

APPLICATION

Fitting occupancy models with E-SURGE: hidden Markov modelling of presence–absence data

Olivier Gimenez^{1*}, Laetitia Blanc¹, Aurélien Besnard¹, Roger Pradel¹, Paul F. Doherty Jr², Eric Marboutin³ and Rémi Choquet¹

¹Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Campus CNRS, 1919 Route de Mende, 34293 Montpellier Cedex 5, France; ²Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523-1474, USA; and ³Office National de la Chasse et de la Faune Sauvage, ZI Mayencin, 5 allée de Béthléem, 38610 Gières, France

Summary

1. Occupancy – the proportion of area occupied by a species – is a key notion for addressing important questions in ecology, biogeography and conservation biology. Occupancy models allow estimating and inferring about species occurrence while accounting for false absences (or imperfect species detection).

2. Occupancy models can be formulated as hidden Markov models (HMM) in which the state process captures the Markovian dynamic of the actual but latent states, while the observation process consists of observations that are made from these underlying states.

3. We show how occupancy models can be implemented in program E-SURGE, which was initially developed to analyse capture–recapture data in the HMM framework. Replacing individuals by sites provides the user with access to several features of E-SURGE that are *not available altogether* or *just not available* in standard occupancy software: i) flexible model specification through a user-friendly syntax without having to write custom code, ii) decomposition of the observation and state processes in several steps to provide flexible parameterisation, iii) up-to-date diagnostics of model identifiability, and iv) advanced numerical algorithms to produce fast and reliable results (including site random effects).

4. To illustrate E-SURGE features, we provide implementation and analysis details for several occupancy models. We also provide simulated and real-world examples as well as further specifications and information in a companion wiki platform <http://occupancyinesurge.wikidot.com/>.

Key-words: Capture–recapture, detectability, detection–non-detection, E-SURGE, hidden Markov models, presence–absence, species occurrence

Introduction

Occupancy models allow estimating and inferring about species occurrence while accounting for false absences or imperfect species detection (MacKenzie *et al.* 2006). These models have been extensively used to address important questions in fields as diverse as conservation biology, biogeography, wildlife epidemiology, metapopulation dynamics and community ecology (review in Bailey, MacKenzie & Nichols 2013).

Following the seminal work of MacKenzie *et al.* (2006), it was soon realized that occupancy models could be formulated as hidden Markov models (HMMs) in which two time series run in parallel: the state process captures the Markovian dynamic of the actual but latent states (e.g. site occupied vs. unoccupied), while the observation process consists of observations that are made from these underlying states (e.g. species detected vs. undetected; e.g. Royle & Kéry 2007). Interestingly, the formulation of capture–recapture models as HMMs was

also witnessed in the capture–recapture literature (Pradel 2005; Gimenez *et al.* 2012). A close connection exists between capture–recapture and occupancy models and can be realized by interchanging individuals in a capture–recapture context and sites in an occupancy context.

Several software programs are available to fit occupancy models, either in the frequentist framework with programs PRESENCE (Hines 2013), MARK (White & Burnham 1999) and the R package unmarked (Fiske & Chandler 2011), or in the Bayesian framework using the BUGS language (WinBUGS, OpenBUGS or JAGS; e.g. Kéry & Schaub 2012). WinBUGS requires writing custom code and knowledge about the Bayesian theory. Programs PRESENCE and MARK often require the construction of design matrices to specify models, a process that can be prone to user's errors. PRESENCE and unmarked do not incorporate random effects, and MARK does so only in limited ways.

Here, by exploiting the equivalence between occupancy and capture–recapture models, we illustrate how occupancy models can be implemented in program E-SURGE (Choquet, Rouan & Pradel 2009), which was initially developed to

*Correspondence author. E-mail: olivier.gimenez@cefe.cnrs.fr

analyse capture–recapture data in the HMM framework. We believe that some users will benefit greatly by realizing that E-SURGE has features that are not implemented together in available occupancy software, namely i) flexible model specification through a user-friendly syntax without having to write custom code, ii) advanced numerical algorithms to produce fast and reliable results, including the incorporation of site random effects, and several other features that are simply not implemented in these other software programs, namely iii) decomposition of the observation and state processes in several steps to provide flexible parameterisation, and iv) up-to-date diagnostics of model identifiability, in other words, a reliable way of counting the number of parameters entering the calculation of the Akaike Information Criterion. We believe that E-SURGE is a powerful option in the suite of software that can be used for occupancy problems; however, few users may recognize this option, even if they are current E-SURGE users. Our goal in this paper is to make E-SURGE more accessible to analysts and field biologists working on occupancy, and related, problems.

Hidden Markov modelling of occupancy data

In Fig. 1, we provide the HMM formulation of the general dynamic occupancy models to carry out inference about occurrence and how extinction and colonization drive changes in occurrence.

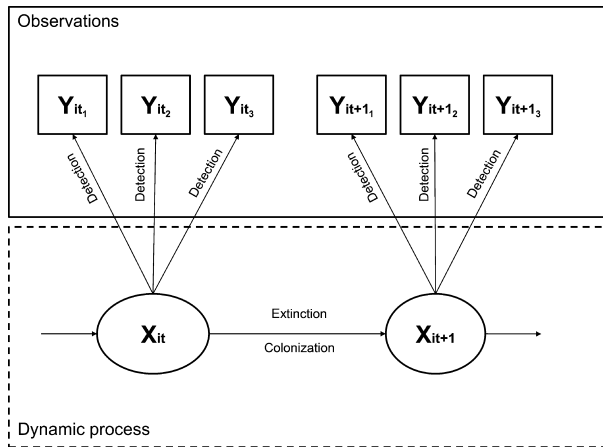


Fig. 1. Schematic hidden Markov representation of a dynamic occupancy model with imperfect detection. We consider site i between seasons (primary periods) t and $t + 1$ and three visits (secondary periods) within each primary period. Each site is closed (i.e. no change in its occupancy status) within primary periods but open (i.e. allowing for changes in occupancy status) between primary periods (robust design *sensu*; Pollock 1982). The first layer (circles) is a succession of hidden states or latent states (0 for unoccupied vs. 1 for occupied) of site i at season $t + 1$ (X_{it+1}) depending on its states at time t (X_{it}). The dynamic of the states is driven by transition probabilities, here probabilities of colonization $\gamma = \Pr(X_{it+1} = 1 | X_{it} = 0)$ and local extinction $\varepsilon = \Pr(X_{it+1} = 0 | X_{it} = 1)$. The second layer (squares) corresponds to the detection ($Y_{itj} = 1$) or not ($Y_{itj} = 0$) of the target species on site i at visit $j = 1, 2$ or 3 conditional on site i being in state X_{it} . These events are driven by the species detection probability $P = \Pr(Y_{itj} = 1 | X_{it} = 1)$. The initial-state probabilities $\psi_1 = \Pr(X_{i1} = 1)$ and $1 - \psi_1$ are not represented.

The parameters of interest are the probability of local extinction ε and of colonization γ as well as the detection probability P and the probability of initial occupancy ψ_1 where we have assumed, for simplicity, all parameters are constant across periods and sites. A HMM is built around three pieces of information: the vector of initial-state probabilities, the matrix of transition probabilities linking states in successive sampling occasions, and the matrix of observation probabilities linking observations and states at a given occasion. At the first sampling occasion $t = 1$, with the first state being ‘unoccupied’ and the second ‘occupied’, the vector of initial-state probabilities is:

$$[1 - \psi_1 \ \psi_1] \quad \text{eqn 1}$$

Then, the states are distributed as a first-order Markov chain governed by the transition matrix with states unoccupied and occupied at t in rows, and states unoccupied and occupied at $t + 1$ in columns:

$$\begin{bmatrix} 1 - \gamma & \gamma \\ \varepsilon & 1 - \varepsilon \end{bmatrix} \quad \text{eqn 2}$$

The observation process conditional on underlying occupancy states is summarized by a matrix with unoccupied and occupied states at t (rows) and undetected and detected observations at visits j (columns):

$$\begin{bmatrix} 1 & 0 \\ 1 - P & P \end{bmatrix} \quad \text{eqn 3}$$

Single-season occupancy models can be reformulated as HMMs and fitted in E-SURGE by imposing no extinction ($\varepsilon = 0$) and no colonization ($\gamma = 0$) in the dynamic model. The extension to multiple states with uncertainty (Nichols *et al.* 2007; MacKenzie *et al.* 2009) is illustrated with breeding states, and we consider three possible states. A site is ‘unoccupied’, ‘occupied by non-breeders’ or ‘occupied by breeders’, while the observations are species undetected (coded 0), species detected without young (coded 1) and species detected with young (coded 2). We use ψ^1 (resp. ψ^2) the probability that the site is occupied by non-breeders (resp. by breeders), P^1 (resp. P^2) the detection probability of non-breeders (resp. of breeders). The model can also accommodate state uncertainty. For example, reproduction may occur on a site, but the young may not be detected. We introduce δ the probability of detecting evidence of reproduction, given the site is occupied with young. Thus, the vector of initial-state probabilities is:

$$[1 - \psi^1 - \psi^2 \ \psi^1 \ \psi^2] \quad \text{eqn 4}$$

while the transition matrix is the identity matrix. The primary modifications are in the observation matrix, which can be written as

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 - P^1 & P^1 & 0 \\ 1 - P^2 & P^2(1 - \delta) & P^2\delta \end{bmatrix} \quad \text{eqn 5}$$

which can be split in two steps to highlight the successive processes of detection and breeding state ascertainment:

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 - P^1 & P^1 & 0 \\ 1 - P^2 & 0 & P^2 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 - \delta & \delta \end{bmatrix} \quad \text{eqn 6}$$

We introduce an intermediate set of observations: undetected, detection of non-breeders, and detection of breeders. The first matrix in the product (6) specifies the probabilities of these intermediate observations conditional on the states, while the second matrix specifies the probabilities of reproduction conditional on the intermediate observations.

Fitting models in E-SURGE

A typical session in E-SURGE follows the workflow provided in Fig. 2. Here, we provide a brief description of the main features of E-SURGE. We refer to the E-SURGE manual (Choquet & Nogué 2013) as well as Choquet (2008) and Choquet, Rouan and Pradel (2009) for additional details.

There are a few standard preliminary steps that we do not detail here. Most importantly, the user needs to select the Occupancy option. The number of events to specify is the number of observations (e.g. undetected, detection of non-breeders and detection of breeders), while the number of states should always be the number of states to be used (e.g. unoccupied, occupied by non-breeders, occupied by breeders) plus one. This additional state is needed because E-SURGE was initially developed to estimate demographic parameters and uses the absorbing state 'dead' to handle mortality as a transition from 'alive' states to the 'dead' state.

The model construction is served in E-SURGE by two modules called GEPAT (Generator of PATtern of elementary matrices) and GEMACO (Generator of Matrices of Constraints; Choquet 2008). In GEPAT, the user specifies the structure of the vector of initial-state probabilities, the matrix of transitions governing the state process and the matrix of observations (conditional on the states) governing the

observation process. To specify a parameter that will be estimated (i.e. that will be assigned an effect in GEMACO, see below), any letter can be used. The minus sign '-' means that the parameter corresponding to this cell will always be set to 0, while the asterisk '*' means the complementary of the sum of all the other parameters on the same row.

Of practical interest, the three elements of a HMM can be specified through a multistep process that proves very useful in accommodating state uncertainty (e.g. eqn 4).

In GEMACO, we specify the effects (*sensu* the design matrix in programs MARK and PRESENCE) using a self-explanatory syntax: for example, a season effect will be specified by 't' for time, a group effect by 'g', while a constant effect will be 'i' for intercept. If the effect of a site covariate needs to be investigated, we use 'i + xind' where xind specifies the slope of the relationship. A site random effect can also be specified in the same spirit using the syntax 'i + ind' (see below for an example). The matrices defined at the GEPAT step can be manipulated using the syntax 'from' for rows and 'to' for columns. For example, f(1).to(2) will pick the element in row 1 and column 2 of the corresponding matrix, and if a time effect is required on this element, then the syntax will be f(1).to(2).t.

Case studies in E-SURGE using simulated data

To illustrate the use of E-SURGE for fitting occupancy models, we used simulated data. Results from fitting the models to these data are provided in Table 1. Estimates for all parameters were close to the true values, and the 95% confidence intervals covered the true values in all cases.

Below, we detail the most important steps of implementing these models in E-SURGE. We refer the reader to the companion wiki website <http://occupancyinesurge.wikidot.com/> for full details on how to simulate the data and implement the model.

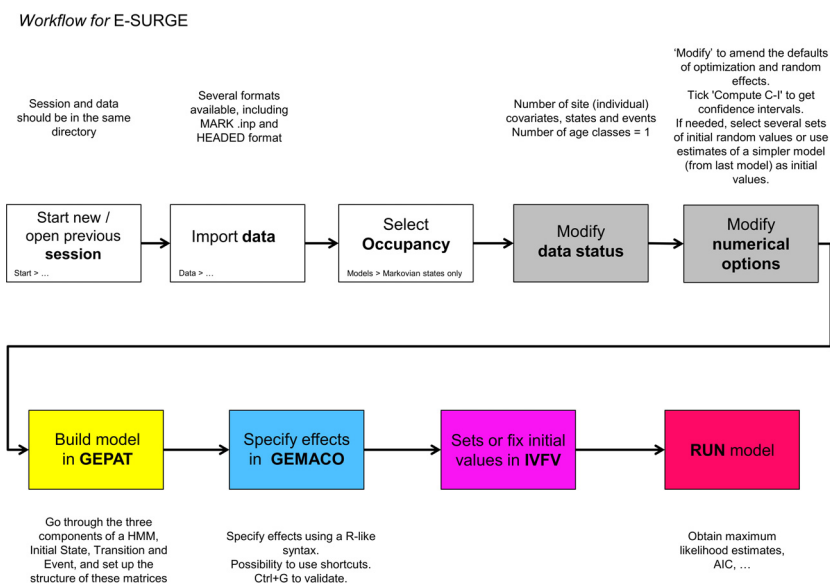


Fig. 2. Workflow diagram for E-SURGE. We describe the successive steps of a typical analysis in E-SURGE, from data input to model fitting through model building and effects specification. Steps that need to be accomplished through pull-down menus are in white boxes; other steps can be done directly from the main interface. We provide details on key steps in using E-SURGE in the text or boxes.

Table 1. Estimates of the parameters in the occupancy models fitted to the simulated data set in E-SURGE. A single data set was generated from the model under study with 250 sites, 3 primary occasions and 3 secondary occasions whenever required. For each parameter, the true value, the maximum likelihood estimate and the 95% confidence interval are provided

Model	Parameter	True value	Estimate	95% Confidence interval
Dynamic model	ψ_1	0.6	0.59	(0.53, 0.65)
	γ	0.3	0.32	(0.27, 0.39)
	ε	0.5	0.53	(0.47, 0.59)
	P	0.7	0.70	(0.67, 0.73)
Single-season	ψ	0.8	0.8	(0.70, 0.87)
	P	0.5	0.5	(0.44, 0.56)
Multistate with uncertainty	ψ^1	0.3	0.30	(0.23, 0.37)
	ψ^2	0.5	0.50	(0.43, 0.57)
	P^1	0.5	0.52	(0.43, 0.61)
	P^2	0.7	0.70	(0.65, 0.76)
	δ	0.8	0.79	(0.73, 0.85)

DYNAMIC MODELS

We start with the dynamic occupancy model described in formulas (1), (2) and (3) above. This mathematical formulation of the model can be translated for E-SURGE as follows. In GEPAT, the vector of initial-state probabilities is specified as

* ψ

which corresponds to (1). The additional dead state is not displayed at this step. Then, the transition matrix in (2) is specified as

* γ –
 ε * –
 – – *

where the last row and column are for the dead state. Finally, the observation matrix in (3) is coded as:

* –
 * P
 * –

where the last row corresponds to the dead state. In GEMACO, we write i at the Initial state and Event steps to impose a constant parameter. Regarding the Transition step, the data were simulated with three seasons (primary sessions) and three visits within each season (secondary session); therefore, we use `to.t(1 2 4 5 7 8)+to.t(3 6)` (or `to.t(1 2 4 5 7 8,3 6)`). The `to` specifies the columns of the matrix to be different, resulting in distinguishing the colonization and extinction parameters (note that from would produce exactly the same result by differentiating the rows). To handle the robust design, `t(1 2 4 5 7 8)` makes the intervals between secondary occasions identical, and the corresponding parameters will be fixed to 0 to impose closure within a primary session (i.e. neither extinction nor colonization). The term `t(3 6)` makes the intervals between the last

visit in a season and the first visit of the next one (between primary sessions) identical.

SINGLE-SEASON MODELS

We view a single-season model as a constrained dynamic model. Thus, we start with the dynamic model as described above. In GEPAT, the specification for the Initial state and the Event steps is exactly the same as for the dynamic model. To satisfy the closure assumption, we impose neither colonization nor extinction between sampling occasions by specifying at the Transition step:

* – –
 – * –
 – – *

In GEMACO, we use i for all steps to obtain constant parameters.

The extension of this model to multiple states with uncertainty is obtained as follows in E-SURGE. In GEPAT, the initial-state probabilities in (4) are:

* ψ ψ

while the Transition step is the same as before. The difference lies in the specification of the observation process that is accomplished in two steps to match (6) in GEPAT: step 1, the detection matrix, is:

* – – –
 * P – –
 * – P –
 * – – –

while step 2, the reproduction assignment matrix, is:

* – –
 – * –
 – * δ
 * – –

In GEMACO, we use i for the Initial state step to distinguish the occupancy probabilities according to states, from for step 1 of the Event step to distinguish the detection probabilities according to states and i for step 2 to have a constant assignment probability.

A real-world example illustrating heterogeneity in species detection

We use a real-world example to further illustrate E-SURGE capabilities. In particular, we show how E-SURGE can easily accommodate site-specific covariates, deal with heterogeneity in the species detection process and identify parameter redundancy issues.

Based on signs of presence collected between 2002 and 2006 by a network of trained volunteers, Rolland *et al.* (2011) modelled the occurrence of the Eurasian lynx (*Lynx lynx*) in the French Jura mountains. Briefly, a grid with cells of 9×9 km

was used, resulting in 197 sites, and the study period was split into five 1-year periods. Observers searched for signs of lynx presence, and each cell was assumed to be visited at least once each year.

The authors used a single-season occupancy model as described above. They, via model selection, found that the probability that a site was used by lynx was best explained by the proportion of forest cover x_i in site i . Thus, on the logit scale, we have:

$$\log \text{it}(\psi_i) = \beta_0 + \beta_1 x_i \quad \text{eqn 7}$$

where the β 's are regression parameters to be estimated.

To further incorporate site heterogeneity in the species detection process, we implemented a model with a site random effect and another one with a 2-class finite mixture.

A feature of E-SURGE is that it allows accommodating site random effects. Let us consider site-to-site variation in the detection probability as follows:

$$\log \text{it}(P_i) = \mu + \eta_i \quad \text{eqn 8}$$

where μ is the mean detection probability and η_i are normally distributed with mean 0 and variance σ^2 . Maximum likelihood estimates of μ and σ^2 are obtained through numerical integration via the Gauss–Hermite quadrature (Gimenez & Choquet 2010). More complex random structures can also be accommodated (Choquet & Gimenez 2012; Choquet *et al.* 2013).

As an alternative, we consider a 2-class finite mixture model (Royle 2006). We considered four states: unoccupied in class 1, unoccupied in class 2, occupied in class 1 and occupied in class 2. There are two observations: species undetected (0) or detected (1). The vector of initial-state probabilities is:

$$[\pi(1 - \psi_1) \quad (1 - \pi)(1 - \psi_2) \quad \pi\psi_1 \quad (1 - \pi)\psi_2] \quad \text{eqn 9}$$

where π is the proportion of sites that belong to class 1 and ψ_1 (resp. ψ_2) is the occupancy probability for sites that belong to class 1 (resp. 2). We obtain π by summing the first and third components in (9), while the ratio of the fourth component over π gives $1 - \psi_2$, hence ψ_2 , and the ratio of the third component over π gives ψ_1 .

However, when this mixture model is fitted, E-SURGE detects that the model is parameter redundant. More precisely, parameters π , ψ_1 and ψ_2 cannot be estimated separately, and the number of estimable parameters is 4, while the model has 5. To make these parameters estimable, the constraint $\psi_1 = \psi_2$ can be imposed to ensure that the occupancy probability is no longer redundant. This means disentangling i) the assignment of sites to a class of heterogeneity and ii) the occupancy process. Here, again, as in the multistate occupancy model with uncertainty in (6), the ability in E-SURGE to specify the state or the observation process in several steps proves useful. More precisely, we write the vector of initial-state probabilities (9) as the matrix product of X and Y in GEPAT where:

$$X = [\pi \quad 1 - \pi] \quad \text{eqn 10}$$

$$Y = \begin{bmatrix} 1 - \psi_1 & 0 & \psi_1 & 0 \\ 0 & 1 - \psi_2 & 0 & \psi_2 \end{bmatrix} \quad \text{eqn 11}$$

then force the ψ 's to be the same in GEMACO. The matrix of observation probabilities is:

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 - P_1 & P_1 \\ 1 - P_2 & P_2 \end{bmatrix} \quad \text{eqn 12}$$

where the states are in rows and the observations in columns.

We fitted three models to the lynx data, $M_0 = [\psi(\text{forest}), P]$, $M_{\text{ran}} = [\psi(\text{forest}), P(\text{random})]$ and $M_{\text{mix}} = [\psi(\text{forest}), P(\text{mixture})]$, with self-explanatory notation. The covariate forest cover had a similar positive effect on the occupancy probability whatever the structure on the detection probability ($\beta_1 = 7.4$ (SE = 1.3) in M_0 , $\beta_1 = 9.3$ (SE = 0.5) in M_{ran} and $\beta_1 = 9.1$ (SE = 1.8) in M_{mix}). Site-to-site variation in the detection process was equally captured by both models with heterogeneity (AICc[M_0] = 1000.1 vs. AICc[M_{ran}] = 978.6 and AICc[M_{mix}] = 980.2). In M_{ran} , the mean detection probability was estimated at $\hat{\mu} = 0.10$ (SE = 0.20) (hence $\log \text{it}^{-1}(\hat{\mu}) = 0.53$ (SE = 0.53)), while the estimated standard deviation of the random effect was $\hat{\sigma} = 1.25$ (SE = 0.22). In M_{mix} , the detection probability in state 1 was $\hat{P}_1 = 0.25$ (SE = 0.05) in proportion $\hat{\pi} = 0.54$ (SE = 0.10), while the detection probability in state 2 was $\hat{P}_2 = 0.74$ (SE = 0.08). By having the ability to easily construct heterogeneity models for detection probability in conjunction with occupancy, we have more confidence in our results. The relationship between lynx occupancy and forest cover is stronger with the highest ranked models (M_{ran} and M_{mix}) than with a lower ranking model that does not include any additional structure on detection (M_0). By modelling heterogeneity in detection, we are better able to uncover and describe important covariates of occupancy, which leads to better management.

Further E-SURGE capabilities

From a numerical perspective, E-SURGE provides a reliable number of parameters via an algorithm described in Choquet and Cole (2012), which is crucial in particular with parameter-redundant models or boundary estimates, and is one of the key steps for correct model selection using AIC. To date, we are aware of a single study investigating parameter redundancy in occupancy models which, actually, used E-SURGE (Fréville *et al.* 2013). With the ongoing development of complex models, the issue of parameter redundancy is likely to occur more often, and this algorithm in E-SURGE should be useful.

From a modelling perspective, E-SURGE offers the possibility to fit multiple species models (MacKenzie, Bailey & Nichols 2004; Waddle *et al.* 2010). Additionally, users may often be interested in deriving the probability that a site is occupied for sites without detections, either to build maps of presence or to calculate the so-called finite-population occupancy (MacKenzie *et al.* 2006). To do so, the latent states of HMMs need to be estimated, and this can be achieved in a frequentist framework via the Viterbi algorithm (Rouan *et al.* 2009; Fiske,

Royle & Gross 2013), which is available in E-SURGE (Rouan *et al.* 2009).

The companion wiki website <http://occupancyinesurge.wikidot.com/> presents such examples.

Conclusions

Although initially developed for capture–recapture data, E-SURGE can be efficiently used to build and analyse a variety of occupancy models via the HMM framework. E-SURGE includes a user-friendly syntax for specifying models without having to write custom code and uses advanced numerical algorithms to produce fast and reliable results. By strengthening the link between the two fast growing user communities of capture–recapture and occupancy, E-SURGE has the potential to provide a unified framework for the construction and analysis of hidden Markov models in ecology.

Acknowledgements

We thank the associate editor, Sophie Véran and three referees for comments on our work. We are indebted to Marc Kéry for a careful and detailed review of our manuscript. L. Blanc was supported by a PhD fellowship from Labex CeMeB.

Data accessibility

We provide the simulated data and the lynx data in the companion wiki platform <http://occupancyinesurge.wikidot.com/>.

References

- Bailey, L.L., MacKenzie, D.I. & Nichols, J.D. (2013) Advances and applications of occupancy models. *Methods in Ecology and Evolution*, doi: 10.1111/2041-210X.12100.
- Choquet, R. (2008) Automatic generation of multistate capture recapture models. *The Canadian Journal of Statistics*, **36**, 43–57.
- Choquet, R. & Cole, D.J. (2012) A hybrid symbolic-numerical method for determining model structure. *Mathematical Biosciences*, **236**, 117–125.
- Choquet, R. & Gimenez, O. (2012) Towards built-in capture-recapture mixed models in program E-SURGE. *Journal of Ornithology*, **152**, 625–639.
- Choquet, R. & Nogué, E. (2013) *E-SURGE 1.8 User's Manual*. Montpellier. Available from http://www.cefe.cnrs.fr/images/stories/DPTEvolution/bio_statistiques/LOGICIELS/E-SURGE-MANUAL.pdf.
- Choquet, R., Rouan, L. & Pradel, R. (2009) Program E - SURGE: a Software Application for Fitting Multievent Models. *Environmental and Ecological Statistics* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 845–865. Springer US, New York.
- Choquet, R., Sanz-Aguilar, A., Doligez, B., Nogué, E., Pradel, R., Gustafsson, L. & Gimenez, O. (2013) Estimating Demographic Parameters in the Wild with Dependence among Individuals within Cluster. *Methods in Ecology and Evolution*, **4**, 474–482.
- Fiske, I.J. & Chandler, R.B. (2011) Unmarked: an R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal Of Statistical Software*, **43**, 1–23.
- Fiske, I.J., Royle, J.A. & Gross, K. (2013) Inference for finite-sample trajectories in dynamic multi-state site-occupancy models using hidden Markov model smoothing. *Environmental and Ecological Statistics*. doi:10.1007/s10651-013-0256-1.
- Fréville, H., Choquet, R., Pradel, R. & Cheptou, P.-O. (2013) Inferring seed bank from hidden Markov models: new insights into metapopulation dynamics in plants. *Journal of Ecology*, **101**, 1572–1580.
- Gimenez, O. & Choquet, R. (2010) Incorporating individual heterogeneity in studies on marked animals using numerical integration: capture-recapture mixed models. *Ecology*, **91**, 951–957.
- Gimenez, O., Lebreton, J.-D., Gaillard, J.-M., Choquet, R. & Pradel, R. (2012) Estimating demographic parameters using hidden process dynamic models. *Theoretical Population Biology*, **82**, 307–316.
- Hines, J.E. (2013) PRESENCE 5.9 – Software to estimate patch occupancy and related parameters.
- Kéry, M. & Schaub, M. (2012) *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective*. Academic Press, Oxford, UK.
- MacKenzie, D.I., Bailey, L.L. & Nichols, J.D. (2004) Investigating species co-occurrence patterns when species are detected imperfectly. *Journal of Animal Ecology*, **73**, 546–555.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press, Oxford, UK.
- MacKenzie, D.I., Nichols, J.D., Seamans, M.E. & Gutiérrez, R.J. (2009) Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology*, **90**, 823–835.
- Nichols, J.D., Hines, A.J.E., Mackenzie, D.I., Seamans, M.E. & Gutiérrez, R.J. (2007) Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology*, **88**, 1395–1400.
- Pollock, K.H. (1982) A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, **46**, 752–757.
- Pradel, R. (2005) Multievent: an extension of multistate capture-recapture models to uncertain states. *Biometrics*, **61**, 442–447.
- Rolland, J., Basille, M., Marboutin, E. & Gaillard, J.-M. (2011) Comparing profile methods and site-occupancy modelling for the study of occurrence of an elusive species. *European Journal of Wildlife Research*, **57**, 1115–1118.
- Rouan, L., Gaillard, J.-M., Guédon, Y. & Pradel, R. (2009) Estimation of lifetime reproductive success when reproductive status cannot always be assessed. *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 867–879. Springer, Berlin, Germany.
- Royle, J.A. (2006) Site occupancy models with heterogeneous detection probabilities. *Biometrics*, **62**, 97–102.
- Royle, J.A. & Kéry, M. (2007) A Bayesian state-space formulation of dynamic occupancy models. *Ecology*, **88**, 1813–1823.
- Waddle, J.H., Dorazio, R.M., Walls, S.C., Rice, K.G., Beauchamp, J., Schuman, M.J. & Mazzotti, F.J. (2010) A new parameterization for estimating co-occurrence of interacting species. *Ecological Applications*, **20**, 1467–1475.
- White, G.C. & Burnham, K.P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, 120–139.

Received 27 September 2013; accepted 24 March 2014

Handling Editor: Nick Isaac