Dealing with overdispersed count data in applied ecology

Shane A. Richards*

Department of Biological and Biomedical Sciences, University of Durham, South Road, Durham DH1 3LE, UK

Summary

1. The ability to identify key ecological processes is important when solving applied problems. Increasingly, ecologists are adopting Akaike's information criterion (AIC) as a metric to help them assess and select among multiple process-based ecological models. Surprisingly, however, it is still unclear how best to incorporate AIC into the selection process in order to address the trade-off between maximizing the probability of retaining the most parsimonious model while minimizing the number of models retained.

2. Ecological count data are often observed to be overdispersed with respect to best-fitting models. Overdispersion is problematic when performing an AIC analysis, as it can result in selection of overly complex models which can lead to poor ecological inference. This paper describes and illustrates two approaches that deal effectively with overdispersion. The first approach involves modelling the causes of overdispersion implicitly using compound probability distributions. The second approach ignores the causes of overdispersion and uses quasi-AIC (QAIC) as a metric for model parsimony.

3. Simulations and a novel method that identifies the most parsimonious model are used to demonstrate the utility of the two overdispersion approaches within the context of two ecological examples. The first example addresses binomial data obtained from a study of fish survival (as related to habitat structure) and the second example addresses Poisson data obtained from a study of flower visitation by nectarivores.

4. Applying either overdispersion approach reduces the chance of selecting overly complex models, and both approaches result in very similar ecological inference. In addition, inference can be made more reliable by incorporating model nesting into the selection process (i.e. identifying which models are special cases of others), as it reduces the number of models selected without significantly reducing the probability of retaining the most parsimonious models.

5. Synthesis and applications. When data are overdispersed, inference can be improved by either modelling the causes of overdispersion or applying QAIC as a metric for model parsimony. Inference can also be improved by adopting a model filtering procedure based on how models are nested. The general simulation approach presented in this paper for identifying the most parsimonious model, as defined by information theory, should help to improve our understanding of the reliability of model selection when using AIC, and help the development of better selection rules.

Key-words: AIC, model selection, overdispersion, QAIC, quasi-likelihood

Introduction

Abstracting ecological complexity by identifying key biological processes, and the factors that affect them, is important when solving applied ecological problems (e.g. Hilborn & Mangel 1997; McPherson & DeStefano 2003; Clark 2007), as the effectiveness of a management action is likely to depend upon the processes that dominate a system (Walters 1986). Hence, it is important that ecologists should be able to compare their data reliably with multiple, process-derived, ecological models. It is common for ecologists to choose the factors and processes that best explain their data using stepwise multiple regression approaches; however, biased parameter estimates and inconsistencies among model selection algorithms (Whittingham et al. 2006) have led many ecologists to adopt
alternative model selection techniques (Johnson & Omland 2004; Hobbs & Hilborn 2006). Akaike’s information criterion (AIC), which is based on information theory, is a common metric used by ecologists to evaluate and select among alternative ecological models (Burnham & Anderson 2002; Richards 2005; Stephens et al. 2005). In this paper I shall consider only information-theoretic approaches for selecting ecological models, although other model selection approaches also exist that are based on subjective rules of thumb (Ginzberg et al. 2004) and Bayesian statistics (Hoeting et al. 1999; Hobbs & Hilborn 2006).

Many ecological problems involve the analysis of count data that exhibit variation greater than that predicted by the stochastic component of a model. For example, consider the case where insect abundance is assessed across a range of habitats using pit-fall traps. Data from this type of monitoring could be assumed to be Poisson distributed. If this assumption is true, then the variation in insect abundance among pit-fall traps within each habitat should be very similar to the mean abundance within the habitat; however, the variance in these types of data is often observed to be greater than the mean by a factor of two or more (Bliss & Fisher 1953). Such data are referred to as overdispersed with respect to the model distribution. Overdispersion may be due to the model not accounting for important covariates, or a lack of independence among study subjects or treatments (Williams 1975; Eberhardt 1978; Cox & Snell 1989). Correctly quantifying variation can be very important in applied ecology (Clark 2007). For example, quantifying how individual fish differ in their learning responses to being caught will influence the effectiveness of catch-and-release regulations for promoting the conservation of a fishery while maintaining high angling quality. If most fish learn quickly to avoid hooks after being caught and subsequently released, then angling quality will drop quickly, whereas quality will be retained if many fish fail to change their feeding habits after being caught. Askey et al. (2006) identified high variation in feeding responses to being caught among rainbow trout (Oncorhynchus mykiss), as the variation in daily catches of trout was greater than that predicted by a binomial distribution, which assumed no individual variation. Unfortunately, published studies that adopt model selection often do not report if their data are overdispersed with respect to their best-fitting models. Ignoring overdispersion is problematic, as it can cause overestimates of the precision of model parameters which can lead to the selection of overly complex models (Anderson, Burnham & White 1994), resulting in poor ecological inference (Burnham & Anderson 2002).

Rejection of a goodness-of-fit test can be used to detect the presence of overdispersion (Burnham & Anderson 2002). If overdispersion is detected for all models considered, then there are four approaches to model selection when AIC is used. First, its presence can be ignored and the analysis continued anyway. The second approach is to collect data on additional independent variables that explain the inflated variation and then repeat the AIC analysis. This approach is preferable, as additional information can enhance ecological insight (Stephens et al. 2007), although one must also be wary of false inference as a result of data dredging (Burnham & Anderson 2002). If additional data are unavailable, then a third approach is to model implicitly the causes of the inflated variation (e.g. missing covariates or non-independence) and then apply AIC to the new models (Lindsey 1999). Lastly, quasi-likelihood theory can be applied, modifying the AIC value associated with each model. Model selection is then based on the quasi-AIC (QAIC) values (Lebreton et al. 1992; Anderson et al. 1994). Although QAIC has general applicability, it is applied typically to mark–recapture studies. In this paper I will focus upon the last two approaches.

Surprisingly, despite an increase in the use of AIC, and to a lesser degree QAIC, it is still unclear when these approaches are likely to perform well (Stephens et al. 2007). For example, in the literature it is most common to see all models selected that have an AIC (QAIC) value within a threshold of the minimum AIC (QAIC) calculated. However, it is not clear if a single threshold is appropriate for all types of studies or models (Burnham & Anderson 2002: 71). Also, Lindsey (1999) suggested that when data are overdispersed and additional data are unavailable, then the causes of overdispersion should be modelled implicitly whenever possible (i.e. the third approach described above); however, Liang & McCullagh (1993) and Burnham & Anderson (2002) suggest that the quasi-likelihood approach is likely to be adequate for many situations.

Part of the reason for the uncertainty in AIC performance is a lack of clarity in the literature regarding what exactly defines the best model according to AIC. To address this ambiguity, Richards (2005) used computationally intensive simulations to demonstrate how well AIC estimates the relative, expected Kullback–Leibler distance (EKLD) of a model, which is a measure of the mean discrepancy between the model and the unknowable truth (see next section for more details). This study provided the first rigorous test of AIC using ecological examples where the test was based on the theory from which AIC was developed. For a variety of ecological problems, Richards (2005) demonstrated that if the best model is defined as the one with the lowest EKLD, then a selection threshold of at least six was required for it to be selected with 95% confidence using AIC differences.

In this paper, I extend the work of Richards (2005) and demonstrate how model selection can best be performed when count data are overdispersed. First, I review briefly the theory behind AIC and QAIC. I then show how overdispersion can be modelled using the beta-binomial and negative-binomial distributions, for binomial and Poisson data, respectively. Although the application of these two distributions to the analysis of biological data is not new (Bliss & Fisher 1953; Williams 1975; May 1978), they are under-utilized in applied ecology (Lindsey 1999). I illustrate the utility of both distributions by presenting plausible ecological studies and appropriate approximating models derived from simple ecological theory. I then demonstrate how AIC and QAIC can be best used to select the most parsimonious models, as defined by information theory. In particular, I demonstrate how incorporating model nesting into the selection process.
can reduce the number of models selected and reduce the chance of poor inference. Finally, I summarize my findings and give some recommendations.

**Model selection and overdispersion**

Model selection first involves proposing a set of models (prior to data analysis) that predict outcomes likely to be consistent with the data collected. Often, these models are based on relevant ecological theory (Richards 2005; Hobbs & Hilborn 2006). Suppose an ecological study is performed that involves the collection of count data. For typical studies the number of possible outcomes is likely to be very large or even infinite, where an outcome is the set of all counts observed during the study. Let \( p_i \), be the probability that if the study were performed, then the outcome indexed by \( i \) would be observed \((\sum p_i = 1)\). In this paper, the probability distribution defined by the \( p_i \), which I denote \( p \), is referred to as the truth. Suppose a stochastic model is proposed which predicts that outcome \( i \) will be observed with probability \( \pi_i \). Let the distribution of predicted probabilities be denoted \( \pi \). An information-theoretic measure of the difference between the truth and the approximating model is the Kullback–Leibler distance (KLD):

\[
I(p, \pi) = -\sum_i p_i \log \left( \frac{p_i}{\pi_i} \right) \quad \text{eqn 1}
\]

(Kullback & Leibler 1951). \( I \) is often interpreted as the information lost when the truth \( p \) is approximated by \( \pi \). The smaller the value of \( I \), the better the model approximates the truth.

Models typically have parameters that require values in order for them to make quantitative predictions. Let \( \theta \) denote the set of model parameters (i.e. the \( \pi \) depend on \( \theta \)). The best parameter values for a model according to information theory are those that minimize \( I \). Unfortunately, because the truth, \( p \), is unknowable in realistic cases, it is not possible to apply equation 1 directly to find the best \( \theta \). However, parameter values can be estimated readily by fitting the model to the data using maximum likelihood. For realistic ecological studies the truth will be too complex to model exactly, and so we would expect \( I(p, \pi) > 0 \) for all parameterizations, \( \theta \).

Suppose the study was repeated an infinite number of times and the processes generating the data did not change from one study to the next. If the model’s parameters were re-estimated each time using maximum likelihood, then the EKLD of the model would be:

\[
E^* \{ I(p, \pi) \} = -\sum_j p_j I(p, \pi|\theta_j) \quad \text{eqn 2}
\]

where \( \theta_j \) is the set of maximum likelihood parameter estimates, given outcome \( j \). A model having a low EKLD can be considered a parsimonious model because, on average, it has the lowest KLD when fitted to data. The proposed model having the lowest EKLD is referred to as the best EKLD model and is the model we wish to identify. Akaike (1973) derived an approximate relationship between the maximum likelihood of a model and the EKLD. Suppose a study resulted in outcome \( j \). For a proposed model, \( M \), its AIC value is defined as:

\[
\text{AIC}(M) = -2 \ln L(\hat{\theta}_j) + 2K, \quad \text{eqn 3}
\]

where \( K \) is the number of model parameters estimated using maximum likelihood and \( \ln L(\hat{\theta}_j) \) is the maximum log-likelihood of the model, given outcome \( j \). If outcome \( j \) is a set of \( \ell \) independent counts, then:

\[
\ln L(\hat{\theta}_j) = \sum_k \ln L(\hat{\theta}_j | y_{j,k}), \quad \text{eqn 4}
\]

where \( \ln L(\hat{\theta}_j | y_{j,k}) \) is the maximum log-likelihood of the model, given the \( k \)th observed count, \( y_{j,k} \).

It can be shown that AIC estimates twice the model’s relative EKLD:

\[
\text{AIC}(M) \approx 2 \{ E^* \{ I(p, \pi) \} - c \}, \quad \text{eqn 5}
\]

where \( c = \sum_k p_k \ln p_k \) is a constant common to all models that depends only on \( p \) (Richards 2005). Hence, models with a low AIC value are more likely to be the best EKLD model. An important point is that AIC is only an estimate of the relative EKLD, because of sampling error, a model not having the lowest AIC value may in fact be the best EKLD model. Rather than simply retaining the best AIC model (i.e. the model with the lowest AIC value), it has been suggested that all models having an AIC value within a threshold of the best AIC model should be retained and used for inference. For convenience, a \( \Delta \)-value is often calculated for each model, which is simply the AIC value of the model minus the lowest AIC value. Burnham & Anderson (2002: 70) recommend the following widely adopted rule of thumb for model selection. All models with a \( \Delta \)-value < 2 should be used for inference. Models with a \( \Delta \)-value between about 4 and 7 are less likely to be the best EKLD model but probably should not be discounted. Models with a \( \Delta \)-value > 10 are extremely unlikely to be the best EKLD model and can be safely discounted.

The rule of thumb presented above is problematic, as it may result in the selection of overly complex models. To see why, consider the case where a model is reduced to a simpler model by setting one of its parameters to a constant (possibly zero). In this case, I refer to the simpler model as being nested within the more complex model. If the additional parameter provides little or no increase in fit (i.e. both models have near equivalent maximum likelihood), then the more complex model will have a \( \Delta \)-value equal to, or slightly less than, 2. Clearly, in this case it makes no sense to select the more complex model and make inference from it, as nothing is explained by the additional complexity. This simple result has been noted before (Burnham & Anderson 2002: 131), but whether or not it has occurred is rarely reported in published AIC analyses. One possible way to avoid this problem of selecting overly complex models is to apply the above rule of thumb but, in addition, select a model only if it has a \( \Delta \)-value less than the \( \Delta \)-values of all of its simpler nested
models. In this paper I examine the consequences of this modified rule.

If all the proposed models fail to explain the degree of variation in the data, then the rule of thumb proposed by Burnham & Anderson (2002) may result in the selection of overly complex models (Anderson et al. 1994). The degree of overdispersion can be quantified by the variance inflation factor (VIF), which is the variation observed in the data divided by the variation expected by the most complex proposed model (Lindsey 1999). Lebreton et al. (1992: 107) cautiously suggested the use of quasi-likelihoods when assessing overdispersed data using the information-theoretic framework. The quasi-likelihood is the likelihood divided by an estimate of the VIF. The QAIC value of a model is calculated by replacing the likelihood with the quasi-likelihood:

\[
\text{QAIC}(M) = -\frac{2}{\hat{v}} \ln L(\hat{\theta}_c) + 2K, \quad \text{eqn 6}
\]

where \(\hat{v}\) is the estimated VIF. The VIF can be estimated by comparing the fit of the most complex model proposed (C) with the saturated model (S), which is the most complex model that could be fitted to the available data. The estimate is given by:

\[
\hat{v} = \frac{2}{\text{df}} [\ln L(\hat{\theta}_{S,j}) - \ln L(\hat{\theta}_{C,j})], \quad \text{eqn 7}
\]

where \(L(\hat{\theta}_{C,j})\) and \(L(\hat{\theta}_{S,j})\) are the maximum likelihoods of the most complex model proposed and the saturated model, respectively, and \(\text{df}\) (degrees of freedom) is the additional number of estimated parameters needed to specify the saturated model. When the data are overdispersed, model selection can be performed as described above but replacing AIC with QAIC (Burnham & Anderson 2002).

Testing AIC estimates and model selection

Identifying which of the proposed models is truly the most parsimonious according to information theory (i.e. the model having the lowest EKLD) is not possible for real cases because the truth, \(p\), is unknown. However, by considering plausible \(p\) and using simulations, it is possible to explore AIC performance. Equation 1 suggests a simple, but computationally intensive, way to estimate the KLD of a model. Suppose a plausible \(p\) is proposed and a random sample from \(p\) results in study outcome \(j\). For sufficiently large \(Z\), the KLD of a model, given outcome \(j\), can be estimated using Monte Carlo integration:

\[
\hat{I}(p, \pi | j) = \frac{1}{Z} \sum_{z=1}^{Z} [\ln p(z) - \ln \pi(z|j)], \quad \text{eqn 8}
\]

where \(p(z)\) is the \(z\)th randomly generated study outcome based on the distribution \(p\). Similarly, \(c\) can be estimated using:

\[
\hat{c} = \frac{1}{Z} \sum_{z=1}^{Z} \ln p(z), \quad \text{eqn 9}
\]

The results presented in this paper were generated using \(Z = 10^4\), which gave estimates of \(I(p, \pi | j)\) and \(c\) to 1 decimal place (dp). The EKLD of a model, \(E_k(I)\), can be estimated by averaging its estimated KLD, \(\hat{I}\), across many simulated outcomes, \(j\). Here, 1000 repeated simulations were sufficient to estimate expectations to 1 dp (i.e. estimating \(E_k(I)\) involved \(10^4\) evaluations of the square-bracketed term in equation 8). Given estimates of \(E_k(I)\), it is possible to identify the best EKLD model, and given the estimate for \(c\), the AIC estimate given by equation 5 can also be assessed. A similar approach has been used to investigate model selection in a Bayesian framework (Hoeting et al. 1999), but surprisingly it has not been applied previously to AIC. Monte Carlo integration allows AIC performance to be examined for much more complex and realistic ecological studies than considered by Richards (2005).

Knowing the best EKLD model it is possible to explore how best to use AIC values when selecting models. A good set of selection rules will select the best EKLD with high probability, and avoid the selection of overly complex models. In this paper I investigate the model selection rule of thumb proposed by Burnham & Anderson (2002) and also its proposed modification that accounts for model nesting. In addition, the quasi-likelihood approach is evaluated by comparing the models selected by an AIC analysis with those selected by a QAIC analysis. If QAIC is useful, then the probability of selecting a model that implicitly incorporates the sources of overdispersion (e.g. a model derived from a compound probability distribution) when selection is based on AIC values, will be similar to the probability of selecting the equivalent model that ignores sources of overdispersion when selection is based on QAIC values.

Binomial count data

In many ecological studies subjects are grouped among replicated treatments and each subject may exhibit a binary response to the treatment (e.g. survive or die). If all subjects of a treatment have the same probability of exhibiting a positive response (or success), then the number of successes among replicated treatments will follow a binomial distribution. However, in many cases the variation observed among replicates is often greater than that predicted by a binomial distribution. In this situation, one way to model overdispersion is to use compound probability distributions. Suppose an experiment involves \(n\) subjects being exposed to a treatment, and let \(p\) be the probability each individual in the replicate treatment exhibits a positive response. In addition, suppose that \(p\) varies randomly among replicated treatments due to unknown covariates and the variation in \(p\) can be described by the probability density function \(f(p)\). In this case, the probability \(y\) of the \(n\) subjects will exhibit a positive response is:

\[
\Pr(y) = \int_0^1 f(p) \left(\begin{array}{c} n \\ y \end{array}\right) p^y(1-p)^{n-y} dp. \quad \text{eqn 10}
\]
In this paper I assume that \( f(p) \) can be well described by a smooth parametric function. Experiments involving a randomized design may best fit this assumption. The beta distribution is a simple and relatively flexible choice for \( f(p) \). In this case, \( y \) is described by the beta-binomial distribution, which may be expressed in multiple ways. Here, I consider the beta distribution parameterized by its mean, \( \bar{p} \), and a non-negative dispersion coefficient, \( \phi \), which quantifies the variation in \( p \) among replicated treatments. This parameterization is convenient because hypotheses are expressed typically according to how the treatment affects the mean probability of success, \( \bar{p} \) (Williams 1975). The beta distribution, parameterized by positive \( \bar{p} \) and \( \phi \), is given by:

\[
f(p; \bar{p}, \phi) = \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} p^{a-1}(1 - p)^{b-1}, \quad \text{eqn 11}
\]

where \( \Gamma(x) \) is the complete gamma function, \( a = \bar{p}/\phi \), and \( b = (1 - \bar{p})/\phi \). The beta distribution has variance \( \bar{p}(1 - \bar{p})\phi/(1 + \phi) \). The beta-binomial distribution has mean \( n\bar{p} \) and variance \( vn\bar{p}(1 - \bar{p})/((1 + \phi)(1 + 2\phi)) \), where \( v \) is the VIF, given by \( v = 1 + (n - 1)\phi/((1 + \phi)(1 + 2\phi)) \) (Cox & Snell 1989).

Let \( \theta \) be the set of model parameters needed to calculate \( \bar{p} \) plus the dispersion coefficient \( \phi \). Substituting the beta distribution into equation 10 gives the likelihood of \( \theta \), given that \( y \) successes were observed among the \( n \) subjects in the treatment; namely:

\[
L(\theta|y) = \frac{\Gamma(n+1)\Gamma(a+b)\Gamma(y+a)\Gamma(n-y+b)}{\Gamma(y+1)\Gamma(n-y+1)\Gamma(a)\Gamma(b)\Gamma(n+a+b)}, \quad \text{eqn 12}
\]

As \( \phi \) approaches zero, equation 12 approaches the binomial distribution. Assuming independent observations among treatments, equation 12 can be applied to all study treatments, and their product is the overall likelihood of the study. Note that statistical software is not necessary to perform any of the analyses presented here; for example, Microsoft Office Excel provides the natural logarithm of the complete gamma function, \textsc{gammaln}(), and maximum likelihood parameter estimates can be obtained using the \textsc{solver} add-in.

Suppose a study has \( R \) replicates in total and the VIF is to be estimated in order to calculate QAIC values. In this case, the saturated model will have \( R \) estimated parameters because \( p \) can be estimated independently for each replicate. For a replicate with \( n \) subjects of which \( y \) exhibit successes, the best estimate for \( p \) is \( y/n \), which implies the following maximum likelihood term for the saturated model:

\[
L(\hat{\theta}_s|y) = L\left(\hat{p} = \frac{y}{n} \right) = \frac{n!y^y(n-y)^{n-y}}{y!(n-y)!n^n}, \quad \text{eqn 13}
\]

The total maximum log-likelihood for the saturated model is calculated by summing in \( L(\hat{\theta}_s|y) \) across all \( R \) replicates. The df for the VIF estimate calculation is \( R - K_C \), where \( K_C \) is the number of parameters estimated by the most complex binomial model proposed.

**BINOMIAL EXAMPLES**

Consider the case where the objective was to understand how habitat structure affects the survival of kelp perch (\textit{Brachyistius frenatus}). To address the question the following experiment was conducted where \textit{B. frenatus} were exposed to the predatory kelp bass (\textit{Paralabrax clathratus}) in different environments (see Anderson 2001 for results from experiments that produced data consistent with the simulations presented here). A single randomly chosen bass and \( n \) perch were placed into 5000-L pools containing a number of equal-sized bunches of giant kelp (\textit{Macrocystis pyrifera}). Four treatments were considered, defined by the number of kelp bunches in the pool (\( x = 1, 2, 3 \) or 4 bunches), and each treatment had \( r \) pool replicates. After time \( t \) the number of perch that had avoided being eaten in each of the \( R = 4r \) pools was noted. This experiment could result in overdispersion among the replicated treatments because of differences among the bass or perch (e.g. their relative body size) or differences among the pools (e.g. kelp structure).

Suppose the average rate at which bass capture and eat perch when there are \( x \) bunches of kelp in the pool is \( \tilde{E}(x) \). Assuming that prey handling time is negligible compared with the time to stalk and capture prey, the probability that each perch will survive the experiment is \( \tilde{p} = \exp[-\tau\tilde{E}(x)] \). To test AIC and QAIC I proposed that the true capture rate was given by \( \tilde{E}(x) = \alpha(1 + \beta x) \), where \( \alpha \) and \( \beta \) were positive constants. Hence, increasing perch refuges increased perch survival, and the effectiveness of the kelp was quantified by \( \beta \). I also proposed that the number of perch surviving each replicate was described by the beta-binomial distribution with dispersion coefficient, \( \phi \), common to all treatments.

The following six models, denoted \( M_1 - M_6 \), were proposed to explain the data. The first three models assumed that overdispersion was absent (i.e. were binomial models) and the mean probability of survival for each model was \( M_1: \tilde{p} = \exp(-\tau\alpha); M_2: \tilde{p} = \exp(-\tau\alpha(1 + |\beta|)); \) and \( M_3: \tilde{p} = \exp(-\tau\alpha + \beta x)/(1 + \exp(-\tau\alpha + \beta x)) \). Model \( M_1 \) assumed \( \alpha > 0 \) and \( M_2 \) also assumed \( \beta > 0 \); whereas \( M_3 \) had no restrictions on \( \alpha \) and \( \beta \). Model \( M_1 \) assumed that the abundance of kelp did not affect survival, whereas \( M_2 \) and \( M_3 \) assumed that it did. Models \( M_4 - M_6 \) were equivalent to models \( M_1 - M_3 \) except that they allowed for overdispersion (i.e. were beta-binomial models) with \( \phi \) constant across the four treatments. Models \( M_1 - M_6 \) have \( K = 1, 2, 2, 2, 3 \) and 3 estimated parameters, respectively. Models \( M_5 \) and \( M_6 \) are the most complex, parameterized by \( \theta = \{\alpha, \beta, \phi\} \). In this example, model \( M_5 \) can describe the truth exactly. When estimating \( \gamma \) for the QAIC analyses I let \( M_1 \) be the most complex model (df = \( R - 2 \)).

Three scenarios (A–C) were considered to investigate the general utility of QAIC. Model nesting was not considered during selection in order to simplify the analysis of the results. For scenario A, 20 prey fish were placed into each tank, and the four kelp treatments each had 10 replicates, giving \( R = 40 \) replicates in total. I assumed that the truth was described by \( \theta = 0.1 \), which resulted in highly overdispersed data (\( \gamma = 2.73 \) for all four treatments). I also set \( t = 15 \) h, \( \alpha = 0.075 \) h\(^{-1}\) and...
\( \beta = 0.5 \) bunch\(^{-1} \), which resulted in kelp abundance having a strong positive effect on mean prey survival (\( \bar{p} = 0.47 \) and \( \bar{p} = 0.69 \) for \( x = 1 \) and \( x = 4 \) kelp bunches, respectively). For this scenario, the effect of kelp on survival and the presence of overdispersion were strong enough so that model M5 was the best EKLD model (Fig. 1a1). Applying QAIC resulted in very similar probabilities of selecting models M1–M3 compared with selecting models M4–M6, respectively, when AIC was applied (Fig. 1a2). Hence, in this case QAIC correctly described the effect of overdispersion without having to model it. For this scenario, applying a \( \Delta \)-threshold of \( \approx 6.5 \) was required to select the best EKLD model, M5, 95% of the time for the AIC analysis, whereas a \( \Delta \)-threshold of \( \approx 6 \) was required to select the corresponding model, M1, 95% of the time for the QAIC analysis (Fig. 1b2).

For scenario C, I reset \( \beta = 0.5 \) bunch\(^{-1} \), and set \( \phi = 0.025 \), which described the situation where the effect of kelp on survival was strong and overdispersion was weak (\( v = 1.22 \)). Model M5 was the best EKLD model but was only marginally better than model M2, which ignored overdispersion (Fig. 1c1). In this scenario the overdispersion parameter typically provided only a minimal increase in fit, and as a result, AIC overestimated the relative EKLD of the models that included overdispersion (Fig. 1c1). Although M5 was the best EKLD model, AIC favoured selection of the corresponding simpler model, M2, that ignored overdispersion (Fig. 1c2). The conclusion regarding whether or not kelp was important for prey survival was nearly always the same when either an AIC or a QAIC analysis was performed (Fig. 1c2).

**Poisson count data**

Another commonly encountered situation is where study subjects may experience zero or more events during the study. Data from this type of experiment can be described by a Poisson distribution. Poisson data are also often overdispersed. Even though different subjects may be exposed to the same treatment, they may experience events at different rates because of uncontrolled covariate variation among subjects or treatments. Events may also not be independent,
but depend upon the outcome of previous events. Suppose for a given treatment a study subject experiences independent events at constant rate \( \lambda \), drawn from a probability density function \( f(\lambda) \). The probability the subject will experience \( y \) events during period \( \tau \) is:

\[
\Pr(y) = \int_{0}^{\infty} f(\lambda) \frac{e^{-\lambda \tau}(\lambda \tau)^y}{y!} d\lambda.
\]  

eqn 14

In this paper I assume that \( \tau \) has units of time, but these results also apply to spatial problems. For example, \( y \) may be the number of objects or events counted in some habitat of area \( \tau \).

It is convenient to assume that \( f(\lambda) \) is the gamma distribution (Lindsey 1999) parameterized by its mean, \( \bar{\lambda} \), and dispersion coefficient, \( \phi \):

\[
f(\lambda;\bar{\lambda},\phi) = \frac{\lambda^{a-1} e^{-\lambda \phi}}{\Gamma(a)},
\]  

eqn 15

where \( a = \frac{\bar{\lambda}}{\phi} \) and \( b = 1/\phi \). The gamma distribution has variance \( \bar{\lambda} \bar{\phi} \). Substituting equation 15 into equation 14 gives the negative-binomial distribution, which has mean \( \bar{\lambda} \) and variance \( \bar{\lambda} + \bar{\lambda}^2/\phi \), where \( \bar{\lambda} = 1 + \lambda/b \). Let \( \theta \) be the set of model parameters needed to calculate \( \bar{\lambda} \) plus the dispersion coefficient \( \phi \). The likelihood of \( \theta \), given that \( y \) events are observed during time \( \tau \), is:

\[
L(\theta|y) = \frac{\Gamma(y + a)}{\Gamma(a+1)\Gamma(y+1)} \left( \frac{b}{b + \tau} \right)^a \left( \frac{1}{1 + b/\tau} \right)^y.
\]  

eqn 16

Equation 16 approaches the Poisson distribution as \( \phi \) approaches zero. Assuming independent observations among subjects, equation 16 can be applied to all study subjects, and their product is the overall likelihood of the study.

For Poisson data, the VIF can be estimated by considering the following saturated model, which assumes that the rate of events may differ for each replicate. If \( y \) events were observed for a replicate, then the best estimate for the rate of events associated with the replicate is \( \hat{\lambda} = y/\tau \). The maximum likelihood associated with the replicate is:

\[
L(\theta|y) = \frac{e^{-y/\tau} (y/\tau)^y}{y!}.
\]  

eqn 17

The maximum log-likelihood for the saturated model is the log of equation 17 summed across all replicates, making \( R \) the number of estimated parameters.

**POISSON EXAMPLES**

Simple foraging theory suggests that nectarivores should visit plants (or inflorescences) at a rate proportional to the number of open flowers they contain and the rate their flowers produce nectar. Such foraging will tend to equalize nectar volume among all flowers, thereby preventing highly rewarding flowers, which could be exploited by other nectarivores (Possingah 1992). To test this theory, consider a study where bees were observed visiting the perennial plant, blue columbine (Aquilegia breville), in an open forest. Plants were categorized according to flower number (\( f = 1–4 \)) and whether or not they were shaded (\( s = 1 \) or \( 0 \)), with the expectation that shaded plants produced nectar at a lower rate; \( r \) plants from each of the eight categories (\( R = 8 \)) were chosen at random within the forest, observed for \( \tau = 10 \) min, and the number of bee visits to each plant noted. This study could be used to test if flower number or shading status affects the likelihood of pollination. Data from this study might be overdispersed because visits are unlikely to be independent. Bees often move among neighbouring plants, and so the rate at which a plant is visited may be influenced strongly by local plant density or other habitat characteristics (Heinrich 1979).

To compare QAIC with the implicit modelling approach I considered the scenario where the true mean rate of arrival to \( f \)-flowered plants with shading status \( s \) was \( \hat{\lambda}, f, s \), where \( \alpha \), was the rate of arrival to one-flowered plants, and \( \beta \), described how flower number affected bee arrival rate (\( \beta > 0 \) implied a positive affect). I also assumed that the number of bee visits within plant categories were negative-binomial and overdispersed, described by a constant dispersion coefficient, \( \phi \). Richards (2005) considered a similar study, but did not include overdispersion.

Based on the foraging theory described above, I considered eight candidate models, denoted M_1–M_8, which differed in their assumptions regarding flower and shading effects on the mean rate of plant visitation, and the presence of overdispersion. Models M_1–M_4 assumed no overdispersion (i.e. were Poisson models). Model M_1 assumed \( \hat{\lambda}, f = \alpha ( \text{both flower number and shading status had an effect on plant visitation}; K = 1) \), model M_2 assumed \( \hat{\lambda}, f = \alpha, s \), \( ( \text{only shading status had an affect}; K = 2) \), model M_3 assumed \( \hat{\lambda}, f = \alpha, f \), \( ( \text{visitation rate was linearly proportional to flower number}; K = 1) \) and model M_4 assumed \( \hat{\lambda}, f = \alpha, s, f \), \( ( \text{visitation rate was linearly proportional to flower number, and differed among shaded and unshaded plants}; K = 2) \). Models M_5–M_8 were equivalent to models M_1–M_4, respectively, but incorporated overdispersion (i.e. were negative-binomial models), and \( \phi \) was the same for all eight plant categories.

As a baseline scenario I considered a truth where bees visited shaded plants 10% less often than unshaded plants. Realistically, I generated the data using a model not among the candidate models by setting \( \beta_0 = \beta = 0.8 \), which might describe the situation where plant clumping lessened the effect of flower number on bee arrivals. Unless stated otherwise, parameters were set to \( \alpha_0 = 0.5 \), \( a_1 = 0.45 \), \( \phi = 0.1 \) and \( r = 10 \). For this scenario models M_1 and M_8 were both estimated to have the lowest EKLD (Table 1). The \( \Delta \)-values of models M_1–M_8, when using QAIC were very similar to the \( \Delta \)-values of M_1–M_4, respectively, when using AIC (Fig. 2a). This result was also observed when bees visited shaded plants 20% less often and M_5 was the best EKLD model (Fig. 2b). Hence, QAIC was again able to estimate well the effect of overdispersion on model selection.

Simulations were performed to investigate the utility of selecting models based on their \( \Delta \)-values when performing an
AIC analysis. Given the results of the previous section and the findings of Richards (2005), models were selected when they had $\Delta$-values $\leq 6$. To investigate the generality of this selection approach I considered six scenarios that differed in the true strength of shading (i.e. $\alpha_1$ relative to $\alpha_0$), sample size ($r$) and the degree of overdispersion ($\phi$), and re-simulated each scenario 1000 times. Specifically, I investigated (1) on average, how many models were selected; (2) how often an AIC analysis would select the best EKLD correctly; and (3) how often a model was selected that incorporated a shading effect it inferred incorrectly that shading had a positive effect on bee visitation. These three statistics were also recorded when, in addition, the selection process involved disregarding models that had a $\Delta$-value greater than any of the simpler models within which they were nested.

Ignoring model nesting resulted typically in the best EKLD model being selected over 96% of the time and, on average, each simulation resulted in about two models being selected. Unfortunately, if shading effect was relatively weak, there were few data or overdispersion was high, then 20% or more of the simulations resulted in the selection of a model that inferred incorrectly a positive shading effect (Table 1). Disregarding the more complex nested models that did not have a lower AIC value reduced the probability of incorrect inference to 0.07 or less (Table 1). Consideration of model nesting also reduced the number of models selected, particularly when incorrect inference was more likely (Table 1). The cost of applying the model nesting rule was that it reduced the probability of selecting the best EKLD, particularly when the strength of shading was intermediate, and model $M_8$ was the best EKLD model. However, in these cases when the best EKLD model, $M_8$, was not selected, the rule always selected the more conservative second-best EKLD model, $M_7$.

### Discussion and recommendations

This paper demonstrates two practical ways to deal with overdispersion to improve the selection of parsimonious ecological models. The first approach involves implicit modelling of the processes that lead to overdispersion. Here I have focused on modelling uncontrolled (or uncontrollable) variation among replicates using compound probability distributions. Alternatively, it may be possible to model explicitly non-independent responses among subjects within a treatment (see Edwards 1960 for an example involving binomial data). Teasing apart which processes contribute most to overdispersion may be possible, depending on the type of data collected. For example, it may be possible to test for non-independence if the times of events are recorded in addition to their number. Martin et al. (2005) provide some examples of how to build appropriate models for situations where overdispersion is represented by inflated zero counts, which is observed commonly in survey data. The advantage of modelling and quantifying the degree of variation (e.g. via the estimate for $\phi$) is that the analysis may lead to a greater understanding of ecological mechanisms, which may be important for prediction (Lindsey 1999; Martin et al. 2005). For example, Fox & Kendall (2002) demonstrated that understanding the mechanisms generating variation among individuals can affect the predicted risk of extinction when performing a population viability analysis. Ruel & Ayres (1999) give examples of how different sources of environmental variation can also alter predictions for a wide variety of nonlinear ecological and evolutionary processes.

Unfortunately, it is not always trivial or possible to model the sources of variation implicitly using compound probability distributions and derive a likelihood equation for an AIC analysis. This is common for mark–recapture studies, where variation can occur among individuals in both their survival and capture probabilities (Anderson et al. 1994; Askey et al. 2006). Here, I have shown that selecting among models that ignore overdispersion, based on their QAIC values, results in very similar inference to an AIC analysis when overdispersion is modelled. These results provide the first quantitative support for the use of QAIC, despite it being well known that estimates of the VIF, upon which QAIC is dependent, are often inflated (Anderson et al. 1994; Lindsey 1999).

The results presented here are generally consistent with the selection rule of thumb proposed by Burnham & Anderson (2002); however, all models having a $\Delta$-value $\leq 6$ should be
Involving the models M1–M4 (see text for model details). The legend both simulated 1000 times with

\[
\begin{align*}
\alpha &= 0.5, \\
\beta &= 0.8 \\
\phi &= 0.1 \quad (r = 2.0)
\end{align*}
\]

Fig. 2. Model selection results for two nectarivore foraging studies, both simulated 1000 times with \( r = 10, \) \( \alpha = 0.5, \) \( \beta = 0.8 \) and \( \phi = 0.1 \) (\( r = 2.0 \)). (a) A scenario where the effect of shading on bee arrivals is weakly negative (arrivals are 10% less frequent to shaded plants, \( \alpha = 0.45 \)) and (b) a scenario where the shading effect is strong (arrivals are 20% less frequent, \( \alpha = 0.4 \)). For (a) both model M7 and M8 are the best EKLD (Table 1), whereas for (b), model M4 is the best EKLD model. Presented is the proportion of times each model was selected based simply on a threshold, when using either an AIC analysis involving all eight models M1–M8 or a QAIC analysis involving the models M1–M4 (see text for model details). The legend in (a) also applies to (b) (only selection statistics for the four best EKLD models are presented).

should be selected only if its AIC (QAIC) value is less than the AIC (QAIC) value of all the simpler models within which it is nested. Adopting this rule will increase the chance of correctly identifying whether a factor has a positive or negative effect when the factor is included in a selected model (Table 1). Care must still be taken, however, when inferring the magnitude of a factor’s effect if it is selected because it may be inflated due to model selection bias (Whittingham et al. 2006). The cost of using this more conservative rule is that it will reduce the chance of selecting the best EKLD, particularly when the best EKLD is relatively complex (Table 1). Incorporating model nesting can also substantially reduce the number of models selected, even when adopting a threshold, when using either EKLD models are presented).

In summary, all ecological analyses involving count data should test for the presence of overdispersion when performing model selection with AIC. If overdispersion is detected and no additional data are available that might explain it, then implicit modelling of the causes of overdispersion should be performed whenever possible. However, if it is not possible to model overdispersion, then selecting models based on QAIC also appears sound. In general, irrespective of the presence of overdispersion, I suggest that all models having a threshold of 6 should initially be selected when using AIC or QAIC. I also suggest that the resulting set of selected models should be reduced by removing the more complex models that do not have a threshold which is lower than all the simpler models within which they are nested. Ecological inference should primarily be based on the simplest models selected, and factors uniquely associated with the more complex models selected can be considered to have weaker support. Stephens et al. (2005, 2007) highlight that information-theoretic approaches, like all model selection approaches, are vulnerable to misuse. Estimating EKLDS using Monte Carlo integration provides a promising framework for further quantifying AIC and QAIC performance, which in turn should help the development of better rules for their use.

Acknowledgements

The author gratefully thanks the reviewers for their valuable comments and suggestions. This research was supported in part by a G8 legacy chair in wildlife ecology, The University of Calgary.

References


