

# Statistical Report

*Ecology*, 99(2), 2018, pp. 281–288  
© 2017 by the Ecological Society of America

## Identifiability in $N$ -mixture models: a large-scale screening test with bird data

MARC KÉRY\*

*Swiss Ornithological Institute, 6204 Sempach, Switzerland*

**Abstract.** Binomial  $N$ -mixture models have proven very useful in ecology, conservation, and monitoring; they allow estimation and modeling of abundance separately from detection probability using simple counts. Recently, doubts about parameter identifiability have been voiced. I conducted a large-scale screening test with 137 bird data sets from 2,037 sites. I found virtually no identifiability problems for Poisson and zero-inflated Poisson (ZIP) binomial  $N$ -mixture models, but negative-binomial (NB) models had problems in 25% of all data sets. The corresponding multinomial  $N$ -mixture models had no problems. Parameter estimates under Poisson and ZIP binomial and multinomial  $N$ -mixture models were extremely similar. Identifiability problems became a little more frequent with smaller sample sizes (267 and 50 sites), but were unaffected by whether the models did or did not include covariates. Hence, binomial  $N$ -mixture model parameters with Poisson and ZIP mixtures typically appeared identifiable. In contrast, NB mixtures were often unidentifiable, which is worrying since these were often selected by Akaike's information criterion. Identifiability of binomial  $N$ -mixture models should always be checked. If problems are found, simpler models, integrated models that combine different observation models or the use of external information via informative priors or penalized likelihoods, may help.

**Key words:** *binomial  $N$ -mixture model; estimability; hierarchical model; identifiability; infinite abundance estimate; maximum likelihood; multinomial  $N$ -mixture model; nonidentifiable; unmarked zero-inflation.*

### INTRODUCTION

Upon their development a little over a decade ago, binomial  $N$ -mixture models (Royle 2004a) achieved what before was believed impossible: to estimate abundance, corrected for imperfect detection, with data from unmarked individuals. Using only replicated counts and assuming population closure, these models combine a Poisson or similar distribution for spatial variation of latent abundance ( $N$ ) with a binomial distribution, conditional on  $N$ , for the counts. Binomial  $N$ -mixture models (hereafter Binmix models) have since proven extremely useful in ecology, conservation, and monitoring. They can be applied to much “cheaper” (because more easily obtainable) data and typically over much larger spatiotemporal extents than other protocols for abundance estimation, which require extra information that is more difficult to come by, e.g., individual

identification (capture–recapture; Williams et al. 2002) or distance measurements (distance sampling; Buckland et al. 2015).

Use of Binmix models has greatly increased over the years and many model extensions have been developed (Kéry and Royle 2016). These include different observation protocols such as removal sampling (Dorazio et al. 2005), distance sampling (Royle et al. 2004) and capture–recapture (Royle et al. 2007), which all lead to a multinomial  $N$ -mixture model, where a multinomial observation model is mixed over a Poisson or similar distribution. Other important developments have been dynamic models (Dail and Madsen 2011, Hostetler and Chandler 2015, Bellier et al. 2016), spatial models (Royle et al. 2007), and models with zero-inflation (Wenger and Freeman 2008) and correlated detections (Martin et al. 2011). Code for Bayesian analysis is widely available (Kéry 2008, Kéry and Schaub 2012), and Binmix models have been implemented in three of the most widely used computer software for capture–recapture and occupancy data; MARK (White and Burnham 1999), PRESENCE (Hines 2012), and unmarked (Fiske and Chandler 2011), where the user may choose between three abundance

Manuscript received 10 May 2017; revised 25 October 2017; accepted 30 October 2017. Corresponding Editor: Paul B. Conn.

\* E-mail: marc.kery@vogelwarte.ch

mixtures, Poisson (P), zero-inflated Poisson (ZIP) and negative-binomial (NB), and use Akaike's information criterion (AIC) to guide that choice.

Recently, however, two issues with these models have been found. First, doubts have been expressed about parameter identifiability in the Binmix model. Its likelihood contains an infinite sum over the latent abundance states, which in practice is achieved by replacing infinity by some large number  $K$  beyond which parameter estimates do not change appreciably (Royle 2004a). In unmarked,  $K$  is chosen by default as the maximum count plus 100, while PRESENCE sets  $K$  at 200 (J. Hines, *personal communication*). Dennis et al. (2015) show that the Binmix model with Poisson mixture can sometimes yield estimates of abundance and detection probability ( $p$ ) that are infinite and zero, respectively. They found this especially for small abundance, detection probability and number of repeat visits. No value of  $K$  then exists at which estimates become stable; rather, as  $K$  increases so does the abundance estimate. Hence, stability of the estimates for increasing  $K$  can be used as a criterion for parameter identifiability (Dennis et al. 2015, Haines 2016). Barker et al. (2017) show that parameter identifiability in the P-Binmix model is tenuous and that in some cases, the model may collapse to a simpler random-effects Poisson model with a random site-specific intercept, which is the product of the abundance and the detection intercepts in the more general Binmix model. They compare the Binmix model with a classical capture-recapture model fitted to a typical detection-history matrix where there are no identifiability problems. Modeling replicated counts in the Binmix model amounts to fitting a model to an information-reduced *summary* of the original detection history. Barker et al. (2017) claim that the associated loss of information is crucial and causes problems for parameter identifiability in the Binmix model.

A second issue is the recurrent observation that NB mixtures may lead to unrealistically high abundance estimates, even though the NB may be strongly preferred by AIC over P or ZIP mixtures (Kéry et al. 2005, Joseph et al. 2009). This has been called the “good fit/bad prediction dilemma” (Kéry and Royle 2016:264). No solution or even explanation for it has been discovered so far. Given the theoretical and practical importance of the Binmix model, both issues give cause for concern.

I conducted a large screening test of parameter identifiability in the Binmix model with P, ZIP, and NB mixtures, using 137 bird data sets comprising territory detection-history data from 2,037 sites. I fitted “realistically complex” Binmix models with multiple covariates to those from analogous multinomial  $N$ -mixture models (here called Multimix models) for the full detection-history data (Royle 2004b, Dorazio et al. 2005). These are hierarchical extensions of simple capture-recapture models and as such their identifiability should not be under question (Barker et al. 2017). For both models, I

varied  $K$  over a large range. I then repeated this for two much smaller data subsets to identify potential identifiability problems for more real-world sample sizes. I also repeated all of this with models without covariates since covariates are known to help or even permit identifiability in similar models (Dorazio 2014, Matechou et al. 2014).

I use insensitivity to increasing  $K$  (series of  $K$ ) as a criterion for parameter identifiability (Dennis et al. 2015, Haines 2016) and compare identifiability and abundance estimates among the three mixtures (P, ZIP, and NB) for both the Binmix and the Multimix models. Since estimability in multimix models is not under question (Barker et al. 2017), agreement of inferences between the two model classes can be taken as another criterion for identifiability in the Binmix model. For P and NB mixtures in the model without covariates, I also compare the series of  $K$  criterion with the two identifiability diagnostics developed by Dennis et al. (2015).

## MATERIALS AND METHODS

### *A screening test with 137 data sets on Swiss breeding birds*

In the new Swiss breeding bird atlas (Knaus et al. 2018), 2,318 1-km<sup>2</sup> quadrats laid out randomly across the 41,285 km<sup>2</sup> of Switzerland were surveyed by experienced volunteers along a prescribed, irregular survey transect route of typically 4–6 km length. Each transect was surveyed two or three times during one breeding season (15 April–30 June) in the years 2012–2016 using the territory mapping method (Bibby et al. 2000), where all bird detections are recorded on a map and territories are then identified based on behavior, clustering of detections and inter-detection distances. Use of territory mapping yields the equivalent of capture-recapture data for every identified territory (Royle et al. 2007, Kéry and Royle 2010). I analyzed data from the 2,037 quadrats surveyed in 2012–2015, where a total of 167 species were detected.

For each site, I first aggregated the territory-level capture-recapture data by observable capture history, e.g., 111 or 10 denote a territory detected during each of three surveys or during the first and missed during the second in a quadrat surveyed only twice. In the Swiss atlas survey, there are seven and three observable detection histories in quadrats surveyed three and two times, respectively:  $\omega = \{111, 110, 101, 011, 100, 010, 001\}$  and  $\omega = \{11, 10, 01\}$ , where  $\omega$  denotes the vector of observable capture histories. I fit Multimix models directly to site-specific detection history frequencies  $y_{i,\omega}$ , i.e., the frequency of an encounter history of type  $\omega$  observed at site  $i$ . The multinomial vector of frequencies of all possible encounter histories is the vector  $\mathbf{y}_i$  for site  $i$ . For the Binmix models, I further aggregated the detection history frequencies to survey- and site-specific counts  $C_{it}$ , denoting the number of territories detected at site  $i$

during survey *t*. Hence, I used exactly the same information for both models *except* for the individual identity across surveys, which is lost in the second aggregation and therefore is information that cannot be exploited by the Binmix model. See Appendix S1: Table S1 for how one obtains detection history frequencies and replicated counts from territory-level detection histories.

*Binomial and multinomial N-mixture (Binmix and Multimix) models*

For the Binmix model, the input data are aggregated counts  $C_{it}$ : the number of birds counted at site *i* during survey *t*. We can write the model in a hierarchical fashion as a sequence of two linked probability statements:  $N_i \sim \text{Poisson}(\lambda_i)$ , and  $C_{it}|N_i \sim \text{Binomial}(N_i, p_{it})$ . That is, the spatial variation of abundance  $N_i$  is described by a Poisson distribution with mean  $\lambda_i$ , and the temporal variation of the observed counts,  $C_{it}$ , is described by a conditional Binomial distribution with index  $N_i$  and detection probability  $p_{it}$ . Implicit are the two assumptions of population closure, i.e.,  $N_i$  does not vary over surveys, and absence of false-positives, i.e., there are no double counts of individuals. Variation in the intensity and detection probability can be specified by covariates via log and logit link transformations, respectively, as is customary in generalized linear models. Instead of a Poisson (P) mixing distribution, we can specify a zero-inflated Poisson (ZIP; Wenger and Freeman 2008) or any other suitable discrete-valued distribution such as a negative binomial (NB; Royle 2004a). We can fit this hierarchical model directly in Bayesian software such as BUGS (Kéry 2008). In a classical analysis, the marginal likelihood is maximized, where the latent states, or random effects,  $N_i$  are eliminated by summation and the product is over *M* sites, which, for the Poisson mixture, yields

$$L(\mathbf{p}, \boldsymbol{\lambda}; \{C_{it}; i = 1 \dots M, t = 1 \dots T\}) = \prod_{i=1}^M \left\{ \sum_{N_i=\max, C_{it}}^{\infty} \left( \prod_{t=1}^T \text{Binomial}(C_{it}; N_i, p_{it}) \right) \text{Poisson}(N_i; \lambda_i) \right\}$$

where  $\mathbf{p} = \begin{pmatrix} p_{11} & \dots & p_{1T} \\ \vdots & \ddots & \vdots \\ p_{M1} & \dots & p_{MT} \end{pmatrix}$  and  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_M)$ .

In practice, the infinite sum over  $N_i$  is restricted to a finite large bound *K* (Royle 2004a).

The Multimix model also exploits the information about individual identity of a territory contained in the individual capture–recapture data by modeling the vector of detection history frequencies,  $\mathbf{y}_i$ , as a realization from a multinomial distribution:  $\mathbf{y}_i|N_i \sim \text{Multinomial}(N_i, \boldsymbol{\pi}_i)$ . The cell probabilities  $\boldsymbol{\pi}_i$  denote the expected frequencies of each possible detection history and are a deterministic function of a site- and occasion-specific detection probability  $p_{it}$  (Kéry and Royle 2016: Chapter

7). Only the observation model of the Multimix model differs from the Binmix model, while the model for the latent parameters  $N_i$  is typically again chosen to be a discrete-valued distribution such as P, ZIP, or NB. Thus, the marginal likelihood of the model is

$$L(\mathbf{p}, \boldsymbol{\lambda}; \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M\}) = \prod_{i=1}^M \left\{ \sum_{N_i=\max, C_{it}}^{\infty} \left( \prod_{t=1}^T \text{Multinomial}(\mathbf{y}_i; N_i, \boldsymbol{\pi}_i) \right) \text{Poisson}(N_i; \lambda_i) \right\}$$

For a Poisson model, the marginal likelihood has an analytic representation as the product of Poisson random variables (Dorazio et al. 2005) while, for the ZIP and NB models, the likelihood contains again an infinite sum and must be maximized numerically.

*Assessment of parameter identifiability*

I used maximum likelihood in the R package unmarked (Fiske and Chandler 2011) to fit realistically complex binomial and multinomial *N*-mixture models with P, ZIP, and NB abundance mixtures to the data from 2,037 1-km<sup>2</sup> quadrats for a total of 137 species that were detected in  $\geq 10$  quadrats 2012–2015. I fit models with covariates in both the abundance and the detection models, such that  $\log(\lambda_i)$  was specified as a linear function of elevation linear and squared, and forest cover, and  $\text{logit}(p_{it})$  was a linear function of elevation and of survey date linear and squared, and duration. Each model was fitted with 10 different values of *K* (170, 250, 500, 750, 1,000, 1,250, 1,500, 3,000). Median and maximum counts for the detection histories and the aggregated counts, respectively, were 9 and 165 and 8 and 153. Hence, for most species, even the smallest *K* was greater than the unmarked default.

As a first criterion for parameter identifiability, I used stability of numerical solutions when *K* was increased (Haines 2016), specifically, of the AIC, which is minus twice the negative log-likelihood. Second, for P and NB mixtures, I also computed the two diagnostic statistics in Dennis et al. (2015). For the P case, estimates of expected population size become infinite when  $p = 0$ , and as  $p > 0$ , repeated counts at a site become independent. Diagnostic 1 is based on a sample covariance, which is a convenient statistic to check for that. For a finite abundance estimate, it must be positive. The NB case is approached through moment estimation, where two diagnostics can be obtained by constraining the parameter estimates to their correct ranges. The same diagnostic appears as for the Poisson case, as well as a further one, and both must be positive for the NB to yield finite abundance estimates (B. Morgan and E. Dennis, *personal communication*).

To check whether patterns of parameter identifiability were different for more real-world sample sizes, I repeated all analyses for random subsets of 267 and 50 quadrats, where I again restricted analysis to those 107

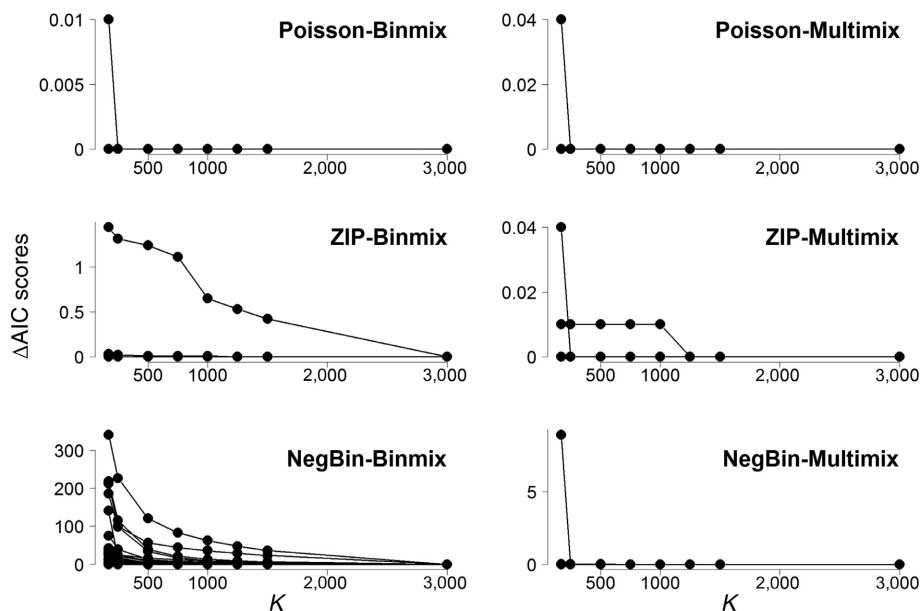


FIG. 1. Plots of the  $\Delta$ AIC scores ( $\Delta$ AIC) (on the y-axis) vs. the value of the summation limit  $K$  in fits of six variants of  $N$ -mixture models with covariates to data from 131 data sets/bird species at 2,037 sites (for a description of these six models see Section “Binomial and multinomial  $N$ -mixture [Binmix and Multimix] models”). Here,  $\Delta$ AIC is the difference in the Akaike’s information criterion (AIC) between one value of  $K$  in a model fit relative to the AIC of the same model with  $K = 3,000$ . Since the number of parameters does not change, it is equivalent to twice the negative log-likelihood difference. A positive value of  $\Delta$ AIC out to  $K = 3,000$  suggests unstable estimates and therefore lack of parameter identifiability. One line is one species.

and 66 species, respectively, with  $\geq 10$  observed occurrences, but only compared five values of  $K$ , {120, 500, 1,000, 2,000, 3,000} for the analysis of 267 quadrats and {105, 500, 1,000, 2,000, 3,000} for that of 50 quadrats. For most species, the smallest  $K$  was again similar to or greater than the unmarked default.

I focused on realistically complex models with multiple covariates in both submodels. Of course, covariates were not always significant, but I was not concerned with this, but simply wanted to work with covariate models, because in practice these are almost always adopted. However, covariates have been found to help or even enable identifiability in similar classes of models (Dorazio 2014, Matechou et al. 2014). Hence, I repeated all of the above with models without covariates, where in addition I also tested a value of  $K$  of 10,000. I used the unmarked function `pcount` for Binmix and a beta version of a revised `gmultmix` function for fitting Multimix models (Kéry and Royle 2016).

## RESULTS

For the full 2,037 sites and Poisson models (Fig. 1 left panel), the AIC of covariate models became stable for most data sets when  $K$  was around the unmarked default (Fig. 1 top). Up to  $K = 3,000$ , solutions stabilized for every single one of the 131 data sets in the screening test for which the covariate models could be fit (Table 1). Some ZIP models required slightly larger than default values of  $K$ ; only for a single data set did solutions remain

unstable up to  $K = 3,000$  (Fig. 1 middle). For the NB-Binmix models, even greater values of  $K$  were required for solutions to stabilize (Fig. 1 bottom) and for 28 data sets, they never stabilized up to  $K = 3,000$ . Hence, P- and ZIP-Binmix models had essentially no problems (0% and 0.1% of data sets), but NB-Binmix models had apparent identifiability problems in 21% of cases. The corresponding Multimix models never had identifiability problems (Fig. 1 right panel). The Dennis et al. (2015) diagnostics indicated no identifiability problems for any data set for the P-Binmix models but did so for 66 (50%) of the data sets for NB-Binmix models (Table 1). While the Dennis and Series of  $K$  criteria agreed broadly in their assessment of the P-Binmix models, the Dennis criterion for the NB-Binmix identified more than twice the number of problems compared with the Series of  $K$  criterion.

Akaike’s information criterion often greatly preferred the NB-Binmix models and these models typically yielded very much higher abundance estimates than the ZIP or P mixtures (Fig. 2). Importantly, both identifiability criteria suggested that most of these NB-Binmix models did *not* have finite abundance estimates. Most such unrealistic estimates were thus based on an unidentified model. Estimates of mean detection probability under the P-Binmix model with  $K = 3,000$  averaged 0.44 per species (range 0.06–0.83). Interestingly, species with infinite abundance estimates under the NB-Binmix models had higher abundance and detection probability than species with finite abundance estimates (Appendix S1: Fig. S1); thus, it was not the rare and elusive species that

TABLE 1. Screening test for identifiability problems in *N*-mixture models with covariates fit to data from the new Swiss breeding bird atlas (Knaus et al. 2018): comparison among two *N*-mixture models (binomial-Binmix and multinomial-Multimix), three mixture distributions (Poisson [P], zero-inflated Poisson [ZIP], negative binomial [NB]), three sample sizes (2,037 sites/131 species, 267 sites/107 species, 50 sites/66 species), and two identifiability criteria (series of *K* criterion and diagnostics of Dennis et al. 2015).

Model (mixture)	Unstable estimates (%)					
	2,037 sites		267 sites		50 sites	
	Series of <i>K</i>	Dennis	Series of <i>K</i>	Dennis	Series of <i>K</i>	Dennis
P-Binmix	0	0	2	0	5	0
ZIP-Binmix	0.1	–	5	–	9	–
NB-Binmix	21	50	25	52	26	62
P-Multimix	0	–	0	–	0	–
ZIP-Multimix	0	–	1	–	0	–
NB-Multimix	0	–	0	–	0	–

Note: Table entries are the percentage of species with unstable estimates, i.e., with apparent lack of identifiability.

created estimation problems under the NB-Binmix model. Abundance estimates under the Binmix and Multimix models were in almost perfect agreement on average for the Poisson and the ZIP mixtures (Fig. 3 top and middle), but less so for NB mixtures (Fig. 3 bottom).

Results for the models with covariates applied to two smaller sample sizes (267 and 50 sites) are summarized in Table 1 and in Appendices S2, S3. They yielded identical patterns but suggested more identifiability problems (Table 1) and also somewhat less agreement between the Binmix and Multimix model-based estimates than results for 2,037 sites. Importantly, results for the same analyses for the intercepts-only models, i.e., without covariates, yielded hardly more cases of parameter unidentifiability than for the models with covariates (compare Table 1 and Appendix S1: Fig. S2).

DISCUSSION

Binomial *N*-mixture (Binmix) models (Royle 2004a) have become widely used because they allow separate estimation and modeling of abundance and detection from replicated counts of unmarked individuals. However, two issues are, first, that they may yield infinite abundance estimates, rendering the model unidentifiable (Dennis et al. 2015, Haines 2016) and second, that Binmix models with negative binomial are often selected over other mixtures by AIC, but then they produce unrealistically high abundance estimates (Kéry et al. 2005, Joseph et al. 2009, Kéry and Royle 2016). Currently, it is not known how widespread identifiability problems with Binmix models are, nor how the “good fit/bad prediction dilemma” (Kéry and Royle 2016:264) can be explained.

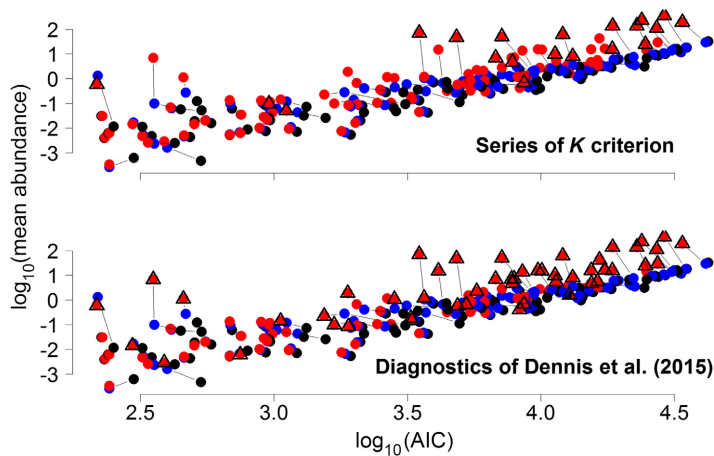


FIG. 2. Plots of abundance estimates vs. Akaike’s information criterion (both on log<sub>10</sub> scale, for *K* = 3,000) for two criteria of parameter identifiability in binomial *N*-mixture models: series of *K* (top) and the two diagnostics of Dennis et al. (2015, bottom). Both panels show results for all three mixture distributions (Poisson, black; zero-inflated Poisson, blue; and negative binomial, red) for 131 Swiss breeding bird species; values for same species are linked by lines. Triangles mark models for which the “series of *K*” approach and the Dennis et al. diagnostics indicate infinite abundance estimates.

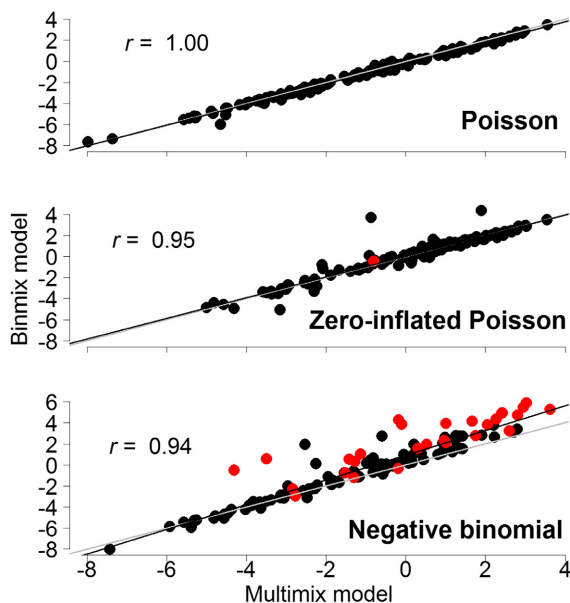


FIG. 3. Comparison of the abundance intercepts between multinomial and binomial  $N$ -mixture models for all three mixture distributions and 131 test species in 2,037 sites (for  $K = 3,000$ ). Black line is regression line of the value on the  $y$ -axis on the value on the  $x$ -axis and gray line is the 1:1 line. Pearson correlation coefficient is given inset. Estimates from non-identifiable Binmix models are shown in red.

I conducted a large-scale screening test of the identifiability of Binmix models in practice, using 137 bird data sets from 2,037 sites (Knaus et al. 2018), for all three commonly used abundance mixture distributions (Poisson, zero-inflated Poisson, and negative-binomial), and for subsets of the data containing 267 and 50 sites. I assessed the identifiability of the Binmix model using two formal criteria (Dennis et al. 2015, Haines 2016) and also informally by comparing the estimates of the Binmix model with those of the related multinomial  $N$ -mixture (Multimix) model (Royle 2004b), a hierarchical variant of a capture–recapture model, and for the latter, identifiability is not under debate (Barker et al. 2017). Finally, I compared “realistic models,” with several covariates in each of the abundance and detection parts of the model, with the analogous intercepts-only “toy models” without any covariates; these latter are the ones dealt with by Dennis et al. (2015) and Barker et al. (2017), but they are rarely adopted in practice.

My results were clear. Regardless of whether the Binmix models did or did not contain covariates, I found virtually no identifiability problems with P or ZIP mixtures, but for NB mixtures there were problems in about one-quarter of all tested data sets. With one exception, no identifiability problems were found in Multimix models. In addition, estimates from models with P and ZIP mixtures, and with identifiable NB mixtures, were almost identical on average between Binmix and Multimix models. These findings all suggest that the P- and ZIP-

Binmix models were overwhelmingly identifiable in my screening test. Next, my results also explain the “good fit/bad prediction dilemma”: typically, unidentifiable NB-Binmix models had the lowest AIC of all three mixtures and yet yielded far higher abundance estimates than the P or ZIP mixtures. This suggests that unrealistically high abundance estimates under NB-Binmix models reported previously (Kéry et al. 2005, Joseph et al. 2009) may simply have been based on unidentifiable models and should be disregarded; it is only the choice of a finite  $K$  that keeps them from becoming infinity (Dennis et al. 2015). As expected, more identifiability problems and less agreement between Binmix and Multimix models were found for smaller sample sizes.

I believe that these are important results, because they now let us apply binomial  $N$ -mixture models with more confidence than before. However, they also suggest that identifiability of Binmix models should always be checked; both the series of  $K$  and the Dennis et al. (2015) diagnostics are useful for this. One of the advantages of the former is that it can be applied to any mixture, including the ZIP, and especially to models with covariates. I conducted my screening test with maximum likelihood, but it is worth emphasizing that in a Bayesian analysis, things normally would not be practically different if the usual vague priors are adopted (but see Barry et al. 2003). Non-identifiability of parameters is more difficult to diagnose in a Bayesian analysis (but see Gimenez et al. 2009), so a series of ML fits of a Binmix model may be insightful about identifiability of a model even when a Bayesian analysis of a Binmix model is carried out.

What should be done in the future? Perhaps the most important thing is that users must become aware of these issues and know how to deal with them in an informed way when they use Binmix models. Checking identifiability ought to become part of the routine workflow when fitting Binmix models. Since the “good fit/bad prediction dilemma” now appears to be solved, we need no longer shy away from using NB-Binmix models. Rather, we know now when we can and cannot use them and hence in the former case use their inference with confidence.

Dennis et al. (2015) criticize the use of a constant value of  $K$ , which may hide identifiability problems by artificially truncating the solutions for abundance. Instead, they use an adaptive approach that which will automatically use larger values of  $K$  if needed. Program MARK uses such an adaptive approach for Binmix models (G. White, *personal communication*), and this will help identification of cases with infinite abundance estimates. Dennis et al. (2015) and Haines (2016) have developed equivalent likelihoods for the P- and NB-Binmix model that have closed solutions. This is interesting, because in these cases infinite abundance estimates could be found directly, without having to rerun the model for multiple  $K$  values or having to rely on a diagnostic. But, so far, the closed form representations of the likelihood have not been implemented in the common software.

The Multimix models never had any problem (with a single exception), regardless of the mixture, which is a definite advantage of that model and one that we did not realize in an earlier comparison of the two models (Kéry and Royle 2010). Thus, when data are available such that a Multimix rather than a Binnmix model may be adopted, we may well want to do this even if only to obtain increased confidence in the Binnmix solutions. On the other hand, we note that the Binnmix is robust against splitting and lumping errors in territory delimitation that leads to the detection histories in territory mapping data such as ours while the Multimix is not (Kéry and Royle 2010).

For statisticians, more research into the issue of parameter identifiability of the Binnmix model, and indeed of many classes of hierarchical models, would be valuable from a practical point of view. If problems with identifiability of a model are found, simply concluding that a certain type of model should be abandoned hardly appears like very useful advice. Rather, a more valuable approach would be to identify ways in which problems may be mitigated. For practitioners, the question remains of what to do when no identifiable model can be found for a given data set. Clearly, recourse to a simpler model such as a random-effects Poisson model (Barker et al. 2017) may or may not be a solution, depending on the objectives of an analysis: if abundance estimates are needed or if the same covariates need to be modelled in both the abundance and the detection parts of the model, then such a model will not be satisfactory. Alternative solutions may be integrated models, e.g., if territory detection-nondetection data are available at a subset of sites then their analysis could be integrated in a single, combined Binnmix and Multimix model and the extra information in the Multimix part may mitigate identifiability problems in the Binnmix part of the model. The opportunities for such combination approaches are vast and may include the addition to simple replicated counts of distance sampling as well. Similarly, external (prior) information may be added in a Bayesian analysis via informative priors or in a likelihood analysis by penalties (Lele et al. 2012). These ideas also suggest a large number of possible investigations about design, which may revolve around how to optimally combine different types of data, e.g., which data types and in what proportions. Thus, hierarchical modeling of abundance from unmarked individuals using  $N$ -mixture models will likely remain a rich ground for both theoretical and applied investigations. This is especially so because of the evident utility of binomial  $N$ -mixture models in many fields including ecology, conservation, and monitoring.

#### ACKNOWLEDGMENTS

For comments I thank E. Dennis, B. Morgan, J. Knappe, F. Plard, M. Schaub, and N. Strebel. I also thank all the volunteers who conducted field work that produced the high-quality data analysed in this paper. Financial support of this study was provided by a grant from the Swiss National Science Foundation (No. 31003A\_1464125 to M. Kéry and M. Schaub).

#### LITERATURE CITED

- Barker, R., M. Schofield, W. Link, and J. Sauer. 2017. On the reliability of  $N$ -mixture models for count data. *Biometrics*. <https://doi.org/10.1111/biom.12734>
- Barry, S. C., S. P. Brooks, E. A. Catchpole, and B. J. T. Morgan. 2003. The analysis of ring-recovery data using random effects. *Biometrics* 59:54–65.
- Bellier, E., M. Kéry, and M. Schaub. 2016. Simulation-based assessment of dynamic  $N$ -mixture models with density-dependence and environmental stochasticity in the vital rates. *Methods in Ecology and Evolution* 7:1029–1040.
- Bibby, C. J., N. D. Burgess, D. A. Hill, and S. Mustoe. 2000. *Bird census techniques*. Academic Press, London, UK.
- Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven. 2015. *Distance sampling: methods and applications*. Springer, Cham, Switzerland.
- Dail, D., and L. Madsen. 2011. Models for estimating abundance from repeated counts of an open population. *Biometrics* 67:577–587.
- Dennis, E. B., B. J. T. Morgan, and M. S. Ridout. 2015. Computational aspects of  $N$ -mixture models. *Biometrics* 71:237–246.
- Dorazio, R. M. 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. *Global Ecology and Biogeography* 23:1472–1484.
- Dorazio, R. M., H. L. Jelks, and F. Jordan. 2005. Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics* 61:1093–1101.
- Fiske, I., and R. Chandler. 2011. unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. *Journal of Statistical Software* 43:1–23.
- Gimenez, O., B. J. T. Morgan, and S. P. Brooks. 2009. Weak identifiability in models for mark-recapture-recovery data. Pages 1055–1067 in D. L. Thomson, E. G. Cooch, and M. J. Conroy, editors. *Modeling demographic processes in marked populations*. Series: Environmental and Ecological Statistics, Volume 3. Springer, Berlin, Germany.
- Haines, L. M. 2016. Maximum likelihood estimation for  $N$ -mixture models. *Biometrics* 72:1235–1245.
- Hines, J. E. 2012. PRESENCE4—software to estimate patch occupancy and related parameters. USGS-PWRC. <http://www.mbr-pwrc.usgs.gov/software/presence.html>.
- Hostetler, J. A., and R. B. Chandler. 2015. Improved state-space models for inference about spatial and temporal variation in abundance from count data. *Ecology* 96:1713–1723.
- Joseph, L. N., C. Elkin, T. G. Martin, and H. Possingham. 2009. Modeling abundance using  $N$ -mixture models: the importance of considering ecological mechanisms. *Ecological Applications* 19:631–642.
- Kéry, M. 2008. Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships. *Auk* 125:336–345.
- Kéry, M., and J. A. Royle. 2010. Hierarchical modeling and estimation of abundance in metapopulation designs. *Journal of Animal Ecology* 79:453–461.
- Kéry, M., and J. A. Royle. 2016. *Applied hierarchical modeling in ecology—modeling distribution, abundance and species richness using R and BUGS*. Volume 1: Prelude and static models. Elsevier/Academic Press, Amsterdam, The Netherlands.
- Kéry, M., J. A. Royle, and H. Schmid. 2005. Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications* 15:1450–1461.
- Kéry, M., and M. Schaub. 2012. Bayesian population analysis using WinBUGS—a hierarchical perspective. Academic Press, Waltham, Massachusetts, USA.

- Knaus, P., S. Antoniazza, J. Guélat, M. Kéry, N. Strebel, and S. Wechsler. 2018. Schweizer brutvogelatlas 2013–2016. Schweizerische Vogelwarte, Sempach, Switzerland.
- Lele, S. R., M. Moreno, and E. Bayne. 2012. Dealing with detection error in site occupancy surveys: What can we do with a single survey? *Journal of Plant Ecology* 5:22–31.
- Martin, J. E., J. A. Royle, B. Gardner, D. I. MacKenzie, H. H. Edwards, and M. Kéry. 2011. Accounting for non-independent detection when estimating abundance of organisms with a Bayesian approach. *Methods in Ecology and Evolution* 2:595–601.
- Matechou, E., E. B. Dennis, S. N. Freeman, and T. Brereton. 2014. Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model. *Journal of Applied Ecology* 51:766–775.
- Royle, J. A. 2004a. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* 60:108–115.
- Royle, J. A. 2004b. Generalized estimators of avian abundance from count survey data. *Animal Biodiversity and Conservation* 27:375–386.
- Royle, J. A., D. K. Dawson, and S. Bates. 2004. Modeling abundance effects in distance sampling. *Ecology* 85:1591–1597.
- Royle, J. A., M. Kéry, R. Gautier, and H. Schmid. 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs* 77:465–481.
- Wenger, S. J., and M. C. Freeman. 2008. Estimating species occurrence, abundance, and detection probability using zero-inflated distributions. *Ecology* 89:2953–2959.
- White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. *Bird Study* 46:120–139.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1002/ecy.2093/supinfo>