

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

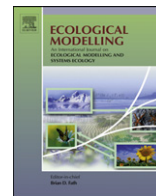
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Estimating dyad association probability under imperfect and heterogeneous detection

M. Javier Klaich^{a,b,*}, Paul G. Kinas^c, Susana N. Pedraza^{a,b}, Mariano A. Coscarella^{a,b}, Enrique A. Crespo^{a,b}^a Centro Nacional Patagónico, Boulevard Brown 2915, Puerto Madryn, Chubut, Argentina^b Universidad Nacional de la Patagonia San Juan Bosco, Boulevard Brown 3150, Sede Puerto Madryn, Chubut, Argentina^c Universidade Federal do Rio Grande, Avenida Itália, km 8, Rio Grande, RS, Brazil

ARTICLE INFO

Article history:

Received 18 December 2010

Received in revised form 24 March 2011

Accepted 25 March 2011

Available online 30 April 2011

Keywords:

Animal behaviour analysis

Association index

Individual detection

Dyad association probability

Information theory

Likelihood based model

Cephalorhynchus commersonii

ABSTRACT

In animal behaviour studies, association indices estimate the proportion of time two individuals (i.e. a dyad) spend in association. In terms of dyads, all association indices can be interpreted as estimators of the probability that a dyad is associated. However, traditional indices rely on the assumptions that the probability to detect a particular individual (p) is either approximately one and/or homogeneous between associated and not associated individuals. Based on marked individuals we develop a likelihood based model to estimate the probability a dyad is associated (ψ) accounting for $p < 1$ and possibly varying between associated and not associated individuals. The proposed likelihood based model allows for both individual and dyadic missing observations. In addition, the model can easily be extended to incorporate covariate information for modeling p and ψ . A simulation study showed that the likelihood based model approach yield reasonably unbiased estimates, even for low and heterogeneous individual detection probabilities, while, in contrast, traditional indices showed moderate to strong biases. The application of the proposed approach is illustrated using a real data set collected from a population of Commerson's dolphin (*Cephalorhynchus commersonii*) in Patagonia Argentina. Finally, we discuss possible extensions of the proposed model and its applicability in animal behaviour and ecological studies.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

When an individual identification process is involved in animal social studies, it is difficult to assure the perfect detection of each individual in the population. For instance, for species like tree-dwelling monkeys (Jack and Fedigan, 2004) or cetaceans (Williams and Thomas, 2009) the individual detection and identification is sometimes a difficult and complex task. Under these conditions, behavioural interactions, the foundation of the conceptual framework of Hinde (1976) for social analysis cannot be always observed. A common way around this difficulty is to use associations instead of interactions or in addition to interactions as the fundamental elements of social analysis. Association usually defined based on spatial or temporal proximity plus, perhaps, some behavioural state measure (Whitehead, 2008a). In this sense, the necessary conditions for considering two individuals in association must be established by the researcher prior the analysis and for the particular study (Michener, 1980).

The proportion of time two individuals spend in association is estimated by the so called association indices (AI) (Cairns and Schwager, 1987; Ginsberg and Young, 1992). In terms of dyads, all AI can be interpreted as estimators of the probability that a dyad is associated at a given time t (Cairns and Schwager, 1987; Whitehead, 2008b). Despite the existence of several types of AI, the estimation of any of these always requires that the individuals involved are individually identified by either artificial (e.g. via tags) (Glander et al., 1991) or natural (e.g. via photoidentification) (Wursig and Jefferson, 1990) marks. When the individual detection, defined as the probability to detect a particular individual (p) is perfect (i.e. $p = 1$), the Simple Index (SI) is the unbiased estimator of the probability that a dyad is associated and therefore, the unbiased estimator of the proportion of time two individuals spend together in association (Cairns and Schwager, 1987; Ginsberg and Young, 1992). Considering the potential sources of bias on SI, two issues have to be mentioned. In first place, when $p < 1$ (i.e. imperfect individual detection), two types of missing data can occur, when only one individual of the dyad is detected vs. when both individuals of the dyad are not detected. Thus, if only one or both individuals were not detected from a dyad, an uncertainty exists of whether the dyad was associated or not. When missing data is present, it is recommended the application of the Half Weight Index (HWI) instead of the SI due to the HWI lowers the effect of low detection (Cairns and

* Corresponding author at: Centro Nacional Patagónico – CONICET, Biología y Manejo de Recursos Acuáticos, Boulevard Brown 2915, U9120ACV Puerto Madryn, Chubut, Argentina. Tel.: +54 02965 451024; fax: +54 02965 451024.

E-mail addresses: mjklaich@cenpat.edu.ar, mjklaich@yahoo.com (M.J. Klaich).

Schwager, 1987; Whitehead, 2008a). The second issue is related to the presence of heterogeneity in individual detection. Heterogeneous individual detection is defined when the individual detection parameter (p) varies by individual (Link, 2003). For instance, in a population with size N , it is assumed N different p values. Another case of heterogeneity in individual detection is when there are g homogeneous groups (e.g. sex, ages) in the population, each with a different probability p (Norris et al., 1996; Pledger, 2000). The later represents the case when associated individuals show a different individual detectability than not associated individuals. The Twice Weight Index (TWI) was proposed to account for individuals having a higher individual detection probability when they are in association (Cairns and Schwager, 1987; Whitehead, 2008a). However, the presence of heterogeneity in individual detection is not necessarily confined to differences between associated and not associated individuals. For instance, young calves of cetacean species like *Tursiops truncatus* are often less identifiable than older animals and in some cases the presence of the mother is necessary to positively identify a calf (Gibson and Mann, 2008a,b). This represents another type of heterogeneity and so far no methodology have been proposed for dealing with such phenomenon in animal behaviour studies.

Under an inferential framework, Cairns and Schwager (1987) developed two different likelihood based estimators for estimating the AI as the probability that a dyad is associated at given sampling occasion when (a) all individuals in the population are equally detectable and (b) when a given group has a higher individual detectability than the others. Despite a clear improvement on the accuracy of these two different estimators, they are rarely used; the simplicity in calculating SI and HWI may explain its enduring preference over the likelihood based approach in the most animal behaviour studies. However, despite the extensive use and good performance in animal behaviour studies, the AI remains very much descriptive, without a possibility to either model or quantify a measure of uncertainty due to an imperfect individual detection.

Lusseau et al. (2008) pointed out the importance of sampling effects on AI and included that uncertainty in the estimation of weighted network measures; and Gibson and Mann (2009) concluded that when sociality patterns are studied, the results differ depending on the sampling method. Accounting for imperfect species detection, Chao et al. (2006) proposed a modification of the Dice's similarity index. Association indices like SI and HWI are derived from these methods for studying the co-occurrence of two or more species. Then, the proposed modification of Chao et al. (2006) would have straightforward adaptations to the indices used in the animal behavioural studies. In addition, several authors discuss the importance of accounting for and dealing with imperfect and heterogeneous individual detection (MacKenzie et al., 2002; Jennelle et al., 2007; MacNeil et al., 2008; Corkrey et al., 2008; Garamszegi et al., 2009; Cubaynes et al., 2010). Yet the explicit modelling of the imperfect and heterogeneous individual detection remains as a pending issue in quantitative animal behaviour analyses.

Here, we develop a new likelihood based model (LBM) for estimating the probability that a dyad is associated while accounting for individual detection probability less than one and heterogeneous individual detection probability. Our approach provides a flexible framework allowing for detected and not detected individuals, time-specific parameters and the inclusion of covariate information (e.g. biotic and abiotic factors) at individual, dyadic and group levels. Because we rely on a model-based inference approach, we test under the information theory competing sociobiological hypotheses by selecting the model best supported by the data (Burnham and Anderson, 2002). In addition, based on the proposed LBM, the model can be readily extended to a Bayesian framework in which a priori information could be included into the estimation of parameters and hypothesis testing (Link and Barker, 2010).

First, we explain how to build an encounter history for a dyad. This type of encounter history is analogous to an individual encounter history used in mark-recapture models (Williams et al., 2002) and the detection history for sites in occupancy models (MacKenzie et al., 2002). Second, we introduce the LBM for estimating the probability that a dyad is associated and an estimation procedure is proposed via the maximum likelihood approach. A simulation study is carried out to investigate the performance of the proposed method. In particular, we compare the performance of our approach with two traditional AI (SI and HWI). Third, we illustrate our method using a real data set from a population of Commerson's dolphin (*Cephalorhynchus commersonii*) in Patagonia Argentina. Finally, we discuss possible extensions of the LBM and its applicability in animal behaviour and ecological studies.

2. Methods

2.1. The association history

In this section, we provide a description for building an encounter history for a dyad when in an animal behaviour study individually marked animals are used for estimating either the probability that a dyad is associated or the proportion of time two individuals spend in association. We explain how the encounter history for a dyad can be used for describing dyadic data across time when both imperfect and heterogeneous individual detection are present.

Suppose a population closed to birth, immigration, mortality and emigration; and for which a given number of individuals can be individually identified by either artificial (e.g. via tags) or natural marks (e.g. via photoidentification). Then, either a recapture or a resight experiment is conducted with 4 sampling occasions in which an individual named " i " was detected at occasions 1, 3 and 4 and an individual named " j " was detected at occasions 1 and 4. This implies that both individuals were detected on occasions 1 and 4, at occasion 2 neither individual i nor individual j were detected and at occasion 3 only individual i was detected. In animal behaviour studies, additionally to the individual capture–recapture data, a definition for association is necessary under which two individuals are considered in association. In this case, we consider that individuals i and j are in association when they are detected in the same group and at the same sampling occasion. This is one among several definitions for association (see Whitehead, 2008a) and is known as "the gambit of the group" (Whitehead and Dufault, 1999). Suppose now, that individuals i and j were detected in the same group (i.e. in association) at occasion 1 and were detected in different groups (i.e. not in association) at occasion 4. Note that given the detection of both individuals is a necessary condition for considering two individuals either in association or not in association, it is not possible to define the association state for individuals i and j at occasions 2 and 3. The information related with both association and detection events across the 4 sampling occasions for individuals i and j can be expressed in a single encounter history in terms of dyads as:

Sampling occasion:	1	2	3	4
Encounter history:	$A_{(ij)}$	$0_{(ij)}$	$I_{(i)}$	$B_{(ij)}$

Where at occasion 1, $A_{(ij)}$ indicates that the dyad composed of the individuals i and j was observed as associated; at occasion 2, $0_{(ij)}$ indicates that neither individual i nor individual j were detected; at occasion 3, $I_{(i)}$ indicates that only individual i of the dyad was detected and at occasion 4, $B_{(ij)}$ indicates that the dyad composed of the individuals i and j was observed as not associated. Note that notations $A_{(ij)}$ and $B_{(ij)}$ imply that both individuals from the dyad were detected being in association and not in association respectively; and that notations $0_{(ij)}$ and $I_{(i)}$ imply that given both individuals were not detected or only one individual

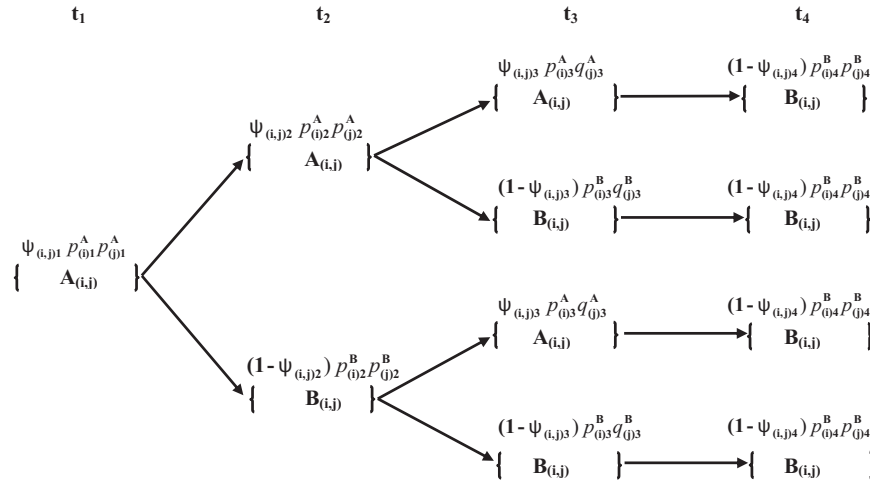


Fig. 1. The possible event probability statements when the association history (*ah*) $A_{(ij)}O_{(ij)}I_{(i)}B_{(ij)}$ is observed. Between brackets are either the observed or the potential association state for the dyad composed of individuals *i* and *j*. Above of each of those, are the probability expressions given the events in the observed *ah*. The dyad composed of individuals *i* and *i* is observed as associated and both individuals are detected at occasion 1 (column t_1). Neither individual *i* nor individual *j* are detected at occasion 2 and the actual association state remains unknown; thus the dyad may be either associated or not associated (column t_2). At occasion 3 only individual *i* is detected; again the actual association state remains unknown and the dyad may be either associated or not associated (column t_3). At occasion 4 both individuals are detected and the dyad is observed as not associated (column t_4). Then, given the observed *ah*, four alternatives *ah* were generated.

was detected respectively, the actual association state of the dyad remains unknown. To distinguish from other types of encounter histories, we named this encounter history as the association history (*ah*) for a particular dyad.

2.2. The likelihood based model formulation

In this section, we describe the LBM formulation for dyadic data by using a series of probabilistic arguments for modelling an *ah* for a particular dyad. To achieve this, we based the LBM formulation on the fundamentals of mark-recapture models (Williams et al., 2002) and occupancy models (MacKenzie et al., 2006). In this sense, the dyad is considered as the equivalent of an individual in mark-recapture analyses or as a site in occupancy analyses. In our case, modelling an *ah* involves the combination of information at both, dyadic and individual levels. This is achieved by using the association parameter $\psi_{(ij)k}$; defined as the probability that the dyad composed of the *i*th and *j*th individual (with *i*th \neq *j*th) is associated at sampling occasion *k*; and the individual detection parameter $p_{(i)k}^s$; defined as the probability to detect the *i*th individual (with *i*th \neq *j*th) at sampling occasion *k* given that it belongs to an associated dyad (*s* = *A*) or to a not associated dyad (*s* = *B*). Assuming time and individual specific parameters, the contribution to the whole data likelihood of the dyad composed of individuals *i* and *j* with association history $A_{(ij)}O_{(ij)}I_{(i)}B_{(ij)}$ is:

$$\psi_{(ij)1}p_{(i)1}^Ap_{(j)1}^A [\psi_{(ij)2}q_{(i)2}^Aq_{(j)2}^A + (1 - \psi_{(ij)2})q_{(i)2}^Bq_{(j)2}^B] \\ [\psi_{(ij)3}p_{(i)3}^Aq_{(j)3}^A + (1 - \psi_{(ij)3})p_{(i)3}^Bq_{(j)3}^B] \dots \times (1 - \psi_{(ij)4})p_{(i)4}^Bp_{(j)4}^B \quad (1)$$

where $q_{(i)k}^s = (1 - p_{(i)k}^s)$. The product of terms with subscript 1 is the probability that the dyad is associated and to detect both individuals given they are in association at sampling occasion 1 (t_1 column in Fig. 1). The expression with subscript 2 indicates the probability for the two possible situations when both individuals in the dyad are not detected. Then, the first term is the probability that the dyad is associated and both individuals are not detected given they are in association at sampling occasion 2. The second term is the probability the dyad is not associated and both individuals are not detected given they are not in association at sampling occasion 2 (t_2 column in Fig. 1). The next expression showing terms with sub-

script 3 represents the probability for all possible situations when only individual *i* is detected. The first term is the probability that the dyad is associated and individual *i* is detected while individual *j* is not given that both individuals are in association at sampling occasion 3; the second term is the probability that the dyad is not associated and individual *i* is detected while individual *j* is not given that both individuals are not in association at sampling occasion 3 (t_3 column in Fig. 1). The last product of terms with subscript 4 represents the probability that the dyad is not associated and to detect both individuals of the dyad given they are not in association at sampling occasion 4 (t_4 column in the Fig. 1).

Assuming the individuals are detected independently and parameters are time and individual specific, the likelihood of the data is:

$$L(\psi, \mathbf{p}^A, \mathbf{p}^B) = \prod_{i=1}^N \prod_{j=(i+1)}^N \prod_{k=1}^t (\psi_{(ij)k}p_{(i)k}^Ap_{(j)k}^A)^{\ln[A]_{(ij)k}} \\ ((1 - \psi_{(ij)k})p_{(i)k}^Bp_{(j)k}^B)^{\ln[B]_{(ij)k}} \dots \\ \times (\psi_{(ij)k}p_{(i)k}^Aq_{(j)k}^A + (1 - \psi_{(ij)k})p_{(i)k}^Bq_{(j)k}^B)^{\ln[I]_{(ij)k}} \\ (\psi_{(ij)k}p_{(i)k}^Aq_{(j)k}^A + (1 - \psi_{(ij)k})p_{(i)k}^Bq_{(j)k}^B)^{\ln[I]_{(ij)k}} \dots \\ \times (\psi_{(ij)k}q_{(i)k}^Aq_{(j)k}^A + (1 - \psi_{(ij)k})q_{(i)k}^Bq_{(j)k}^B)^{\ln[O]_{(ij)k}} \quad (2)$$

where *N* is the total number of detected individuals, *t* is the number of sampling occasions, $\ln[A]_{(ij)k}$, $\ln[B]_{(ij)k}$, $\ln[I]_{(ij)k}$, $\ln[I]_{(ij)k}$ and $\ln[O]_{(ij)k}$ are the usual indicator functions for the *i*th and *j*th individual at occasion *k* for the observational events $A_{(ij)}$, $B_{(ij)}$, $I_{(i)}$, $I_{(j)}$ and $O_{(ij)}$ respectively. For instance, if the dyad composed of the *i*th and *j*th individuals are detected in association at occasion *k* the indicator function $\ln[A]_{(ij)k}$ will take value 1 and $\ln[A]_{(ij)k} = 0$ otherwise.

Note that the formulated likelihood in (2) contains $t[N(N-1)/2 + 2N]$ parameters. For instance, if *N* = 40 and *t* = 4, the total number of parameters to be estimated from the model in (2) will be 3440. In practice, this model could not be fitted to the data because the likelihood model is over-parameterized. However, the model is formulated in general terms in order to allow for detection and/or association parameters modelling by means of either discrete (e.g. sex, age) or continuous (e.g. weight, length) individual covariates (Pollock et al., 1984; Lebreton et al., 1992).

If the association parameter is assumed to be time-specific but homogeneous across individuals and the individual detection parameter is assumed to be time-specific and heterogeneous between associated and not associated individuals, then a reduced model containing $3t$ parameters is:

$$L(\psi, p^A, p^B) = \prod_{i=1}^N \prod_{j=(i+1)}^N \prod_{k=1}^t \left[\psi_k (p_k^A)^2 \right]^{\ln[A]_{(i,j)k}} \left[(1 - \psi_k) (p_k^B)^2 \right]^{\ln[B]_{(i,j)k}} \dots \times \left\{ 2 \left[\psi_k p_k^A q_k^A + (1 - \psi_k) p_k^B q_k^B \right] \right\}^{\ln[I]_{(i,j)k} + \ln[J]_{(i,j)k}} \left[\psi_k (q_k^A)^2 + (1 - \psi_k) (q_k^B)^2 \right]^{\ln[O]_{(i,j)k}} \quad (3)$$

A simpler model is obtained when the association parameters are assumed homogeneous across individuals and constant across time; and individual detection parameter vary only between associated and not associated individuals. Then, a likelihood containing only 3 parameters is:

$$L(\psi, p^A, p^B) = \prod_{i=1}^N \prod_{j=(i+1)}^N \left[\psi (p^A)^2 \right]^{x_{A(i,j)}} \left[(1 - \psi) (p^B)^2 \right]^{x_{B(i,j)}} \left\{ 2 \left[\psi p^A q^A + (1 - \psi) p^B q^B \right] \right\}^{(y_{I(i)} + y_{J(j)})} \dots \times \left[\psi (q^A)^2 + (1 - \psi) (q^B)^2 \right]^{z_{O(i,j)}} \quad (4)$$

where $x_{A(i,j)}$ is the number of times the dyad composed of the i th and j th individuals was observed as associated ($k = 1, \dots, t$); $x_{B(i,j)}$ is the number of times the dyad composed of the i th and j th individuals was observed as not associated ($k = 1, \dots, t$); $y_{I(i)}$ is the number of times only individual i was observed from the dyad composed of the i th and j th individuals ($k = 1, \dots, t$); $y_{J(j)}$ is the number of times only individual j was observed from the dyad composed of the i th and j th individuals ($k = 1, \dots, t$); and $z_{O(i,j)}$ is number of times neither the i th nor j th individual were observed from the dyad they composed ($k = 1, \dots, t$). Applying an optimization algorithm on Eq. (4) along with $[x_{A(i,j)}, x_{B(i,j)}, (y_{I(i)} + y_{J(j)}), z_{O(i,j)}]$ as known data summaries, it is possible to estimate the association parameter ψ and detection parameters p^A and p^B .

By using different likelihood formulations as the examples presented in Eqs. (3) and (4), it is possible under the information theory to investigate the data support (Burnham and Anderson, 2002) for models containing for instance, either time-specific or constant across time parameters; or individual detection parameters being either heterogeneous between associated and not associated individuals or homogeneous across the individuals. In this way, hypotheses concerning the time variation of parameters and/or the parameters heterogeneity between associated and not associated individuals could be tested. Moreover, if either discrete or continuous covariates were available, then the individual heterogeneity could be investigated; and the homogeneity assumption could be tested by comparison the data support from models where parameters were or were not modeled using individual covariates.

3. Simulation study

3.1. Simulation methods

A simulation study was undertaken for evaluating the performance of the proposed LBM for estimating ψ and p under several scenarios. Specifically, we were interested to investigate the effects of five factors: the association probability in the population (ψ), the

association state specific individual detection probabilities (p^A and p^B), the number of sampling occasions (t) and the number of marked individuals in the population (M).

We investigated the effects of three levels of ψ by using 0.3, 0.5 or 0.7 in order to simulate populations showing different levels of association intensity among them. To study the effect of the individual detection for both individuals in association and individuals not in association, we used three levels, being 0.2, 0.4 or 0.6 for both p^A and p^B . The combinations between p^A and p^B represent the scenarios for which there is none, moderate and strong differences between individual detections. Is not uncommon that the number of sampling occasions (t) and/or the total number of marked individuals in the population (M) are restricted due to either weather or funds restrictions. Considering this situation, we investigated the performance of our method using values of 2, 5 or 10 for t and values of 20, 40 or 60 for M respectively. Note that when there are M marked individuals in the population, D is defined as the maximum number of different dyads that can be found across the t sampling occasions. In other words, D is the binomial coefficient of picking 2 individuals (i.e. a dyad) from M marked individuals in the population.

The datasets were simulated using a homogeneous association probability across time occasions and dyads; and the individual detection parameter remained constant across time and varied between associated and not associated individuals. Following the notation of Lebreton et al. (1992), the model to be fitted is expressed as $[\psi(\cdot), p^A(\cdot), p^B(\cdot)]$. The sign (\cdot) indicates that the parameter is constant across time while (t) indicates a parameter is time specific.

The combination of parameters levels yielded 243 scenarios, and for every one 1000 data sets were simulated. The data simulation process was divided in two steps: (1) assign the association state to each dyad across the sampling occasions and (2) apply the effect of individual detection given the association state of the dyad. For each sampling occasion and given M marked individuals in the population, D independent Bernoulli trials with parameter ψ were generated. If the generated number was 0, then the dyad was considered as not associated; if the generated number was 1 the dyad was considered as associated. Given the presence of an imperfect individual detection, for each sampling occasion it was simulated the number of individuals that were detected in each dyad. The number of detected individuals in each dyad was generated from a binomial distribution with parameters 2 and p^A for associated dyads or 2 and p^B for not associated dyads. If the generated number was 0, both individuals of the dyad were not detected, 1 indicated that only one individual was detected of the dyad and 2 indicated that both individuals of the dyad were detected. After t sampling occasions, the values of $[x_{A(i,j)}, x_{B(i,j)}, (y_{I(i)} + y_{J(j)}), z_{O(i,j)}]$ were recorded for each dyad of the N observed dyads. Once this information was recorded, model $[\psi(\cdot), p^A(\cdot), p^B(\cdot)]$ was fitted to the simulated data set. For each simulated data set, the estimated values for ψ , p^A and p^B using model $[\psi(\cdot), p^A(\cdot), p^B(\cdot)]$ were recorded and asymptotic variances were calculated using the second partial derivative of the model likelihood (Lebreton et al., 1992). We developed an R script (R Development Core Team, 2010) for the simulation study, made available in the Supplementary file 1.

In order to compare the performances of the proposed LBM and the SI and HWI indices, the values for SI and HWI for the each dyad were calculated following Cairns and Schwager (1987) as:

$$SI_{(i,j)} = \frac{x_{A(i,j)}}{x_{A(i,j)} + x_{B(i,j)} + (y_{I(i)} + y_{J(j)})}, \text{ and} \\ HWI_{(i,j)} = \frac{x_{A(i,j)}}{x_{A(i,j)} + x_{B(i,j)} + 0.5 * (y_{I(i)} + y_{J(j)})}$$

As a measure of the accuracy for each estimator; the Mean Square Error (MSE) was used for comparing the performance of the proposed LBM with the traditional SI and HWI indices across the 243 scenarios used. The MSE for the mean of each estimator in the r th scenario ($\hat{\alpha}_r$) was calculated as: $MSE(\hat{\alpha}_r) = \text{var}(\hat{\alpha}_r) + (\alpha_r - \hat{\alpha}_r)^2$, where $\text{var}(\hat{\alpha}_r)$ is the variance of $\hat{\alpha}_r$ and $(\alpha_r - \hat{\alpha}_r)^2$ is the square bias of $\hat{\alpha}_r$. The values of $\text{var}(\hat{\alpha}_r)$ and $(\alpha_r - \hat{\alpha}_r)^2$ were calculated across the 1000 estimates of $\hat{\alpha}_r$ for the r th scenario where the real value of the parameter is α_r .

3.2. Simulation results

Fig. 2 shows the results for the means of LBM ($\hat{\psi}$), SI (\bar{SI}) and HWI (\bar{HWI}) as the estimates of ψ for scenarios where $\psi = 0.3, 0.5, 0.7$; $M = 40$; $t = 2, 10$; $p^A = 0.2, 0.6$ and $p^B = 0.2, 0.4, 0.6$; but these are representative of the general simulations results. The full simulation results are included in the Supplementary file 2.

When individual detection was either homogeneous ($p^A = p^B$) or heterogeneous ($p^A \neq p^B$) the LBM provided reasonably unbiased estimates of the probability that a dyad is associated (ψ) in most scenarios considered. However, the precision of $\hat{\psi}$ was affected by all factors except by the values of ψ . The higher t values, the higher the precision of $\hat{\psi}$. The higher the M values the higher precision observed for $\hat{\psi}$. In addition, a slightly improvement in the accuracy for $\hat{\psi}$ was observed for higher values of M . Concerning precision and bias of $\hat{\psi}$, interactions between M and p^A, p^B, ψ and t values were not observed. For scenarios where homogeneous individual detection was present, the lower the p values, the lower the precision of $\hat{\psi}$. Fig. 2 shows that the lowest precisions of $\hat{\psi}$ were obtained for scenarios where $p^A = p^B = 0.2$ and $t \leq 10$; and where $p^A = 0.2, p^B = 0.4$ and $t = 2$. In general, the higher the difference between p^A and p^B the higher the precision of $\hat{\psi}$.

Concerning the performance of \bar{SI} and \bar{HWI} as estimates of the probability that a dyad is associated, these have shown to be biased when either homogeneous ($p^A = p^B$) or heterogeneous ($p^A \neq p^B$) imperfect individual detection were present. In this sense, the number of sampling occasions and the individual detection probability showed to be the most influential factors on the estimates of \bar{SI} and \bar{HWI} . Fig. 2 shows that the higher the t and p values the lower the estimates of \bar{SI} and \bar{HWI} . The same effects of M observed in $\hat{\psi}$, were observed for \bar{SI} and \bar{HWI} . As observed in $\hat{\psi}$, precision of \bar{SI} and \bar{HWI} was affected by all factors except by the values of ψ and was clearly higher than $\hat{\psi}$ (Fig. 2).

Fig. 3 shows the results for the MSE calculated for $\hat{\psi}, \bar{SI}$ and \bar{HWI} in scenarios where $\psi = 0.3, 0.5, 0.7$; $M = 40$; $t = 2, 10$; $p^A = 0.2, 0.6$ and $p^B = 0.2, 0.4, 0.6$. In general, $\hat{\psi}$ showed to be more accurate for estimating ψ than \bar{SI} and \bar{HWI} . The calculated values of MSE for $\hat{\psi}$ showed to be smaller than MSE of \bar{SI} and \bar{HWI} in 162 of the 243 scenarios, which represent the % 66.6 of the total. This was observed for scenarios where $\psi \leq 0.5$ and $t = 2$; $\psi \geq 0.5, t = 10, p^A \geq 0.2$ and $p^B \geq 0.4$ (Fig. 3). For those scenarios where $t = 5, \hat{\psi}$ showed to be more accurate than \bar{SI} and \bar{HWI} when $\psi \neq 0.5$.

4. Example: Commerson's dolphins in Patagonian Northern Sea

In this section we illustrate the application of the proposed LBM for studying: (1) the time effect on both the association parameter (ψ) and the individual detection probability (p); and (2) the variation of p between associated and not associated individuals for a coastal dolphin species. The coastal waters near to the Chubut

Table 1

Model selection for the Commerson's dolphin data analysis. Eleven models were considered combining time and association effects on association and individual detection parameters. The notation used is: np: number of parameters, QAICc: small sample correction for Quasi Akaike Information Criterion, Δ QAICc: difference between the QAICc of the current model and the best model.

Model	np	QAICc	Δ QAICc
$\psi(\cdot), p^A(t) = p^B(t)$	6	20985.36	0.00
$\psi(\cdot), p^A(t), p^B(t)$	11	21191.44	206.08
$\psi(\cdot), p^A(t), p^B(\cdot)$	7	21229.59	244.22
$\psi(\cdot), p^A(\cdot), p^B(t)$	7	21233.14	247.78
$\psi(t), p^A(t), p^B(t)$	15	21291.46	306.10
$\psi(t), p^A(t), p^B(\cdot)$	11	21549.62	564.26
$\psi(t), p^A(\cdot), p^B(t)$	11	21818.25	832.89
$\psi(t), p^A(\cdot), p^B(\cdot)$	7	21990.06	1004.70
$\psi(t), p^A(t) = p^B(\cdot)$	6	22444.90	1459.53
$\psi(\cdot), p^A(\cdot), p^B(\cdot)$	3	22467.15	1481.79
$\psi(\cdot), p^A(\cdot) = p^B(\cdot)$	2	22489.83	1504.47

River mouth, at the Northern Sea of Patagonia Argentina ($43^\circ 20'S$, $65^\circ 00'W$), are frequently used by a local population of Commerson's dolphin (*C. commersonii*) (Coscarella et al., 2003). Using photo-identification techniques, 5 occasions of capture–recapture were performed between October 2 and October 13 of 2007. Using a 5 m rigid hulled boat powered by a 40 hp outboard, two researchers took photographs of dolphin's dorsal fins. The observed patterns of notches on the contour of the dorsal fin were used to identify the individuals (Wursig and Jefferson, 1990). Photographs were classified in four levels of quality: bad, regular, good and very good. For this analysis, only good and very good quality photographs were used. Following Coscarella et al. (2011), for each day two individuals were considered in association (i.e. the dyad is associated) when they were photo-identified at least once within a 44 min interval. Two individuals were considered not in association when they were never photo-identified at least once within a 44 min interval in a day.

Across the five sampling occasions, a total of 71 well marked dolphins were detected. This value for N detected dolphins yielded a total of 2485 observed association histories. Note that the total marked dolphins in the population (i.e. M) is unknown, thus (M choose 2 combinations minus 2485) association histories were not observed in this study. Given that each association history describes joint information from two individuals, dependence among the association histories could be present. In order to measure and correct for such dependence, the inflation factor c was estimated (Burnham and Anderson, 2002). Following White et al. (2002), the estimation of c was achieved by the ratio between the observed Pearson chi-square statistic (X^2_{Obs}) from the saturated model [$\psi(t), p^A(t), p^B(t)$] and the bootstrap mean Pearson chi-square statistic (X^2_B). The estimation of X^2_B ($B = 1000$) was based on the parametric bootstrap routine described by MacKenzie and Bailey (2004). In this case, the parametric bootstrap routine was used for generating association histories instead of occupation histories. The models were ranked according to the small sample correction of the quasi-AIC (QAICc) (Burnham and Anderson, 2002) (Table 1).

Due to the estimate of c was less than 1 we set $c = 1$ according to Burnham and Anderson (2002). According to the Δ QAICc values, the first model has a large data supported than the other models. The best supported model indicated that the association probability was constant across time [$\psi(\cdot)$] while the individual detection probability was time specific but homogeneous regarding the association state of the individuals [$p^A(t) = p^B(t)$]. The estimates obtained from the best supported model indicated that the probability that any two dolphins are associated in a day is 0.6356 (se = 0.01679); and that the individual detection probability was less than one and time-specific with values of $\hat{p}_1 = 0.2676$ (se(\hat{p}_1) = 0.0063), $\hat{p}_2 =$

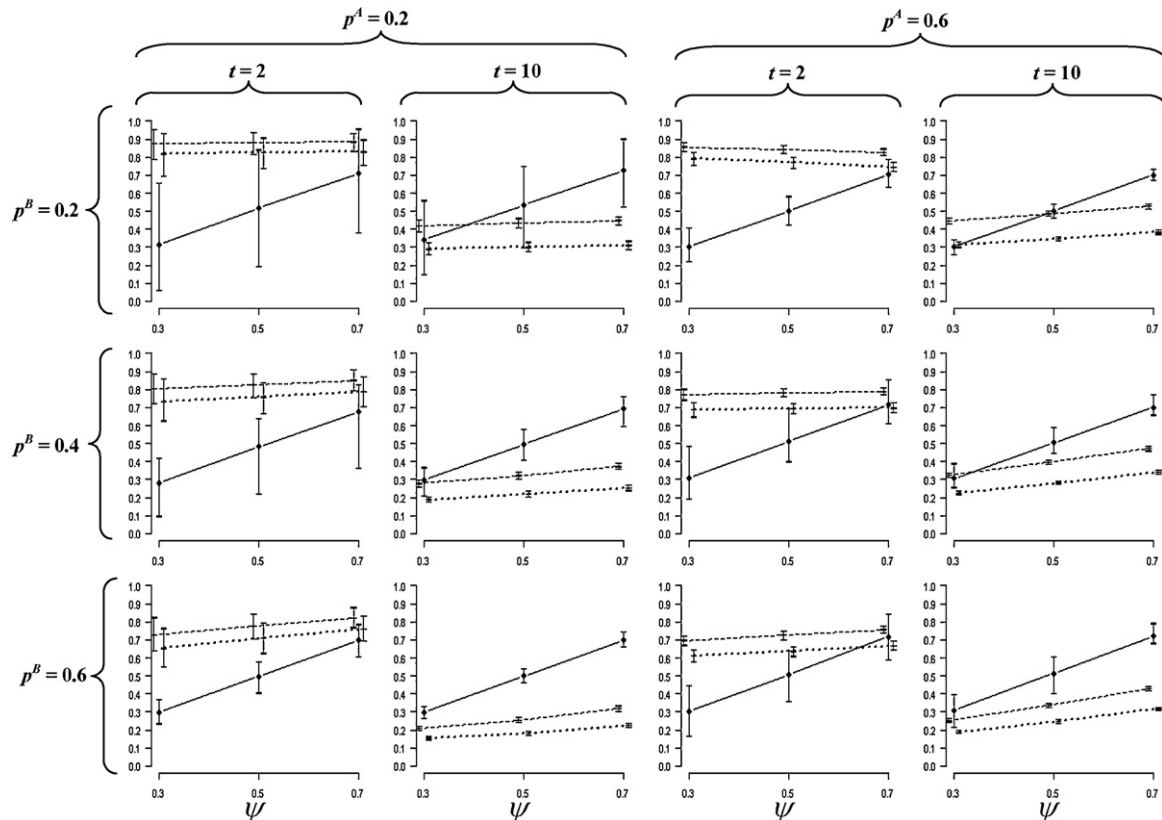


Fig. 2. Results of the simulation study for twelve scenarios. Each single plot shows the relation between the mean of the likelihood based model approach ($\hat{\psi}$) (solid line), the mean Half Weight Index (dashed line) and the mean Simple Index (dotted line) vs. the real value of $\psi = (0.3, 0.5, 0.7)$ in each scenario. Each row of plots shares the same value of p^B . When $p^A = 0.2$, first and second column shows the results for simulation using $t = 2$ and $t = 10$ respectively. When $p^A = 0.6$, third and fourth column show the results for simulation using $t = 2$ and $t = 10$ respectively. Vertical lines represent the 95% confidence interval estimated at the 2.5 and the 97.5 percentiles using the 1000 values of ψ for each scenario.

0.1125 ($se(\hat{p}_2) = 0.0045$), $\hat{p}_3 = 0.4444$ ($se(\hat{p}_3) = 0.0070$), $\hat{p}_4 = 0.1808$ ($se(\hat{p}_4) = 0.0055$) and $\hat{p}_5 = 0.1971$ ($se(\hat{p}_5) = 0.0056$) for sampling occasions 1–5 respectively. The R code (R Development Core Team, 2010) for the model fit, model selection procedure for the fourteen models and the estimation of the inflation factor (c) is available in the [Supplementary file 3](#).

5. Discussion

In this paper, we propose a new LBM for estimating the probability that a dyad is associated under field situations when an imperfect and heterogeneous individual detection is present. Our simulations showed that the proposed LBM is highly accurate for estimating ψ in several scenarios showing either low or heterogeneous individual detection. Under these same scenarios, moderate to strong bias were observed in SI and HWI indices as estimators of the probability a dyad is associated. For some scenarios, it was observed a lower precision of the proposed LBM in comparison to SI and HWI indices. This could be an undesirable feature for those quantitative behavioural analyses in which the relative differences between AI's are used and therefore there is no concern about the actual value of the probability a dyad is associated. However, if these kinds of analyses were performed, the proposed LBM could be used as a complementary tool for testing the data support for the homogenous and almost perfect individual detection assumptions. Then, and even in the presence of an imperfect and heterogeneous individual detection, the researcher would have the necessary information for choosing the lowest possible biased AI. Often, the comparison

between two or more estimated social measures may become a difficult task because different methodologies and/or amount of information may yield different results (Gibson and Mann, 2009). Our simulation study showed that the estimates of the association probability from the proposed LBM showed a quite small variation across scenarios showing different combination of sampling condition variables like the number of sampling occasions, number of marked animals in the population and an imperfect and/or heterogeneous individual detection. These results may indicate that the proposed LBM would represent a robust tool for comparing two or more association probability estimates from populations where these were estimated under different sampling conditions.

When individuals are identified by natural marks (e.g. photo-identification or DNA fingerprints) misidentification may lead to a substantial bias in estimates from capture–recapture models (Yoshizaki et al., 2009). The formulation of the LBM in this work assumed that individuals are correctly identified across the whole study; and if natural marks were used for estimating the association probability, misidentification could be an issue. Therefore, we acknowledge the importance to allow in future works for identification error in the estimation of the association probability.

As it was stated in the introduction, the necessary conditions for considering two individuals in association must be established by the researcher prior the analysis and for each particular study (Michener, 1980). Therefore, the estimate of ψ will be conditional on the specific association definition currently used by the researcher. However, even with a specific association definition, the condition of either associated or not associated for a dyad can be erroneously assigned. This misclassification may

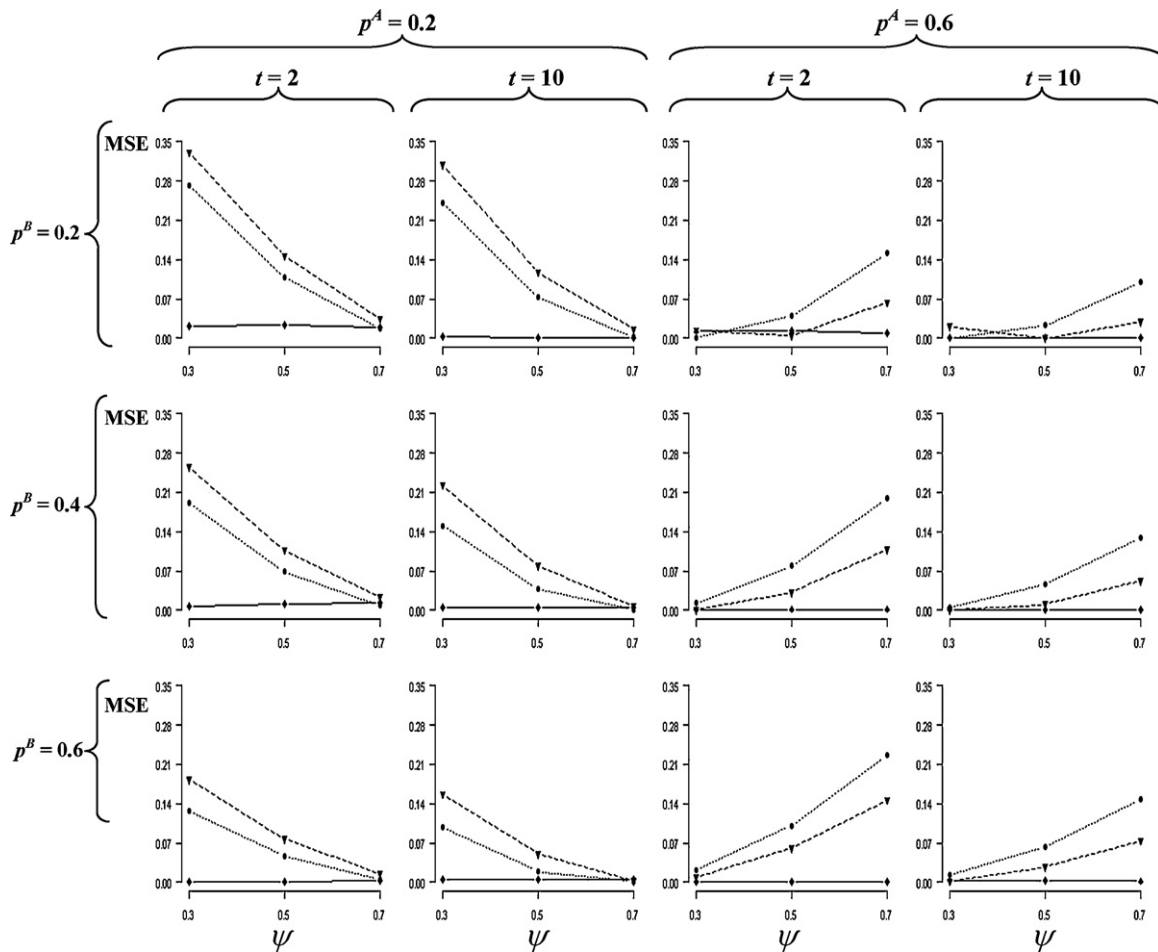


Fig. 3. Mean square error for the estimates from the likelihood based model, the simple index and the half weight index for twelve scenarios. Each single plot shows the relation between the mean square error (MSE) of the likelihood based model approach ($\hat{\psi}$) (solid line and diamond), the MSE of the Half Weight Index (dashed line and triangle) and the MSE of the Simple Index (dotted line and circle) vs. the real value of $\psi = 0.3, 0.5, 0.7$ in each scenario. Each row of plots shares the same value of p^B (0.2, 0.4, 0.6) and the number of marked individuals is constant ($M = 40$). When $p^A = 0.2$, first and second column show the results for simulation using $t = 2$ and $t = 10$.

be due to either misidentification or incomplete information to assign the correct association state. Pradel (2005) and Nichols et al. (2007) proposed two general frameworks for incorporating state uncertainty in capture-recapture multistate models and occupancy multistate models respectively. Based on these general frameworks, the estimates of ψ could be improved by taking association state uncertainty into account. However, more work is necessary to investigate this approach.

The LBM presented here assumes both geographic and demographic closure. However, association between two individuals could be affected by demographic processes like survival or migration across years or seasons (Whitehead, 2008a). A more complex model for estimating association probability could be developed incorporating individual survival and individual temporary migration parameters (Kendall et al., 1997; Schaub et al., 2004). If survival and migration are allowed, the activation or the deactivation of a particular dyad needs to be accounted for. This is related to situations in which either one or both individuals die or permanently emigrate, in which case the dyad is no longer active. However, a dyad can be temporarily deactivated when some of the individuals leave the study area and reactivated when those individuals return to the study area. Then, the estimation of ψ would be made conditionally on either the survival or the presence of the individuals in the study area.

Bejder et al. (1998) developed a method to test if individuals are associated in a nonrandom fashion or not. We have not addressed

the nonrandom association testing although we acknowledge the importance to address this issue in future works. However, given how an association history is built, it contains information about the randomness of association dynamics in a given population. This particular issue could be assessed by extending our model to consider the event a dyad is associated as a first-order Markov chain (see Schwarz and Schweigert (1993) for an application in the mark-recapture modelling framework). This would allow the estimation of the transition probability that a dyad is associated at occasion k given it was associated at occasion $k - 1$ (ψ_k^{AA}) and the transition probability that a dyad is associated at occasion k given it was not associated at occasion $k - 1$ (ψ_k^{BA}). The most important feature of these probabilities is that they give to us information about how an association state at the occasion $k - 1$ influences the probability to be in a given association state at the next occasion k . Then, a model with $\psi_k^{AA} = \psi_k^{BA}$ would assume that the previous association state does not influence the next association state and therefore, the transitions between association states are random. On the other hand, a model with $\psi_k^{AA} \neq \psi_k^{BA}$ would assume that the previous association state does influence the next association state and therefore, the transitions between association states are not random (i.e. Markovian process).

Concerning the potential applications in other fields than behavioural ecology, in disease ecology, the imperfect and heterogeneous individual detection of the infectious status of individuals is an increasingly recognized problem (McClintock et al., 2010). For

instance, the estimation of dyad association probabilities between two specific diseases in a given individual or between individuals would be useful information in studies addressing the potential relation between diseases and disease dynamics within social groups. For community ecology studies, MacKenzie et al. (2006) proposed a LBM for estimating the co-occurrence species probability when species are imperfectly detected for a given number of inspected sites. Based on individually marked individuals of 2 or more species, the proposed LBM would allow the estimation of the dyad association probability between individuals from different species instead of different individuals from the same species. In this way, it would be possible to study the data support for models representing competing hypotheses concerning either behavioural or community ecology processes.

6. Conclusion

The proposed LBM represents the first approach for dealing with imperfect and heterogeneous individual detection in quantitative studies of animal behaviour when marked individuals are involved. When imperfect individual detection is not accounted for in the estimation of the probability that a dyad is associated, such as is the case for traditional association indices, biased estimates are obtained. We have shown that a LBM that explicitly models the imperfections and heterogeneity of individual detection yields reasonably unbiased estimates of the probability a dyad is associated. In addition, the proposed LBM can be extended for including external information and/or demographic parameters such as survival, movement and migration rates. Finally, the application of the proposed approach for testing ecological hypotheses in other fields than behavioural ecology would be possible due to the general formulation of the model.

Acknowledgements

The authors wish to express their gratitude to Olivier Gimenez, Michael Conroy and Rob Williams for their useful comments which improved an earlier version of the manuscript. Also, the authors wish to thank two reviewers for their helpful comments and suggestions. This work was funded by the BBVA Foundation BIOCON04 and Cetacean Society International.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ecolmodel.2011.03.027.

References

- Bejder, L., Fletcher, D., Brøger, S.A., 1998. Method for testing association patterns of social animals. *Animal Behaviour* 56, 719–725.
- Burnham, K.P., Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York.
- Cairns, S.J., Schwager, S.J., 1987. A comparison of association indices. *Animal Behaviour* 35, 1454–1469.
- Chao, A., Chazdon, R.L., Colwell, R.K., Shen, T.-J., 2006. Abundance-based similarity indices when there are unseen species in samples. *Biometrics* 62, 361–371.
- Coscarella, M.A., Dans, S.L., Crespo, E.A., Pedraza, S.N., 2003. Potential impact of dolphin watching unregulated activities in Patagonia. *Journal of Cetacean Research and Management* 5 (1), 77–84.
- Coscarella, M.A., Gowans, S., Pedraza, S.N., Crespo, E.A., 2011. Influence of body size and ranging patterns on delphinid sociality: associations among Commerson's dolphins. *Journal of Mammalogy* 92 (3).
- Corkrey, R., Brooks, S., Lusseau, D., Parsons, K., Durban, W.J., Hammond, P.S., Thompson, P.M., 2008. A Bayesian capture–recapture population model with simultaneous estimation of heterogeneity. *Journal of American Statistics Association* 103 (483), 948–960.
- Cubaynes, S., Pradel, R., Choquet, R., Duchamp, C., Gaillard, J.-M., Lebreton, J.-D., Marboutin, E., Miquel, C., Reboulet, A.-M., Poillot, C., Taberlet, P., Gimenez, O., 2010. Importance of accounting for detection heterogeneity when estimating abundance: the case of French wolves. *Conservation Biology* 24 (2), 621–626.
- Garamszegi, L.Z., Calhim, S., Doctermann, N., Hegyi, G., Hurd, P.L., Jørgensen, C., Kutsukake, N., Lajeunesse, M.J., Pollard, K.A., Schielzeth, H., Symonds, M.R.E., Nakagawa, S., 2009. Changing philosophies and tools for statistical inferences in behavioural ecology. *Behavioural Ecology* 20, 1363–1375.
- Gibson, Q.A., Mann, J., 2008a. Early social development in wild bottlenose dolphins: sex differences, individual variation, and maternal influence. *Animal Behaviour* 76, 375–387.
- Gibson, Q.A., Mann, J., 2008b. The size, composition, and function of wild bottlenose dolphin (*Tursiops* sp.) mother–calf groups in Shark Bay, Australia. *Animal Behaviour* 76, 389–405.
- Gibson, Q.A., Mann, J., 2009. Do sampling method and sample size affect basic measures of dolphin sociality? *Marine Mammals Science* 25 (1), 187–198.
- Ginsberg, J.R., Young, T.P., 1992. Measuring association between individuals or groups in behavioural studies. *Animal Behaviour* 44, 377–379.
- Glander, K.E., Fedigan, L.M., Fedigan, L.M., Chapman, C.A., 1991. Capture techniques for three species of monkeys in Costa Rica. *Folia Primatologica* 57, 70–82.
- Hinde, R.A., 1976. Interactions, relationships and social structure. *Man* 11, 1–17.
- Jack, K.M., Fedigan, L., 2004. Male dispersal patterns in white-faced capuchins, *Cebus capucinus*. Part 1. Patterns and causes of natal emigration. *Animal Behaviour* 67, 761–769.
- Jennelle, C.S., Cooch, E.G., Conroy, M.J., 2007. Senar JCState-specific detection probabilities and disease prevalence. *Ecological Applications* 17, 154–167.
- Kendall, W.L., Nichols, J.D., Hines, J.E., 1997. Estimating temporary emigration using capture–recapture data with Pollock's robust design. *Ecology* 78, 563–578.
- Lebreton, J.D., Burnham, K.P., Clobert, J., Anderson, D.R., 1992. Modelling survival and testing biological hypotheses using marked animals. A unified approach with case studies. *Ecological Monographs* 62, 67–118.
- Link, W.A., 2003. Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities. *Biometrics* 59, 1123–1130.
- Link, W.A., Barker, R.J., 2010. *Bayesian Inference: With Ecological Applications*. Elsevier, Amsterdam.
- Lusseau, D., Whitehead, H., Gero, S., 2008. Incorporating uncertainty into the study of animal social networks. *Animal Behaviour* 75, 1809–1815.
- MacKenzie, D.I., Bailey, L.L., 2004. Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics* 9 (3), 300–318.
- MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Royle, J.A., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology* 83, 2248–2255.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L., Hines, J.E., 2006. *Occupancy Estimation and Modelling: Inferring Patterns and Dynamics of Species Occurrence*. CA Elsevier, San Diego.
- McClintock, B.T., Nichols, J.D., Bailey, L.L., MacKenzie, D.I., Kendall, W.L., Franklin, A.B., 2010. Seeking a second opinion: uncertainty in disease ecology. *Ecology Letters* 13, 659–674.
- MacNeil, M.A., Tyler, E.H.M., Fonnesbeck, C.J., Rushton, S.P., Polunin, N.V.C., Conroy, M.J., 2008. Accounting for detectability in reef-fish biodiversity estimates. *Marine Ecology Progress Series* 367, 249–320.
- Michener, G.R., 1980. The measurement and interpretation of interaction rates: an example with Richardson's ground squirrels. *Biology of Behaviour* 5, 371–384.
- Nichols, J.D., James, E.H., Mackenzie, D.I., Seamans, M.E., Gutiérrez, R.J., 2007. Occupancy estimation and modeling with multiple states and state uncertainty. *Ecology* 88, 1395–1400.
- Norris, J.L., Pollock, K.H., 1996. Nonparametric MLE under two closed capture–recapture models with heterogeneity. *Biometrics* 52, 639–649.
- Pledger, S., 2000. Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics* 56, 434–442.
- Pollock, K.H., Hines, J.E., Nichols, J.D., 1984. The use of auxiliary variables in capture–recapture and removal experiments. *Biometrics* 40, 329–340.
- Pradel, R., 2005. Multievent: an extension of multistate capture–recapture models to uncertain states. *Biometrics* 61, 442–447.
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2010. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Schaub, M., Gimenez, O., Schmidt, B.R., Pradel, R., 2004. Estimating survival and temporary emigration in the multistate capture–recapture framework. *Ecology* 85, 2107–2113.
- Schwarz, C.J., Schweigert, J.F., 1993. Arnason ANEstimating migration rates using tagrecovey data. *Biometrics* 49, 177–193.
- White, G.C., Burnham, K.P., Anderson, D.R., 2002. Advanced Features of Program MARK. In: Fields, R (Ed.), *Integrating People and Wildlife for a Sustainable Future: Proceedings of the Second International Wildlife Management Congress*. Bethesda, MD: The Wildlife Society.
- Whitehead, H., Dufault, S., 1999. Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Advances in the Study of Behaviour* 28, 33–74.
- Whitehead, H., 2008a. *Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis*. University of Chicago Press, Chicago.
- Whitehead, H., 2008b. Precision and power in the analysis of social structure using association. *Animal Behaviour* 75, 1093–1099.
- Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. *Analysis and Management of Animal Populations*. Academic Press, San Diego California, USA.

- Williams, R., Thomas, T., 2009. Cost-effective abundance estimation of rare animals: Testing performance of small-boat surveys for killer whales in British Columbia. *Biological Conservation* 142, 1542–1547.
- Wursig, B., Jefferson, T.A., 1990. Methods of photo-identification for small cetaceans. In: Hammond, P.S., Mizroch, S.A., Donovan, G.P. (Eds.), *Individual Recognition of Cetaceans: Use of Photo-identification and other Techniques to Estimate Population Parameters*, 12. Report of the International Whaling Commission, pp. 43–52, special issue.
- Yoshizaki, J., Pollock, K.H., Brownie, C., Webster, R.A., 2009. Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. *Ecology* 90 (1), 3–9.