

Assessing individual heterogeneity using model selection criteria: how many mixture components in capture–recapture models?

Sarah Cubaynes^{1,2}, Christian Lavergne², Eric Marboutin³ and Olivier Gimenez^{1*}

¹Centre d'Ecologie Fonctionnelle et Evolutive, UMR 5175, Campus CNRS, 1919 Route de Mende, 34293 Montpellier Cedex 5, France; ²Institut de Mathématiques et Modélisation de Montpellier, UMR 5149, place Eugène Bataillon, 34095 Montpellier Cedex 5, France; and ³Office National de la Chasse et de la Faune Sauvage, ZI Mayencin, 5 Allée de Béthléem, 38610 Gières, France

Summary

1. Capture–recapture mixture models are important tools in evolution and ecology to estimate demographic parameters and abundance while accounting for individual heterogeneity. A key step is to select the correct number of mixture components i) to provide unbiased estimates that can be used as reliable proxies of fitness or ingredients in management strategies and ii) classify individuals into biologically meaningful classes. However, there is no consensus method in the statistical literature for selecting the number of components.

2. In ecology, most studies rely on the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) that has recently gained attention in ecology. The Integrated Completed Likelihood criterion (ICL; *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22, 719) was specifically developed to favour well-separated components, but its use has never been investigated in ecology.

3. We compared the performance of AIC, BIC and ICL for selecting the number of components with regard to a) bias and accuracy of survival and detection estimates and b) success in selecting the true number of components using extensive simulations and data on wolf (*Canis lupus*) that were used for management through survival and abundance estimation.

4. Bias in survival and detection estimates was < 0.02 for both AIC and BIC, and more than 0.09 for ICL, while mean square error was < 0.05 for all criteria. As expected, bias increased as heterogeneity increased. Success rates of AIC and BIC in selecting the 'true' number of components were better than ICL (68% for AIC, 58% for BIC, and 16% for ICL). As the degree of heterogeneity increased, AIC (and BIC in a lesser extent) overestimated the number of components, while ICL often underestimated this number. For the wolf study, the 2-class model was selected by BIC and ICL, while AIC could not decide between the 2- and 3-class models.

5. We recommend using AIC or BIC when the aim is to estimate parameters. Regarding classification, we suggest taking the classification quality into account by using ICL in conjunction with BIC, pending further work to adapt its penalty term for capture–recapture data.

Key-words: capture–recapture, classification, individual heterogeneity, information criteria, mixture models, simulation experiment

Introduction

Studying natural populations is of major interest for understanding the functioning of biological systems. In particular, the assessment of demographic parameters is essential for

studying population dynamics and assessing fitness in wild populations. In practice, however, individuals may or may not be detected (seen or recaptured) at various occasions during their lifetime, which raises the issue of detectability < 1 (Gimenez *et al.* 2008). Capture–recapture (CR) models were developed to estimate demographic parameters while accounting for the imperfect detection of individuals (Lebreton *et al.* 1992). While standard CR models consider populations as

*Correspondence author. E-mail: olivier.gimenez@cefe.cnrs.fr
Correspondence site: <http://www.respond2articles.com/MEE/>

communities of homogeneous individuals sharing the same traits, this assumption cannot be expected to hold when working on non-clonal species. Several studies have shown that ignoring individual heterogeneity (IH) can lead to substantial bias in demographic rates (Carothers 1973) and population abundance (Cubaynes *et al.* 2010). Moreover, considering all individuals as homogeneous can impede the study of evolutionary processes, individual variability in traits being a necessary condition for natural selection. With evolutionary questions in wild populations being of growing interest (e.g. Kingsolver *et al.* 2001), the use and development of CR models incorporating IH in demographic parameters are increasing (Tuljapurkar, Steiner & Orzack 2009; Gimenez & Choquet 2010; Péron *et al.* 2010; Pledger, Pollock & Norris 2010).

Several approaches are available to cope with IH in CR models. First, IH can explicitly be integrated into CR models using covariates or states (Pollock 2002; Lebreton *et al.* 2009). However, individual characteristics cannot always be measured in wild populations, so some sources of IH may remain unknown. Second, discrete IH can be incorporated in CR models using finite-mixture models (Pledger 2000; Pledger, Pollock & Norris 2003). In these models, a latent variable is used to assign individuals to one of the mixture components characterized by specific parameters. CR mixture models have had several applications in both ecology and evolution (Dorazio & Royle 2003; Véran *et al.* 2007; Bunge & Barger 2008; Pledger & Phillpot 2008; Morgan & Ridout 2009; Tyrrell *et al.* 2009; Wanger *et al.* 2009; Cubaynes *et al.* 2010; Péron *et al.* 2010; Pradel *et al.* 2010; Oliver *et al.* 2011).

The key steps in fitting a mixture model are (i) to determine the number of mixture components and (ii) estimate the parameters characterizing these components, typically using maximum likelihood (Pledger, Pollock & Norris 2003; Pradel 2005). Step (i) is crucial in model building as, besides affecting the parameter estimates, components often represent 'true' classes of individuals sharing same survival or detection parameters in the population, for example, 'good quality' vs. 'bad quality' individuals in a senescence analysis (Péron *et al.* 2010), or 'infected' and 'healthy' individuals in disease ecology. Indeed, biologists often aim at identifying the 'true' number of components, to classify (i.e. assign to a component) individuals into biologically meaningful classes.

The number of mixture components is usually accomplished through model selection by comparing several candidate models with different numbers of components. In the statistical literature, however, there is no consensus regarding the choice of method (McLachlan & Basford 1988; Wedel, Kamakura & Bockenholt 2000; Andrews & Currim 2003; Brame, Nagin & Wasserman 2006). In ecology and evolution, model selection generally relies on the Akaike Information Criterion (AIC; Akaike 1974; Johnson & Omland 2004), the selection of mixtures in CR models being no exception (Burnham & Anderson 2002; Pledger, Pollock & Norris 2003). AIC has been proven to be efficient in the sense that 'it behaves "almost as well," in terms of mean square error [...] as the theoretically best model' (Claeskens & Hjort 2008) but, by construction, it tends to

select too complex model (Kass & Raftery 1995), so that it may overestimate the 'true' number of mixture components (McLachlan & Peel 2000). The Bayesian Information Criterion (BIC; Schwarz 1978) is another commonly used criterion that has recently gained attention in ecology (Link & Barker 2006). There is strong support for BIC in mixture modelling (Roeder & Wasserman 1997; Fraley & Raftery 1998) as BIC has been shown to be consistent (Keribin 2000), i.e. to select the actual model if it is in the set of candidate models. However, when dealing with real data, for which there is no true model, BIC may also overestimate the number of components, as it does not account for the separation of the mixture components (Biernacki, Celeux & Govaert 2000). The Integrated Completed Likelihood criterion (ICL; Biernacki, Celeux & Govaert 2000) was recently developed to overcome these limitations, and its use has been recommended in mixture modelling (McLachlan & Peel 2000) but its potential has never been investigated in ecological studies. ICL was derived from BIC by including an extra term called entropy, which quantifies the degree of separation of the mixture components, hence favouring well-separated components. By accounting for the quality of the classification, ICL should avoid overestimating the number of components, but may underestimate this number if the components are poorly separated (Biernacki, Celeux & Govaert 2000; McLachlan & Peel 2000). Although the effects of overestimating or underestimating the number of components on the estimation of survival and detection parameters are not clear a priori, identification of biologically meaningful classes is obviously affected.

Given the lack of consensus in the literature, there is a need to evaluate the performance of model selection criteria with respect to *i*) bias in demographic parameters and (ii) classification of individuals. There have been several earlier attempts to do this in the literature (Fraley & Raftery 1998; Andrews & Currim 2003; Brame, Nagin & Wasserman 2006; Fonseca & Cardoso 2007; Lukociene & Vermunt 2010), but nothing to our knowledge for CR data. In addition, there are aspects of CR data that might affect the performance of the criteria. Encounter histories are right-censored, which makes it hard-to-classify an individual that is captured for the first time close to the end of the study. Besides, because of the limited size of CR data sets, models with more than three classes are not worth fitting, as confidence intervals become too wide to be useful in an applied context, for example, to produce reliable abundance estimates (Pledger 2000).

We performed an extensive simulation study to evaluate the performance of AIC, BIC and ICL in selecting the number of components in CR studies. We paid particular attention to distinguish the aims of bias reduction and classification, where the aim is to identify the 'true' number of components, and correctly assign individuals into the different components or classes. We considered a set of 240 scenarios generated from a 2-class distribution, covering a wide range of biological situations. In addition, we compared the three model selection criteria using real CR data on wolves (*Canis lupus*) that were used to estimate survival and abundance in a management setting.

Material and methods

SIMULATIONS

We conducted a simulation study to evaluate the performance of AIC, BIC and ICL at selecting the number of mixture components in CR models. First, we generated CR data from a 2-class model with heterogeneous survival or detection probability for a wide range of scenarios. Second, each generated data set was analysed using 1-, 2- and 3-class mixture models. Third, we performed model selection using AIC, BIC and ICL. Fourth, we determined the criterion that led to (i) minimizing the bias and maximizing the precision for both survival and detection parameters, and (ii) selecting the 'true' number of components used to generate the data. We then investigated the effect of different factors including skewness of the mixture distribution, degree of heterogeneity, value of the heterogeneity parameter, number of sampling occasions and individuals, on the performance of the criteria. We used 500 iterations, and all simulations were performed in MATLAB (Supporting Information).

Simulation of CR histories

For each situation, we simulated $n = 150$ individual CR histories over $t = 15$ sampling occasions, with 10 individuals released at each sampling occasion. At each sampling occasion, the probability that a given individual i is alive was determined by a Bernoulli distribution with probability of survival set to a specific value (scenarios 1 and 2) or determined by a mixture of beta distributions (scenarios 3 and 4). Among individuals alive at a given occasion, the event being detected was governed by a Bernoulli distribution with probability of detection set to a specific value (scenarios 3 and 4) or determined by a mixture of beta distributions (scenarios 1 and 2).

Fitting finite-mixture models to CR data

Capture–recapture mixture models allow for heterogeneity in the survival or detection process by considering the study population to be a mixture of a finite number of latent classes of individuals (Pledger 2000; Pledger, Pollock & Norris 2003; Pradel 2005, 2009). Considering a 2-class model with heterogeneous detection, an animal may be in one of the three states: alive in class 1, alive in class 2 or dead, and the following observations may be made: 1 if detected and 0 otherwise. We define parameters π (respectively $1 - \pi$) for the proportion of newly marked individuals in class 1, (resp. class 2), ϕ the survival probability and p_1 and p_2 the detection probabilities for individuals alive in classes 1 and 2, respectively. Starting from a 2-class model, assuming all individuals have equal detection probabilities (i.e. $p_1 = p_2$) gives the 1-class model, while considering an extra mixture leads to a 3-class model. A model with heterogeneous survival can easily be obtained by setting ϕ to be class-specific and p homogeneous. Under the assumption of independence among individuals, the likelihood is the product of the probability of all encounter histories (Lebreton *et al.* 1992). As an example of calculation of an individual contribution to the likelihood, let us consider a CR history 101 of an individual encountered on the first and third sampling occasions but missed on the second. The probability of this particular CR history, under a C-class model with heterogeneous detection, is $\sum_{i=1}^C \pi_i \cdot \phi \cdot (1 - p_i) \cdot \phi \cdot p_i$. We fitted all models using maximum likelihood.

Generating distributions

We generated CR data from a 2-class mixture model with parameters constant through time. The proportion of individuals in component 1

and the non-heterogeneous parameter (survival or detection) were set to a specific value and the heterogeneous parameter (survival or detection) was determined by a mixture of two beta distributions. We, hence, avoided cases where the true model was in the set of candidate models, which would have favoured BIC. We adjusted μ_i and σ_i , the mean and standard deviation of component i , to obtain various levels of heterogeneity. To cover a wide range of situations, we determined parameters x_1 and x_2 of $\beta(x_1; x_2)$ for each mixture from 60 generating distributions, by forming all possible combinations with (i) $\pi = 0.2, 0.5$ or 0.8 , (ii) $\mu_1 = 0.1, 0.3$ or 0.7 and $\mu_2 = 0.3, 0.5$ or 0.9 , (iii) $\sigma_1 = 0.0001$ or 0.05 and $\sigma_2 = 0.0001$ or 0.05 . To characterize each distribution, we calculated the mean value of the heterogeneity parameter $\mu = \pi\mu_1 + (1 - \pi)\mu_2$, the heterogeneity coefficient $\eta = \frac{\sigma^2}{\mu(1-\mu)}$, the variance between components $\sigma^2 = \pi(\mu_1 - \mu)^2 + (1 - \pi)(\mu_2 - \mu)^2$, and the skewness coefficient $\gamma = \frac{E[(\theta_i - \mu)^3]}{\sigma^3}$ (Dorazio & Royle 2003). We further considered four biological scenarios as follows:

- Detection heterogeneity in a short-lived species, with survival fixed at 0.6;
- Detection heterogeneity in a long-lived species, with survival fixed at 0.95;
- Survival heterogeneity with relatively low detectability, with detection fixed at 0.7;
- Survival heterogeneity with high detectability, with detection fixed at 0.9.

In total, this design led to 240 different situations.

We performed additional simulations to assess the effects of the number of sampling occasions ($t = 15$ or 30) and the number of individuals ($n = 150$ or 300) on model selection. We tested these effects in one situation with heterogeneous detection (scenario 2) and one situation with heterogeneous survival (scenario 4). Both situations presented a high degree of heterogeneity ($\eta > 0.74$), no skewness ($\gamma = 0$) and a medium mean value of the heterogeneity parameter ($\mu = 0.5$). Data were simulated with $\pi = 0.5$, $\mu_1 = 0.1$ and $\mu_2 = 0.9$.

Model selection

Akaike Information Criterion, BIC and ICL are criteria based on a penalized likelihood of the general form:

$$IC_M = -2\log L + p_M$$

where $p_M > 0$ is the penalty applied to the likelihood L of model M . The three criteria aim to find the best balance between the fit of the model to the data and its complexity. This balance is achieved for the model with minimal IC_M . The difference between the three criteria lies in the value of the penalty.

One of the most commonly used information criteria in ecology (Johnson & Omland 2004) is AIC (Akaike 1974) that provides an estimate of the 'distance' between an approximate model and the truth, and for which $p_M = 2k$, where k is the number of parameters in model M . Another widely used information criterion is BIC (Schwarz 1978) that was designed to find the most probable model given the data, as an estimate of the Bayes factor for two competing models (Schwarz 1978; Kass & Raftery 1995). For BIC, $p_M = k \ln(n)$ where n is the sample size; for CR data, this is the number of individuals sampled at least once. ICL (Biernacki, Celeux & Govaert 2000) was designed to select the model leading to the greatest evidence for clustering the data, by maximizing the integrated likelihood. We used the BIC like approximation of ICL (Biernacki, Celeux & Govaert 2000), which is derived from the BIC, but involves an extra penalty for poor

classification quality. For ICL, $p_M = k \ln(n) - 2 \text{ENT}$ where the so-called entropy term, $\text{ENT} = \sum_{i=1}^n \sum_{s=1}^s p_{is} \ln p_{is}$, quantifies the ability of a mixture model to provide well-separated classes, with p_{is} being the estimated posterior probability that individual i belongs to component s of model M . Classification was achieved by assigning individuals to a component *a posteriori*, i.e. individual i is assigned to component s of the model if $p_{is} > 0.5$. If the components are well separated, $\text{ENT} \approx 0$ and the classification is almost perfect; if not, ENT is large and positive and the rate of error of classification of individuals increases.

STATISTICAL ANALYSES

AIC, BIC and ICL performance with respect to parameter estimates

We assessed bias and precision with the 1-, 2- and 3-class models. Let $\hat{\theta}_i$ be the estimate of parameter θ ($\theta = \phi$ or $\theta = p$) for simulation i . Bias was calculated as $B = \frac{1}{J} \sum_{i=1}^J (\hat{\theta}_i - \theta)$, J being the number of simulations. To assess precision, we calculated mean square error (MSE) as $\text{MSE} = \frac{1}{J} \sum_{i=1}^J (\hat{\theta}_i - \theta)^2$. A low MSE means a good trade-off between low bias and low variance. We calculated the bias and MSE on survival and detection estimates obtained with the model selected by AIC, BIC and ICL for each scenario. Hereafter, we refer to as, for example, 'the bias of AIC', the value of bias on parameter estimates obtained with the model selected by AIC. Then, we performed linear regressions to test the effect of μ , η and γ on bias or MSE for AIC, BIC and ICL.

AIC, BIC and ICL performance at selecting the 'true' number of components

We calculated AIC, BIC and ICL percentage of success, underestimation and overestimation of the 'true' number of mixture components. A success occurred when the 2-class model was selected, an underestimation when the 1-class model was selected and an overestimation when the 3-class model was selected. We performed multinomial regressions to test the effect of μ , η and γ on model choice for AIC, BIC and ICL using the R package *mlogit*. The effect of scenario was also included as a factor, and the 2-class model served as a reference. The regression coefficients β are the log of the ratio of the two probabilities of choosing the 1-class or 3-class model over choosing the reference model. For example, if β_μ represents the effect of μ on the probability of choosing the 1-class model over the 2-class model, we expect that for a unit change in μ , the log of the ratio of the probability of underestimating the actual number of components increases by β_μ and the relative risk of choosing the 1-class over 2-class model to increase by $\exp(\beta_\mu)$.

CASE STUDY

As a case study, we analysed a CR data set obtained from the non-invasive monitoring of wolves in the French Alps based on DNA genotyping. The data set included the capture history of 160 different individuals that were monitored over 35 3-month sessions from 1995 to 2003. A previous analysis revealed the existence of detection heterogeneity most likely related to social status (Cubaynes *et al.* 2010). The objectives were to (i) estimate survival, (ii) estimate detection probabilities to derive population size and (iii) identify individuals belonging to the 'high detection' class (we expected them to be dominant individuals) vs. 'low detection' class (subordinates and young). This data set corresponds to scenario 2

in the simulations, i.e. detection heterogeneity in a long-lived species.

Results

AIC, BIC AND ICL PERFORMANCE WITH RESPECT TO PARAMETER ESTIMATES

On average, AIC and BIC both performed well in minimizing bias and MSE on detection and survival estimates, while ICL did worse (Table 1). Bias was generally < 0.02 for AIC and BIC, while up to > 0.09 for ICL and $\text{MSE} < 0.05$ for all criteria.

INFLUENCE OF THE DATA SET CHARACTERISTICS ON AIC, BIC AND ICL PERFORMANCE WITH RESPECT TO PARAMETER ESTIMATES

We assessed the effect of the mean value of the heterogeneity parameter μ , heterogeneity η and skewness γ on bias in detection (scenarios 1 and 2; Appendix S1) and survival (scenarios 3 and 4; Appendix S1). In all scenarios, bias in detection (scenarios 1 and 2) or survival (scenarios 3 and 4) was mainly affected by both η and μ (in interaction) for the 3 criteria. For all criteria, bias increased as the degree of heterogeneity increased, even more for low values of the heterogeneity parameter. This was particularly important in scenarios with heterogeneous detection, rather than survival (Appendix S2).

In the presence of heterogeneous detection in a short-lived species (scenario 1), BIC performed better at minimizing bias than AIC when mean detection was relatively low (Fig. 1).

Akaike Information Criterion selected models with a larger bias in detection than BIC did, which was even worse for high values of η . When mean values of detection were more than 0.5, BIC and AIC performed equally well, even better for lower values of η , and both criteria performed even better in the case of a long-lived species (scenario 2; Appendix S2).

In the presence of survival heterogeneity with relatively low detection (scenario 3), BIC performed better than AIC at minimizing bias in survival, especially for short-lived species (Appendix 2). BIC and AIC performed equally well and provided almost unbiased estimates when detection was high (scenario 4).

As expected, ICL performed less well than AIC and BIC in all scenarios, in particular for high values of η . MSE was relatively low whatever the model was selected, although a bit higher with the 3-class model (results not shown).

AIC, BIC AND ICL PERFORMANCE AT SELECTING THE 'TRUE' NUMBER OF COMPONENTS

Overall, AIC performed better than BIC and ICL in selecting the 'true' number of components, except with long-lived species with heterogeneous detection, for which BIC did better (Table 2). Mean success rate varied between 53% and 81% for AIC, between 28% and 72% for BIC and was always lower

Table 1. Mean absolute bias (B) and MSE for estimates of survival ($\hat{\phi}$) and detection (\hat{p}) probabilities calculated for AIC, BIC and ICL in the four scenarios

	B ($\hat{\phi}$)	MSE ($\hat{\phi}$)	B (\hat{p})	MSE (\hat{p})
Detection heterogeneity in a short-lived species (scenario 1)				
AIC	-0.007	0.001	-0.006	0.046
BIC	-0.02	0.001	0.030	0.009
ICL	-0.046	0.002	0.071	0.005
Detection heterogeneity in a long-lived species (scenario 2)				
AIC	-0.001	< 0.001	-0.012	0.036
BIC	-0.003	< 0.001	-0.007	0.035
ICL	-0.013	< 0.001	0.027	0.008
Survival heterogeneity with relatively low detection (scenario 3)				
AIC	-0.016	0.044	0	< 0.001
BIC	-0.002	0.038	-0.002	< 0.001
ICL	0.096	0.015	-0.019	0.001
Survival heterogeneity with high detection (scenario 4)				
AIC	0.016	0.041	0.001	< 0.001
BIC	-0.006	0.037	0	< 0.001
ICL	0.076	0.012	-0.005	< 0.001

Values of bias and MSE for the heterogeneous parameter are in bold font.

AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; ICL, Integrated Completed Likelihood; MSE, mean square error.

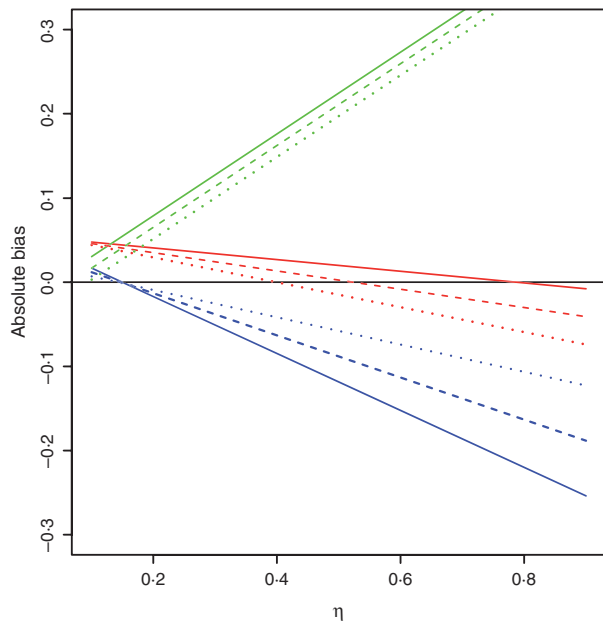


Fig. 1. Predicted bias as a function of the heterogeneity coefficient (η) for AIC (blue), BIC (red) and ICL (green). We assumed heterogeneity in a relatively low detection ($\mu = 0.3$) in a short-lived species (scenario 1). The relationships are displayed for a skewness coefficient of $\gamma = 1.5$ (solid line), $\gamma = 0$ (dashed line) and $\gamma = -1.5$ (dotted line). The horizontal line stands for no bias.

than 33% for ICL. While both AIC and BIC performed well in presence of survival heterogeneity, they tended to underestimate the number of components in the case of detection heterogeneity for a short-lived species – by about 45% for AIC and 72% for BIC – and to overestimate it in the case of a long-lived

Table 2. Mean percentage of success, underestimation and overestimation of the number of mixture components for AIC, BIC and ICL in the presence of detection heterogeneity in a short-lived or long-lived species (respectively, scenarios 1 and 2) and survival heterogeneity with relatively low or high detection (respectively, scenarios 3 and 4). A success was reported when the 2-class model was selected, an underestimation when the 1-class model was selected and an overestimation when the 3-class model was selected

	Scenario 1 $\hat{\phi}_{(0.60)}$ P(het)	Scenario 2 $\hat{\phi}_{(0.95)}$ P(het)	Scenario 3 $\hat{\phi}_{(het)}$ P(0.7)	Scenario 4 $\hat{\phi}_{(het)}$ P(0.9)
Success				
AIC	54.0	54.3	79.2	80.7
BIC	28.7	66.7	66.2	71.3
ICL	0.0	32.2	13.1	18.4
Underestimating				
AIC	44.5	4.3	19.4	16.6
BIC	71.3	11.0	33.8	28.7
ICL	100	66.6	86.9	81.6
Overestimating				
AIC	1.6	41.4	1.4	2.8
BIC	0.1	22.2	0.0	0.0
ICL	0.0	1.2	0.0	0.0

AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; ICL, Integrated Completed Likelihood.

species – by about 42% for AIC and 23% for BIC. In contrast, ICL tended to underestimate the number of components by a factor ranging from 81% up to 100%.

INFLUENCE OF THE DATA SET CHARACTERISTICS ON AIC, BIC AND ICL PERFORMANCE AT SELECTING THE 'TRUE' NUMBER OF COMPONENTS

The risk of overestimating the number of components was mainly affected by an interaction of mean value of the heterogeneity parameter (μ) and heterogeneity coefficient (η) for AIC, and by an effect of η for BIC (Appendix S3). The risk increased as the degree of heterogeneity increased, and even more for high values of the heterogeneity parameter for AIC. For both criteria, this effect was stronger in the presence of detection heterogeneity in a long-lived species (scenario 2). Regarding ICL, the risk of underestimating the number of components was mainly and negatively affected by μ , independently of the scenario considered, and it reached 100% almost for low heterogeneous survival, i.e. short-lived species with heterogeneous survival (Appendix S4).

In scenarios with heterogeneous survival, AIC and BIC both performed generally well (> 80% of success) for $\eta > 0.4$, and AIC performed slightly better than BIC for lower values of η . ICL performed generally poorly, but its success rate increased for $\eta > 0.6$ and even more for skewness $\gamma < 0$, i.e. when a large proportion of the population has a higher survival (Appendix 4). The same pattern was observed in the scenario with heterogeneous detection in a short-lived species (scenario 1), but the criterion's performance was reduced.

On the contrary, in the case of heterogeneous detection in a long-lived species (scenario 2), BIC did almost always better

than AIC especially when skewness γ was highly negative, while ICL did better than BIC for high values of η , for which BIC and AIC tended to overestimate the number of components. These discrepancies between the criteria were even bigger when γ and μ increased (Fig. 2).

INFLUENCE OF THE NUMBER OF SAMPLING OCCASIONS (T) AND NUMBER OF INDIVIDUALS (N) ON AIC, BIC AND ICL PERFORMANCE AT SELECTING THE 'TRUE' NUMBER OF COMPONENTS

There was no clear effect of the number of sampling occasions t and the number of individuals n on AIC and BIC model selection. In contrast, ICL selected more complex models as t increased and to a lesser extent as n increased (Table 3).

With heterogeneous detection, AIC and BIC overestimated the number of components, and ICL showed the same trend as t increased. In contrast, with heterogeneous survival, AIC and BIC were successful, whereas ICL generally underestimated the number of components, except when t increased.

STUDY CASE: CLASSIFYING INDIVIDUALS AND ESTIMATING PARAMETERS IN A LONG-LIVED SPECIES IN PRESENCE OF DETECTION HETEROGENEITY

We fitted the 1-class, 2-class and 3-class models to the wolf data and calculated AIC, BIC and ICL for each model (Table 4).

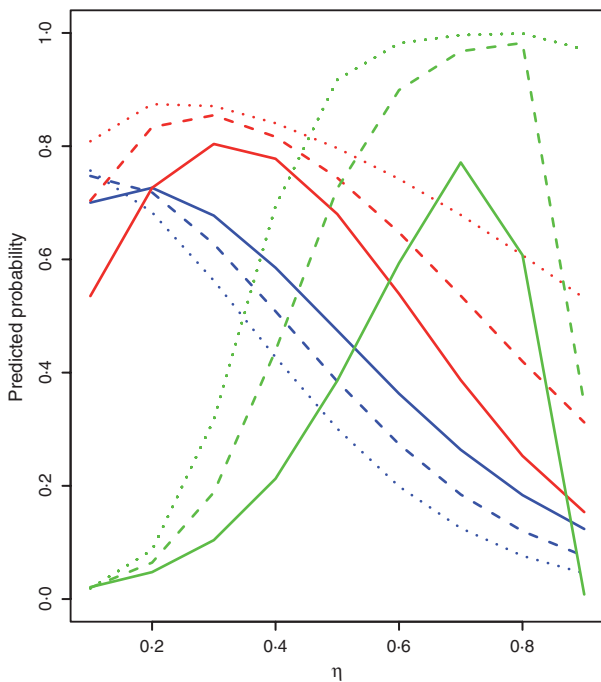


Fig. 2. Probability of selecting the correct number of mixtures as a function of the heterogeneity coefficient (η) for AIC (blue), BIC (red) and ICL (green). We assumed detection heterogeneity in a relatively high detection ($\mu = 0.7$) for a long-lived species (scenario 2). The relationships are displayed for a skewness coefficient of $\gamma = 1.5$ (solid line), $\gamma = 0$ (dashed line) and $\gamma = -1.5$ (dotted line).

Table 3. Effects of the number of individuals (n) and sampling occasions (t) on mean percentage of success, underestimation and overestimation of the number of mixture components for AIC, BIC and ICL in the presence of heterogeneous detection (scenario 2; $\eta = 0.84$; $\mu = 0.5$; $\gamma = 0$) and heterogeneous survival (scenario 4; $\eta = 0.74$; $\mu = 0.5$; $\gamma = 0$)

	Heterogeneous detection			Heterogeneous survival		
	AIC	BIC	ICL	AIC	BIC	ICL
$n = 150, t = 15$						
Success	3.4	7.4	73.4	98.8	100	25.8
Overestimating	96.6	92.6	26.4	1.2	0	0
Underestimating	0	0	0.2	0	0	74.2
$n = 150, t = 30$						
Success	0.4	0.6	11.6	97.0	99.6	89.4
Overestimating	99.6	99.4	88.4	3.0	0.4	0
Underestimating	0	0	0	0	0	10.6
$n = 300, t = 15$						
Success	0.2	0.6	71.6	97.4	99.8	20.2
Overestimating	99.8	99.4	28.0	2.6	0.2	0
Underestimating	0	0	0.4	0	0	79.8
$n = 300, t = 30$						
Success	0	0	1.4	96.4	100	77.0
Overestimating	100	99.6	98.6	3.6	0	0
Underestimating	0	0	0	0	0	23.0

AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; ICL, Integrated Completed Likelihood.

Table 4. Values of AIC, BIC and ICL for the 1-class, 2-class and 3-class model for the wolf data

	1-class model	2-class model	3-class model
AIC	1390.1	1268.3	1269.6
BIC	1396.2	1280.6	1288.1
ICL	1396.2	1368.7	1482.7

Smaller values of the criteria are highlighted in bold font. AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; ICL, Integrated Completed Likelihood.

While AIC could not distinguish between the 2-class and 3-class models ($\Delta AIC = 1.3$), BIC ($\Delta BIC = 7.5$) and ICL ($\Delta ICL = 114$) both clearly selected the 2-class model. When considering only winter sampling sessions (eight occasions), ICL selected the 1-class model, while AIC and BIC were not affected (results not shown).

Discussion

SIMULATION STUDY

Parameter bias and precision

Overall, AIC and BIC were both appropriate means of selecting the number of components that minimized bias with a reasonable precision, whereas ICL was clearly not. Bias was generally < 0.02 for AIC and BIC, while up to > 0.09 for ICL. In heterogeneous detection scenarios, BIC performed as well

as AIC and better for a short-lived species, and both AIC and BIC provided almost unbiased estimates for heterogeneous survival (Appendix S4). By contrast, ICL showed a positive bias in survival and detection estimates, especially when the degree of heterogeneity was high.

Survival is an important parameter in population demography and in evolution as a proxy for individual fitness. On the other hand, biased estimates of detection probability can lead to substantial bias in abundance estimates (Pledger 2000; Cubaynes *et al.* 2010). Therefore, these results give strong support to the use of AIC and BIC in the analysis of CR data when the focus is on parameter estimation, in line with recent recommendations in ecology (Burnham & Anderson 2002; Link & Barker 2006).

More generally, our results confirmed that CR mixture models with two or more components worked generally well at minimizing bias on survival and detection and provided a significant improvement over the 1-class homogeneous model. Nevertheless, as pointed out by Pledger (2000, 2005), situations with low and highly heterogeneous detection constitute a challenge for CR studies, particularly when a large proportion of the population has a low detection probability (positive skewness). From our simulations, this issue was more problematic for short-lived species (scenario 1; Appendix S2). Hence, such situations should be avoided by increasing sampling effort for hard-to-detect individuals.

Finally, MSE were generally small regardless of the model considered, suggesting an acceptable loss of accuracy associated with the use of heterogeneous models, at least for the scenarios considered in our simulations.

Selection of the 'true' number of components

Finite-mixture models are a powerful modelling technique in the analysis of clustered data (McLachlan & Basford 1988; Fraley & Raftery 1998) and CR mixture models hold promising applications in evolution and ecology. Examples are the identification of 'high' and 'low' quality individuals, which needs to be accounted for in senescence analyses (Péron *et al.* 2010), and the incorporation of social structure (Cubaynes *et al.* 2010) – e.g. dominant vs. subordinate (hidden) status – via the exploitation of behavioural information, which is still rarely seen in demographic studies. By construction, ICL is a good candidate model selection criterion to determine the number of components, as its penalty accounts for the classification quality. In our simulations, ICL almost never overestimated the number of components, but was rarely successful, except in heterogeneous detection in a long-lived species (Appendix S4). Although its success rate increased with the degree of heterogeneity, especially for negative skewness, ICL generally underestimated the number of components, even when the components were well separated ($\eta > 0.5$) (Table 2 and Appendix S4). While ICL appeared to work well for Gaussian mixtures (Biernacki, Celeux & Govaert 2000), results similar to ours have been observed for Poisson mixtures (McLachlan & Peel 2000; Brame, Nagin & Wasserman 2006), confirming the importance of evaluating data-specific perfor-

mance of the criteria. Hence, the ICL penalty term seems to not be suitable for CR data. This might be due to some individuals that are 'hard-to-classify', entering the data set towards the end of the study, thus increasing the entropy term.

Akaike Information Criterion and BIC generally had a higher success rate than ICL (Table 2). In contrast with ICL, AIC almost never underestimated the number of components, but it tended to overestimate this number as η increased, especially in heterogeneous detection in a long-lived species (Appendix S4). The same pattern was observed for BIC, but to a lesser extent, as BIC was more conservative than AIC in adding components.

Overall, AIC outperformed BIC and ICL in short-lived species, while BIC outperformed AIC and ICL for $\eta < 0.5$, and ICL outperformed AIC and BIC for $\eta > 0.5$ in a long-lived species, even more when skewness was negative (Appendix S4). As none of the criteria could perfectly identify the true number of components, additional research is needed to develop an optimal criterion. In particular, further work is needed to assess the benefits of modifying ICL penalty to give less weight to individuals that are 'hard-to-classify'.

Interpretation of the relative performance of the criteria

The discrepancies observed in the relative performance of the criteria to select the 'true' number of components (Appendix 4) were because of differences in the construction of the criteria as well as to specificities of CR data. The quantity of information required to distinguish CR histories arising from different components is proportional to the number of individuals (n) and to the number of 1's in each CR history, which in turns depends on the number of possible detection events over the individual lifetime. The number of possible detection events increases with the survival probability (longer CR histories), even more with the detection probability (more 1's than 0's in CR histories) and the number of sampling occasions (more chances to sample more individuals and longer CR histories). It also increases as skewness decreases (i.e. the proportion of longer CR histories increases). Consequently, scenario 2 (in which all individuals survive well, and detection is heterogeneous) was the most informative. Scenario 4 (in which detection is relatively high, and survival is heterogeneous) and scenario 3 (in which detection is relatively low and survival is heterogeneous) were less informative. Scenario 1 (in which all individuals have a reduced survival, and detection is heterogeneous) was the least informative. For each of these scenarios, situations with a negative skewness (i.e. a large proportion of the population has a high detection probability) were the most informative. To illustrate these differences among scenarios, we calculated the percentage of errors of classification (the number of individuals assigned to the wrong component, over the total number of individuals) using the estimated posterior probabilities that each individual belong to each component of the model involved in the calculation of the entropy (see Material and methods section). In a situation with $\eta = 0.84$, $\gamma = 0$ and $\mu = 0.5$, the error rate was 12.9% in scenario 2, 14.6% in scenario 3, 14.3% in scenario 4 and 23.9% in scenario 1. As

expected, in all scenarios, the error rate increased with γ , and decreased with η and μ (results not shown).

All three criteria tended to select more complex models as the amount of information increased, but to a different level depending on their penalty. AIC has the least severe penalty ($2k$), so it rarely selected the 1-class model, selected the 2-class model when the amount of information was reduced (scenario 1, and scenario 3 and 4 with $\eta < 0.5$), and selected the 3-class model when the amount of information increased (scenario 2, and to a lesser extent scenarios 3 and 4 with $\eta > 0.5$). The BIC penalty term is larger [$k \ln(n)$], so it was more conservative than AIC, but performed less well than AIC when the amount of information was reduced as it selected the 1-class model more often (scenario 1), and better than AIC when the amount of information increased as it selected the 2-class model more often (scenario 3 and 4 with $\eta > 0.5$, and scenario 2 with $\eta < 0.5$). As n increases, we expect BIC (and ICL), which involves a penalty based on sample size, to do better than AIC which would select too numerous components, although results showed that for realistic changes in n , none of the criteria was strongly affected (Table 4). ICL has the largest penalty [$k \ln(n) - 2\text{ENT}$] where the entropy quantifies the quality of the classification. Because of this additional penalty, ICL was more sensitive than AIC and BIC to the amount of information contained in the data, especially the number of sampling occasions (Table 4) and the skewness of the distribution that strongly affects the length of CR histories. When skewness was positive, a large proportion of the population had short and relatively uninformative CR histories, inflating the entropy term so that ICL underestimated the number of components. Hence, the ICL penalty was high for the less informative scenarios so that it was much more conservative than BIC and AIC and often selected the 1-class model in scenarios 1, 2 and 3. This is why ICL, in contrast with AIC and BIC, provided more biased parameter estimates. On the contrary, ICL worked well and outperformed AIC and BIC that tended to select the 3-class model in scenario 2 with $\eta > 0.5$, and even more when skewness was negative (Fig. 2). Hence, we expect the benefits of using ICL to increase with number of sampling occasions, species longevity, detection probability and even more when the skewness is negative.

CASE STUDY

These data were analysed to provide reliable estimates of survival and detection parameters, to derive reliable estimates of population size (Cubaynes *et al.* 2010). Moreover, because IH was suspected and likely to be related to social structure (dominants are supposed to be more prone to detection than others), we also aimed at identifying the 'true' number of components to allow the identification of meaningful classes of individuals sharing the same social status. In agreement with simulations mimicking this situation (scenario 2), both BIC and ICL selected the 2-class model, while AIC could not distinguish between the 2-class and the 3-class models. As it was selected by ICL, the 2-class model was expected to perform well at assigning individuals to components with high probability. This was partly con-

firmed, as those individuals known as being dominant from field observations were all assigned to the 'high detectability' class.

Limitations of the study

To deal with IH, we have focused on finite-mixture models, as they have been available for a decade, hence numerous applications and a relatively easy implementation in standard CR pieces of software. As an alternative, one could use a model with individual random effects (Royle 2008; Gimenez & Choquet 2010). This type of model assumes a continuous source of heterogeneity and leads to the concept of a mean value for the trait in the population, with some variation around this mean. However, populations often consist of finite classes of individuals (e.g. juveniles vs. adults, dominant vs. subordinate individuals, males vs. females, breeders vs. non-breeders, healthy vs. sick individuals). In such situations, heterogeneity among individuals can be explicitly dealt with in finite-mixture models. Whether individual random effects CR models perform well in this context needs to be investigated. Besides, performing model selection with random effects models is not an easy task as it involves parameters on the boundary and so renders classical inference questionable (Bolker *et al.* 2009), CR models being no exception (Gimenez & Choquet 2010). An extensive simulation study similar to ours would be useful to compare the performance of model selection criteria or hypothesis testing approaches.

There are also limitations inherent in the design of our simulation study. First, we considered data arising from a 2-class mixture only. Although different results may be expected with more classes, estimating class-specific parameters and the proportion of individuals in each mixture component is costly in terms of sampling occasions, which is clearly a constraint in CR studies, for which the time unit is often the year. Second, we considered survival and detection parameters were constant through time, but different results might be obtained with time-dependent parameters. This requires further investigation.

Finally, our choice of model selection criteria was based on the popularity of AIC and BIC and the expected better performance of ICL in the context of finite-mixture models. We acknowledge that numerous other criteria are available. In particular, the mixture regression criterion (Naik, Shi & Tsai 2007), a variant of AIC involving a penalty for poor classification quality, was developed for the simultaneous determination of the number of components and variables in finite-mixture regression models. Another candidate, the Deviance Information Criterion (Spiegelhalter *et al.* 2002), often seen as a Bayesian counterpart of AIC, is easily obtained with the population WinBUGS computer program (Spiegelhalter *et al.* 2003), although its calculation for mixture models requires amendments (Celeux *et al.* 2006).

Summary and recommendations

This study confirmed that CR mixture models are powerful tools for modelling IH. To select the number of components, we recommend making the objectives of the study explicit. We

encourage the use of AIC and BIC when the focus is on estimation and inference about the parameters. When the aim is to assign individuals to meaningful classes, we warn that none of the criteria we considered did better than the others and we suggest taking the classification quality into account by using ICL in conjunction with BIC, although it appears that further work is needed to adapt its penalty term for CR data.

Acknowledgements

The authors are indebted to all the fieldworkers who gathered the wolf biological samples and thank the associate editor and the three reviewers for their comments, which helped improving the manuscript. This work was supported by a grant from the French Research National Agency (ANR), reference ANR-08-JCJC-0028-01.

References

- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.
- Andrews, R.L. & Currim, I.S. (2003) A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, **40**, 235–243.
- Biernacki, C., Celeux, G. & Govaert, G. (2000) Assessing a mixture model for clustering with the Integrated Completed Likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**, 719–725.
- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H. & White, J.-S.S. (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, **24**, 127–135.
- Brame, R., Nagin, D.S. & Wasserman, L. (2006) Exploring some analytical characteristics of finite mixture models. *Journal of Quantitative Criminology*, **22**, 31–59.
- Bunge, J. & Barger, K. (2008) Parametric models for estimating the number of classes. *Biometrical Journal*, **50**, 971–982.
- Burnham, K.P. & Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York.
- Carothers, A. D. (1973) The effects of unequal catchability on Jolly-Seber estimates. *Biometrics*, **29**, 79–100.
- Celeux, G., Forbes, F., Robert, C. P. & Titterton, D. M. (2006). Deviance information criteria for missing data models (with Discussion). *Bayesian Analysis*, **1**, 651–706.
- Claeskens, G. & Hjort, N.L. (2008) *Model Selection and Model Averaging*. Cambridge University Press, New York.
- Cubaynes, S., Pradel, R., Choquet, R., Duchamp, C., Gaillard, J.M., Lebreton, J.-D., Marboutin, E., Miquel, C., Reboulet, A.-M., Poillot, C., Taberlet, P. & Gimenez, O. (2010) Importance of accounting for detection heterogeneity when estimating abundance: the case of French wolves. *Conservation Biology*, **24**, 621–626.
- Dorazio, R.M. & Royle, J.A. (2003) Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, **59**, 351–364.
- Fonseca, J.R.S. & Cardoso, M.G.M.S. (2007) Mixture-model cluster analysis using information theoretical criteria. *Intelligent Data Analysis*, **11**, 155–173.
- Fraley, C. & Raftery, A.E. (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis *The Computer Journal*, **41**, 578–588.
- 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., Viallefont, A., Charmantier, A., Pradel, R., Cam, E., Brown, C.R., Anderson, M.D., Bomberger Brown, M., Covas, R. & Gaillard, J.-M. (2008) The risk of flawed inference in evolutionary studies when detectability is less than one. *The American Naturalist*, **172**, 441–448.
- Johnson, J.B. & Omland, K.S. (2004) Model selection in ecology and evolution. *Trends in Ecology & Evolution*, **19**, 101–108.
- Kass, R.E. & Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.
- Keribin, C. (2000) Consistent estimation of the order of mixture models. *Sankhya Series A*, **62**, 49–66.
- Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C., Hoang, A., Gibert, P. & Beerli, P. (2001) The strength of phenotypic selection in natural populations. *The American Naturalist*, **157**, 245–261.
- Lebreton, J.D., Burnham, K.P., Clobert, J. & Anderson, D.R. (1992) Modeling survival and testing biological hypotheses using marked animals – a unified approach with case studies. *Ecological Monographs*, **62**, 67–118.
- Lebreton, J.D., Nichols, J.D., Barker, R.J., Pradel, R. & Spendlow, J.A. (2009) Modeling individual animal histories with multistate capture-recapture models. *Advances in ecological research*, **41**, 87–173.
- Link, W.A. & Barker, R.J. (2006) Model weights and the foundations of multi-model inference. *Ecology*, **87**, 2626–2635.
- Lukociene, O. & Vermunt, J.K. (2010). Determining the number of components in mixture models for hierarchical data. In *Advances in Data Analysis, Data Handling and Business Intelligence* (eds A. Fink, B. Lausen, W. Seidel & A. Ultsch), pp. 241–249. Springer-Verlag, Berlin.
- McLachlan, G.J. & Basford, K.E. (1988) *Mixture Models. Inference and Applications to Clustering*. Marcel Dekker, New York.
- McLachlan, G. & Peel, D. (2000) *Finite Mixture Models*. John Wiley, New York.
- Morgan, B.J.T. & Ridout, M.S. (2009) Estimating N: a robust approach to capture heterogeneity. In *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 1069–1080. Springer-Verlag, New York.
- Naik, P.A., Shi, P. & Tsai, C. (2007) Extending the Akaike information criterion to mixture regression Models. *Journal of the American Statistical Association*, **102**, 244–254.
- Oliver, L.J., Morgan, B.J.T., Durant, S.M. & Pettorelli, N. (2011) Individual heterogeneity in recapture probability and survival estimates in cheetah. *Ecological Modelling*, **222**, 776–784.
- Péron, G., Crochet, P.A., Choquet, R., Pradel, R., Lebreton, J.D. & Gimenez, O. (2010) Capture–recapture models with heterogeneity to study survival senescence in the wild. *Oikos*, **119**, 524–532.
- Pledger, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, **56**, 434–442.
- Pledger, S. (2005) The performance of mixture models in heterogeneous closed population capture-recapture. *Biometrics*, **61**, 868–873.
- Pledger, S. & Phillpot, P. (2008) Using mixtures to model heterogeneity in ecological capture-recapture studies. *Biometrical Journal*, **50**, 1022–1034.
- Pledger, S., Pollock, K.H. & Norris, J.L. (2003) Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, **59**, 786–794.
- Pledger, S., Pollock, K.H. & Norris, J.L. (2010) Open capture–recapture models with heterogeneity: II. Jolly–Seber model. *Biometrics*, **66**, 883–890.
- Pollock, K.H. (2002) The use of auxiliary variables in capture-recapture modeling: an overview. *Journal of Applied Statistics*, **29**, 85–102.
- Pradel, R. (2005) Multievent: an extension of multistate capture-recapture models to uncertain states. *Biometrics*, **61**, 442–447.
- Pradel, R. (2009) The stakes of capture-recapture models with state uncertainty. In *Modeling Demographic Processes in Marked Populations* (eds D.L. Thomson, E.G. Cooch & M.J. Conroy), pp. 781–795. Springer-Verlag, New York.
- Pradel, R., Choquet, R., Lima, M., Merritt, J. & Crespín, L. (2010) Estimating population growth rate from capture-recapture data in presence of capture heterogeneity. *Journal of Agricultural Biological and Environmental Statistics*, **15**, 248–258.
- Roeder, K. & Wasserman, L. (1997) Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, **92**, 894–902.
- Royle, J.A. (2008) Modeling individual effects in the Cormack-Jolly-Seber model: a state-space formulation. *Biometrics*, **64**, 364–370.
- Schwarz, G. (1978) Estimating the number of components in a finite mixture model. *Annals of Statistics*, **6**, 461–464.
- Spiegelhalter, D., Thomas, A., Best, N.G. & Lunn, D. (2003). *WinBUGS User Manual*. Version 1.4. Technical report, Medical Research Council Biostatistics, Cambridge. <http://www.mrc-bsu.cam.ac.uk/bugs> (accessed January 2003).
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van der Linde, A. (2002). Bayesian measures of model complexity and fit (with Discussion). *Journal of the Royal Statistical Society, Series B*, **64**, 583–616.
- Tuljapurkar, S., Steiner, U.K. & Orzack, S.H. (2009) Dynamic heterogeneity in life histories. *Ecology Letters*, **12**, 93–106.
- Tyrell, C.L., Christy, M.T., Rodda, G.H., Adams, A.A., Ellingson, A.R., Savidge, J.A., Dean-Bradley, K. & Bischof, R. (2009) Evaluation of trap capture in a geographically closed population of brown treesnakes on Guam. *Journal of Applied Ecology*, **46**, 128–135.
- Véran, S., Gimenez, O., Flint, E., Kendall, W.L., Doherty, P.F. & Lebreton, J.D. (2007) Quantifying the impact of longline fisheries on adult survival in the black-footed albatross. *Journal of Applied Ecology*, **44**, 942–952.

- Wanger, T.C., Motzke, I., Furrer, S.C., Brook, B.W. & Gruber, B. (2009) How to monitor elusive lizards: comparison of capture-recapture methods on giant day geckos (Gekkonidae, *Phelsuma madagascariensis grandis*) in the Masoala rainforest exhibit, Zurich Zoo. *Ecological Research*, **24**, 345–353.
- Wedel, M., Kamakura, W. & Bockenholt, U. (2000) Marketing data, models and decisions. *International Journal of Research in Marketing*, **17**, 203–208.

Received 29 April 2011; accepted 30 October 2011
Handling Editor: Robert Freckleton

Supporting Information

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

Data S1. Matlab code.

Appendix S1. Results of regressions of bias on parameter estimates for AIC, BIC and ICL.

Appendix S2. Bias on parameter estimates for AIC, BIC and ICL.

Appendix S3. Results of the generalized multinomial linear regressions showing the influence of the mean value of the heterogeneous parameter, the coefficients heterogeneity and skewness on model selection.

Appendix S4. Predicted probability of success for AIC, BIC and ICL.

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