# Estimating the effects of detection heterogeneity and overdispersion on trends estimated from avian point counts

Matthew A. Etterson,[1] Gerald J. Niemi, and Nicholas P. Danz[2]

*Center for Water and the Environment, Natural Resources Research Institute, 5013 Miller Trunk Highway,*
*Duluth, Minnesota 55811-1442 USA*

*Abstract.* Point counts are a common method for sampling avian distribution and abundance. Although methods for estimating detection probabilities are available, many analyses use raw counts and do not correct for detectability. We use a removal model of detection within an *N*-mixture approach to estimate abundance trends corrected for imperfect detection. We compare the corrected trend estimates to those estimated from raw counts for 16 species using 15 years of monitoring data on three national forests in the western Great Lakes, USA. We also tested the effects of overdispersion by modeling both counts and removal mixtures under three statistical distributions: Poisson, zero-inflated Poisson, and negative binomial. For most species, the removal model produced estimates of detection probability that conformed to expectations. For many species, but not all, estimates of trends were similar regardless of statistical distribution or method of analysis. Within a given combination of likelihood (counts vs. mixtures) and statistical distribution, trends usually differed by both stand type and national forest, with species showing declines in some stand types and increases in others. For three species, Brown Creeper, Yellow-rumped Warbler, and Black-throated Green Warbler, temporal patterns in detectability resulted in substantial differences in estimated trends under the removal mixtures compared to the analysis of raw counts. Overall, we found that the zero-inflated Poisson was the best distribution for our data, although the Poisson or negative binomial performed better for a few species. The similarity in estimated trends that we observed among counts and removal mixtures was probably a result of both experimental design and sampling effort. First, the study was originally designed to avoid confounding observer effects with habitats or time. Second, our time series is relatively long and our sample sizes within years are large.

*Key words: abundance indices; avian point counts; detection heterogeneity; overdispersion; point counts; removal sampling.*

## INTRODUCTION

Understanding the distribution and abundance of animals is a fundamental goal of basic ecological research. Knowledge of how these factors depend on environmental conditions and how they are affected by anthropogenic activities is also a key goal in applied ecology and conservation biology. However, estimating abundance and density is difficult because many animals are cryptic and survey methods require observers with specialized skills such as knowledge of bird song. Due in part to these difficulties, it is rarely possible to count all animals present in a survey area; typically, we do not know what proportion of the total number of individuals present are actually counted. Thus it is easy to question the utility of incomplete count data (Rosenstock et al. 2002).

Point counts are a common method for sampling avian distribution and abundance. Although protocols vary, the basic method involves recording all birds seen or heard within a specified time limit, usually 3, 5, or 10 minutes, often within specified distance classes (Reynolds et al. 1980, Howe et al. 1997). Common criticisms of point count methods include bias in selection of sampling locations (Bart et al. 1995, Hanowski and Niemi 1995*a*), bias due to variation in observer ability (Sauer et al. 1994), and failure to account for detection heterogeneity (Burnham 1981, Johnson 1995). The latter topic has been of considerable interest among ornithologists recently. Although some have expressed skepticism over the practicality of routinely estimating detectability (Verner 1985), numerous sampling protocols have been proposed that allow estimation of detection probabilities with point count data. Johnson (2008) reviews the assumptions and potential shortcomings of these methods in comparison to naïve estimation of trends using only the number of birds counted.

[1] Present address: U.S. Environmental Protection Agency, Mid-Continent Ecology Division, 6201 Congdon Boulevard, Duluth, Minnesota 55804 USA.
E-mail: etterson.matthew@epa.gov
[2] Present address: Department of Biology and Earth Sciences, University of Wisconsin–Superior, Belknap and Catlin Ave., Superior, Wisconsin 54880 USA.

### How does imperfect detection affect point count data?

A sufficient condition underpinning the robust estimation of trends from uncorrected point counts is that the number of birds counted is strongly correlated with the actual number of birds present, e.g., $E(C) = pN$, where $C$ is the number of birds counted, $p$ is the detection probability, and $N$ is true abundance (Johnson 2008). The validity of this assumption is often asserted to depend, in turn, on constant detection probability (Barker and Sauer 1995, Nichols et al. 2000, Rosenstock et al. 2002, Williams et al. 2002). However, Johnson (2008) shows that two weaker conditions are sufficient for valid trend inference: (1) that $p$ be independent of $N$, and (2) that the variance of $p$ be "small" in relation to the variance of $N$ (Johnson 2008). Implicit in condition $i$, is the further requirement that $p$ be independent of any covariate (i.e., time) on which $N$ depends.

If one of the above two conditions is violated, for example, if there is a trend in detection probability, then trends in counts may not reflect trends in true abundance. Furthermore, even when the two conditions do hold, plotted trend lines may not, in general, be accurate descriptions of mean abundance, especially for species with low detectability. Thus trend analysis using counts, even when justified, provides limited information about abundance, which is an ecological state variable of primary interest.

Royle (2004a, b) showed how a general "N-mixture" approach can be used to jointly estimate abundance and detection probability, where the resulting abundance estimate is corrected for imperfect detection. This requires an assumed statistical distribution for site-specific abundance, for example Poisson or negative binomial (Royle 2004b, Royle et al. 2004), and is intuitively appealing because it is relatively straightforward to implement using most sampling protocols that allow estimation of detection probabilities. However, an important assumption is that $N$ and $p$ are conditionally independent, which, if satisfied, may also justify the index assumption (Johnson 2008). Thus if the mixture approach is to provide different trend estimates than naïve analysis of counts, such differences must arise through trends in detection probability.

### When and why are count data overdispersed?

Although the Poisson distribution is often considered the natural distribution for animal abundance (Hilborn and Mangel 1997), the equality of mean and variance that is specified under a Poisson distribution is often violated in animal count data (White and Bennetts 1996). When the observed variance exceeds the mean, the data are said to be overdispersed (here and hereafter we use "overdispersion" in reference to the Poisson expectation). Imperfect detection of Poisson-distributed abundance may result in either overdispersion or underdispersion, depending on the distribution of detection probabilities. If $N \sim \text{Poisson}(\lambda)$ then $E(C) = E(p)\lambda$. However, $\text{Var}(C) = \text{Var}(p)\,\lambda(1 + \lambda) + \lambda E(p)^2$. In general, this will not be equal to $E(p)\lambda$. For example, if $p$ is constant, $\text{Var}(C) = \lambda p^2$, and counts will be underdispersed. More generally, counts will be overdispersed when

$$\text{Var}(p) > \frac{E(p)[1 - E(p)]}{1 + \lambda}.$$

Thus detection heterogeneity is more likely to cause overdispersion in counts of abundant species (large $\lambda$) with either high or low mean detection probability.

Environmental heterogeneity may also result in overdispersion if it is not controlled through the use of appropriate covariates. For an overall abundance distribution that is a mixture of Poisson-distributed subpopulations ($N_i \sim \text{Poisson}(\lambda_i)$), $E(N) = E_i(\lambda)$. However, $\text{Var}(N) = E_i(\lambda) + \text{Var}_i(\lambda)$. Therefore Poisson mixtures are overdispersed by the amount $\text{Var}_i(\lambda)$. For example, an equal number of (perfectly detected) counts among two populations, with $\lambda_1 = 4$ and $\lambda_2 = 6$, will have $E(N) = 5$, but $\text{Var}(N) = 6$. Thus, environmental heterogeneity will *always* result in overdispersion if subpopulations are conditionally Poisson distributed and the environmental effects are not incorporated into covariate analysis.

The negative binomial distribution is a mixture of Poisson distributions that is often used for modeling overdispersed count data (Johnson et al. 2005). However, Kéry et al. (2005) argued that the negative binomial distribution may not be the most appropriate distribution for counts of territorial birds. Point count data may also show a different pattern of overdispersion if some subset of counts occurs on unsuitable habitat in which the species is absent. In this case, a zero-inflated Poisson distribution may better capture the pattern of overdispersion (Johnson et al. 2005). Although overdispersion generally gets less attention than detectability in point counts, it may also influence conclusions about abundance and trends. Ver Hoef and Boveng (2007) compared covariate models of harbor seal abundance using quasi-Poisson and negative binomial models. Covariate patterns differed considerably between the two distributional assumptions and the authors concluded the quasi-Poisson was a better assumption for their data.

### Objectives

Here we compare trends inferred from analysis of raw point counts to those inferred using the N-mixture approach with the removal model of Farnsworth et al. (2002). We use hierarchical models to explore the dependence of counts, abundance, detection probability, and overdispersion on several covariates to investigate the potential for obtaining conflicting estimates of trends when correcting for detection probability and overdispersion or not. We apply the models to 16 bird species using a large sample (17 127 surveys) of point counts conducted over 16 years (mean 1070 points surveyed per year) in three national forests in the
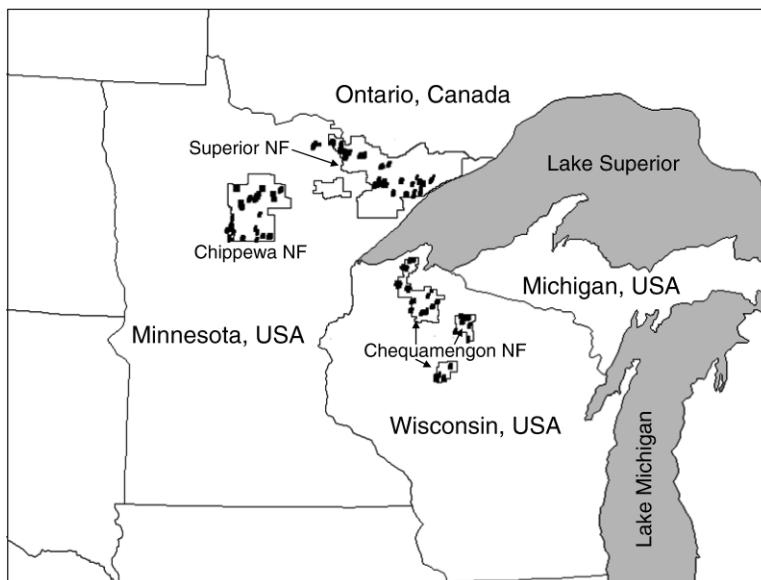
Fɪɢ. 1. Map of western Great Lakes, USA, showing location of the three national forests (NF) surveyed.

western Great Lakes region, USA (Fig. 1). We have five main objectives: (1) to develop a statistical model for preliminary estimation of trends, corrected for detection on the three national forest sites; (2) to test the model using empirical data from species that vary in abundance, trends, and detection probability; (3) to determine whether the removal mixtures give a substantially different estimate of landscape-scale trends than the analysis of raw counts; (4) to investigate the suitability of the Poisson, zero-inflated Poisson, and negative binomial distributions as the underlying abundance distribution for our count data; and (5) to inform the decision-making process about whether and how to control for detection heterogeneity in this ongoing monitoring program.

## Methods

### Site selection and field methods

Data for the analyses presented in this manuscript come from a long-term monitoring program for forest birds in three national forests of the western Great Lakes (Hanowski and Niemi 1995b). Experimental design considerations for the distribution and abundance of samples have been described at length elsewhere (Hanowski and Niemi 1995b) and we briefly summarize the salient points here. Sampling locations were placed across the forests following a stratified random design, with stratum definitions depending on both dominant tree species and stocking density according to existing national forest inventories. For each national forest, a set of stands (minimum area ≥16 ha) was selected from each stratum (hereafter referred to as stand types) so that the final proportion of stands of each type was equal to the proportion of that stand type within its respective national forest. In 1991 and 1992 an initial set of 133, 135, and 169 sample stands were established in the Chequamegon, Chippewa, and Superior National Forests (Fig. 1), respectively, and have been monitored annually, with a few (<5%) additions and deletions, until the present. Stands are large enough to accommodate three sampling points a minimum of 220 m apart. However, even though attention was devoted to avoid double-counting of individual birds, we include here only the two most distal sampling points in each stand.

At each point between June and early July, 10-minute point counts were conducted (Howe et al. 1997) by trained observers (see Hanowski and Niemi 1995b) from approximately 0.5 h before to 4 h after sunrise on days with little wind (<15 km/h) and little or no precipitation. The counting protocol changed in the early years of the study to incorporate unlimited radius counts. In the first three years of the study, all birds within 100 m of the point were recorded. Beginning in 1995, all birds were recorded regardless of distance, but were classified as being greater or less than 100 m from the point. Birds were classified by first detection into the three temporal intervals defined by upper bounds at the end of 3, 5, and 10 minutes (Ralph et al. 1995, Farnsworth et al. 2002). Thus the total number of birds counted at a given site is divided into three counts ($x_1, x_2, x_3 \geq 0$) corresponding to the three initial detection intervals. Each observer sampled a similar number of stands of each stand type to avoid observer differences within stand types.

### Likelihoods

For the analysis of counts we assumed that the total number of birds counted at site $i$ ($C_i = x_{i_1} + x_{i_2} + x_{i_3}$) followed a Poisson, zero-inflated Poisson (ZIP), or a negative binomial (NB) distribution. Specific forms for

all of the likelihoods, log-likelihoods, and gradients are provided in Appendix A. Assuming that counts are Poisson distributed with mean $\lambda$, then the expected variance of $C$ is also $\lambda$. Under a ZIP distribution, with abundance parameter $\lambda$ and zero-inflation parameter $\phi$, the expected count $C$ is $\mu = \lambda(1 - \phi)$ and the expected variance is $\mu(1 + \lambda\phi)$, whereas under an NB distribution with mean $\mu$ and overdispersion parameter $k$, the expected variance is $\mu + (\mu^2/k)$ (Hilborn and Mangel 1997). Thus the ZIP and NB distributions both specify greater variance than their mean, although the structure of the variance and its relationship to the mean differ from each other. For the ZIP distribution, variance grows with $\phi$ and $\lambda$, whereas for the NB it grows with $\mu$ and the inverse of $k$.

Farnsworth et al. (2002) showed how the temporally stratified counts used for this study allow estimation of detection probabilities by application of a removal model. Their model has two parameters ($c$ and $q$). The first parameter, $c$, is the probability that a bird is hard to detect, and all easy-to-detect birds are assumed to be detected during the first three minutes of sampling. The second parameter, $q$, is the probability that a hard-to-detect bird is *not* detected during one minute of sampling (Farnsworth et al. 2002). Thus the overall probability that a bird is detected during 10 minutes of sampling is $1 - cq^{10}$, which we will refer to throughout the article as the detection probability ($p$). Farnsworth et al. (2002) listed five assumptions underlying the removal model, which we abbreviate as follows, (1) population closure, (2) no double counting, (3) easy-to-detect birds are detected in the first 3 minutes, (4) $q$ is constant within counts, and (5) observers correctly assign birds to distance classes. To these we would add a sixth assumption that birds are correctly identified to species.

Royle (2004a) elaborated a likelihood approach for estimating abundance conditional on both an assumed distribution for abundance and an underlying detection process (in this case the removal model). This framework allows simultaneous estimation of detection parameters ($c$ and $q$), abundance ($\mu$ or $\lambda$) and overdispersion ($\phi$ or $k$), conditional on the temporally stratified counts ($x_1$, $x_2$, and $x_3$). For these analyses we maximized Royle's (2004a:381) integrated likelihood over our point count data using MATLAB's (Math-Works 2007) optimization toolbox with Poisson, ZIP, and NB assumptions, as previously described.

To compare trend inference for counts vs. abundance under each of these six likelihoods ("Poisson counts," "ZIP counts," "NB counts," "Poisson mixture," "ZIP mixture," and "NB mixture") we chose 16 songbirds that we thought would vary in detectability, abundance, trends, and stand affinities (Appendix B).

### Hierarchical models

For all six likelihoods we specified an abundance parameter from each distribution as a log-linear function of covariates (Appendix B). For unadjusted counts, the resulting models were Poisson regression, zero-inflated Poisson regression, and negative binomial regression, respectively. (Although the log-link is not the canonical link for negative binomial regression, it is nevertheless the most commonly used link function; Hilbe 2007). This approach was also used to investigate trends in the detection parameters ($c$, $q$) and dependence of the overdispersion parameters ($k$ and $\phi$) on stand types. For $c$, $q$, and $\phi$ we used a logit-link, and for $\lambda$, $\mu$, and $k$ we used a log-link. Throughout this paper we treat $c$, $q$, $k$, and $\phi$ as nuisance parameters; our interest in them is restricted to their potential influence on inference about abundance and trends.

### Strategy for modeling abundance

For all hierarchical models we adopt the notation $\theta\{y\}$, indicating that likelihood parameter "$\theta$" is taken to be a function of covariate "$y$," where the dot notation "$\theta\{\cdot\}$" indicates that the parameter is assumed to be constant. When referencing a model for an abundance parameter we will generally write "$a\{y\}$," where $a$ indicates $\lambda$ for the Poisson and ZIP models and $\mu$ for the NB models. For each species, under each of the six likelihoods, we compared 24 models that specified abundance as varying functions of stand type, national forest, and year (Appendix B). Lowland deciduous forest was not surveyed in Superior National Forest because it is a rare type and thus is not represented in these data. Because we were certain that mean abundance would vary with either stand type, $a\{st\}$, or national forest, $a\{fr\}$, all models in the full model set included at least one of these two as a categorical covariate (Appendix B). We also hypothesized that abundance would vary with year, but the most interesting annual variation would be a trend (either linear or quadratic). Thus, when year was included as a covariate to abundance, we specified it as a continuous variable, $a\{yq\}$ (Appendix B). We also included models with quadratic trends, $a\{yq + yq^2\}$. We further considered a limited number of models containing interactions, particularly interactions between national forest and stand type, $a\{fr \times st\}$, between either stand type and year, $a\{yq(st)\}$, or forest and year, $a\{yq(fr)\}$, and two models with three-way interactions, $a\{yq(fr \times st)\}$, $a\{yq^2(fr \times st)\}$ (Appendix B). Although the models with interactions, particularly the three-way interactions, contained many parameters (Appendix B), previous work had suggested that some species would likely show discordant trends at regional scales (Blake et al. 1994).

### Strategy for modeling detection probability

To limit the number of detection probability models, we always optimized the same covariate function for $q$ as was used for $c$. Thus $p\{yq\} \equiv c\{yq\}$, $q\{yq\}$. To more closely examine the pattern of annual variation in detection probability, we specified year first as a continuous covariate, $p\{yq\}$, and second as a categorical

TABLE 1. Distribution parameters, number of models considered, best performing distribution, and mean number of models required to account for 99% of model weights among 16 bird species analyzed in national forests of the western Great Lakes, USA.

| Metric | Counts | | | Mixtures | | |
|---|---|---|---|---|---|---|
| | Poisson | ZIP | NB | Poisson | ZIP | NB |
| Abundance parameter | $\lambda$ | $\lambda$ | $\mu$ | $\lambda$ | $\lambda$ | $\mu$ |
| Detection parameters | n/a | n/a | n/a | $c, q$ | $c, q$ | $c, q$ |
| Overdispersion parameter | n/a | $\phi$ | $k$ | n/a | $\phi$ | $k$ |
| No. models considered | 24 | 48 | 48 | 72 | 144 | 144 |
| Best distribution (of 16 spp.) | 0 | 16 | 0 | 2 | 12 | 2 |
| Mean no. models in "best" set (SD) | 1.3 (0.6) | 2.4 (1.9) | 1.9 (1.1) | 1.9 (1.0) | 3.5 (3.9) | 2.4 (1.4) |

*Notes:* The influence of detection, heterogeneity, and overdispersion on estimated trends was assessed by comparing trend estimates from raw counts with corrected trend estimates from removal mixtures, modeled under Poisson, zero-inflated Poisson, and negative binomial statistical distributions; n/a means not applicable. Full likelihood details are provided in Appendix A.

covariate $p\{yc\}$. However, we also suspected that for some of our 16 selected species there would not be enough data to fully parameterize annual variation in detection, and we also included models with constant detection probabilities across years, $p\{\cdot\}$ (Appendix B). We did not consider models with $c$ constrained to unity, which we deemed implausible.

### Strategy for modeling overdispersion

We assumed that overdispersion in our data was caused primarily through unexplained environmental heterogeneity and that this might differ across our broad stand types. Therefore overdispersion parameters were modeled either as fixed across all stand types, $k\{\cdot\}$, $\phi\{\cdot\}$, or separately for each stand in each forest $k\{fr \times st\}$, $\phi\{fr \times st\}$. In preliminary analyses we had difficulty optimizing the NB distribution with large $k$ (little overdispersion) and in general we found it more efficient to optimize its inverse (i.e., set $r = 1/k$ and substitute $1/r$ into the NB likelihoods). With covariate models this was achieved by changing the sign of the log-link and the gradient of the link function (Appendix A). For both ZIP and NB models, upper and lower limits were placed on $\phi$ and $k$ to facilitate convergence. In most cases an optimal solution was found within this range and we took this to be the global optimum (MLE, maximum likelihood estimate). In a few cases, the optimal solution occurred at one of the imposed limits.

### Model selection and comparisons

Our final model sets for each of our six likelihood approaches contained varying multiples of the set of 24 abundance models (Table 1). For each species and each likelihood we used Akaike's Information Criterion, AIC (Akaike 1973), to rank all models; we report only the minimal set of models that account for ≥99% of model weights, where weights are calculated following Burnham and Anderson (2002). For all likelihoods, all combinatoric terms were retained to allow comparison of AIC values across Poisson, ZIP, and NB distributions that specify the same response data (i.e., to compare the raw count models to each other and to compare the removal mixtures to each other across the three

distributional assumptions, Burnham and Anderson 2002:318).

## RESULTS

### Presentation of results

Here we will present selected results to highlight important patterns. Full results for all species are provided in online appendices. We begin with a brief survey of basic patterns of model selection. We then provide detailed results about detection probability, overdispersion, and finally, abundance patterns in relation to observed results for detection and over-dispersion. All statistical computing was done in MATLAB 7.4 (MathWorks 2007). Note that plots of overdispersion ($k$) for the NB distribution show $r = 1/k$, which is how the NB likelihoods were optimized. Visually this also facilitates comparison of the ZIP results to NB results because $\phi$ and $r$ should both increase with variance in their respective distributions.

### General distributional and model selection patterns

For all 16 species, the best ZIP count model had a lower AIC than the best Poisson or NB count models (Table 1). For two species (Red-eyed Vireo and Black-and-white Warbler) the best Poisson mixture had a lower AIC than the best ZIP or NB mixture. Similarly, for two species (Yellow-rumped Warbler and White-throated Sparrow) the best NB mixture had a lower AIC score than the best Poisson or ZIP mixture. The average number of models per species required to account for at least 99% of model weight was greatest for ZIP models (both counts and mixtures) and least for Poisson models (both counts and mixtures; Table 1). Within distributions, the detection mixtures required a larger average number of models per species than the count models to account for at least 99% of model weight (Table 1). The standard deviation of the number of models required to account for at least 99% of model weight followed the same patterns (Table 1).

### Detection probabilities

The total number of detections by species varied almost 20-fold, from a minimum of 1223 (Golden-
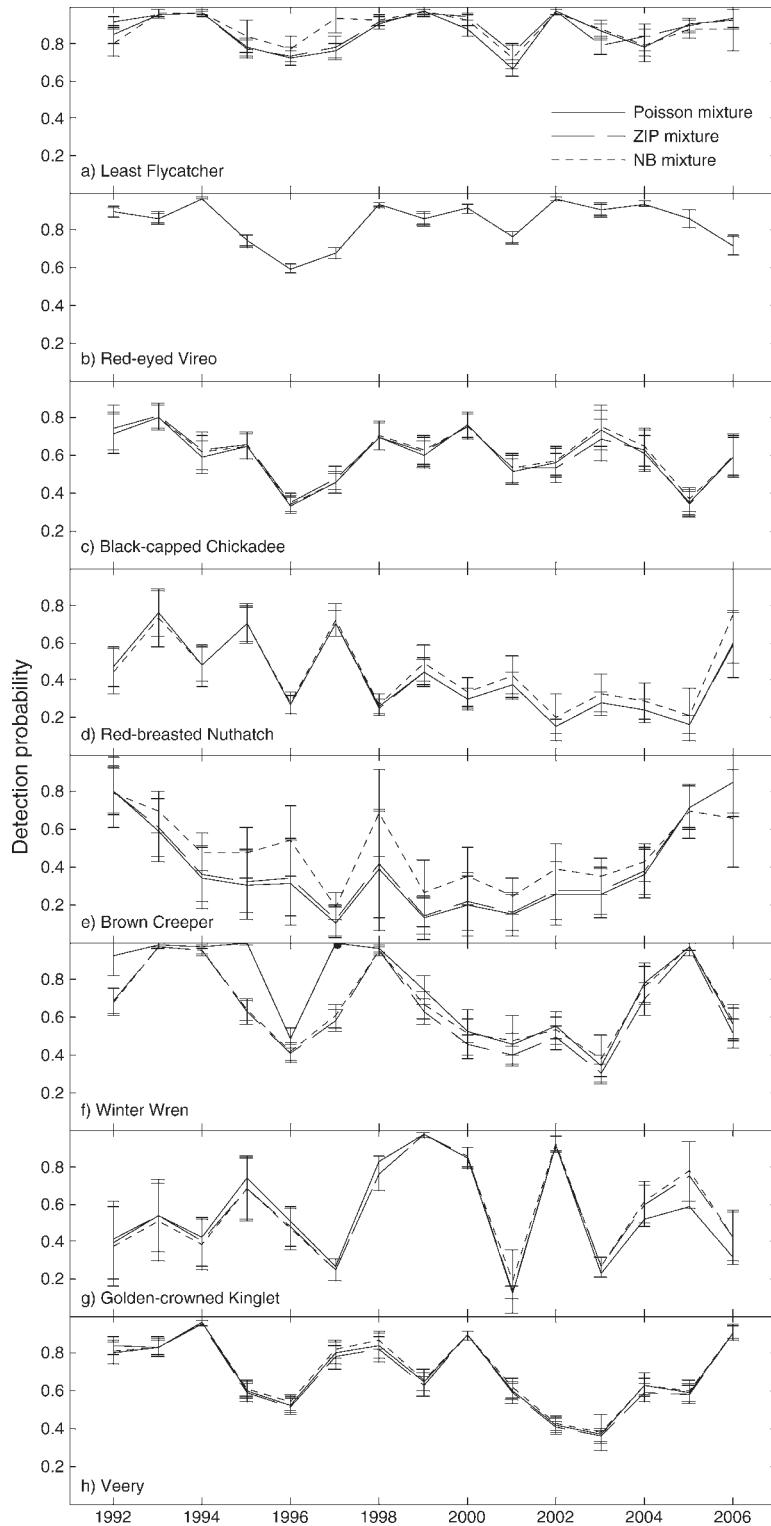
FIG. 2. Fitted overall detection probability (mean ± SE) by year, 1992–2006, for 16 bird species surveyed in the three national forests using three removal mixtures modeled under Poisson, zero-inflated Poisson (ZIP), and negative binomial (NB) statistical distributions.
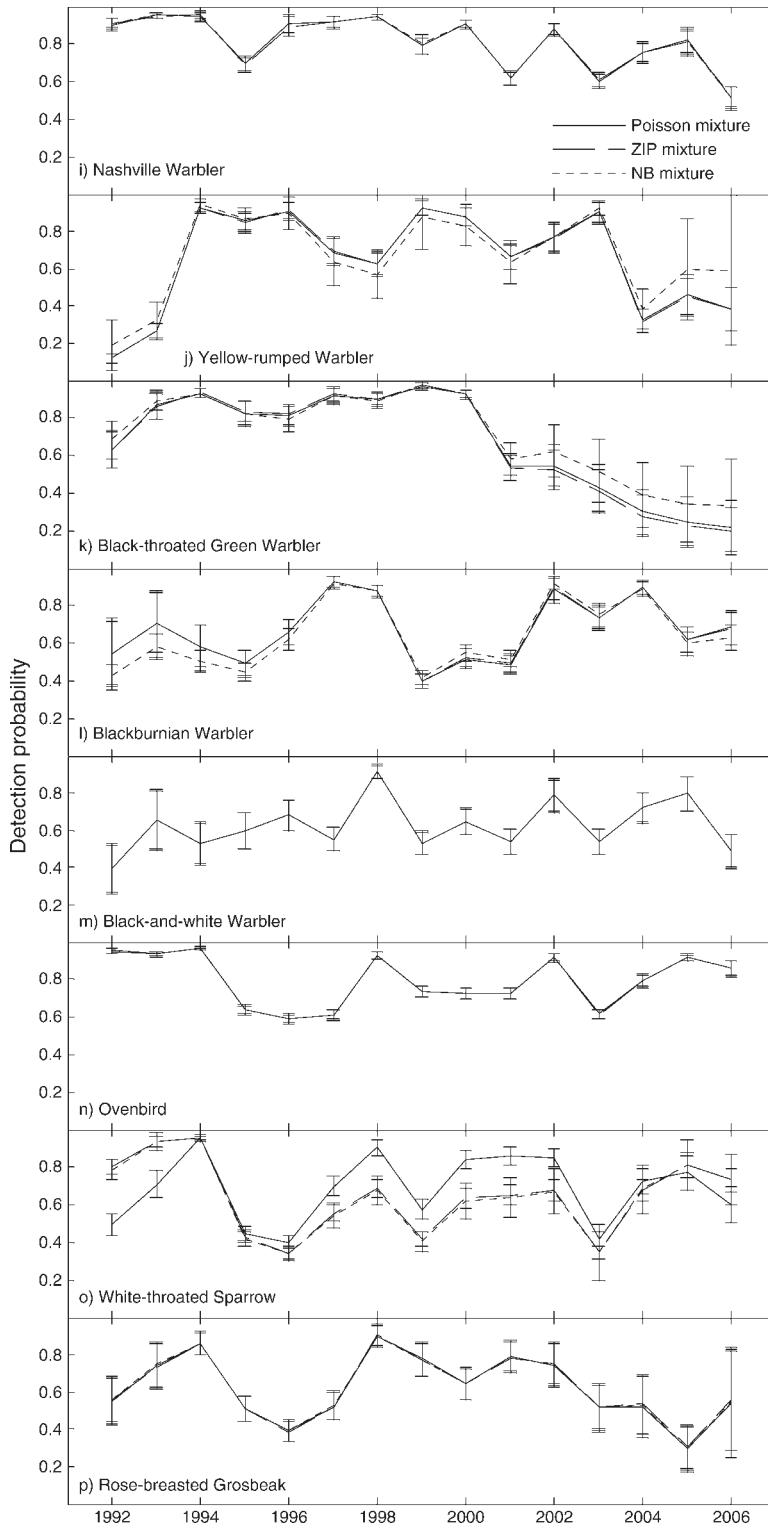
Fig. 2. Continued.
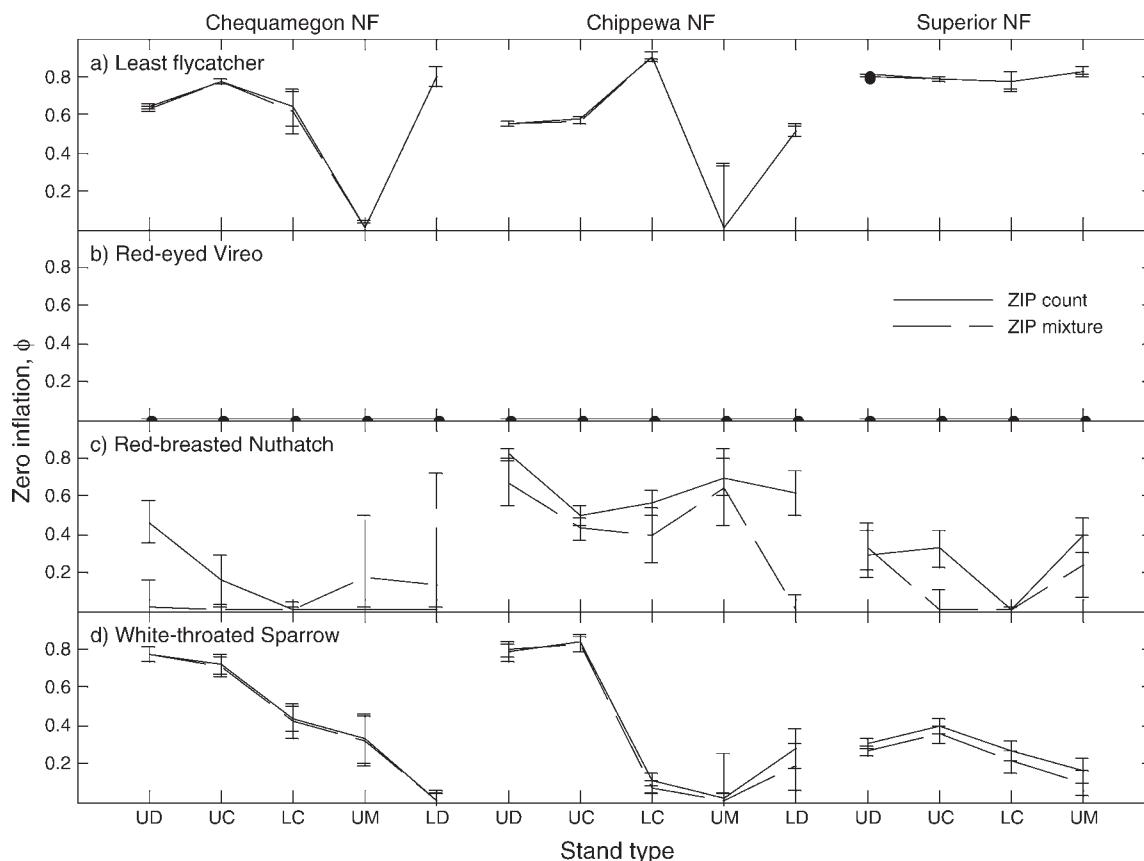
FIG. 3.  Zero-inflation patterns (mean ± SE) for selected species by national forest (NF: Chequamegon, Chippewa, Superior) and stand type (upland deciduous, UD; upland conifer, UC; lowland conifer, LC; upland mixed, UM; lowland deciduous, LD). The y-axis gives the actual estimated value of the zero-inflation parameter ($\phi$), which can be interpreted as the proportion of sites that are "unsuitable."

crowned Kinglet) to a maximum of 21 039 (Ovenbird; Appendix B). The estimated overall probability that a species was detected during 10 minutes of sampling ($\hat{p} = 1 - \hat{c}\hat{q}^{10}$) ranged from 0.105 (Brown Creeper, 1997, Poisson mixture) to 0.981 (Winter Wren, 1993, NB mixture; Fig. 2). Regardless of whether the underlying abundance distribution was assumed to be Poisson, ZIP, or NB, the best model for detection probability always included yearly variation as a categorical covariate (Appendix B). The models $p\{\cdot\}$ and $p\{yq\}$ (i.e., detection probability is constant or has log-linear trend, respectively) were never competitive. Fig. 2 shows the magnitude of annual variation in estimated detection probabilities for all 16 species. Year-to-year variation in detection probability was large, sometimes extreme (e.g., Golden-crowned Kinglet; Fig. 2g). Over the full 15 years, no species showed monotonic trends in detection probability, although for some (e.g., Black-capped Chickadee, Nashville Warbler, and Black-throated Green Warbler; Fig. 2c, i, k) the model $p\{yq\}$ would show a decreasing trend in detection probability. Over shorter timescales (3–5 years) fitted detection probabilities often showed alternating patterns of steep declines and increases (Fig. 2). In general, the fitted detection

probabilities were little influenced by the assumed distribution for abundance (close agreement of solid, dashed, and dotted lines; Fig. 2).

### Overdispersion

Estimates of zero inflation ($\phi$) were usually congruent, and often virtually identical, between ZIP counts and ZIP mixtures (Fig. 3). A notable exception occurred with Red-breasted Nuthatch (Fig. 3c). Fitted values of $\phi$ ranged from 0.007 (the lower estimation constraint, Red-eyed Vireo and Winter Wren, all stand types, Fig. 3b) to 0.993 (the upper estimation constraint, Golden-crowned Kinglet, deciduous forest stands, Appendix C). Patterns of zero inflation among stand types generally differed among forests (Fig. 3). Visual comparison of patterns suggests a greater similarity between Cheque-megon and Chippewa NF than between either of those two forests and Superior NF (Fig. 3, Appendix C).

As with zero inflation, estimates of $r$ ($1/k$) were often congruent between NB counts and NB mixtures (Fig. 4). Fitted values of $r$ varied considerably from 0.01 (the lower estimation constraint, little or no overdispersion, Red-eyed Vireo, Winter Wren, Black-and-white Warbler; Fig. 4b, Appendix C) to approximately 54.6 (the
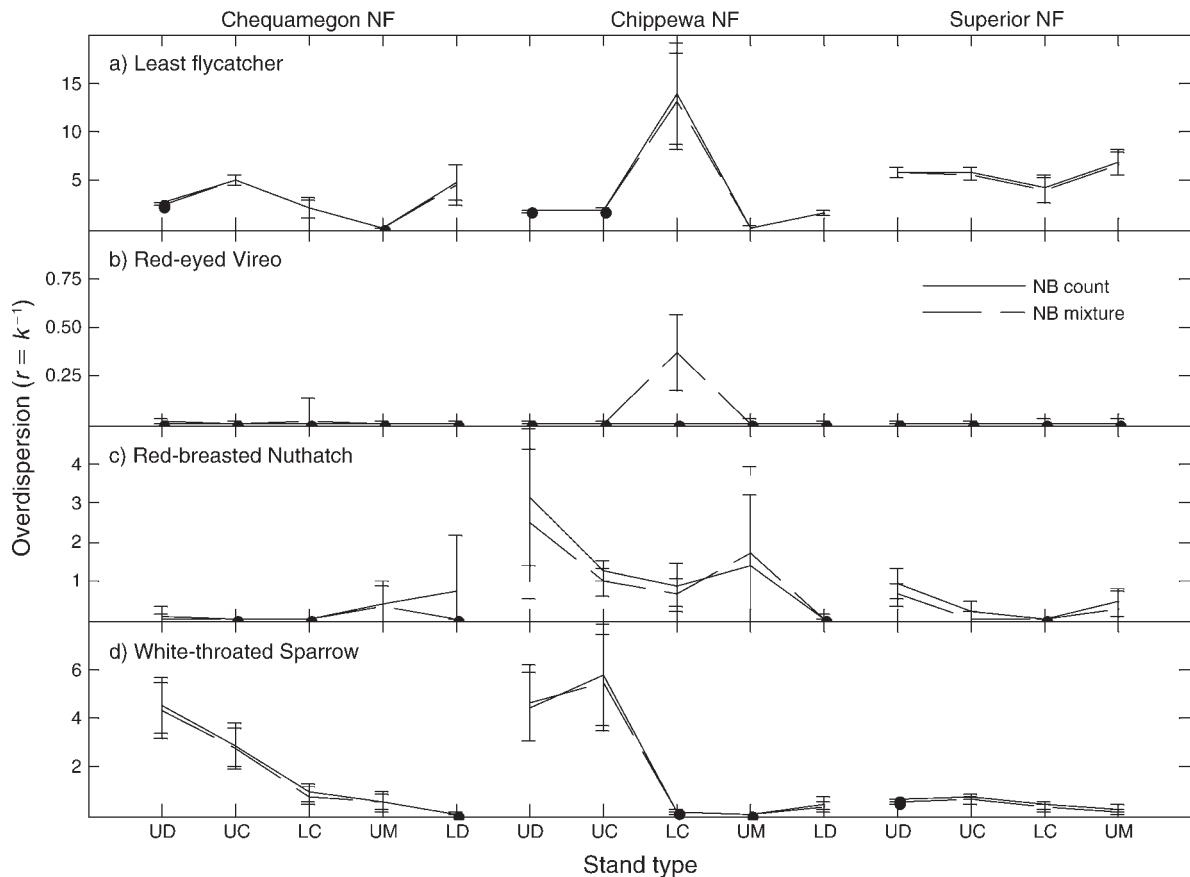
FIG. 4. Overdispersion patterns (mean $\pm$ SE) for selected species by national forest and stand type. Top and bottom labels are as in Fig. 3. The estimated parameter $k$ is the negative binomial overdispersion parameter. Variance in the negative binomial distribution grows with $r$, the inverse of $k$.

upper estimation constraint, extreme overdispersion, Golden-crowned Kinglet, deciduous forest stands, Chippewa NF, Appendix C). Estimated patterns of overdispersion varied among forests, and the similarity between Chequamegon and Chippewa NFs for zero inflation did not appear to hold for NB overdispersion. In some cases, patterns of NB overdispersion were closely congruent to patterns of zero inflation (e.g., White-throated Sparrow; Figs. 3d and 4d), but this was not generally the case.

The two species for which the best Poisson mixture had a lower AIC score (Red-eyed Vireo and Black-and-white Warbler) gave estimates of $r$ that were virtually zero across the full time series (Fig. 4b, Appendix C). For Red-eyed Vireo, zero inflation was also close to zero, although this was not true for Black-and-white Warbler (Appendix C). For the two species for which the best NB mixture had a lower AIC score (Yellow-rumped Warbler and White-throated Sparrow, Appendix C, Figs. 3d and 4d) patterns of overdispersion ($r$) were similar to patterns for $\phi$, although in neither case (ZIP or NB mixtures) did the fitted values represent extremes among all species analyzed.

For most species showing high zero inflation (large $\phi$) and/or high overdispersion (large $r$), the estimated value of the respective parameter was usually smaller for the best mixture than it was under the respective count model (Figs. 3 and 4). This suggests that some overdispersion does result from detection heterogeneity. However, we did not see any evidence that controlling for detection heterogeneity effectively removed overdispersion altogether.

### Abundance

Estimates of mean counts ($\lambda$, $\mu$) ranged from zero (Least Flycatcher, upland mixed forest, all years, Chippewa NF, all distributions) to 2.35 (Ovenbird, all years upland conifer, Chequamegon NF, ZIP counts). Estimates of mean detection-corrected abundance ($\lambda$, $\mu$) ranged from zero (Least Flycatcher, upland mixed forest, all years, Chippewa NF, all distributions) to 5.10 (Black-throated Green Warbler, 2006, upland mixed forest, Chippewa NF, Poisson mixture). Of the 24 abundance models considered (Appendix B), all but two ($a\{st + fr\}$ and $a\{st + fr + yq\}$) occurred at least once in the best model set for at least one combination

TABLE 2. Frequency of occurrence of bird abundance models in the best model set.

| Abundance model | Occurrences | Frequency |
|---|---|---|
| $a\{st \times fr + yq(st \times fr) + yq^2(st \times fr)\}$ | 74 | 0.35 |
| $a\{st \times fr + yq(fr) + yq^2(fr)\}$ | 48 | 0.23 |
| $a\{st \times fr + yq(st \times fr)\}$ | 20 | 0.09 |
| $a\{st \times fr + yq(fr)\}$ | 15 | 0.07 |
| $a\{st + fr + yq(fr) + yq^2(fr)\}$ | 7 | 0.03 |
| $a\{st \times fr + yq + yq^2\}$ | 6 | 0.03 |
| $a\{fr + yq(fr) + yq^2(fr)\}$ | 5 | 0.02 |
| $a\{st + fr + yq(st) + yq^2(st)\}$ | 4 | 0.02 |
| $a\{st + yq + yq^2\}$ | 4 | 0.02 |
| $a\{st + yq(st) + yq^2(st)\}$ | 4 | 0.02 |
| $a\{st + fr + yq(fr)\}$ | 4 | 0.02 |
| $a\{st \times fr + yq\}$ | 4 | 0.02 |
| $a\{fr + yq(fr)\}$ | 3 | 0.01 |
| $a\{st + fr + yq + yq^2\}$ | 3 | 0.01 |
| $a\{st + fr + yq(st)\}$ | 3 | 0.01 |
| $a\{fr + yq + yq^2\}$ | 2 | 0.01 |
| $a\{st + yq(st)\}$ | 2 | 0.01 |
| $a\{fr\}$ | 1 | <0.01 |
| $a\{fr + yq\}$ | 1 | <0.01 |
| $a\{st\}$ | 1 | <0.01 |
| $a\{st \times fr\}$ | 1 | <0.01 |
| $a\{st + yq\}$ | 1 | <0.01 |
| Total | 213 | 1 |

*Notes:* Model terms are *a*, abundance; st, stand type (upland deciduous, upland conifer, lowland conifer, upland mixed, and lowland deciduous); fr, national forest (Chequamegon, Chippewa, Superior); yq, year (1992–2006), treated as a quantitative variable; and yc, year treated as a categorical variable; $\{\cdot\}$ indicates an intercept-only model. An interaction between two covariates is indicated by "×," yq(fr) indicates that the year effect is estimated separately for each national forest. All interactive models also contained the corresponding main effects, and all models contained a global intercept.

of species and likelihood (Table 2). However, only four abundance models occurred in the best model set with a frequency greater than 0.05 (Table 2) and these four models always included the stand type × forest interaction, $a\{st \times fr\}$. These four models also always included a nested linear trend, either within forest or within the stand type × forest interaction (Table 2).

Plotted trends can be quickly separated into two classes. First, for some species, estimated trends differed little between distributions (e.g., Ovenbird; Fig. 5) or between count models vs. removal mixtures (Red-eyed Vireo; Fig. 6) or both (Least Flycatcher; Figs. 7–9). The three most commonly counted species (e.g., Red-eyed Vireo, Nashville Warbler, and Ovenbird) fall in this group (Appendix C). Second, for other species, plotted trends differed by distributional assumption and likelihood method in at least one national forest (Brown Creeper, Fig. 10; Veery, Appendix C). In general, when differences occurred, the ZIP model trends usually differed from the Poisson or NB trends more than the latter two differed from each other (e.g., Least Flycatcher, Figs. 7–9; Black-capped Chickadee, Red-breasted Nuthatch, and Blackburnian Warbler, Appendix C).

Two consistent upward shifts can be observed on all trend graphs. First, within distributions, the mixtures

are shifted upward relative to the count models, reflecting the correction for imperfect detection that occurs with the removal mixtures. As you would expect, this shift is greatest for those species with generally lower estimated detection probabilities (e.g., Red-breasted Nuthatch, Brown Creeper, Veery, and White-throated Sparrow, Appendix C). However, it is also evident in the most commonly detected species (Ovenbird, Red-eyed Vireo, Figs. 5 and 6; Nashville Warbler, Appendix C). A second shift can be seen in most comparisons between ZIP trends and either Poisson or NB trends. The most conspicuous example occurs with Least Flycatcher (Figs. 7–9). This upward shift occurs because a substantial fraction of sites with zero counts (i.e., φ) is not included in the plotted abundance distribution.

AIC often selected the same structural abundance model as the best model across distributions and across count vs. mixture models. Extreme examples of this occurred with Nashville Warbler and Ovenbird, for which a single abundance model accounted for more than 99% of model weight in all six likelihood groups (Appendix B). These are two of the three most commonly counted species in our study. The greatest exceptions to this pattern usually occurred in comparisons between ZIP models and Poisson or NB models. For two species (Black-capped Chickadee and Brown Creeper), the most parsimonious ZIP model (both count and mixture) included no effect of stand type on abundance (Appendix B). Thus under the ZIP models for these two cavity-nesting species, all differences among stand types were accounted for by the zero-inflation factor (Fig. 10, Appendix C).

Three species, Brown Creeper, Yellow-rumped Warbler, and Black-throated Green Warbler, showed temporal patterns of detection that did have a substantial effect on the resulting estimate of abundance trends. For Brown Creeper, the best detection model estimated relatively high detection probability early and late in the time series, but low detection probability in between (Appendix C). This changed the relatively flat count-estimated trends to hump-shaped trends in the mixtures (Fig. 10). The opposite pattern occurred with Yellow-rumped Warbler (Appendix C). Black-throated Green Warbler showed a general trend of decreasing detection probability over the 15 years of the study (Fig. 2k). This resulted in a considerable correction in the abundance curve at the end of the time series for the removal mixtures, although the count models also showed an upward inflection at the end of the time series (Appendix C). The Black-throated Green Warbler results also illustrate a more general pattern across species, in which increases or decreases in detection probability at the beginning or end of the time series imposed much stronger pattern on the corrected abundance trajectories than when they occurred in the middle of the time series (Appendix C).
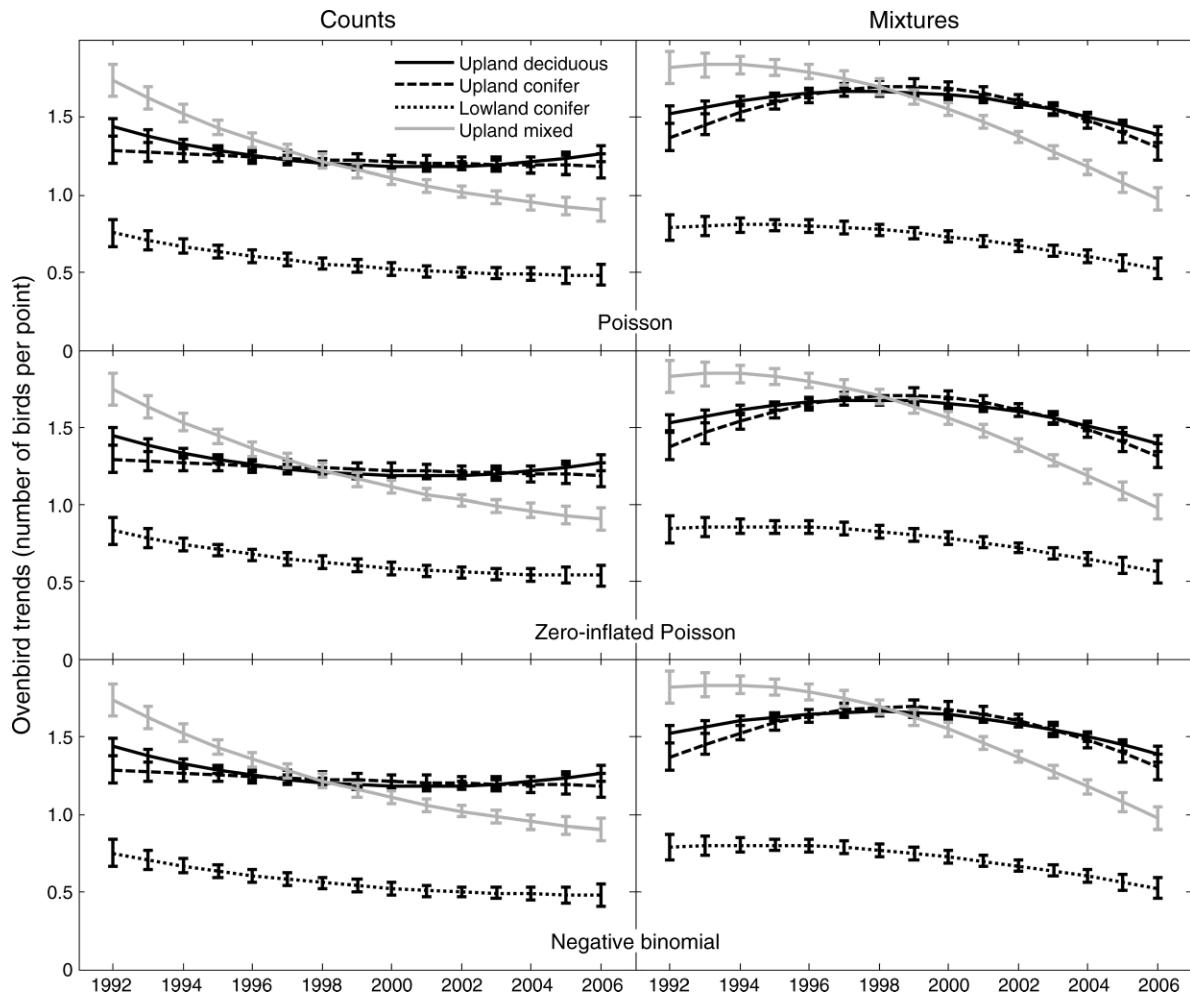
Fig. 5. Ovenbird trends (number of birds per point, mean ± SE) in Superior National Forest, Minnesota, USA, in different stand types, for counts vs. mixture models. The *y*-axis shows trends in the expected number of birds per point under the three assumed distributions (Poisson, zero-inflated Poisson, and negative binomial). Count models use only birds counted to fit the distributions, whereas the mixture models correct for imperfect detection.

## DISCUSSION

### Detection probabilities

For all species in all removal mixtures, the best model for detection probability included annual variation as a categorical effect, $p\{yc\}$. Neither the constant model, $p\{\cdot\}$, nor one of continuous annual variation, $p\{yq\}$, was competitive. Therefore strongly monotonic trends in detection probability did not appear to occur over the full 15-year time series. Nevertheless, some patterns in detection probability did result in disagreement between trends in counts and trends in removal mixtures. Also, there was large annual variation in detection probability (Fig. 2), which would tend toward violations of Johnson's (2008) second condition that trends in counts should accurately reflect trends in abundance when $\mathrm{Var}(p)$ is small in relation to $\mathrm{Var}(N)$. The range of variability that we observed using removal estimates of detectability is similar to that reported by Pacifici et al.

(2008) using an experimental bird-song simulation system in conjunction with distance estimation.

Several species showed a conspicuous pattern of high detectability during the first three years of sampling (Fig. 2). This might be due to a change in sampling protocol that occurred between the 1994 and 1995 surveys. During the first three years observers were instructed to count only those birds within 100 m of themselves and to ignore birds outside this detection radius. Beginning in 1995, observers were instructed to count all birds, regardless of distance, but to record whether or not the bird was within 100 m. For 1995 and all subsequent years, we included only detections classified within 100 m, so the data should be consistent with 1992–1994. However, it is possible that, when told to ignore birds outside the 100-m detection radius, observers were more likely to classify birds close to the boundary as being inside rather than outside the boundary.
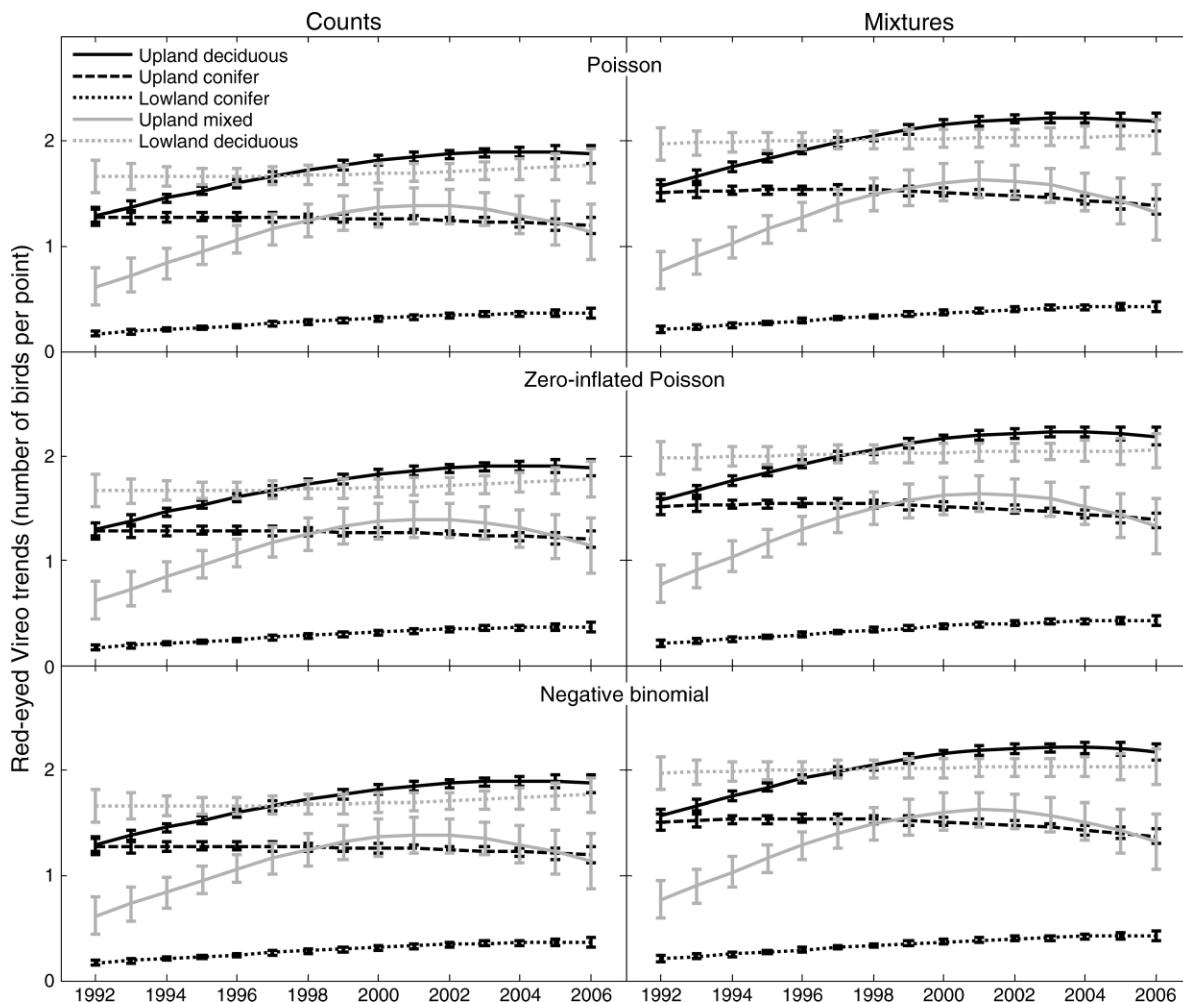
FIG. 6. Red-eyed Vireo trends (number of birds per point, mean ± SE) in Chippewa National Forest, Minnesota, USA.

If real, this heaping effect may be a good example of the potential for methodology to interact with observer psychology to result in misclassification of birds as being within the 100-m detection radius (Rosenstock et al. 2002). However, if the heaping effect occurred, it did not appear to have a great effect on inference about abundance, probably due to the relative length of our total time series compared to the three-year period during which heaping may have occurred. Nevertheless, our results support the recommendation of Rosenstock et al. (2002) that when performing variable-distance circular point counts, an unlimited distance ring should be included even if the corresponding data are not used. Our results reinforce this recommendation regardless of whether distance sampling per se is the intended methodology.

### The removal model

Strictly speaking, we cannot be certain that any of the six assumptions that we have listed are met in our data (Alldredge et al. 2007*a*, Johnson 2008), although we

believe any violations were minor. Indeed, we have presented evidence for heaping during the first three years of our study that suggests observers were not correctly classifying birds to within 100 m of the observer, but this is probably much less an issue since 1995. Of the six assumptions listed in *Methods*, the two that are unique to the removal likelihood are probably the most questionable (all easy-to-detect birds detected in the first three minutes, and constant *q*). These assumptions could be relaxed under the capture–recapture formulation of the time-to-detection methods described by Alldredge et al. (2007*a*, *c*). However, our data were not collected in a way that will allow comparative analyses; the original design motivation was intended for comparing data across studies using different sampling durations, not for estimating detection probability. Nevertheless the analyses do recover information about detectability that conforms to our expectations. Birds with conspicuous, loud songs are generally classified with high detection probability and birds with high-pitched, quiet songs are usually classified
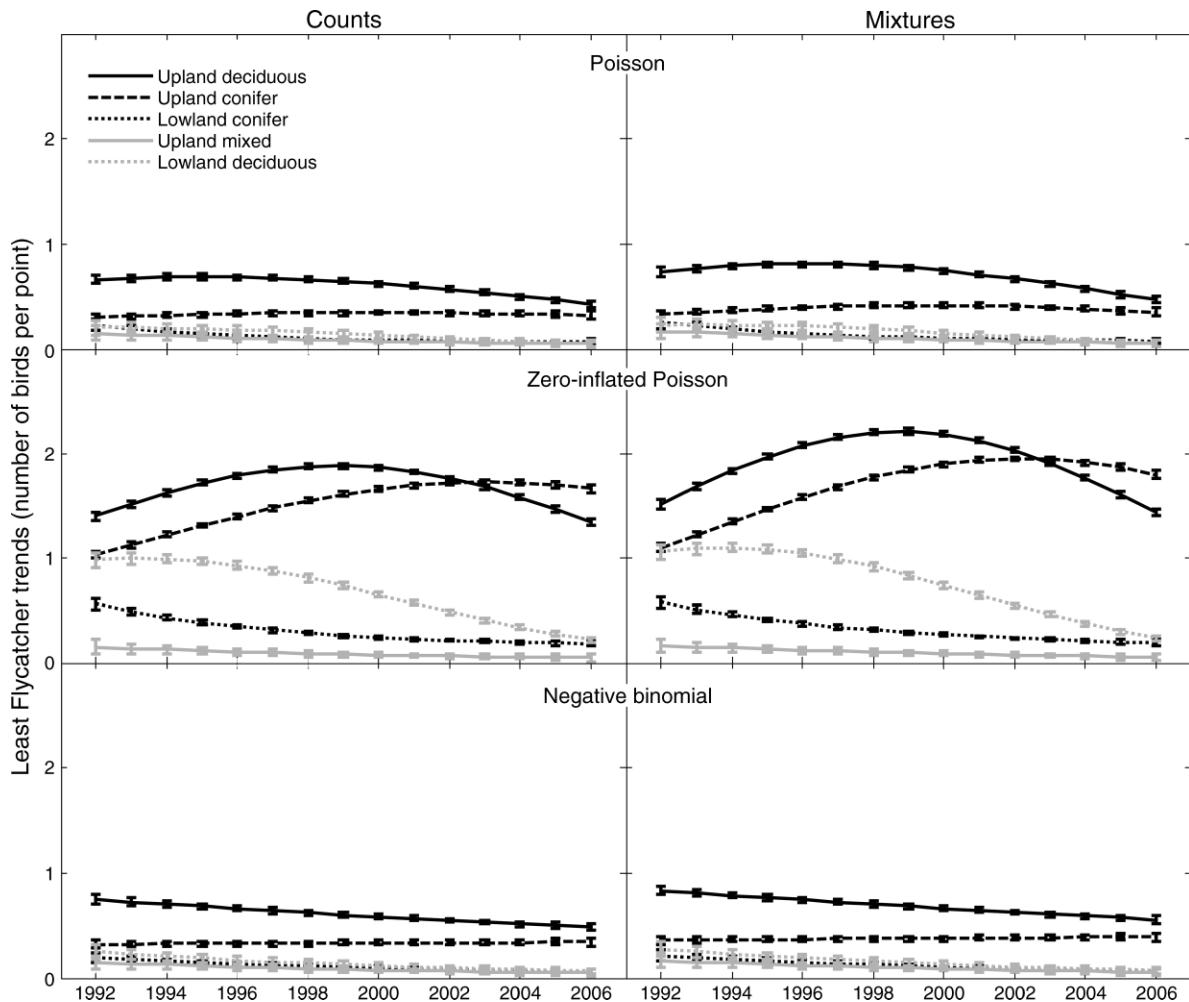
FIG. 7. Least Flycatcher trends (number of birds per point, mean ± SE) in Chequamegon National Forest, Wisconsin, USA.

with low detection probability (Fig. 2), although the latter species usually have more pronounced annual variation in detectability.

### Overdispersion

Two patterns in our results reinforce our hypothesis that overdispersion in our point counts is caused by environmental heterogeneity rather than detection heterogeneity. First, common and highly detectable species are less likely to show overdispersion than are rare and/or difficult to detect species. This is in contradiction to the expectation based on the derived relationship between variance in detection probability and abundance presented earlier:

$$\left\{ \mathrm{Var}(p) > \frac{E(p)[1 - E(p)]}{1 + \lambda} \right\}.$$

Second, the ZIP distribution was preferred for count analyses for all species and for mixtures for most (75%) species. Thus there was a clear excess of zero counts in our data, suggesting that some proportion of habitat

(often substantial) within our broad stand types was unsuitable. Least Flycatcher and Golden-crowned Kinglet are extreme examples (Figs. 7–9, Appendix C).

The high performance of the ZIP distribution also has implications for occupancy modeling. In the ZIP mixtures there are three classes of points at which zero counts are expected. First, there are the $\phi$ empty sites, which we loosely attributed to the presence of unsuitable habitat within modeled covariate classes (national forest and stand type). Second, of the $1 - \phi$ suitable sites, a further fraction ($e^{-\lambda}$) are expected to be unoccupied. Finally, for occupied sites with $n$ birds present, no birds will be detected at approximately $(1 - p)^n = (cq^{10})^n$ of them. If these parameters ($\phi$, $c$, $q$, and $\lambda$) depend on different factors, as seems likely, then any model that tries to predict distributions using simple presence–pseudo-absence data (e.g., Elith et al. 2006) seems likely to find incomplete (at best) or spurious (at worst) covariate relationships. MacKenzie et al. (2006) provide examples of spurious conclusions about habitat relationships that arise through failure to account for
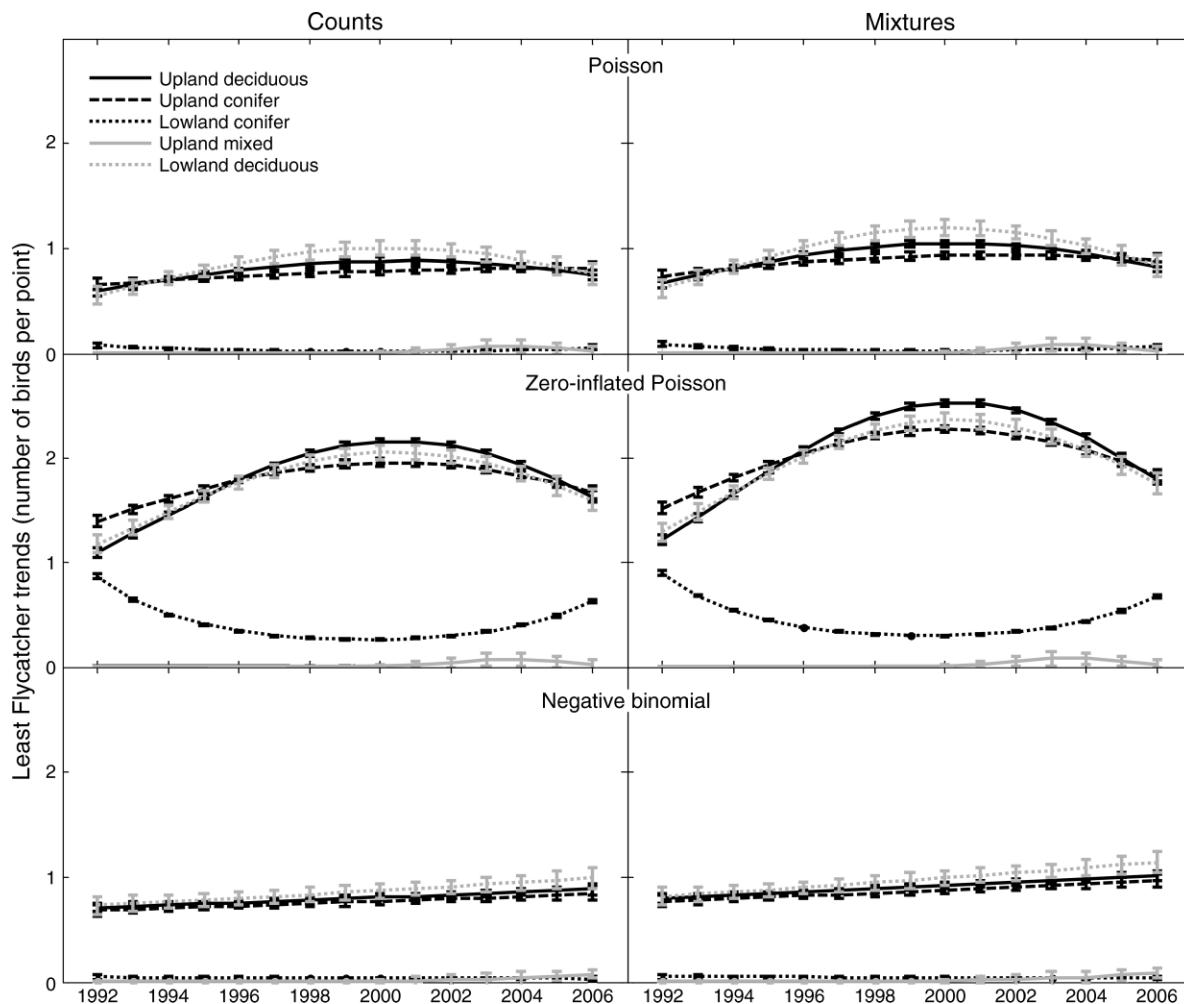
FIG. 8. Least Flycatcher trends (number of birds per point, mean ± SE) in Chippewa National Forest.

detection heterogeneity. Our results suggest that erroneous conclusions about habitat relationships could also arise through failure to properly account for patterns of overdispersion.

### Abundance

Our results showed that relatively little correction in abundance curves occurred between the raw count models and the removal mixtures (Figs 5–10, Appendix C). This may have occurred for several reasons. First, with 10-minute counts it is plausible that a relatively large proportion of individuals available for detection were actually detected. Second, the removal method may not fully correct for imperfect detection. Alldredge et al. (2007c) showed that a related method (time-to-detection) was negatively biased in experimental tests with auditory detection when used with 100 m radius circular point counts. Third, the method may not completely account for availability if there are some individual birds present that sing only rarely or not at all (Johnson 2008).

For most species, estimates of mean abundance using the ZIP distribution were substantially higher than with either the Poisson or NB distributions, and this was true across both counts and mixtures. This is in part due to our presentation of results. The ZIP means are means among the $1 - \phi$ "suitable" sites, whereas the means of the Poisson and NB distributions are means across all sites. Nevertheless, we found the ZIP means, with excess zeros removed, to be more informative of true habitat relationships and trends in the presence of overdispersion. We also found that the specific form of the most preferred abundance model (the preferred covariate model for abundance) as well as the shape of the abundance curve was strongly affected by whether or not the ZIP distribution was used (e.g., Figs. 7–10). Thus, like Ver Hoef and Boveng (2007), we found that the choice of method for modeling overdispersion could have a large influence on inference about covariates to abundance.
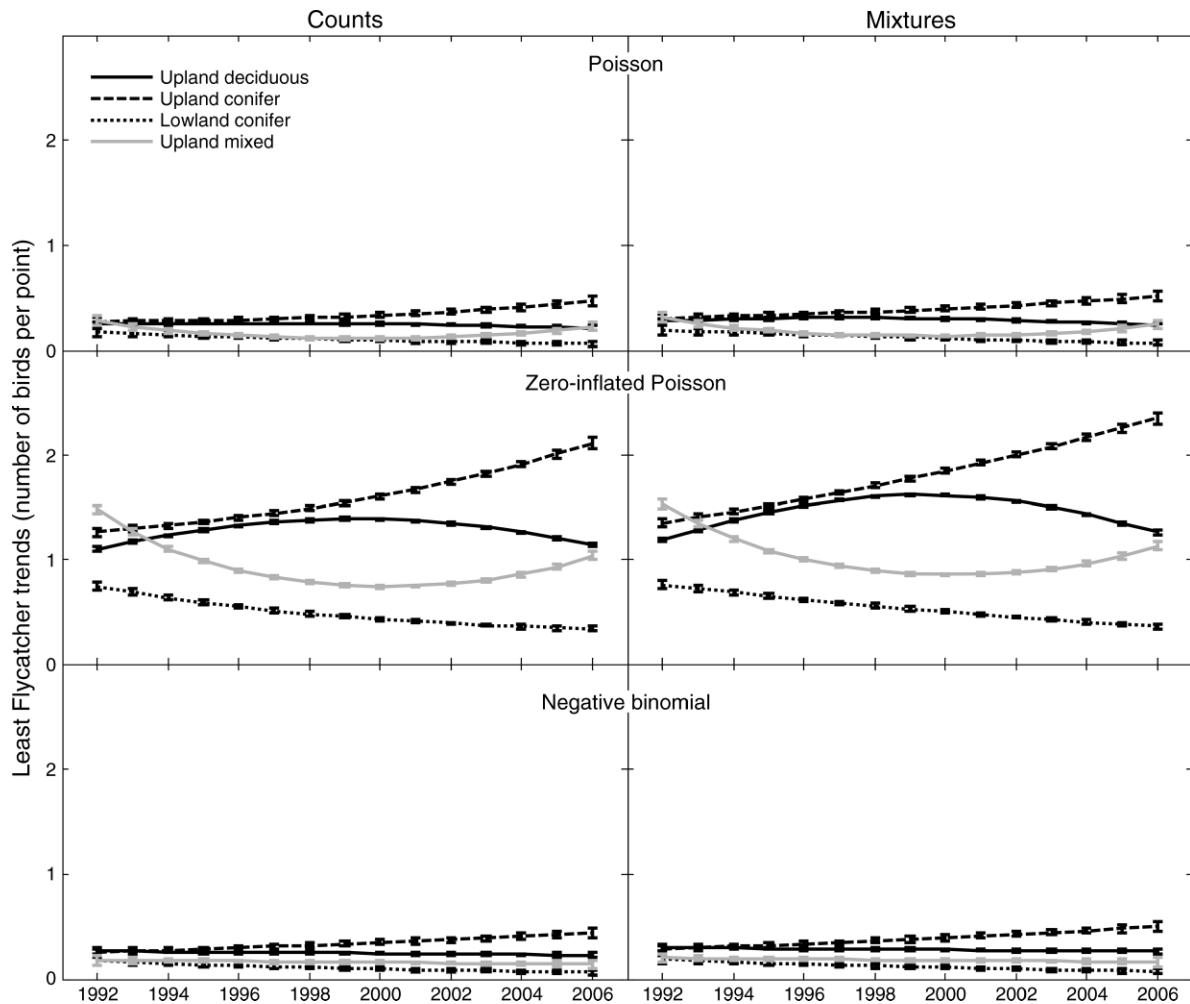
FIG. 9.   Least Flycatcher trends (number of birds per point, mean ± SE) in Superior National Forest.

In our results, trends inferred from counts were similar to trends inferred from the removal mixtures for most species. Therefore, either the analysis of raw counts provided a reliable estimate of trend or the removal mixtures did not. Several patterns in estimated detection probabilities argue in favor of the former. First, estimated detection probabilities generally conformed to our expectations across species, with conspicuous, easily detected species showing high detection probabilities, and species generally considered more cryptic showing low detection probabilities. Second, the change in our survey methodology in 1995 resulted in a predictable change in estimated detection probability that was confirmed in our analyses. Third, in our results the assumption of constant $p$ is never satisfied; the model $p\{\cdot\}$ is never competitive, but neither is $p\{yq\}$. Thus, with the exceptions noted here, heterogeneous detection probabilities did not impose systematic trends on counts. Our results broadly support the assertion of Johnson (2008) that even when the strong assumption of

homogeneity is not met, raw counts may be a reliable index of abundance.

### Why didn't we observe greater trend differences in counts vs. mixtures?

A growing body of evidence suggests that many factors, including weather, habitat structure, phenology, background noise, anthropogenic noise, time of day, and observer differences can all affect the detectability of birds during point counts (Alldredge et al. 2007b, Simons et al. 2007, Pacifici et al. 2008). How can we reconcile our results here with those results? First, our results do not prove that anthropogenic and environmental factors have little influence on detectability in our system. On the contrary, we believe that they do and are responsible for a considerable amount of the annual variation in detectability that is depicted in Fig. 2. Thus, we attribute much of the lack of differences in trends between counts and mixtures to the way in which our survey protocols were standardized during the design
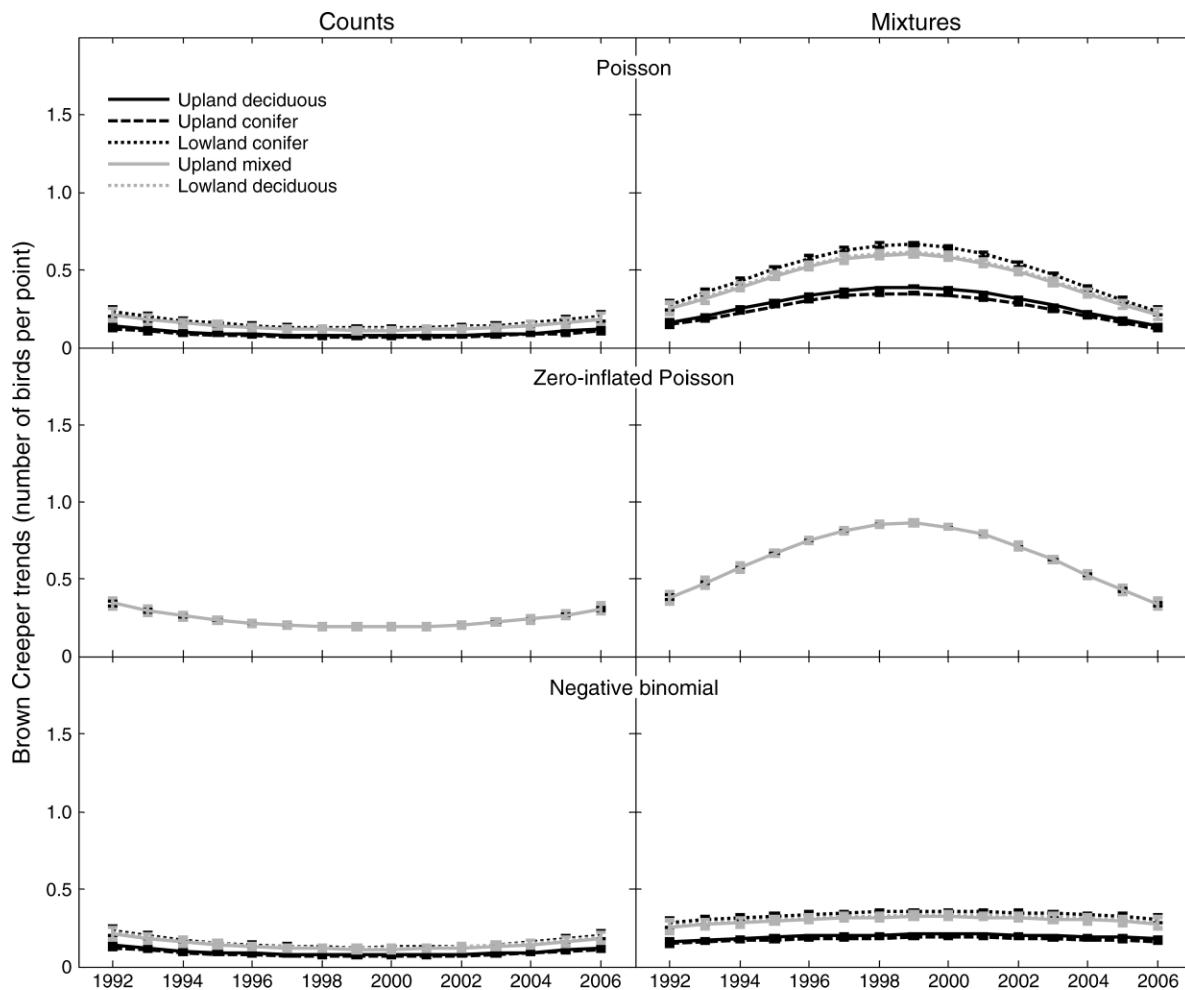
Fig. 10. Brown Creeper trends (number of birds per point, mean ± SE) in Chequamegon National Forest.

phase to anticipate detection heterogeneity. Clearly it would be useful to understand the conditions under which our comparative results might generalize, whether across species, studies, or both. We will now provide some heuristic guidance, noting in advance that the conditions we cite generally conform to the two criteria proposed by Johnson (2008): $p$ independent of $N$ and Var($p$) small in relation to Var($N$).

From a methodological point of view, there was little or no a priori reason to suspect systematic changes in detectability in our study. Our observers are trained consistently across years; they change from year to year; and within years they are distributed among habitats so as to avoid confounding observers with habitat types or particular points. Thus, changes in the ability of returning observers and differences among observers within years are far less likely to impose a pattern than, for example, in the North American Breeding Bird Survey (Sauer et al. 1994). Similarly, our points are all off-road, making changes in anthropogenic noise (e.g., Buckland 2006) an unlikely confounding factor. Inter-

estingly, the change in methodology between 1994 and 1995 was an a priori design consideration that we had flagged as having a possible effect on trends. Our detection probability analyses largely confirmed our suspicions, but the effect appears to have been of short enough duration relative to the full time series that it did not substantially alter the fitted trends.

Several data considerations from our study also help to explain our result. First, our time series is long enough that short-term patterns of detectability, clearly evident over 3–5 year time scales, appear much more like random fluctuations in detectability over 10–15 year timescales. Second, the large sample sizes (numbers of birds detected, Appendix B) result in relatively small degrees of sampling uncertainty, further reducing the likelihood of spurious trends in abundance, detectability, or both. In general, agreement between trends inferred from counts vs. mixtures increased with the number of detections, and the greatest disagreement among the two methods was observed with Brown Creeper (Fig. 10), one of two smallest samples included

in our study. The latter considerations suggest that Johnson's (2008) second condition [$\text{Var}(p) \ll \text{Var}(N)$] can be meaningfully applied to sampling variance as well as process variance.

Doubts about the efficacy of controlling detection heterogeneity through standardization of protocols have been used as a justification for incorporating detection probability estimation into the analysis of animal monitoring data (MacKenzie et al. 2006, Simons et al. 2007). Our results suggest that standardization can be effective. However, we believe that our results also depend on large sample sizes and that standardization alone would not be sufficient to overcome detection heterogeneity with limited data. Nevertheless, it is worth noting that careful a priori attention to survey protocols to anticipate confounded results constitutes sound experimental design regardless of whether or not detection probabilities will be estimated and/or controlled.

### Literature Cited

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov and F. Csaki, editors. Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Alldredge, M. W., K. H. Pollock, T. R. Simons, J. A. Collazo, and S. A. Shriner. 2007a. Time-of-detection method for estimating abundance from point-count surveys. Auk 124: 653–664.

Alldredge, M. W., T. R. Simons, and K. H. Pollock. 2007b. Factors affecting aural detections of songbirds. Ecological Applications 17:948–955.

Alldredge, M. W., T. R. Simons, K. H. Pollock, and K. Pacifici. 2007c. A field evaluation of the time-of-detection method to estimate population size and density for aural avian point counts. Avian Conservation and Ecology 2:13.

Barker, R. J., and J. R. Sauer. 1995. Statistical aspects of point count sampling. Pages 125–130 in C. J. Ralph, J. Sauer, and S. Droege, editors. Monitoring bird populations by point counts. USDA U.S. Forest Service General Technical Report GTR-PSW 149. Albany, California, USA.

Bart, J., M. Hofschen, and B. G. Peterjohn. 1995. Reliability of the breeding bird survey: effects of restricting surveys to roads. Auk 112:758–761.

Blake, J. G., J. M. Hanowski, G. J. Niemi, and P. T. Collins. 1994. Annual variation in bird populations in mixed conifer northern hardwoods forests. Condor 96:381–399.

Buckland, S. T. 2006. Point-transect surveys for songbirds: robust methodologies. Auk 123:345–357.

Burnham, K. P. 1981. Summarizing remarks: environmental influences. Pages 324–325 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. Studies in avian biology 6. Blackburn Press, Caldwell, New Jersey, USA.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical information-theoretic approach. Second edition. Springer, New York, New York, USA.

Elith, J., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29: 129–151.

Farnsworth, G. L., K. H. Pollock, J. D. Nichols, T. R. Simons, J. E. Hines, and J. R. Sauer. 2002. A removal model for estimating detection probabilities from point-count surveys. Auk 119:414–425.

Hanowski, J. M., and G. J. Niemi. 1995a. A comparison of on-road and off-road bird counts: do you need to go off road to count birds accurately? Journal of Field Ornithology 66:469–483.

Hanowski, J. M., and G. J. Niemi. 1995b. Experimental design considerations for establishing an off-road habitat-specific monitoring program using point-counts. Pages 145–150 in C. J. Ralph, J. Sauer, and S. Droege, editors. USDA U.S. Forest Service General Technical Report GTR-PSW 149. Albany, California, USA.

Hilbe, J. M. 2007. Negative binomial regression. Cambridge University Press, New York, New York, USA.

Hilborn, R., and M. Mangel. 1997. The ecological detective. Princeton University Press, Princeton, New Jersey, USA.

Howe, R. W., G. J. Niemi, S. J. Lewis, and D. A. Welsh. 1997. A standard method for monitoring songbird populations in the Great Lakes region. Passenger Pigeon 59:183–194.

Johnson, D. H. 1995. Point counts of birds: what are we estimating? Pages 117–123 in C. J. Ralph, J. Sauer, and S. Droege, editors. Monitoring bird populations by point counts. USDA U.S. Forest Service General Technical Report GTR-PSW 149. Albany, California, USA.

Johnson, D. H. 2008. In defense of indices: the case of bird surveys. Journal of Wildlife Management 72:857–868.

Johnson, N. L., A. W. Kemp, and S. Kotz. 2005. Univariate discrete distributions. Third edition. Wiley-Interscience, Hoboken, New Jersey, USA.

Kéry, M., A. J. Royle, and H. Schmid. 2005. Modeling avian abundance from replicated counts using binomial mixture models. Ecological Applications 15:1450–1461.

MacKenzie, D. I., J. D. Nichols, J. A. Royle, K. H. Pollock, L. L. Bailey, and J. E. Hines. 2006. Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence. Academic Press, Burlington, Massachusetts, USA.

MathWorks. 2007. MATLAB: The language of technical computing. R2007b. MathWorks, Natick, Massachusetts, USA.

Nichols, J. D., J. E. Hines, J. R. Sauer, F. W. Fallon, J. E. Fallon, and P. J. Heglund. 2000. A double-observer approach for estimating detection probability and abundance from point counts. Auk 117:393–408.

Pacifici, K., T. R. Simons, and K. H. Pollock. 2008. Effects of vegetation and background noise on the detection process in auditory avian point-count surveys. Auk 125:600–607.

Ralph, C. J., S. Droege, and J. R. Sauer. 1995. Managing and monitoring birds using point counts: standards and applications. Pages 161–168 in C. J. Ralph, J. Sauer, and S. Droege, editors. Monitoring bird populations by point counts. USDA U.S. Forest Service General technical Report GTR-PSW 149. Albany, California, USA.

Reynolds, R. T., J. M. Scott, and R. A. Nussbaum. 1980. A variable circular-plot method for estimating bird numbers. Condor 82:309–313.

Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. 2002. Landbird counting techniques: current practices and an alternative. Auk 119:46–53.

Royle, J. A. 2004a. Generalized estimators of avian abundance from count survey data. Animal Biodiversity and Conservation 27:375–386.

Royle, J. A. 2004b. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60:108–115.

Royle, J. A., D. K. Dawson, and S. Bates. 2004. Modeling abundance effects in distance sampling. Ecology 85:1591–1597.

Sauer, J. A., B. G. Peterjohn, and W. A. Link. 1994. Observer differences in the North American breeding bird survey. Auk 111:50–62.

Simons, T. R., M. W. Alldredge, K. H. Pollock, and J. M. Wettroth. 2007. Experimental analysis of the auditory detection process on avian point counts. Auk 124:986–999.

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.

Verner, J. 1985. Assessment of counting techniques. Pages 247–302 in R. F. Johnston, editor. Current ornithology. Volume 2. Plenum, New York, New York, USA.

White, G. C., and R. E. Bennetts. 1996. Analysis of frequency count data using the negative binomial distribution. Ecology 77:2549–2557.

Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations: modeling, estimation, and decision making. Academic Press, New York, New York, USA.

## APPENDIX A

Likelihood functions, negative log-likelihood functions, and gradients of the negative log-likelihood function for all likelihoods (*Ecological Archives* A019-085-A1).

## APPENDIX B

Species list, models, and AIC tables for all sixteen species reporting the minimal (best) set of models accounting for at least 99% of model weight (*Ecological Archives* A019-085-A2).

## APPENDIX C

Plots of zero inflation, overdispersion, and abundance trends from the best model for each species and each likelihood (*Ecological Archives* A019-085-A3).