

# Field evaluation of abundance estimates under binomial and multinomial *N*-mixture models

YVES BÖTSCH\*  LUKAS JENNI & MARC KÉRY

Swiss Ornithological Institute, Seerose 1, CH-6204, Sempach, Switzerland

Assessing and modelling abundance from animal count data is a very common task in ecology and management. Detection is arguably never perfect, but modern hierarchical models can incorporate detection probability and yield abundance estimates that are corrected for imperfect detection. Two variants of these models rely on counts of unmarked individuals, or territories (binomial *N*-mixture models, or *binmix*), and on detection histories based on territory-mapping data (multinomial *N*-mixture models or *multimix*). However, calibration studies which evaluate these two *N*-mixture model approaches are needed. We analysed conventional territory-mapping data (three surveys in 2014 and four in 2015) using both *binmix* and *multimix* models to estimate abundance for two common avian cavity-nesting forest species (Great Tit *Parus major* and Eurasian Blue Tit *Cyanistes caeruleus*). In the same study area, we used two benchmarks: occupancy data from a dense nestbox scheme and total number of detected territories. To investigate variance in estimates due to the territory assignment, three independent ornithologists conducted territory assignments. Nestbox occupancy yields a minimum number of territories, as some natural cavities may have been used, and *binmix* model estimates were generally higher than this benchmark. Estimates using the *multimix* model were slightly more precise than *binmix* model estimates. Depending on the person assigning the territories, the *multimix* model estimates became quite different, either overestimating or underestimating the ‘truth’. We conclude that *N*-mixture models estimated abundance reliably, even for our very small sample sizes. Territory-mapping counts depended on territory assignment and this carried over to estimates under the *multimix* model. This limitation has to be taken into account when abundance estimates are compared between sites or years. Whenever possible, accounting for such hidden heterogeneity in the raw data of bird surveys, via including a ‘territory editor’ factor, is recommended. Distributing the surveys randomly (in time and space) to editors may also alleviate this problem.

**Keywords:** abundance estimation, cavity nesters, detection probability, forest birds, territory assignment, *unmarked*.

Describing spatiotemporal patterns in abundance and investigating reasons for these patterns is a cornerstone of ecology, including many of its applications in conservation and management. However, assessments of abundance in the field are almost always affected by false-negative and false-positive measurement errors (Williams *et al.* 2002, Royle & Dorazio 2008, Kéry & Schaub

2012, Kéry & Royle 2016). False-negatives are particularly common and correspond to the case where detection probability (*P*) of an individual is less than one. This means that abundance will be underestimated when inference is based solely on counts without any correction for *P*. In addition, if there are patterns in *P*, for example over time or among different habitats, existing patterns in abundance may be masked or non-existent patterns may spuriously appear in analyses of relative abundance (Barker *et al.* 2018). Thus, in many applications it appears prudent to try to correct for

\*Corresponding author.

Email: yves.boetsch@vogelwarte.ch

Twitter: @YBoetsch

imperfect detection when making inferences about abundance and its patterns in space and time.

Several statistical models and associated sampling protocols have been developed to correct abundance estimates for imperfect detection. They all require the collection of extra information; for example, distance measurements for distance sampling (Buckland *et al.* 2001, 2015) or individual identity during repeated observations for capture–recapture methods when, typically, individuals need to be captured and marked with tags (Williams *et al.* 2002). This may be costly and cumbersome and, for capture–recapture, may harm the animals during capture and handling. Hence, methods that do not require capture and tagging or the error-prone measurement of distance have been developed, which enable estimation of abundance and detection probability from observations of individuals that are not physically captured and marked or not individually identified at all. These include double-observer methods (Nichols *et al.* 2000) and (mental) removal sampling (Farnsworth *et al.* 2002, 2005), which have been applied to monitoring data of breeding birds. In essence, in these applications of capture–recapture modelling, the location of an individual is taken as its mark.

For spatially replicated breeding bird surveys, Royle *et al.* (2007) (see also Webster *et al.* 2008) have also developed a hierarchical variant of a capture–recapture model that directly exploits the individual ‘capture-histories’ of all detected territories that are individually identified as part of the commonly used ornithological survey method called territory, or spot, mapping (Bibby *et al.* 2000). These multinomial *N*-mixture models (Royle 2004a, Dorazio & Royle 2005), henceforward called multimix models, allow estimation of abundance at each surveyed site and introduction of covariates to explain variation in abundance as well as detection probability (Kéry & Royle 2016), but they require the actual survey results to be aggregated into putative territories. In addition, they are clearly not suitable for species that do not have well-defined territories.

Finally, for spatially and temporally replicated counts of unmarked individuals, Royle (2004b) developed the binomial *N*-mixture (henceforward binmix) models, which do not require any decision to be made about aggregating the raw survey data into territories. Thus, it is likely that they provide less precise abundance estimates than multimix models, but they require ‘cheaper’ data only and

have been found to be insensitive to errors in the territory assignment (Kéry & Royle 2010). Recently, however, the practical utility of binmix models has been questioned because, with small sample sizes, boundary estimates are not rare (i.e. estimates of a probability that are at 0 or 1 or of abundance that are 0 or infinity) (Dennis *et al.* 2015) and because abundance estimates may be over-sensitive to even small violations of the parametric assumptions (Barker *et al.* 2018, Duarte *et al.* 2018, Knappe *et al.* 2018, Link *et al.* 2018).

In the above context, calibration studies are urgently needed in which the quality of abundance estimates under both *N*-mixture models can be evaluated, in other words where estimates of both model types can be compared with a known truth or where the perhaps ‘riskier’ binmix estimates can be compared with those of the more ‘robust’ multimix estimates. Although this can be done using simulated data, simulations can never encompass the full range of complications that may and will arise in nature. Unfortunately, though, the truth is only extremely rarely known in the field.

Here we exploit a rare situation where we can assume that the truth is at least approximately known in the field. Over 2 years we conducted territory-mapping surveys for two common cavity-nesting passerines (Great Tit *Parus major* and Eurasian Blue Tit *Cyanistes caeruleus*) in 24 forest plots, in which we also conducted an intensive nestbox study. A few tit pairs may have nested in natural cavities, but we still assume that true abundance is approximately known or at least that we have a good minimal estimate of population size in each plot. Hence, we first formatted our territory-mapping survey data into territory detection histories (for analysis by the multimix model) and, secondly, aggregated these to replicated counts without any individual ID (for analysis by the binmix model). We then compared the estimates from both models with each other as well as with the primary benchmark provided by our nestbox study and the secondary benchmark provided by the total number of assigned territories. In addition, to investigate the variance introduced into the estimates by the territory assignment from the raw survey data, we let three different ornithologists conduct the territory assignment independently and then compared the estimates from all three variants of the resulting territory detection histories.

## METHODS

### Study site, breeding bird and nestbox surveys

The study took place in the Forêt domaniale de Chaux, a large (200 km<sup>2</sup>) oak–hornbeam (*Quercus robur* and *Carpinus betulus*) forest near Dole, France (47°05'N, 05°40'E), in 2014 and 2015. Within the forest, we selected 12 plots averaging 9.3 ha (range 7.5–13 ha), where no timber harvesting occurred during the study period. Between April and early June, two observers conducted bird surveys in all plots using the territory-mapping method of Bibby *et al.* (2000), with three survey visits in 2014 and four in 2015. Each observer surveyed six sites, and the sites were surveyed by the same observer in both years. All territorial (e.g. singing) birds were recorded on a map; pairs or families with fledged young were noted as one contact. Plots were surveyed using lawnmower-pattern transects (60-m distance between parallel passages), going back and forth, to cover the entire area thoroughly (for details see Böttsch *et al.* 2017). As the 12 plots served as an experimental study, where half of a plot (split-plot) received a treatment and the other half served as a control (Böttsch *et al.* 2017), we here treated each split-plot as an independent 'site' in our analyses ( $n = 24$ ). We acknowledge that in doing so we commit a minor pseudo-replication, because our software (see below) could not accommodate random block effects, and we could therefore not account for some slight non-independence between the 24 plots. However, we were not interested in testing hypotheses, and so any underestimation of the standard errors was deemed fairly inconsequential in our study. In addition, goodness of fit results (see below) indicated that any non-independence may not have been marked.

In early February 2014, 210 nestboxes (at an average density of two boxes/ha, not exceeding natural breeding densities (Krebs 1971)) for cavity nesters (Schwegler, Type 1B, 32-mm entrance hole diameter) were mounted, removed in autumn and remounted again in February 2015. The first nestbox checks were conducted in late April; from then on, all nestboxes were checked every second week to assess occupation. A nestbox was considered to be occupied if at least one egg had been laid. The nesting bird species (Great Tit, and Eurasian Blue Tit, hereafter 'Blue Tit') were identified

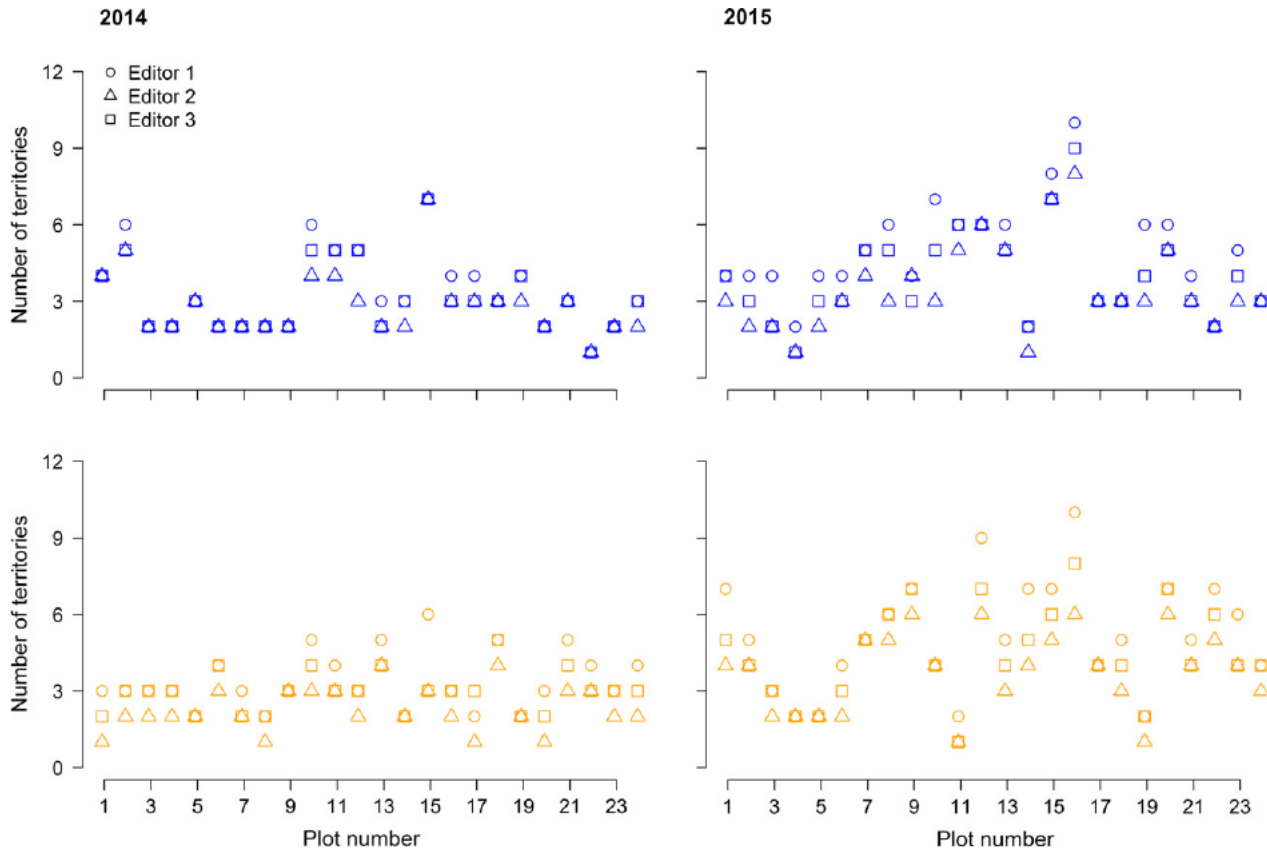
either by examination of the nestlings or by observation of the parents entering the nestboxes. We therefore evaluated the performance of multimix and binmix models for two species and 2 years.

### Data preparation and statistical analysis

For the multinomial  $N$ -mixture models (Royle 2004a, Dorazio *et al.* 2005), territory detection history frequencies are needed, and therefore the raw mapping data need to be aggregated to territories, exactly as is customarily done in territory mapping. We therefore we copied all contacts of Great and Blue Tits from the multiple surveys of a given year on to one map, using different colours for each survey. Later, three independent 'editors' assigned the raw data to territories, mostly based on clustering of contacts and their distance from one another. After that, Y.B. extracted the territory detection histories. Territory detection histories are  $n \times t$  tables for a total of  $n$  territories detected over  $t$  surveys for each species, year and plot. Cell  $(i, j)$  contains a 1 when territory  $i$  was detected on occasion  $j$  and a 0 otherwise.

The replicated count data (number of contacts per survey occasion and species) required as the input for a binomial  $N$ -mixture model can easily be condensed from these territory detection histories (Royle 2004b). As binmix models do not rely on territory assignment, they were run only once a year and species, and not separately for each 'territory editor'.

We used the R package *unmarked* (Fiske & Chandler 2011) in R version 3.3.0 (R Core Team 2016) to fit multimix (function *gmmultmix*) and binmix (function *pcount*) models to the territory-specific detection histories or to the replicated counts of unmarked individuals, respectively. All models estimating abundance were run with three different abundance mixture distributions (Poisson, negative binomial and zero-inflated Poisson). As the Akaike information criterion (AIC; Akaike 1987) always favoured the Poisson mixtures, we present results from these. Dennis *et al.* (2015) have shown that binmix estimates can be sensitive to the summation limit ( $K$ ) in the likelihood evaluation of the binmix model in *unmarked*. Specifically, in some cases, the maximum likelihood estimates (MLEs) are on the boundary of parameter space, i.e. infinite abundance and zero detection probability. So, in addition to the default



**Figure 1.** An illustration of the editor effect in territory-mapping data. Number of assigned territories per plot for the two study years and species (Blue Tits in blue (above), Great Tits in orange (below)), depending on the editor. Breeding density of both species was higher in 2015 than in 2014.

choice of  $K$  in *unmarked*, we re-ran all binmix models with several higher values of  $K$  to see whether the estimates were stable and to diagnose any cases of infinite abundance MLEs.

We ran separate models for each species and each year and accounted for slight variation in plot area via an offset of  $\log(\text{area})$  in the model for expected abundance ( $\lambda$ ). For the detection model ( $P$ ), we accounted for the daily and seasonal phenology of song activity by including linear and quadratic effects of both time of the day ( $\text{time} + \text{time}^2$ ) and Julian date ( $jdate + jdate^2$ ). As two observers collected the survey data, we accounted for this potential confounding factor by including 'observer' as a two-level fixed factor in all detection models. The number of detected birds is likely to depend on the duration spent within a given plot. As the time spent in a given plot might depend on its accessibility (e.g. dense or sparse understorey vegetation), we accounted for this heterogeneity by incorporating a measure

of survey intensity. The 'intensity' was expressed as the time spent during a survey round divided by the plot surface area. All continuous variables were standardized to have a mean of 0 and an sd of 1 to enhance model convergence.

To quantify the 'editor effect' (i.e. the variation in results due to the identity of the person aggregating the raw detections on the maps into territories), we ran an analysis of variance (ANOVA) model on the number of territories determined and assessed the proportion of the total sums of squares explained by year, species, plot and the editor. Additionally we computed coefficients of variation separately for the two years and the two study species.

## RESULTS

In 2014, 151 of the available 210 nestboxes were occupied by one of the two tit species (76 Great Tits and 75 Blue Tits). In 2015, the dominant

**Table 1.** Mean *N*-mixture model estimates  $\pm$  se for the full model (abundance and detection probability, separately for each editor in the case of the multimix models (multimix 1 to 3)).

	2014		2015	
	Blue Tit	Great Tit	Blue Tit	Great Tit
Mean number of occupied nestboxes per ha	0.68 $\pm$ 0.37	0.69 $\pm$ 0.35	0.33 $\pm$ 0.20	1.26 $\pm$ 0.31
Mean number of assigned territories per ha (multimix-specific)	0.67 $\pm$ 0.29	0.62 $\pm$ 0.19	0.89 $\pm$ 0.38	0.97 $\pm$ 0.37
Mean number of contacts per ha (binmix-specific)	0.53 $\pm$ 0.22	0.51 $\pm$ 0.18	0.67 $\pm$ 0.28	0.76 $\pm$ 0.31
Abundance ( $\lambda$ )				
binmix	1.22 $\pm$ 0.75		1.33 $\pm$ 0.47	1.78 $\pm$ 0.73
multimix 1	1.27 $\pm$ 0.22	1.32 $\pm$ 0.22	1.57 $\pm$ 0.19	1.43 $\pm$ 0.14
multimix 2	0.75 $\pm$ 0.10	0.53 $\pm$ 0.07	0.86 $\pm$ 0.10	0.85 $\pm$ 0.09
multimix 3	1.00 $\pm$ 0.15	0.79 $\pm$ 0.11	1.05 $\pm$ 0.12	1.07 $\pm$ 0.11
Detection ( <i>P</i> )				
binmix	0.16 $\pm$ 0.14		0.28 $\pm$ 0.10	0.27 $\pm$ 0.11
multimix 1	0.22 $\pm$ 0.04	0.23 $\pm$ 0.04	0.23 $\pm$ 0.03	0.31 $\pm$ 0.03
multimix 2	0.38 $\pm$ 0.05	0.62 $\pm$ 0.05	0.42 $\pm$ 0.03	0.54 $\pm$ 0.03
multimix 3	0.28 $\pm$ 0.04	0.39 $\pm$ 0.05	0.34 $\pm$ 0.03	0.42 $\pm$ 0.03

The binmix model for Great Tits in 2014 was not identifiable and therefore estimates are not shown. Abundance estimates are the mean number of territories per hectare across all plots. For comparison we also computed the mean number of occupied nestboxes, the mean number of assigned territories and the mean number of contacts per hectare, respectively (first three rows).

species was the Great Tit (see Löhrl 1977), which occupied 139 nestboxes, whereas only 37 nestboxes were occupied by Blue Tits (in total 176 occupied nestboxes).

The territory assignment, necessary for the multimix models, showed some between-editor variation (Fig. 1), with more variation between editors in territory assignment at higher population size (2014 had fewer territories than 2015). This editor variation was also mirrored in the coefficients of variation for each species and year where high values of up to 22% were reached (Table S1).

One of the four binmix models (Great Tits in 2014) showed sensitivity of the estimates to the choice of *K* in the likelihood evaluation in *unmarked* (see Fig. S2). Hence, for practical purposes its parameters were not estimable in the sense of being on the boundary of parameter space, infinity for abundance and zero for detection. For the other binmix models we detected no such problems. Parametric bootstrap goodness-of-fit (GOF) tests based on a Freeman–Tukey discrepancy measure indicated that the Poisson mixture models all fitted the data adequately, whereas the unidentified binmix model yielded a GOF *P*-value of almost 1 (Table S2).

The models revealed that detection probabilities were quite variable between models, ranging from 16% to 62% (Table 1), whereas abundance estimates for both species were very similar between models (except for multimix 2), with a slight increase in abundance estimates from 2014 to 2015 (Table 1).

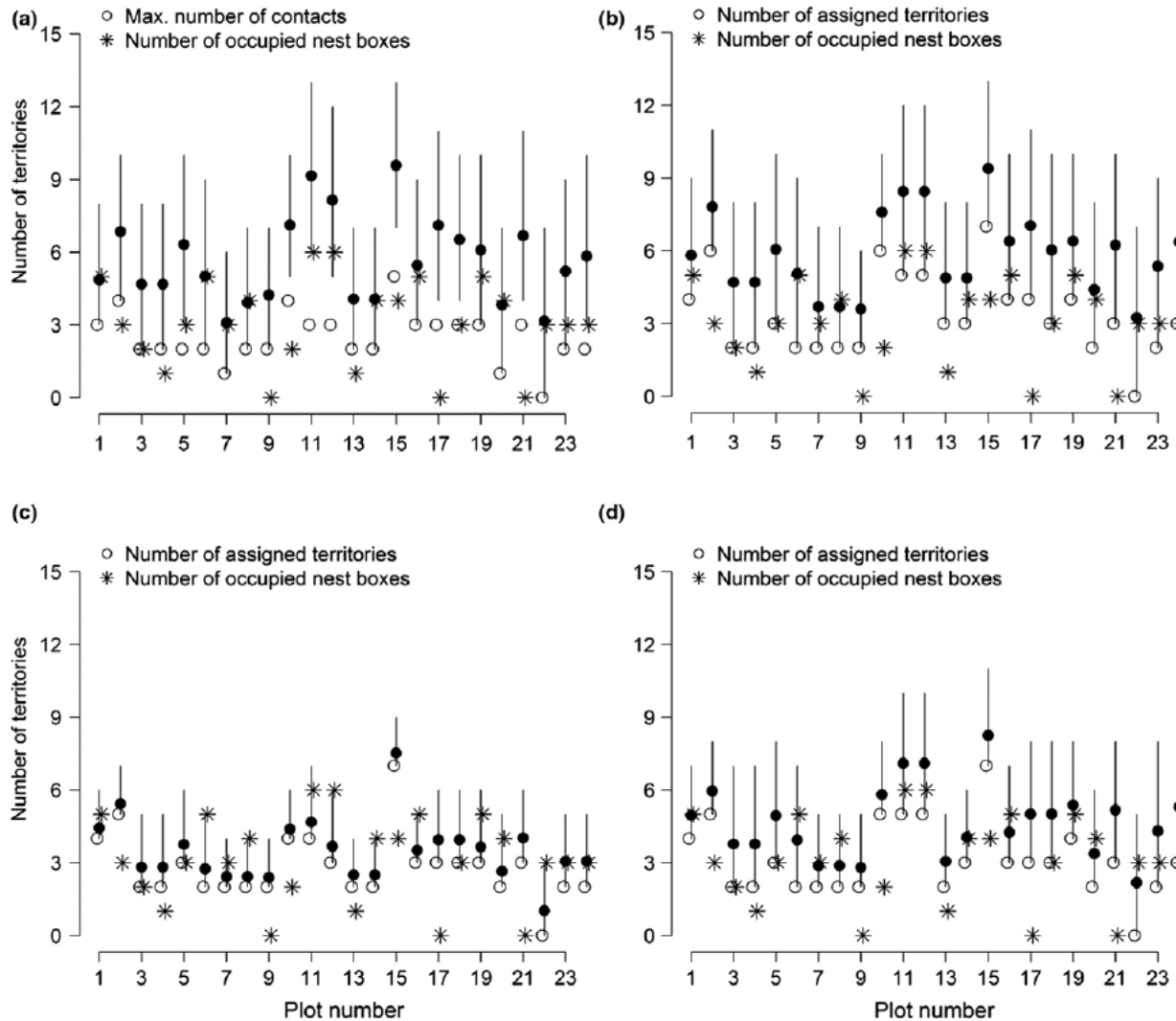
The estimates of *N*-mixture models were consistent in their accuracy across sites in comparison with nestbox occupancy (Figs 2 & 3, Figs S1 & S2). The multinomial *N*-mixture model estimates depended heavily on the editor assigning the territories, as shown by the variation between estimates for the same plots (for the same year, same species and same plot, Figs 2b–d & 3b–d, Figs S1 & S2). The ANOVA sums-of-squares decomposition indicated that editor identity explained up to 8% of the variation in the number of assigned territories, much more than the species or the plot (Table S3).

## DISCUSSION

Overall, the *N*-mixture models predicted abundance fairly well, even for such low sample sizes as ours. We used nestbox occupancy as a benchmark for assessing model prediction accuracy, although there may be several weaknesses to this approach, including birds breeding in natural cavities, counting of edge territories, predation (Nilsson 1984) and departures from monogamy.

From our data we have some evidence for such confounding factors. Depending on the plot, some abundance estimates were substantially higher or lower in comparison with the occupied number of nestboxes (e.g. Fig. 2b, plot 5 or 14). So, either males were present and singing but not related to an occupied nestbox or we failed to record birds



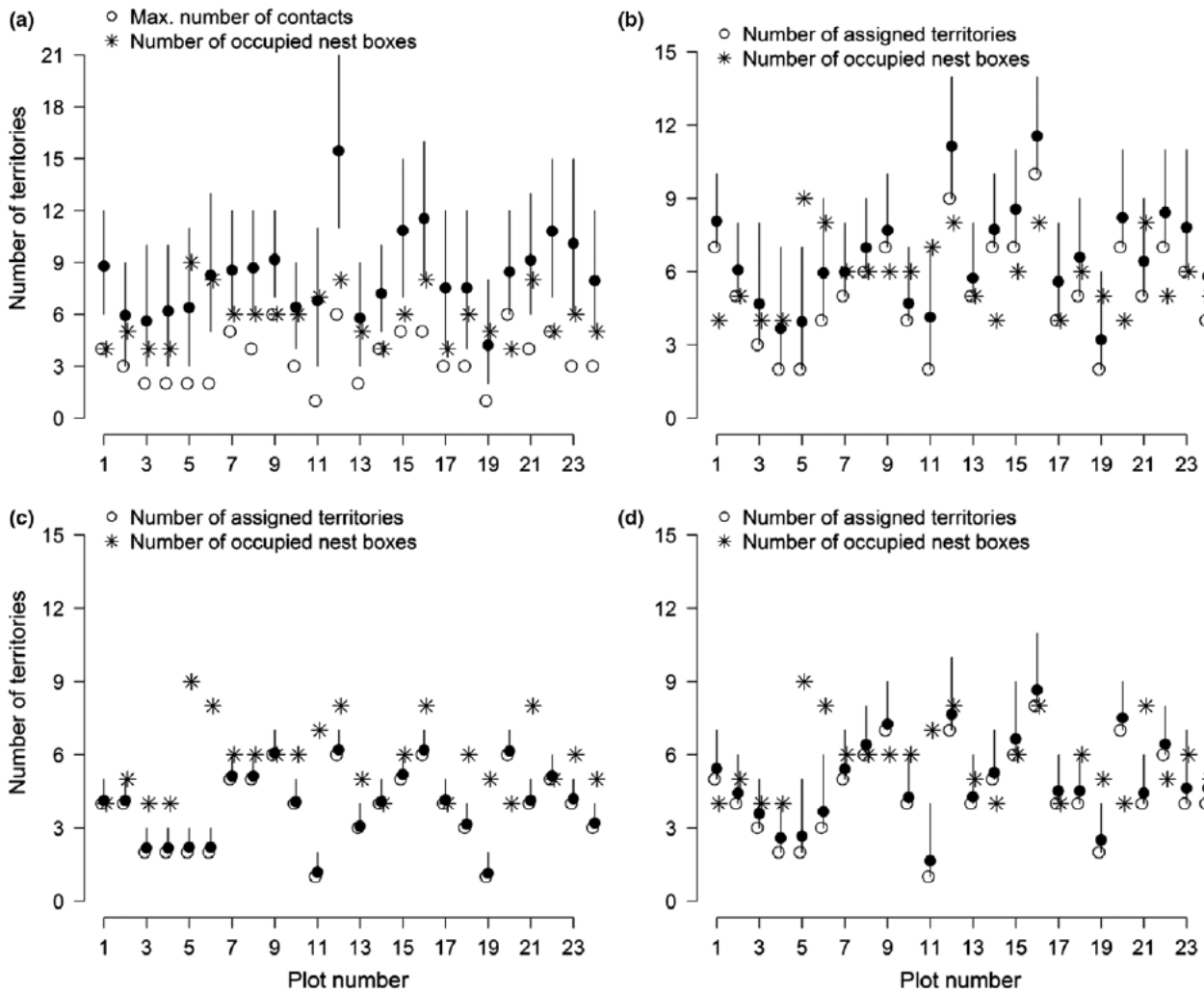


**Figure 2.** Plot-specific Blue Tit abundance estimates for 2014 (see Fig. S1 for 2015). The black dots are estimates from either binomial models (a) or multinomial models for each territory editor (b–d). Error bars are 95% confidence intervals.

that were occupying nestboxes. The cause of these discrepancies cannot be assessed with our data.

One reason for a lower number of occupied nestboxes compared with the abundance estimates could be that some pairs of tits were breeding in natural cavities. Natural cavities were certainly present in our study area, but the high number of occupied nestboxes shows that nestboxes were preferentially used. Nevertheless, this factor would lead to higher estimates than the benchmark even for a perfect model. Furthermore, the surveyed population might not exactly match the nestbox population because, in the former, territories at the edge of a plot may have been included in the count (i.e. the ‘sunflower effect’ of Kéry & Royle 2016: pp. 279–282). Hence, even for

perfect model estimates, we might have slightly higher estimates than the benchmark. Finally, at our study sites, natural predators of cavity-nesting birds included several woodpecker species, Eurasian Jay *Garrulus glandarius* and Pine Marten *Martes martes*, all of which were widespread. We had some losses that we could clearly assign to predation, but this did not influence our analyses, as we designated a nestbox as occupied as soon as one egg was laid. Other ecological processes, such as competition for cavities, affected our study in the sense that for the second year (2015) the density of Great Tits was much higher than in 2014, and therefore Blue Tits occupied a smaller number of nestboxes and potentially switched to natural cavities. This may explain the



**Figure 3.** Plot-specific Great Tit abundance estimates for 2015 (see Fig. S2 for 2014). The black dots are estimates from either binomial models (a) or multinomial models for each territory editor (b–d). Error bars are 95% confidence intervals. Note the different y-axis range for (a).

‘bad fit’ between the number of territorial contacts, the number of assigned territories and modelled abundance, and our benchmark, the number of occupied nestboxes (especially obvious for Blue Tits in 2015, see Fig. S1). The number of occupied nestboxes, therefore, depended on multiple ecological processes and is not a true ‘gold standard’ for evaluating model performance. However, we still think that it is a useful benchmark in practice.

Estimates using multinomial-mixture models were more precise than those using the binomial mixture models, presumably because they exploit somewhat more information (territory detection history frequencies provide better information about minimal abundance than do maximum

counts) (Kéry & Royle 2010, Kéry 2018). From this point of view, one might recommend the use of multimix models for estimating abundance, if applicable (e.g. for territorial species). However, as abundance estimates depend strongly on the territory assignment behaviour of a given ‘editor’ (see Table 1), we recommend paying attention to this issue, especially when comparisons between different study sites are the main interest (Best 1975, Verner & Milne 1990), because an editor’s tendency to ‘lump or split’ observations when assigning territory identities then influences the multimix abundance and detection estimates. If strong lumping occurs, the detection probability will be high, and therefore the ‘detection-correction’ for abundance will be low

(see Table 1, Editor 2, or Fig. S1c). On the other hand, if splitting occurs, the estimate of detection probability will be low, and therefore abundance estimates will be substantially corrected, resulting in high abundance estimates (Table 1, Editor 1). The binomial mixture models are independent of this source of variation (at least when multiple detections during the same survey are not allowed within one territory). However, due to their use of input data with less information, estimates are generally larger and more imprecise, as the larger confidence intervals in this study show (Figs 2 & 3, Figs S1 & S2; see also Kéry 2018, Kéry & Royle 2010).

Accounting for variation between territory editors is highly recommended, although it is only a solution within a study and not between studies. Other solutions might include an automated territory assignment algorithm, which would be faster and more reliable, in the sense that everywhere at any time the same criteria are used for assigning a territory. This idea is not new (North 1977, Gerß 1984, Scheffer 1987) but until now, only one algorithm has been used regularly in the Netherlands, as far as we know (Van Dijk *et al.* 2013). On the other hand, this solution only creates one well-defined territory assignment and is not necessarily better than simply using one human territory editor. The best solution may be to distribute the surveys randomly (time-wise and spatially) to editors, so that no systematic biases are introduced. In future, it seems possible that current work on spatial partial identity models in the context of spatial capture–recapture models (Augustine *et al.* 2019) may be extended to deal with the uncertainty associated with territory assignment of individual bird detections. To obtain further insights into this problem and assess its extent, a survey could be done by a 'naïve' observer/editor on a plot where all territories/nests are known (e.g. with ringed birds).

In summary, we conclude that both multinomial and binomial N-mixture models may represent good methods for estimation and modelling of abundance from unmarked individuals, even with fairly small sample sizes, as shown in our study. Multinomial models have been known to provide more precise estimates but are sensitive to splitting and lumping errors in the assignment of records to territory (Kéry & Royle 2010). Here, we emphasize also their sensitivity to interobserver variation in the degrees of these assignment errors. Thus, deciding on the best combination of field and analytical protocols for abundance assessments for unmarked populations will have to be done carefully.

We thank M. Romanski and J.-L. Dessolin from the Office National des Forêts (ONF) for allowing us to perform this study in the marvellous Forêt domaniale de Chaux. Many thanks to all trainees for helping during the nestbox checks, and special thanks to H. F. Lemke, who surveyed half of the plots in both years. We would like to thank two colleagues at our institute (D. Hagist and Ch. Rogenmoser) for assigning the territories from the survey-data maps. Thanks also go to S. Wechsler, who supported us with literature on territory assignment of territory-mapping data. We thank the two anonymous reviewers for their suggestions.

## AUTHORS' CONTRIBUTIONS

Y.B. and L.J. conceived the ideas and designed the study. Y.B. collected and, together with M.K., analysed the data. Y.B. and M.K. led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## DATA AVAILABILITY STATEMENT

The data used for this manuscript will be uploaded on the DRYAD digital repository (Bötsch *et al.* 2019).

## REFERENCES

- Akaike, H. 1987. Factor analysis and AIC. *Psychometrika* **52**: 317–332.
- Augustine, B.C., Royle, J.A., Murphy, S.M., Chandler, R.B., Cox, J.J. & Kelly, M.J. 2019. Spatial capture–recapture for categorically marked populations with an application to genetic capture–recapture. *Ecosphere* **10**: e02627.
- Barker, R.J., Schofield, M.R., Link, W.A. & Sauer, J.R. 2018. On the reliability of N-mixture models for count data. *Biometrics* **74**: 369–377.
- Best, L.B. 1975. Interpretational errors in the 'Mapping Method' as a census technique. *Auk* **92**: 452–460.
- Bibby, C.J., Burgess, N.D., Hill, D.A. & Mustoe, S.H. 2000. *Bird Census Techniques*, 2nd edn. London: Academic Press.
- Bötsch, Y., Jenni, L. & Kéry, M. 2019. Field evaluation of abundance estimates under binomial and multinomial N-mixture models. Dryad, Dataset, <https://doi-org.eres.library.manoa.hawaii.edu/dryad.fttdz08nx>
- Bötsch, Y., Tablado, Z. & Jenni, L. 2017. Experimental evidence of human recreational disturbance effects on bird-territory establishment. *Proc. R. Soc. B* **284**: 20170846.
- Buckland, S.T., Anderson, D., Burnham, K., Laake, J., Thomas, L. & Borchers, D. 2001. *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford: Oxford University Press.
- Buckland, S.T., Rexstad, E.A., Marques, T.A. & Oedekoven, C.S. 2015. *Distance Sampling: Methods and Applications*. Berlin: Springer.



- Dennis, E.B., Morgan, B.J.T. & Ridout, M.S. 2015. Computational aspects of N-mixture models. *Biometrics* **71**: 237–246.
- Dorazio, R.M. & Royle, J.A. 2005. Estimating size and composition of biological communities by modeling the occurrence of species. *J. Am. Stat. Assoc.* **100**: 389–398.
- Dorazio, R.M., Jelks, H.L. & Jordan, F. 2005. Improving removal-based estimates of abundance by sampling a population of spatially distinct subpopulations. *Biometrics* **61**: 1093–1101.
- Duarte, A., Adams, M.J. & Peterson, J.T. 2018. Fitting N-mixture models to count data with unmodeled heterogeneity: bias, diagnostics, and alternative approaches. *Ecol. Modell.* **374**: 51–59.
- Farnsworth, G.L., Pollock, K.H., Nichols, J.D., Simons, T.R., Hines, J.E. & Sauer, J.R. 2002. A removal model for estimating detection probabilities from point-count surveys. *Auk* **119**: 414–425.
- Farnsworth, G.L., Nichols, J.D., Sauer, J.R., Fancy, S.G., Pollock, K.H., Shriner, S.A. & Simons, T.R. 2005. *Statistical Approaches to the Analysis of Point Count Data: A Little Extra Information Can Go a Long Way*. Washington, DC: U.S. Forest Service.
- Fiske, I. & Chandler, R. 2011. unmarked: an R package for fitting hierarchical models of wildlife occurrence and abundance. *J. Stat. Softw.* **43**: 1–23.
- Gerß, W. 1984. Automatische Revierabgrenzung bei Siedlungsdichteuntersuchungen. *J. Ornithol.* **125**: 189–199.
- Kéry, M. 2018. Identifiability in N-mixture models: a large-scale screening test with bird data. *Ecology* **99**: 281–288.
- Kéry, M. & Royle, J.A. 2010. Hierarchical modelling and estimation of abundance and population trends in metapopulation designs. *J. Anim. Ecol.* **79**: 453–461.
- Kéry, M. & Royle, J.A. 2016. *Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS: Volume 1: Prelude and Static Models*. London: Elsevier/Academic Press.
- Kéry, M. & Schaub, M. 2012. *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective*. New York: Academic Press.
- Knape, J., Arlt, D., Barraquand, F., Berg, A., Chevalier, M., Part, T., Ruete, A. & Zmihorski, M. 2018. Sensitivity of binomial N-mixture models to overdispersion: the importance of assessing model fit. *Methods Ecol. Evol.* **9**: 1–13.
- Krebs, J.R. 1971. Territory and breeding density in the Great Tit, *Parus major* L. *Ecology* **52**: 2–22.
- Link, W.A., Schofield, M.R., Barker, R.J. & Sauer, J.R. 2018. On the robustness of N-mixture models. *Ecology* **99**: 1547–1551.
- Löhr, H. 1977. Nistökologische und ethologische Anpassungserscheinungen bei Höhlenbrütern. *Vogelwarte* **29**: 92–101.
- Nichols, J.D., Hines, J.E., Sauer, J.R., Fallon, F.W. & Jane, E. 2000. A double-observer approach for estimating detection probability and abundance from point counts. *Auk* **117**: 393–408.
- Nilsson, S.G. 1984. The evolution of nest-site selection among hole-nesting birds: the importance of nest predation and competition. *Ornis Scand.* **15**: 167–175.
- North, P.M. 1977. A novel clustering method for estimating numbers of bird territories. *Appl. Stat.* **26**: 149–155.
- R Core Team 2016. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>
- Royle, J.A. 2004a. Generalized estimators of avian abundance from count survey data. *Anim. Biodivers. Conserv.* **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. & Dorazio, R.M. 2008. *Hierarchical Modeling and Inference in Ecology: the Analysis of Data from Populations, Metapopulations and Communities*. London: Elsevier.
- Royle, J.A., Kéry, M., Gautier, R. & Schmid, H. 2007. Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecol. Monogr.* **77**: 465–481.
- Scheffer, M. 1987. An automated method for estimating the number of bird territories from an observation map. *Ardea* **75**: 231–236.
- Van Dijk, A.J., Troost, G., Vergeer, J.W., Sierdsema, H., Van Turnhout, C. & Noback, M. 2013. De introductie van autocluster in het broedvogel monitoring project. *Limosa* **86**: 94–102.
- Verner, J. & Milne, K.A. 1990. Analyst and observer variability in density estimates from spot mapping. *Condor* **92**: 313–325.
- Webster, R.A., Pollock, K.H. & Simons, T.R. 2008. Bayesian spatial modeling of data from avian point count surveys. *J. Agric. Biol. Environ. Stat.* **13**: 121–139.
- Williams, B.K., Nichols, J.D. & Conroy, M.J. 2002. *Analysis and Management of Animal Populations*. San Diego: Academic Press.

Received 10 February 2019;  
revision accepted 4 November 2019.  
Associate Editor: Auriel Fournier.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Figure S1.** Plot-specific Blue Tit abundance estimates for 2015.

**Figure S2.** Plot-specific Great Tit abundance estimates for 2014.

**Table S1.** Coefficients of variation for the number of assigned territories for the two study years and species, respectively.

**Table S2.** *P*-values for Goodness of Fit test statistics (with Freeman–Tukey discrepancy measure) for the four models (binmix and three multi-mix models for the three different editors) separately per study year and species.

**Table S3.** Sums-of-squares decomposition (%) for three linear models (ANOVA) showing how much editor identity influences the number of assigned territories: first an overall model and then two additional models for each study year.