 DOI: 10.1111/j.1541-0420.2011.01681.x

Simple Estimation and Test Procedures in Capture–Mark–Recapture Mixed Models

J. D. Lebreton,* R. Choquet, and O. Gimenez

CEFE, UMR 5175, CNRS, 1919 Route de Mende, 34293 Montpellier cedex 5, France
**email:* jean-dominique.lebreton@cefe.cnrs.fr

Summary. The need to consider in capture-recapture models random effects besides fixed effects such as those of environmental covariates has been widely recognized over the last years. However, formal approaches require involved likelihood integrations, and conceptual and technical difficulties have slowed down the spread of capture–recapture mixed models among biologists. In this article, we evaluate simple procedures to test for the effect of an environmental covariate on parameters such as time-varying survival probabilities in presence of a random effect corresponding to unexplained environmental variation. We show that the usual likelihood ratio test between fixed models is strongly biased, and tends to detect too often a covariate effect. Permutation and analysis of deviance tests are shown to behave properly and are recommended. Permutation tests are implemented in the latest version of program `E-SURGE`. Our approach also applies to generalized linear mixed models.

Key words: Capture–mark–recapture; Environmental covariates; GLMM; Mixed models; Population dynamics; Random effects.

1. Introduction

The basis of modern capture–mark–recapture (CMR) models aiming at estimating survival is the Cormack–Jolly–Seber (CJS) model (Cormack, 1964; Jolly, 1965; Seber, 1965), which considers survival and recapture probabilities varying over time. Based on generalized linear model (GLM) ideas (McCullagh and Nelder, (1989), models considering linear constraints on these two types of parameters in the CJS model have become a standard tool for data analysis in population biology (Lebreton et al., 1992). The cell probabilities of the multinomial distributions inherent in the model are functions of the parameters of interest, generally estimated by numerical maximization of the likelihood. As a typical example, survival probabilities ϕ_i can be represented as varying over time (indexed as i) in relation with values of an environmental covariate x_i through a link function in a fashion analogous to logistic regression, as $\text{logit}(\phi_i) = \mu + \beta x_i$ (North and Morgan, 1979; Clobert and Lebreton, 1985). This type of model provided for instance evidence of a relationship between survival and rain in the Sahel wintering area of the white stork population breeding in Alsace (Eastern France) (Kanyamibwa et al., 1990). A shortcoming of this quite useful formulation is to consider that the survival probability is entirely determined by the environmental covariate value. The time-dependent CJS model and these constrained models can thus be viewed as fixed effect models. However, in the CJS model, it seems natural to consider that the variation over time in survival probability can be represented by values drawn from a random distribution rather than by a fixed set of parameter values (Burnham et al., 1987; Lebreton, 1995; Barker, Fletcher, and Scofield, 2002; Burnham and White, 2002; Barry et al., 2003). In the same spirit, when variation in survival is driven by an environmental covariate, a model more general than the fixed effect model considers that the variation in only partly driven by the environmental covariate, as $\text{logit}(\phi_i) = \mu + \beta x_i + \varepsilon_i$. The last term, representing unexplained temporal variation in survival, is a random effect. Including such an effect in the model amounts to considering that the sources of variation in survival beyond that induced by the environmental covariate are reasonably randomized over the study years. Moreover, in nearly all cases, the covariate can be viewed as a proxy to the actual latent variable driving the variation in survival. In this case, adaptations of regression models accounting for error in variable are recommended (Barker et al., 2002), but the model above appears as a reasonable approximation. The need to consider random effects beside any fixed effect in capture–recapture models is well emphasized by Barry et al. (2003) and Gimenez and Choquet (2010).

The model $\text{logit}(\phi_i) = \mu + \beta x_i + \varepsilon_i$, with a fixed and a random effect, is a mixed model, in the usual sense of analysis of variance (Searle, Casella, and McCulloch, 1992), the sampling variability coming on top of that represented by ε_i, through the capture–recapture sampling. Mixed models are a natural and often relevant generalization of fixed effect models in all models in which a GLM philosophy can be developed. It seems thus natural to give a central role to mixed models in capture–recapture analyses. In theory, the likelihood of such models with random effects can be obtained by integrating the likelihood of the corresponding fixed effect model over the distribution of the random effects. In practice, technical difficulties prevent a straightforward implementation of random effects, as the multiple integration required is sufficiently involved to be in general impractical (Johnson and Hoeting, 2003; see, however, Gimenez and Choquet, 2010; Choquet and Gimenez, 2011). Burnham and White (2002) propose simple procedures (available in program MARK, White and Burnham, 1999) for estimating the parameters, deviance, and Akaike information criterion (AIC) of models with random effects. However, model selection based on an information criterion may be not fully appropriate when studying environmental covariates. First, when a single major covariate is of interest, a formal test with an a priori P-level may be preferable, in particular as a one-tailed test when a clear prediction on the direction of the covariate effect is available. This will often be the case in the context of climate change, such as, e.g., with the expected detrimental effect of an increase in sea surface temperature on the survival of king penguins *Eudyptes patagonicus* (Le Bohec et al., 2008). Second, when several covariates have to be considered, the number of parameters remains constant whatever the covariate considered. Then, information-based model selection reduces to preferring the model with the lowest deviance, and can thus produce grossly misleading results in a form of data dredging, a trap rather commonly associated to excessive confidence in hypothesis testing (Stephens et al., 2005). A set of tests corrected for multiple testing by, say, a Bonferroni correction, is then clearly preferable.

Both estimation and tests in nonlinear mixed models frequently require specialized approaches, such as Bayesian procedures (Brooks et al., 2000; Barry et al., 2003), that currently tend to limit the spread of mixed capture–recapture models among biologists (Gimenez, 2008). Bayesian approaches indeed appear as the natural framework to handle capture–recapture mixed models (King et al., 2009, p. 259 ff) and should become a privileged approach in the long run. However, difficulties with the concepts of mixed models and their implementation in WinBUGS or R (King et al., 2009, chapter 7) may discourage population biologists accustomed to the flexibility and easiness of fixed effect models implementation in software such as MARK or E-SURGE. Grueber et al. (2011) emphasize in a broader context some of the practical difficulties in using information theoretic approaches with complex models. Indeed, in a review of CMR applications linking survival to environmental variables, Grosbois et al. (2008) found only 2 papers (out of 78) considering a random effect for unexplained temporal variation. One of these papers (Milner, Elston, and Albon, 1999) considers the probability of detection is 1 and thus does not concern a CMR framework, whereas the other one (Schaub, Kania, and Koppen, 2005) provides estimates of variance components but does not consider any test or information-based model selection. Indeed, the few uses of CMR mixed models have been published by methodologists (e.g., da Silva et al. 2008), population biologists being far from having gained full autonomy. Finally, all approaches, whether based on empirical Bayes (shrunk) estimates (Burnham and White, 2002), numerical integration (Gimenez and Choquet, 2010; Choquet and Gimenez, 2011), or Monte Carlo Markov chain algorithms are based on an approximation of some kind. It is thus necessary to pursue the evaluation of such tools and bring new approaches in the toolbox.

The purpose of this article is to propose and evaluate simple estimation and test procedures for testing for fixed effects (such as a test of H_0 $\beta = 0$ in the covariate model) in presence of a random effect. The simplicity of the procedures proposed lies in that nothing is needed beyond the deviance of fixed effect models. The analysis of deviance (ANODEV) statistic (Skalski, Hoffman, and Smith, 1993) was proposed as an alternative to the usual likelihood ratio test (LRT) statistic to test for a fixed effect in presence of overdispersion, based on an asymptotic F distribution. The ANODEV statistic has also been used in an ad hoc fashion in the context we consider here, i.e., in presence of unexplained variation modeled as a random factor (Grosbois et al., 2006; Hénaux, Bregnballe, and Lebreton, 2007). Moreover some simulations empirically show that the ANODEV statistic and the corresponding thresholds deduced from a Fisher–Snedecor distribution seems to behave properly in the case we are considering (Grosbois et al., 2008).

We investigate in what follows distributional properties and behavior of the LRT and ANODEV statistics in simple CMR mixed models under simple null and alternative hypotheses, using a generalized least squares (GLS) approach. GLS have been already used in the context of random effects in CMR models, with a focus on estimation, by Burnham et al. (1987), Gould and Nichols (1998), and Burnham and White (2002). Our contribution complements that by Burnham and White (2002) in several respects:

- We incorporate in our approach the statistical dependencies between estimates of the parameters of interest and the other parameters in the model (such as recapture probabilities), a point not accounted for by Burnham and White (2002).
- We show through simple formulas how the deviances and differences in deviance are modified by the presence of a random effect, and account for the presence of overdispersion, two points not covered by Burnham and White (2002). We demonstrate in particular that the usual LRT is biased by the presence of an unaccounted random effect, and that LRT results can thus be grossly misleading.
- Based on these formulas, we obtain an explicit estimate of the temporal process variance, simpler than the numerical procedure of Burnham and White (2002).
- We concentrate on test procedures, as a complement to the focus on AIC by Burnham and White (2002), because, as discussed above, the potential effect of environmental variables will often preferably be examined through a formal test, often one tailed, rather than through an information-index-based model selection.

We also examine the performance of permutation tests. We complement formal calculations by simulations and present two illustrative examples. Although, for the sake of clarity, we restrain our attention to a single time-dependent covariate in simple CJS models, the material presented here can easily be extended to more general situations, and is also applicable to GLM models.

2. Models, Parameters, and Notation

Let $\theta = \begin{bmatrix} \eta \\ \varsigma \end{bmatrix}$ be the n x 1 column matrix (vector) of parameters of a partly time-dependent CMR model. The two submatrices η and ς have p and q parameters, respectively, i.e.,

$n = p + q$. The submatrix η is formed of the time-dependent parameters. For instance η is a set of logit-transformed time-dependent survival probabilities, and ς is a set of recapture parameters. The model is considered to be full rank, i.e., to have all its parameters separately identifiable. Generalized inverses have to be used in the less-than-full-rank case (Searle et al., 1992, p. 415). We assume overdispersion, with a coefficient denoted as c, can be present. This coefficient is estimated as $\hat{c}$, in general from goodness-of-fit procedures (Lebreton et al., 1992). If there is no overdispersion, one simply uses $c = \hat{c} = 1$ in what follows.

Block matrix notation is used throughout, with straightforward notation of dimensions.

We consider three fixed effect models (e.g., Lebreton et al., 1992):

- Model F_t, in which η is left unconstrained. The maximum likelihood estimates (MLEs) of θ is denoted as $\hat{\theta}_t$. Its full-rank estimated variance–covariance matrix is denoted as $\hat{c}\Sigma_t$, with no circumflex on the matrix for the sake of simplicity. Σ_t is derived under the assumption of independence of individuals, i.e., under the standard product-multinomial likelihood of the CJS model, and is multiplied by $\hat{c}$ to account for overdispersion.
- Model F_x in which η is constrained as a function of a covariate x. The constraint matrix for η is $\begin{bmatrix} 1_{p,1} & X \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_p \end{bmatrix}$. The overall constraint matrix for θ is, in block-matrix notation, the $(p+q) \times (q+2)$ matrix $Z = \begin{bmatrix} 1_{p,1} & X & 0_{p,q} \\ 0_{q,1} & 0_{q,1} & I_q \end{bmatrix}$. Hence, in this model, $\theta = Z \begin{bmatrix} \mu \\ \beta \\ \varsigma \end{bmatrix}$, which constrains η as $\eta = \begin{bmatrix} 1_{p,1} & X \end{bmatrix} \begin{bmatrix} \mu \\ \beta \end{bmatrix}$ and leaves ς unconstrained. The MLE of θ is $\hat{\theta}_x = Z \begin{bmatrix} \hat{\mu} \\ \hat{\beta} \\ \hat{\varsigma} \end{bmatrix}$, with estimated covariance matrix $\hat{c}\Sigma_x$ with rank q+2.
- Model F_i in which η is constrained to be constant. The variable i stands for *intercept*. The constraint matrix for θ is, in block-matrix notation, $S = \begin{bmatrix} 1_{p,1} & 0_{p,q} \\ 0_{q,1} & I_q \end{bmatrix}$. In this model, $\eta = 1_{p,1}\mu$ and ς is unconstrained. The MLE of θ is $\hat{\theta}_i = S \begin{bmatrix} \hat{\mu} \\ \hat{\varsigma} \end{bmatrix}$ with estimated covariance matrix $\hat{c}\Sigma_i$, with rank q+1.

The deviances at the MLEs are denoted as $D(\hat{\theta}_t) \leq D(\hat{\theta}_x) \leq D(\hat{\theta}_i)$, for models F_t, F_x, and F_i, with $n = p + q \geq 2 + q > 1 + q$ identifiable parameters, respectively. Again, these numbers of parameters for the full rank case can be readily adapted to the rank-deficient case. The deviances may be relative, i.e., have an arbitrary origin, as only differences in deviance will be used. These three fixed effect models can easily be fitted using standard CMR software such as `MARK` (White and Burnham, 1999), `M-SURGE` (Choquet et al., 2004), and `E-SURGE` (Choquet, Rouan, and Pradel, 2009).

Two mixed models are considered:

- Model M_i in which $\eta_i = \mu + \varepsilon_i$, $i = 1,...p$, i.e., $\eta = 1_{p,1}\mu + \varepsilon$. The random effect vector ε is assumed to be distributed as $N_p(0, \sigma^2 I_p)$. The overall model is $\theta = \begin{bmatrix} \eta \\ \varsigma \end{bmatrix} = N(\begin{bmatrix} 1_{p,1}\mu \\ \varsigma \end{bmatrix}, \sigma^2 V = \sigma^2 \begin{bmatrix} I_p & 0_{p,q} \\ 0_{q,p} & 0_{q,q} \end{bmatrix})$, with an obvious definition of V, and in which the distribution of the last q components, ς, is degenerate. This model corresponds in what follows to the absence of a covariate effect, i.e., to H_0.
- Model M_x in which $\eta = \begin{bmatrix} 1_{p,1} & X \end{bmatrix} \begin{bmatrix} \mu \\ \beta \end{bmatrix} + \varepsilon$. The random effect ε is again assumed to be distributed as $N_p(0, \sigma^2 I_p)$. If $\beta \neq 0$, this model corresponds to a covariate effect, i.e., the alternative hypothesis H_1. If $\beta = 0$, model M_x reduces to model M_i.

3. LRT and ANODEV Statistics

We want to investigate the properties of the LRT and the ANODEV statistics (Skalski et al., 1993) when used to test for H_0: $\beta = 0$ when $\sigma^2 \neq 0$, i.e., in presence of a random effect and, possibly, overdispersion. It is a test of H_0 $\beta = 0$ in model M_x, or, equivalently a test of model M_x vs model M_i although it is based on statistics derived from fitting models F_t, F_x, and F_i.

The LRT statistic is $G = D(\hat{\theta}_i) - D(\hat{\theta}_x)$. We define $H = D(\hat{\theta}_x) - D(\hat{\theta}_t)$ and will use $G + H = D(\hat{\theta}_i) - D(\hat{\theta}_t)$

The ANODEV statistic is

$$F = \frac{D(\hat{\theta}_i) - D(\hat{\theta}_x)}{(D(\hat{\theta}_x) - D(\hat{\theta}_t))/(p-2)} = \frac{G}{H/(p-2)}$$

.

The three quantities G, H, and G+H, are analogous to regression (between), within, and total sums of squares, respectively, in ordinary linear regression and we will indeed approximate them below by quadratic forms. Under the assumptions of model F_i and in presence of overdispersion only, a different context, the ANODEV statistic F is considered as asymptotically following a Fisher–Snedecor distribution $F(1, p-2)$ (Skalski et al., 1993) although, to our knowledge, no formal proof was ever published.

4. Approximating the Deviance and MLEs by a Generalized Least Squares Approach

The deviance, calculated in the same fashion under overdispersion or not, can be asymptotically approximated from the asymptotic normal distribution of the MLEs assuming no overdispersion, as minus twice the log probability density (Lebreton et al., 1995), which in the full rank case is $(\hat{\theta}_\mathrm{t} - \theta)'\Sigma_\mathrm{t}^{-1}(\hat{\theta}_\mathrm{t} - \theta) + \log(\det(\Sigma_\mathrm{t})) + (p+q)\log(2\pi)$, assuming Σ_t as known without uncertainty, an assumption asymptotically acceptable.

The GLS constrained estimates are obtained from $\hat{\theta}_t$ by applying Σ_t^{-1}-orthogonal projectors on $Im(Z)$ and $Im(S)$, $P_Z = Z(Z'\Sigma_t^{-1}Z)^{-1}Z'\Sigma_t^{-1}$ and $P_S = S(S'\Sigma_t^{-1}S)^{-1}S'\Sigma_t^{-1}$, respectively.

The MLE of θ under model F_x, $\hat{\theta}_x$ is asymptotically equivalent to $\tilde{\theta}_x = P_Z\hat{\theta}_t$, the GLS estimate derived by projection from $\hat{\theta}_t$, the MLE under model F_t: Similarly, $\hat{\theta}_i$, the MLE of

θ under model F_i, is asymptotically equivalent to the GLS estimate $\tilde{\theta}_i = P_s\hat{\theta}_t$.

The constant terms disappear in differences in deviance and one may use:

$$D(\theta) \approx (\theta - \hat{\theta}_t)\Sigma_t^{-1}(\theta - \hat{\theta}_t). \quad (1)$$

Using Σ_t^{-1}-orthogonality, one gets in turn the following asymptotic approximations:

$$\begin{aligned} G &= D(\hat{\theta}_i) - D(\hat{\theta}_x) \approx D(\tilde{\theta}_i) - D(\tilde{\theta}_x) \\ &\approx ((P_Z - P_S)\hat{\theta}_t)'\Sigma_t^{-1}((P_Z - P_S)\hat{\theta}_t), \end{aligned} \quad (2)$$

$$\begin{aligned} H &= D(\hat{\theta}_x) - D(\hat{\theta}_t) \approx D(\tilde{\theta}_x) - D(\hat{\theta}_t) \\ &= ((I - P_Z)\hat{\theta}_t)'\Sigma_t^{-1}((I - P_Z)\hat{\theta}_t) \end{aligned} \quad (3)$$

5. Results

5.1 *Distributional Results*

We get the following distributional results (see Web Appendix A).

Under H_0 $\beta = 0$:

$$(I - P_Z)\hat{\theta}_t \approx N_{p+q}(0_{p+q,1}, (I - P_Z)(c\Sigma_t + \sigma^2 V)(I - P_Z)'), \quad (4)$$

$$(P_Z - P_S)\hat{\theta}_t \approx N_{p+q}(0_{p+q,1}, (P_Z - P_S)(c\Sigma_t + \sigma^2 V)(P_Z - P_S)'), \quad (5)$$

and, under H$_1$ $\beta \neq 0$:

$$(I - P_Z)\hat{\theta}_t \approx N_{p+q}(0_{p+q,1}, (I - P_Z)(c\Sigma_t + \sigma^2 V)(I - P_Z)'), \quad (6)$$

$$\begin{aligned} &(P_Z - P_S)\hat{\theta}_t \approx N_{p+q}((P_Z - P_S)\begin{bmatrix} X\beta \\ 0_{q,1} \end{bmatrix}, \\ &(P_Z - P_S)(c\Sigma_t + \sigma^2 V)(P_Z - P_S)'). \end{aligned} \quad (7)$$

One can show then (see Web Appendix A) that, if $\sigma^2 > 0$, the quadratic forms (2) and (3) cannot be distributed as proportional to χ^2. Thus, even in the absence of overdispersion, the LRT statistic, G, does not follow under the mixed model assumptions its usual asymptotic χ^2 distribution valid under the fixed model assumptions. Similarly, no distributional results such as noncentral chi-squared distributions can be obtained under H$_1$. Indeed, Σ_t is not a multiple of the identity matrix, inducing complications similar to those arising in random or mixed ANOVA models with unbalanced data (Searle et al., 1992, p. 76).

When $\sigma^2 = 0$, i.e., under fixed effects only and in presence of overdispersion:

Under H$_0$ $\beta = 0$:

$$(I - P_Z)\hat{\theta}_t \approx N_{p+q}(0_{p+q,1}, c(I - P_Z)\Sigma_t(I - P_Z)'), \quad (8)$$

$$(P_Z - P_S)\hat{\theta}_t \approx N_{p+q}(0_{p+q,1}, c(P_Z - P_S)\Sigma_t(P_Z - P_S)'), \quad (9)$$

and, under H$_1$ $\beta \neq 0$:

$$(I - P_Z)\hat{\theta}_t \approx N_{p+q}(0_{p+q,1}, c(I - P_Z)\Sigma_t(I - P_Z)'), \quad (10)$$

$$\begin{aligned} &(P_Z - P_S)\hat{\theta}_t \approx N_{p+q}((P_Z - P_S)\begin{bmatrix} X\beta \\ 0_{q,1} \end{bmatrix}, \\ &c(P_Z - P_S)\Sigma_t(P_Z - P_S)'). \end{aligned} \quad (11)$$

Then, under both H$_0$ and H$_1$, H is asymptotically distributed as c times a chi-squared distribution with degrees of freedom (df) equal to $tr(I - P_Z) = p - 2$, hence with expectation $c(p - 2)$. Similarly, under H$_0$, G is asymptotically distributed as c times a chi-squared distribution with df equal to $tr(P_Z - P_S) = 1$, hence with expectation c. Cochran's theorem in its general form (Rao, 1952, p. 55 (iii)) ensures the independence of the chi-squared distributions to which G and H are proportional. Under H$_1$, the nonnull expectation in (11) shifts the quadratic form distribution by

$$\begin{aligned} U &= \beta^2 \begin{bmatrix} X' & 0 \end{bmatrix} (P_Z - P_S)'\Sigma_t^{-1}(P_Z - P_S)\begin{bmatrix} X \\ 0 \end{bmatrix} \\ &= \beta^2 \begin{bmatrix} X' & 0 \end{bmatrix} \Sigma_t^{-1}(P_Z - P_S)\begin{bmatrix} X \\ 0 \end{bmatrix} \end{aligned}$$

alike a weighted version of $\beta^2 SSE(X)$ in ordinary regression (see below). G is then asymptotically distributed as c times a noncentral chi-squared distribution with 1 df and noncentrality factor U/c (Rao, 1952, p. 57 (vii)). It follows that the ANODEV statistic F is, under the fixed model assumptions with overdispersion, asymptotically distributed as a Fisher–Snedecor distribution with 1 and p-2 df, centrally and noncentrally, under H$_0$ and H$_1$, respectively. This provides to our knowledge the first formal proof of the rationale for using ANODEV in presence of overdispersion, as proposed by Skalski et al. (1993). The central point is that $H/(p - 2)$ provides an estimate of the overdispersion coefficient c asymptotically independent of G. When another independent estimate of c is available, derived, e.g., from goodness-of-fit statistics, one can pool the estimates and modify the F-test accordingly. The increase in the df of the denominator will tend to increase power. However, data sparseness often tends to bias toward low values the estimate of the overdispersion coefficient deduced from goodness-of-fit statistics. When sparseness is a concern, the original ANODEV test might then be preferable, and the user will have to exert his or her judgment. We show below that the issue is also complicated by potential unexplained environmental variation, i.e., when $\sigma^2 > 0$.

5.2 *Expectations under Mixed Models*

In the absence of straightforward distributional results under the mixed model assumptions, one remaining possibility is to get expectations. A classical result on the expectation of quadratic forms and trace is (Seber, 1977, p. 13):

if$E(Y) = \nu$ and if the covariance matrix of Y is Σ, then: $E(Y'AY) = tr(A\Sigma) + \nu' A\nu$

Then, under H$_0$:

$$E(H) = tr(\Sigma_t^{-1}(I - P_Z)'(c\Sigma_t + \sigma^2 V)(I - P_Z)), \quad (12)$$

$$E(G) = tr(\Sigma_t^{-1}(P_Z - P_S)'(c\Sigma_t + \sigma^2 V)(P_Z - P_S)), \quad (13)$$

and, under H$_1$:

$$E(H) = tr(\Sigma_t^{-1}(I - P_Z)'(c\Sigma_t + \sigma^2 V)(I - P_Z)), \quad (14)$$

$$E(G) = tr\left(\Sigma_t^{-1}(P_Z - P_S)'(c\Sigma_t + \sigma^2 V)(P_Z - P_S)\right) + U, \quad (15)$$

with again $U = \beta^2 \left[X' \; 0 \right] \Sigma_t^{-1}(P_Z - P_S) \begin{bmatrix} X \\ 0 \end{bmatrix}$.

After some algebra, under H_0:

$$E(G) = c + \sigma^2 tr(V\Sigma_t^{-1}(P_Z - P_S)), \tag{16}$$

and, under both H_0 and H_1:

$$E(H) = c(p-2) + \sigma^2 tr(V\Sigma_t^{-1}(I - P_Z)). \tag{17}$$

As a consequence, under H_0, $E(G+H) = c(p-1) + \sigma^2 tr(V\Sigma_t^{-1}(I - P_S))$.

As apparent in (16), the expectation of the LRT statistic, G, is thus inflated compared with its value under the fixed model assumptions as soon as $\sigma^2 > 0$. It departs thus from the fixed model χ^2 distribution in a severe fashion, even in the absence of overdispersion.

The relevance of equations (16) and (17) to the conjecture $E(G)/E(H)/(p-2) \approx 1$ under H_0, as expected for an F-statistic under H_0, comes from $tr(P_Z - P_S) = rk(P_Z - P_S) = 1$ and $tr(I - P_Z) = rk(I - P_Z) = p - 2$. Were $V\Sigma_t^{-1}$ equal to αI_{p+q}, the result would hold exactly. One can thus expect it to hold approximately. Under H_1, $E(G)$ is further increased by a term proportional to β^2, and in turn, as expected for a test statistic in this case, $E(G)/E(H)/(p-2)$is a monotonously increasing function of $|\beta|$. Often, a one-sided test will be relevant, based on predictions of the expected direction of an environmental covariate effect on the demographic trait under investigation.

The distribution of the ANODEV statistic when used to test for a covariate effect can thus be expected to be approximated at least grossly by a Fisher–Snedecor distribution, whatever the inflation of the LRT statistic induced by unexplained environmental variation and the level of overdispersion. Because we are dealing with statistical tests, the quality of the approximation has in particular to be checked for the tails of the distribution, and is the subject of simulations below.

Provided an estimate $\hat{c}$ of c is available, equations (16) and (17) lead also to straightforward empirical estimates of σ^2. One gets the estimates $\sigma^2 = \frac{G+H-\hat{c}(p-1)}{tr(V\Sigma_t^{-1}(I-P_S))}$ and $\sigma^2 = \frac{H-\hat{c}(p-2)}{tr(V\Sigma_t^{-1}(I-P_Z))}$, under H_0 and H_1, respectively. In the absence of overdispersion, one has to use $\hat{c} = 1$. As underdispersion is infrequent in practice (McCullagh and Nelder, 1989, p. 73) and cannot be generated by dependencies or heterogeneities among individuals, there are no reasons to use an estimate $\hat{c} < 1$.

5.3 *Permutation Distributions*

Another approach is to obtain the distribution of the statistic of interest over permutations of the fixed effect X, i.e., of the p covariate values. The resulting test is distribution free and bears a relationship to the previous approach by the fact that the usual F-tests in linear models are asymptotically permutation tests (Kazi-Aoual et al., 1995). The combinatorial approach by Kazi-Aoual et al. (1995) failed here to lead to explicit results for the moments of the ANODEV and LRT statistics under the permutation distribution. One thus has to use a complete enumeration of all permutations when feasible, or a random subsample of the $p!$ permutations. $D(\hat{\theta}_t)$ and $D(\hat{\theta}_i)$, the deviances of the fixed models F_t and F_i, respectively, are fixed under any permutation of X. The ANODEV F-statistic is then a monotonous function of the deviance of the fixed model with covariate $D(\hat{\theta}_x)$. Hence $D(\hat{\theta}_x)$ or the LRT-statistic $G = D(\hat{\theta}_i) - D(\hat{\theta}_x)$ can be used equivalently to the ANODEV statistic in a permutation test. The paradox is solved by noticing that the expectation of the LRT statistic is widely inflated by the random effect, and cannot be referred to a χ^2, even as an approximation, whereas the expectation results and, as shown below, simulation results suggest the Fisher–Snedecor distribution is a reasonable approximation for the ANODEV statistic. However the threshold of the ANODEV and LRT statistics and, as a consequence, that of the conditional ANODEV–LRT test lead to the same rejection and acceptance regions under permutation distributions. The performance of permutation tests is examined in the two examples below.

5.4 *Simulation Results*

We present in Web Appendix B an examination through simulation of the quality of the Fisher–Snedecor approximation for the ANODEV statistic, by investigating the P-levels obtained by using rejection thresholds based on the Fisher–Snedecor distribution. We also check the performance of the LRT, and that of the conditional test using either the ANODEV or the LRT, respectively, depending upon $H > p - 2$ or $H \leq p - 2$, respectively.

In presence of unexplained environmental variation:

- The level of the ANODEV always remain close to nominal;
- The level of the LRT is biased whatever the amount of data and time intervals;
- The conditional test brings no further advantage: i.e., the moderate departure of ANODEV from the nominal level does not seem to result from an accidentally low denominator; and
- These results do not depend on the continuous or discrete character of the covariate.

The departure of the LRT from the nominal test level rapidly increases with the process standard error σ. The ANODEV resists much better, with again, no clear advantage brought by the conditional test, and no difference between a continuous and discrete covariate. As expected, the LRT can thus be quite misleading, rejecting too often the null hypothesis.

The power of ANODEV rapidly increases with the slope value, and is quite high even under high levels of unexplained environmental variation. Under $\sigma > 0$, the apparent power of the LRT is high even for low values of the slope, because, as the P-level, it is biased, a result emphasizing again the risk of misleading conclusions.

6. Illustrative Examples

6.1 *Dipper*

Our first example is based on the CMR data obtained by Marzolin (1988) in a study of the European dipper *Cinclus cinclus*, and used by Lebreton et al. (1992) as a simple example of constrained CMR modeling (see also Marzolin, Charmantier, and Gimenez, 2011, for a treatment of more extensive data). A total of 254 adult dippers were ringed, released, and

Table 1
Results for the three fixed effect and the two CMR mixed models for the dipper data (see text)

Survival probability	Model	Relative deviance DEV	Number of identifiable parameters NP	AIC= DEV+2 NP
Time dependent	F_t	659.730	7	673.730
Logit-linear relationship with flood covariate	F_x	660.103	3	666.103
Constant	F_i	666.838	2	670.838
Mixed model with flood covariate	M_x	660.103*	3*	666.103
Mixed model, constant	M_i	661.274*	4.531*	670.336

*Based on the approach by Burnham and White (2002).

Table 2
Results for the three fixed effect and the two CMR mixed models for the white stork data (see text)

Survival probability	Model	Relative deviance DEV	Number of identifiable parameters NP	AIC= DEV+2 NP
Time dependent	F_t	1314.760	17	1348.760
Logit-linear relationship with Sahel rain covariate	F_x	1345.318	3	1351.318
Constant	F_i	1352.082	2	1354.082
Mixed model with Sahel rain covariate	M_x	1317.751*	12.968*	1343.687
Mixed model, constant	M_i	1317.362*	13.149*	1343.660

*Based on the approach by Burnham and White (2002).

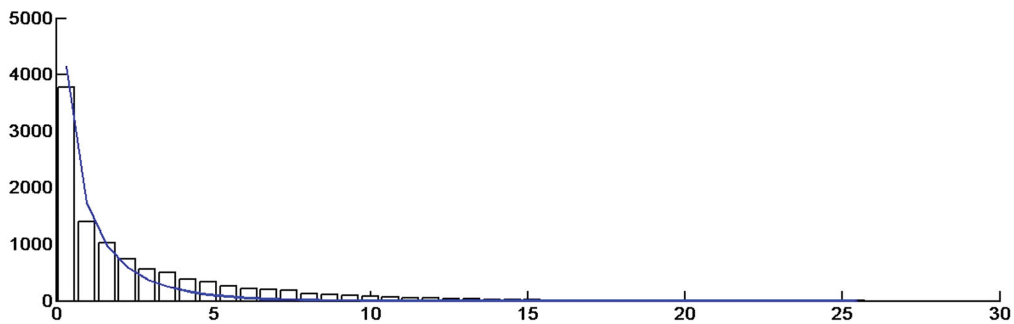

Figure 1. Permutation distribution of the LRT statistic for the effect of the Sahel rain covariate on white stork survival, referred to the standard $\chi^2(1)$ distribution. The inflation of the LRT distribution due to the presence of unexplained variation among annual survival probabilities is obvious.

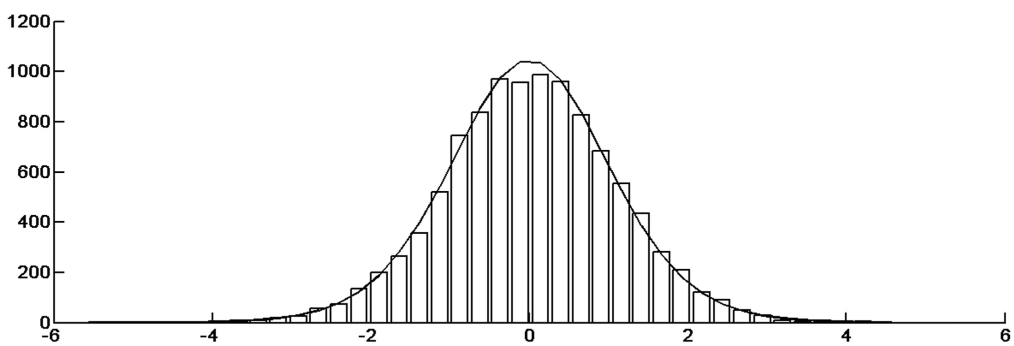

Figure 2. Permutation distribution of the signed square-root ANODEV statistic for the effect of the Sahel rain covariate on white stork survival, referred to a Student distribution $t(p\text{-}2) = t(14)$. The unexplained environmental variation is totally accounted for, as shown by the excellent match.

recaptured each spring from 1981 to 1987, i.e., over seven occasions of capture (six intervals). We start from model F_t with time dependence in survival and constant probability of capture, which fits the data, to investigate if survival probabilities over the second and third intervals were influenced by a flood that took place around the third occasion of capture. This potential flood impact can be examined as a potential logit-linear effect of an indicator variable on time-dependent survival. The CMR data and covariate values are given in Web Appendix C. The discreteness of the environmental covariate, the moderate number of individuals released, and the low number of occasions make this example an extreme case for applying the techniques proposed here. From the table of the three relevant fixed effect models (F_t, F_x, and F_i; Table 1), fitted using `M-SURGE` (Choquet et al., 2004), one gets $G = D(F_i) - D(F_x) = 6.7349$, df = 1 and $H = D(F_x) - D(F_t) = 0.3725$, df = p-2 = 4, from which one deduces $F = 72.3211$. The quadratic approximations (equations 2 and 3) compare quite favorably, with $G = 6.6894$, $H = 0.3755$ and, in turn, $F = 71.2586$. The LRT between models F_t and F_x ($\chi^2(4) = 0.0375$, P = 0.9998) indicates there is nearly no unexplained variation in survival left beyond the flood effect. This low value is indeed extreme and is discussed below. The LRT is thus adequate, and is indeed selected in the conditional test. It indicates a significant flood effect ($\chi^2(1) = 6.7349$, P = 0.0095). The probability level of the ANODEV statistic based on the Fisher–Snedecor distribution is $Pr(F(1,4) > 72.3211) = 0.0010488$. An interesting feature of this example is that because of the low number of occasions, the distribution of the statistics over all 6! = 720 permutations of the covariate values over the years can be examined. The P-level of the permutation test is 0.0667, as 48 out of the 720 permutations have an F-value at least equal to the observed one. This P-level is affected by the discreteness of the covariate, as the same maximum F value is obtained for the 48 = 2! × 4! permutations that place the two values of the covariate equal to one on intervals 2 and 3. In spite of this discreteness problem, it is clear here that in relation with the evidence for the absence of unexplained environmental variation, the LRT behaves better than the ANODEV, with a P-level closer to the permutation test P-level. The estimated variance $\sigma^2 = \frac{H - \hat{c}(p-2)}{tr(V\Sigma_t^{-1}(I-P_Z))} = \frac{0.3725-4}{tr(V\Sigma_t^{-1}(I-P_Z))} = \frac{-3.6275}{56.1921} = -0.0646$ is negative, leading to assume $\sigma^2 = 0$. This will always happen for $H < p$-2, i.e., when the LRT between models F_t and F_x points to the absence of unexplained environmental variation and overdispersion. The Burnham and White (2002) estimate of the process variance, −0.0511, is also negative, leading too to assume $\sigma^2 = 0$. One may thus consider that the survival probability is entirely determined by the flood effect. The AIC for model M_x is then indeed equal to that of model F_x (Table 1). Indeed, model M_x under $\sigma^2 = 0$ is not anymore nested in model M_i under $\sigma^2 \neq 0$ and the resulting number of identifiable parameters is lower for model M_x than for model M_i.

The origin of the low H value is unknown. In this particular case as well as in a general fashion, there is no reason to believe in underdispersion: some departure from independence between individuals is expected, in particular because of a tendency to simultaneously recapture both members of each pair, but would tend then to induce overdispersion, by a factor close to 2. The conditional test here protects again an optimistic P-level by ANODEV, and appears thus as conservative. From our simulations, we conclude it will be a fairly rare case. Finally, the smooth tail of the F-distribution does not perfectly approximate the extremely discrete permutation distribution of the ANODEV statistic, as would happen in any context, e.g., in that of usual linear regression.

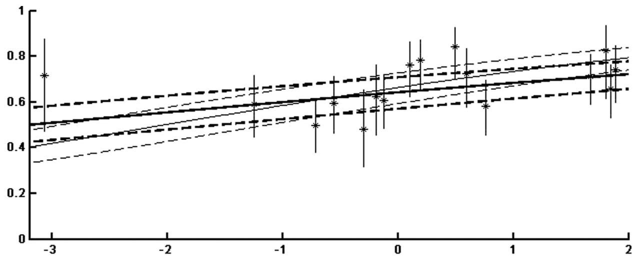

Figure 3. relationship between the first PCA of Sahel rain amounts (abscissa) and estimated survival probability (ordinate) (see Table 3). Points: estimates from model (φ_t, p) also denoted as F_t; Plain line: logit-linear relationship; Dotted lines: logit-linear relationship ±1 and ±2 random unexplained environmental variation standard error.

6.2 *White Stork*

Our second example concern resightings as breeders of white storks *Ciconia ciconia* in Bäden-Wurttemberg (Kanyamibwa, Bairlein, and Schierer, 1993), kindly made available by F. Bairlein. The birds were ringed as chicks and the first resighting as a breeder was considered as the first capture. The resulting data consist of 321 individuals over 17 years, from 1956 to 1972. In relation with a strong decline in the west European populations of the white stork in the 1960s, several papers examined the relationship between survival and measures of rain amount in the Sahel (Kanyamibwa et al., 1990, 1993). We use here as a covariate the first component of a principal component analysis (PCA) of five standardized annual amounts of rain (in Kayes, Segou, Sikasso, Tombouctou, and Mopti) that is positively correlated with all five variables (Grosbois et al., 2008, p. 263). The CMR data and the covariate are given in Web Appendix C. One expects that increasing drought, i.e., a decrease in the amount of rain, tends to decrease survival, inducing thus a positive slope (for further analyses of the same data, see Grosbois et al., 2008; Gimenez et al., 2009). The covariate being continuous and the number of occasions fairly large, one expects a fair behavior of the approaches proposed here, in particular ANODEV, despite a moderate number of individuals. The overall goodness-of-fit test ($\chi^2(42) = 38.2147$, P = 0.6379) and its components indicated that the CJS model (φ_t, p_t) was a satisfying starting point from which to proceed to model selection. In a first step, the capture probability could be considered as constant, i.e., model (φ_t, p) is our model F_t in what follows.

For the fixed models (Table 2), $G = D(F_i) - D(F_x) = 6.7640$, df = 1, and $H = D(F_x) - D(F_t) = 30.5580$, df = p-2 = 14, from which $F = 6.7640/(30.5580/14) = 3.0989$. The quadratic approximations (equations 2 and 3) are reasonably close ($G = 6.7563$ and $H = 27.0055$) and lead to $F = 3.5026$.

The P-levels of ANODEV and the LRT, at this stage as two-tailed tests, are equal to 0.1002 and 0.0053, respectively, showing thus a major discrepancy. However the LRT comparing F_x and F_t ($\chi^2(14) = H = 30.5580$, P= 0.0064) clearly indicates some among-year variation unexplained by the covariate. The significance of this test invalidates model F_x and the LRT between model F_x and F_i, the LRT chi-squared distribution being valid only under the assumption that both models

Table 3
Estimates from fixed and mixed CMR models, based on various approaches, for the white stork data

Estimates	Fixed model F_x	Burnham and White (2002)	GLS approximation (this article)	Bayesian (MCMC) approach
Intercept	0.634	0.647	0.585	0.675
Slope of rain index+	0.177	0.164	0.183	0.166
Process s.e.	—	0.335	0.341	0.356
Process variance	—	0.113	0.116	0.127

fit the data. The ANODEV–LRT conditional test points here to the ANODEV. The permutation distribution (sampled over 10,000 random permutations) clearly shows the inflation of the LRT statistic compared to the chi-squared distribution (Figure 1). A test assuming the existence of a random effect on top of the potential variation induced by the covariate is needed, and this is where the ANODEV statistic is useful.

We obtained a permutation P-level of the ANODEV (and LRT) statistics equal to 0.1013. As it was estimated from 10,000 random permutations, its standard error, based on a binomial distribution, was $\approx \frac{\sqrt{1013}}{10000} = 0.032$. The resulting 95 % confidence interval [0.0951, 0.1075] contains the Fisher–Snedecor P-level, 0.1002. In this example, under H_0, based on equations (16) and (17), the ratio $\frac{E(G)}{E(H)/(p-2)}$ increases from 1 to 1.04 when σ varies from 0 to 0.4, remaining thus very close to 1. The F-test can be converted into a t-test, as $t = sgn(\hat{\beta})\sqrt{F}$, suitable for one-tailed tests. Indeed, the permutation distribution of the t-statistic closely matches a Students distribution with 14 df (Figure 2), confirming that the ANODEV adequately accounts for unexplained random variation in logit-linear relationships. The one-tailed resulting P-level is 0.0506 (permutation test) or 0.0501 (Students distribution), at the limit of significance at the usual 0.05 level. The resulting evidence for an effect of rain is thus much weaker than under the—misleading—usual chi-squared based LRT.

Burnham and White (2002) use shrunk estimates of survival probabilities to obtain the deviance of mixed models and a matrix trace to obtain an equivalent number of parameters. An approximate AIC value can be deduced from these two quantities. Although the resulting AIC points to the mixed models (with or without covariate), the approximation inherent in the shrunk estimates gives ambiguous results as the estimated deviance of the mixed model without covariate is lower than that for the model with a covariate, although the latter is nested in the former (Table 2). Hence, the difference in deviance between models M_x and M_i cannot be used to get an equivalent of our tests.

The variance component estimate $\sigma^2 = \frac{H - \hat{c}(p-2)}{tr(V\Sigma_t^{-1}(I-P_Z))} = \frac{30.5580-14}{142.8288} = 0.1159$ is valid even in presence of a covariate effect. This process variance and the intercept and slope estimates closely match the estimates based on more sophisticated approaches (Table 3).

The relationship between survival and the covariate based on the fixed model estimates and our explicit estimate of the process variance is given in Figure 3. In this example, using blindly the LRT would have lead to a strongly exaggerated level of significance for the covariate of interest and the AIC would have lead to prefer a model without covariate, whereas the ANODEV test based on a Fisher–Snedecor distribution and the permutation test provide in a coherent fashion fair evidence for a covariate effect.

7. Discussion

For biologists to extract the full information from mark–recapture studies, it is important that software for random effects models is developed and is widely distributed (Barker et al., 2002). For such progress, we advocate the use of the simple and straightforward approaches used here, which amount to using asymptotic normal distributions to simplify otherwise impractical likelihood integrations. In particular, the ANODEV F-test will clearly behave properly in most cases, and the amount of departure from the nominal levels will be quite alike that observed in unbalanced linear mixed models. The dipper example is a counterexample, in which the severe appearance of underdispersion would deserve a deeper investigation.

Altogether we recommend a permutation test is used whenever possible, and the ANODEV F-test (or its t-test version when a one-sided test is suitable) is used in other cases. Permutation tests are implemented in the current version of `E-SURGE` (Choquet et al. (2009).

Following Skalski et al. (1993), and as proved for the first time formally here, the ANODEV also offers protection against overdispersion.

Although the statistics used here are those obtained from the fixed effect models, the conclusions can be markedly different from that based on a naive AIC-based model selection among fixed models. Although Burnham and White (2002) provide approaches to AIC for mixed capture–recapture models, the approximation inherent in their approach and the focus on covariate effects may often lead to prefer formal tests to information theoretic approaches. Although the interest of the GLS approach has been largely recognized for estimation purposes in CMR models, to our knowledge, it had never been used to investigate properties of test procedures. Finally we note that our approach equally applies to GLM as our results rely on general quantities such as the deviance and rank of models, and never on the peculiarities of CMR likelihoods.

8. Supplementary Materials

Web appendices on (A) distributional results, (B) simulation results, and (C) the data used in examples referenced in sections 5 and 6 are available under the Paper Information link at the *Biometrics* website http://www.biometrics.tibs.org/.

Acknowledgements

This research was supported by a grant (reference: ANR-08-JCJC-0028-01) from the "Jeunes Chercheuses et Jeunes Chercheurs" program of the French ANR to OG on mixed CMR models. We thank G. Marzolin and F. Bairlein for having made available the dipper and white stork data sets, respectively. We acknowledge comments by two anonymous referees that greatly helped improving this article.

References

Barker, R., Fletcher, D., and Scofield, P. (2002). Measuring density dependence in survival from mark-recapture data. *Journal of Applied Statistics* **29**, 305–313.

Barry, S. C., Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2003). The analysis of ring-recovery data using random effects. *Biometrics* **59**, 54–65.

Brooks, S. P., Catchpole, E. A., Morgan, B. J. T., and Barry, S. C. (2000). On the Bayesian analysis of ring recovery data. *Biometrics* **56**, 951–956.

Burnham, K. P. and White, G. C. (2002). Evaluation of some random effects methodology applicable to ringing data. *Journal of Applied Statistics* **29**, 245–262.

Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., and Pollock, K. H. (1987). *Design and analysis methods for fish survival experiments based on release-recapture*. Bethesda, Maryland: American Fisheries Society.

Choquet, R. and Gimenez, O. (2011). Towards built-in capture–recapture mixed models in program `E-SURGE`. *Journal of Ornithology* in press, DOI 10.1007/s10336-010-0613-x.

Choquet, R., Reboulet, A.-M., Pradel, R., Gimenez, O., and Lebreton, J.-D. (2004). `M-SURGE`: New software specifically designed for multistate capture-recapture models. *Animal Biodiversity and Conservation* **27**, 207–215.

Choquet, R., Rouan, L., and Pradel, R. (2009). Program `E-SURGE`: A software application for fitting multievent models. In *Environmental and Ecological Statistics*, Thomson D. L., Cooch E. G. and Conroy M. J. (eds), Vol**.** **3**, 845–865. New-York: Springer.

Clobert, J. and Lebreton, J.-D. (1985). Dependence on environmental factors in estimation of survival rates using capture-recapture (in French). *Biometrics* **41**, 1031–1037.

Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika* **51**, 429–438.

da Silva, C. Q., Martins, E. G., Bonato, V., and dos Reis, S. F. (2008). Bayesian capture-recapture analysis: An application in modeling semelparity of a neotropical didelphid marsupial. *Communications in Statistics-Simulation and Computation* **37**, 816–828.

Gimenez, O. (2008). Discussion: Towards a Bayesian analysis template? *The Canadian Journal of Statistics* **36**, 21–24.

Gimenez, O., Bonner, S. J., King, R., Parker, R. A., Brooks, S. P., Jamieson, L. E., Grosbois, V., Morgan, B. J. T., and Thomas, L. (2009). WinBUGS for population ecologists: Bayesian modeling using Markov Chain Monte Carlo methods. In *Environmental and Ecological Statistics*, Thomson D. L., Cooch E. G. and Conroy M. J. (eds), Vol. **3**, 883–915. New-York: Springer.

Gimenez, O. and Choquet, R. (2010). Incorporating individual heterogeneity in studies on marked animals using numerical integration: Capture-recapture mixed models. *Ecology* **91**, 951–957.

Gould, W. R. and Nichols, J. D. (1998). Estimation of temporal variability of survival in animal populations. *Ecology* **79**, 2531–2538.

Grosbois, V., Henry, P. Y., Blondel, J., Perret, P., Lebreton, J. D., Thomas, D. W., and Lambrechts, M. M. (2006). Climate impacts on Mediterranean blue tit survival: An investigation across seasons and spatial scales. *Global Change Biology* **12**, 2235–2249.

Grosbois, V., Gimenez, O., Gaillard, J. M., Pradel, R., Barbraud, C., Clobert, J., Moller, A. P., and Weimerskirch, H. (2008). Assessing the impact of climate variation on survival in vertebrate populations. *Biological Review* **83**, 357–399.

Grueber, C. E., Nakagawa, S., Laws, R. J., and Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology* **24**, 699–711.

Hénaux, V., Bregnballe, T., and Lebreton, J.-D. (2007). Dispersal and recruitment during population growth in a colonial bird, the great cormorant *Phalacrocorax carbo sinensis*. *Journal of Avian Biology* **38**, 44–57.

Johnson, D. S. and Hoeting, J. A. (2003). Autoregressive models for capture-recapture data: A Bayesian approach. *Biometrics* **59**, 340–349.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika* **52**, 225–247.

Kanyamibwa, S., Schierer, A., Pradel, R., and Lebreton, J.-D. (1990). Changes in adult annual survival rates in a Western European population of the white stork (*Ciconia ciconia*). *Ibis* **132**, 27–35.

Kanyamibwa, S., Bairlein, F., and Schierer, A. (1993). Comparison of survival rates between populations of white stork in central Europe. *Ornis Scandinavica* **24**, 297–302.

Kazi-Aoual, F., Hitier, S. Sabatier, R., and Lebreton, J.-D. (1995). Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis* **20**, 643–656.

King, R., Morgan, B. J. T. Gimenez, O., and Brooks, S. P. (2009). *Bayesian Analysis for Population Ecology*. Boca Raton, Florida: Chapman & Hall/CRC.

Le Bohec, C., Durant, J. M., Gauthier-Clerc, M., Stenseth, N. C., Park, Y. H., Pradel, R., Gremillet, D., Gendner, J. P., and Le Maho, Y. (2008). King penguin population threatened by Southern Ocean warming. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 2493–2497.

Lebreton, J.-D. (1995). The future of population dynamic studies using marked individuals: A statistician's perspective. *Journal of Applied Statistics* **22**, 1009–1030.

Lebreton, J.-D., Burnham, K. P., Clobert, J., and 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., Morgan, B. J. T., Pradel, R., and Freeman, S. N. (1995). A simultaneous analysis of dead recovery and live recapture data. *Biometrics* **51**, 1418–1428.

Marzolin, G. (1988). Polygyny of the European Dipper (Cinclus cinclus) in the Lorraine hills (in French). *L'Oiseau et la revue Française d'Ornithologie* **58**, 277–286.

Marzolin, G., Charmantier, A., and Gimenez, O. (2011). Frailty in state-space models: Application to actuarial senescence in the dipper. *Ecology* **92**, 562–567.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. New York: Chapman and Hall.

Milner, J. M., Elston, D. A., and Albon, S. D. (1999). Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of Soay sheep. *Journal of Animal Ecology* **68**, 1235–1247.

North, P. M. and Morgan, B. J. T. (1979). Modelling heron survival using weather data. *Biometrics* **35**, 667–681.

Rao, C. R. (1952). *Advanced Statistical Methods in Biometric Research*. New-York: Wiley.

Schaub, M., Kania, W., and Koppen, U. (2005). Variation of primary production during winter induces synchrony in survival rates in migratory white storks *Ciconia ciconia*. *Journal of Animal Ecology* **74**, 656–666.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*. New-York: Wiley.

Seber, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika* **52**, 249–259.

Seber, G. A. (1977). *Linear Regression Analysis*. New York: Wiley.

Skalski, J. R., Hoffman, A. and Smith, S. G. (1993). Testing the significance of individual- and cohort-level covariates in animal survival studies. In *Marked Individuals in the Study of Bird Population*, J. D. Lebreton and P. M. North (eds), 9–28. Basel, Switzerland: Birkhäuser Verlag.

Stephens, P. A., Buskirk, S. W., Hayward, G. D., and Del Rio, C. M. (2005). Information theory and hypothesis testing: A call for pluralism. *Journal of Applied Ecology* **42**, 4–12.

White, G. C. and Burnham, K. P. (1999). Program `MARK`: Survival estimation from populations of marked animals. *Bird Study* **46**, 120–139.

Received August 2010*. Revised July* 2011*.*
Accepted July 2011*.*