Journal of Applied Ecology

British Ecological Society

Journal of Applied Ecology 2009, **46**, 1163–1172 doi: 10.1111/j.1365-2664.2009.01724.x

Trend estimation in populations with imperfect detection

Marc Kéry*[1], Robert M. Dorazio[2], Leo Soldaat[3], Arco van Strien[3], Annie Zuiderwijk[4] and J. Andrew Royle[5]

[1]*Swiss Ornithological Institute, 6204 Sempach, Switzerland;* [2]*U.S. Geological Survey, Florida Integrated Science Center, Department of Statistics, University of Florida, PO Box 110339, Gainesville, FLA 32611, USA;* [3]*Statistics Netherlands, PO Box 24500, 2490 HA The Hague, The Netherlands;* [4]*RAVON – Monitoring Network, PO Box 94766, 1090 GT Amsterdam, The Netherlands; and* [5]*U.S. Geological Survey, Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, MD 20708, USA*

Summary

1. Trends of animal populations are of great interest in ecology but cannot be directly observed owing to imperfect detection. Binomial mixture models use replicated counts to estimate abundance, corrected for detection, in demographically closed populations. Here, we extend these models to open populations and illustrate them using sand lizard *Lacerta agilis* counts from the national Dutch reptile monitoring scheme.

2. Our model requires replicated counts from multiple sites in each of several periods, within which population closure is assumed. Counts are described by a hierarchical generalized linear model, where the state model deals with spatio-temporal patterns in true abundance and the observation model with imperfect counts, given that true state. We used WinBUGS to fit the model to lizard counts from 208 transects with 1–10 (mean 3) replicate surveys during each spring 1994–2005.

3. Our state model for abundance contained two independent log-linear Poisson regressions on year for coastal and inland sites, and random site effects to account for unexplained heterogeneity. The observation model for detection of an individual lizard contained effects of region, survey date, temperature, observer experience and random survey effects.

4. Lizard populations increased in both regions but more steeply on the coast. Detectability increased over the first few years of the study, was greater on the coast and for the most experienced observers, and highest around 1 June. Interestingly, the population increase inland was not detectable when the observed counts were analysed without account of detectability. The proportional increase between 1994 and 2005 in total lizard abundance across all sites was estimated at 86% (95% CRI 35–151).

5. *Synthesis and applications*. Open-population binomial mixture models are attractive for studying true population dynamics while explicitly accounting for the observation process, i.e. imperfect detection. We emphasize the important conceptual benefit provided by temporal replicate observations in terms of the interpretability of animal counts.

Key-words: abundance, binomial mixture model, detectability, GLM, hierarchical model, lizard, metapopulation design, monitoring, trend

Introduction

The study of population dynamics lies at the heart of ecology (Andrewartha & Birch 1954; Krebs 2001). The simplest characterization of population dynamics is by the magnitude of a sustained population change, or trend. Population trends are extremely important in conservation biology (Caughley 1994; Norris 2004) and widely assessed in monitoring programmes (Balmford, Green & Jenkins 2003; Gregory *et al.* 2005). Time-series of counts of organisms are therefore analysed frequently to infer trends or other temporal patterns, and the factors governing them, in populations where these counts have been conducted (Meyer, Schmidt & Grossenbacher 1998; Lande, Engen & Saether 2003).

*Correspondence author. E-mail: marc.kery@vogelwarte.ch

Unfortunately, observed counts cannot usually be equated with true abundance; rather, it is useful to distinguish formally between two linked, stochastic processes that together produce the observed, raw counts. The first stochastic process is that of the true biological state, about which inference is desired; and the second is the observation process consisting of double-counting, misidentification and overlooking, i.e. imperfect detection, of individuals. Clearly, the observation process affects the way in which the true state of a population is perceived. Overlooking individuals in particular occurs in many if not most studies of natural populations. Failure to properly account for the observation process in the analysis of population count data may lead one to overestimate true variation in the biological process (Link & Nichols 1994), to underestimate abundance (Schmidt 2005) and to obtain distorted inference about patterns in the biological process such as the strength of density dependence (Shenk, White & Burnham 1998), habitat relations (Kéry 2004) or temporal trends (Hochachka & Fielder 2008).

To improve inference from population counts about patterns in the underlying biological process, such as trends, state-space (or hierarchical) models honour the hierarchical genesis of ecological field data. These models originate in the classical structural time series literature (see e.g. Soldaat *et al.* 2007 for an application) and formally partition the observed data into a stochastic state and a stochastic observation component, and may thus access features of the state, such as density-dependence, free from effects of the observation process (Lande *et al.* 2003; de Valpine 2003; Dennis *et al.* 2006). However, they can only describe, perhaps on a transformed scale such as the log, random deviations of the observations around the mean of the state *and the expectation of the observation processes combined.* Assuming detection error dominates the observation process and that detection probability is denoted p, the mean of the modelled state process at time i, N'_i, is the product of N_i and the expectation $E(p)$ of p, rather than abundance N alone. Thus, these models cannot infer absolute abundance and its dynamics unless one has additional information about the observation process, such as that provided by replicate counts (Royle 2004a), distance sampling (Buckland *et al.* 2001) or capture–recapture data (Williams, Nichols & Conroy 2002). Hence, in the presence of imperfect detection, this first kind of hierarchical model contains only an implicit description of the biological (state) process (Royle & Dorazio 2008, p. 16–17).

Inferences about the state process using such traditional state-space models will be biased by any systematic pattern present in the observation process. For instance, increasing detection probability over time might lead one to falsely conclude an increasing population trend, even when accounting for sampling variation about the product Np. Standardized field protocols may help to make p constant and therefore to prevent biased inference about apparent patterns in the state that in reality are attributable to the observation process. However, even the tightest standardization may fail to produce constant detection probability (Schmidt 2005; Brown, Hines & Kéry 2007; Hochachka & Fielder 2008). To adjust inference about population trends for distortions caused by systematic variation in detection probability, Link & Sauer (1997, 2002) developed hierarchical models that directly use information about the observation process contained in measurable covariates such as observer identity, start-up years or survey effort. Thus, an important distinction of Link & Sauer's work from hierarchical models mentioned earlier is that they incorporate information about the detection process, even if the latter is not directly modelled. However, they cannot correct for effects of detectability that are not contained in measured covariates.

Here, we develop a hierarchical model for trend estimation in the presence of imperfect detection that contains an explicit description of the state process. We build on a similar explicit hierarchical model developed by Royle & Dorazio (2008, pp. 4–7) which we extend considerably. By explicit hierarchical model we mean that the quantities contained in our state process have a clear biological meaning (Royle & Dorazio 2008); we directly model abundance (N) and its change, rather than the change of an index to abundance, such as the expected count, that is confounded with detection. Furthermore, our description of the observation process is a mechanistic representation of the detection process leading to the observed counts. Our model represents an extension of a closed-population, binomial mixture model (Royle 2004a,b; Dodd & Dorazio 2004) to open populations and is applicable when counts are available for multiple sites and multiple seasons, and when there is some degree of replication within a season, when the population is assumed closed. Our model is a hierarchical extension of a generalized linear model (GLM) and hence enjoys the flexibility of this class of models (McCullagh & Nelder 1989). We illustrate our model using counts of sand lizards *Lacerta agilis* Linnaeus 1758 from the Dutch national reptile monitoring scheme. Importantly, our analysis uncovers increases in both Dutch regions, the coast and inland, while a simple analysis of the population indices, i.e. the raw counts without accounting for detectability, fails to reveal the trend inland.

Methods

BINOMIAL MIXTURE MODELS FOR OPEN POPULATIONS

We first briefly review the simple binomial mixture model for a single season before extending it to multiple seasons. We assume that replicated counts c_{ij} are available from a number of sites (or transects, i) and surveys (j) conducted during a single season within which the population at each site is assumed to be demographically closed. We call this data format a metapopulation design, since the same quantities are surveyed at a number of spatial replicates (Royle 2004b; Royle & Dorazio 2006). Perhaps the conceptually simplest description of counts obtained in a metapopulation design is the hierarchical model that partitions the variation in c_{ij} into one component due to spatial variation in the true numbers present at each site, and observation error induced by imperfect detection. This model is described by two simple conditional probability statements:

$$\text{State process}: N_i|\lambda_i \sim \text{Poi}(\lambda_i) \qquad \text{eqn 1}$$

$$\text{Observation process}: c_{ij}|N_i \sim \text{Bin}(N_i, p_{ij}) \qquad \text{eqn 2}$$

That is, first, the spatial variation in the unobserved true state, local population size or abundance (N_i), is described by a Poisson distribution with rate parameter λ_i. Second, given a particular realization of the state process, repeated observations c_{ij} of the local population N_i follow a binomial distribution with index parameter N_i and success parameter p_{ij}. The latter is the per-individual detection probability, which is assumed to be identical for all individuals within the same site, but may vary by site i and survey j.

This basic hierarchical model first described by Royle (2004a) is a simple hierarchical extension of a familiar Poisson GLM to account for imperfect detection induced by $p_{ij} < 1$. Hence, the basic binomial mixture model is represented by two coupled GLMs, where the first (Poisson) GLM describes the only partially observed true state of the population (i.e. local population size N_i) and the second (Binomial) GLM describes the results of the observation process (c_{ij}), conditional on the true state N_i.

Here, we generalize the basic binomial mixture model to open populations where N_i is allowed to change; for instance, when counts are conducted across several seasons (years). This generalization is conceptually simple, but of great practical importance, since trends in abundance are the focus of a large and increasing number of monitoring schemes (Balmford, Green & Jenkins 2003; Gregory *et al.* 2005). To accommodate the temporal dimension, an additional index k is needed for year, resulting in c_{ijk}, the number (c) of individuals counted at site i, during survey j and in year k (the 'robust design' sampling protocol in capture–recapture studies; Williams *et al.* 2002). This introduces slight changes into the description of the basic hierarchical model (eqns 1 and 2).

The state process becomes $N_{ik} \sim \mathrm{Poi}(\lambda_{ik})$, i.e. a separate, independent Poisson intensity (λ_{ik}) is assumed for each site i and year k. Initially, this does not stipulate any relationship in the state process across years. However, a ubiquitous question in ecology, and especially in monitoring applications such as our case study, is that for sustained change, which is typically answered by fitting some sort of linear regression. Thus, to generalize the basic model to multiple years and to estimate a trend in the unobserved abundances, one may consider the following extension of the state process:

$$N_{ik} \sim \mathrm{Poi}(\lambda_{ik}) \qquad \text{eqn 3}$$

$$\log(\lambda_{ik}) = \alpha_i + r_i \times (k-1) \qquad \text{eqn 4}$$

That is, the log-transform of the Poisson rate parameter λ_{ik} in year k is the sum of an intercept parameter α_i, representing the log rate in year 1, and the effect of the annual population growth rate r_i. Both parameters are indexed also by i, indicating they may differ by site. This is the familiar log-linear Poisson regression model that forms the basis of many analyses of count data (e.g. Link & Sauer 1997, 2002; Gregory *et al.* 2005). It can easily be extended to include effects of other explanatory variables on abundance by adding additional terms such as $\beta_w \times x_{ikw}$ to the right-hand side of equation 4, where β_w represents the coefficient and x_{ikw} the value of the w-th covariate measured at site i in year k.

Similar extensions can be envisioned for the observation process which becomes:

$$c_{ijk}|N_{ik} \sim \mathrm{Bin}(N_{ik}, p_{ijk}) \qquad \text{eqn 5}$$

$$\mathrm{logit}(p_{ijk}) = \alpha_{ijk} + \beta_w \times x_{ijkw} \qquad \text{eqn 6}$$

Given a realized value N_{ik} of the state process at site i in year k, the logit transform of detection probability at site i, during survey j in year k is represented by an intercept α_{ijk} plus the effect β_w of the w-th covariate with value x_{ijkw}.

To account for unobserved sources of variation (overdispersion) in abundance or detection, random effects parameters may be added into equations 4 and 6. Random effects for abundance would necessarily be indexed by site i (and possibly also by year k), while for detection, they may vary by any combination of site i, survey j and year k. Customarily, they would be assumed to follow a zero-mean normal distribution with variance σ^2 to be estimated from the data.

CASE STUDY: NATIONAL TRENDS IN DUTCH SAND LIZARDS

The Dutch reptile monitoring scheme is a volunteer-based programme launched in 1994 by RAVON and Statistics Netherlands (Smit & Zuiderwijk 2003). It is based on transect counts replicated typically seven times within each activity season (March–October) for an arbitrary number of years, with individual surveys spaced ≥5 days. Observers are guided in choosing their transect location, but only sites are selected that are likely to contain reptiles. Transect length is about 2 km. Surveys are conducted during weather conditions favourable for the observation of reptiles, i.e. without rain, neither too cold nor too hot and no strong wind. Reptiles are counted up to 5 m on either side of a transect, and each survey lasts about 2 h. Observers must know reptiles well, but their experience is rated by the coordinator at three increasing levels (1–3). Level 1 is for novices, level 2 for observers with at least 1 year experience and level 3 for observers with at least 2–3 years' experience. On average, these levels were associated with mean sand lizard counts (see below) of 2·4, 4·0 and 7·1 respectively. Most observers count only one transect, a few do several transects on their own and some other transects are counted by two or sometimes three persons. Overall, for the 208 transects in our study (see below), about 100–150 observers are involved annually.

For each survey, observers record date, start and stop time, number of observers, temperature and wind speed. However, as not all observers recorded temperature, and because the recorded temperatures often were mere estimates, we used daily mean temperatures recorded at the meteorological station De Bilt in the centre of The Netherlands. As daily mean temperatures are based on registrations during night and day they are on average 4·0 °C below the actual temperature in the field during the observations (correlation between daily mean temperatures and field temperature $r = 0{\cdot}684$, $P = 0{\cdot}001$).

The sand lizard *Lacerta agilis* is a middle-sized (SVL 70–100 mm), brown and green lizard widespread from Western Europe into Central Asia. Preferred habitats in The Netherlands contain dry, sandy soils, southern exposed slopes and a mosaic of living and dead, herbaceous vegetation, open sandy patches and small bushes, including forest edges. It is the second most widespread Dutch reptile species, occurring widely in open, but still relatively densely vegetated areas such as the coastal dunes and inland heaths (including road- and railway-sides if southern exposed and supporting heather). Coast and inland habitat differ in many respects, such as in their vegetation, relief and aerial nitrogen load. Mating season in The Netherlands is in April–May, egg-laying in June and young of the year do not hatch until the end of July. The date and duration of seasonal activities depend very much on weather conditions, especially temperature and sunshine.

To illustrate trend estimation using the open-population binomial mixture model, we chose sand lizard counts conducted from 1994 to 2005 at all 208 sites where the species had ever been detected between 1985 and 2005. We restricted our analysis to surveys conducted in

spring (April–June), because the population was most probably closed during that period and no young had yet hatched. We note that although we model open populations we require replicate observations within a period of closure to inform estimates of detection probability.

We used the following submodels for inference about the state (population size) and the observation processes (detection probability) involved in the genesis of the observed counts of Dutch sand lizards. Since *a priori* different population trajectories were expected in the two regions where sand lizards occur in The Netherlands, the coast and inland, we specified two independent log-linear Poisson regressions as descriptions of population dynamics at coastal and inland sites. To take account of unexplained heterogeneity among sites in abundance (overdispersion), we added a normally distributed random site effect (ε_i) into equation 4. This site effect allowed for additional variation in the counts beyond that stipulated by a conditional Poisson distribution. Hence, eqn 4 became $\log(\lambda_{ik}) = \alpha_{\text{region}(i)} + r_{\text{region}(i)} \times (k-1) + \varepsilon_i$,where k indexes years as before.

As a description of the observation process giving rise to the counts (eqn 6), we fitted separate intercepts for each year, additive effects of region (to take account of probable different observability in the open dunes on the coast), observer experience, and linear and quadratic effects of both date and temperature. To allow for additional heterogeneity (overdispersion) among individual surveys in detection probability that was not captured by these explanatory variables, we fitted a normally distributed random effect that differed by site, survey and year. Thus, this overdispersion term in detection probability provided considerable flexibility to our model for detection. (See Appendix S1 for the specification of our model in the BUGS language.)

We fitted the open-population binomial mixture model in a Bayesian mode of inference. That is, we combined the model just described with prior distributions for each of its parameters to specify our prior belief about their likely magnitudes. Markov chain Monte Carlo (MCMC) methods were used to obtain an arbitrarily large sample of draws from the joint posterior distribution (Brooks 2003). We chose suitable 'vague', independent priors for all parameters to express the absence of prior information about the model parameters. Specifically, we chose uniform(−5, 5) distributions for *loglam.coast*, *beta.inland*, *beta.p.inland* as well as for all coefficients of the day, temperature and experience covariates (see Table 1 for further explanation of the meaning of these parameters), uniform(−1, 1) distributions for the two independent population growth rates (r_{coast} and r_{inland}), and uniform(0, 1) distributions for the year-specific intercepts of detection probability. Finally, we chose uniform(0, 10) distributions for the standard deviation of the two normal distributions used to account for unexplained variation in log(abundance) and logit(detection) respectively. To enhance convergence of the MCMC sampler, we standardized the day and temperature covariates and truncated both normal distributions for overdispersion effects to within the range (−20, 20); see Kéry & Royle (2009). Inspection of the marginal posterior distributions for each parameter suggested that the truncation implied by the use of uniform prior distributions did not affect our estimates. We used WinBUGS (Spiegelhalter, Thomas & Best 2003) to run three parallel chains with 500 000 iterations each, discarded the first half as burn-in and thinned the remainder by one in 150 to obtain 5001 draws of the joint posterior for inference. This led to acceptable convergence for all structural parameters [Gelman-Rubin statistic (Gelman & Hill 2007) $|\hat{R} - 1| < 0.1$]. To gauge the adequacy of the model chosen for our data set, we computed a Bayesian *P*-value for a discrepancy measure constructed from summed Pearson residuals (see Appendix S1). Its value of 0·40 suggested a good fit.

Table 1. Parameter estimates based on summaries of the marginal posterior distributions (mean, SD, 2·5% and 97·5% percentiles) under the open-population binomial mixture model fitted to Dutch sand lizard counts (1994–2005)

	Post. mean	Post. SD	95% CRI
State model			
loglam.coast	1·670	0·154	1·378, 1·970
beta.inland	1·083	0·241	0·622, 1·562
r_{coast}	0·082	0·012	0·058, 0·105
r_{inland}	0·038	0·017	0·003, 0·074
sigma.site	1·229	0·074	1·094, 1·384
Observation model			
p0[1994]	0·177	0·027	0·129, 0·234
p0[1995]	0·161	0·025	0·117, 0·211
p0[1996]	0·228	0·031	0·170, 0·290
p0[1997]	0·186	0·026	0·138, 0·241
p0[1998]	0·261	0·032	0·200, 0·324
p0[1999]	0·226	0·027	0·175, 0·283
p0[2000]	0·240	0·029	0·185, 0·301
p0[2001]	0·234	0·030	0·180, 0·297
p0[2002]	0·213	0·025	0·168, 0·264
p0[2003]	0·244	0·031	0·189, 0·308
p0[2004]	0·209	0·027	0·160, 0·264
p0[2005]	0·217	0·030	0·164, 0·280
beta.p.inland	−1·006	0·182	−1·396, −0·664
bday1	0·116	0·025	0·066, 0·165
bday2	−0·137	0·022	−0·180, −0·095
btemp1	−0·003	0·026	−0·053, 0·047
btemp2	−0·038	0·016	−0·068, −0·007
bexp2	0·183	0·109	−0·029, 0·401
bexp3	0·644	0·112	0·426, 0·870
sigma.p	0·936	0·042	0·859, 1·023

Parameters are batched into those of the state model describing structure in the expected true abundance, and those of the observation model, describing structure in detection probability. Notation: *loglam.coast*: intercept for abundance at coastal sites, *beta.inland*: coefficient for inland sites (abundance), r_{coast} and r_{inland}: independent population trends at coastal and inland sites, *sigma.site*: overdispersion SD (abundance), p0[1994]–p0[2005]: mean detection probability at coastal sites for observers of experience 1, *beta.p.inland*: coefficient for inland sites (detection), *bday1* and *bday2*: linear and quadratic effects of season, *temp1* and *temp2*: linear and quadratic effects of temperature, *bexp2* and *bexp3*: effects of observer experience class 2 and 3, *sigma.p*: overdispersion SD (detection). These parameters are the same as those shown in the model description in the BUGS language in Appendix S1.

Results

During 12 years (1994–2005) of the Dutch reptile monitoring scheme, a total of 3822 sand lizard surveys were conducted at 132 coastal and 76 inland sites. The resulting data set was fairly imbalanced with counts available from only 51% of all site-year combinations. Individual sites were surveyed for 1–12 (mean 6·2) years. Individual lizard counts ranged 0–61 (median 3), annual mean counts per site 0–41 (median 3) and annual maximum counts per site 0–61 (median 5).

The observed average count per site and survey increased over the years in the coastal but not in the inland region (Fig. 1a). A conventional Poisson regression of mean counts

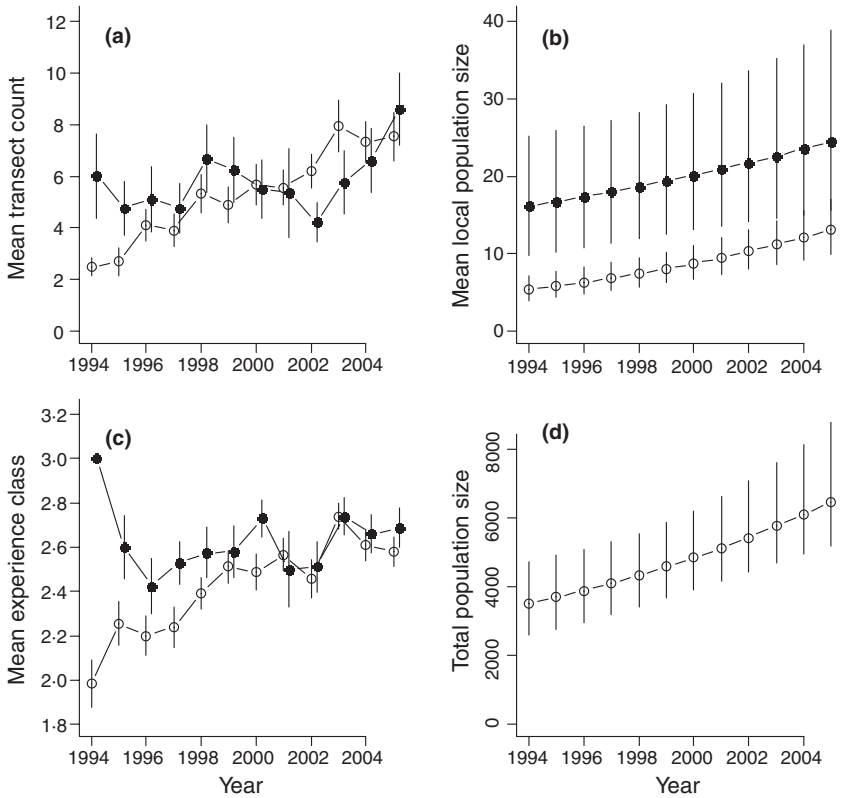

Fig. 1. Comparison of Dutch coastal and inland sites in terms of (a) mean observed counts of sand lizards *Lacerta agilis* per transect (with 1 SE), (b) trends in mean local population size corrected for detection probability as estimated under a binomial mixture model, (c) mean experience class of observers (with 1 SE), and (d) estimated total population size across all 208 surveyed sites estimated under our model. Open symbols, coastal sites; solid symbols, inland sites. Posterior means and 95% CRI are shown in (b) and (d).

on year yielded a significant trend estimate at coastal sites only ($\hat{r}_{\text{coast}} = 0{\cdot}092$, $z = 2{\cdot}46$, $P = 0{\cdot}01$; $\hat{r}_{\text{inland}} = 0{\cdot}025$, $z = 0{\cdot}73$, $P = 0{\cdot}46$).

However, estimates of abundance obtained using our hierarchical model to correct for imperfect and variable detection indicated that sand lizards increased in both regions, although more so on the coast ($\hat{r}_{\text{coast}} = 0.083$) than inland (Table 1 'State model': $\hat{r}_{\text{inland}} = 0.038$; Fig. 1b). We estimated the regional difference in population trends at 0·044 (95% CRI 0·019–0·068). In addition, mean local population size at inland sites was estimated about three times higher than for coastal sites (*beta.inland* = 1·083; Fig. 1b), although average counts were fairly similar in both regions (Fig. 1a). Relative to a Poisson distribution, true lizard abundances were overdispersed (*sigma.site* = 1·229).

That lizards inland were also found to be increasing may be partly because of different trends in average observer experience at coastal and inland sites (Fig. 1c). Average experience increased from 2 to 2·6 at coastal sites, but decreased from 3 to 2·6 at inland sites between 1994 and 2005. On average, observers in experience classes 2 and 3 counted 4·0 and 7·1 sand lizards, respectively, which was reflected also in the estimated detection probability for each experience class (Fig. 2a–b).

We found effects of region, year, season and observer experience on, as well as unexplained survey-specific heterogeneity in, detection probability of Dutch sand lizards (Table 1 'Observation model'). Lizards were less easily observed inland (Table 1, *beta.p.inland* = −1·006; Fig. 2). Observers in experience class 1 and 2 were not statistically distinguishable (Fig. 2a–b; Table 1, *bexp2*), but observers in classes 1 and 3 were (*bexp3*), as were those in class 2 and 3 (95% CRI for *bexp2-bexp3*: 0·325–0·593). Moreover there was a suggestion of an increased detection probability over the first few years of the survey (Fig. 2a–b). Detection of an individual sand lizard was easiest around day 150 (~June 1; Fig. 2c) and at national daily mean temperatures around 15 °C, corresponding to about 19 °C in the field (Fig. 2d), although the latter effect was not significant (Table 1). In addition, there was heterogeneity in detection probability that could not be ascribed to any recognized factor and was subsumed into survey-specific 'overdispersion' effects (Table 1, *sigma.p* = 0·936).

Our modelling framework enables site-specific estimates of the true population trajectories to be obtained (see Appendix S2). Noteworthy is the smoothing of the predicted population trajectory compared with the more ragged observed mean counts. Another advantage of the model-based integration of information across sites is the ability to obtain estimates of abundance even for years where a site was not surveyed. Missing values are automatically imputed (estimated) as part of the updating via MCMC, with full accounting for all components of uncertainty involved. As a result, population totals can easily be estimated for subsets of, or for all surveyed sites with full accounting for imperfect detection as well as any imbalance in the data set. For instance, total abundance across all sampled sites in 1994 was estimated at 3517 (95% CRI 2582–4715) and in 2005 at 6449 (95% CRI 5174–8772) sand lizards (Fig. 1d). This increase was estimated at 2932 (95% CRI 1467–5035), corresponding to a proportional population increase of 86% (95% CRI 35–151).

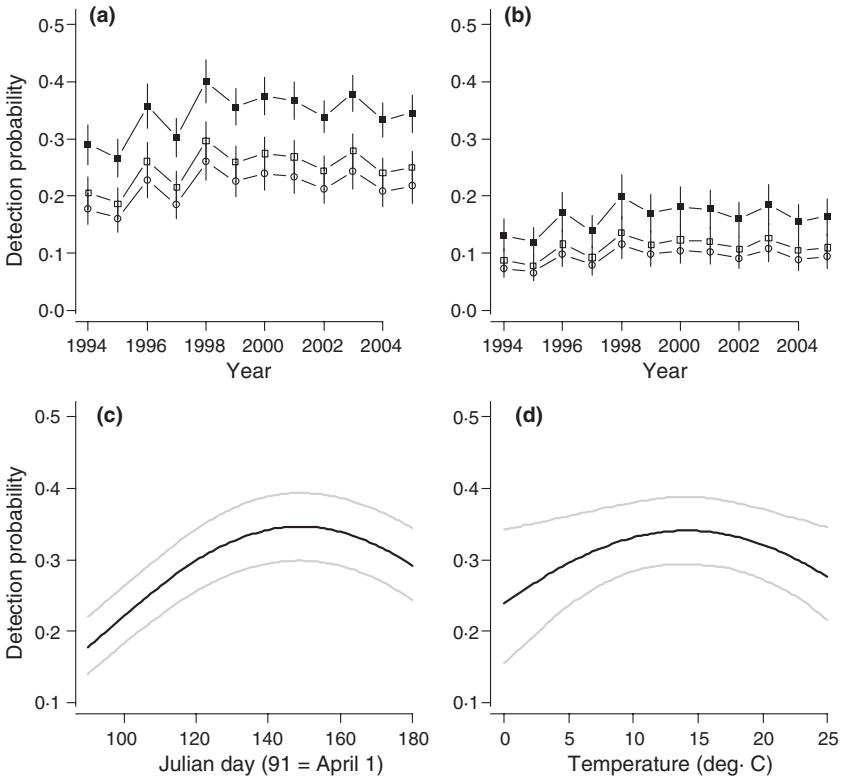

Fig. 2. Detection probability (posterior means and SD) of Dutch sand lizards in relation to year, region and observer experience (experience 1: open circle, 2: open square, 3: solid square) (a) at coastal and (b) at inland sites, and in relation to (c) season (Julian day) and (d) national daily average temperature (°C). Predictions in (c) and (d) were formed by averaging over year, region and experience effects and show the posterior mean and 95% CRI.

Discussion

OPEN-POPULATION BINOMIAL MIXTURE MODELS

We describe a simple, but important extension to open populations of binomial mixture models (Royle 2004a) to estimate trends of abundance, corrected for overlooked individuals. Our model is a hierarchical model as the process generating the observed counts is partitioned into two stochastic subprocesses, the un- or only partly observed state (or biological) process, and the observation process that depends on the particular realization of the state process. As the quantities in our model have an explicit biological meaning, our model is an explicit, or mechanistic, hierarchical model (Royle & Dorazio 2008), in contrast to the mechanistically more implicit hierarchical models of Link & Sauer (2002), Lande *et al.* (2003), Dennis *et al.* (2006), Ver Hoef & Jansen (2007), Cressie *et al.* (2009) and many others. These latter models confound the product of the state and the observation process, i.e. describe the product of N and $E(p)$ in our equations 1 and 2 as their state process, and simple sampling variation around that product as their observation process.

A log-linear trend as in our example is the simplest description of population dynamics. For the most typical question in monitoring studies, 'are things getting better or worse?' (Jenkins, Green & Madden 2003), this is probably sufficient. However, explicitly incorporating more biological realism into the description of interannual changes in abundance by honouring demographic relationships may yield better inference than a purely phenomenological population dynamics model such as ours (de Valpine 2003; Mazzetta, Brooks & Freeman 2007; Brooks *et al.* 2008). Hence, a useful further development would be to specify a population dynamics model that incorporates density-dependence into the state submodel, such as the Ricker, Gompertz or theta-logistic equations (Saether, Engen & Matthysen 2003; Dennis *et al.* 2006), i.e. to embed an explicit population dynamics model into an estimation approach that corrects for bias introduced by imperfect detection and factors associated with the observation process (Buckland *et al.* 2007). This would be interesting for the two reasons of providing better biological inference, e.g. about the prevalence of density-dependence, and for improved population diagnosis, e.g. to detect a decline (de Valpine 2003). Finally, a yet more exciting possible future development of our modelling framework would be to express interannual change directly using demographic rates of survival and fecundity, i.e. by imposing a Leslie matrix on the dynamics of the state process (e.g. Besbeas *et al.* 2002; Buckland *et al.* 2007; Schaub *et al.* 2007). However, this would probably require the injection of additional demographic information, for instance about fecundity or survival, into the analysis or else some parameters would not be identifiable. Alternatively, strongly informative prior distributions may be adopted. Given suitable data, i.e. short term, replicate observations, these and further generalizations can all be cast within the framework of an open-population binomial mixture model.

In contrast to simpler analyses that do not separate the state and observation processes when modelling animal counts, our model explicitly specifies trends in abundance, rather than expected counts, by correcting for the average level as well as any modelled non-random patterns in detection probability. However, in a sense, we still only have an index of abundance,

as the effective sampling area associated with each transect is not known. This would be the case even when the transect lengths were known. How to estimate the effective sampling area relevant for a density estimate is a widespread but under-appreciated challenge in animal surveys. Distance sampling (Buckland *et al.* 2001) or spatially explicit capture–recapture models (Borchers & Efford 2008; Royle & Dorazio 2008, chaps 6–7) are needed when absolute density is required. In addition, we expect individuals to move on and off the transects between sample occasions, producing an effect similar to that of temporary emigration (Kendall, Nichols & Hines 1997). In the case of our model, the mean of the abundance distribution applies to the super-population of individuals that are ever exposed to sampling. Explicitly accounting for movement-induced bias of this sort again requires additional information in the form of spatial encounter history data.

Although arguably a maximum likelihood implementation of open-population binomial mixture models would be possible and might be advantageous in terms of model selection via AIC or deviance statistics, we think that the advantages of the Bayesian approach outweigh those of an ML approach. For instance, the ease with which the Bayesian approach handles error propagation for functions of parameters is compelling. Thus, the uncertainty around the estimated difference of the total lizard population sizes in 1994 and 2005 is obtained trivially from the MCMC output of a Bayesian analysis. In addition, we think that the average biologist will feel more comfortable with the BUGS model code for an open-population binomial mixture model than with, say, R code that maximizes the likelihood for the same model.

Predictions of population trajectories for transects with missing years (i.e. where there are no counts in some years; see examples in Appendix S2) provide one particularly striking example of the flexibility of a Bayesian analysis. The imputation of such missing values given the model and the available data happens naturally and almost as a byproduct. Indeed, estimation of a site-specific population trajectory is possible even for sites with counts available in a single year only. To the degree that observations from other sites are, indeed, informative about that site, these predictions form the best possible interpretation of the data at that site with respect to population trends. We note that applied biologists often do much the same in an informal way. For instance, given a few observations from a site on the coast, and knowing that coastal sites generally support increasing sand lizard numbers, a biologist might informally draw similar conclusions about a site with sparse data. The difference with our approach is that an explicit model does this in a completely transparent way and with full accounting for all modelled components of uncertainty, much in contrast to an informal interpretation of the same sparse data.

CASE STUDY: DUTCH SAND LIZARDS

Our most striking finding was the discrepancy in the inference about population levels and trends based on raw counts (Fig. 1a) and under our model, which corrects for imperfect detection (Fig. 1b). We estimated sand lizards to be two to three times more numerous at inland sites than at coastal sites, despite similar average raw counts. In addition, after accounting for detection probability, sand lizards were seen to increase also at inland sites, even though no significant trend was discernible at inland sites in the mean observed counts. These findings should provide ample motivation to account for imperfect detectability when making inferences about population trends.

Not surprisingly, the site random effect for abundance introduced in eqn 4 turned out to be highly significant. This effect probably included two components of variation in abundance that were otherwise unmodelled in our analysis: true habitat differences in density and variation in the size of the sampled area owing to the lack of transect standardization. We believe that it will frequently be unrealistic to assume that the entire variation in density among sites or over years is adequately captured by a Poisson distribution alone and that it is important to model part of extra-Poisson variation either explicitly (using covariates) or implicitly (by site random effects). Alternatives would include the adoption of a negative-binomial distribution for abundance (Royle 2004a; but see Joseph *et al.* 2009) or a 'non-parametric', data-based representation of the variation in density among sites (see Dorazio *et al.* 2008).

In spite of the number of factors affecting detection probability of Dutch sand lizards, it may be interesting to compare some directly observable population statistic with the estimated local abundance. The mean daily sand lizard count at a site averaged 21% (range 0–71%) of the estimated local population size, and the maximum daily sand lizard count averaged 34% (range 0–95%). This again stresses the fact that detectability was far from perfect, something particularly striking in reptiles (Kéry 2002; Brown *et al.* 2007; Kéry & Schmidt 2008).

As for density, there was clear evidence for random effects on detection probability. That is, there was temporal variation in conditions that affected detection beyond what was captured by the fixed covariates in the model, year, date, temperature and observer experience. Again, we believe that it may be advantageous to account for this additional source of variation as the assumption of constant probability of detection, even conditional on the values of measured covariates, may be overly restrictive. Analyses not shown here suggest that unmodelled temporal variation in detection yields a positive bias in abundance. However, the ability to model such diffuse variation in detection probability may well be limited by the number of temporal replicates, which were important in the Dutch survey. These topics deserve to be investigated further by simulation.

There appears little data with which to compare our estimates. A Dutch capture–recapture study conducted in 2001 used photo ID to estimate daily detection probabilities (as a ratio of the average daily numbers seen and total estimated population) at 0·10 in an inland heath and 0·27 and 0·16 in two coastal dunes (A. Boere, unpublished data), fairly similar to what we find here (Fig. 2a–b). A Swiss capture–recapture study conducted in 2007 (C. Berney, unpublished data) estimated mean first-capture probability at 0·18 ($n = 6$ occasions,

range 0·06–0·28), again broadly comparable with estimates from our study. That study found a very large negative trap response, which illustrates one big advantage of the field protocol underlying our model: by not having to actually catch the animals, the binomial mixture model permits much less invasive sampling than does capture–recapture.

There may have been a hint of a startup effect in the Dutch reptile monitoring in Fig. 2a–b; detection probability appeared to be lower during surveys in the first 4 years and only thereafter seemed to stabilize. This may be an example of the usefulness of explicitly accounting for detection in a model for trend estimation, since uncorrected for, this pattern would exaggerate population increase. Detection was easiest during mid to late spring; consistent with conventional wisdom that the species is most visible during the mating season. Similarly, there was a tendency for detection probability to be higher at medium daily temperatures that was not quite significant. However, this is unsurprising since we could only use the same national daily average for all sites. This effect suggests that lizards remain more inactive at lower temperatures and behave more concealed at higher ones. Male visibility is higher in May during mating and lower in April and after mid-June. Basking females are best seen when pregnant between late May and egg-laying in late June.

STUDY DESIGN

It is hard to give specific recommendations on study design when analysing counts using a binomial mixture model, but we offer some general advice. The best estimates in terms of precision and lack of bias will clearly be achieved with a large number of sites (R) and within-season replicates (T), with high detection probability (p) and with as little, especially unexplained, variation in the system as possible (in the sense of covariate and especially overdispersion effects that need to be modelled). In addition, there is some interaction between these quantities: for instance, lower p may partly be compensated with higher T and less unexplained variation with smaller values of R, T or p. So how small is still enough?

Even with good-quality data, it may make little sense to fit the model with $R < 20$. It is not necessary for every site to be surveyed more than once within a season, but if the average within-season replication (T) is much < 2, biased estimates of abundance will result (unpublished simulations). On the other hand, $T > 5$ is seldom warranted and may make the within-season closure assumption suspect. With regard to detection, it may be questionable to fit the model with about $p < 0.1$. However, these can hardly be more than extremely simplistic rules of thumb, and much more needs to be learnt about the performance of the binomial mixture model in 'marginal data situations'. Much of this can be usefully studied using simulation. For instance, MacKenzie & Royle (2005) investigated design issues for related occupancy models using simulation. The binomial mixture model shares the basic structure with this class of occupancy models, which suggests that R should be emphasized much more than T to improve precision. Indeed, simulations can quite easily be tailored to the particular study design and species at hand, to ensure that an intended survey design can, indeed, yield inference of the desired quality.

Although our analysis was able to accommodate much nuisance variation and distortion in the main quantity of interest (here, abundance and its trend) by allowing for 'overdispersion', it is much better to eliminate such variation from the outset. Thus, the magnitude of the random site effect on abundance would surely have been reduced had the transect lengths been properly standardized, and this would have made the uncertainty around the trend estimates smaller. Alternatively, actual transect length could be measured and used as an offset in the analysis and this would probably increase precision of the trend estimates again. Similarly, for the observation part of the model, survey duration and perhaps average vegetation height appear promising variables to explain some of the variation in detection probability.

Finally, our model assumes that individuals are detected independently, so it may not be applicable to aggregated populations (e.g. animals in flocks). Some lack of fit of a Poisson distribution in this case may be accounted for by introduction of an overdispersion correction in the linear predictor of abundance, as in our analysis, or by adoption of a Negative Binomial distribution. If most animals are distributed in groups, then the number of groups might be modelled instead of individuals and group size could be used as a detection covariate.

A final issue to consider is the optimal size of the spatial sample. When that is too small, there will be movement onto and off the sample unit, which violates the closure assumption and will lead to a positive bias in abundance estimates. On the other hand, when the spatial sample is too large, the assumption that all individuals within it have the same detection probability may be violated, which we believe would lead to a negative bias in abundance estimates. Thus, the choice of scale of the spatial sampling unit is an important consideration, but one where it is difficult to give a general advice.

Conclusion

Trends, or more generally, temporal patterns in the abundance of populations, are the focus of a vast number of ecological studies and are crucial to ecological applications such as monitoring schemes. Alas, the interpretation of trends from raw counts is always complicated by imperfect detection and especially the possibility of concomitant trends in detection, even in strongly standardized schemes. This may bias inference unless detection probability is explicitly accounted for. If counts are available over multiple sites and years, and if there are within-year replicates, the open-population binomial mixture model offers a flexible analysis framework that explicitly corrects for variable levels of imperfect detection. Finally, we believe that the advantages of field protocols using repeat observations within a short-time period are hard to overstate and we would hope that they entice an increasing number of new studies to apply this method of data collection.

Acknowledgements

The Dutch reptile monitoring is a joint scheme of RAVON and Statistics Netherlands financed by the Dutch Ministry of Agriculture, Nature and Food Quality. We thank all volunteers for providing such a rich data set to serve as a case study. We furthermore thank F. Korner, M. Schaub, B.R. Schmidt and two referees for valuable comments.

References

Andrewartha, H.G. & Birch, L.C. (1954) *The Distribution and Abundance of Animals*. University of Chicago Press, Chicago, IL.

Balmford, A., Green, R.E. & Jenkins, M. (2003) Measuring the changing state of nature. *Trends in Ecology and Evolution*, **18**, 326–330.

Besbeas, P., Freeman, S.N., Morgan, B.J.T. & Catchpole, E.A. (2002) Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, **58**, 540–547.

Borchers, D.L. & Efford, M.G. (2008) Spatially explicit maximum likelihood methods for capture–recapture. *Biometrics*, **64**, 377–385.

Brooks, S.P. (2003) Bayesian computation: a statistical revolution. *Philosophical Transactions of Royal Society of London A*, **361**, 2681–2697.

Brooks, S.P., Freeman, S.N., Greenwood, J.J.D., King, R. & Mazzetta, C. (2008) Quantifying conservation concern – Bayesian statistics, birds and the red lists. *Biological Conservation*, **141**, 1436–1441.

Brown, W.S., Hines, J.E. & Kéry, M. (2007) Survival of timber rattlesnakes (*Crotalus horridus*) estimated by capture–recapture models in relation to age, sex, color morph, time, and birthplace. *Copeia*, **2007**, 656–671.

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.

Buckland, S.T., Newman, K.B., Fernandez, C., Thomas, L. & Harwood, J. (2007) Embedding population dynamics models in inference. *Statistical Science*, **22**, 44–58.

Caughley, G. (1994) Directions in conservation biology. *Journal of Animal Ecology*, **63**, 215–244.

Cressie, N., Calder, C.A., Clark, J.S., Ver Hoef, J.M. & Wikle, C.K. (2009) Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications*, **19**, 553–570.

Dennis, B., Ponciano, J.M., Lele, S.R., Taper, M.L. & Staples, D.F. (2006) Estimating density dependence, process noise, and observation error. *Ecological Monographs*, **76**, 323–341.

Dodd, C.K. Jr & Dorazio, R.M. (2004) Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica*, **60**, 468–478.

Dorazio, R.M., Mukherjee, B., Zhang, L., Ghosh, M., Jelks, H.L. & Jordan, F. (2008) Modeling unobserved sources of heterogeneity in animal abundance using a Dirichlet process prior. *Biometrics*, **64**, 635–644.

Gelman, A. & Hill, J. (2007) *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, Cambridge.

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*, **360**, 269–288.

Hochachka, W.M. & Fielder, W. (2008) Trends in trappability and stop-over duration can confound interpretations of population trajectories from long-term migration ringing sites. *Journal of Ornithology*, **149**, 375–391.

Jenkins, M., Green, R.E. & Madden, J. (2003) The challenge of measuring global change in wild birds: are things getting better or worse? *Conservation Biology*, **17**, 20–23.

Joseph, L.N., Elkin, C., Martin, T.G. & Possingham, H.P. (2009) Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. *Ecological Applications*, **19**, 631–642.

Kendall, W.L., Nichols, J.D. & Hines, J.E. (1997) Estimating temporary emigration using capture–recapture data with Pollock's robust design. *Ecology*, **78**, 563–578.

Kéry, M. (2002) Inferring the absence of a species – a case study of snakes. *Journal of Wildlife Management*, **66**, 330–338.

Kéry, M. (2004) Extinction rate estimates for plant populations in revisitation studies: the importance of detectability. *Conservation Biology*, **18**, 570–574.

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*. (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 639–656. Series: Environmental and Ecological Statistics, Vol. 3, Springer, New York, NY.

Kéry, M. & Schmidt, B.R. (2008) Imperfect detection and its consequences for monitoring for conservation. *Community Ecology*, **9**, 207–216.

Krebs, C.J. (2001) *Ecology: The Experimental Analysis of Distribution and Abundance*, 5th edn. Benjamin Cummings, Menlo Park, CA.

Lande, R., Engen, S. & Saether, B.-E. (2003) *Stochastic Population Dynamics in Ecology and Conservation*. Oxford University Press, Oxford, UK.

Link, W.A. & Nichols, J.D. (1994) On the importance of sampling variance to investigations of temporal variation in animal population size. *Oikos*, **69**, 539–544.

Link, W.A. & Sauer, J.R. (1997) Estimation of population trajectories from count data. *Biometrics*, **53**, 488–497.

Link, W.A. & Sauer, J.R. (2002) A hierarchical analysis of population change with application to Cerulean warblers. *Ecology*, **83**, 2832–2840.

MacKenzie, D.I. & Royle, J.A. (2005) Designing occupancy studies: general advice and allocating survey effort. *Journal of Applied Ecology*, **42**, 105–1114.

Mazzetta, C., Brooks, S. & Freeman, S.N. (2007) On smoothing trends in population index modeling. *Biometrics*, **63**, 1007–1014.

McCullagh, P. & Nelder, J.A. (1989) *Generalized Linear Models*, 2nd edn. Chapman and Hall/CRC, London.

Meyer, A.H., Schmidt, B.R. & Grossenbacher, K. (1998) Analysis of three amphibian populations with quarter-century long time-series. *Proceedings of the Royal Society of London, Series B-Biological Sciences*, **265**, 523–528.

Norris, K. (2004) Managing threatened species: the ecological toolbox, evolutionary theory, and declining-population paradigm. *Journal of Applied Ecology*, **41**, 413–426.

Royle, J.A. (2004a) N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, **60**, 08–115.

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*. Academic Press, Amsterdam.

Saether, B.-E., Engen, S. & Matthysen, E. (2003) Demographic characteristics and population dynamical patterns of solitary birds. *Science*, **295**, 2070–2073.

Schaub, M., Gimenez, O., Sierro, A. & Arlettaz, R. (2007) Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, **21**, 945–955.

Schmidt, B.R. (2005) Monitoring the distribution of pond-breeding amphibians when species are detected imperfectly. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **15**, 681–692.

Shenk, T.M., White, G.C. & Burnham, K.P. (1998) Sampling-variance effects on detecting density dependence from temporal trends in natural populations. *Ecological Monographs*, **68**, 445–463.

Smit, G. & Zuiderwijk, A. (2003) *Handleiding Voor het Tellen van Reptielen in Nederland (Manual for Counting Reptiles in The Netherlands)*. RAVON and Statistics Netherlands, Amsterdam.

Soldaat, L., Visser, H., van Roomen, M. & van Strien, A. (2007) Smoothing and trend detection in waterbird monitoring data using structural time-series analysis and the Kalman filter. *Journal of Ornithology*, **148**(Suppl. 2), S351–S357.

Spiegelhalter, D., Thomas, A. & Best, N.G. (2003) *WinBUGS User Manual, Version 1.4*. MCR Biostatistics Unit, Cambridge.

de Valpine, P. (2003) Better inference from population-dynamics experiments using Monte Carlo state-space likelihood methods. *Ecology*, **84**, 3064–3077.

Ver Hoef, J.M. & Jansen, J.K. (2007) Space-time zero-inflated count models of Harbour seals. *Environmetrics*, **18**, 697–712.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego.

Received 14 April 2009; accepted 14 September 2009
Handling Editor: Phil Stephens

Supporting Information

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

Appendix S1. BUGS model specification for the open-population binomial mixture model fitted to the Dutch sand lizards as described in the article by Kéry *et al.* 2009.

Appendix S2. Estimates of site-specific abundance (posterior means and SDs) of Dutch sand lizards under the open-population binomial mixture model (black circles) and comparison with mean observed counts for a selection of sites (blue squares; inset: site number and region).

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials may be reorganized 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.