



## U-CARE: Utilities for performing goodness of fit tests and manipulating Capture–REcapture data

Rémi Choquet, Jean-Dominique Lebreton, Olivier Gimenez, Anne-Marie Reboulet and Roger Pradel

*R. Choquet (remi.choquet@cefe.cnrs.fr), J.-D. Lebreton, O. Gimenez, A.-M. Reboulet and R. Pradel, Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, UMR 5175, 1919 Route de Mende, FR-34293 Montpellier cedex 5, France.*

In any statistical analysis, assessing the goodness of fit of a model to the data is crucial to avoid drawing incorrect conclusions. U-CARE is a computer application that deals with the mandatory first steps of the analyses of capture–recapture data: the preparation of the data set and the assessment of the fit of a general model (Cormack–Jolly–Seber and variants for single-state data; Jolly–Move and variants for multi-state data). U-CARE implements the current state of the art in goodness-of-fit testing by incorporating components aimed at detecting the most likely departures from assumptions (Pradel et al. 2003, 2005). It is a free and stand-alone application for Windows.

The estimation and comparison of demographic rates in animal and plant populations (Lebreton et al. 1992, Gregg and Kery 2006) is currently based on following marked individuals over time and possibly space and analyzing the resulting data with adequate statistical models. Such models, in which individuals may move among states (such as geographical sites or breeding statuses) over discrete time periods, must account for the non-exhaustive detection of individuals. These so-called capture–recapture models are still evolving (Pradel et al. 2005). In their current form, they are based on two key assumptions. 1) The individuals are assumed to be independent, 2) within the history of each individual, the successive pairs of each release associated with its subsequent reencounter (if any) are assumed to be independent (Burnham 1991).

While fitting multi-state capture–recapture models can now be carried out without much difficulty with a variety of computer software applications (M-SURGE (Choquet et al. 2004), E-SURGE (Choquet et al. 2009) or MARK (White and Burnham 1999) for a frequentist approach, MARK for Bayesian inference), the examination of the above mentioned assumptions can be done in an optimal way (Pradel et al. 2003) only with program U-CARE. And yet this step is critical. For instance, if a major structural effect such as the presence of transient individuals on the study site is overlooked, survival will be severely underestimated. More generally, spurious effects will often be detected if no model fits the data. This is best seen when considering how model selection proceeds.

In the frequentist paradigm, model selection is generally based on the Akaike information criterion (AIC, (Burnham

and Anderson 2002)). The AIC is equal to the deviance plus two times the number of estimable parameters. In presence of lack of fit, the deviance tends to be inflated, thus leading to the selection of over-parameterized models and potentially to erroneous biological conclusions. Moreover, the estimate of precision of the maximum likelihood estimators (MLE) will be overly optimistic. The consequences of lack-of-fit are thus too deleterious to be ignored.

A preliminary assessment of goodness-of-fit (GOF) is thus a crucial prerequisite, just as in any statistical analysis (D'Agostino and Stephens 1986). Unlike in regression, residuals are not easily available in CR to check the validity of a model. Furthermore, the omnibus approach to goodness-of-fit testing that consists of comparing expected vs observed sufficient statistics is impractical due to the sparseness of the data. A specific approach was developed (Pradel et al. 2003). By making this approach available in an easy-to-use software application, we aim at encouraging practitioners to assess the validity of their assumptions. Sound model selection, i.e. preceded by appropriate GOF assessment, is indeed becoming more and more common in the literature (Henaux et al. 2007, Jenouvrier et al. 2008) but it is still far from being systematic. We describe here the main features of program U-CARE. More details on tests and tools can be found in the user's manual (Choquet et al. 2005).

U-CARE (<<http://purl.oclc.org/NET/U-CARE>>) is a free, easy-to-use, stand-alone, menu-driven computer application for Windows, with a set of options for managing data and GOF tests capabilities for both single-state and multi-state capture-recapture models (Fig. 1). The two

kinds of test are presented as the sum of components examining different aspects of the data through a range of contingency tables. This structure guides the choice of an appropriate model. For instance, if the only significant subcomponent of the multistate goodness-of-fit test is TEST 3G.SR, a two age-class structure for survival should be used (Pradel et al. 1995), while, if this is TEST WhereBeforeWhereAfter (WBWA), a memory model is recommended (Pradel et al. 2003). Due to historical reasons, biological relevance is particularly furthered for single-state data through the computation of directional tests. We briefly examine hereafter the following four aspects of data analysis for which U-CARE can be very useful. 1) How to prepare the data? 2) How to detect overdispersion? 3) How to correct for overdispersion? 4) How to conduct more specific tests?

### 1) How to prepare the data?

First, the menu TRANSFORM DATA offers some tools for selecting a data subset, such as a particular range of years or some groups (e.g. males or females, “ringed as young” or “ringed as adults”), and for recoding data, such as pooling states, groups or years. Among other features, the menu FILE makes it possible to convert data between two main formats (BIOMEKO (Lebreton and Roux 1989) and MARK (White and Burnham 1999)).

### 2) How to detect overdispersion?

When the data are ready, the next step should be the assessment of the fit of a general model to the data at hand. For single-state data, the classical model is the model with time dependent only parameters (Cormack-Jolly-Seber) (CJS); for multistate data, this is a time and state dependent model (Arnason-Schwarz) (AS). For both of them, the probability of the encounter history of each individual is calculated conditional on its first capture. Optimal goodness of fit tests of the assumptions inherent in the

CJS or the JollyMove (JMV) model, this last being a slight generalization of the AS model, are then derived based on the classical partitioning, according to sufficient statistics  $T$ , of the likelihood  $P(\text{data} | \text{parameters})$ :

$$P(\text{data} | \text{parameters}) = P(\text{data} | T) \times P(T | \text{parameters}).$$

The compatibility of the data with the hypergeometric distributions in  $P(\text{data} | T)$  are tested asymptotically by contingency table chi-squared tests. These tests are organized into several interpretable components by further partitioning  $P(\text{data} | T)$  (Pradel et al. 2005). If systematic departures from the general model can be ruled out, lack of fit may result from the lack of independence among individuals (the two members of a pair, social groups, a contagious disease ...). Then, the value of the Pearson statistic is much larger than the residual degrees of freedom. One simple measure for overdispersion is defined as the ratio of the Pearson statistic  $X^2$  by its number of degrees of freedom  $df$

$$\hat{c} = \frac{X^2}{df}$$

This ratio can be calculated for each component individually or overall. Although there is no clear cut decision rule to decide that an observed lack of fit results solely from overdispersion, a reasonable rule of thumb is: the  $\hat{c}$  ratio is greater than 1 for all components, and there is no component for which the  $\hat{c}$  ratio greatly exceeds the others. This procedure, although not perfect, works generally well as long as the overall ratio does not exceed 3, sometimes 5 (Burnham and Anderson 2002); if it does, an important factor has likely been left out of the model and should be identified. On the other hand, small structural effects that result in an overall  $\hat{c}$  between 1 and 3 can without much damage be assimilated to overdispersion and treated as noise (see below for how to do this in practice).

To illustrate these general ideas, we will now treat the example of the study of movements of Canada geese *Branta canadensis* between three wintering regions, mid-Atlantic,

TRANSFORM DATA	GOODNESS-OF-FIT	GOODNESS-OF-FIT for Multistate	TOOLS
Clean data		TEST WBWA	
Suppress first encounter		TEST 3G.SR	
Select occasions		TEST 3G.Sm	
Gather occasions		TEST M.ITEC	
Reverse time		TEST M.LTEC	
Multistate to single state		Sum of Multistate TESTs over groups	
Recode states		OPTIONS	
Encounter histories layout		TOOLS	
Individuals layout		M-ARRAY	
Extract groups		P-Value	
Suppress groups		Test of mixtures for contingency table	
Pool groups		Test of homogeneity for contingency table	
Split for trap-dep analysis		Power divergence family statistics used for contingency table test	
Build age-groups for trap-dep analysis			
Build groups for trap analysis (Fletcher)			
Sorting cohorts into BIOMEKO files			

Figure 1. U-CARE has 5 menus showing utilities for input-output of capture-recapture files (FILE menu), manipulation of data (TRANSFORM DATA menu), goodness-of-fit testing for single-state (GOODNESS OF FIT menu) and multi-state (GOODNESS OF FIT FOR MULTI-STATE menu) data and various additional tools (TOOLS menu). Three of them are shown.

Chesapeake and Carolinas, between 1984 and 1989 (Hestbeck et al. 1991).

The option “GOODNESS-OF-FIT for Multi-state” of the main menu of U-CARE opens onto the goodness-of-fit test of the JMVmodel (Brownie et al. 1993). Multi-state models (Hestbeck et al. 1991) allow for transitions between states (the wintering sites for the geese), survival probabilities and encounter probabilities. In the JMV model, transitions vary by state of departure, state of arrival and time interval. Survival probabilities vary by state of departure and time interval. Encounter probabilities vary by previous state, current state and date. This model, in contrast with the better-known Arnason-Schwarz (AS) model, allows encounter probabilities to vary by previous state. However, there is currently no optimal GOF test available for the AS model. Program U-CARE contains specific tests for transience (test 3G.SR, null hypothesis H0: “there is no difference in the probability of being later reencountered between ‘new’ and ‘old’ individuals encountered simultaneously”), trap-dependence (test M.ITEC, H0(i): “there is no difference in the probabilities of being reencountered in the different states at  $i+1$  between the animals in the same state at occasion  $i$  whether encountered or not encountered at this date, conditional on presence at both occasions”), and memory (test WBWA, H0: “there is no difference in the expected state of next reencounter among individuals previously encountered in the different states”). The  $\hat{c}$  ratio for the JMV model is computed from these three main components plus two complementary tests 3G.SM and M.LTEC (Pradel et al. 2005).

$$\hat{c} = \frac{X_{WBWA}^2 + X_{3G.SR}^2 + X_{3G.SM}^2 + X_{M.ITEC}^2 + X_{M.LTEC}^2}{df_{WBWA} + df_{3G.SR} + df_{3G.SM} + df_{M.ITEC} + df_{M.LTEC}}$$

If we were to consider that the lack of fit is due solely to overdispersion, the overall  $\hat{c}$  ratio for the geese would be:

$$\hat{c} = \frac{472.8 + 117.7 + 302.7 + 68.2 + 21.9}{20 + 12 + 119 + 27 + 19} = 5$$

However, the  $\hat{c}$  ratio relative to the WBWA component is huge: 23.6 (Fig. 2)!

```

TEST WBWA (test of memory: Where Before vs Where After) , group =1
===== Summary of TEST WBWA results =====
  sta,  pval,  df,  occ,  state,  1 if fisher 0 otherwise
19.591 0.000 2.000 2.000 1.000 0.000
37.868 0.000 2.000 2.000 2.000 0.000
 4.487 0.034 1.000 2.000 3.000 1.000
80.590 0.000 1.000 3.000 1.000 0.000
98.761 0.000 4.000 3.000 2.000 0.000
 0.807 0.369 1.000 3.000 3.000 1.000
27.705 0.000 1.000 4.000 1.000 0.000
53.694 0.000 2.000 4.000 2.000 0.000
25.293 0.000 1.000 4.000 3.000 0.000
43.655 0.000 1.000 5.000 1.000 0.000
50.926 0.000 2.000 5.000 2.000 0.000
29.476 0.000 2.000 5.000 3.000 0.000
----- TEST WBWA (sta, pval, df) -----
472.855 0.000 20.000

```

Figure 2. Details of test WBWA for the Canada goose *Branta canadensis* data from U-CARE. The Pearson statistic (sta) is calculated per occasion and state. The  $\hat{c}$  ratio is 23.6. This suggests strong memory effects in accordance with the overall p-value (pval).

### 3) How to correct for overdispersion?

Given the large value of the overall  $\hat{c}$  ratio obtained for the Canada geese, we must try and find a more general model that takes into account the effects unveiled by the GOF test: the strong memory effect and a transient effect. The  $\hat{c}$  ratio of test 3G.SR is indeed 9.8. A model studied in (Rouan et al. in press) combines these two effects. After discarding the components corresponding to the effects incorporated in this model (memory: component WBWA; transience: component 3G.SR), the new  $\hat{c}$  ratio is

$$\hat{c} = \frac{302.7 + 68.2 + 21.9}{119 + 27 + 19} = 2.4$$

This value is typical of large data sets with several thousands of individuals where individual differences inherent in any animal population are inevitably detected. This ratio can be used as a variance inflation factor for the model of Rouan and derived models. There is an option in M-SURGE, E-SURGE and MARK to introduce this factor. The impact of the correction factor  $\hat{c}$  is null on parameter estimates, but estimated (co)variances are multiplied by  $\hat{c}$ , widths of the Wald-CI-intervals are  $\sqrt{\hat{c}}$  times larger and the deviance is divided by  $\hat{c}$ . This in turn affects model selection where the AIC is replaced with the QAIC (Burnham and Anderson 2002). This procedure of identifying a starting model is valid more generally and should be pursued until an acceptable value for the  $\hat{c}$  ratio is reached.

### 4) How to conduct more specific tests?

U-CARE provides details that may give some additional information, notably the tables of observed and expected numbers. For example, Fig. 3 shows that newly marked individuals are less recaptured than previously marked individuals at occasion 5 on sites 1 and 3 but not 2. This is a general pattern over all occasions which suggests that transients may be absent from the central site 2. A more specific model could be considered. For an application to seasonal data, see Gauthier et al. (2001). U-CARE also gives directional tests (Lebreton et al. 1992) focusing more closely on the detection of transience and trap-dependence.

TABLE 3G.SR(5,1): animals caught at occasion5 in state1 seen earlier and later.		
	Seen again	Never seen again
New	93	317
Old	279	602
Table of expected values		
	Seen again	Never seen again
New	118.141	291.859
Old	253.859	627.141
TABLE 3G.SR(5,2): animals caught at occasion5 in state2 seen earlier and later.		
	Seen again	Never seen again
New	297	748
Old	512	1330
Table of expected values		
	Seen again	Never seen again
New	292.8317	752.1683
Old	516.16834	1325.8317
TABLE 3G.SR(5,3): animals caught at occasion5 in state3 seen earlier and later.		
	Seen again	Never seen again
New	157	504
Old	134	197
Table of expected values		
	Seen again	Never seen again
New	193.9022	467.0978
Old	97.09778	233.9022

Figure 3. Tables of expected and observed values for test 3G.SR at occasion 5 for sites 1, 2, 3.

Eventually, the option “TOOLS” contains some specific tools, not easily found, like the ability to test for mixtures of multinomial in a contingency table (Pradel et al. 2003), with an improved algorithm.

To cite U-CARE or acknowledge its use, cite this Software Note as follows, substituting the version of the application that you used for “Version 2.3”:

Choquet, R., Lebreton, J.-D., Gimenez, O., Reboulet, A.-M. and Pradel, R. 2009. U-CARE: Utilities for performing goodness of fit tests and manipulating CAPTURE-RECAPTURE data. – *Ecography* 32: 1071–1074 (Version 2.3).

## References

- Brownie, C. et al. 1993. Capture–recapture studies for multiple strata including non-Markovian transitions. – *Biometrics* 49: 1173–1187.
- Burnham, K. P. 1991. On a unified theory for release-resampling of animal populations. – In: Chao, C. (ed.), *Symposium in statistics*. Inst. of Statistical Science, Academia sinica, Taipei, pp. 11–35.
- Burnham, K. P. and Anderson, D. R. 2002. *Model selection and Inference: a practical information-theoretic approach*, 2nd ed. – Springer.
- Choquet, R. et al. 2004. M-SURGE: new software specifically designed for multistate recapture models. – *Anim. Biodivers. Conserv.* 27: 207–215.
- Choquet, R. et al. 2005. U-CARE 2.2 user’s manual. – CEFE, UMR 5175, Montpellier.
- Choquet, R. et al. 2009. Program E-SURGE: a software application for fitting multievent models. – In: Thomson, D. L. et al. (eds), *Modeling demographic processes in marked populations*. Springer, pp. 845–865.
- D’Agostino, R. and Stephens, A. 1986. *Goodness-of-fit techniques*. – Marcel Dekker.
- Gauthier, G. et al. 2001. Seasonal survival of greater snow geese and effect of hunting under dependence in sighting probability. – *Ecology* 82: 3105–3119.
- Gregg, K. B. and Kery, M. 2006. Comparison of size vs life-state classification demographic models for the terrestrial orchid *Cleistes bifaria*. – *Biol. Conserv.* 129: 50–58.
- Hénaux, V. et al. 2007. Dispersal and recruitment during population growth in a colonial bird, the great cormorant *Phalacrocorax carbo sinensis*. – *J. Avian Biol.* 38: 44–57.
- Hestbeck, J. B. et al. 1991. Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. – *Ecology* 72: 523–533.
- Jenouvrier, S. et al. 2008. Recruitment processes in long-lived species with delayed maturity: estimating key demographic parameters. – *Oikos* 117: 620–628.
- Lebreton, J. D. and Roux, M. 1989. BIOMEKO (Biometrie-Ecologie). – Avenix, Montpellier.
- Lebreton, J.-D. et al. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. – *Ecol. Monogr.* 62: 67–118.
- Pradel, R. et al. 1995. Transient animals in a resident population of snow geese: local emigration or heterogeneity? – *J. Appl. Stat.* 22: 695–710.
- Pradel, R. et al. 2003. A proposal for a goodness-of-fit test to the Arnason-Schwarz multistate capture–recapture model. – *Biometrics* 59: 43–53.
- Pradel, R. et al. 2005. Principles and interest of GOF tests for multistate capture–recapture models. – *Anim. Biodivers. Conserv.* 28: 189–204.
- Rouan, L. et al. in press. A general framework for modeling memory in capture–recapture data. – *J. Agric. Biol. Environ. Stat.*
- White, G. C. and Burnham, K. P. 1999. Program MARK: survival estimation from populations of marked animals. – *Bird Study* 46 (Suppl.): 120–139.