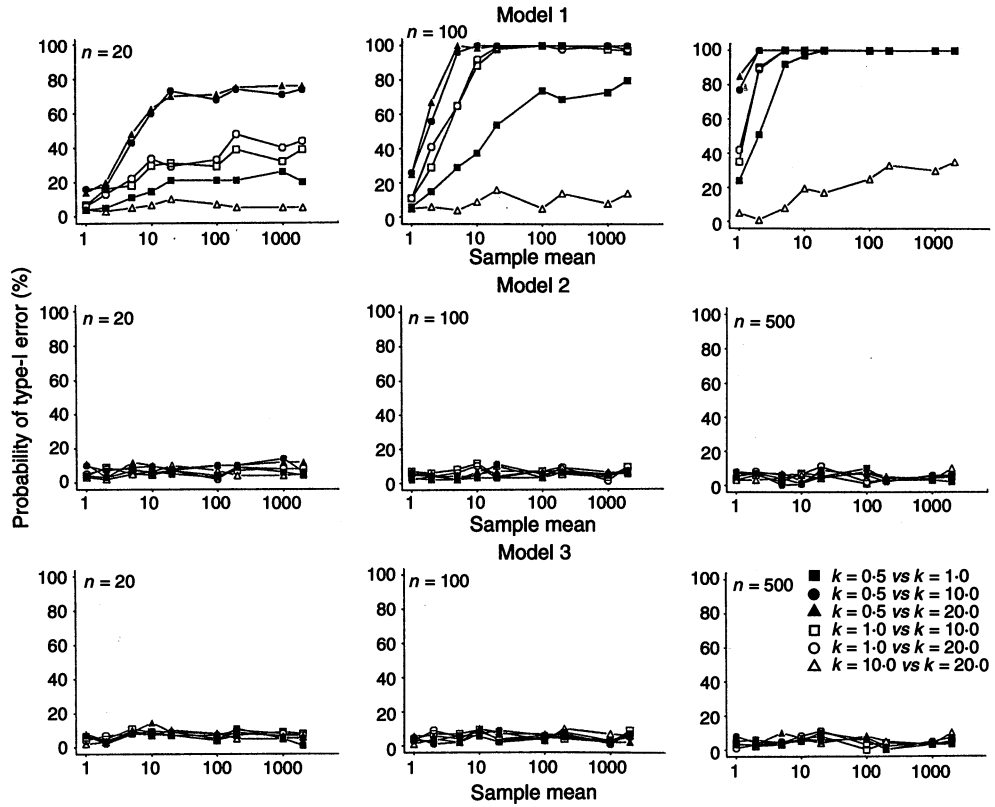
Fig. 2. Probability of producing type I errors when the component distributions have the same mean but different *k* values. See legend to Fig. 1. When the means are identical, models 2 and 3 have a low probability of producing type I errors, even when the *k* values differ. Model 1, on the other hand, generates errors over a range of sample sizes, sample means and component *k* values.
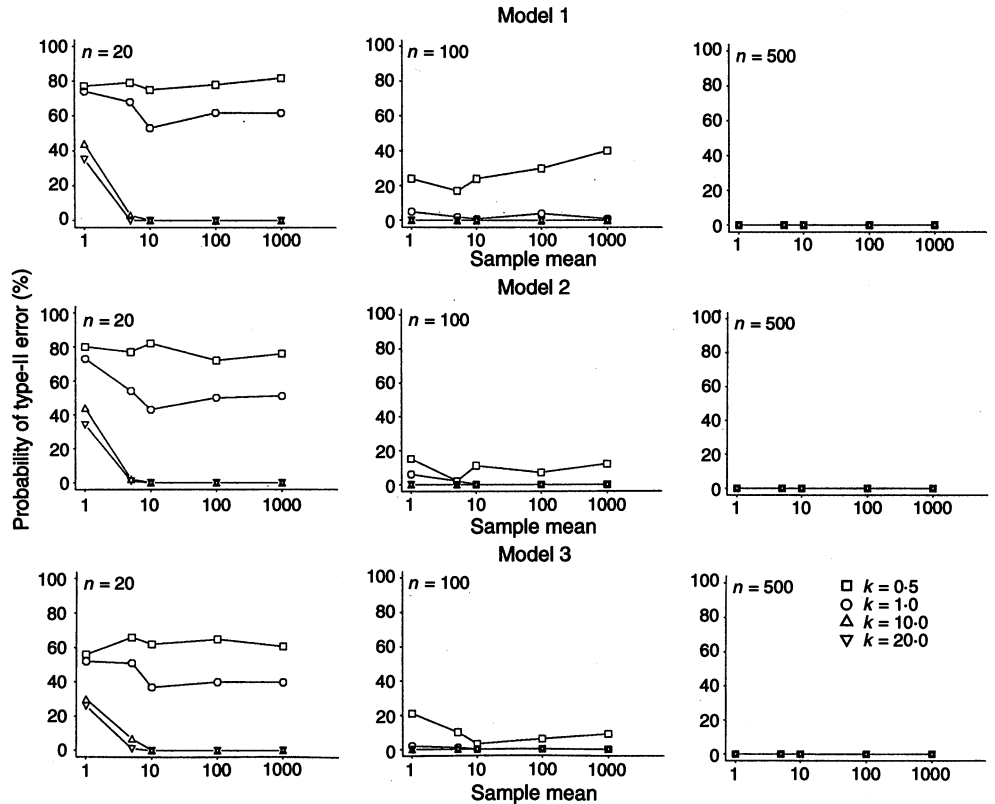
Fig. 3. Probability of producing type II errors when the component distributions have different means but the same *k* values. See legend to Fig. 1. When the means differ by 100%, but the *k* values are identical, all three models have a low probability of producing type II errors when sample sizes are large. When sample sizes are small and *k* values are small, all three models generate type II errors.

Model 1
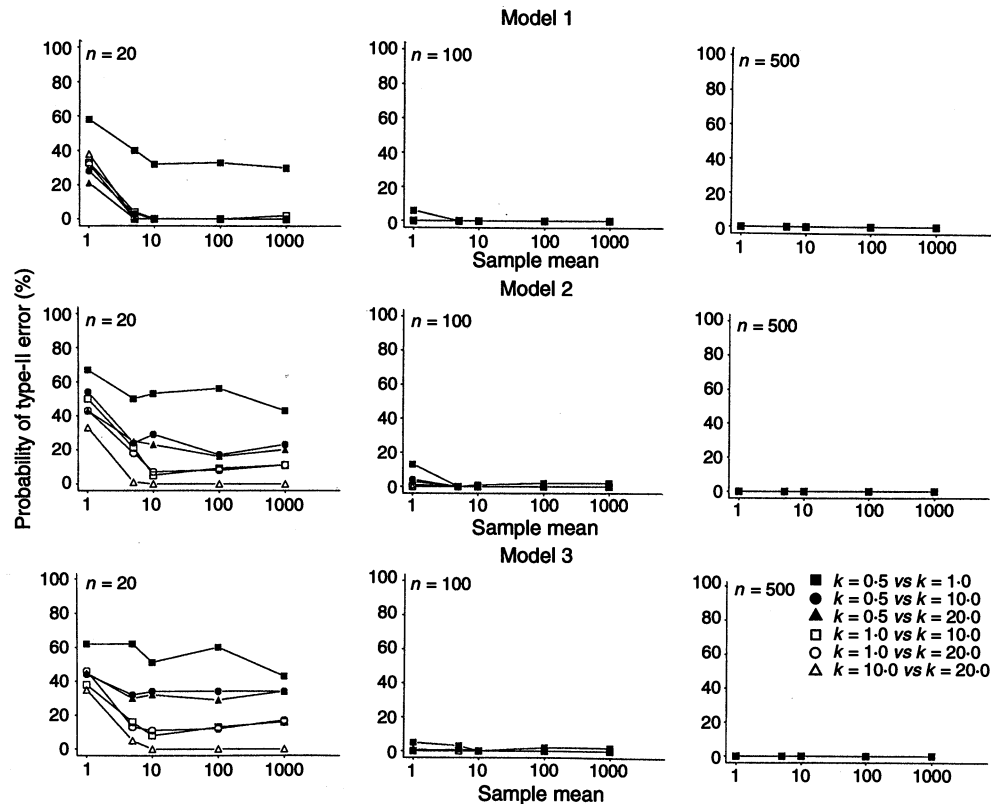
Fig. 4. Probability of producing type II errors when the component distributions have both different means and different k values, and means and k values are positively related. See legend to Fig. 1. When the means and k values differ and are *positively* related to each other, as appears to occur in nature, all three models have a low probability of producing type II errors when sample sizes are reasonably large ($n \geq 100$). However, when sample sizes are small, all three models (and especially the GLMs) generate type II errors, particularly when the component k values are markedly different.

(d) Comparisons of distributions with different means and different k *values, where the* k *values and means are positively related*

When both the means and the k values of the two distributions differ and the k values and means are positively related, all three models have a low probability of producing type II errors, except when sample sizes are small (n = 20). Paradoxically, Model 1 produced fewer errors than either Model 2 or Model 3 at this time, but did so only when the k values of both distributions were very small (e.g. 0·5 and 1·0) or the sample means were low ($\bar{x} < 5$) (Fig. 4). This counterintuitive result is due to the fact that the coefficients of variation for log-transformed distributions are generally smaller than those for the non-transformed distributions (Fig. 5) and hence, when sample sizes are small, Model 1 is less likely than the other two models to indicate significant differences between similar means (as witnessed by the higher probability of type I errors, see above). Under these circumstances it is therefore advisable to use alternative modelling methods, such as non-linear maximum-likelihood analysis.

At low sample sizes (n = 20), the probability of Model 2 and Model 3 producing type II errors ranged between 0 and 70%, and declined as the overall k of the combined distribution increased. There seemed to

be little effect of sample mean on the probability of error when the means of both distributions were greater than 5, but below this point the probability of error was generally much higher. As illustrated in Fig.
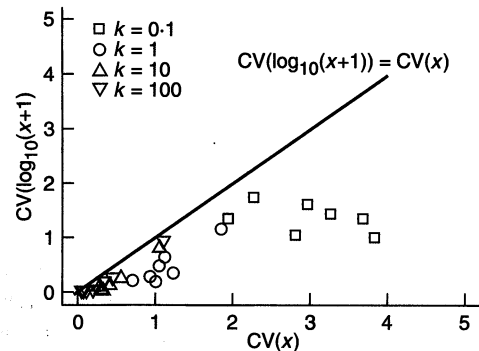
Fig. 5. Relationship between the coefficient of variation (CV) of data randomly generated from the negative binomial distribution (CV(x)) and that for the same data following logarithmic transformation (CV[log10(x+1)]). The means of the untransformed distributions ranged between 1 and 1000, and k values between 0·1 and 100. The degree of variation in the $\log_{10}$-transformed data is consistently less than that in the untransformed data, especially when k values are low, thus increasing the probability of Model 1 generating type I errors and reducing the probability of it generating type II errors.
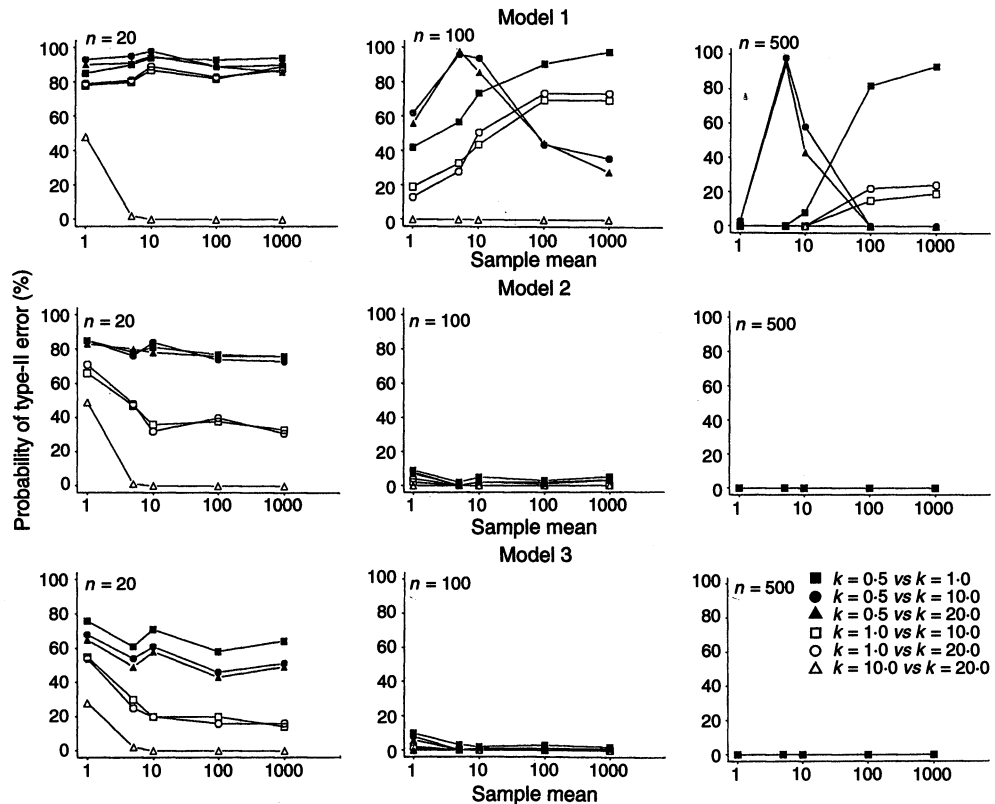
Fig. 6. Probability of producing type II errors when the component distributions have both different means and different k values, and means and k values are negatively related. See legend to Fig. 1. When the means and k values differ and are *negatively* related to each other, models 2 and 3 have a low probability of producing type II errors and have a high probability of producing type II errors only when sample sizes are small ($n = 20$). However, Model 1 has a high probability of generating type II errors even when sample sizes are large ($n = 500$).

4, there was little difference in the efficacies of the two GLMs.

(e) Comparisons of distributions with different means and different k values, where the k values and means are negatively related

When the k values and means are *negatively* related, Model 1 performs badly (Fig. 6). Over the range of k values, sample means and sample sizes employed in this study, Model 1 was much more likely to accept the null hypothesis incorrectly than either Model 2 or Model 3. At very small sample sizes ($n = 20$), the probability of error ranged between about 80 and 100%, except when both samples were approximately Poisson distributed (k values ≥ 10). When sample sizes were larger (≥ 100), the probability of a type II error ranged between 0 and 100% and either *increased* with sample mean (when distributions had similar k values); *decreased* with sample mean, after an initial increase (when distributions had very dissimilar k values); or was *unaffected* by sample mean and equal to zero (when both samples were approximately Poisson distributed).

For both Model 2 and Model 3, the probability of type II error was negligible for sample sizes greater than or equal to 100. At low sample sizes ($n = 20$), the

probability of error increased as the overall k value declined and was unaffected by sample mean when both means were greater than or equal to 5. Model 3 performed significantly better than Model 2. For example, when two samples comprising 20 data points were compared, one with a sample mean of 100 and a k value of 10 and another with a mean of 200 and a k value of 1, the probabilities of type II errors were 82%, 38% and 20% for Models 1-3, respectively.

SOAY SHEEP PARASITE DATA

Worm burdens

The three models were used to examine the effects of AGECLASS, SEX and AGECLASS:SEX interaction on the burdens of six species or genera of adult helminth worms (Table 1). Two species of worm were excluded from this analysis either because their distributions failed to conform to the negative binomial (*Moniezia* sp.) or because Model 3 failed to reach convergence (*Capillaria* sp.). This analysis identified three qualitative discrepancies between the results of the three models. In two of these (*Trichuris ovis* and *Dictyocaulus filaria*), the two GLMs (Model 2 and Model 3) identified SEX as a significant heterogeneity in worm burdens, when the conventional linear model (Model 1)

Table 1. Comparison of three statistical models for assessing age and sex differences in worm burdens of sheep that died during the population crash of 1991–92. Age and sex differences in the burdens of six taxa of adult worms are compared using Model 1 (conventional linear model with $\log_{10}$-transformed data and Guassian errors), Model 2 (GLM with Poisson errors and an empirical scale parameter, and Model 3 (GLM with negative binomial errors). Worm taxa are arranged in terms of increasing mean worm burden and declining aggregation (increasing $\hat{k}$). Significance levels from F-tests (Model 1 and Model 2) and χ^2-tests (Model 3), following stepwise deletion (Crawley 1993) are as follows: NS, $P > 0.05$, $* P < 0.05$, $** P < 0.001$, $*** P < 0.0001$. Two other taxa of worms were excluded from this table. *Moniezia* sp. was excluded because it failed to conform to the negative binomial distribution ($\hat{k} = 42.70 \pm 268.17$), and *Capillaria* sp. was excluded because Model 3 failed to reach convergence. Worm burdens were calculated for sheep that died in the population crash of 1991–92. These comprised 38 female lambs, seven yearling or adult ewes, 15 male lambs and 15 yearling or adult rams. Factors for which the three models give different significance levels are shown in boxes

Species	Location	Mean	Variance	$\hat{k} \pm SE$	Empirical scale parameter (Model 2)	AGECLASS (2 levels)			SEX (2 levels)			AGECLASS:SEX		
						Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
Trichuris ovis	L. intestine	22	705	0.79 ±0.15	18	***	***	***	NS	**	*	NS	NS	NS
Bunostomum trigonocephalum	S. intestine	47	1 597	1.18 ±0.20	31	NS	NS	NS	*	*	*	NS	NS	NS
Dictyocaulus filaria	Lungs	137	8 157	1.70 ±0.28	56	***	*	*	NS	*	*	NS	NS	NS
Chabertia ovina	L. intestine	569	148 180	3.32 ±0.52	226	***	***	***	NS	NS	NS	NS	NS	NS
Trichostrongylus spp.	S. intestine	2 631	2 157 287	3.40 ±0.53	813	NS	NS	NS	NS	NS	NS	NS	NS	NS
Teladorsagia spp.	Abomasum	14 658	41 658 107	5.00 ±0.79	2696	NS	NS	NS	NS	NS	NS	NS	*	*

failed to do so. In the third (*Teladorsagia* spp.), Models 2 and 3 identified a significant AGE:SEX interaction, whereas Model 1 failed to do so. Thus, it appears that the conventional linear model committed type II errors in these analyses. This was checked for *T. ovis* and *D. filaria* using non-linear maximum-likelihood methods (for details see Shaw 1994). This indicated that for *T. ovis*, the sexes differed both in their negative binomial k values ($\chi^2_1 = 4.70$, $P = 0.030$), and in their means ($\chi^2_1 = 3.70$, $P=0.040$), and for *D. filaria* there was no significant heterogeneity in their negative binomial k values ($\chi^2_1 = 3.03$, $P = 0.082$), but there was in their means ($\chi^2_1 = 82.13$, $P < 0.001$).

August faecal egg counts

The variation in ($\log_{10}$-transformed) August faecal egg counts, with respect to sheep age, sheep sex and year of sample, is illustrated in Fig. 7. Analyses were performed in which the maximal models included the following three factors and their interactions: AGE-CLASS (four levels), SEX (two levels) and YEAR (six levels). Analysis of deviance tables for the three maximal models are shown in Table 2(a).

Although the three models give similar results for the maximal model (a model including all of the factors and their interactions), stepwise deletion tests (e.g. Crawley 1993) indicate that the SEX:YEAR interaction is non-significant for Model 1, whereas it is statistically significant for Model 2 and Model 3. Thus, conventional methods appear to fail to pick up a significant interaction between two important factors.

The interpretation of this analysis is complicated by the large number of significant interaction terms, and it is therefore constructive to repeat the above analysis for the two sexes separately. This indicates that all three models identify the same heterogeneities in female faecal egg counts (AGECLASS, YEAR and AGE-CLASS:YEAR interaction; Table 2b), but that they differ in the identified heterogeneities in male egg counts (Table 2c). Whilst all three models indicate important between-year variation in male faecal egg counts, only Model 1 identifies significant differences between the four age-classes ($P < 0.01$). Thus, the conventional linear regression model appears to have made a type I error that is not made by the two GLMs. This can be verified by using a non-linear maximum-likelihood model. This indicates that, in males, there is significant heterogeneity between age-classes in negative binomial k values ($\chi^2_3 = 12.27$, $P = 0.0065$), but not in mean faecal egg counts ($\chi^2_3 = 3.20$, $P = 0.36$), thus confirming the results of models 2 and 3.

Discussion

The results of the simulation studies suggest that, overall, standard parametric tests have low utility even when the data are first log-transformed. Although type II errors are unlikely when sample

Table 2. Analysis of deviance table for comparing three statistical models to examine the factors influencing August faecal egg counts of Soay sheep between 1988 and 1993. Age, year and sex differences in the August faecal egg counts of Soay sheep are compared using Model 1 (conventional linear model with $\log_{10}$-transformed data and Gaussian errors), Model 2 (GLM with Poisson errors and an empirical scale parameter) and Model 3 (GLM with negative binomial errors). Significant terms are shown in bold type. Factors for which the three models give different significance levels are shown in boxes

		Deviance			Probability		
Factor	DF	Model 1	Model 2	Model 2	Model 1 (F-test)	Model 2 (F-test)	Model 3 (χ^2-test)
(a) Both sexes							
NULL MODEL	715	90·554	2775·16	1148·65			
AGECLASS (A)	3	6·613	192·43	67·63	**< 0·0001**	**< 0·0001**	**< 0·0001**
SEX (S)	1	6·249	150·02	60·61	**< 0·0001**	**< ·00001**	**< 0·0001**
YEAR (Y)	5	6·527	207·58	64·84	**< 0·0001**	**< 0·0001**	**< 0·0001**
A:S	3	1·626	83·23	24·51	**0·0003**	**< 0·0001**	**< 0·0001**
A:Y	15	2·679	79·91	46·91	**0·0073**	**0·0116**	**< 0·0001**
S:Y	5	0·823	32·83	12·87	0·0814	**0·0293**	**0·0249**
A:S:Y	11	1·903	78·46	30·67	**0·0182**	**0·0014**	**0·0012**
RESIDUAL	672	55·232	1712·77	720·54			
(b) Females only:							
NULL MODEL	492	49·645	1653·19	681·52			
AGECLASS	3	8·421	265·15	67·68	**< 0·0001**	**< 0·0001**	**< 0·0001**
YEAR	5	4·295	161·75	50·50	**< 0·0001**	**< 0·0001**	**< 0·0001**
A:Y	15	3·045	118·19	58·36	**0·0003**	**< 0·0001**	**< 0·0001**
RESIDUALS	469	33·884	1053·46	446·48			
(c) Males only:							
NULL MODEL	222	27·363	785·36	300·59			
AGECLASS	3	1·277	14·95	6·03	**0·0086**	0·2093	0·1103
YEAR	5	3·063	66·39	20·93	**< 0·0001**	**0·0015**	**0·0008**
A:Y	11	1·537	40·24	13·77	0·2104	0·3420	0·2460
RESIDUALS	203	21·348	659·31	256·27			

sizes are large and k is positively related to the mean, when k is negatively related to the mean, type II errors are probable even when sample sizes are large. Moreover, type I errors (which are arguably the more serious) are always likely when using classical linear models, especially when the distributions being compared differ markedly in their degree of aggregation (as indicated by k of the negative binomial distribution).

The two generalized linear models have a low probability of producing type I and type II errors when sample sizes are large, and Model 3 is slightly less likely to incorrectly accept the null hypothesis than Model 2 when sample sizes are small. Then, the probability of producing type II errors increases with the overall k of the aggregated distribution. The two GLMs do not always outperform the classical linear model; Model 1 appears to produce much fewer type II errors when sample sizes are small and k values and means are positively related, as is often the case with real parasite data (Shaw & Dobson 1995, Grenfell *et al.* 1995).

Analysis of the Soay Sheep data confirms the findings of the simulations. For example, the two GLMs and the maximum-likelihood model indicate that male August faecal egg counts do not vary between the four age-classes, whereas the classical linear model shows a highly significant age effect ($P < 0·009$). Thus, using the standard linear model we would have erroneously concluded that male faecal egg counts in August vary with age, perhaps because of increased acquired immunity in older males or parasite-induced host mortality. However, the GLM approach is not infallible; it sometimes fails to detect heterogeneities identified using non-linear maximum-likelihood methods (K. Wilson & D. J. Shaw, unpublished observation).
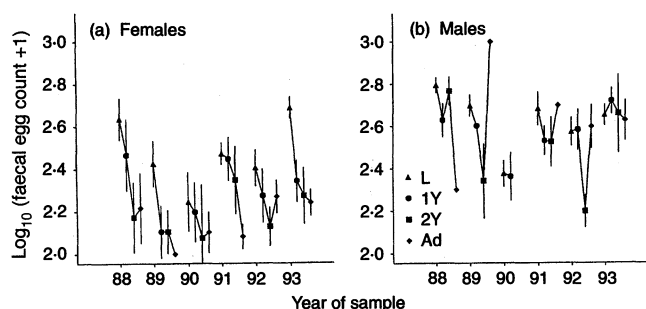
Fig. 7. Variation in August faecal egg counts in Soay Sheep on St Kilda. Within each year of sample, faecal egg counts ($\log_{10}$-transformed count +100) are shown for lambs ('L', aged *c.* 4 months), yearlings ('1Y', aged *c.* 16 months), two-year-olds ('2Y', aged *c.* 28 months) and adults ('Ad', aged $\geq$ 40 months). Means of $\log_{10}$-transformed counts are shown, together with their standard errors. Analyses are described in the main text and Table 2.

Pacala & Dobson (1988) have also cautioned against the use of classical linear regression models to identify heterogeneities in parasite loads, and show that the patterns that emerge with non-linear maximum-likelihood models can differ markedly from those of more conventional techniques, particularly when sample sizes are small. Gregory and Woolhouse (1992) have also examined the utility of various methods for estimating parasite abundance and aggregation and concluded that the average estimated geometric mean (the back-transformed mean of the logarithmic data) was independent of sample size, k value and population mean. However, the confidence limits around each of these averages are generally large, especially when sample sizes are less than 100 (see Figure 1(h) in Gregory & Woolhouse 1992). Thus, when sample sizes are small, comparisons between any two random samples taken from the negative binomial distribution are likely to have a high probability of producing errors, as demonstrated by the present study.

In conclusion, when sample sizes are very small ($n \simeq 20$), the difference between the three models is marginal, except in the probability of producing type I errors when the k values of the component distributions differ (in which case the classical linear model using log-transformed data fails most frequently). However, when sample sizes are greater than this ($n \simeq 100$), only the classical linear model produces significant numbers of type I and type II errors (cf. the central boxes for Models 1–3 in Fig 2, 3 and 5). Clearly, bootstrap simulations or non-linear maximum-likelihood models would provide greater accuracy in testing for heterogeneities in parasite burdens (Pacala & Dobson 1988, Fulford 1994, Williams & Dye 1994). However, both methods are impractical for complex model designs. In their absence, generalized linear models offer the best alternative, particularly if negative binomial errors can be explicitly defined. We have shown in this study that the marginally greater inconvenience of GLMs over conventional linear models is certainly worthwhile and we strongly urge their greater use by both ecologists and parasitologists.

Acknowledgements

References

Aitkin, M., Anderson, D., Francis, B. & Hinde, J. (1989) *Statistical Modelling in GLIM*. Clarendon Press, Oxford.

Anderson, R.M. (1974) Population dynamics of the cestode *Caryophyllaeus laticeps* (Pallas, 1781) in the bream (*Aramis brama* L.). *Journal of Animal Ecology* **43**, 305–321.

Anderson, R.M. (1978) The regulation of host population growth by parasitic species. *Parasitology* **76**, 119–157.

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

Anderson, R.M. & May, R.M. (1991) *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, Oxford.

Clutton-Brock, T.H., Illius, A.W., Wilson, K., Grenfell, B.T., MacColl, A.D.C. & Albon, S.D. (in press) Stability and instability in ungulate populations: an empirical analysis. *The American Naturalist* (in press).

Crawley, M.J. (1993) *GLIM for Ecologists.* Blackwell Scientific Publications, Oxford.

Crofton, H.D. (1971) A quantitative approach to parasitism. *Parasitology* **63**, 343–364.

Elliott, J.M. (1977) *Statistical Analysis of Samples of Benthic Invertebrates*, 2nd edn. Freshwater Biological Association, Ambleside.

Fulford, A.J.C. (1994) Dispersion and bias: can we trust geometric means? *Parasitology Today* **10**, 446–448.

Gregory, R.D. & Woolhouse, M.E.J. (1992) Quantification of parasite aggregation: a simulation study. *Acta Tropica* **54**, 131–139.

Grenfell, B.T., Price, O.F., Albon, S.D. & Clutton-Brock, T.H. (1992) Overcompensation and population cycles in an ungulate. *Nature London* **355**, 823–826.

Grenfell, B.T., Wilson, K., Isham, V., Boyd, H.E.G. & Dietz, K. (1995) Modelling patterns of parasite aggregation in natural populations: trichostrongylid nematode-ruminant interactions as a case study. *Parasitology* **111**, 5135–5151.

Gulland, F.M.D. (1992) The role of nematode parasites in Soay sheep (*Ovis aries* L.) mortality during a population crash. *Parasitology* **105**, 493–503.

Gulland, F.M.D. & Fox, M. (1992) Epidemiology of nematode infections of Soay sheep (*Ovis aries* L.) on St. Kilda. *Parasitology* **105**, 481–492.

Gulland, F.M.D., Albon, S.D., Pemberton, J.M., Moorcroft, P.R. & Clutton-Brock, T.H. (1993) Parasite-associated polymorphism in a cyclic ungulate population. *Proceedings of the Royal Society of London, Series B* **254**, 7–13.

Pacala, S.W. & Dobson, A.P. (1988) The relation between the number of parasites/host and host age: population dynamic causes and maximum likelihood estimation. *Parasitology* **96**, 197–210.

Pennycuick, L. (1971) Frequency distributions of parasites in a population of three-spined sticklebacks, *Gasterosteus aculeatus* L., with particular reference to the negative binomial distribution. *Parasitology* **63**, 389–406.

Shaw, D.J. (1994) *Distribution of macroparasites in naturally-fluctuating host populations.* PhD thesis, University of Cambridge, UK.

Shaw, D.J. & Dobson, A.P. (1995) Patterns of macroparasite abundance and aggregation in wildlife populations: a quantitative review. *Parasitology* **111**, S111–S153.

Southwood, T.R.E. (1978) *Ecological Methods*, 2nd edn. Chapman & Hall, London.

Venables, W.N. & Ripley, B.D. (1994) *Modern Applied Statistics with S-Plus.* Springer-Verlag, New York.

Williams, B.T. & Dye, C. (1994) Maximum likelihood for parasitologists. *Parasitology Today* **10**, 489–493.

Wilson, K. & Grenfell, B.T. (in press) Generalized linear modelling for parasitologists. *Parasitology Today* (in press).