

Hierarchical modelling and estimation of abundance and population trends in metapopulation designs

Marc Kéry^{1*} and J. Andrew Royle²

¹Swiss Ornithological Institute, 6204 Sempach, Switzerland; and ²U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD 20708, USA

Summary

1. Population assessment in changing environments is challenging because factors governing abundance may also affect detectability and thus bias observed counts. We describe a hierarchical modelling framework for estimating abundance corrected for detectability in metapopulation designs, where observations of 'individuals' (e.g. territories) are replicated in space *and* time. We consider two classes of models; first, we regard the data as independent binomial counts and model abundance and detectability based on a product-binomial likelihood. Secondly, we use the more complex detection–non-detection data for each territory to form encounter history frequencies, and analyse the resulting multinomial/Poisson hierarchical model. Importantly, we extend both models to directly estimate population trends over multiple years. Our models correct for any time trends in detectability when assessing population trends in abundance.

2. We illustrate both models for a farmland and a woodland bird species, skylark *Alauda arvensis* and willow tit *Parus montanus*, by applying them to Swiss BBS data, where 268 1 km² quadrats were surveyed two to three times during 1999–2003. We fit binomial and multinomial mixture models where log(abundance) depended on year, elevation, forest cover and transect route length, and logit(detection) on year, season and search effort.

3. Parameter estimates were very similar between models with confidence intervals overlapping for most parameters. Trend estimates were similar for skylark (-0.074 ± 0.041 vs. -0.047 ± 0.019) and willow tit (0.044 ± 0.046 vs. 0.047 ± 0.018). As expected, the multinomial model gave more precise estimates, but also yielded lower abundance estimates for the skylark. This may be due to effects of territory misclassification (lumping error), which do not affect the binomial model.

4. Both models appear useful for estimating abundance and population trends free from distortions by detectability in metapopulation designs with temporally replicated observations. The ability to obtain estimates of abundance and population trends that are unbiased with respect to any time trends in detectability ought to be a strong motivation for the collection of replicate observation data.

Key-words: Bayesian, Poisson GLMM, hierarchical model, population dynamics, population trends, WinBUGS

Introduction

Detection and explanation of spatial and temporal patterns in abundance lie at the heart of ecology (Krebs 2001) as well as of its applications such as conservation biology, pest management or monitoring science (Caughley 1994; Norris 2004). Abundance N (also called local abundance N_i at site i) is the key state variable for describing populations, but it can

hardly ever be measured without error due to imperfect detection probability p —in most situations, some individuals will be missed, that is, detectability $p < 1$. Hence, simple counts C are not equivalent to N but are related to abundance by the well-known relationship $E(C) = Np$ (Williams, Nichols & Conroy 2002); they are only *indices* to abundance with the expected value E of a count being a proportion p of N . Absent double counting, simple counts almost always underestimate abundance. In addition, spatial and temporal patterns in simple counts will be due to both patterns in abundance *and* patterns in detection (Kéry 2008). Hence,

*Correspondence author. E-mail: marc.kery@vogelwarte.ch, marc_kery@hotmail.com

when unbiased estimates of abundance are required or when population trends need to be assessed free from possible distorting patterns in detectability, abundance must be estimated separately from detectability (MacKenzie & Kendall 2002; Kéry & Schmid 2004; Kéry & Schmidt 2008).

Over the last decades, an armada of protocols and associated statistical models have been developed to 'adjust' simple counts by an estimate of detection probability and thus arrive at an estimate of abundance. Examples include distance sampling (Buckland *et al.* 2001) and a large array of capture–recapture protocols (Borchers, Buckland & Zucchini 2002; Williams *et al.* 2002). Distance sampling uses the distribution of detection distances to provide information about detection probability, while capture–recapture uses the pattern of detection/non-detection over replicated surveys from a period during which a population can be assumed static or closed. In both frameworks, the estimate of detection probability provides the direct link between the observed count and the estimate of population size.

Much of ecology and its applications is concerned with comparisons of abundance in space, and consequently abundance is frequently assessed at multiple sites and using similar protocols at each. This, in essence, represents a metapopulation design (Royle 2004b). Analysis of abundance and detection on a site-by-site basis would be highly inefficient or sometimes impossible due to locally small sample size or even zero counts. Instead, an integrated analysis is required for the most efficient use of available information and to directly model patterns of abundance.

Recently, hierarchical models have been developed for inference about abundance and detection that explicitly account for a metapopulation design (Royle & Dorazio 2006, 2008; also see Borchers *et al.* 2002). These models have one stochastic component to describe spatial and possibly temporal variation in abundance and another stochastic component that specifies the stochastic outcome of the observation process. The beauty of these models is that, conceptually, they provide a truly mechanistic rendering of how counts of organisms arise as a result of two linked stochastic processes, an ecological process and a dependent observation process. Different descriptions of these two processes can simply be combined in a modular fashion as needed for the particular study at hand. This means that a large variety of animal sampling protocols can all be subsumed into a generic hierarchical model by simply using different stochastic descriptions of the observation process (Royle 2004b). Examples include distance sampling (Royle, Dawson & Bates 2004), point counts (Royle 2004a), removal sampling (Dorazio, Jelks & Jordan 2005), detection–non-detection (a.k.a. 'presence–absence') sampling (MacKenzie *et al.* 2002; Royle & Nichols 2003; Dorazio 2007; Royle & Kéry 2007) and capture–recapture proper (Royle *et al.* 2007; Webster, Pollock & Simons 2008).

Furthermore, these hierarchical models are just variations of the Poisson generalized linear model (GLM) or generalized linear mixed model (GLMM), both of which are commonly used to model population change from count data (Link & Sauer 2002; Gregory *et al.* 2005; Link, Sauer &

Niven 2006). Thus, all what is known from this important class of models (see e.g. McCulloch & Searle 2001; Lee, Nelder & Pawitan 2006) could be directly carried over to these hierarchical models. Examples include the introduction of covariate information by means of a link function (McCullagh & Nelder 1989), the introduction of random effects and correlation among parameters (Lee *et al.* 2006) or the extension to additive models with smooth terms (Wood 2006).

In this article, we compare two classes of hierarchical models that are particularly useful for inference about abundance in metapopulation designs, that is, when observations of 'individuals' are replicated in space *and* time; the binomial (Royle 2004a) and the multinomial mixture model (Royle *et al.* 2007). 'Individuals' may be any individually recognizable units, such as individual animals or plants, breeding pairs or territories, or even species (Kéry 2009). Here, we use territory-mapping data from the Swiss Breeding bird survey MHB (Schmid, Zbinden & Keller 2004) for two species; hence, individuals represent individual territories. First, in the binomial mixture model, we regard the data as independent binomial counts and inference is based on a product-binomial/Poisson hierarchical model. Secondly, we use the more complex detection–non-detection data for each territory to form encounter history frequencies for each site, and our analysis is based on a multinomial/Poisson hierarchical model. As the data for the former are just an aggregated form of the more detailed data format used in the latter, we expect very similar inferences under these two models. We hypothesize better precision for the multinomial model because it uses a more detailed (i.e. information-rich) format of the data. However, the data collection assumptions are somewhat more strict as we describe subsequently.

Importantly, in our comparison of the binomial and the multinomial mixture models, we extend both models to directly estimate population trends over multiple years (see also Royle & Dorazio 2008, pp. 4–7 and Kéry *et al.* 2009 for similar models). That is, our models enable one to estimate population trends corrected for any (parallel) time trends that might exist in detectability. We believe that the ability to directly model population dynamics (here, a simple log-linear population trend) embedded within a framework that fully accounts for the observation process (here, imperfect detection) will be of great value to monitoring and ecological studies alike.

Materials and methods

SWISS BREEDING BIRD SURVEY MHB

The Monitoring Häufige Brutvögel (MHB) is the Swiss national breeding bird survey and has been described extensively elsewhere (e.g. Kéry & Schmid 2004, 2006; Schmid *et al.* 2004; Kéry & Royle 2009). Here, we only describe some key features relevant to this study. MHB is based on 268 1-km² sampling units (quadrats) laid out as a grid. Since 1999, every quadrat has been surveyed two to three times during most breeding seasons (15 April to 15

July) using territory mapping (Bibby *et al.* 2000), a protocol that yields territory encounter histories. An example might be (0,1,0) for a territory encountered on the second of three surveys only, and another (1,0,–) for one encountered on the first survey in a quadrat surveyed only twice. Surveys are conducted along irregular transects that aim to cover as large a proportion of the quadrat as possible. These transects differ among quadrats (mean 5.1; range 1.2–9.4 km) but are constant across years. Survey duration, expressed as time spent per km, is not standardized but varies over 11-fold (mean 48; 15–167 min km⁻¹).

STUDY SPECIES AND DATA

For our model comparison we chose MHB data from 1999 to 2003 for a farmland and a woodland species, skylark (*Alauda arvensis* L.) and willow tit (*Parus montanus* Conrad). Both are widespread but moderately rare species. Observed mean counts in quadrats where they were ever encountered during 1999–2003 ranged 0.07–14.6 (median 0.8) for the lark and 0.07–12.5 (median 1.0) for the tit, and the annual total number of mapped territories ranged 263–321 and 400–557 respectively. We chose these two species because they did not challenge our model assumptions: closed populations during the entire sampling period (though see later) and fairly well recognizable territories. Owing to its high and wide-ranging, conspicuous song-flight and loud song, the skylark is easy to detect. The willow tit on the other hand has a weak song and elusive behaviour and can be rather difficult to detect. Also see Kéry, Royle & Schmid (2005) and Royle, Nichols & Kéry (2005) for analyses of a 2002 subset of these data. For analysis under the binomial mixture model (Royle 2004a), we aggregated the encounter history data to territory counts. In contrast, for the multinomial mixture model, we formed territory encounter history frequencies for each site (for an example, see Royle *et al.* 2007; Table 1).

STATISTICAL MODELS

We describe our data, either territory counts or territory encounter history frequencies, by means of hierarchical models that specify the two stochastic mechanisms, or processes, involved in their genesis. First, the ecological process distributes territories in space, i.e. across R sample quadrats. This results in a latent distribution of realized local population sizes N_i . The result of this stochastic process is latent because local population size N_i is only imperfectly observed due to possibly overlooked territories. Secondly, given the realized local population size N_i in quadrat i , the observation process describes the observations as a result of binomial sampling – at every survey j , each territory can be either detected with probability p_j or missed with $1 - p_j$. Depending on the observation process, or the format of the data that are analysed (i.e. unaggregated frequencies of encounter history or aggregated to counts), we have a binomial or a multinomial mixture model.

Multinomial observation model

The territory encounter history data are analogous to data that arise in classical capture–recapture studies. For a study based on $J = 3$ surveys of a population, there are eight possible encounter histories representing all possible combinations of ‘encountered’ ($y = 1$) and ‘not encountered’ ($y = 0$). For example, a possible encounter history h is $h = (1,0,0)$ for an individual encountered in the first survey but not subsequently. We denote by $n_{ik}; k = 0,1,2,\dots,7$ the number of each type of encounter history for quadrat i , where we use the integers k here to index the eight nominal categories representing possible encounter histories. We denote by n_{i0} the number of individuals never encountered, i.e. those having encounter history $h = (0,0,0)$. The encounter history frequencies for a specific quadrat i are described by a multinomial distribution having ‘sample size’ N_i and cell probabili-

Table 1. Comparison of parameter estimates (posterior means, SD, lower (2.5%) and upper credible (97.5%) limits) under the binomial and the multinomial mixture models for Swiss skylarks (*Alauda arvensis*) 1999–2003

	Posterior							
	Binomial mixture				Multinomial mixture			
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
loglam0	–2.177	0.919	–3.926	–0.210	–2.325	0.739	–3.915	–0.980
r	–0.074	0.041	–0.153	0.012	–0.047	0.019	–0.086	–0.010
bele1	–0.747	0.295	–1.324	–0.191	–0.688	0.266	–1.221	–0.177
bele2	–0.205	0.315	–0.819	0.416	–0.165	0.276	–0.730	0.370
bforest	–1.885	0.351	–2.614	–1.243	–1.707	0.344	–2.413	–1.050
blength	–2.895	3.757	–11.05	4.273	–2.858	3.147	–8.813	3.476
sigma.lam	3.132	0.270	2.646	3.701	2.994	0.269	2.509	3.581
p0.1999	0.445	0.048	0.351	0.536	0.622	0.027	0.569	0.674
p0.2000	0.484	0.041	0.403	0.562	0.678	0.025	0.627	0.725
p0.2001	0.502	0.038	0.431	0.578	0.690	0.025	0.640	0.739
p0.2002	0.463	0.038	0.388	0.539	0.643	0.030	0.584	0.699
p0.2003	0.545	0.049	0.442	0.634	0.656	0.027	0.600	0.708
bday1	–0.028	0.037	–0.099	0.043	–0.033	0.051	–0.132	0.068
bday2	0.027	0.031	–0.032	0.088	0.065	0.048	–0.029	0.158
brate	0.192	0.070	0.055	0.328	0.079	0.061	–0.040	0.199

Notation (also see main text): loglam0, log(saturation density); r , annual log-linear trend in abundance; bele1 and bele2, linear and quadratic log-linear effects of elevation on abundance; bforest, log-linear effect of forest cover on abundance; blength, log-linear effect of inverse route length on abundance; sigma.lam, standard deviation of Normal distribution used to account for overdispersion in abundance; p0.1999–p0.2003, mean detection probability per territory and survey for the years 1999–2003; bday1 and bday2, logit-linear effect of Julian day on detection probability; brate, logit-linear effect of survey effort on detection probability.

ties π_k , which are functions of the parameters that describe the encounter process. As an example, for the encounter history $h = (1, 0, 0)$, we have that $\pi_h = p_1(1-p_2)(1-p_3)$. The cell probabilities are constructed as in standard capture–recapture models (Williams *et al.* 2002). The probability mass function for the observations made at quadrat i is given by

$$f(\mathbf{n}_i | N_i, \mathbf{p}) \propto \frac{N_i!}{n_{0i}!} \prod_{k=1}^7 \pi_k^{n_{ik}}$$

For this model, the parameters are n_{0i} for each quadrat as well the detection probabilities for each survey that are implied by π_k . Note that $N_i = n_{0i} + \sum_{k=1}^7 n_{ik}$ where the encounter history frequencies $k = 1, 2, \dots, 7$ are observed.

Binomial observation model

While multinomial observation models are very common in animal sampling applications (e.g. capture–recapture), such data are generally more difficult to obtain because it is (usually) required that an individual's identity can be preserved across the replicate surveys. In the case of the MHB, we are asserting that the identity of each territory is established upon first observation and that it can be reconciled unambiguously with previous and subsequent observations.

A simpler data structure that makes less stringent assumptions about the data collection process arises by considering the counts of unique individuals during each of the J sample periods, say $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iJ})$ for site i . A natural view is that such counts have a binomial distribution with sample size N and parameter p . The probability mass function for such independent counts is a product of binomials:

$$f(\mathbf{y}_i | N_i, \mathbf{p}) \propto \prod_{j=1}^J \frac{N_i!}{y_{ij}!} p_j^{y_{ij}} (1-p_j)^{N_i-y_{ij}}$$

The binomial model is common in many animal sampling situations. For example, it also underlies most capture–recapture methods (as a model for the number of unique individuals counted). However, unlike capture–recapture models for which inference about parameters can be carried out effectively from a single multinomial sample, the information about model parameters from the binomial mixture model is obtained from a repeated-measures type of design: replicate counts are obtained at a sample of spatial locations (Royle 2004a).

For the MHB, with $J = 3$ surveys, the data obtained under the binomial counting view consists of three counts instead of the seven encounter history frequencies under the multinomial observation view. As such, we would expect some loss of information. In fact, binomial counts are linear functions of the multinomial cell frequencies. Specifically,

$$\begin{aligned} y_1 &= n_{100} + n_{110} + n_{101} + n_{111} \\ y_2 &= n_{010} + n_{011} + n_{110} + n_{111} \\ y_3 &= n_{001} + n_{011} + n_{101} + n_{111} \end{aligned}$$

Modelling variation in N

We used log-linear models to accommodate suspected structure in our latent variables N_i . Specifically, we fitted the effects of elevation (linear and quadratic) and forest cover, as these variables are known to explain much spatial variation in avian densities in Switzerland (Kéry *et al.* 2005). In addition, we introduced into the model the reciprocal of transect route length to account for incomplete quadrat

coverage achieved by the variable transect routes. This parameterisation of the covariate route length has the effect that the intercept of the analysis (loglam0, see) becomes saturation density, i.e. the number of territories that would be exposed to detection were the quadrat covered completely. Accordingly, the coefficient of the reciprocal of route length will normally be negative (Royle *et al.* 2007). Finally, to account for spatial variation in abundance that is not captured by these covariates and the Poisson assumption, we added a normally distributed site random effect into the linear predictor for abundance. This allows for overdispersion relative to the conditional Poisson distribution.

Most importantly, our latent variables differ by year T , hence, N is indexed by site i and year t . The essential difference between the models developed in this article and virtually all of their previous published applications is that we fit a log-linear trend to our multi-year data (see Royle & Dorazio 2008, pp. 4–7 and Kéry *et al.* 2009 for similar models). That is, we fit the simplest possible population dynamics model, a log-linear trend, within a model that fully accounts for imperfect detection. For the expected count λ_{it} in quadrat i and year t we specified $\log(\lambda_{it}) = \text{loglam0} + rt$. Here, r is just the slope of a log-linear Poisson regression of the unknown, true local population size, N_{it} on year t , i.e. the population trend.

Modelling variation in p

In many practical situations, it may be possible to identify explicit covariates that influence the ability of observers to detect individuals during surveys, but not the number of individuals *exposed* to sampling. For example, time of day or temperature may influence detectability of many species for behavioural reasons. Also, certain basic elements of the protocol may influence detectability (e.g. 5 min counts vs. 10 min counts; whether playbacks are used, etc.). To accommodate these considerations, we can develop models for the detection probability parameters p_j in the models described above. In general, we can index detection probability by both sample location, i and survey occasion j and then model factors thought to influence detection probability on the logit- p scale. For the MHB study, the logit of detection probability was expressed as a function of year, survey rate [i.e. duration (min)/transect length (km)] as well as linear and quadratic effects of the day in the season. All covariates except for one (inverse route length) were standardized before analysis to produce more efficient estimation by the Markov chain Monte Carlo (MCMC) algorithms used. As for abundance, we accounted for overdispersion in detection probability by introducing into the linear predictor for detection a survey-specific normal random effect. We did this for the willow tit only, as accounting for overdispersion in abundance alone yielded fitting models for the skylark.

The 'likelihood'

The model as described here with either the multinomial or binomial observation model *and* a model for the latent abundance parameters N_{it} is a special type of hierarchical model (Royle & Dorazio 2008, see Ch. 8) which can be analysed either by classical likelihood or Bayesian methods. Likelihood analysis of hierarchical models can be carried out based on the integrated likelihood. Alternatively, Bayesian analysis using modern methods of MCMC is also straightforward using specialized software (e.g. WinBUGS, Spiegelhalter, Thomas & Best 2003). Both methods of inference are described in Royle & Dorazio (2008), Ch. 8) and elsewhere.

When N is Poisson, the latent N parameters can be removed from the likelihood analytically. Specifically, the encounter history

frequencies n_{ik} have a Poisson distribution with mean $\lambda_i \pi_{ik}$. This is computationally convenient for analysis whether by likelihood methods or MCMC (Dorazio *et al.* 2005; Royle *et al.* 2005).

ANALYSIS OF THE MODELS

We used WinBUGS run from R (R Development Core Team 2009) via the R2WinBUGS interface (Sturtz, Ligges & Gelman 2005) to fit the models. We used conventional 'vague' priors throughout (see also BUGS code in Appendices 1 and 2). In our analysis, the number of surveys per quadrat and year varied from 2 to 3 (for treatment of missing values, see Appendix 3).

We based our inference on 3000 random draws from the joint posterior distribution (see Appendix 4 for details on MCMC analysis). We assessed the goodness-of-fit of our four models using a Bayesian P -value based on a chi-squared discrepancy measure (Gelman *et al.* 2004). This resulted in values of 0.81 and 0.86, respectively, for the binomial and the multinomial mixture models and the lark data and in corresponding values of 0.18 and 0.67 for the tit data. All four indicated acceptable to good fit of the models to our data sets.

Results

For both species, estimates of key parameters were similar under the binomial and the multinomial mixture models (Tables 1 and 2). In particular, trend estimates (\pm SE) were very similar for the skylark (-0.074 ± 0.041 vs. -0.047 ± 0.019 , for the binomial and the multinomial mixture model respectively) and the willow tit (0.044 ± 0.046 vs. 0.047 ± 0.018). For most parameter estimates, and especially for the trend estimates, credible intervals under the two models overlapped. Sole exceptions were four annual detection intercepts in the skylark, where estimates were higher

under the multinomial than under the binomial mixture model (Table 1). There was a slight tendency of the same pattern also in the willow tit (Table 2). As a consequence, the multinomial model yielded lower estimates of abundance than did the binomial mixture model (Fig. 1a,b). (Note that predictions for response to elevation are made at the mean value of all other covariates in the model which means, for instance, that they show the hypothetical response to elevation that *would* be expected at a constant forest cover of 35%.) As expected, the multinomial mixture model yielded estimates with greater precision than the binomial mixture model (Tables 1 and 2). In particular, SEs for the trend estimates under the binomial were about twice as large as those under the multinomial mixture model.

Results from both analyses concurred well with what we know about these species. The skylark is declining ($r < 0$) and avoids forest (bforest < 0), while the willow tit is increasing ($r > 0$) and is a forest bird (bforest > 0 ; Tables 1 and 2). The skylark was most abundant at lower elevations (Fig. 1c), while the willow tit had an intermediate optimum (Fig. 1d). For both species, abundance showed the expected negative response to increasing inverse route length, although the 95% credible interval for that parameter covered zero in the skylark. For both species, average detection probability varied among years, and the skylark was much easier detectable than the willow tit. Seasonal variation in detection probability was small in both species (Fig. 1e,f). The effect of search effort as measured by survey rate (min km^{-1}) was not quite significant in the frequent singer skylark under the multinomial model, but clearly positive in the less frequent, and weak, singer, the willow tit.

Table 2. Comparison of parameter estimates (posterior means, SD, lower (2.5%) and upper credible (97.5%) limits) under the binomial and the multinomial mixture models for Swiss willow tits (*Parus montanus*) 1999–2003

	Posterior							
	Binomial mixture				Multinomial mixture			
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
loglam0	0.242	0.550	−0.793	1.354	0.080	0.526	−0.933	1.158
r	0.044	0.046	−0.041	0.140	0.047	0.018	0.009	0.083
bele1	2.349	0.234	1.903	2.811	2.318	0.234	1.876	2.793
bele2	−1.221	0.231	−1.660	−0.774	−1.228	0.219	−1.659	−0.813
bforest	1.293	0.192	0.940	1.702	1.240	0.190	0.878	1.623
blength	−6.035	2.207	−10.44	−1.824	−4.865	2.208	−9.708	−0.838
sigma.lam	1.986	0.169	1.686	2.348	1.891	0.155	1.608	2.205
p0.1999	0.319	0.045	0.229	0.409	0.325	0.024	0.278	0.374
p0.2000	0.267	0.031	0.209	0.330	0.292	0.021	0.253	0.334
p0.2001	0.319	0.032	0.258	0.381	0.333	0.021	0.293	0.375
p0.2002	0.289	0.033	0.229	0.357	0.300	0.021	0.261	0.343
p0.2003	0.363	0.046	0.276	0.453	0.381	0.021	0.339	0.423
bday1	−0.172	0.043	−0.255	−0.085	−0.061	0.044	−0.145	0.022
bday2	0.067	0.029	0.011	0.126	0.056	0.034	−0.009	0.124
brate	0.213	0.066	0.078	0.340	0.035	0.047	−0.059	0.125
sigma.p	0.384	0.065	0.257	0.520	0.609	0.050	0.511	0.709

Notation as in Table 1 except for sigma.p, which is the standard deviation of the Normal distribution used to account for overdispersion in detection probability.

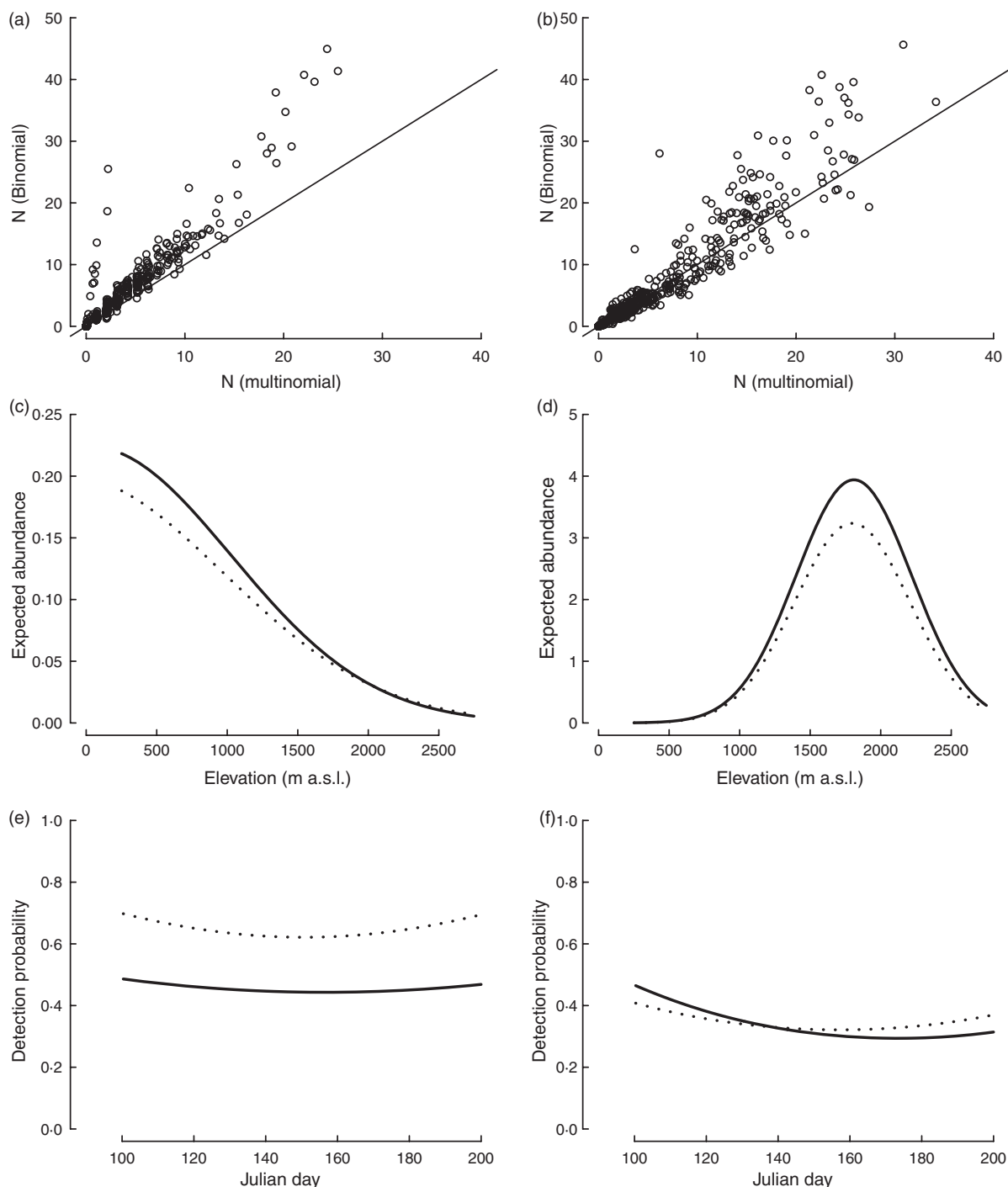


Fig. 1. (a, b) Predicted population size (N_i) under the binomial and the multinomial mixture models for the skylark (left panels) and the willow tit (right panels). (c, d) Expected abundance in relation to elevation under the binomial (solid line) and the multinomial mixture model (dotted line). (e, f) Expected detection probability per survey in relation to season (solid: binomial; dotted: multinomial; day 1 is 1 January). The relationships with season are averaged over the five years (1999–2003).

Discussion

We presented a comparison of two new hierarchical modelling frameworks to estimate abundance in metapopulation designs with temporal replicate observations. The multinomial mixture model (Dorazio *et al.* 2005; Royle *et al.* 2007;

Webster *et al.* 2008) is a multi-site, integrated version of the classical multinomial model widely used for capture–recapture data (e.g. Williams *et al.* 2002). That is, it requires replicate observations of individually recognizable units, such as individuals or, here, territories. These are expensive data, because individual identification may not be possible under

all circumstances or may be costly in time or effort. In contrast, the binomial mixture model (Royle 2004a) is based on the integration over multiple sites and replicate visits of counts without individual identification. These are much cheaper data, as just tallying up all detected individuals on each occasion separately will often be hardly more difficult than simply recording presence/absence (actually, detection/non-detection) in an occupancy study (MacKenzie & Nichols 2004).

Both models are well suited for inference about data from metapopulation designs in monitoring and similar studies, where the same observation protocol is applied to an array of spatial replicates (Royle 2004b). By specifying a weak stochastic relationship among the abundance parameters at different sites, they represent a much more flexible and parsimonious way of integrating data across replicate sites than by assuming that they are either all equal or all different (Gelman & Hill 2007). Temporal replicate observations enable one to decompose the observed variation in counts into effects from the unobserved biological process, represented by the abundance parameters, and those from the observation process, represented by the detection parameters.

Notably, we extend most previous applications of the binomial and multinomial mixture model to open populations. This allows us to directly model population trends fully embedded in an estimation framework that accounts for imperfect detection. This would appear to make these models attractive for assessing population change in a changing environment, where not only abundance but also detection in animal or plant populations may change over time.

So which one is the more useful model? Most results of our comparison were as expected: we found concurring estimates that were, however, more precise under the multinomial mixture model. In addition, unpublished simulation results show that mixing of the Markov chains in a Bayesian analysis is greatly improved for the multinomial as compared to the binomial mixture model. This is most likely due to the higher information content of encounter frequency data compared to replicated counts without individual identification. On the other hand, computational costs were about eight times smaller for the binomial mixture model (which meant hours instead of days on a fast laptop), and this may well be decisive in the analysis of very large data sets. So one might say that where individual encounter data are available, they are best analysed under the multinomial mixture model unless sample sizes are too large.

There is one further interesting issue, though: for both studied species, abundance estimates under the binomial mixture model tended to be higher than under the multinomial mixture model (although the 95% credible intervals of log₁₀ under both models overlapped). This was particularly surprising as in a similar comparison, Webster *et al.* (2008) found a much greater similarity between the abundance estimates under these two model classes than we did. We believe that the discrepancies in the N estimates between binomial and multinomial mixture model may be due to different effects of territory misclassification.

Table 3 shows how errors in territory identification between replicate visits affect the observed encounter history frequency data. Consequently, when applying a multinomial model to encounter history frequency data, lumping errors will bias abundance estimates low and splitting errors will bias them high. Interestingly, when analysing the same encounter histories aggregated to replicated counts under the binomial mixture model, estimates will be unaffected (Table 3). Individual identification can be very difficult, especially for acoustic detections (Allredge, Simons & Pollock 2007; Simons *et al.* 2007). Mistakenly attributing records of two birds from different territories to the same territory may thus account for the observed discrepancy between abundance estimates. It would seem that this was more pronounced for the skylark, which has a wide-ranging song-flight where individuals may be more difficult to assign to a territory than for the willow tit. Territory misclassification might also explain the differences between our results and those of Webster *et al.* (2008). Their occasions were contiguous 3, 2 and 5 min intervals, so presumably there was much less chance for misclassification.

As one would expect for a GLM-based class of models (McCullagh & Nelder 1989), our modelling framework is extremely flexible to extensions. First, as an alternative to an overdispersed Poisson distribution (i.e. the log-normal Poisson we assumed for abundance), other distributions could be used to specify the unstructured variation in the latent state, N_i , across quadrats i , such as the negative binomial or a zero-inflated Poisson distribution. Secondly, spatial correlation among site random effects may be added (Royle *et al.* 2007; Webster *et al.* 2008). Thirdly, one could also employ non-parametric modelling of the abundance distribution, e.g. by use of Dirichlet process priors (Dorazio *et al.* 2008) or by adding smooth terms as in a generalised additive model (GAM) (Wood 2006). Finally, truly individual effects at the level of the individual territory could be introduced if each

Table 3. Effects of two types of territory misclassification, lumping and splitting, on estimates of abundance under the binomial mixture model, based on territory counts, and on the multinomial mixture model, based on encounter frequencies. In the example, truth (true abundance) is represented by three territories with encounter histories (0,1,1), (1,0,0) and (0,0,1). The lumping error consists of erroneously assuming that detections in territory 2 and 3 represent a single territory. The splitting error consists of mistakenly assuming that the two detections in territory 1 belong to two different territories. Inference under the binomial mixture model is unaffected by either kind of error

	Lumping error		Splitting error	
	Truth	Observation	Truth	Observation
Encounter histories	0-1-1	0-1-1	0-1-1	0-1-0
	1-0-0	1-0-1	1-0-0	0-0-1
	0-0-1		0-0-1	1-0-0
				0-0-1
Territory counts	1-1-2	1-1-2	1-1-2	1-1-2

individual encounter history is modelled individually. In the context of the Swiss bird survey MHB, territory- and survey-specific covariates that appear useful include daytime (for a fine-scale modelling of the temporal patterns in detection probability within a morning), or territory-specific covariates such as the proximity to a road or another noise source (river, torrent, stream) to account for habitat-specific detection. As a special case of an individual effect, the coordinates of each detection could be formally integrated into an analysis for a spatial capture–recapture model (Efford 2004; Borchers & Efford 2008; Royle & Young 2008). In conclusion, where temporal replicate observations are available for at least part of a data set produced in a metapopulation design, we believe that the hierarchical models presented here offer an extremely powerful framework for inference about population dynamics free from the possibly distorting effects of imperfect detection. We would hope that this ability makes a strong argument for obtaining replicate observations in at least a subsample in ecological or applied studies that employ metapopulation designs.

Acknowledgements

We thank the volunteers in the Swiss Breeding bird survey Monitoring Häufige Brutvögel. We also thank Jérôme Duplain and Jérôme Guélat as well as two anonymous referees for useful comments.

References

- Allredge, M.W., Simons, T.R. & Pollock, K.H. (2007) A field evaluation of distance measurement error in auditory avian point count surveys. *Journal of Wildlife Management*, **71**, 2759–2766.
- Bibby, C.J., Burgess, N.D., Hill, D.A. & Mustoe, S. (2000) *Bird Census Techniques*. Academic Press, San Diego.
- Borchers, D.L. & Efford, M.G. (2008) Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics*, **64**, 377–385.
- Borchers, D.L., Buckland, S.T. & Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. Springer, New York.
- Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001) *Introduction to Distance Sampling*. Oxford University Press, Oxford.
- Caughley, G. (1994) Directions in conservation biology. *Journal of Animal Ecology*, **63**, 215–244.
- Dorazio, R.M. (2007) On the choice of statistical models for estimating occurrence and extinction from animal surveys. *Ecology*, **88**, 2773–2782.
- Dorazio, R.M., Jelks, H.L. & Jordan, F. (2005) Improving removal-based estimates of abundance by sampling a population at spatially distinct subpopulations. *Biometrics*, **61**, 1093–1101.
- Dorazio, R.M., Mukherjee, B., Zhang, L., Ghosh, M., Jelks, H.L. & Jordan, F. (2008) Modeling unobserved sources of heterogeneity in animal abundance using Dirichlet process prior. *Biometrics*, **64**, 635–644.
- Efford, M.G. (2004) Density estimation in live-trapping studies. *Oikos*, **106**, 598–610.
- Gelman, A. & Hill, J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge.
- Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. (2004) *Bayesian Data Analysis*, 2nd edn. CRC/Chapman & Hall, Boca Raton, FL.
- Gregory, R.D., van Strien, A., Vorisek, P., Gmelig Meyling, A.W., Noble, D.G., Foppen, R.P.B. & Gibbons, D.W. (2005) Developing indicators for European birds. *Philosophical Transactions of the Royal Society, Series B*, **360**, 269–288.
- Kéry, M. (2008) Estimating abundance from bird counts: binomial mixture models uncover complex covariate relationships. *Auk*, **125**, 336–345.
- Kéry, M. (2009) Species richness and community dynamics – a conceptual framework. *Camera Traps in Animal Ecology: Methods and Analyses* (eds A.F. O'Connell, J.D. Nichols & U.K. Karanth). Springer, New York.
- Kéry, M. & Royle, J.A. (2009) Inference about species richness and community structure using species-specific occupancy models in the national Swiss Breeding bird survey MHB. *Modeling Demographic Processes in Marked Populations*. Series: Environmental and Ecological Statistics, Vol. 3 (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 639–656. Springer, New York.
- Kéry, M. & Schmid, H. (2004) Monitoring programs need to take into account imperfect species detectability. *Basic and Applied Ecology*, **5**, 65–73.
- Kéry, M. & Schmid, H. (2006) Estimating species richness: calibrating a large-scale avian monitoring program. *Journal of Applied Ecology*, **43**, 101–110.
- Kéry, M. & Schmidt, B.R. (2008) Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**, 207–216.
- Kéry, M., Royle, J.A. & Schmid, H. (2005) Modeling avian abundance from replicated counts using binomial mixture models. *Ecological Applications*, **15**, 1450–1461.
- Kéry, M., Dorazio, R.M., Soldaat, L., van Strien, A., Zuiderwijk, A. & Royle, J.A. (2009) Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, **46**, in press.
- Krebs, C.J. (2001) *Ecology: The Experimental Analysis of Distribution and Abundance*, 5th edn. Benjamin Cummings, Menlo Park, CA.
- Lee, Y., Nelder, J.A. & Pawitan, Y. (2006) *Generalized Linear Models with Random Effects*. Chapman and Hall/CRC, Boca Raton.
- Link, W.A. & Sauer, J.R. (2002) A hierarchical analysis of population change with application to Cerulean warblers. *Ecology*, **83**, 2832–2840.
- Link, W.A., Sauer, J.R. & Niven, D.K. (2006) A hierarchical model for regional analysis of population change using Christmas bird count data, with application to the American black duck. *The Condor*, **108**, 13–24.
- MacKenzie, D.I. & Kendall, W.L. (2002) How should detection probability be incorporated into estimates of relative abundance? *Ecology*, **83**, 2387–2393.
- MacKenzie, D.I. & Nichols, J.D. (2004) Occupancy as a surrogate for abundance estimation. *Animal Biodiversity and Conservation*, **27.1**, 461–467.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A. & Langtimm, C.A. (2002) Estimating site occupancy rates when detection probability rates are less than one. *Ecology*, **83**, 2248–2255.
- McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*. Chapman and Hall, London.
- McCulloch, C.E. & Searle, S.R. (2001) *Generalized, Linear, and Mixed Models*. Wiley, New York.
- Norris, K. (2004) Managing threatened species: the ecological toolbox, evolutionary theory and declining-population paradigm. *Journal of Applied Ecology*, **41**, 413–426.
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.
- 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 spatially replicated count data. *Animal Biodiversity and Conservation*, **27.1**, 375–386.
- Royle, J.A. & Dorazio, R.M. (2006) Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 249–263.
- Royle, J.A. & Dorazio, R.M. (2008) *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations, and Communities*. Academic Press, San Diego.
- Royle, J.A. & Kéry, M. (2007) A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**, 1813–1823.
- Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.
- Royle, J.A. & Young, K.V. (2008) A hierarchical model for spatial capture–recapture data. *Ecology*, **89**, 2281–2289.
- Royle, J.A., Dawson, D.K. & Bates, S. (2004) Modeling abundance effects in distance sampling. *Ecology*, **85**, 1591–1597.
- Royle, J.A., Nichols, J.D. & Kéry, M. (2005) Modeling occurrence and abundance of species with imperfect detection. *Oikos*, **110**, 353–359.
- Royle, J.A., Kéry, M., Gautier, R. & Schmid, H. (2007) Hierarchical spatial models of abundance and occurrence from imperfect survey data. *Ecological Monographs*, **77**, 465–481.
- Schmid, H., Zbinden, N. & Keller, V. (2004) *Überwachung der Bestandsentwicklung häufiger Brutvögel in der Schweiz*. Swiss Ornithological Institute, Sempach, Switzerland.
- Simons, T.R., Allredge, M.W., Pollock, K.H. & Wettroth, J.M. (2007) Experimental analysis of the auditory detection process on avian point counts. *Auk*, **124**, 986–999.

- Spiegelhalter, D., Thomas, A. & Best, N.G. (2003) *WinBUGS User Manual, Version 1.4*. MCR Biostatistics Unit, Cambridge.
- Sturtz, S., Ligges, U. & Gelman, A. (2005) R2WinBUGS: a package for running WinBUGS from R. *Journal of Statistical Software*, **12**, 1–16.
- Webster, R.A., Pollock, K.H. & Simons, T.R. (2008) Bayesian spatial modeling of data from avian point count surveys. *Journal of Agricultural, Biological and Environmental Statistics*, **13**, 121–139.
- Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego.
- Wood, S.N. (2006). *Generalized Additive Models. An Introduction with R*. Chapman and Hall/CRC, Boca Raton.

Received 6 March 2009; accepted 7 October 2009
Handling Editor: Tim Coulson

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Appendix S1. BUGS model description for the binomial mixture model with Normal random effects in the linear predictors of both abundance and detection

Appendix S2. BUGS model description for the multinomial mixture model with Normal random effects in the linear predictors of both abundance and detection. This code allows survey number to vary between 2 and 3

Appendix S3. How to treat unbalanced data for the multinomial mixture model

Appendix S4. Details on the analysis by Markov chain Monte Carlo

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.