

WAS IT THERE? DEALING WITH IMPERFECT DETECTION FOR SPECIES PRESENCE/ABSENCE DATA[†]

DARRYL I. MACKENZIE¹

Proteus Wildlife Research Consultants

Summary

Species presence/absence surveys are commonly used in monitoring programs, metapopulation studies and habitat modelling, yet they can never be used to confirm that a species is absent from a location. Was the species there but not detected, or was the species genuinely absent? Not accounting for imperfect detection of the species leads to misleading conclusions about the status of the population under study. Here some recent modelling developments are reviewed that explicitly allow for the detection process, enabling unbiased estimation of occupancy, colonization and local extinction probabilities. The methods are illustrated with a simple analysis of presence/absence data collected on larvae and metamorphs of tiger salamander (*Ambystoma tigrinum*) in 2000 and 2001 from Minnesota farm ponds, which highlights that misleading conclusions can result from naïve analyses that do not explicitly account for imperfect detection.

Key words: *Ambystoma tigrinum*; colonization; detection; local extinction; occupancy; presence/absence; tiger salamander.

1. Introduction

Presence/absence surveys for species are commonly used in a wide variety of ecological applications. In large-scale monitoring programs it is often infeasible to collect sufficient information from the population to enable reasonable estimates of abundance or density to be obtained; the costs may be prohibitive. Alternatively, the proportion of area occupied (PAO) may be seen as a low-cost surrogate, assuming that this is an acceptable metric for management/conservation purposes. Presence/absence surveys can be conducted at a number of sites across a broad landscape, with records being kept of the number of locations where the species is detected. In metapopulation studies, data collected on the presence/absence of a species at discrete population ‘patches’ are often used to derive colonization and local extinction probabilities (Hanski, 1992, 1994, 1997; Moilanen, 1999, 2002). Other quantities such as turnover rate can also be calculated, and relationships between population persistence and patch characteristics (such as patch size or habitat quality) are often postulated. In habitat modelling the intent is often to build a model relating species distribution to certain habitat characteristics, then infer whether the species exhibits particular habitat preferences, or identify other areas of potentially suitable habitat.

However, in all such applications an observed species absence (or more correctly, the non-detection of the species), does not imply the species is genuinely absent from a sampling location. Frequently, a species can be present at an area but go undetected due to random

Received February 2003; revised September 2003; accepted September 2003.

¹ Proteus Wildlife Research Consultants, PO Box 5193, Dunedin, New Zealand. e-mail: Darryl@proteus.co.nz

[†] This paper was presented at the 4th Conference on Statistics in Ecology and Environmental Monitoring, ‘Population dynamics: the interface between models and data’, 9–12 December 2002, University of Otago, Dunedin, New Zealand.

chance. Such ‘false absences’ lead to incorrect inferences if the imperfect detection of the species is not accounted for: PAO is underestimated; colonization and local extinction rates are biased; and habitat relationships may well be misleading, particularly if detectability also changes with different habitats (e.g. open forest vs scrubby undergrowth). Some may argue that they are not interested in absolute measures, and that a relative index suffices; hence detectability is not an issue. This argument is only valid if detectability is always constant, and experience suggests that is not the case, especially given the dynamic nature of many ecological systems, particularly those impacted upon by humans in some form.

This paper reviews some recently developed methods that explicitly incorporate imperfect detection into the modelling process, allowing unbiased estimation of the parameters of interest. A requirement of these methods is that some locations are surveyed on more than one occasion within a relatively short period of time. Below, the models are briefly detailed and implications on the study designs are discussed.

2. Practical sampling situation

Presence/absence surveys are conducted at N sampling sites within the area of interest. These sites may represent discrete patches of habitat, such as ponds or vegetation patches; potential nesting sites; established monitoring stations; or some randomly chosen points from a map. Naturally, the methods used to select the sites directly relate to the generality of the results of subsequent modelling. The target species is detected or not detected during each survey if present, and is never falsely detected when absent (through species misidentification).

Further, it is biologically reasonable to assume that for some appropriate time interval (a ‘season’) the area is closed to any changes with respect to occupancy of the sampling sites, i.e. within a season sites are either always occupied or always unoccupied by the species (this may be relaxed in some circumstances, see the section on model assumptions for details). Changes in the occupancy state of sites may occur between seasons. Multiple presence/absence surveys are conducted at each site within each season, with the series of detections and non-detections for each site being recorded as a sequence of 1s and 0s respectively: a detection history.

3. A single season model

There have been a number of different approaches to the problem of estimating the fraction of sites occupied by a species that is imperfectly detected (Giessler & Fuller, 1987; Azuma, Baldwin & Noon, 1990; MacKenzie *et al.*, 2002; Tyre *et al.*, 2003), although here the method of MacKenzie *et al.* (2002) is reviewed as it allows for the simultaneous estimation of occupancy and detectability, and associated variances and covariances. The independently developed methods of Tyre *et al.* (2003) are closely related, but not as flexible.

MacKenzie *et al.* (2002) describe a method that allows for unbiased estimation of the proportion of area occupied by a species in a single season, in scenarios where the species cannot always be detected with certainty. They use straightforward probabilistic arguments to describe the underlying processes that may have caused a given detection history to be observed. They do so by defining the following parameters. Let ψ denote the probability a site is occupied by the species (constant across all sites), and p_j denote the probability of detecting the species (given presence) in the j th survey of a site ($q_j = 1 - p_j$). For example,

consider the detection history $H_i = 101$ (denoting site i was surveyed three times, with the species being detected in the first and third surveys). Clearly the species must be present; it is detected in survey 1, not detected in survey 2, and detected in survey 3. The probability of this history would therefore be

$$\Pr(H_i = 101) = \psi p_1 q_2 p_3.$$

A similar expression can be obtained for all detection histories where the species is detected in at least one survey, as there is only one possible process that generates an observed history (insofar as we have defined them here). For sites where the species is never detected ($H_i = 000$, for example) there are two possibilities however. Either the species is present but never detected (a ‘false absence’), or the species is genuinely absent from the site. In terms of the model parameters, this can be written as

$$\Pr(H_i = 000) = \psi q_1 q_2 q_3 + (1 - \psi),$$

where the first term is the probability of a false absence, and the second term is the probability of the site being unoccupied.

By deriving such an expression for each of the N observed detection histories, assuming independent observations, the likelihood for the data is

$$L(\psi, \mathbf{p} \mid H_1, H_2, \dots, H_N) = \prod_{i=1}^N \Pr(H_i), \quad (1)$$

where \mathbf{p} is a vector of detection probabilities.

4. A multiple season model

MacKenzie *et al.* (2003) extend the above single season model to multiple seasons by introducing two additional parameters representing the processes of colonization (γ_t) and local extinction (ε_t). These parameters are defined as follows: γ_t is the probability an unoccupied site becomes occupied by the species between seasons t and $t + 1$; and ε_t is the probability a site that was occupied by the species in season t is unoccupied in $t + 1$. As in the single season model, the parameters are used to express the probability of observing any given detection history, by considering the underlying dynamic processes. For example, consider the detection history $H_i = 101\ 000\ 001$, denoting the sequence of detections and non-detections over three seasons, with three surveys per season. The species is detected twice in the first season, never in the second season and once in the third. Assuming closure of the sites within seasons, one of two distinct processes may have occurred. Either (i) the species has become locally extinct at the site between seasons 1 and 2 (so is not there to be detected in the second season), then recolonizes the site between seasons 2 and 3, or (ii) the species has not become locally extinct between seasons 1 and 2, has gone undetected in the second season, and has continued to occupy the site in the third season. Using the parameters defined above, the probability of this occurring could be expressed as

$$\Pr(H_i = 101\ 000\ 001) = (\psi_1 p_{11} q_{12} p_{13}) (\varepsilon_1 \gamma_2 + (1 - \varepsilon_1) q_{21} q_{22} q_{23} (1 - \varepsilon_2)) (q_{31} q_{32} p_{33}),$$

where the term $\varepsilon_1 \gamma_2 + (1 - \varepsilon_1) q_{21} q_{22} q_{23} (1 - \varepsilon_2)$ represents the two possible situations for the site in the second season as described above.

Often it is more convenient to consider the model using matrix notation, as the number of possible explanations for a given detection history can be large, especially when the data have been collected over a large number of seasons. Therefore, let ϕ_t ($t \geq 1$) denote a matrix that determines the probability of a site transitioning between occupancy states, between seasons t and $t + 1$, e.g.

$$\phi_t = \begin{bmatrix} 1 - \varepsilon_t & \varepsilon_t \\ \gamma_t & 1 - \gamma_t \end{bmatrix},$$

and let ϕ_0 denote a row vector that determines the initial occupancy state immediately before the first survey of the first season, e.g.

$$\phi_0 = [\psi_1 \quad 1 - \psi_1].$$

In each case, columns represent the occupancy state in season $t + 1$ (1 = occupied, 2 = unoccupied), and rows represent the current occupancy state in season t . Also, let p_{Ht} denote a column vector, which contains the probability of observing the given sequence of detections and non-detections in season t , conditional upon occupancy state. For example,

$$p_{001,t} = \begin{bmatrix} q_{t1}q_{t2}p_{t3} \\ 0 \end{bmatrix}, \quad \text{and} \quad p_{000,t} = \begin{bmatrix} q_{t1}q_{t2}q_{t3} \\ 1 \end{bmatrix}.$$

Whenever the species is detected at least once during a season, the second element of p_{Ht} is always 0 as it is impossible to observe such a history if the site is in the unoccupied state. Similarly, given the site is unoccupied the species is never detected at the site, and hence the second element of $p_{000,t}$ is always 1.

The probability of observing the detection history H_t can then be expressed as

$$\Pr(H_t) = \phi_0 \prod_{t=1}^{T-1} D(p_{Ht}) \phi_t p_{HT},$$

where T is the total number of seasons and $D(p_{Ht})$ is a diagonal matrix with the elements of p_{Ht} along the main diagonal (top left to bottom right), zero otherwise. The model likelihood is then defined as the product of all such terms for all sites, similar to (1).

The seasonal occupancy probability can be calculated using the relationship

$$\psi_t = \psi_{t-1}(1 - \varepsilon_{t-1}) + (1 - \psi_{t-1})\gamma_{t-1},$$

and in some instances it may be advantageous to reparameterize the model in terms of ψ_t so that the seasonal occupancy probability can be modelled in terms of covariates directly. Similarly, sometimes the rate of change in occupancy may be of interest (in fact in some monitoring situations it may be of primary interest); it can be defined as

$$\lambda_t = \frac{\psi_{t+1}}{\psi_t}.$$

Again, the rate of change in occupancy may be derived or the model reparameterized so that it can be estimated directly.

Under an assumption of perfect detection (as in a naïve approach), ε_t could be estimated as the fraction of sites apparently occupied in season t that appear to be unoccupied in $t + 1$, and γ_t could be estimated as the fraction of sites that appeared unoccupied in season t which appear occupied in $t + 1$.

5. Estimation and modelling

MacKenzie *et al.* (2002) and MacKenzie *et al.* (2003) use standard maximum likelihood techniques to obtain estimates of the model parameters. Alternatively, using the model likelihoods defined above, we could take a Bayesian approach and define appropriate prior distributions for the parameters. Posterior distributions could then be obtained (probably a computer intensive method such as Markov chain Monte Carlo would be required), and inferences could be drawn in the light of these results.

Regardless of how the estimation is done, researchers can develop a suite of models that they wish to fit to the data, each representing a different hypothesis about the system under study. The suite of models can then be compared using information-theoretic methods (such as Akaike's Information Criterion for example) to determine which model(s) best describes the data. Alternatively, if the interest is in comparing two specific models representing competing hypotheses, e.g. whether there is evidence that local extinction probabilities are affected by El Niño events, or where there is a desire to test the effect of a factor being investigated via an appropriate experimental design, standard likelihood ratio tests can be used.

6. Extensions

Missing observations

A likely feature of many ecological studies is missing observations. In some instances it might not be possible to collect the required data: weather conditions may prevent access to some sites; vehicles may breakdown en route; or logistically it may not be possible to sample all sites within a suitably small time frame. MacKenzie *et al.* (2002) and MacKenzie *et al.* (2003) show that missing observations can easily be incorporated into the models described above. In effect, the detection probability for the respective survey of a site is set to zero, which fairly reflects the fact that the species could not be detected (even if present) as no survey was conducted at that time. Essentially, this removes the detection probability parameter from the model likelihood (with respect to the site in question). The ability of the model to handle missing observations has important ramifications for study designs, as it enables different sites to have different sampling intensities.

Incorporating covariates

Often researchers are interested in potential relationships between the model parameters (occupancy, colonization, local extinction and detection probabilities) and characteristics of the sites or generalized weather patterns (e.g. drought years). Further, the surveyors' ability to detect the species during any given survey may also be affected by localized conditions at the sampling site (e.g. weather conditions or intensity of nearby traffic noise). Using the logistic model (2),

$$\theta_i = \frac{\exp(Y_i\beta)}{1 + \exp(Y_i\beta)}, \quad (2)$$

MacKenzie *et al.* (2002, 2003) detail how such covariate information can be incorporated. The logistic model allows the relationship between the probability of interest for site i (θ_i) and the covariates, Y_i , to be modelled, with β denoting the covariate coefficients to be estimated. Analyses of this type could be considered as generalized logistic regression analyses, where

allowance has been made for uncertainty in the binary observation of occupancy state (due to imperfect detection).

The logistic model is not the only possible method for including covariate information; other functional forms may be used if desired.

7. Model assumptions and consequences of their violation

There are two main assumptions related to the above models that could be violated in practice: (i) the closure of sites with respect to changes in occupancy within the seasons; and (ii) the constancy of model parameters across sites (i.e. no heterogeneity).

The closure assumption is required for estimating detection probabilities correctly, so that for those sites where the species is detected at least once during a season, a non-detection means the 'species is present, but undetected'; it is not necessary to also include the possibility that during that particular survey the species may have been absent from the site. Kendall (1999) investigated the effect of violating the closure assumption in mark-recapture studies, and due to the strong similarities of mark-recapture models and the models reviewed above we can make some educated guesses about the effect of violating the closure assumption in the current situation. If the species occupies sites within a season in a random manner (i.e. the species randomly 'chooses' which sites it will occupy on any given day), then parameter estimates are unbiased although their interpretation must change. The occupancy parameter, ψ , should now be interpreted as the probability the species *uses* a site during a season, while the detection probability, p , now relates to the probability that the species both occupies the site and is detected in the survey. Other, more Markovian forms of movement between sites within seasons may cause parameter estimates to be biased, and indeed if one thinks this might be the case then a rethink of the study objectives and sampling design may be required. In fact, it may be appropriate to apply the 'multiple season model' above, within a single year or breeding season, where 'seasons' (as defined for the model) constitute a shorter time interval where the closure assumption is more likely to hold.

One of the more likely violations of the assumptions would be heterogeneity in detection probability due to differing abundances of the species at different sites: usually (but not always) the species is more detectable at sites with a higher abundance. This may or may not be problematic. Again, based upon experience with mark-recapture methods, unmodelled heterogeneity in detection probabilities causes occupancy probabilities to be underestimated (which has been confirmed via trial simulations), yet the dynamic probabilities of colonization and local extinction can be unbiased. Pollock *et al.* (1990) reviewed the literature available at the time with respect to the degree of bias in the survival estimates of the Jolly-Seber model caused by heterogeneous capture probabilities, with the consensus suggesting any bias was small. The same may apply in the current situation although this needs to be formally investigated.

Two important points need to be made. First, in many situations it might be possible to model the heterogeneity. While we are not able to include information on the actual abundances at each site (as these are unknown of course), there may be suitable covariates that could be used as surrogates, such as quality of habitat, which could reflect that different abundances are caused by different habitat qualities. Alternatively, Royle & Nichols (2003) suggest a modification of the single season model that allows for heterogeneity caused by varying levels of abundance, and also enables abundance to be estimated. Another approach would be to use finite mixture models such as those used by Norris & Pollock (1996) and Pledger (2000)

TABLE 1

Estimates (standard errors) of occupancy, local extinction and colonization probabilities for tiger salamander larvae and metamorphs in 40 Minnesota farm ponds. The 'Naïve 1' estimates are based upon the outcome of only the first survey of each pond; 'Naïve 2' estimates are based upon the outcome of all surveys conducted at the sites; and the 'Model' estimates are obtained from the described multiple-season model that explicitly accounts for detectability.

Method	$\hat{\psi}_{2000}$	$\hat{\epsilon}_{2000}$	$\hat{\gamma}_{2000}$	$\hat{\psi}_{2001}$
Naïve 1	0.08 (0.04)	0.00 (0.00)	0.11 (0.05)	0.18 (0.06)
Naïve 2	0.20 (0.06)	0.25 (0.15)	0.16 (0.06)	0.28 (0.07)
Model	0.26 (0.09)	0.24 (0.16)	0.13 (0.08)	0.29 (0.08)*

* This estimate and standard error have been derived and are not estimated directly from the model

in mark–recapture. Here the logic is that occupied sites consist of two (or more) different unknown types each with different detection probabilities. For example sites occupied by the species may have either low or high abundances, but the researchers cannot tell which type a particular site belongs to. However, this uncertainty can be expressed as a probability and included in the model to be estimated. Another approach might be to use Markov chain Monte Carlo, and assume the detection probabilities for each site come from some undefined (but reasonably well behaved) distribution. The main point is that there are options available for dealing with heterogeneity caused by variations in abundance.

Second, even if parameter estimates are biased by heterogeneous detection probabilities, the argument could be made that such an analysis is still better than using naïve counts that do not account for detectability at all. As discussed in MacKenzie *et al.* (2003), naïve approaches (based solely upon counts) are also likely to be affected by heterogeneity, but their assumptions are not explicitly outlined (or even realized) by many users. For accurate inferences to be made, the above methods assume closure of the sites with respect to occupancy during the surveying period, and that parameters are constant across sites; a naïve count requires these assumptions plus perfect detection of the species.

8. Example

A simple analysis of the tiger salamander (*Ambystoma tigrinum*) data, considered by MacKenzie *et al.* (2003), illustrates the methods. During the northern hemisphere spring and summer months of 2000 and 2001, dipnet surveys were conducted for larvae and metamorphs of various amphibian species at 40 farm ponds in south-eastern Minnesota. Biologically, it is reasonable to expect tiger salamander larvae and metamorphs to emerge at ponds during mid-spring, and continue to occupy ponds until late summer. Therefore, to satisfy our within-season closure assumption, the data have been truncated so that only surveys conducted from 1 May to 7 August are included in the analysis. Each day has been considered as a potential opportunity for surveying the farm ponds, hence a missing observation has been defined as any day on which a site was not surveyed. The 40 ponds were surveyed 5.0 times on average (range 1–9) in 2000; in 2001, 39 of the ponds were surveyed 3.0 times on average (range 2–7). Table 1 presents estimates of occupancy, local extinction and colonization probabilities using three approaches. The first two are naïve methods, that do not account for detectability, based upon (i) only the first survey of a site each year (representing the type of data expected if sites were only surveyed once per season); and (ii) all surveys each year. In both cases,

tiger salamander larvae or metamorphs were defined to (a) occupy a site if they were detected there at least once during the season; (b) have become locally extinct from a site between years if they were not detected at least once in 2001, having been detected there during 2000; and (c) have colonized a site if they were detected there in 2001 having not been detected there during 2000. The third set of estimates is obtained by applying the data to the multiple season model described above, where detection probabilities may vary between seasons but are constant within seasons.

The first set of naïve estimates is quite different to the other results, highlighting the dangers of basing inferences about the state of the species on a single presence/absence survey when the species is not detected with certainty. Based on this set of results we would conclude that the proportion of area occupied has increased by 139%. There is little difference between the second naïve method and the multiple season model estimates of local extinction and colonization probabilities, and level of occupancy in 2001. However there is a reasonable difference in the level of occupancy in 2000 which, for the naïve approach, translates to an apparent 41% increase in occupancy, but allowing for species detectability there is an apparent increase of only 15%. This is because in the 2000 season the probability of detecting the species in a dipnet survey of an occupied site was estimated to be 0.29, whereas in the 2001 season detection probability was estimated to have increased to 0.66. This has inflated the apparent rate of change in proportion of area occupied when detectability is not explicitly accounted for in the naïve approach. Without an omniscient viewpoint, we cannot judge which set of estimates most closely resembles the truth, although it would seem most prudent to use those that allow for uncertain detection.

An added advantage of using the modelling approach is that it can be used to constrain some parameters, to investigate, for example, whether there is variation over time. MacKenzie *et al.* (2003) fitted a suite of models to the data above to represent various constraints on parameter values. They found the most parsimonious model (in terms of Akaike's Information Criterion) to be one in which the overall level of occupancy was constant, while allowing for changes in the state of occupancy of the ponds between years.

9. Discussion

The methods reviewed above offer very flexible analysis for species presence/absence surveys, using well established statistical techniques (namely, likelihood theory). The ability to incorporate missing observations or unequal surveying effort across sites allows a wide range of sampling designs to be implemented. The most robust is likely to be a design in which all sites are visited an equal number of times. This design provides the most data, so more complex models could be investigated during an analysis, and checks of assumption violations would probably be more accurate. Repeated surveys allow detection probabilities to be estimated, but there is undoubtedly a point beyond which it is inefficient to collect further information on detectability. An alternative design might randomly choose a subset of sites that are repeatedly surveyed and survey other sites only once: such a design relies more heavily on the now untestable assumption that the detection function constructed for sites surveyed more often is equally appropriate for the sites surveyed once. Another possible design might repeatedly survey all sites until the species is detected for the first time (up to some maximum number of surveys). Once the species has been detected at a site, that site is never surveyed again during that season. Such a design could be useful where access to sites is difficult, and detection probabilities are constant across all surveys or modelled accurately by

a covariate relationship. These examples (there are many more that could be used in practice) are compatible with the methods described above, although there are implicit assumptions involved with each that need to be considered during analysis of the data.

There is also a great deal of flexibility in how the repeated surveys are conducted in practice. More than one survey could be conducted with each visit to a site, by either the same or multiple observers. However it is important to ensure that the surveys remain independent and that the methods used, and any assumptions implicit in the choice of method, are consistent across all sites.

Recently, this general approach of analysing species presence/absence data with imperfect detections has been extended to multiple species (MacKenzie, Bailey & Nichols, 2004). The joint modelling of two or more species allows questions of species co-occurrences to be addressed, while accounting for the imperfect detection of the species. Over the last 30 years, a substantial body of literature has detailed various methods for determining whether patterns observed in a species presence/absence matrix could have occurred by chance or show strong evidence of non-randomness (Connor & Simberloff, 1979, 1983; Gilpin & Diamond, 1982, 1984; Kelt, Taper & Mesevire, 1995; Manly, 1995; Gotelli, 2000; Gotelli & McCabe, 2002). However, by not accounting for imperfect detection, inferences can be misleading, as an apparent species absence may actually be due to the non-detection of the species. Furthermore, non-random patterns can be caused by different habitat preferences, or by ability to detect different species varying across habitat types. By extending our models above, the level of co-occurrence between species can be assessed while allowing for such habitat differences. In theory, the models could be extended to a large number of species; but the number of parameters to be estimated would be huge (requiring a very large dataset) and difficult to interpret. In practice, multiple-species models should be useful for assessing co-occurrences between a small number of species (< 4). A multiple-season version should also be able to estimate the rate at which one species is being replaced by another at the monitoring sites, which might be particularly useful for quantifying the impact an invasive species is having on a native, for example.

The above models could be applied to the monitoring of biodiversity or species richness. Rather than monitoring a single species at multiple sites, multiple species could be monitored at a single site. To develop a list of species that are of interest, multiple presence/absence surveys could be conducted within a common 'season' for all species, and the proportion of species on the list that are thought to be present could be estimated (allowing for imperfect detectability). The multiple season model would then be useful for assessing, and quantifying, changes in the community structure. Such an approach would be feasible for low-cost biodiversity monitoring.

Specialized computer software has been developed to enable the above methods to be applied. Program PRESENCE is freely available over the Internet, and may be downloaded from <http://www.proteus.co.nz>. The methods have also been implemented in Program MARK (<http://www.cnr.colostate.edu/~gwhite/mark>), although the multiple season model uses one of the alternative parameterizations where occupancy is estimated for each season, along with local-extinction probabilities.

References

- AZUMA, D.L., BALDWIN, J.A. & NOON, B.R. (1990). *Estimating the Occupancy of Spotted Owl Habitat Areas by Sampling and Adjusting Bias*. USDA Forest Service General Technical Report PSW-124.

- CONNOR, E.F. & SIMBERLOFF, D. (1979). The assembly of species communities: chance or competition? *Ecology* **60**, 1132–1140.
- CONNOR, E.F. & SIMBERLOFF, D. (1983). Interspecific competition and species co-occurrence patterns on islands: null models and the evaluation of evidence. *Oikos* **41**, 455–465.
- GEISSLER, P.H. & FULLER, M.R. (1987). Estimation of the proportion of area occupied by an animal species. In *Proceedings of the Section on Survey Research Methods of the American Statistical Association 1986*, pp. 533–538.
- GILPIN, M.E. & DIAMOND, J.M. (1982). Factors contributing to nonrandomness in species co-occurrences on islands. *Oecologia* **52**, 75–84.
- GILPIN, M.E. & DIAMOND, J.M. (1984). Are species co-occurrences on islands non-random, and are null hypotheses useful in community ecology. In *Ecological Communities: Conceptual Issues and Evidence*, eds D.R. Strong, Jr, D. Simberloff, L.G. Abele & A.B. Thistle, pp. 297–315. Princeton, NJ: Princeton University Press.
- GOTELLI, N.J. (2000). Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621.
- GOTELLI, N.J. & MCCABE, D.J. (2002). Species co-occurrence: a meta-analysis of J.M. Diamond's assembly rules model. *Ecology* **83**, 2091–2096.
- HANSKI, I. (1992). Inferences from ecological incidence functions. *American Naturalist* **139**, 657–662.
- HANSKI, I. (1994). A practical model of metapopulation dynamics. *J. Animal Ecology* **63**, 151–162.
- HANSKI, I. (1997). Metapopulation dynamics: from concepts and observations to predictive models. In *Metapopulation Biology: Ecology, Genetics, and Evolution*, eds I.A. Hanski & M.E. Gilpin, pp. 69–91. New York: Academic Press.
- KELT, D.A., TAPER, M.L. & MESEVRE, P.L. (1995). Assessing the impact of competition on community assembly: a case study using small mammals. *Ecology* **76**, 1283–1296.
- KENDALL, W.L. (1999). Robustness of closed capture–recapture methods to violations of the closure assumption. *Ecology* **80**, 2517–2525.
- MACKENZIE, D.I., BAILEY, L.L. & NICHOLS, J.D. (2004). Investigating species co-occurrence patterns when species are detected imperfectly. *J. Animal Ecology* **73**, 546–555.
- MACKENZIE, D.I., NICHOLS, J.D., HINES, J.E., KNUTSON, M.G. & FRANKLIN, A.D. (2003). Estimating site occupancy, colonization and local extinction probabilities when a species is not detected with certainty. *Ecology* **84**, 2200–2207.
- MACKENZIE, D.I., NICHOLS, J.D., LACHMAN, G.B., DROEGE, S., ROYLE, J.A. & LANGTIMM, C.A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology* **83**, 2248–2255.
- MANLY, B.F.J. (1995). A note on the analysis of species co-occurrences. *Ecology* **76**, 1109–1115.
- MOILANEN, A. (1999). Patch occupancy models of metapopulation dynamics: efficient parameter estimation using implicit statistical inference. *Ecology* **80**, 1031–1043.
- MOILANEN, A. (2002). Implications of empirical data quality for metapopulation model parameter estimation and application. *Oikos* **96**, 516–530.
- NORRIS, J.L. & POLLOCK, K.H. (1996). Nonparametric MLE under two closed capture–recapture models with heterogeneity. *Biometrics* **52**, 639–649.
- PLEDGER, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* **56**, 434–442.
- POLLOCK, K.H., NICHOLS, J.D., BROWNIE, C. & HINES, J.E. (1990). Statistical inference for capture–recapture experiments. *Wildlife Monographs* **107**.
- ROYLE, J.A. & NICHOLS, J.D. (2003). Estimating abundance for repeated presence absence data or point counts. *Ecology* **84**, 777–790.
- TYRE, A.J., TENHUMBERG, B., FIELD, S.A., NIEJALKE, D., PARRIS, K. & POSSINGHAM, H.P. (2003). Improving precision and reducing bias in biological surveys by estimating false negative error rates in presence–absence data. *Ecological Applications* **13**, 1790–1801.