Principles and interest of GOF tests for multistate capture–recapture models

R. Pradel, O. Gimenez & J.–D. Lebreton

Abstract

Principles and interest of GOF tests for multistate capture–recapture models.— Optimal goodness–of–fit procedures for multistate models are new. Drawing a parallel with the corresponding single–state procedures, we present their singularities and show how the overall test can be decomposed into interpretable components. All theoretical developments are illustrated with an application to the now classical study of movements of Canada geese between wintering sites. Through this application, we exemplify how the interpretable components give insight into the data, leading eventually to the choice of an appropriate general model but also sometimes to the invalidation of the multistate models as a whole. The method for computing a corrective overdispersion factor is then mentioned. We also take the opportunity to try to demystify some statistical notions like that of Minimal Sufficient Statistics by introducing them intuitively. We conclude that these tests should be considered an important part of the analysis itself, contributing in ways that the parametric modelling cannot always do to the understanding of the data.

Key words: Memory, Transients, Trap–dependence, Test WBWA, Contingency tables partitioning, Test M.

Resumen

Principios e interés de los test Bondad de Ajuste (GOF) para los modelos de captura–recaptura multiestado.— Los procedimientos óptimos de bondad de ajuste, aplicados a los modelos multiestado, son nuevos. Trazando un paralelismo con los correspondientes procesos de uniestado, presentamos sus particularidades y mostramos como el test general puede descomponerse en componentes susceptibles de ser interpretados. Todos los desarrollos teóricos están ilustrados con una aplicación del ya clásico estudio de los desplazamientos de la barnacla canadiense entre sus lugares de invernada. Mediante esta aplicación, presentamos un ejemplo de cómo los componentes susceptibles de ser interpretados nos proporcionan una idea de los datos que nos pueden llevar a la elección de un modelo general apropiado, pero también a veces a la invalidación de los modelos de multiestados en su conjunto. Se menciona entonces el método para calcular un factor de corrección de la sobredispersión. Aprovechamos esta ocasión para intentar también desmitificar algunas nociones estadísticas, como las Estadísticas Suficientes Mínimas, introduciéndolas intuitivamente. La conclusión es que estas pruebas deberían considerarse una parte importante del propio análisis, contribuyendo a la comprensión de los datos, de un modo que el modelaje paramétrico no siempre consigue.

Palabras clave: Memoria, Transeúntes, Dependencia del trampeo, Test WBWA, Partición de tablas de contingencia, Test M.

Roger Pradel, Olivier Gimenez & Jean–Dominique Lebreton, CEFE, CNRS, 1919 Route de Mende, 34293 Montpellier cedex 5, France.

Introduction

Multistate capture–recapture models are very appealing for studying a variety of biological questions such as dispersal where states are geographical sites (Hestbeck et al., 1991), trade–off between reproductive status and survival where states are breeder *vs.* non–breeder (Nichols et al., 1994), rate of accession to reproduction where states are experienced *vs.* inexperienced breeders (Pradel & Lebreton, 1999), etc. Furthermore, different types of demographic information, such as live recaptures and recoveries of dead individuals by the public can be analyzed simultaneously using adequate multistate models (Lebreton et al., 1999). A general review of the biological relevance of multistate capture–recapture models can be found in Lebreton & Pradel (2002). In multistate capture–recapture models (Arnason, 1972, 1973; Hestbeck et al., 1991), marked individuals can move among a finite number of states, or die, between discrete time occasions. Survivors are detected ("encountered") in each state, not exhaustively at each occasion. Based on parameters which are the transition, survival and encounter probabilities, the probability of an individual encounter history —conditional on the date and state of first encounter, marking and release— can be calculated in a way similar to that used for the classical one–state Cormack–Jolly–Seber (CJS) model. Under the assumption of independence between individuals, the likelihood for a particular data set is then obtained as the product of the probabilities for each individual encounter history.

The rationale of model selection, based on the AIC, assumes that the set of models considered encompasses a model that fits the data (Burnham & Anderson, 1998). If not, the deviance will tend to be inflated, favoring the incorrect selection of overparametrized models and thus leading to erroneous biological conclusions. Moreover, the precision of the final estimates will also be biased if some lack–of–fit or overdispersion is ignored. The consequences of lack–of–fit are thus too deleterious to be ignored. Yet, difficulties with goodness–of–fit issues have been recurrent in the application of capture–recapture methodology. In a survey of the literature, Begon (1983) concluded that fewer than 11% of the applications of the Jolly–Seber model addressed in a quantitative way or discussed the assumptions inherent in the model. This state of fact was the consequence of the absence at that time of any general goodness–of–fit procedure. The simplest approach, which consists of comparing observed and expected numbers of animals with a particular encounter history, was hampered by the large number of encounter histories (in the one–site case, with 10 occasions, there are more than 1,000 different encounter histories), and as a consequence by the very low expected numbers (the resulting sparseness makes χ^2 distributions for quadratic X^2 statistics or for the deviance quite inadequate (McCullagh & Nelder,

1989)). Today, bootstrap procedures may be a way around distributional problems. However, another weakness of the omnibus approach of comparing expected and observed numbers is that it lacks power against specific alternatives and that it is not informative when it rejects. Specialized tests have been built to address frequent causes of departure. Examples are the Leslie–Carothers test of equal catchability (Carothers, 1971), the Brownie–Robson test of marking–induced deaths (Robson, 1969; Brownie & Robson, 1983), which has later been shown to test also for the presence of transients (Pradel et al., 1997), and, in the context of multistate models, a test of memory (Pradel et al., 2003). However, the relationships between the particular tests will remain unknown until a careful study of the likelihood is carried out. Only such a study can provide the basis for a sound partitioning of the information.

A major step in this direction was the development of optimal goodness–of–fit procedures for the CJS model (Pollock et al., 1985). The global test, organized into several interpretable components and based on adequately pooled tables, was implemented in the RELEASE programme (Burnham et al., 1987). Since then, several specialized tests have been shown to be components of this general test (the test for the presence of transients (Pradel et al., 1997): that for trap–dependence (Pradel, 1993)) and a slightly different version of the general test is now proposed in program U–CARE (Choquet et al., 2005). Recently, Pradel et al. (2003) have developed a similar approach for the multistate model called JMV (Brownie et al., 1993), a model which generalizes the Arnason–Schwarz model by allowing encounter probabilities to vary by site occupied at the previous occasion. The AS model, regarded as the reference model for multistate capture–recapture, has not yet received a specific treatment.

The purpose of this paper is to review the principles on which the goodness–of–fit tests for CJS and JMV are based, underlining their similarities and differences, and to examine how alternatives of interest can be embedded within the general tests. This paper is intended for the biologist with some experience of capture–recapture analysis but no deep statistical training. Thus, we assume that the reader knows what the CJS and the AS models are. On the other hand, we have tried to use everyday words in place of statistical terms. For instance, we seek to introduce notions like minimal sufficient statistics from a practical angle. Most of the paper is illustrated with one example, that of the Canada goose data originally analyzed by Hestbeck et al. (1991). We proceed by steps. First, we present and discuss the features of the goodness–of–fit test of the simpler CJS model and specialized tests embedded within it. For the specialized tests, we examine some statistics particularly suitable to address the alternatives of interest. The second section presents and discusses the goodness–of–fit test of the JMV model, drawing a parallel —as far

Table1. *m*–array for the Canada goose data (Hestbeck et al., 1991) pooled over sites: *i*. Occasion of release*; R~i~.* Number released at *i*; *m~ij~.* Number reencountered at *j* among those released at *i; r~i~.* Number ever reencountered among those released at *i*; *m~j~.* Total number reencountered at occasion *j*.

Tabla 1. Serie m *para los datos de la barnacla canadiense (Hestbeck et al., 1991) una vez reunidos los de diversas localidades:* i. *Ocasión de liberación;* R~i~. *Número de liberaciones en la ocasión* i*;* m~ij~. *Número de reencuentros siguientes en la ocasión* j *con una liberación dada en la ocasión* i*;* r~i~. *Número que se han vuelto a ver con una liberación dada en la ocasión* i*;* m~j~. *Número de reencontrados en la ocasión* j.

i	R_i	2	3	4	5	6	r_i
1	3,494	1,138	309	159	64	42	1,722
2	7,098		1,941	734	345	154	3,174
3	7,603			2,180	740	329	3,249
4	6,804				1,905	702	2,607
5	5,170					1,472	
m_j		1,138	2,250	3,073	3,054	2,699	

Table 2. A fake *m*–array with the same margins as that of the Canada goose data. (For abbreviations see table 1.)

Tabla 2. Una serie m *simulada, con los mismos márgenes que la de los datos de la barnacla canadiense. (Para las abreviaturas ver tabla 1.)*

i	R_i	2	3	4	5	6	r_i
1	3,494	1,138	9	459	64	42	1,722
2	7,098		2,241	34	745	154	3,174
3	7,603			2,580	40	629	3,249
4	6,804				2,205	402	2,607
5	5,170					1,472	
m_j		1,138	2,250	3,073	3,054	2,699	

as is possible— with the goodness–of–fit test of the CJS model. Finally, the last section is devoted to proposals for the improvement of the present situation and tries to identify future directions of research. The material presented in this paper has been implemented in program U–CARE, and is freely available at http://ftp.cefe.cnrs.fr/biom/Soft–CR/.

Likelihood–based goodness–of–fit test for the single–site CJS model

A perfect segregation of information between "estimation of parameters" and "test of assumptions"

For the sake of illustration, let us consider the observations of 28,849 Canada Geese (*Branta canadensis*) banded with individually–coded neck bands and re–observed at three locations: the mid–Atlantic (New York, Pennsylvania, New Jersey), the Chesapeake (Delaware, Maryland, Virginia), and the Carolinas (North and South Carolina) (Hestbeck et al., 1991). Ignoring the locations for the moment, the data can be summarized in what is called an m–array (table 1). At the beginning of each row is the number of geese released on each occasion, followed by the numbers of them reencountered for the

first time on each subsequent occasion. The *m*–array is an interesting summary because it turns out that any set of encounter histories that produces the same *m*–array yields the same maximum likelihood estimates (MLE) of survival and encounter probabilities under the Cormack–Jolly–Seber (CJS) model. For this reason, the *m*–array is said to be a *sufficient statistic* for the CJS model. Actually, even the margins of the *m*–array, i.e. the total number of reencounters per occasion m_j's and the numbers ever seen again among those released at every occasion r_i, are sufficient (Burnham et al., 1987). This is in fact the maximum reduction possible and these margins are thus logically called *minimal sufficient statistics* (MSS). Table 2 presents a different *m*–array with the same margins as the Canada Geese data set. Thus, this *m*–array leads to the same MLE's under the CJS model. However, of two data sets that lead to the same estimates one may respect the model assumptions while the other may not. For instance, of the 3,494 individual geese released at occasion 1, we know for sure that 309 + 159 + 64 + 42 = 574, which had not been encountered at occasion 2 but were encountered later, were still alive at occasion 3. At the same time, 1,941 + 734 + 345 + 154 = 3,174 of the 7,098 geese released at occasion 2 were also alive. The two groups had experienced distinct encounter histories up to occasion 3 but under the assumptions of the CJS model, this should be irrelevant as regards their future; for instance, each of them should have an equal chance of being encountered at occasion 3. This may be tested using a contingency table:

	Seen at 3	Seen later
Last seen at 1	309	265
Last seen at 2	1,941	1,233

As it turns out, this table is slightly unbalanced but far less than the corresponding table from the fake *m*–array:

	Seen at 3	Seen later
Last seen at 1	9	565
Last seen at 2	2,241	933

Thus, while knowing the MSS suffices to estimate the parameters, details of the data i.e. the encounter histories, are needed to test the model assumptions. One may wonder, on the other hand, whether something can be learned from the MSS about the respect of model assumptions.

The answer to this question depends on the model. In general, there is indeed something to learn from the examination of the MSS but not in the case of the CJS model. The CJS model has indeed a peculiarity: its number of MSS is exactly equal to its number of parameters. For instance, the Canada goose study spans 6 years; thus there are 5 *m*'s and 5 *r*'s, and hence a total of 10 margins. However, the sum of the *m*'s and that of the *r*'s are both equal to the total number of animals in the data set. Therefore, there are only 9 minimal sufficient statistics (one margin can be spared). The CJS model with 6 occasions has 5 survival and 5 encounter parameters, 10 parameters in total; but again, at the last time step, only the product of the last survival by the last encounter is estimable, and hence there are only 9 true parameters in total. It can be shown that every time that the number of individual statistics making up the MSS is exactly equal to the number of parameters in the model — as is true of the CJS model— there is nothing to learn from the MSS with respect to model assumptions. The likelihood can always be factorized into two terms: one, the probability of the encounter histories given the MSS, and the other, the probability of the MSS given the parameters.

$$\text{Pr (data; } \theta) = \text{Pr (data / MSS) Pr (MSS; } \theta) \quad (1)$$

In the case of the CJS model, (1) corresponds to a perfect separation of the information. Pr (data / MSS) serves to check the model assumptions, and Pr (MSS; θ) serves *solely* to estimate the parameters. The construction of an optimal goodness–of–fit test is thus based on the sole first part, Pr (data / MSS).

The CJS model makes several assumptions. Based on the encounter histories of otherwise similar individuals, not all are verifiable: for instance, the assumption that the marked animals are representative or that the band codes are not misread. In fact, based on the study of the part Pr (data / MSS) of the likelihood, it can be shown that the verifiable assumptions come down to essentially one thing: *all animals present at any given time are assumed to behave the same*. Pollock et al. (1985) have further shown that this, in turn, can be divided into two (conditionally) independent main

points to be checked: 1) all animals released together have the same expected future whatever their past encounter history and 2) all animals alive at the same date that will be seen again do not differ in the timing of their reencounters whether they are currently encountered or not. The first point leads to what is known as TEST 3; the second to TEST 2, which is also known as the Jolly–Balser test (Balser, 1984). This is actually not the only way of breaking down the general test (Pollock et al. do propose another form of their goodness–of–fit test) but it is the most commonly used and the one we will consider. Starting from this decomposition, it becomes possible to see how tests of specific hypotheses articulate with the general test and among themselves. This has not been done systematically and to our knowledge, the Leslie–Carothers test of unequal catchability (Carothers, 1971) for instance has never been related to the optimal GOF test of CJS. There are already at least two specific tests which have been fully incorporated into the GOF test of the CJS model and to which alternative models have been attached. We examine them now in turn.

A test of transience

TEST 3 theoretically compares, at each occasion, the future history of encountered individuals with respect to their previous encounter history. In practical implementations, the comparison is limited to newly marked and previously marked individuals. That these two categories should have similar expectations implies an equal chance of being seen again. It is thus possible to distinguish two steps in TEST 3: first, the check that newly and already–marked animals have an equal chance of being seen again and then, for those seen again, the check that the spread of next reencounters over time is similar in the two categories. (This corresponds in practice to the partitioning of contingency tables, a very classical statistical technique.) The first subcomponent (table 3) has been suggested many times and has been known since at least 1969 (Robson, 1969). It has received an interpretation as a test for an effect of marking on immediate survival i.e. in the period immediately following release (Brownie & Robson, 1983). It has also been shown to be the adequate test to detect the presence of transients, animals that are passing through the study site en route to other locations (Pradel et al. 1997). This test is called the Brownie–Robson test or TEST 3.SR. An interesting point is that it is the test of comparison of CJS —or (ϕ_t, p_t) in the notation of Lebreton et al. (1992)— with the more general model that provides for 2 age–classes in survival (ϕ_{a2*t}, p_t). As a consequence, a GOF test for model (ϕ_{a2*t}, p_t) is readily available from the GOF test of CJS by ignoring subcomponent TEST 3.SR.

If the alternative of interest is the presence of transients, the direction of departure is predictable. In this case a directional test is appropriate. One such test can be computed by taking the square

Table 3. TEST 3.SR. This subcomponent of the CJS goodness–of–fit test is also a specific test of transience. The signs indicate the expected difference between observed and expected values if there are transients caught in the samples: Sl. Seen later; Nsa. Never seen again; Nsb. Never seen before; Sb. Seen before.

Tabla 3. TEST 3.SR. Este subcomponente del test de bondad de ajuste CJS es también un test específico de divagancia. Los signos indican la diferencia esperada entre los valores observados y esperados, si en las muestras existen transeúntes: Sl. Visto posteriormente; Nsa. No se ha vuelto a ver; Nsb. Nunca visto con antelación; Sb. Visto con antelación.

	Sl	Nsa
Nsb ("new" or "newly marked")	–	+
Sb ("old" or "already marked")	+	–

Table 4. Results of TEST 3.SR for the Canada goose data. The test can be calculated at each of the 4 intermediate occasions. The table gives the Pearson chi–square statistics (X^2) and the corresponding P–value (P) as well as the signed square root (z) of the Pearson chi–square statistic which is normally N(0,1) distributed. z is positive when there is an excess of never seen again among the newly marked.

Tabla 4. Resultados del TEST 3.SR para los datos de la barnacla canadiense. Este test puede calcularse en cada una de las cuatro ocasiones intermedias. La tabla proporciona la ji–cuadrado de Pearson (X^2) y su correspondiente valor P (P), así como la raíz cuadrada provista de signo (z) de la ji–cuadrado de Pearson, que presenta una distribución normal N(0,1). z es positiva cuando existe un exceso de individuos nunca vistos antes entre los que acaban de ser marcados.

Occasion	z	X^2	P
2	1.11	1.24	0.27
3	5.16	26.58	0.00
4	3.79	14.33	0.00
5	3.48	12.09	0.00

root of the Pearson chi–square statistics and giving it a conventional sign (see table 3).

For the Canada geese (table 4), the overall test is highly significant ($\chi^2(4) = 54.24$; $P < 10^{-10}$). A more specific and thus more powerful overall test of transience can be based on the statistic

$$z = \frac{1}{\sqrt{p}} \sum_{i=1}^{p} z_i \qquad \text{(for } p \text{ components)}$$

which is standardized normal N(0,1) under H_0 (no transience or age effect) and will tend to be positive under H_1. Here $z = 6.766$ is highly significant.

A test of short–term trap–dependence

The other test of a specific alternative that has been fully incorporated in the general GOF test of CJS is directed at detecting immediate trap–dependence on encounter probability, meaning that animals that are encountered at occasion i have a different, higher (in case of trap–happiness) or lower (in case of trap–shyness), probability of encounter than the rest of the population at the next occasion $i+1$ (table 5). The tables built in Section "A perfect segregation of information between 'estimation of parameters' and 'test of assumptions'" are examples of this test. As mentioned earlier, TEST 2 compares the future of animals alive at the same occasion which are then encountered or not encountered. Just as TEST 3.SR was obtained as a subcomponent of TEST 3, the test of trap–dependence, called 2.CT, is obtained as a subcomponent of TEST 2. (The complement of TEST 2.CT in TEST 2 investigates the timing of next encounters

Table 5. TEST 2.CT. This subcomponent of the CJS goodness–of–fit test is also a specific test of immediate trap–dependence. The signs indicate the expected difference between observed and expected values in case of trap–happiness. They should be reversed for trap–shyness: Ei+1. Encountered at i+1; El. Encountered later; NEi. Not encountered at i; Ei. Encountered at i.

Tabla 5. TEST 2.CT. Este subcomponente del test de bondad de ajuste CJS es también un test específico de la dependencia del trampeo inmediato. Los signos indican la diferencia esperada entre los valores observados y esperados en el caso de animales habituados a la trampa. Deberían ser contrarios en el caso de individuos no habituados a la trampa: Ei+1. Encontrado en i+1; El. Encontrado más tarde; NEi. No encontrado en i; Ei. Encontrado en i.

	Ei+1	El
NEi	–	+
Ei	+	–

Table 6. Results of TEST 2.CT for the Canada goose data. This test can be calculated at each intermediate occasions but the penultimate. The table gives the Pearson chi–square statistics (X^2) and the corresponding P–value (P) as well as the signed square root (z) of the Pearson chi–square statistics which is normally N(0,1) distributed. z is negative when there is an excess of "encountered at i+1" among the "encountered at i".

Tabla 6. Resultados del TEST 2.CT para los datos de la barnacla canadiense. Este test puede calcularse en cada una de las ocasiones intermedias exceptuando la penúltima. La tabla proporciona la ji–cuadrado de Pearson (X^2) y su correspondiente valor P (P), así como la raíz cuadrada provista de signo (z) de la ji–cuadrado de Pearson, que presenta una distribución normal N(0,1). z es negativa cuando existe un exceso de individuos "encontrados en i+1" entre los "encontrados en i".

Component	z	χ^2	P
2	-3.29	10.86	0.00
3	-5.02	25.16	0.00
4	-3.13	9.80	0.00

of the animals not encountered at i+1.) The alternative model to CJS here is the generalization, noted (ϕ_t, p_{t^*m}), allowing for a different encounter probability of animals just released. A GOF test for this model can be obtained by leaving out subcomponent 2.CT from the GOF test of CJS.

Unlike for transients, the direction of departure can be in any direction with trap–dependence. Yet, we expect the effect to be consistent over occasions. Thus the signed z statistic remains useful when combining the TESTs 2.CT of the different occasions: the evidence for a trap effect accumulates with tables repeatedly unbalanced in the same direction (table 6).

For the geese, there is overwhelming evidence that encounter probability is much higher for a goose encountered at the previous occasion. Both the omnibus chi–squared statistics and the directional test are highly significant ($X^2 = 45.8212$, $P < 10^{-9}$; $z = -6.6061$, $P < 10^{-10}$). However, we have up to now ignored the site of observation. If, as is likely, the effort of observation is unequal and the geese tend to be faithful to the same site from year to year, a goose that frequents the site with high observation pressure will tend to be reobserved consistently more often, leading to a spurious trap effect. To get around this problem, we now turn our attention toward multisite (also multistate) models, more specifically, the JMV model.

Likelihood–based goodness–of–fit test for the multistate model JMV

In multisite protocols, the individuals are sampled over K occasions and s sites. In the example of the Canada goose, there are 3 main areas in the Atlantic flyway, which we will now distinguish. The data can again be summarized in a multisite or multistate m–array (Brownie et al., 1993) (table 7), a generalization of the m–array for one–site data. The comparison of table 7 with table 1 should make clear how the multistate m–array is built. Therefore, we introduce here another approach to the m–array. Each encounter history can be split into several pieces, from the first release to the next reencounter, from the subsequent release to the next reencounter and so on until the end of the study period. For instance, the capture history 302300 over 6 occasions may be seen as made of the three pieces: 302000, 002300 and 000300. Each time that the individual is reencountered (the first two pieces), it is treated as if removed from the data set; this insures that only one individual remains present at the same time in the data set. Each piece is then treated as if coming from an independent individual. The m–array is essentially the tally of these pieces arranged by rows according to the occasion and state of release, and by columns according to the occasion and state of next reencounter (plus a "never–reencountered" column). Obviously, for a model that assumes that the fate of an individual is not affected by its past capture history, the information retained is sufficient. However, because of the loss of information accompanying the construction of the m–array, some assumptions can no longer be checked; for instance, whether some individuals are encountered significantly more often than others. This explains why the objectives of checking the model assumptions and that of estimating the parameters tend to use the complementary part of the total information.

An imperfect segregation of information between "estimation" and "test of assumptions"

The basic assumptions inherent in the JMV model are similar to those of the CJS model except that differences between individuals in the different states are now acknowledged. Again, the fate of the individuals that are in the same state at the same time does not depend on their past. A consequence is that the multistate m–array is a sufficient statistic. Moreover, it can be shown that, unlike the one–site m–array, the multistate m–array is minimally sufficient. Now, the number of sufficient statistics, i.e. the number of independent cells in the multistate m–array, is K (K – 1) s² / 2. This is greater than the number of identifiable parameters as soon as K > 3: there are indeed (K – 1) s² transition probabilities, plus (K – 1) s² encounter probabilities minus s² because the encounter probabilities of the last occasion are not estimable separately from the last transitions; a total of (2K – 3) s² true parameters. The JMV model does not therefore have the nice properties of the CJS model. For instance, it is

Table 7. Multisite *m*–array for the Canada goose data. Sites are North Atlantic (1), the Chesapeake (2) and the Carolinas (3). Only the first 2 and the last occasions of release are shown: m_{ij}^{rs}. Number of next reencounter at occasion *j* in site *s* given release at occasion *i* in site *r; i.* Release occasion; *r.* Release sites; R_i^r. Number released.

Tabla 7. Serie m multilocalidad para los datos de la barnacla canadiense. Las localidades son el Atlántico Norte (1), la región de Chesapeake (2) y las dos Carolinas (3). Sólo se muestran las dos primeras y la última ocasión de liberación: m_{ij}^{rs}. *Número del siguiente reencuentro en la ocasión* j *en la localidad* s *dada la liberación en la ocasión* i *en la localidad* r; i. *Ocasión de liberación;* r. *Localidades de liberación;* R_i^r. *Número de liberaciones.*

			m_{ij}^{rs}														
			2			3			4			5			6		
i	*r*	R_i^r	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	1	785	239	53	0	36	18	0	13	6	0	6	5	1	5	2	0
1	2	2,086	85	615	6	36	158	2	22	92	3	7	32	2	3	22	0
1	3	623	24	49	67	11	30	18	3	10	10	0	8	3	2	5	3
2	1	2,082				491	134	0	149	71	3	51	42	3	21	13	0
2	2	3,918				159	869	15	63	335	10	41	164	3	18	74	2
2	3	1,698				14	101	158	8	47	48	7	16	18	1	14	11
…	…	…													…	…	…
5	1	1,291													271	99	2
5	2	2,887													137	654	18
5	3	992													18	105	16

no longer possible to factorize the likelihood into a term used solely for parameter estimation and another for assessing the goodness of fit; the term Pr(MSS;θ) of formula (1) retains some information about the respect of model assumptions and has to be examined when assessing the fit of the JMV model. There is, however, an analogy with the CJS model which still holds. The verifiable assumptions come down again to one very similar thing: *all animals present at any given time at the same site behave the same*. And this is again equivalent to the verification of two (conditionally) independent points: 1) animals released together have the same expected future whatever their past encounter history and 2) animals present at the same site at the same date that are eventually reencountered do not differ in the timing of their reencounters whether they are currently encountered or not. Thus, apart for the precision of a common site, the exact same two main components are retrieved.

Past encounter history should not matter (TEST 3G)

The first main component of the GOF test of the JMV model, called TEST 3G, examines the effect of the past capture history on the future of animals captured and released at the same time on the same site (Pradel et al., 2003). It is thus the equivalent of TEST 3 of which it is a generalization. Again, there are many possible past capture histo-

ries and the practical implementation of this test as found in program U–CARE version 2.0 (Choquet et al., 2003) considers only a limited number of situations: the newly caught animals are on the first row while the previously caught ones are dispatched over the subsequent rows according to their site of most recent encounter (see table 8); the columns correspond to the particulars (time and site) of the next encounter if any.

As can be seen in table 8, even with a large data set like that of the Canada geese, empty cells easily occur and some sort of pooling is needed. The results in table 9 were obtained with U–CARE version 2.0 which has an automatic pooling algorithm built in. They show that the Canada geese caught together differ strongly depending on their past (over all TEST 3G: $X^2(103) = 749.27$; $P < 10^{-14}$). A close examination of the individual contingency tables like that of table 8, especially the comparison of expected and observed numbers in each cell, might prove useful in understanding the reasons for the departure. However, the breaking up of TEST 3G into meaningful subcomponents is a better option.

A generalized test of transience
A first subcomponent can be built to test for the presence of transients in each sample defined by a site and an occasion (table 10). This straightfor-

Table 8. Component 3G (2,1) of the JMV goodness–of–fit test applied to the Canada goose data. This component is based on the animals caught at occasion 2 on site 1. They are dispatched according to the site of most recent encounter in rows and the particulars (time and location) of the next encounter in columns. The "–" sign is used for animals that are caught for the first time (first row) or that will never be encountered again (last column).

Tabla 8. Componente 3G (2,1) del test de bondad de ajuste JMV aplicado a los datos de la barnacla canadiense. Este componente se basa en los animales capturados en la ocasión 2 en la localidad 1. Se han distribuido en filas según el lugar de encuentro más reciente y en columnas según las circunstancias (tiempo y localidad) del siguiente hallazgo. El signo "–" se utiliza para los animales capturados por primera vez (primera fila) o que nunca serán vueltos a encontrar (última columna).

		Time (j) and location (v) of first reencounter												
Location	j =	3			4			5			6			–
at time 1	v =	1	2	3	1	2	3	1	2	3	1	2	3	–
–		390	124	0	122	64	3	46	35	3	18	9	0	920
1		75	3	0	21	4	0	5	2	0	1	0	0	128
2		19	6	0	4	3	0	0	2	0	1	3	0	47
3		7	1	0	2	0	0	0	3	0	1	1	0	9

Table 9. Results of TEST 3G for the Canada goose data. The table gives the Pearson chi–square statistic (χ^2) and the corresponding P–value (P) as well as the number of degrees of freedom after pooling (df): Oc. Occasion; S. Site.

Tabla 9. Resultados del TEST 3G para los datos de la barnacla canadiense. Las tablas presentan la ji–cuadrado de Pearson (X^2) y su correspondiente valor P (P), así como el número de grados de libertad (df) tras la reunión: Oc. Ocasión; S. Localidad.

X^2	P	df	Oc	S
40.85	0.000	14	2	1
6.73	0.566	8	2	2
17.57	0.007	6	2	3
115.35	0.000	12	3	1
72.64	0.000	15	3	2
57.64	0.000	7	3	3
89.19	0.000	8	4	1
94.49	0.000	12	4	2
50.57	0.000	5	4	3
62.33	0.000	6	5	1
53.75	0.000	6	5	2
88.16	0.000	4	5	3

Table 10. A suitable partitioning of TEST 3G isolates TEST 3G.SR. This subcomponent is exactly similar to TEST 3.SR but involves a further stratification by site (and not only by date). It is also a specific test of transience. The signs indicate the expected difference between observed and expected values when transients are caught in the samples: Sl. Seen later, Nsa. Never seen again; Nsb. Never seen before; Sb. Seen before.

Tabla 10. Una partición adecuada del TEST 3G aísla al TEST 3G.SR. Este subcomponente tiene una similitud exacta con el TEST 3.SR, pero presenta una mayor estratificación respecto a la localidad (y no sólo a la fecha). También constituye un test específico de la divagancia. Los signos indican la diferencia esperada entre los valores observados y esperados cuando en las muestras se capturan transeúntes: Sl. Visto posteriormente; Nsa. No se ha vuelto a ver; Nsb. Nunca visto con antelación; Sb. Visto con antelación.

	Sl	Nsa
Nsb ("new" or "newly marked")	–	+
Sb ("old" or "already marked")	+	–

Table 11. Results of TEST 3G.SR for the Canada goose data: Oc. Occasion; S. Site.

Tabla 11. Resultados del TEST 3G.SR para los datos de la barnacla canadiense: Oc. Ocasión; S. Localidad.

X^2	P	Oc	S
0.004	0.95	2	1
0.000	0.99	2	2
8.130	0.00	2	3
11.394	0.00	3	1
2.708	0.10	3	2
33.459	0.00	3	3
10.608	0.00	4	1
0.353	0.55	4	2
10.168	0.00	4	3
11.013	0.00	5	1
0.129	0.72	5	2
29.785	0.00	5	3

Table 12. TEST WBWA, this subcomponent of TEST 3G tests for a memory effect. It compares the site of the most recent observation in row ("Where Before") to the site of the next observation in column ("Where After"). The "+" signs indicate where the observed values should exceed the expected values when animals tend to return to previously visited sites: Nss. Next seen on site; Lss. Last seen on site.

Tabla 12. TEST WBWA, este subcomponente TEST 3G comprueba el efecto de la memoria. Compara la localidad de la observación más reciente en la fila ("Where Before"), con la localidad de la siguiente observación en la columna ("Where After"). El signo "+" indica dónde los valores observados deberían exceder a los valores esperados, cuando los animales tienden a volver a las localidades previamente visitadas: Nss. Siguiente avistamiento en la localidad; Lss. Último avistamiento en la localidad.

	Nss 1	Nss 2	…	Nss s
Lss 1	+			
Lss 2		+		
…			+	
Lss s				+

ward generalization of TEST 3.SR, called TEST 3G.SR, provides a GOF test of a generalized JMV model, a model with two age classes on survival. This generalization of the classical JMV model is denoted $F_{a2*from*t}$, $\psi_{from*to*t}$, $p_{from*to*t}$ in the notation of Choquet et al. (2004): F is survival, ψ transition and p encounter probability ; *from* is the site of departure, *to* the site of arrival. The goodness–of–fit test of this model is obtained by leaving out subcomponent 3G.SR when calculating the overall GOF test.

Applied to the Canada geese, TEST 3G.SR reveals an interesting feature (table 11). Although globally significant ($X^2(12) = 117.753$; $P < 10^{-13}$), the test is not significant when restricted to site 2 only ($X^2(4) = 3.19$; $P = 0.53$). Thus, there seem to be no transients in the central Chesapeake region! A directional *z* statistic could be calculated in the same manner as with TEST 3.SR.

A test of memory

Animals may make decisions of movement based on the knowledge of previously visited sites. Hestbeck et al. (1992) have identified this phenomenon in the election of wintering sites by Canada geese. This "memory effect", which is probably common in many long–lived species, is a violation of the assumption of the JMV model of the sort that TEST 3G examines: it leads to different behaviour for animals belonging to the same sample depending on which sites they had visited previously. This memory effect is detectable by the specific test of

memory, called WBWA, proposed by Pradel et al. (2003). We will show in the next section that TEST WBWA presented in table 12 is a subcomponent of TEST 3G.

Applied to the Canada geese (table 13), TEST WBWA confirms the very strong role of memory in the movements of these birds ($X^2(20) = 472.86$; $P < 10^{-14}$): the overdispersion factor calculated for this test alone is $472.86/20 = 23.6$, much higher than that of the overall TEST 3G (7.27) or that of the test for transience (9.81).

In order to more specifically target the departures which are expected under the memory effect along the diagonal, an alternative statistic to the Pearson Chi–square can be used. One such possibility is Cohen's kappa (Cohen, 1960), which has a standardized normal N(0,1) distribution. The individual kappa tests can be combined in the same manner as the *z* tests of section 1.2 to get an overall test of memory. They are added and their sum is divided by the square root of the number of components p (Gimenez, 2003):

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} \kappa_i$$

Table 13. Results of TEST WBWA for the Canada goose data. The table gives the Pearson chi–square statistic (X^2) and the corresponding P–value (P) as well as the number of degrees of freedom after pooling (df): Oc. Occasion; S. Site.

Tabla 13. Resultados del TEST WBWA para los datos de la barnacla canadiense. Las tablas presentan la ji–cuadrado de Pearson (X^2) y su correspondiente valor P (P), así como el número de grados de libertad (df) tras la reunión: Oc. Ocasión; S. Localidad.

X^2	P	df	Oc	S
19.59	0.000	2	2	1
37.87	0.000	2	2	2
4.49	0.034	1	2	3
80.59	0.000	1	3	1
98.76	0.000	4	3	2
0.81	0.369	1	3	3
27.71	0.000	1	4	1
53.69	0.000	2	4	2
25.29	0.000	1	4	3
43.66	0.000	1	5	1
50.93	0.000	2	5	2
29.48	0.000	2	5	3

Table 14. Directional test of memory applied to the Canada goose data. This test is distributed as N(0,1) and looks at a consistent excess (or lack) on the diagonal.

Tabla 14. Test direccional de memoria aplicado a los datos de la barnacla canadiense. Este test se distribuye como N(0,1) y demuestra un exceso (o una falta) consistente en la diagonal.

κ	P	κ	P
3.87	0.00	4.33	0.00
4.33	0.00	7.22	0.00
1.81	0.04	4.22	0.00
8.59	0.00	6.36	0.00
5.98	0.00	6.47	0.00
1.01	0.21	4.44	0.00

This directional test of memory confirms the strong positive correlation between the previous and the next sites of observation of the Canada geese ($\kappa = 16.92$; $P < 10^{-13}$). There is as yet no simple alternative model associated to TEST WBWA (like the 2–age model associated to TEST 3G.SR), but the model that accounts for the location at i–1 in the transitions (Brownie et al., 1993) will probably treat most of the "memory effect". Unfortunately, this model cannot be fitted in the framework of multistate models for the full data as it belongs to a more general family of capture–recapture models (Pradel, 2005).

The full decomposition of TEST 3G
TEST 3G.SR and TEST WBWA are two independent subcomponents of TEST 3G but they do not make up for TEST 3G alone. We illustrate here how the original table of TEST 3G is partitioned to isolate these specific tests with the example of the Canada geese encountered at occasion 2 and site 1 (table 8). This procedure of partitioning is very general with contingency tables (Everitt, 1977). In a first step, table 8 is replaced with two tables. This step consists in setting aside the previously

captured geese (first 3 rows) in a separate table and then confronting them all pooled together against the newly captured geese in a second table.

(see 2 contingency tables of step 1 below)

Then, within each one of these two new tables, the never–seen–again geese (last column) are set aside leading to four new tables, one of which is the component of TEST 3.SR relative to occasion 2 and site 1. In this step, the timing of first reencounters is compared among the different rows in a first table, and then the total of reencounters is compared with the number of never–seen–again animals among the same rows in a second table.

(see 4 contingency tables of step 2 below)

Eventually, the first of the four previous tables, which summarizes the first reencounters of the previously encountered individuals, is replaced with four tables: one contains the reencounters made at site 1, one those made at site 2, one those made at site 3, and the last one contrasts the number of reencounters at each site depending on the site of most recent encounter (in rows). This last one is the component of TEST WBWA relative to occasion 2 and site 1. Of the 7 tables obtained at this stage (the last three tables of step 2 remain unchanged in the last step), only two belong to the specific tests described in the previous sections.

(see 7 contingency tables of step 3 below)

Step1. Two contingency tables.

Paso 1. Dos tablas de contingencia.

75	3	0	21	4	0	5	2	0	1	0	0	128
19	6	0	4	3	0	0	2	0	1	3	0	47
7	1	0	2	0	0	0	3	0	1	1	0	9

| 390 | 124 | 0 | 122 | 64 | 3 | 46 | 35 | 3 | 18 | 9 | 0 | 920 |
| 101 | 10 | 0 | 27 | 7 | 0 | 5 | 7 | 0 | 3 | 4 | 0 | 184 |

The remaining tables constitute together TEST 3G.Sm of U–CARE version 2.2. To summarize, TEST 3G is made up of 3 subcomponents: TEST 3G.SR, which tests specifically for transients; TEST WBWA, which aims at detecting a memory effect, and the complementary composite TEST 3G.Sm.

To be caught or not should have no effect (TEST M)

The second main component of the JMV model GOF test, called TEST M, contrasts the animals not caught at a given occasion —yet known to be alive— to those caught at the same occasion. Again, the JMV assumptions imply that there should be no difference between two animals when one is caught and the other is not. However, the exact location of the animals not encountered remains unknown. This is the most far–reaching difference with the one–site context and the reason why the multistate JMV model has not all the nice properties of the single–site CJS model (see Section "An imperfect segrega-

Step 2. Four contingency tables.

Paso 2. Cuatro tablas de contingencia.

75	3	0	21	4	0	5	2	0	1	0	0
19	6	0	4	3	0	0	2	0	1	3	0
7	1	0	2	0	0	0	3	0	1	1	0

111	128
38	47
15	9

| 390 | 124 | 0 | 122 | 64 | 3 | 46 | 35 | 3 | 18 | 9 | 0 |
| 101 | 10 | 0 | 27 | 7 | 0 | 5 | 7 | 0 | 3 | 4 | 0 |

(TEST 3G.SR)

| 814 | 920 |
| 164 | 184 |

Step 3. Seven contingency tables.

Paso 3. Siete tablas de contingencia.

75	21	5	1
19	4	0	1
7	2	0	1

(TEST WBWA)

102	9	0
24	14	0
10	5	0

3	4	2	0
6	3	2	3
1	0	3	1

111	128
38	47
15	9

0	0	0	0
0	0	0	0
0	0	0	0

(TEST 3G.SR)

814	920
164	184

390	124	0	122	64	3	46	35	3	18	9	0
101	10	0	27	7	0	5	7	0	3	4	0

Table 15. Component of TEST M relative to date 2 for the Canada geese. The first three rows correspond to the geese not observed at occasion 2 that were released at sites 1, 2 and 3 respectively at date 1. The last three rows are for the geese observed at date 2 on the three sites in the same order. The columns correspond to the particulars (site within date) of the next reencounter.

Tabla 15. Componente del TEST M, relativa a la fecha 2, para la barnacla canadiense. Las tres primeras filas corresponden a los gansos no observados en la ocasión 2, que fueron liberados en la fecha 1 en las localidades 1, 2 y 3, respectivamente. Las tres últimas filas corresponden a los gansos observados en la fecha 2 en las tres localidades en el mismo orden. Las columnas corresponden a los datos (localidad en una fecha determinada) del siguiente reencuentro.

		Time (j) and location (v) of first reencounter											
$j =$		3			4			5			6		
$v =$		1	2	3	1	2	3	1	2	3	1	2	3
last seen at site													
	1	36	18	0	13	6	0	6	5	1	5	2	0
	2	36	158	2	22	92	3	7	32	2	3	22	0
	3	11	30	18	3	10	10	0	8	3	2	5	3
currently seen at site													
	1	491	134	0	149	71	3	51	42	3	21	13	0
	2	159	869	15	63	335	10	41	164	3	18	74	2
	3	14	101	158	8	47	48	7	16	18	1	14	11

Table 16. Results of TEST M for the Canada geese. This test cannot be computed at the first occasion or less than 2 occasions before the end of the study: Oc. Occasion.

Tabla 16. Resultados del TEST M para la barnacla canadiense. Este test no puede calcularse a la primera ocasión, o a menos de dos ocasiones antes del final del estudio.

X^2	P	df	Oc
24.119	0.044	14	2
36.037	0.007	18	3
23.098	0.006	9	4

tion of information between 'estimation' and 'test of assumptions'"). As regards the tests, because of the uncertain location of the not–encountered individuals, homogeneity tests of contingency tables are now replaced with more complex tests of mixtures.

Let us consider the table retained in U–CARE version 2.0 for the component of TEST M relative to date 2 for the Canada geese (table 15). The first three rows correspond to the geese that were last released at date 1 at sites 1, 2 and 3 respectively; the last three rows to the geese currently released at the same sites. The columns correspond to the timing and place of the next reencounter. In this table, the first three rows do not play the same role as the last ones; they should be approximate linear combinations of the last ones. The rationale for this is as follows: the animals not observed at date 2 may have moved since they were last released; hence, their current location can be any one of the three sites. These animals are thus a mixture of animals in the different sites in unknown proportions. In accordance with the model assumptions, those at site 1 (resp. at site 2 and 3), i.e. on rows 1 (resp. 2 and 3), should behave like those caught and released at site 1 (resp. at site 2 and 3), i.e. on rows 4 (resp. 5 and 6).

The results concerning the Canada geese (table 16) are significant (overall test: $\chi^2(41) = 83.254$; $P < 10^{-3}$), although not as strong as those from TEST 3G and its subcomponents. It is difficult to know the reason for departure simply by examining a complex table like table 15, even if the expected numbers were given. A suitable partitioning is again the key to a better understanding.

A test of short–term trap–dependence

Drawing a parallel with the CJS GOF test, a test for trap–dependence can be considered. The immediate question that arises is what trap–dependence means when there are several sites. Once an animal has been caught, is it expected

Table 17. TEST M.ITEC: this subcomponent of TEST M tests for trap–dependence. In case of local trap–happiness, observed numbers should exceed expected numbers where there are "++" signs. If trap–dependence is present even if the animal has moved to another site (or more presumably state), the excesses will show also where there are "+" signs: Sns1. Seen next occasion at site 1; Sns2. Seen next occasion at site 2; Sls1. Sen later at site 1; Sls2. Seen later at site 2; Lss1. Last seen at site 1; Lss2. Last seen at site 2. Cs1. Currently seen at site 1; Cs2. Currently seen at site 2.

Tabla 17. TEST M.ITEC: este subcomponente de los TEST M está destinado a demostrar la dependencia a la trampa. En el caso de una habituación a la trampa local, las cifras observadas deberían exceder a las esperadas en los lugares en los que existen signos "++". Si la dependencia a la trampa existe incluso si el animal se ha desplazado a otro lugar (o más probablemente a otro estado), los excesos se pondrán también de manifiesto donde existan signos "+": Sns1. Visto la siguiente ocasión en la localidad 1; Sns2. Visto la siguiente ocasión en la localidad 2; Sls1. Visto más tarde en la localidad 1; Sls2. Visto más tarde en la localidad 2; Lss1. Visto la última vez en la localidad 1; Lss2. Visto la última vez en la localidad 2. Cs1. Visto habiatualmente en la localidad 1; Cs2. Viso habitualmente en la localidad 2.

	Sns1	Sns2	…	Sls1	Sls2	...
Lss1						
Lss2						
…						
Cs1	++	+		+		
Cs2	+	++		+		
…		+		+		++

to change its behaviour the next time only if it remains at the same site, or should it change even if it moves to a different site? Presumably, the first option is more reasonable. However, when dealing with states instead of sites, the second option may be better: the animal will be faced with the same trap whatever its state. The table for testing for an immediate trap response is the same in both cases (table 17). It is the region of the table where departure is expected tath differs. In the second case, the whole lower left quarter of the table should exhibit high (resp. low) numbers observed in case of trap–happi-

Table 18. Component of TEST M.ITEC for the Canada geese relative to date 3. There is evidence of local trap–happiness with the number in bold greater than expected: Sdi–sj. Seen at day i and site j.

Tabla 18. Componente del TEST M.ITEC para la barnacla canadiense, relativa a la fecha 3. Existen pruebas de una adicción a la trampa local, siendo las cifras en negrita mayores de lo esperado: Sdi–sj. Visto el día i en la localidad j.

	Sd4–s1	Sd4–s2	Sd4–s3	Sd>4–s1	Sd>4–s2	Sd>4–s3
Last seen at site 1	162	77	3	83	62	4
Last seen at site 2	85	427	13	69	292	7
Last seen at site 3	11	57	58	10	43	35
Currently at site 1	**564**	200	8	202	162	7
Currently at site 2	125	**1,017**	36	82	471	20
Currently at site 3	7	45	**178**	12	48	65

ness (resp. trap–shyness); in the first case, only the diagonal in the same quarter is expected to be affected. The table corresponding to date 3 for the Canada geese is given in table 18.

There is evidence of local trap–happiness in the geese with a systematic excess on the diagonal of the lower left quarter of the table (table 18) and a globally highly significant test ($X^2(27) = 68.177$; $P < 10^{-4}$) (table 19).

Discussion

Goodness–of–fit testing is not the most popular part of a capture–recapture analysis, probably because it is neither automatic nor very appealing. Although, some automatic procedures, such as the bootstrap procedure built in MARK (White, 2001), are available, they have their limitations (White, 2002) and above all do not suggest what may be wrong when a model is rejected. On the other hand, optimal goodness–of–fit procedures exist only for a very

limited number of models, and have long been entirely missing for the multistate models. However, we believe that such procedures can be made more user–friendly and interpretable than they currently are, and that they have a great potential in helping understand capture–recapture data. There is certainly a lot of work yet to be done in this direction, but we have tried to show in this paper that there is already a lot to be learned from them. We have shown in particular that the goodness–of–fit multistate test of the JMV model as proposed by Pradel et al. (2003) can be partitioned in subcomponents directly related to some frequent violations of the assumptions (transience, trap–dependence, memory). Some

Table 19. Results of TEST M.ITEC for the Canada geese: Oc. Occasion.

Tabla 19. Resultados del TEST M.ITEC para la barnacla canadiense: Oc. Ocasión.

X^2	P	df	Oc
14.242	0.114	9	2
30.837	0.000	9	3
23.098	0.006	9	4

Table 20. Overdispersion factor calculated for different components (straight police) and subcomponents (italics) of the JMV and CJS goodness–of–fit tests.

Tabla 20. Factor de sobredispersión calculado para diferentes components (redondilla) y subcomponentes (cursiva) de los test de bondad de ajuste JMV (primeras dos columnas) y CJS (últimas dos columnas).

	JMV		CJS
3G	7.25		
3GSR	*9.8*	*13.5*	*3SR*
WBWA	*23.64*		
M	2.03		
M.ITEC	*2.53*	*15*	*2.CT*

of these effects can be treated more (transience) or less (trap–dependence) easily by fitting appropriate multistate models; others (memory) call for entirely different models. The fit of the alternative model to be used in case of transience, namely the generalized JMV model with two age classes on survival, is itself exactly testable by deducting TEST 3.SR from the overall goodness–of–fit test of JMV. Thus, this is a new model for which an optimal goodness–of–fit test is available. It would be interesting to examine whether a model with an optimal goodness–of–fit test could be identified as well in case of trap–dependence. The memory effect is a more difficult challenge. If this effect is strong, like for the Canada geese, all multistate models are invalidated. However, if it is weak, it can be kept out of the structural part of the model provided an overdispersion factor (the ratio of the X^2 statistic to its number of degrees of freedom) calculated from the goodness–of–fit tests is used in the analysis. An overdispersion factor can be used more generally whenever there is no obvious structural explanation for a lack of fit. Suitably partitioned goodness–of–fit tests are thus a more general tool than initially apparent for a correct assessment of the situation.

Beyond their purely technical usage, partitioned goodness–of–fit tests can serve to unveil some biological information. For instance, the intensity of transit is likely related to dispersion (Perret et al., 2003; Cam et al., 2004); heterogeneity of capture (a test of which has yet to be incorporated within a general goodness–of–fit test) may be a reflection of the intensity of social structuration; the role of memory helps understand how the organism apprehends its environment. This potential has yet to be fully exploited. The analysis of the Canada goose data set that we have used throughout this paper yields examples of the insight gained from the different components and subcomponents of the goodness–of–fit tests. A simple way to rank the relative strength of different effects is to calculate an overdispersion factor per components or subcomponents of the CJS and the JMV goodness–of–fit tests (table 20).

A first remark is that the corresponding subcomponents for transients, 3.SR and 3G.SR, and particularly trap–dependence, 2.CT and M.ITEC, present higher overdispersion coefficients in the one–site than in the multisite context. Obviously, taking account of the location has removed part of the heterogeneity. This is not surprising as encounter probabilities tend to be higher at some sites and at the same time the geese exhibit a high fidelity to their wintering sites; hence, the same individual geese tend to be consistently reencountered. The examination of the subcomponents of the multisite TEST 3G reveals in turn that memory is by far the most important cause of departure confirming the need for specific generalized models (Hestbeck et al., 1991; Brownie et al., 1993). Going through the occasion– and site–specific tables, we have also gained along the way new insights into the data: transit seems to affect only the peripheral sites and

trap dependence is more precisely local trap happiness. All this information has been obtained without fitting a single model so that, at the onset of modelling, we know for instance that a model with transients on the two peripheral sites is appropriate. The risk of overfitting, which must be kept in mind, is limited here by the consistency of the effects through several occasions. Another safeguard is provided by the use of even more specialized tests more precisely targeting the alternative of interest. The z–score tests of transience and the Cohen's kappa test of memory are two examples, but more can be developed, in particular for the detection of trap–dependence.

Although the Arnason–Schwarz model (AS) is generally considered as the reference for multistate analyses, we have not examined it specifically. This is because there is currently no specific goodness–of–fit test for it. The best approach is to treat the AS model as a particularization of the JMV model. After assessing the fit of JMV, JMV and AS can be fitted using program M–SURGE (Choquet et al., 2004) and the two models compared with the AIC criterion (possibly modified to incorporate an overdispersion factor). However, there is no more a priori reason to fit the AS model than any other multistate model. At present, the most urgent need is the study of the statistical properties of the new tests, notably the tests of mixture. For instance, in presence of sparse data, there is no equivalent to the Fisher's exact test. Another very promising extension is the use of the multistate tests with recovery data. This is possible because recoveries can be presented as multistate data with two states: 'alive' and 'dead' (Lebreton et al., 1999). However, as the state 'dead' is absorbing, the tests have first to be modified accordingly. There are more generally various potential original applications of the non–parametric tests presented in this paper (see for instance Gauthier et al. (2001) for seasonal trap–dependence). We believe that these tests should no longer be considered only as the necessary routine first step of a capture–recapture analysis but also as an important part of the analysis itself, contributing in ways that the parametric modelling cannot always do to the understanding of the data.

References

Arnason, A. N., 1972. Parameter estimates from mark–recapture experiments on two populations subject to migration and death. *Researches on Population Ecology*, 13: 97–113.
– 1973. The estimation of population size, migration rates and survival in a stratified population. *Researches on Population Ecology*, 15: 1–8.
Balser, J. P., 1984. *confidence interval estimation and tests for temporary outmigration in tag–recapture studies*. Doctoral dissertation, Cornell Univ., Ithaca, New York.
Begon, M., 1983. Abuses of mathematical techniques in ecology: applications of Jolly's cap-

ture–recapture method. *Oikos*, 40: 155–158.

Brownie, C., Hines, J. E., Nichols, J. D., Pollock, K. H. & Hestbeck, J. B., 1993. Capture–recapture studies for multiple strata including non–Markovian transitions. *Biometrics*, 49: 1173–1187.

Brownie, C. & Robson, D. S., 1983. Estimation of time–specific survival rates from tag–resighting samples: a generalization of the Jolly–Seber model. *Biometrics*, 39: 437–453.

Burnham, K. P. & Anderson, D. R., 1998. *Model selection and inference: a practical information–theoretic approach*. Springer–Verlag, New York.

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

Cam, E., Oro, D., Pradel, R. & Jimenez, J., 2004. Assessment of hypotheses about dispersal in a long–lived seabird using multistate capture–recapture models. *Journal of Animal Ecology*, 73: 723–736.

Carothers, A. D., 1971. An examination and extension of Leslie's test of equal catchability. *Biometrics*, 27: 615–630.

Choquet, R., Reboulet, A.–M., Pradel, R., Gimenez, O. & Lebreton, J.–D., 2003. *U–CARE (Utilities – CApture–REcapture)*. CEFE, Montpellier, France.

– 2004. M–SURGE: new software specifically designed for multistate capture–recapture models. *Animal Biodiversity and Conservation*, 27.1: 207–215.

Choquet, R., Reboulet, A.–M., Lebreton, J.–D., Gimenez, O. & Pradel, R., 2005. *U–CARE 2.2 User's Manual*. CEFE, Montpellier, France.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20: 37–46.

Everitt, B. S., 1977. *The analysis of contingency tables*. Chapman & Hall, Londres.

Gauthier, G., Pradel, R., Menu, S. & Lebreton, J.–D., 2001. Seasonal survival of greater snow geese and effect of hunting under dependence in sighting probabilities. *Ecology*, 82: 3105–3119.

Gimenez ,O., 2003. Estimation et tests d'adéquation pour les modèles de capture–recapture multiétats. Ph. D. thesis. Univ. Montpellier II, Montpellier.

Hestbeck, J. B., Nichols, J. D. & Hines, J. E., 1992. The relationship between annual survival and migration distance in mallards: an examination of the time–allocation hypothesis for the evolution of migration. *Canadian Journal of Zoology*, 70: 2021–2027.

Hestbeck, J. B., Nichols, J. D. & Malecki, R. A., 1991. Estimates of movement and site fidelity using mark–resight data of wintering canada geese. *Ecology*, 72: 523–533.

Lebreton, J.–D., Alméras, T. & Pradel, R., 1999. Competing events, mixture of information and multistrata recapture models. *Bird Study*, 46: 39–46.

Lebreton, J.–D., Burnham, K. P., Clobert, J. & Anderson, D. R., 1992. Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, 62: 67–118.

Lebreton, J.–D & Pradel, R., 2002. Multistate recapture models: modelling incomplete individual histories. *Journal of Applied Statistics*, 29: 353–369.

McCullagh, P. & Nelder, J. A., 1989. *Generalized linear models*, second edition. Chapman and Hall, New York, USA.

Nichols, J. D., Hines, J. E., Pollock, K. H., Hinz, R. L. & Link, W. A., 1994. Estimating breeding proportions and testing hypotheses about costs of reproduction with capture–recapture data. *Ecology*, 75: 2052–2065.

Perret, N., Pradel, R., Miaud, C., Grolet, O. & Joly, P., 2003. Transience, dispersal, and survival rates in newt patchy populations. *Journal of Animal Ecology*, 72: 567–575.

Pollock, K. H., Hines, J. E. & Nichols, J. D., 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*. Birkhaüser 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. & Lebreton, J.–D., 1999. Comparison of different approaches to the study of local recruitment of breeders. *Bird Study*, 46: 74–81.

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

Robson, D. S., 1969. *Mark–recapture methods of population estimation*. BU–168, Cornell Univ.

White, G. C., 2001. Advanced features of Program Mark. Pages 368–377 *in* Field, Warren & Sievert, editors. *Wildlife, Land, and People: Priorities for the 21st Century. Proceedings of the Second International Wildlife Management Congress.* The Wildlife Society, Bethesda, Maryland, USA.

– 2002. Discussion comments on: the use of auxiliary variables in capture–recapture modelling. An overview. *Journal of Applied Statistics*, 29: 103–106.