

Using the negative binomial distribution to model overdispersion in ecological count data

ANDREAS LINDÉN^{1,3} AND SAMU MÄNTYNIEMI²

¹Department of Biology, Centre for Ecological and Evolutionary Synthesis, University of Oslo, P.O. Box 1066 Blindern, NO-0316 Oslo, Norway

²Department of Environmental Sciences, Fisheries and Environmental Management Group (FEM), P.O. Box 65, FIN-00014 University of Helsinki, Finland

Abstract. A Poisson process is a commonly used starting point for modeling stochastic variation of ecological count data around a theoretical expectation. However, data typically show more variation than implied by the Poisson distribution. Such overdispersion is often accounted for by using models with different assumptions about how the variance changes with the expectation. The choice of these assumptions can naturally have apparent consequences for statistical inference. We propose a parameterization of the negative binomial distribution, where two overdispersion parameters are introduced to allow for various quadratic mean–variance relationships, including the ones assumed in the most commonly used approaches. Using bird migration as an example, we present hypothetical scenarios on how overdispersion can arise due to sampling, flocking behavior or aggregation, environmental variability, or combinations of these factors. For all considered scenarios, mean–variance relationships can be appropriately described by the negative binomial distribution with two overdispersion parameters. To illustrate, we apply the model to empirical migration data with a high level of overdispersion, gaining clearly different model fits with different assumptions about mean–variance relationships. The proposed framework can be a useful approximation for modeling marginal distributions of independent count data in likelihood-based analyses.

Key words: bird migration; count data; environmental stochasticity; flocking; generalized linear models; mean–variance relationship; negative binomial distribution; overdispersion; Poisson process; sampling error.

INTRODUCTION

Counts of organisms are one of the most common types of data used in ecological studies (Pielou 1984, O'Hara and Kotze 2010), when temporal and/or spatial patterns of abundance are in focus. In models for discrete count data, unexplained stochastic variation around the model expectation (for example a regression line) is commonly modeled with the Poisson process as a starting point, assuming that the probability of observing the next individual or event is constant in time or space for each sampling unit. In a Poisson distribution the variance equals the mean. However, for various reasons the amount of variation for each sampling unit is typically higher than expected by a pure Poisson process. This extra variation, termed as overdispersion, is caused by spatiotemporal heterogeneity in the process that produces data, typically due to observation errors (sampling effects and observation inaccuracy) or process errors (variation in the Poisson intensity).

Overdispersed Poisson processes can be modeled in many alternative ways. The most common approaches

used for count data include quasi-likelihood-based Poisson models (Wedderburn 1974), random-effects models (Bolker et al. 2008, O'Hara 2009), and negative binomial models (e.g., McCullagh and Nelder 1989, White and Bennetts 1996, Alexander et al. 2000), all extensively used in ecological studies and readily available in statistical software packages. In quasi-Poisson methods the amount of overdispersion is estimated under the assumption that the variance is proportional to the mean, after which, e.g., standard errors are corrected for the estimated overdispersion. On the other hand, mixed-effects Poisson models and typical negative binomial models assume that the extra-Poisson variance is a quadratic function of the mean. Within the Bayesian framework these types of models are often referred to as hierarchical models. The statistical literature also proposes other types of overdispersed generalized Poisson models (e.g., Famoye 1993, Famoye et al. 2004) and zero-inflated variants of the Poisson and negative binomial models (e.g., Lambert 1992, Sileshi et al. 2009).

Ver Hoef and Boveng (2007) made a comparison between quasi-Poisson and negative binomial regressions as two contrasting approaches for dealing with overdispersed count data in ecology. With an example on harbor seal data they showed that the choice of

Manuscript received 20 September 2011; revised 22 February 2011; accepted 28 February 2011; final version received 21 March 2011. Corresponding Editor: M. Fortin.

³ E-mail: andreas.linden@iki.fi

approach can affect the outcome of the analysis. The authors recommended sound scientific reasoning and graphical investigation of the data as the basis for model choice. Yet, different processes underlying overdispersion in ecological data and resulting in various mean–variance relationships have not been thoroughly investigated.

We present here a specific parameterization of the negative binomial distribution, which can be used to approximate overdispersed Poisson processes with a wide range of mean–variance relationships. We investigate different scenarios of observational, behavioral, and environmental processes that are likely to produce overdispersion in ecological data and report the resulting mean–variance relationships. Further, we present an empirical example where the proposed error structures are applied to fit phenological models to autumn migration counts of Woodlarks (*Lullula arbor-rea*). Finally, we propose how to handle situations where the type of overdispersion is difficult to specify on biological grounds.

AN ALTERNATIVE PARAMETERIZATION OF THE NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial distribution has been parameterized in a number of different ways in the statistical and applied literature. Perhaps the most common way to parameterize is to see the negative binomial distribution arising as a distribution of the number of failures (X) before the r th success in independent trials, with success probability p in each trial (consequently, $r \geq 0$ and $0 \leq p \leq 1$). In such a case the probability mass function can be expressed as

$$P(X = x | r, p) = \frac{\Gamma(x+r)}{x! \Gamma(r)} p^r (1-p)^x \quad (1)$$

and the random variable X has the expectation (theoretical mean) $\mu = r(1-p)/p$ and variance $\sigma^2 = r(1-p)/p^2$. Corresponding to this parameterization, parameters r and p can be written in terms of the mean and the variance as

$$r = \frac{\mu^2}{\sigma^2 - \mu} \quad (2)$$

$$p = \frac{\mu}{\sigma^2}. \quad (3)$$

We use Eqs. 2 and 3 as a starting point for a more flexible parameterization of the negative binomial distribution that can be used to represent various kinds of overdispersed Poisson processes. By writing the variance as a quadratic function of the mean,

$$\sigma^2 = \omega\mu + \theta\mu^2 \quad (4)$$

diverse mean–variance relationships are obtained using the parameters ω and θ , hereafter called overdispersion parameters. This process is defined as long as $\sigma^2 > \mu$, or

phrased in terms of the overdispersion parameters, when $\omega + \theta\mu > 1$. With any valid combination of expectation $E(X)$ and overdispersion parameters, the parameters r and p of the negative binomial distribution can be derived as

$$r = \frac{\mu}{\omega - 1 + \theta\mu} \quad (5)$$

$$p = \frac{1}{\omega + \theta\mu} \quad (6)$$

and then be substituted into Eq. 1 in order to obtain the relevant likelihood. Supplement 1 contains MATLAB code for calculating the log-likelihoods for combinations of θ , ω , and parameters defining the mean, given any data set. A linear mean–variance relationship, corresponding to the assumption of quasi-Poisson regression, is obtained by fixing θ to zero. Similarly, fixing ω to 1 gives the quadratic mean–variance relationship of the classic negative binomial regression.

The negative binomial distribution can also be seen as an explicit overdispersed Poisson process, where the Poisson intensity is drawn from a gamma distribution (Gelman et al. 2004). The description corresponding to this parameterization is given in Appendix A.

The proposed parameterizations allow both Bayesian and maximum likelihood estimation of the overdispersion parameters and model parameters that define the expectation, with any statistical software package that can handle user-defined likelihood functions. Due to straightforward calculation of likelihoods, it is possible to use analytical forms of model selection techniques such as the Akaike information criterion (AIC), deviance information criterion (DIC), and Bayes factors without any additional corrections for overdispersion (Akaike 1973, Spiegelhalter et al. 2002, Gelman et al. 2004).

PROCESSES CAUSING OVERDISPERSION

In the following section, we describe scenarios of how count data might arise and what kind of overdispersion this can lead to. We compose the scenarios from multiple different processes that are assumed to take place when the data are being generated, and model each of these processes using a stochastic model (summarized in Table 1). The models are chosen to represent commonly applied choices in ecological modeling, such as the log-normal distribution for multiplicative variation. However, most of the scenarios are more general in that, unless stated otherwise, the resulting mean–variance relationships depend only on the means and variances of the assumed processes and not on the specific distributional assumptions. We have used standard methods of deriving means and variances of functions of random variables to deduce the marginal mean and variance of the resulting observable data, as functions of parameters of contributing processes. These functions can

TABLE 1. Scenarios for how sampling error, flocking, and environmental stochasticity might cause overdispersion in daily bird migration count data. Each scenario is described with a mathematical notation, the expected daily count, $E(X_i)$, and overdispersion parameters ω and θ giving the variance as $V(X_i) = \omega E(X_i) + \theta E(X_i)^2$.

Scenario	Count data	Migration numbers	No. flocks (detection rate)	$E(X_i)$	ω	θ
1	$X_i = M_i$	$M_i \sim \text{Pois}(\lambda_i)$		λ_i	1	0
Counting- and sampling error						
2	$X_i \sim \text{Bin}(M_i, q)$	$M_i \sim \text{Pois}(\lambda_i)$		$\lambda_i q$	1	0
3	$X_i \sim \text{Pois}(M_i q)$	$M_i \sim \text{Pois}(\lambda_i)$		$\lambda_i q$	$1 + q$	0
4	$X_i \sim \text{Bin}(M_i, q_i)$	$M_i \sim \text{Pois}(\lambda_i)$	$q_i \sim \text{Beta}(a, b)$	$\lambda_i E(q)$	1	A
Flocking behavior						
5	$X_i = M_i$	$M_i = Z_i s$	$Z_i \sim \text{Pois}(\lambda_i / s)$	λ_i	s	0
6	$X_i = M_i$	$M_i = \sum_{j=1}^{Z_i} s \varepsilon_{ij}$	$Z_i \sim \text{Pois}(\lambda_i / s)$	λ_i	$s + sV(\varepsilon)$	0
7	$X_i = M_i$	$M_i = \sum_{j=1}^{Z_i} f_j$	$Z_i \sim \text{Pois}(1 / d)$	λ_i	g	d
Environmental stochasticity						
8	$X_i = M_i$	$M_i \sim \text{Pois}(\lambda_i \varepsilon_i)$		λ_i	1	$V(\varepsilon)$
9	$X_i = M_i q \phi_i$	$M_i \sim \text{Pois}(\lambda_i)$		$\lambda_i q$	$q + qV(\phi)$	$V(\phi)$
10	$X_i = M_i q \phi_i$	$M_i \sim \text{Pois}(\lambda_i \varepsilon_i)$		$\lambda_i q$	$q + qV(\phi)$	$V(\phi) + V(\varepsilon) + V(\phi)V(\varepsilon)$
Combinations of several factors						
11	$X_i \sim \text{Bin}(M_i, q)$	$M_i \sim \text{Pois}(\lambda_i \varepsilon_i)$		$\lambda_i q$	1	$V(\varepsilon)$
12	$X_i \sim \text{Pois}(M_i q)$	$M_i \sim \text{Pois}(\lambda_i \varepsilon_i)$		$\lambda_i q$	$1 + q$	$V(\varepsilon)$
13	$X_i \sim \text{Bin}(M_i, q_i)$	$M_i \sim \text{Pois}(\lambda_i \varepsilon_i)$	$q_i \sim \text{Beta}(a, b)$	$\lambda_i E(q)$	1	$V(\varepsilon) + A + V(\varepsilon)A$
14	$X_i = M_i$	$M_i = Z_i s$	$Z_i \sim \text{Pois}(\lambda_i \varepsilon_i / s)$	λ_i	s	$V(\varepsilon)$
15	$X_i = M_i$	$M_i = \sum_{j=1}^{Z_i} s \phi_j$	$Z_i \sim \text{Pois}(\lambda_i \varepsilon_i / s)$	λ_i	$s + sV(\phi)$	$V(\varepsilon)$
16	$X_i = M_i$	$M_i = \sum_{j=1}^{Z_i} f_j$	$Z_i \sim \text{Pois}(\varepsilon_i / d)$	λ_i	g	$d + V(\varepsilon)$

Notes: Key to notation: λ_i , migration intensity, i.e., the expected number of migrating individuals on day i ; X_i , number of observed individuals on day i ; M_i , actual number of migrating individuals; q , detection probability or efficiency; s , average flock size; Z_i , number of flocks; f_j , size of flock j , such that $E(f_j) = d\lambda_i$, $V(f_j) = gE(f_j)$, $0 < d < 1$, $g > 0$; a, b , parameters of the beta distribution; $E(q) = a/(a+b)$; $A = b/(a^2 + ab + a)$; ε_i , ε_j , and ϕ_i are lognormally distributed random variables with unit mean [$\ln \varepsilon_i \sim \mathcal{N}(-0.5v, v)$, $E(\varepsilon) = 1$, and $V(\varepsilon) = e^v - 1$].

then be used to understand the mean–variance relationship of the assumed multistage process. To keep our explanations more concrete, we refer to the number of migrating birds counted at a specific migration locality on a given day, which is a typical example of overdispersed ecological count data (Miller et al. 2002, Knudsen et al. 2007, Lehtikoinen et al. 2010). However, many of the scenarios are far more general and can be exported to other sampling-based census methods where the data consist of counts. Naturally, the applicability in other areas than bird migration might vary from case to case.

The following notation is used in all scenarios (Table 1). The variable of main interest is the migration intensity λ_i , defined as the expected number of migrating birds on day i . These expectations are assumed to be specified by a model, e.g., a regression line or plane. The daily number of migrating and observed birds (the data), are denoted as M_i and X_i , respectively.

In the simplest scenario (scenario 1), the number of migrating individuals M_i is a pure Poisson process with intensity λ_i , and all individuals are observed and identified correctly. This can be thought as a starting point, which provides the lower bound for the amount of variation: at a given day there is some constant probability of appearance for the next bird individual of a given species. If the pool of potentially passing

individuals is far bigger than the actually observed numbers during one day, this should be a good approximation, as the day-specific probability will not decrease much with each observed individual. However, a pure Poisson process is rather unrealistic, since the detection and identification of the birds, variation in observer quality and observation conditions, flocking behavior of the birds, and environmental variation affecting migration are potential causes of overdispersion in the data.

Counting and sampling error

Most ecological count data are inflated with some kind of counting or sampling error compared to the actual variable of interest. A simple scenario of counting or sampling error occur when the count data arise from a binomial process, where each bird is found and counted with a given probability and the number of trials is the actual number of birds (scenario 2). The same result will apply when counts are based on temporal sampling, e.g., if migration is counted only during the first 10 minutes each hour (counting probability is $\leq 10/60$). The outcome of this scenario is another Poisson process, where the variance equals the mean, and is hence technically easy to handle.

One alternative is that the counting can also be described as a Poisson process (scenario 3). This



PLATE 1. The counted numbers of migrating Woodlarks (*Lullula arborea*) at Hanko bird observatory (Finland) is an example of highly overdispersed data. Photo credit: Jorma Tenovu.

scenario could be plausible in cases where the same bird can be counted several times, for example due to several people involved in the counting. This can be the case in databases, where data are gathered from amateur ornithologists through the Internet. A nested (double) Poisson process is not the most typical for bird migration. However, in microbiology the framework of multiple nested Poisson processes can be used to model results from dilution series, a standard technique for estimating densities of bacteria in liquid (Niemelä 1983). In the presented scenario, the variance will be larger than the expected value, but still increasing linearly with the mean. If the total detection probability is very small relative to the true number of migrating birds, the result will be close to that of scenario 2.

A probably more typical type of counting error includes variability in detection efficiency. This can be due to different observers on different days, or variable counting conditions. We model this (scenario 4) by viewing the detection probability as a beta-distributed random variable. This leads to a quadratic mean–variance relationship, similar to the classic parameterization in negative binomial regression. The mean–variance relationship of this scenario holds under the assumption of beta-distributed detection probability, but may not apply for other probability density functions with the same mean and variance.

Flocking behavior and aggregation

Another source of overdispersion in count data is social behavior and other kinds of spatiotemporal

aggregation of certain species. At a large scale, the spatial distribution of an organism is usually aggregated to some extent, simply because the landscape is typically a mosaic of suitable and less suitable habitats for the species in question. Here we consider the mean–variance relationships of data on bird species migrating in flocks.

A simple example arises when individuals are assumed to form groups of a certain size (scenario 5). If the groups are then also assumed to behave independently in respect to observing them, then one can see the number of observed flocks as a Poisson process. If all flocks are observed, the number of observed individuals is the number of flocks times flock size and the variance increases linearly with the mean.

A more realistic scenario occurs when flock size is allowed to vary (scenario 6). The number of observed birds is given by the sum of the number of individuals in each flock. As long as the distribution for flock size remains constant, the type of overdispersion will remain unchanged, compared to the case of constant flock size, i.e., a linear mean–variance relationship. This is also the kind of overdispersion that Ver Hoef and Boveng (2007) found in their example of spatially aggregated seals. If the expected flock size is a constant fraction of the Poisson intensity, and the variance of the flock size is assumed to depend linearly on the expected flock size (scenario 7), the data will show a quadratic mean–variance relationship similar to that presented in Eq. 4.

Environmental stochasticity

Often unknown environmental factors, such as weather, affect the occurrence of animals or the counting process. We first consider the probably very common scenario where the actual migration intensity varies according to environmental variation (scenario 8), here modeled as a log-normally distributed Poisson intensity. This type of stochasticity is conceptually equal to what we call process errors, and actually corresponds to the most common type of (log-linear) mixed-effects Poisson model. This leads to a variance increasing quadratically with the mean in the same manner as in classic negative binomial regression and random-effects Poisson models.

Environmental variation can also affect observability. For example, weather typically affects the height at which birds migrate, and during beneficial conditions the migration height of some species is far above that where the birds are easy to observe. We present here a scenario where observability varies according to a log-normal distribution, representing environmental stochasticity (scenario 9). Here we relax the definition of M_i (the actual number of birds) to the expected number of birds migrating at an observable distance in average conditions (given the migration intensity λ_i). According to this model, the number of counted birds (X_i) can exceed M_i , if environmental conditions increase the observability or canalize the migration close to the counting spot. The scenario is conceptually very close to scenario 4 and could have been presented under the heading of counting- and sampling error as well. This process will lead to data with a more complex quadratic mean–variance relationship, which still can be described with the two parameters of Eq. 4.

More likely than scenarios 8 and 9, environment can affect both migration intensity and observability simultaneously, here modeled again as multiplicative log-normal effects (scenario 10). For simplicity, we base our results in Table 1 on the assumption that these effects are uncorrelated with each other, although that might in many cases not be true. The resulting overdispersion will have the same type of quadratic mean–variance relationship as scenario 9, where two parameters for overdispersion are needed to sufficiently model the variation. The principal result will remain unchanged when allowing linear covariation between $\ln \phi$ and $\ln \epsilon$. Under the fairly good approximation that M_i is log-normally distributed, the overdispersion parameters can be calculated as $\omega = C^2\omega_{10}$ and $\theta = C\theta_{10} + C - 1$, where $C = \exp[2\text{Cov}(\ln \phi, \ln \epsilon)]$, while ω_{10} and θ_{10} are the overdispersion parameters from the uncorrelated case (scenario 10). These equations apply if $\ln \phi$, $\ln \epsilon$, and $\ln M_i$ are assumed to be normally distributed.

Combinations of several sources of overdispersion

Different sources of error are likely to affect count data simultaneously. A considerable amount of the variation not explained in a regression type of a model is

not necessarily due to counting error, flocking behavior, or environmental stochasticity, but due to lack of important predictors or insufficiency to model the migration intensity. This kind of process error is likely to be of the same type as environmental stochasticity in migration intensity (scenario 8). The scenarios representing combinations of different sources of overdispersion (scenarios 11–16) correspond to scenarios 2–7, but with log-normal stochasticity of the migration intensities.

According to our results, all of these models require a quadratic term to describe the mean–variance relationship ($\theta > 0$ in Eq. 4), and in most cases the relationship is more complex (also $\omega > 1$ in Eq. 4), yet such that the mean–variance relationship can be explained by the negative binomial distribution as explained in this paper. In scenario 11 and scenario 13, $\omega = 1$, which is the case for all parts involved in the model (both scenario 8 and scenarios 2 and 4).

AN EMPIRICAL EXAMPLE: MIGRATING WOODLARKS

To illustrate and justify the use of the negative binomial distribution with the proposed parameterizations, we fit phenological curves to the autumn migration data of Woodlark (*Lullula arborea*; see Plate 1) at Hanko bird observatory (59°49' N, 22°54' E). The observatory is located in southwestern Finland on a peninsula known to be a bottleneck for terrestrial bird migration in autumn. We use migration data from the autumn migration season (1 September–10 November) in 2007–2009. The data (Supplement 2) consist of daily standardized counts on actively migrating birds during the first four hours of daylight, starting from sunrise.

To model phenology, i.e., the gradual change in migration intensity throughout the season, we define the daily expected number of migrants (λ_i) as a Gaussian function:

$$\lambda_i = c \frac{1}{\sqrt{2\pi}v} \exp \left[-\frac{(i-m)^2}{2v^2} \right]. \quad (7)$$

This corresponds to a nonlinear regression curve describing bird numbers (response variable) against day of year (explanatory variable). The parameter c scales the function height to match the number of expected migrants, while parameters m and v correspond to the Gaussian mean and standard deviation, respectively. Since the daily data are discrete and truncated by season beginning and end, relevant quantities can be calculated from the fitted function for λ_i . In this example, we use m to measure central tendency, v for spread and $n = \sum \lambda_i$ for expectation of the total number of migrants.

Under the assumption that the Gaussian function (Eq. 7) appropriately captures the temporal expectation of the migration curve, the data points are assumed to be spread around the expectation according to different

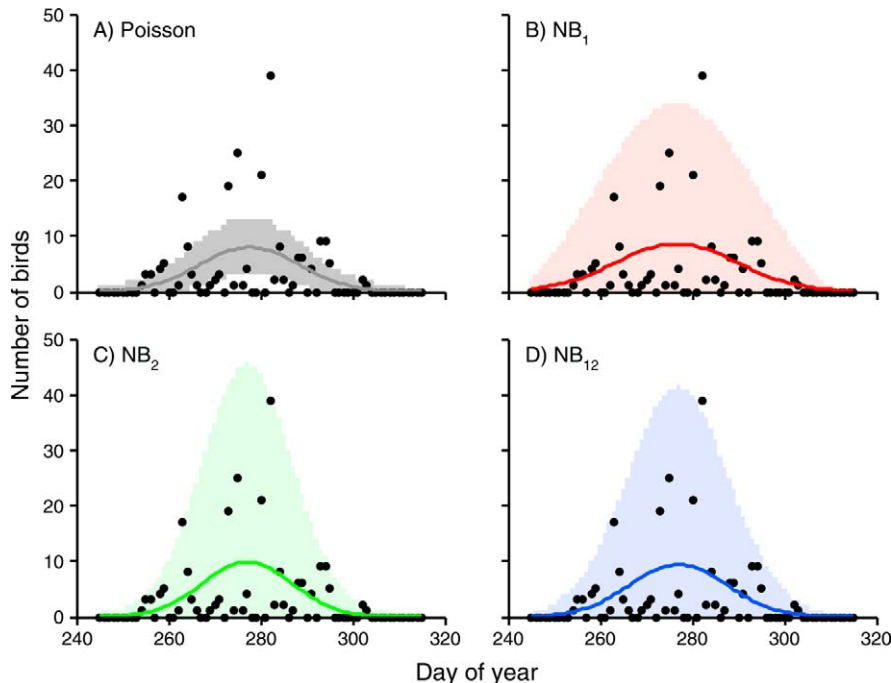


FIG. 1. Seasonal variation in intensity of autumn migration of Woodlarks at Hanko bird observatory, Finland, in 2009. The number of counted birds is plotted as black dots against the autumn migration season, here described as day of year, starting from 1 September (day 245) to 10 November (day 315). The solid lines are fitted Gaussian functions describing the expected number of observed birds, assuming different error distributions: (A) Poisson, (B) negative binomial distribution with a linear mean–variance relationship $\sigma^2 = \omega\mu$ (NB_1), (C) negative binomial distribution with a quadratic mean–variance relationship $\sigma^2 = \mu + \theta\mu^2$ (NB_2), and (D) negative binomial distribution with $\sigma^2 = \omega\mu + \theta\mu^2$ (NB_{12}). The parameters ω and θ are overdispersion parameters, μ is the expectation, and σ^2 is the variance. The shaded areas are 95% prediction intervals for the observations.

error distributions: (1) the Poisson distribution, (2) negative binomial distribution with a linear mean–variance relationship $\sigma^2 = \omega\mu$ (NB_1), (3) negative binomial distribution with a quadratic mean–variance relationship $\sigma^2 = \mu + \theta\mu^2$ (NB_2), and (4) negative binomial distribution with $\sigma^2 = \omega\mu + \theta\mu^2$ (NB_{12}). Overdispersion parameters are assumed to be constant across years. The models are fitted by maximum likelihood estimation using the Nelder–Mead simplex algorithm for numerical search over the parameter space. For all analyses and model-fitting purposes we use MATLAB (version 7.6.0; The MathWorks, Natick, Massachusetts, USA).

Despite large day-to-day variation, the fitted curves seem to describe the phenological data reasonably well (Fig. 1). Two-dimensional profile likelihoods for the overdispersion parameters ω and θ (Appendix B) provide strong evidence for overdispersion, but show a strong negative association, indicating that the type of overdispersion is difficult to identify. In this example, the best option is either to include both overdispersion parameters (maximum likelihood estimates and SE: $\omega = 3.87 \pm 2.63$; $\theta = 2.56 \pm 0.70$), or alternatively to use NB_2 , since ω is not significantly different from 1 ($\theta = 3.29 \pm 0.43$). The mean–variance relationships are also fairly well described using these two approaches

(Appendix C), in particular using NB_{22} . Point estimates of parameters and quantities of interest, e.g., quantiles of the fitted distribution, vary somewhat according to the approach for overdispersion (Appendix D). However, estimates of uncertainty (bootstrap standard errors, Appendix D) vary strongly according to the chosen approach. These results confirm the points raised by Ver Hoef and Boveng (2007): the approach for modeling overdispersion matters.

CONCLUSIONS AND RECOMMENDATIONS

In decisions for specifying model structure, we believe it is wisest to stick to some mechanistic justification, such as existence of flocking behavior, for using a certain approach. The best option is to explicitly model the spatiotemporal process of interest to capture the underlying nature of overdispersion. However, often hierarchical models with variable model parameters (as “random effects”) described with another set of (hyper) parameters, can be associated with technical difficulties in model fitting and/or convergence. This can happen when there are no data on intermediate layers of the hierarchical process, like observations on flock sizes. Given that the underlying model describes the spatiotemporal dependencies in the data, the negative binomial distribution can be a good approximation for the

marginal distribution of data in many of the presented scenarios. At least the mean–variance relationships can in all considered cases be described appropriately. The method is straightforward to use, both in the Bayesian framework and in maximum likelihood based methods.

In addition to the proposed scenarios, a potential area of application for the proposed parameterization of the negative binomial distribution is population dynamics, where process error in population size (or density) is likely to take a quadratic mean–variance relationship. Typically, demographic stochasticity will induce linear mean–variance relationships, while that of environmental stochasticity will be quadratic (Engen et al. 1998, Brännström and Sumpter 2006). The combination of potentially correlated demographic and environmental stochasticity will still correspond to the quadratic mean–variance relationship in Eq. 4 (Engen et al. 1998).

It is not always easy to choose the model for overdispersion solely on theoretical and biological grounds. Knowing the type of data used is of course important and general exploration of data in terms of, e.g., plotting may be useful or recommendable (Chatfield 2004). As Ver Hoef and Boveng (2007) stated, plotting residuals against the expectation can be a useful step while deciding how to model overdispersion. However, learning to know the data, for example by plotting, can change the beliefs in what model may be the most appropriate and what modeling decisions to make. It is important to acknowledge the risk for data dredging, i.e., making implicitly multiple use of the information in the data (Burnham and Anderson 2002).

One alternative approach for cases with uncertain type of overdispersion is model selection, possibly further extended to multimodel inference through model averaging. Model selection can be done in a Bayesian context, e.g., using Bayes factors (Gelman et al. 2004), or by applying information theoretical criteria (Burnham and Anderson 2002, Johnson and Omland 2004). For quasi-likelihood-based models, Burnham and Anderson (2002) present an information theoretical criterion corrected for overdispersion (QAIC). Sileshi (2006) compared the performance of quasi-Poisson models with negative binomial models in a study on insect count data, using model selection mixing AIC and QAIC. However, this approach is questionable, since QAIC is developed only for comparison of models with essentially the same kind of overdispersion (Burnham and Anderson 2002). One benefit of using the approach proposed here is that AIC can be calculated directly from the likelihoods and different types of overdispersion can be compared without any such problems. Whenever desired, likelihood ratio tests can be done to test null hypotheses on the absence of certain types of overdispersion.

Because stochastic variation in the Poisson intensity (i.e., process errors) are likely to affect the variable of interest and typically introduce variation increasing quadratically with the mean, ignoring such randomness

can be a dangerous choice in general. If there is a doubt about the kind of overdispersion present, a simple alternative to model selection is to take two parameters for overdispersion into consideration in any case. Such a model is especially useful in the Bayesian context, if there is prior information available about the sources of overdispersion, for example, average flock size and/or the variability of flock size.

As shown in this paper, the type of overdispersion in ecological data can take different forms. Choices on how to model such overdispersion has obvious impact on the results, especially on estimates of uncertainty. The best option might be careful consideration of what the overdispersion should look like based on biological understanding about the phenomenon, and using a model corresponding to that logic. The results presented in this paper are targeted to help with that type of biological model choice. Whenever this is difficult to implement, flexible approaches for overdispersion and model selection can provide suitable solutions. The negative binomial distribution is a viable option for describing many common types of mean–variance relationships seen in ecological data.

ACKNOWLEDGMENTS

We are grateful to Jonas Knappe and an anonymous referee for useful comments on the manuscript. We acknowledge Hanko bird observatory and the voluntary ornithologists who gathered the Woodlark data. The work was funded by the University of Helsinki and Academy of Finland (A. Lindén).

LITERATURE CITED

- Akaike, H. 1973. Information theory and the extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov, and F. Czaki, editors. *Proceedings of the International Symposium on Information Theory*. Akademia Kiadoo, Budapest, Hungary.
- Alexander, N., R. Moyeed, and J. Stander. 2000. Spatial modeling of individual-level parasite counts using the negative binomial distribution. *Biostatistics* 1:453–463.
- Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. 2008. *Generalized linear mixed models: a practical guide for ecology and evolution*. Trends in Ecology and Evolution 24:127–135.
- Brännström, Å., and D. J. T. Sumpter. 2006. Stochastic analogues of deterministic single-species population models. *Theoretical Population Biology* 69:442–451.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, Berlin, Germany.
- Chatfield, C. 2004. *The analysis of time series—an introduction*. Sixth edition. Chapman and Hall, London, UK.
- Engen, S., Ø. Bakke, and A. Islam. 1998. Demographic and environmental stochasticity—concepts and definitions. *Biometrics* 54:840–846.
- Famoye, F. 1993. Restricted generalized Poisson regression model. *Communications in Statistics—Theory and Methods* 22:1335–1354.
- Famoye, F., J. T. Wulu, Jr., and K. P. Singh. 2004. On the generalized Poisson regression model with an application to accident data. *Journal of Data Science* 2:287–295.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin. 2004. *Bayesian data analysis*. Second edition. Chapman and Hall/CRC, Boca Raton, Florida, USA.

- Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. *Trends in Ecology and Evolution* 19:101–108.
- Knudsen, E., A. Lindén, T. Ergon, N. Jonzén, J. O. Vik, J. Knape, J. E. Røer, and N. C. Stenseth. 2007. Characterizing bird migration phenology using data from standardized monitoring at bird observatories. *Climate Research* 35:59–77.
- Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* 34: 1–14.
- Lehikoinen, A., P. Saurola, P. Byholm, A. Lindén, and J. Valkama. 2010. Life history events of Eurasian sparrowhawk in a changing climate. *Journal of Avian Biology* 41:627–636.
- McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall, London, UK.
- Miller, M. W., E. M. Greenstone, W. Greenstone, and K. L. Bildstein. 2002. Timing and magnitude of broad-winged hawk migration at Montclair Hawk Lookout, New Jersey, and Hawk Mountain Sanctuary, Pennsylvania. *Wilson Bulletin* 114:479–484.
- Niemelä, S. 1983. Statistical evaluation of results from quantitative microbiological examinations. Nordic Committee on Food Analysis Report 1. Second edition. Ord and Form AB, Uppsala, Sweden.
- O'Hara, R. B. 2009. How to make up models add up—a primer on GLMMs. *Annales Zoologici Fennici* 46:124–137.
- O'Hara, R. B., and D. J. Kotze. 2010. Do not log-transform count data. *Methods in Ecology and Evolution* 1:118–122.
- Pielou, E. C. 1984. The interpretation of ecological data. a primer on classification and ordination. John Wiley and Sons, New York, New York, USA.
- Sileshi, G. 2006. Selecting the right statistical model for analysis of insect count data by using information theoretic measures. *Bulletin of Entomological Research* 96:479–488.
- Sileshi, G., G. Hailu, and G. I. Nyadzi. 2009. Traditional occupancy–abundance models are inadequate for zero-inflated ecological count data. *Ecological Modelling* 220:1764–1775.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 64:583–639.
- Ver Hoef, J. M., and P. L. Boveng. 2007. Quasi-Poisson vs. negative binomial regression: How should we model over-dispersed count data? *Ecology* 88:2766–2772.
- Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika* 61:439–447.
- White, G. C., and R. E. Bennetts. 1996. Analysis of frequency count data using the negative binomial distribution. *Ecology* 77:2549–2557.

APPENDIX A

The negative binomial distribution viewed as a hierarchical Poisson–gamma model (*Ecological Archives* E092-120-A1).

APPENDIX B

A figure showing two-dimensional profile likelihoods of ω and θ (*Ecological Archives* E092-120-A2).

APPENDIX C

A figure with observed vs. fitted mean–variance relationships (*Ecological Archives* E092-120-A3).

APPENDIX D

A table with parameter estimates and standard errors of fitted models (*Ecological Archives* E092-120-A4).

SUPPLEMENT 1

MATLAB code for the log-likelihood (*Ecological Archives* E092-120-S1).

SUPPLEMENT 2

Count data on migrating woodlarks at Hanko bird observatory (*Ecological Archives* E092-120-S2).