

R2ucare: An R package to perform goodness-of-fit tests for capture-recapture models

Olivier Gimenez, Jean-Dominique Lebreton, Rémi Choquet, Roger Pradel

CEFE UMR 5175, CNRS—Université de Montpellier—Université Paul-Valéry Montpellier—EPHE, 1919 Route de Mende, 34293 Montpellier cedex 5, France

Summary:

1. Assessing the quality of fit of a statistical model to data is a necessary step for conducting safe inference.
2. We introduce R2ucare, an R package to perform goodness-of-fit tests for open single- and multi-state capture-recapture models. R2ucare also has various functions to manipulate capture-recapture data.
3. We remind the basics and provide guidelines to navigate towards testing the fit of capture-recapture models. We demonstrate the functionality of R2ucare through its application to real data.

Keywords: Arnason-Schwarz, capture-mark-recapture, Cormack-Jolly-Seber, model validation, R2ucare

Introduction

Capture–recapture (CR) models have become a central tool in population ecology for estimating demographic parameters under imperfect detection of individuals (Lebreton et al. 1992; 2009). These methods rely on the longitudinal monitoring of individuals that are marked (or identifiable) and then captured or sighted alive over time.

Single-state CR models, and the Cormack-Jolly-Seber (CJS) model in particular (Lebreton et al. 1992), have been used to assess the effect of climate change (e.g. Guéry et al. 2017) or study senescence (e.g. Péron et al. 2016). The extension of single-state models to situations where individuals are detected in several geographical sites or equivalently states (e.g. breeding/non-breeding or sane/ill) are called multi-state CR models (Lebreton et al. 2009). Multistate CR models, and the Arnason-Schwarz model in particular (Lebreton et al. 2009), are appealing for addressing various biological questions such as metapopulation dynamics (e.g. Spindel et al. 2016) or life-history trade-offs (e.g. Supp et al. 2015).

A necessary step for correct inference about demographic parameters is to assess the fit of single- and multi-state models to CR data, regardless of whether a Bayesian or a frequentist framework is adopted.

Two families of methods exist to perform goodness-of-fit (GOF) tests for CR models. First, an omnibus test of the null hypothesis that a given model fits the data adequately can be conducted using resampling methods and the deviance as a metric (White 2002). However when the null hypothesis is rejected, this omnibus approach does not inform about an alternative model that could be fitted. Second, specialized tests have been built to address biologically meaningful causes of departure from the null hypothesis. A global test for single- and multi-state CR models is decomposed into several interpretable components based on contingency tables, for example the presence of transients (Pradel et al., 1997; Pradel et al. 2003) or that for trap-dependence (Pradel, 1993; Pradel et al. 2003). These GOF tests are implemented in the Windows application U-CARE (Choquet et al. 2009).

Here, we introduce the R (R Development Core Team 2014) package R2ucare to perform GOF tests for single- and multi-state CR models. R2ucare also includes various functions to help manipulate CR data. As a package in the CRAN database, R2ucare provides full advantage of R's many features (e.g. simulations, model fitting), while being multi-platform. We go through the theory first, then illustrate the use of R2ucare with wolves in France for single-state models and geese in the U.S. for multi-state models.

Theory

Once a model has been specified, GOF testing is the procedure that controls model assumptions. GOF testing and model fitting are two complementary procedures that share and compete for the information contained in the data. More liberal models require more information to be fitted (there are more parameters to estimate) but also fewer assumptions need to be verified. For instance, the time-dependent CJS model is merely content with the numbers of individuals captured at each occasion and the numbers never seen again from those released at each occasion when it comes to estimating its parameters. These summary statistics leave much of the details of the capture histories available to test its assumptions.

There are several ways in which this remaining information may be exploited to test the assumptions. The implementation retained in R2ucare builds on the optimal approach originally devised by Pollock et al. (1985) and later modified by Pradel (1993). It is based on contingency tables and aims at testing with chi-squared tests (and Fisher's exact tests when needed) for transients and trap-dependence. These aspects are examined specifically in two independent component tests called respectively Test 3.SR and Test 2.CT. The component tests directed at transients and trap-dependence actually address features of the data that are consequences of respectively the presence of transients and trap-dependence, so that these features may also be caused by other, completely different phenomena. They do verify respectively that:

- Newly encountered individuals have the same chance to be later reobserved as recaptured (previously encountered) individuals; this is the null hypothesis of Test 3.SR.
- Missed individuals have the same chance to be recaptured at the next occasion as currently captured individuals; this is the null hypothesis of Test 2.CT.

Although these components are often called 'test of transience' and 'test of trap-dependence', when it comes to interpretation, one should keep in mind that transience and trap-dependence are just two specific reasons why the tests respectively called 3.SR and 2.CT might be significant.

Interestingly, these two components provide formal tests for comparing the CJS model with more general models, namely a model with an interaction between age (2 classes) and time in the survival probability for Test 3.SR (Pradel et al. 1997) and a model allowing for a different recapture probability of individuals just released for Test 2.CT (Pradel 1993).

Beyond these two oriented components, the remaining information is distributed and structured into two additional components: Test 3.Sm and Test 2.CL. Those examine long-term features of the data:

- Among those individuals seen again, when they were seen does not differ among previously and newly marked individuals; this is the null hypothesis of Test 3.Sm.
- There is no difference in the timing of reencounters between the individuals encountered and not encountered at occasion i , conditional on presence at both occasions i and $i + 2$; this is the null hypothesis of Test 2.CL.

Data are generally sparse for these components and scattered over many occasions. Despite the implementation of some automatic pooling (see Choquet et al. 2005 for more details about the pooling rules), they are rarely significant alone.

Although many situations can lead to similar test results, we propose here a decision tree (Figure 1) that should lead to reasonable solutions in most cases.

The theory for the GOF test of the multistate Arnason-Schwarz model was developed along similar lines as for the CJS model (Pradel et al. 2003). This test has yet more components and some components have a more complex structure (hence our non attempt to build a decision tree as for the CJS model), but for all that concerns us, the reasoning remains very similar. The test implemented in R2ucare is actually a test of the Jolly-Move model, a slightly more general model than the Arnason-Schwarz model in that it allows detection parameters to depend on the previous state occupied. This is biologically irrelevant in most common situations (Pradel et al. 2003), so that we may reason as if we were examining the Arnason-Schwarz model. Components here have been designed to detect transients, trap-dependence, and the memory of past states. This last point means that the component examines whether transitions to a new state depend on previous states beyond the current one. The corresponding components are respectively Test 3.GSR, Test M.ITEC, and Test WBWA. Like for the CJS case, they actually examine features of the data, namely that:

- Newly encountered individuals have the same chance to be later reobserved as recaptured (i.e. previously encountered) individuals; this is the null hypothesis of Test 3.GSR which is the exact equivalent of 3.SR.
- Individuals currently in the same state, whether captured or missed, have the same chance to be recaptured in each state at the next occasion; this is the null hypothesis of Test M.ITEC.
- Individuals currently captured in the same state have the same chance to be next reobserved in the different states independently of their observed state at the most recent capture; this is the null hypothesis of Test WBWA.

These interpretable components are complemented by two composite components with no clearly identified interpretation, Test 3.GSm and Test M.LTEC. We do not attempt to give a de-

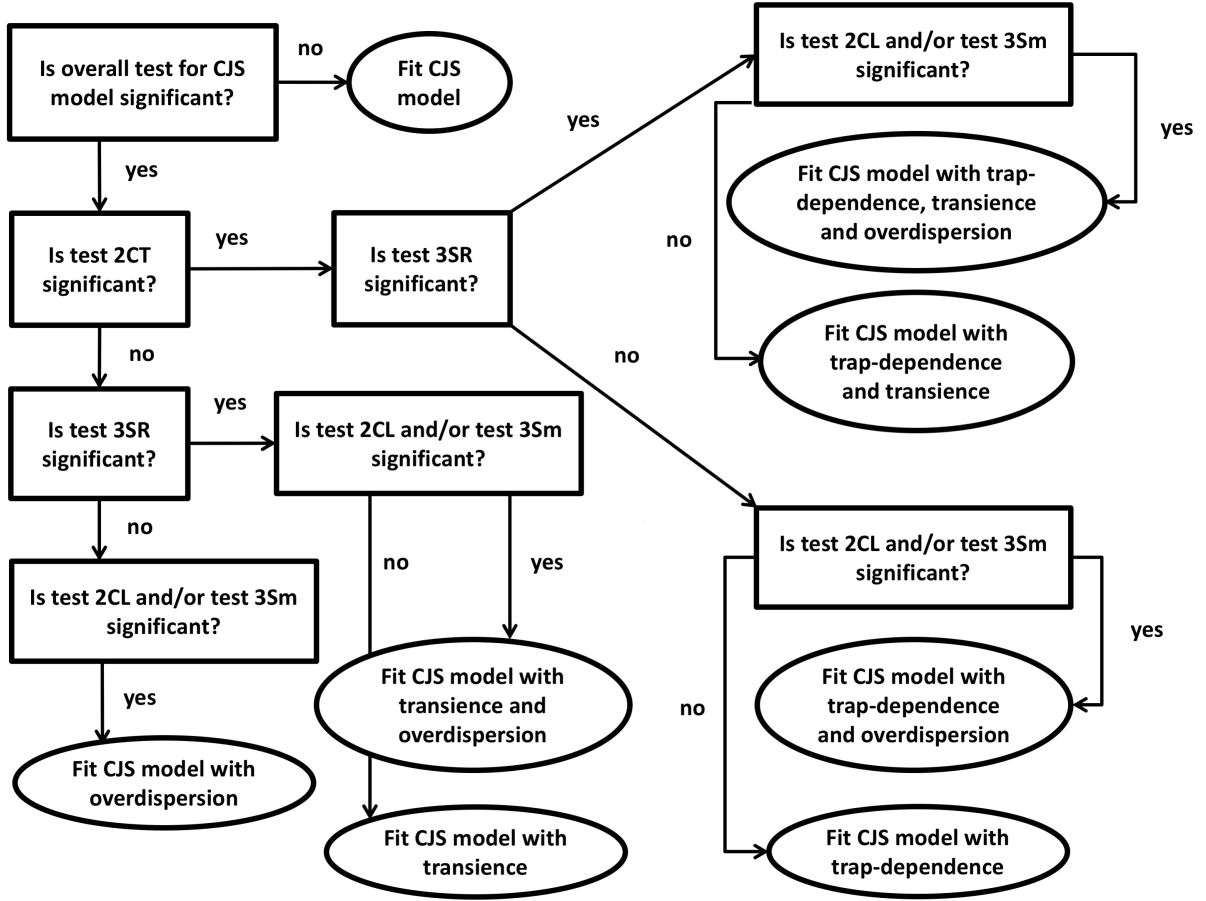


Figure 1: Decision tree to navigate towards testing the fit of single site/state capture-recapture models, with the Cormack-Jolly-Seber (CJS) model as a reference. Questions are in the rectangles, actions in the ellipses. We start by asking the question in the top-left corner. The coefficient of overdispersion is calculated as the ratio of the goodness-of-fit test statistic over the number of degrees of freedom (Pradel et al. 2005). *Remark 1:* we begin by testing for the presence of trap-dependence, then that of transience; these steps could be permuted without affecting the final outcome. *Remark 2:* the overall goodness-of-fit test may be significant while none of the four sub-components is; in this situation, we recommend fitting the CJS model and correcting for overdispersion. *Remark 3:* we do not cover the issue of heterogeneity for which a formal test does not exist. When both the tests for the presence of transience and trap-dependence are significant, and only them, there is suspicion of heterogeneity in detection (Péron et al. 2010). Péron et al. (2010) implemented an approximate procedure to assess the presence of heterogeneity in the detection process, and Jeyam et al. (2017) developed a formal test for the same purpose. Cubaynes et al. (2012) recommended using the Akaike Information Criterion (AIC) to compare models with and without heterogeneity. *Remark 4:* To account for the presence of transience, that of trap-dependence or an effect of heterogeneity, we refer to Pradel et al. (1997), Pradel and Sanz-Aguilar (2012; see also Pradel 1993 and Gimenez et al. 2003) and Gimenez et al. (2017) respectively.

115 scription of these; let it suffice to say that Test 3.GSm is concerned with comparing newly and
 116 previously encountered, while Test M.LTEC contrasts missed and encountered individuals. For-
 117 tunately, these components play a secondary role as they are usually not significant alone.

118 For more details about the theory of GOF testing for CR models, we strongly encourage users
 119 to read Pradel et al. (2005) and Cooch and White (2006).

120 The R2ucare package

121 The R2ucare package contains R functions to perform GOF tests for CR models as well as various
 122 functions to manipulate CR data (see Table 1 and the vignette of the package named vignette_R2ucare).
 123 It ensures reproducibility which was not possible with the U-CARE (Choquet et al. 2009) Windows
 124 standalone application. Besides, it can be used in combination with other R packages for fitting
 125 CR data like RMark (Laake 2013) or marked (Laake et al. 2013) or to carry out simulations to assess
 126 statistical power (e.g. Bromaghin et al. 2013; Fletcher et al. 2012).

Table 1: The main functions of R2ucare and their description. See main text for more details.

Function	Description
marray	build a m-array for single-site/state capture-recapture data
multimarray	build a m-array for multi-site/state capture-recapture data
group_data	pool together individuals with the same encounter capture-recapture history
ungroup_data	split encounter capture-recapture histories into individual ones
read_inp	read MARK formatted files
read_headed	read E-SURGE formatted files
test3sr	implement Test 3.SR for single-site/state models (presence of transients)
test3sm	implement Test 3.Sm for single-site/state models
test2ct	implement Test 2.CT for single-site/state models (presence of trap-dependence)
test2cl	implement Test 2.CL for single-site/state models

Function	Description
test3Gsr	implement Test 3.GSR for multi-site/state models (presence of transients)
test3Gsm	implement Test 3.GSm for multi-site/state models
test3Gwbwa	implement Test WBWA for multi-site/state models (presence of memory)
testMitec	implement Test M.ITEC for multi-site/state models (presence of trap-dependence)
testMltec	implement Test M.LTEC for multi-site/state models

Goodness-of-fit tests for single-site/state models

We illustrate the use of R2ucare to assess the GOF of the CJS model to a dataset on wolves (*Canis lupus*) in France (e.g., Fletcher et al. 2012). Briefly, the data consist of capture histories for 160 individuals, partitioned into 35 3-month intervals (from spring 1995 to autumn 2003).

We first load the R2ucare package:

```
library(R2ucare)
```

Then we read in the wolf data that is provided with the package. To do so, R2ucare contains two functions that accomodate the most frequent CR formats: `read_inp` deals with the MARK format (Cooch and White 2006) while `read_headed` deals with the E-SURGE format (Choquet et al. 2009). The wolf dataset has the MARK format, therefore:

```
wolf = system.file("extdata", "wolf.inp", package = "R2ucare")
wolf = read_inp(wolf)
```

We then get the matrix and number of CR encounter histories:

```
ch = wolf$encounter_histories
n = wolf$sample_size
```

Following the procedure described in Figure 1, we first assess the overall fit of the CJS model by using the function `overall_CJS`:

```
overall_CJS(ch,n)
```

```
139 ##                                chi2 degree_of_freedom p_value
140 ## Gof test for CJS model: 180.73                115          0
```

141 Clearly, the CJS model does not fit the data well ($\chi^2_{115} = 180.73$, $P < 0.01$). We then test for an
142 effect of trap-dependence:

```
test2ct(ch,n,verbose = FALSE)
```

```
143 ## $test2ct
144 ##      stat      df      p_val sign_test
145 ##    64.451    31.000    0.000    -5.641
```

146 Test 2.CT is significant ($\chi^2_{31} = 64.45$, $P < 0.01$). We also provide the signed square root
147 (sign_test) of the Pearson chi-square statistic as a directional test of the null hypothesis (Pradel et
148 al. 2005), which is negative when there is an excess of individuals encountered at a given occasion
149 among the individuals encountered at the previous occasion.

150 Note that, by default, the GOF functions in R2ucare returns all the contingency tables that
151 compose the test under scrutiny, which might not be of immediate use and rather cumbersome
152 on screen, hence the use of verbose=FALSE in the call to the test2ct function above. Now we ask
153 whether there is a transient effect:

```
test3sr(ch,n,verbose = FALSE)
```

```
154 ## $test3sr
155 ##      stat      df      p_val sign_test
156 ##    65.414    29.000    0.000     5.037
```

157 Test 3.SR is also significant ($\chi^2_{29} = 65.41$, $P < 0.01$). We also provide the signed square root
158 (sign_test) of the Pearson chi-square statistic (Pradel et al. 2005), which is positive when there is
159 an excess of never seen again among the newly marked.

160 Navigating through the decision tree in Figure 1 suggests we should perform the two remain-
161 ing tests:


```
test3sm(ch,n,verbose = FALSE)
```

```
162 ## $test3sm
163 ##      stat      df  p_val
164 ## 22.977 25.000  0.579
```

```
test2cl(ch,n,verbose = FALSE)
```

```
165 ## $test2cl
166 ##      stat      df  p_val
167 ## 27.888 30.000  0.576
```

168 Neither Test 3.Sm ($\chi^2_{25} = 22.98, P = 0.58$) nor Test 2.CL ($\chi^2_{30} = 27.89, P = 0.58$) is significant,
169 therefore we recommend fitting a CJS model incorporating both a transience effect and a trap-
170 dependence effect and start the analysis from there. In passing, it is possible to calculate a GOF
171 test for this new model by removing the two components Test 3.SR and Test 2.CT to the overall
172 GOF test (Pradel et al. 2005):

```
# subtract the components 3SR and 2CT to the CJS test statistic
stat_new = overall_CJS(ch,n)$chi2 - (test3sr(ch, n)$test3sr[[1]]
                                     + test2ct(ch, n)$test2ct[[1]])

# calculate degree of freedom associated with the new test statistic
df_new = overall_CJS(ch,n)$degree_of_freedom -
         (test3sr(ch, n)$test3sr[[2]] + test2ct(ch, n)$test2ct[[2]])

# compute p-value
1 - pchisq(stat_new, df_new)
```

```
173 ## [1] 0.6332861
```

174 This new model incorporating transient and trap-dependence effects fits the wolf data well
175 ($\chi^2_{55} = 50.87, P = 0.63$).

176 To date, no GOF test exists for models with individual covariates (unless we discretize them
177 and use groups), individual time-varying covariates (unless we treat them as states) or temporal

178 covariates; therefore, these covariates should be removed from the dataset before using R2ucare.
179 For groups, we recommend treating the groups separately (see e.g. the example in the help file for
180 overall_CJS).

181 **Goodness-of-fit tests for the Arnason-Schwarz model**

182 We now wish to assess the GOF of the Arnason-Schwarz model to a dataset on Canada Geese
183 (*Branta canadensis*) (Pradel et al. 2005). Briefly, the data consist of capture histories for 28,849 indi-
184 viduals marked and re-observed at wintering locations in the US between 1984 and 1986.

185 We first read in the geese data that are provided with the package:

```
geese = system.file("extdata", "geese.inp", package = "R2ucare")  
geese = read_inp(geese)
```

186 We then get the matrix and number of CR encounter histories:

```
ch = geese$encounter_histories  
n = geese$sample_size
```

187 Then we assess the quality of fit of the Arnason-Schwarz model to the geese CR data with
188 the overall_JMV function. Beware that it takes a minute or so to run the test because an iterative
189 optimization procedure is involved to perform Test M.ITEC and Test M.LTEC (Pradel et al. 2003)
190 that is repeated several times to try and avoid local minima.

```
overall_JMV(ch,n)
```

```
191 ##                                chi2 degree_of_freedom p_value  
192 ## Gof test for JMV model: 982.593                197          0
```

193 The null hypothesis that the Arnason-Schwarz provides an adequate fit to the data is clearly
194 rejected ($\chi^2_{197} = 982.59$, $P < 0.01$). In a second step, we further explore each component of the
195 overall test:

```
test3Gsr(ch,n,verbose=FALSE) # transience
```

```
196 ## $test3Gsr
197 ##      stat      df   p_val
198 ## 117.753  12.000   0.000
```

```
test3Gsm(ch,n,verbose=FALSE)
```

```
199 ## $test3Gsm
200 ##      stat      df   p_val
201 ## 302.769 119.000   0.000
```

```
test3Gwbwa(ch,n,verbose=FALSE) # memory
```

```
202 ## $test3Gwbwa
203 ##      stat      df   p_val
204 ## 472.855  20.000   0.000
```

```
testMitec(ch,n,verbose=FALSE) # short-term trap-dependence
```

```
205 ## $testMitec
206 ##      stat      df   p_val
207 ## 68.225  27.000   0.000
```

```
testMltec(ch,n,verbose=FALSE) # long-term trap-dependence
```

```
208 ## $testMltec
209 ##      stat      df   p_val
210 ## 20.989  19.000   0.337
```

211 It appears that all components are significant but the test for a long-term trap-dependence
212 effect. By setting the verbose argument to TRUE (by default argument), one could closely examine
213 the individual contingency tables and better understand the reasons for the departure to the null
214 hypotheses. For example, let us redo the test for transience Test 3.GSR:

```
test3Gsr(ch,n,verbose=TRUE)
```

```
215 ## $test3Gsr
216 ##      stat      df    p_val
217 ## 117.753  12.000    0.000
218 ##
219 ## $details
220 ##      occasion site      stat df      p_val  test_perf
221 ## 1           2      1 3.894777e-03  1 9.502378e-01 Chi-square
222 ## 2           2      2 2.715575e-04  1 9.868523e-01 Chi-square
223 ## 3           2      3 8.129814e+00  1 4.354322e-03 Chi-square
224 ## 4           3      1 1.139441e+01  1 7.366526e-04 Chi-square
225 ## 5           3      2 2.707742e+00  1 9.986223e-02 Chi-square
226 ## 6           3      3 3.345916e+01  1 7.277633e-09 Chi-square
227 ## 7           4      1 1.060848e+01  1 1.125702e-03 Chi-square
228 ## 8           4      2 3.533332e-01  1 5.522323e-01 Chi-square
229 ## 9           4      3 1.016778e+01  1 1.429165e-03 Chi-square
230 ## 10          5      1 1.101349e+01  1 9.045141e-04 Chi-square
231 ## 11          5      2 1.292013e-01  1 7.192616e-01 Chi-square
232 ## 12          5      3 2.978513e+01  1 4.826802e-08 Chi-square
```

233 By inspecting the data.frame containing the details of the test, we see that there is no transients
234 in site 2.

235 Future directions

236 R2ucare allows evaluating the quality of fit of standard capture-recapture models for open pop-
237 ulations. Future developments will focus on implementing goodness-of-fit tests for models com-
238 bining different sources of data (McCrea et al. 2014) and residual-based diagnostics (Choquet et
239 al. 2013, Warton et al. 2017).

Availability

The current stable version of the package requires R 3.4.3 and is distributed under the GPL license. It can be installed from CRAN (<https://cran.r-project.org/web/packages/R2ucare/>) and loaded into a R session as follows:

```
install.packages("R2ucare",dependencies=TRUE)
library("R2ucare")
```

The repository on GitHub <https://github.com/oliviergimenez/R2ucare> hosts the development version of the package, it can be installed as follows:

```
if(!require(devtools)) install.packages("devtools")
library("devtools")
install_github("oliviergimenez/R2ucare")
```

We also maintain a forum at https://groups.google.com/forum/#!forum/esurge_ucare to which questions can be asked.

Acknowledgments

Replication files (paper and code) are available on the first author's Github account (<https://github.com/oliviergimenez>). This work was supported by a grant from the French National Research Agency, reference ANR-16-CE02-0007. We warmly thank E. Marboutin and J. Hestbeck for sharing the wolf and geese datasets, respectively.

Authors' contributions

OG, JDL and RP conceived the ideas and designed methodology; OG, JDL, RC and RP wrote the code; OG and RP led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

References

Bromaghin, J. F., T. L. McDonald, and S. C. Amstrup (2013) Plausible combinations: An improved method to evaluate the covariate structure of Cormack-Jolly-Seber mark-recapture models. *Open Journal of Ecology* 3: 11-22.

- 260 Choquet, R., Reboulet, A.-M., Lebreton, J.-D., Gimenez, O., and R. Pradel. (2005). U-CARE 2.2 user's
261 manual. CEFE, UMR 5175, Montpellier.
- 262 Choquet, R., Lebreton, J.-D., Gimenez, O., Reboulet, A.-M., and R. Pradel. (2009). U-CARE: Utilities for
263 performing goodness of fit tests and manipulating CAPture-REcapture data. *Ecography* **32**: 1071-1074.
- 264 Choquet, R., Rouan, L., Pradel, R. (2009). Program E-SURGE: A software application for fitting multievent
265 models. Modeling Demographic Processes in Marked Populations. D. L. Thomson, E. G. Cooch and M.
266 J. Conroy. Berlin, Germany, Springer. **3**: 845-865.
- 267 Choquet, R., Carrie, C., Chambert, T. & Boulinier, T. (2013). Estimating transitions between states using
268 measurements with imperfect detection: application to serological data. *Ecology* **94**: 2160-2165.
- 269 Cubaynes, S., C. Lavergne, E. Marboutin, and O. Gimenez (2012). Assessing individual heterogeneity us-
270 ing model selection criteria: How many mixture components in capture-recapture models? *Methods in*
271 *Ecology and Evolution* **3**: 564-573.
- 272 Fletcher, D., Lebreton, J.-D., Marescot, L., Schaub, M., Gimenez, O., Slooten, E. and S. Dawson (2012). Bias
273 in estimation of adult survival and asymptotic population growth rate caused by undetected capture
274 heterogeneity. *Methods in Ecology and Evolution* **3**: 206-216.
- 275 Gimenez O., Choquet R. and J.-D. Lebreton (2003). Parameter redundancy in multistate capture-recapture
276 models. *Biometrical Journal* **45**: 704-722.
- 277 Gimenez, O., Cam, E., and J.-M. Gaillard (2017). Individual heterogeneity and capture-recapture models:
278 what, why and how? *Oikos*. In press. doi: 10.1111/oik.04532.
- 279 Guéry, L., Descamps, S., Pradel, R., Hanssen, S. A., Erikstad, K. E., Gabrielsen, G. W., Gilchrist, H. G. and
280 Bêty, J. (2017). Hidden survival heterogeneity of three Common eider populations in response to climate
281 fluctuations. *Journal of Animal Ecology* **86**: 683–693.
- 282 Jeyam, A., R. S. McCrea, T. Bregnballe, M. Frederiksen, and R. Pradel (2017). A test of positive association
283 for detecting heterogeneity in capture for capture-recapture data. Submitted to *Journal of Agricultural,*
284 *Biological, and Environmental Statistics*.
- 285 Laake, J. L. (2013). RMark: An R Interface for Analysis of Capture-Recapture Data with MARK. AFSC
286 Processed Rep 2013-01, 25p. Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., 7600 Sand Point Way
287 NE, Seattle WA 98115.
- 288 Laake, J. L., Johnson, D. S. and Conn, P. (2013). marked: An R package for maximum-likelihood and MCMC

analysis of capture-recapture data. *Methods in Ecology and Evolution* **4**: 885-890.

Lebreton, J.-D. et al. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs* **62**: 67-118.

McCrea, R. S., B. J. T. Morgan, R. Pradel (2014). Diagnostic Goodness-of-Fit Tests for Joint Recapture and Recovery Models. *Journal of Agricultural, Biological, and Environmental Statistics* **19**: 338-356.

Péron, G., Crochet, P.A.C., Choquet, R., Pradel, R., Lebreton, J.-D. and O. Gimenez. (2010). Capture-recapture models with heterogeneity to study survival senescence in the wild. *Oikos* **119**: 524-532.

Péron, G., J.-M. Gaillard, C. Barbraud, C. Bonenfant, A. Charmantier, R. Choquet, T. Coulson, V. Grosbois, A. Loison, G. Marzolin, N. Owen-Smith, D. Pardo, F. Plard, R. Pradel, C. Toïgo, O. Gimenez (2016). Evidence of reduced individual heterogeneity in adult survival of long-lived species. *Evolution* **70**: 2909-2914.

Pollock, K., J. Hines, J. Nichols (1985). Goodness-of-Fit Tests for Open Capture-Recapture Models. *Biometrics* **41**: 399-410.

Pradel, R. (1993). Flexibility in Survival analysis from recapture data: Handling trap-dependence. Pages 29-37 in Lebreton & North, editors. *Marked individuals in the study of bird population*. Birkhäuser Verlag, Basel, Switzerland.

Pradel, R., Hines, J. E., Lebreton, J.-D. & Nichols, J. D. (1997). Capture-recapture survival models taking account of transients. *Biometrics* **53**: 60-72.

Pradel R., Wintrebert C.M.A. and Gimenez O. (2003). A proposal for a goodness-of-fit test to the Arnason-Schwarz multisite capture-recapture model. *Biometrics* **59**: 43-53.

Pradel, R., Gimenez O. and J.-D. Lebreton (2005). Principles and interest of GOF tests for multistate capture-recapture models. *Animal Biodiversity and Conservation* **28**: 189-204.

Pradel R, Sanz-Aguilar A (2012) Modeling Trap-Awareness and Related Phenomena in Capture-Recapture Studies. *PLoS ONE* **7**: e32666.

R Development Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Spendelov, J. A., D. Monticelli, J. D. Nichols, J. E. Hines, I. C. T. Nisbet, G. Cormons, H. Hays, J. J. Hatch, and C. S. Mostello (2016). Roseate Tern breeding dispersal and fidelity: responses to two newly restored

317 colony sites. *Ecosphere* **7**: e01510.

318 Supp, S. R., D. N. Koons, and S. K. M. Ernest (2015). Using life history trade-offs to understand core-
319 transient structuring of a small mammal community. *Ecosphere* **6**: 187.

320 Warton, D. I., J. Stoklosa, G. Guillera-Aroita, D. I. MacKenzie and A. H. Welsh (2017). Graphical diagnostics
321 for occupancy models with imperfect detection. *Methods in Ecology and Evolution* **8**: 408-419.

322 White, G. C. (2002). Discussant: The Use of Auxiliary Variables in Capture-Recapture Modeling: An
323 Overview. *Journal of Applied Statistics* **29**: 103-106.