

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biostatistique

École doctorale Information, Structures et Systèmes

Unité de recherche Centre d'Écologie Fonctionnelle et Évolutive

## Démographie des espèces animales sociales

Présentée par Lorena MANSILLA

Le 25 novembre 2019

Sous la direction de Roger PRADEL  
et Olivier GIMENEZ

Devant le jury composé de

Jean-Noël Bacro, Institut Montpelliérain Alexandre Grothendieck, Montpellier

Président du jury

Emmanuelle Cam, Laboratoire des Sciences de l'Environnement Marin, Brest

Rapportrice

Rachel McCrea, School of Mathematics, Statistics and Actuarial Science, University of Kent,  
Canterbury (Angleterre)

Rapportrice

Marie Nevoux, Écologie et Santé des Écosystèmes, Rennes

Examinateuse

Roger Pradel, Centre d'Écologie Fonctionnelle et Évolutive, Montpellier

Directeur de thèse

Olivier Gimenez, Centre d'Écologie Fonctionnelle et Évolutive, Montpellier

Co-directeur de thèse



UNIVERSITÉ  
DE MONTPELLIER

---

## **Agradecimientos**

Mis agradecimientos a R. Pradel y O. Giménez por ser los directores de esta tesis, a quienes nos prestaron los datos y colaboraron con nosotros para el desarrollo de los diferentes capítulos de esta tesis y que son autores y coautores de los artículos que lo componen. También agradezco a CONICYT (PFCHA 2016) por el financiamiento que me permitió realizar mi tesis, a mi familia por el apoyo durante estos tres años. Y por último, gracias a mi D' amado, *HaKadosh Baruj Hu* que hace posible todo.

## **Remerciements**

Mes remerciements à R. Pradel et O. Gimenez pour être les directeurs de cette thèse, à ceux qui nous ont prêté les données et collaboré avec nous pour le développement des différents chapitres de cette thèse et qui sont auteurs et co-auteurs des articles qui la composent. Mes remerciements aussi à CONICYT (PFCHA 2016) pour le financement pour développer ma thèse, à ma famille pour le soutien pendant ces trois années. Et enfin merci à mon D' bien aimé *HaKadosh Barukh Hu* qui rend tout possible.

---

## Résumé

Cette thèse est centrée sur l'étude de la dynamique des populations animales, avec comme principal objectif d'inclure des aspects de socialité dans les modèles démographiques. On qualifie des animaux de sociaux lorsque des individus s'associent à d'autres individus de la même espèce pas exclusivement pour se reproduire. Certains aspects de la socialité sont importants au cours des différentes étapes de la vie d'un individu, par exemple des événements de toilettage, d'agressivité, d'allaitement, etc. et d'autres aspects jouent à l'échelle de la population, comme le nombre et la taille des groupes par exemple pour le cas des animaux avec structure hiérarchique, le nombre de petits et la taille des harems, ou encore le temps d'association pour les cas des agrégations d'animaux. Dans cette thèse, j'étudie les composantes de la socialité et ses effets sur la dynamique des populations à travers, entre autres, des modèles de capture-recapture (CR) qui permettent d'estimer les principaux paramètres démographiques en conditions naturelles. On s'intéresse à trois espèces avec des structures sociales différentes. Pour une première espèce, le loup (*Canis lupus*), on développe une approche pour estimer la taille et le nombre de meutes grâce à une combinaison de modèles de capture-recapture spatialement explicites et de méthodes de hiérarchisation. Pour une deuxième espèce, l'éléphant de mer du sud (*Mirounga leonina*), on analyse l'influence de la structure sociale sur la dynamique des populations dans un cadre de modélisation intégrée. Sur un troisième cas d'étude, le dauphin de Commerson (*Cephalorhynchus commersonii*), nous modélisons le réseau social des animaux et estimons les principales caractéristiques de ce réseau (e.g., degré, centralité) grâce à une formulation à espace d'état des modèles de CR. Globalement, mes travaux illustrent le potentiel des méthodes modernes de dynamique des populations pour l'étude de la démographie des espèces sociales en conditions naturelles.

**Mots-clés:** Ecologie statistique, dynamique des populations, démographie, modèle de capture-recapture, socialité, structure social.

---

## Abstract

This thesis focuses on the study of social animal populations, with the main objective of including aspects of sociality in demographic models. Animals are termed social when individuals associate with other individuals of the same species not exclusively for reproduction. Some aspects of sociality are important in the different events of an individual's life, such as grooming, aggression, breastfeeding, etc. and other aspects at the population level, such as the number and size of groups, for example, in the case of animals with hierarchical structure, the number of cubs and the size of the harems, or the duration of association for cases of aggregations of animals. In this thesis, I study the components of sociality and its effects on population dynamics through capture-recapture (CR) models, which make it possible to estimate the main demographic parameters under natural conditions. We are interested in three species with different social structures. For a first species, the wolf (*Canis lupus*), an approach is developed to estimate the size and number of packs through a combination of spatially explicit capture-recapture models and hierarchical methods. For a second species, the Southern elephant seal (*Mirounga leonina*), we analyze the influence of the social structure on population dynamics in an integrated modeling framework. On a third case study, the Commerson's dolphin (*Cephalorhynchus commersonii*), we model the social network of animals and estimate the main characteristics of this network (e.g., degree, centrality) thanks to a state space formulation of CR models. Overall, my work illustrates the potential of modern population dynamics methods for studying the demography of social species under natural conditions.

**Keywords:** Statistical ecology, population dynamics, demography, capture-recapture model, sociality, social structure.



## Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
1.1	Qu'est-ce qu'un animal social? . . . . .	7
1.2	Pourquoi est-il nécessaire de considérer la socialité en démographie? . . . . .	7
1.3	La difficulté de lier la socialité à la démographie . . . . .	8
1.4	Objectifs de thèse . . . . .	8
1.5	Différentes structures sociales pour différentes espèces animales . . . . .	9
1.6	Les espèces dans nos études de cas . . . . .	9
1.7	Probabilité conditionnelle et règle de Bayes . . . . .	10
1.8	Modèle hiérarchique de probabilité . . . . .	11
1.9	Des modèles structurés de population . . . . .	11
1.9.1	Le modèle de Capture-Recapture spatialement explicite . . . . .	11
1.9.2	Modèle à espace d'état . . . . .	12
1.9.3	Modèle de Markov caché . . . . .	12
1.10	Méthodes de Monte Carlo par chaînes de Markov et algorithme de Gibbs . . . . .	13
1.10.1	L'intégration Monte Carlo . . . . .	13
1.10.2	Chaîne de Markov . . . . .	14
1.10.3	L'échantillonneur de Gibbs . . . . .	14
<b>2</b>	<b>OBJECTIF 1</b>	<b>15</b>
2.1	ARTICLE 1. Estimating the number of packs in wolf populations using spatially-explicit capture-recapture models and "clustering" methods . . . . .	15
<b>3</b>	<b>OBJECTIF 2</b>	<b>31</b>
3.1	ARTICLE 2. Integrated population modeling for social species: a case study with elephants seals . . . . .	31
<b>4</b>	<b>OBJECTIF 3</b>	<b>57</b>
4.1	ARTICLE 3. Inferring animal social networks with imperfect detection . . . . .	57
<b>5</b>	<b>Discussion</b>	<b>65</b>
	<b>References</b>	<b>71</b>
<b>A</b>	<b>A propos du Chapitre 2</b>	<b>71</b>
<b>B</b>	<b>A propos du Chapitre 3</b>	<b>87</b>



## 1.1 Qu'est-ce qu'un animal social?

Un animal social est un animal membre d'une société. Ces animaux sont liés à des individus de la même espèce de façon coopérative pendant tout ou partie de sa vie.

Tout comme les principales caractéristiques des systèmes vivants, le comportement coopératif est une propriété émergente (Wilson 1998). La société est formée par des unités fonctionnelles elles-mêmes formées par des membres de cette société en colonies, conglomérats, groupes ou harems. On trouve dans la littérature différentes terminologies liées à un système social, comme l'organisation sociale, la structure sociale et l'accouplement social, considérées comme des composants de ce système (e.g. Kappeler & van Schaik, 2002). Dans d'autres cas, et par simplicité, ces concepts sont présentés sans aucune distinction (e.g. Whitehead 2008).

## 1.2 Pourquoi est-il nécessaire de considérer la socialité en démographie?

Ignorer l'information sur le comportement animal dans les études de démographie des populations peut causer des erreurs dans les résultats et les prédictions, et mener à des politiques de gestion inadaptées pour la conservation. Nous allons voir en détail comment.

Le comportement des membres d'une société détermine les caractéristiques de la structure sociale, par exemple la taille des groupes, la cohésion, la perméabilité, la différenciation des rôles, le flux d'information, ou encore la fraction de temps consacré au comportement social (Wilson 1998).

Le comportement social, par exemple, est un élément critique qui doit être pris en compte dans les modèles de dynamique des populations d'espèces à structure sociale car il peut affecter la survie et la reproduction (Zeigler & Walters 2014).

Les fortes associations entre épaulards (*Orcinus orca*) à l'île de Crozet leur permet d'améliorer leurs performances individuelles (valeur sélective ou "fitness" en anglais), via l'augmentation du succès de la recherche de nourriture et la résilience face à des événements de mortalité par les activités humaines (Busson et al. 2019).

Des études ont montré que le risque d'extinction pour des espèces sociales comme le loup gris (*Canis lupus*), le chien sauvage africain (*Lycaon pictus*) et le pic à queue rouge (*Picoides borealis*) est corrélé négativement au nombre de groupes sociaux dans la population (Vucetich et al. 1997, Walters et al. 2002, Somers et al. 2008, cité par Zeigler et Walters 2014).

Incorporer les aspects de socialité en démographie permet d'améliorer la compréhension et les estimations des patrons de risque d'extinction, par exemple, à travers l'analyse de viabilité de population lié au comportement de sélection d'habitat (González-Suárez & Gerber 2008), et d'identifier des comportements signes d'une population en déclin (Gerber 2006). Cela permet aussi de prévoir de façon plus réaliste le devenir des populations en prenant en compte dans les taux démographiques des événements liés à la stochasticité démographique ou environnementale. On citera comme exemple le loup qui produit un faible nombre de petits par an, et sur lequel donc la stochasticité peut restreindre le nombre d'individus reproducteurs et augmenter le taux de

risque d'extinction (cité par Vucetich et al. 1997). Un autre exemple est la hyène tachetée: pour cette espèce, la probabilité de contagion d'un agent pathogène et de guérison des individus sont fortement dépendantes de la structure sociale hiérarchique (Benhaiem et al. 2018), et le statut social des individus permet d'expliquer la variation de ces paramètres clés (Marescot et al. 2018). Williams & Lusseau (2006) signalent l'importance de modéliser l'hétérogénéité des individus des épaulards pour pouvoir prendre des décisions de gestion en conformité avec l'organisation et les besoins des agrégations, et également de la nécessité de considérer les effets de la probabilité de recapture différente entre individus sur l'identification des liens sociaux.

### 1.3 La difficulté de lier la socialité à la démographie

Prendre en compte les aspects du comportement dans les modèles démographiques n'est pas trivial. Les études démographiques sur les espèces sociales sont relativement rares, alors que ces espèces sont fréquentes. Cela est dû principalement à la difficulté sur le terrain d'observer le comportement des animaux et à accumuler des données sur un temps relativement long pour pouvoir étudier la démographie (Caro 1998). Cette difficulté est directement en lien avec le problème de détectabilité imparfaite, autrement dit l'impossibilité de voir ou capturer avec certitude les individus d'une population (et a fortiori leur comportement) à tout moment (Gimenez et al. 2008). Ce problème d'une probabilité de détection inférieure à 1 constitue la principale motivation qui a mené aux développements méthodologiques proposés dans ma thèse.

En outre, les animaux ne se comportent pas tous de la même façon dans une population. Alors, comme l'écrit González-Suárez (2014), il convient de déterminer dans quelles situations on devrait prendre en compte les variations individuelles car il est crucial de comprendre comment les différentes composantes du comportement affectent les taux vitaux que sont la survie et la reproduction. En résumé, il est difficile de faire se parler dynamique des populations et écologie du comportement. Or ce lien est vital pour déterminer le statut de conservation des espèces sociales. Dans sa revue des études liant biologie de la conservation et comportement des animaux, Sutherland (1998) note que les journaux d'écologie du comportement incluaient peu ou pas les thèmes de la biologie de la conservation, et vice-versa.

### 1.4 Objectifs de thèse

Les objectifs de cette thèse sont les trois suivants: i) estimer le nombre de meutes dans des populations de loups avec des modèles de capture-recapture spatialement explicites et des méthodes de partitionnement; ii) estimer les paramètres démographiques des espèces sociales. Étude de cas sur les éléphants de mer; ii) décrire et quantifier les réseaux sociaux animaux quand la détectabilité est imparfaite.

Chacun d'eux est développé dans les trois chapitres ci-dessous, avec formats d'articles. Ci-dessous je définis les structures sociales des populations que j'ai étudié, quelques aspects généraux de ces populations, et les modèles qui ont été utilisés tout au long de l'élaboration des mes trois objectifs.

## 1.5 Différentes structures sociales pour différentes espèces animales

Nous donnons ici les définitions des structures sociales abordées dans cette thèse (source : Wilson 1998).

**Groupes:** un ensemble d'organismes appartenant à la même espèce qui restent ensemble durant certaines périodes de temps tandis qu'ils interagissent les uns avec les autres à un plus haut degré qu'avec d'autres conspécifiques. Le vocable groupe est utilisé pour se référer à une agrégation, à un type de société ou à un sous-ensemble d'une société. Il est utilisé spécialement dans les descriptions de certaines sociétés de primates dans lesquelles il existe un niveau d'organisation hiérarchique construit par des sous-ensembles d'individus appartenant à la même grande et unique congrégation.

**Harem:** un groupe de femelles gardées par un mâle qui empêche les autres mâles de s'accoupler avec elles.

**Agrégation:** un groupe d'individus de la même espèce, formé par un peu plus qu'une paire ou une famille, réunis dans le même endroit; cette structure n'a toutefois pas une organisation interne ou un comportement coopératif. Cette structure peut fournir une protection pour ses membres.

## 1.6 Les espèces dans nos études de cas

### Loups

**Socialité et territorialité:** Les loups (*Canis lupus*) sont des animaux sociaux qui vivent en meutes avec une structure sociale hiérarchique, formées par un couple reproductif dominant dit alpha, et des individus qui restent dans l'état non reproducteur, qui sont généralement la progéniture de l'année du couple alpha (Mech & Boitani, 2003). Lorsque les juvéniles atteignent la maturité sexuelle, entre un et trois ans, ils dispersent et quittent la meute pour former d'autres meutes (Mech & Boitani, 2003). La taille de meute est généralement de cinq à six chiots (Mech, 1970). Le territoire et le domaine vital d'une meute de loups sont les mêmes. Puisque le territoire est le domaine vital défendu (Mech & Boitani, 2003).

**Nourriture:** Ils sont carnivores, et se nourrissent principalement d'ongulés en Europe (Mech & Boitani, 2003).

**Distribution:** L'espèce s'étend dans l'hémisphère nord, sur le continent européen, américain et asiatique.

**Densité et population:** Caniglia et al. (2012) signale une taille moyenne de population entre 117 et 233 par les années entre 2003 et 2007 respectivement. Cette population a une tendance constante d'augmenter à excepté 2003 et 2008.

Apollonio et al. (2004) ont étudié des cas locaux, avec 3 à 5 groupes, ils ont décrit une moyenne de taille de groupes de  $4.2 \pm 0.9$  loups. Aussi ils ont estimé une densité général pour le domaine de 4.7 individus par  $100 \text{ km}^2$  avec une distance moyenne de groupes adjacents de 11.1 km, sur données du nord des Apennins dans la péninsule italique, entre les années 1993 à 2000.

**Éléphants de mer**

**Socialité et territorialité:** Les éléphants de mer (*Mirounga leonina*) sont les plus grands pinipèdes et ont un dimorphisme sexuel significatif (Oosthuizen et al. 2019), les mâles adultes ont entre 3000 à 4000 kg, tandis que des femelles adultes ont entre 400 à 900 kg (Laws, 1953). Ils se rassemblent sur terre en des harems pendant la saison de reproduction (Oosthuizen et al. 2019). Les femelles ont un petit par an. Le mâle dominant arrive avant les femelles pour protéger son territoire.

**Nourriture:** Leurs proies principales sont les céphalopodes (Daneri et al, 2000, Hindell et al. 2003) et les poissons (Daneri et Carlini, 2002).

**Distribution:** Ils ont une distribution circumpolaire antarctique (Le Boeuf & Laws 1994).

**Population:** La population des éléphants de mer de l'Île Marion a diminué de 83% entre 1951 et 1994 (Laws, 1994), point à partir de quelle la population reste relativement stable (Pistorius et al. 2011), aujourd'hui environ 550 femelles reproductrices (Oosthuizen et al. 2019). Ils sont aujourd'hui catalogués comme une espèce en état global stable (UICN, 2014).

**Dauphins de Commerson** (*Cephalorhynchus commersonii*)

**Socialité et territorialité:** Les individus sont sociaux et sont rassemblés en agrégations.

**Distribution:** Il y a des observations de dauphins de Commerson (*Cephalorhynchus commersonii*) dans les océans de l'hémisphère sud, près des côtes d'Amérique du Sud, les Iles Falkland, et Austral à la Péninsule Antarctique, le Détrict de Magellan, la Terre de Feu, et l'océan Indien du sud (Goodall et al. 1988).

**Populations:** Des publications ont documenté des observations de groupes de 20 à 100 individus (Goodall et al. 1988, Iñíguez & Tossenberger, 2010). Ils sont catalogués comme une espèce en état global non connu (UICN 2017)

## 1.7 Probabilité conditionnelle et règle de Bayes

On part de l'idée que l'on a des quantités connues, les données  $D$ , et des quantités inconnues qui dans cette thèse correspondront à des paramètres,  $\theta$ . Les deux quantités,  $D$  et  $\theta$  sont liées à partir d'une distribution de probabilité conditionnelle:

$$p(D|\theta) = p(D, \theta)/p(\theta),$$

nous multiplions les deux côté par  $p(D)$ , nous arrivons à  $p(D|\theta) \cdot p(D) = p(D, \theta)$ . Nous faisons la même procédure par  $p(\theta|D)$ , et nous écrivons la égalité

$$p(D|\theta) \cdot p(D) = p(\theta|D) \cdot p(\theta)$$

nous divisons les deux côtés par  $p(D)$  et obtenons

$$p(\theta|D) = \frac{p(\theta) \cdot p(D|\theta)}{p(D)}$$

où  $p(D|\theta)$  est la vraisemblance  $L(\theta|D)$ , et  $p(D) = \int_{\theta} p(D) \cdot L(\theta|D)d\theta$  qui ne dépend pas de  $\theta$ .

Ainsi on peut écrire la fonction de distribution a posteriori comme:

$$\pi(\theta, D) \propto p(\theta) \cdot L(\theta|D) \quad 1.1$$

qui est la règle de Bayes. Cette expression traduite en mots dit que:

Probabilité a posteriori des paramètres  $\propto$  probabilité a priori des paramètres · vraisemblance de  $\theta$  sachant les données  $D$  (Kruschke, 2010).

## 1.8 Modèle hiérarchique de probabilité

Dans l'équation au-dessus, on a le cas particulier d'une seule information à priori à propos des données. S'il y a plus d'information à considérer, on a un ensemble de paramètres de dimension  $r$ ,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_r)$  avec une densité a priori  $p(\boldsymbol{\theta})$  et les observations  $\mathbf{x}$  qui ont une densité  $p(\mathbf{x}|\boldsymbol{\theta})$ . Les paramètres  $\boldsymbol{\theta}$  sont déterminés par d'autres paramètres,  $\eta$  qu'on appelle *hyperparamètres*. Ceux-ci peuvent être connus ou inconnus. Lorsque ces paramètres ne sont pas connus, ils ont eux-mêmes une distribution a priori *hyperpriori*  $p(\eta)$ , et dans ce cas on parle de *prior hiérarchique*.

## 1.9 Des modèles structurés de population

Dans cette thèse, on a fait appel à trois types de modèles structurés de population pour analyser les données de capture-recapture : un modèle à espace d'état pour les dauphins, un modèle à espace d'état pour les loups et un modèle intégré de population dont la composante capture-recapture est un modèle de Markov caché pour les éléphants de mer.

### 1.9.1 Le modèle de Capture-Recapture spatialement explicite

Notre base de données de capture-recapture sur laquelle nous avons appliqué le modèle SECR, correspond à registres de individus de une population de loups italiens qui ont été reconnus par ADN prélevé sur crottes, poils (voir Caniglia et al. 2014 pour plus détaillées).

Un modèle de capture recapture peut être vu comme un modèle hiérarchique de probabilité. Capture-recapture est un terme générique qui fait référence à certains protocoles de surveillance des animaux dans la vie sauvage y compris des pièges et marquage des animaux pour les identifier. Il s'agit en général des données dont l'analyse permet d'estimer la survie, la dispersion et l'abondance avec une probabilité inférieure de détection à un (Gimenez et al. 2012). La méthode de capture-recapture spatialement explicite (SECR en anglais pour "spatially-explicit capture-recapture") considère la position géographique de l'individu échantonné, rendant possible l'estimation du centre des domaines vitaux des animaux, ou centres d'activité. Nous avons utilisé les SECR pour l'étude des loups et de leur structure sociale.

La structure du modèle SECR considère les composantes suivantes:

- un ensemble d'individus indexés par  $i = \{1, 2, \dots, n\}$ ,
- $K$  occasions de capture,
- un ensemble de pièges indexées par  $j = \{1, 2, \dots, J\}$ ,

- un tableau  $\mathbf{S}_{i,latitude,longitude}$  qui contient les positions des individus échantillonnés,
- un tableau  $\mathbf{X}_{i,j,1:2}$  de dimension  $n \times J \times 2$  des positions des pièges, c'est-à-dire, une matrice pour les coordonnées,
- une matrice des distances  $\mathbf{D}_{i,t} = \{d_{i,j}\}$ , où  $d_{i,j}$  est la distance euclidienne entre les pièges et les positions des individus échantillonnés,
- une variable aléatoire latente  $Z_i$  de présence et absence de l'individu dans le domaine d'étude,
- des distributions des paramètres à priori non informatives,
- l'équation du modèle d'observation de l'individu  $i$  par le piège  $j$ ,  $Y_{i,j}$  qui dépend des éléments décrits au-dessus.

Dans notre cas, notre base des données ne provient pas de pièges, nous avons ajouté des pièges fictifs pour pouvoir appliquer le modèle SECR.

### 1.9.2 Modèle à espace d'état

Les modèles à espace d'état (SSM pour "state-space models" en anglais) sont composés de deux séries temporelles qui évoluent en parallèle : d'un côté une série capture la dynamique des vrais états (latents) et de l'autre une série correspond à des observations qui sont faites à partir des états sous-jacents et éventuellement inconnus (Gimenez et al. 2012). Nous avons utilisé un SSM pour les dauphins et l'étude d'une population fermée structurée en dyades.

Nous décrivons l'architecture des SSM à partir de l'article de Buckland et al. (2003).

- Un processus d'état  $\mathbf{n}_t$ , avec  $t = 0, 1, \dots, T$  est un vecteur non observable.
- Le processus d'observation  $\mathbf{y}_t$ , avec  $t = 1, \dots, T$  est un vecteur observable complètement, fonction du processus d'état.
- Soient  $g$  et  $f$  des fonctions de densité (ou masse) de probabilité  $\mathbf{n}_t$  et  $\mathbf{y}_t$  respectivement, la structure d'un SSM peut être décrite en considérant les composantes suivantes :
  - la distribution d'état initial  $g_0(\mathbf{n}_0; \Theta)$ , la distribution du processus d'état  $g_t(\mathbf{n}_t | \mathbf{n}_{t-1}; \Theta)$ , associée à une matrice de transition des états, où l'état actuel est dépendant de l'état précédent (dans notre cas c'est un processus Markovien de premier ordre), la distribution du processus d'observation  $f_t(\mathbf{y}_t | \mathbf{n}_t; \Theta)$ , fonction de l'état actuel, où  $t = 1, \dots, T$  et  $\Theta$  est le vecteur de paramètres.

### 1.9.3 Modèle de Markov caché

Les modèles de Markov cachés (HMM pour "hidden Markov models" en anglais) sont un cas particulier de SSM dans lequel les états sont Markoviens, i.e. l'état suivant (futur) dépend uniquement de l'état immédiatement précédent (présent) (Gimenez et al. 2012). Nous avons utilisé les HMM dans le cas des éléphants de mer pour représenter un cycle de vie relativement complexe.

Selon Rabiner (1989) nous pouvons reconnaître 5 composantes dans les HMM :

- un ensemble  $S = \{S_1, S_2, \dots, S_N\}$ , des  $N$  états (cachés), nous appelons  $q_t$  l'état au temps  $t$ ,
- un ensemble  $V = \{v_1, v_2, \dots, v_M\}$ , des  $M$  différents observables,
- une matrice des distributions de probabilité de transition entre états,  $A = \{a_{ij}\}$ , où  $a_{ij} = P(q_{t+1} = S_j | q_t = S_i)$  est la probabilité d'être dans l'état  $q_{t+1}$  quand nous sommes dans l'état  $q_t$  au temps  $t$ , avec  $1 \leq i, j \leq N$ ,  $\sum_{j=1}^N a_{ij} = 1$  et  $a_{ij} \geq 0$ ,

- un ensemble des distributions de probabilité des observables pour chaque état  $B = \{b_j(v_k)\}$ , où  $b_j(k)$ , est la probabilité d'observation de  $v_k$  quand l'état est q:  $b_j(k) = P(v_{kt}|q_t = S_j)$ ,  $\forall j \in \{1, 2, \dots, N\}, \forall k \in \{1, 2, \dots, M\}$ ,
- la distribution des états initiaux  $\pi = \{\pi_i\}$ , où  $\pi_i$  est la probabilité que le modèle soit dans l'état  $S_i$  au temps initial  $t = 0$ , avec  $\pi_i = P(q_1 = S_i), \forall i \in \{1, 2, \dots, N\}$ .

Ces trois familles de modèles (SECR, SSM et HMM) permettent l'analyse des données de capture-recapture.

Comme nous le verrons plus tard dans le développement des chapitres, il existe des différences entre ces modèles qui permettent une grande flexibilité dans la modélisation des processus biologiques. Pour le cas des loups par exemple, nous sommes intéressés à estimer le nombre de groupes présents dans le domaine d'étude, sans inclure la dimension temporelle, contrairement aux autres modèles utilisés dans lesquels les transitions des états sont considérés. Dans le modèle HMM nous considérons des transitions entre des états (partiellement) cachés, et chaque état futur dépend nécessairement de son état actuel, par exemple, la probabilité qu'une femelle d'éléphant de mer, en état de procréer, reste reproductrice d'une année sur l'autre dépend de son état reproducteur l'année courante et non du fait qu'elle était reproductrice l'année précédente.

## 1.10 Méthodes de Monte Carlo par chaînes de Markov et algorithme de Gibbs

Nous cherchons, à partir des données  $\mathbf{x}$ , à inférer la distribution a posteriori des paramètres du modèle en générant des échantillons simulés et en inférant les paramètres de la population, à partir de la règle de Bayes (voir au-dessus).

JAGS est un programme informatique qui implémente des méthodes de Monte Carlo (pour l'intégration) par chaînes de Markov (MCMC), et l'algorithme de Gibbs en particulier. Nous suivons King et al. (2010) et Lee (2012) pour présenter ces méthodes.

### 1.10.1 L'intégration Monte Carlo

Il s'agit d'une méthode d'intégration numérique pour estimer un paramètre  $\theta$  qui est l'espérance d'une fonction  $\psi$ . Si les observations  $\mathbf{x}$  sont données, on a:

$$\mathbb{E}_\pi[\psi(\theta)] = \int \psi(\theta)\pi(\theta|\mathbf{x})d\theta$$

et nous utilisons la technique de simulation Monte Carlo. Nous calculons l'intégrale en la variable d'intérêt. Nous disons, pour les observations  $\theta^{(1)}, \dots, \theta^{(n)} \sim \pi(\theta|\mathbf{x})$ , que l'espérance de la fonction  $\psi$  évalué en  $\theta$  est défini par l'intégrale

$$\mathbb{E}_\pi[\psi(\theta)] = \int \psi(\theta)\pi(\theta|\mathbf{x})d\mathbf{x}$$

que nous pouvons approcher par

$$\bar{\psi}_n = \frac{1}{n} \sum_{i=1}^n \psi(\theta^{(i)}).$$

En d'autres termes, nous tirons des échantillons  $\theta^1, \dots, \theta^T$  dans la distribution à posteriori, puis nous calculons la moyenne de l'échantillon. Si les échantillons sont indépendants, par la Loi des Grands Nombres,  $\bar{\psi}_n \rightarrow \mathbb{E}_\pi[\psi(\theta)]$  quand  $n \rightarrow \infty$ .

### 1.10.2 Chaîne de Markov

Une chaîne de Markov du premier ordre est une séquence de variables aléatoires  $\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(n)}$ , où la valeur de  $\theta^{(0)}$  vient d'une distribution initiale arbitraire. Simuler une chaîne de Markov consiste à générer un nouvel état de la chaîne  $\theta^{(k+1)}$  avec distribution de densité de noyau de transition  $\mathcal{K}$  pour la chaîne et qui dépend seulement de  $\theta^{(k)}$ :

$$\theta^{(k+1)} \sim \mathcal{K}(\theta^{(k)}, \theta) (\equiv \mathcal{K}(\theta | \theta^{(k)})).$$

En supposant que nous avons une chaîne de Markov avec distribution stationnaire  $\pi(\theta) \equiv \pi(\theta | \mathbf{x})$ , le théorème ergodique des états nous donne que, si les échantillons sont indépendants,  $\bar{\psi}_n \rightarrow \mathbb{E}_\pi[\psi(\theta)]$ , quand  $n \rightarrow \infty$ .

### 1.10.3 L'échantillonneur de Gibbs

L'échantillonneur de Gibbs travaille à partir de l'ensemble complet des distributions conditionnelles de  $\pi$ , pour échantillonner indirectement depuis les distributions marginales. Nous supposons qu'il y a  $r$  paramètres  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p) \in \mathbb{R}^p$  avec distribution de probabilité  $\pi(\boldsymbol{\theta})$ , notons  $\pi(\theta_i | \boldsymbol{\theta}_{(i)})$  la distribution conditionnelle complète induite de  $\theta_i$  étant donné les valeurs des autres composantes  $\boldsymbol{\theta}_{(i)} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_r)$ ,  $i = 1, \dots, r$ ,  $1 < r \leq p$ . Alors, étant donné l'état de la chaîne de Markov à l'itération  $k$ ,  $\boldsymbol{\theta}^k = (\theta_1^{(k)}, \theta_2^{(k)}, \dots, \theta_p^{(k)})$ , l'échantillonneur de Gibbs effectue successivement les tirages aléatoires à partir de la distribution conditionnelle postérieure complète comme suit:

- $\theta_1^{(k+1)}$  est échantillonné depuis  $\pi(\theta_1^{k+1} | \boldsymbol{\theta}_{(1)}^{(k)}, \mathbf{x})$ ;
- $\theta_2^{(k+1)}$  est échantillonné depuis  $\pi(\theta_2 | \theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_r^{(k)}, \mathbf{x})$ ;
- $\vdots$
- $\theta_i^{(k+1)}$  est échantillonné depuis  $\pi(\theta_i | \theta_1^{(k+1)}, \dots, \theta_{i-1}^{(k)}, \dots, \theta_{i+1}^{(k+1)}, \dots, \theta_r^{(k)}, \mathbf{x})$ ;
- $\vdots$
- $\theta_p^{(k+1)}$  est échantillonné depuis  $\pi(\theta_p^{k+1} | \boldsymbol{\theta}_{(r)}^{(k+1)}, \mathbf{x})$ .

Cela complète la transition à partir de  $\boldsymbol{\theta}^{(k)}$  jusque  $\boldsymbol{\theta}^{(k+1)}$ , et la probabilité de cette transition est donnée par:

$$\mathcal{K}_G(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}^{(k+1)}) = \prod_{l=1}^r \pi(\theta_l^{k+1} | \theta_j^{k+1}), \quad j < l \quad \text{et} \quad \theta_j^{(k)}, \quad j > l$$

avec comme distribution stationnaire  $\pi$ .

Plus on d'échantillons, plus les valeurs estimées des paramètres sont proches des vraies valeurs des paramètres.

**Estimer le nombre de meutes dans des populations  
de loups avec des modèles de capture-recapture  
spatialement explicites et des méthodes de partitionnement**

## 2.1 ARTICLE 1. Estimating the number of packs in wolf populations using spatially-explicit capture-recapture models and "clustering" methods

### Résumé

Pour la gestion et la conservation des animaux qui passent la plupart de leur vie en groupes, il est important de pouvoir caractériser la population en termes liés à la socialité comme le nombre et la taille de groupes. Toutefois, cette caractérisation est souvent difficile à faire en conditions naturelles car les animaux ne sont pas observables de manière exhaustive à tout moment – on parle d'une détectabilité imparfaite.

Ici, on propose une approche en deux temps dans laquelle i) on utilise un modèle de capture-recapture spatialement explicite (SECR) pour estimer où sont les centres d'activité de chaque individu en tenant compte de la détectabilité imparfaite, et ii) on applique la méthode de partitionnement de Ward pour identifier le nombre et la taille des groupes. Nous présentons une étude de simulation pour valider la méthode, ainsi qu'un cas d'étude sur une population de loups italiens (*Canis lupus*) comme cas d'étude pour illustrer l'approche.

Notre travail montre comment obtenir de l'information sur la socialité des animaux à partir de données imparfaites en combinant deux méthodes bien connues des écologues.

**Mots-clés:** Espèces sociales, socialité, SECR, partitionnement de Ward.



<sup>1</sup> Estimating the number of packs in wolf populations using

<sup>2</sup> spatially-explicit capture-recapture models and clustering methods

<sup>3</sup> Lorena Mansilla<sup>1</sup>, Roger Pradel<sup>1</sup>, Romolo Caniglia<sup>2</sup>, and Olivier Gimenez<sup>1</sup>

<sup>4</sup> <sup>1</sup>*CEFE, CNRS, University of Montpellier, University Paul Valéry Montpellier 3, EPHE, IRD, Montpellier,  
5 France*

<sup>6</sup> <sup>2</sup>*Department of Biology, Geology and Environmental Sciences, University of Bologna, Via Selmi 3, 40126  
7 Bologna, Italy*

<sup>8</sup> *Abstract.*— For the management and conservation of animals that spend most of their lives in groups,  
<sup>9</sup> it is important to be able to characterize the population in terms of sociality such as the number and  
<sup>10</sup> size of groups. However, this characterization is often difficult to do in natural conditions because the  
<sup>11</sup> animals are not observable exhaustively at any time - we speak of imperfect detectability. Here, a two-step  
<sup>12</sup> approach is proposed in which i) a spatially explicit capture-recapture model (SECR) is used to estimate  
<sup>13</sup> where each individual's activity centers are, taking into account imperfect detectability, and ii) clustering  
<sup>14</sup> methods are used to identify the number and size of groups. We present a simulation study to validate  
<sup>15</sup> the method, as well as a case study on a population of Italian wolves (*Canis lupus*) as a case study to  
<sup>16</sup> illustrate the approach.

<sup>17</sup> Our work shows how to obtain information on animal sociality from imperfect data by combining two  
<sup>18</sup> methods well known to ecologists.

<sup>19</sup> *Key words.*— Social species, sociality, SECR, Ward and K-mean clustering methods.

## <sup>20</sup> **1 Introduction**

<sup>21</sup> Estimating animal population density is a key step in providing sound conservation and management  
<sup>22</sup> strategies for wildlife (Williams et al. 2002). For many large carnivores however, estimating density is  
<sup>23</sup> difficult because these species tend to be elusive and wide-ranging, resulting in low detection rates.

<sup>24</sup> To deal with these issues, non-invasive techniques, such as camera trapping and DNA sampling, are in-  
<sup>25</sup> creasingly used. These non-invasive techniques generate data that can be analyzed with capture-recapture

26 methods to estimate densities (Royle et al. 2014). Spatial capture-recapture (SCR; Royle et al. 2014)  
27 models explicitly incorporate spatial locations of detections, therefore accounting for individual hetero-  
28 geneity on the detection due to spatial variation in the distance of home ranges to the traps. Examples  
29 of the use of SCR methods and non-invasive methods for large carnivores include, without pretending to  
30 exhaustivity, mountain lions (Russell et al. 2012), Grizzly bears (Whittington et al. 2015), black bears  
31 (Sollmann et al. 2012, Sun et al. 2014), Eurasian lynx (Gimenez et al. 2019) and gray wolves (Lopez-Bao  
32 et al. 2018).

33 In addition to the issue of imperfect and heterogeneous detection probability, large carnivores are often  
34 social species structured in groups of associated individuals. While this particular biological organisation  
35 may violate the assumption of independence between individuals, the social structure of such species may  
36 be the object of ecological investigation. For example, in the management of wolves, the number of packs  
37 is often the focus of interest to assess population dynamics at a relevant biological scale (Marucco et al.  
38 2010, Chapron et al. 2016).

39 There has been several attempts to account for species social structure while estimating population  
40 abundance and density. Cubaynes et al. (2010) used mixture capture-recapture models to estimate  
41 wolf abundance while accounting for the dominance status. Martin et al. (2011) formally accounted for  
42 dependent detections when estimating abundance of manatees. Byrne et al. (2014) estimated badger  
43 social-group abundance using cross-validated species distribution modelling. Marnewick et al. (2014)  
44 estimated abundance of wild dogs and cheetahs by considering captures and recaptures of packs instead  
45 of individuals. Belant et al. (2016) estimated abundance of lions using N-mixture models incorporating  
46 group-specific detectability. Hickey and Sollmann (2018) proposed a two-step mark-recapture approach  
47 to estimate abundance of mountain gorillas.

48 To the best of our knowledge, however, these studies have not considered estimating the number of  
49 groups in a social species (e.g. the number of packs in wolf). Here, we aimed at estimating population  
50 abundance and density of species that exhibits a social structure. To do so, we develop a two-step  
51 approach: first, we use SECR models to estimate the activity centers of individuals; second, we use  
52 hierarchical clustering to determine the number of groups. To accommodate and propagate uncertainty  
53 in these two steps, as well as estimate latent variables we adopt a Bayesian approach. We performed  
54 a simulation study to formally validate the performance of our approach. In addition, we illustrate our  
55 approach using capture-recapture data on wolf (*Canis lupus*) in Italy for which we also had information  
56 on the relatedness between individuals, therefore allowing an empirical validation of our approach.

57 **2 Material and methods**

58 **2.1 Statistical approach**

59 The statistical approach we propose to estimate the number of groups in a social species (e.g., the number of  
60 packs in a wolf population) is a two-step process. First, we use spatially-explicit capture-recapture (SECR)  
61 models to estimation the individual activity (or home range) centers. Second, we employ hierarchical  
62 clustering on the individual activity centers to determine the number of groups. Below we go through  
63 each step in details.

64 **2.1.1 Estimation of individual activity centers**

65 We used SECR models to estimate the individual activity centers and their coordinates (Royle et al.  
66 2014). SECR models use the spatial locations of captures to infer the activity center (or home range)  
67 of each individual. In technical terms, we consider an observation process that describes the relationship  
68 between individual  $i$  detection probability at trap  $j$  and the Euclidean distance  $d_{ij}$  between trap  $j$  and the  
69 activity center of individual  $i$ . In details, the distance is  $d_{ij} = \|x_j - s_i\|$  where  $x_j$  is a vector of positions  
70 for trap  $j$  and  $s_i$  is a vector of positions for the activity center of individual  $i$  which is considered as a  
71 latent variable to be estimated. We assumed that the detection probability of an individual  $i$  at trap  $j$   
72 decreased as the distance ( $d_{ij}$ ) from its activity center according to a detection function. We assumed  
73 that  $y_{ij}$  the number of times individual  $i$  is detected at trap  $j$  is Poisson distributed with mean  $\lambda_{ij}$ . We  
74 used the standard half-normal detection function:

$$\lambda_{ij} = \lambda_0 \cdot \exp\left(\frac{-d_{ij}^2}{2\sigma^2}\right). \quad (1)$$

75 where in Eq. (1) parameter  $\lambda_0$  is the expected number of detections at the activity center of individual  
76  $i$  and parameter  $\sigma$  is the spatial scale (or movement) parameter that controls the shape of the detection  
77 function and can be directly linked to home range size (Royle et al. 2014).

78 **2.1.2 Clustering individual activity centers**

79 To estimate the numbers of packs present in the area, we group the individuals using hierarchical clustering,  
80 namely the Ward's method. The Ward's method looks at cluster analysis as an analysis of variance  
81 problem. We start with say  $n$  clusters of size 1. In the first step,  $n - 1$  clusters are formed, one of size  
82 two and the remaining of size 1. The pair of activity centers that yield the smallest error sum of squares  
83 forms the first cluster (we minimize the within-cluster variance). The smaller is the error sum of squares,  
84 the closer the activity centers are to their cluster means, and the more similar these units are within that

particular cluster. Then, in the second step of the algorithm,  $n - 2$  clusters are formed from that  $n - 1$  clusters defined in the previous step, and the error sum of squares is minimized. And so on, at each step of the algorithm, clusters or activity centers are combined in such a way as to minimize the error sum of squares. The algorithm stops when all sample units are combined into a single large cluster of size  $n$ . To determine the group (or pack) membership, we cut the tree at a value obtained from the distribution of all distances between estimated activity centers.

### 2.1.3 Model fitting

Estimating the latent positions of the individual activity centers is challenging. When one is only interested in the population density, i.e. to estimate the number of activity centers, marginalization can be used and the latent positions will be integrated over. Here, our interest was precisely on estimating the latent positions of the individual activity centers, therefore we resorted to a Bayesian approach using Monte Carlo Markov chain (MCMC). We used the data augmentation algorithm (Royle et al. 2014). We augment our observed sample of size  $n$  individuals with  $M - n$  individuals (where  $M > n$ ) that have 0's encounter histories. We define a latent variable  $z_i$  that takes value 1 if individual  $i$  ( $i = 1, \dots, M$ ) belongs to the sampled population and 0 otherwise; we assumed  $z_i \sim \text{Bernoulli}(\psi)$  with  $\psi \sim U(0, 1)$  usually referred to as the inclusion probability. An estimate for population size is then  $\hat{N} = \sum_{i=1}^M I_{\{z_i=1\}}$ .

## 2.2 Wolf case study

To illustrate our approach, we used a case study on gray wolf (*Canis lupus*) in Italy. The data were obtained from identifying the individuals applying non-invasive genetic methods on feces and hair samples (see Caniglia et al. 2014 for more information).

We had data on 94 individuals that were captured and recaptured 251 times between October 1, 2006 and March 31, 2007 in the Emilia Romagna and Tuscany Apennines (between 566615 and 743917 Easting, and 4843299 and 4949916 Northing, in UTM32-WGS84). These data were pooled in six 1-month capture occasions.

In contrast with standard camera-trap studies, the data were collected through a search-effort protocol that does not require fixed traps (Royle et al. 2011, Russell et al. 2012). We built a grid cell layer and assigned all genotyped samples to a cell in which we defined the centroid as the detector location (or a trap). We ended up with 150 cells with a distance between each cell that was determined following recommendation by Solmann et al. (2012). Precisely, we considered a distance between traps at least equal to  $2 \times \sigma$  and checked that the size of a cell was smaller than the home range of a wolf. We found

115 that a distance of 13km was appropriate<sup>1</sup>.

116 Regarding parameter estimation, we augmented the sample with 100 individuals so that  $M = 294$ . We  
117 used the following priors on parameters:  $\sigma^2 \sim U(0, 20)$ ,  $\lambda_0 \sim \Gamma(0.1, 0.1)$ ,  $\psi \sim U(0, 1)$ . We used 2 MCMC  
118 chains with 1,000 iterations as a burnin and 2,000 iterations for final inference. To check for convergence,  
119 we visually inspected the MCMC chains for parameters of interest, and calculated the Gelman-Rubin  
120 statistic.

### 121 2.3 Simulation study

122 To validate our approach, we performed a simulation study by mimicking the characteristics of the wolf  
123 case study.

#### 124 2.3.1 Spatial distribution of individuals

125 Wolves have a home range, in which they realize the activities for their survival. The center of the home  
126 range is what we call the activity center. Here, we will refer to space occupied by one group as "disc", and  
127 to central position of a group as "center of disc", and we will assume that the home ranges are circular. In  
128 Figure 1 we represent the movement of an individual in an area and its activity center  $S_i$ . Two individuals  
129 are also represented that share a space because their movement trajectories are overlapping, therefore these  
130 individuals are considered belonging to same group  $P$  with central position  $S_P$ .

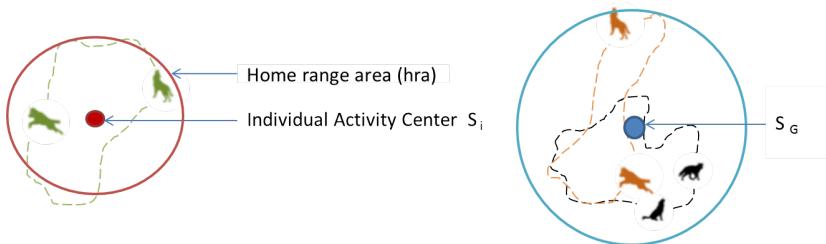


Figure 1: Home range area ( $hra$ ) of an individual  $i$ , individual activity center  $S_i$ , center of disc  $S_P$  used for the pack  $P$ .

131 Now to simulate groups of individuals, we used a Poisson point process. We think of groups of points  
132 that represent groups of individuals, and these groups of points are disposed in circular areas similar to  
133 circular home ranges areas. We make the distinction between these two groups of points: we have the  
134 center points of discs that are called *point patterns* - which are moving in the space - and we have the  
135 points that represent the positions of individuals in the circular areas, *point sons*. Below, we explain how

1We determined  $\sigma$  that allows generating the 95% home range area ( $hra$ ) of an individual as given in Caniglia et al. (2014), where  $\sigma = \sqrt{(hra/\pi)}/\sqrt{\tilde{\chi}^2_{(0.95, 2)}}$

136 we create the disc centers and then the points that represent the individual activity centers.

137

### 138 2.3.2 Central position of groups and movement (point patterns)

139 We generated the positions of the centers of discs and their movements in several steps. First, the number  
140 of groups  $a$  in the space is simulated from a Poisson distribution with mean 19. Second, the centers of  
141 groups  $P$  are spatially disposed with uniform distribution in an area of  $19,171 \text{ km}^2$  (as in the real case  
142 study).  $P$  has dimension  $a \times 2$  corresponding to  $x$  (that represent Easting coordinates) and  $y$  (Northing  
143 coordinates) (see example in Royle et al. 2014, pp. 309 and definitions in Baddeley et al. 2016, pp. 129).  
144 Third, we use these central points as initial points to generate a bidirectional Wiener process  $W_t$ , with  
145  $W_t^1$  and  $W_t^2$  both stationary processes in discrete time with correlation parameters  $\rho_1 = \rho_2 = 0$ , which  
146 is used to capture the movement of groups over time. We provide an example in Figure 2  $a = 11$  groups  
147 and 5 repetitions.

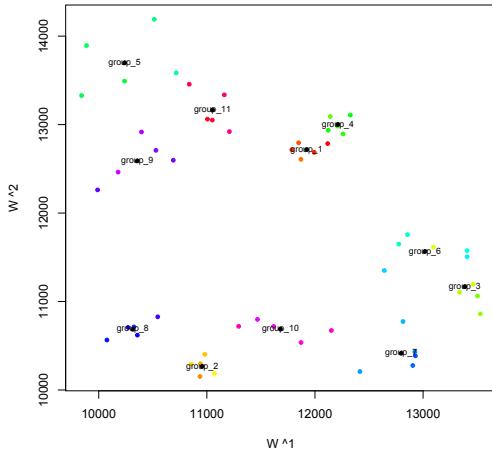


Figure 2: Five positions for  $a = 11$  groups, with their center in black.

### 148 2.3.3 Positions of individuals (points sons)

149 The points that represent the positions of individuals are generated from a truncated multivariate normal  
150 distribution, for which we need to specify the size of groups (or packs), correlation parameter and point  
151 of truncating. The size of groups is drawn from a Poisson distribution with mean 4. We select arbitrarily  
152 the correlation parameter between the axes  $x$  and  $y$  and say  $\rho = 0$ . We use the ratio  $r_{hra}$  as the truncating  
153 parameter. By doing so, the points sons around of a point pattern are less than  $r_{hra}$  farther away from

154 the central point. Regarding the latter, the vector of mean is the matrix of coordinates  $P$  of the centers  
155 of groups over time.

Our vector of coordinates for all individual  $i$ ,  $i = 1, 2, \dots, N$  (where  $N$  is the size of simulated population) is contained in the vector of positions  $XY$  generated from the truncated multivariate normal distribution with mean the vector of means  $W_t$  of coordinates over time, and covariance the matrix  $\Sigma$  such that:

$$\Sigma = \begin{pmatrix} r_x^2 & \rho r_x r_y r_{hra} \\ \rho r_y r_x r_{hra} & r_y^2 \end{pmatrix}$$

156 where  $r_x^2$  is the variance in the axis  $x$ ,  $r_y^2$  is the variance in the axis  $y$  ( $r_y = r_x$  here),  $XY \sim N(P, \Sigma)$  and  
157  $r_{hra}$  is the parameter of truncating. A example of a realization from the truncated multivariate normal  
158 distribution is shown in Figure 3.

159

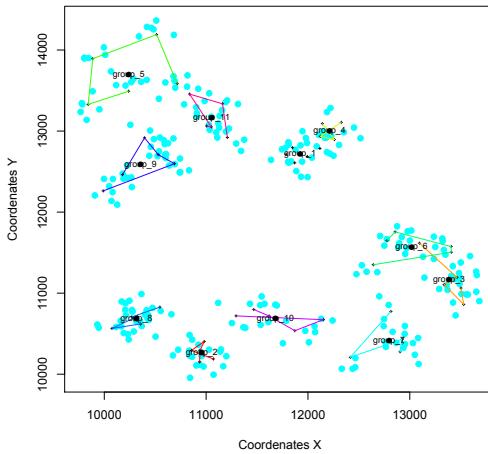


Figure 3: All groups of individuals are represented for color cyan, and generated from a truncated multivariate normal distribution with parameters mean  $P$  and  $\Sigma$  already defined. The positions of centers of discs for 5 times are represent little stars and the displacement of centers disc is showed with lines for each group.

#### 160 2.3.4 Detection process

161 To obtain the individual capture-recapture histories, we apply the Poisson distribution with the detection  
162 function as in the wolf case study.

163 **2.3.5 MCMC details and performance assessment**

164 To simulate the data, we used parameter values  $\sigma^2 = 5$ ,  $\psi = 0.8$  and 100 individuals as the actual  
 165 population size. To estimate the model parameters, we used two MCMC chains of 2,000 iterations with a  
 166 burnin of 1,000 iterations and 100 augmented individuals.

167 To assess the performance of our approach, we generated 100 datasets and estimated the bias and root  
 168 square error mean (RMSE) of the estimate  $\hat{a}$  of the number of groups  $a$ .

169 **3 Results**

170 **3.1 Wolf case study**

171 We ended up with a grid made of 150 cells that was overlapped on the study area; the data are displayed  
 172 in Figure 4 with respect to the traps and the individuals.

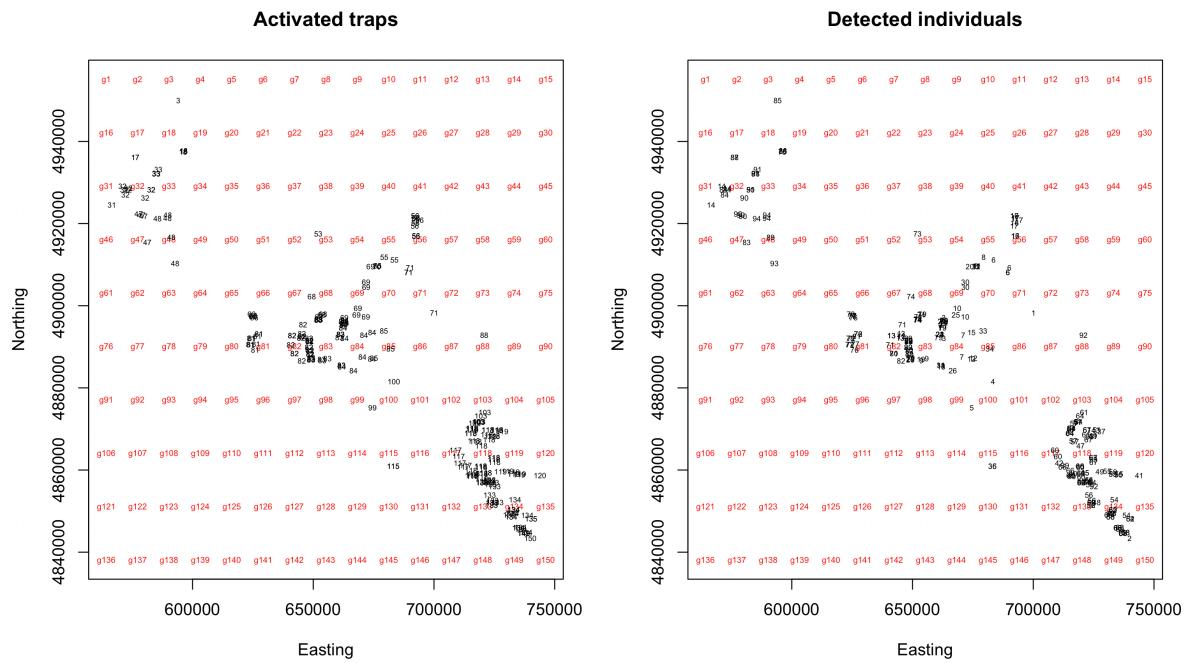


Figure 4: Wolf observations in the study area labeled by name (a number) of activated trap (left) or of detected individual (right).

173 Numerical summaries of parameter estimates are given in Table 1, while posterior distributions are  
 174 displayed in Figure 5. We estimated 130 wolves in total (95% credible interval [116, 130]) for 94 individuals  
 175 actually observed at least once.

Parameter	mean	sd	2.5%	median	97.5%
$\lambda_0$	0.12	0.02	0.09	0.12	0.15
$\sigma^2$	0.37	0.03	0.31	0.37	0.44
N	130.52	7.77	116.00	130.00	147.00

Table 1: Wolf posterior estimates for the spatially-explicit capture-recapture model. We provide the posterior mean and median, standard deviation (sd), lower (2.5%) and upper (97.5%) bounds of the credible interval.

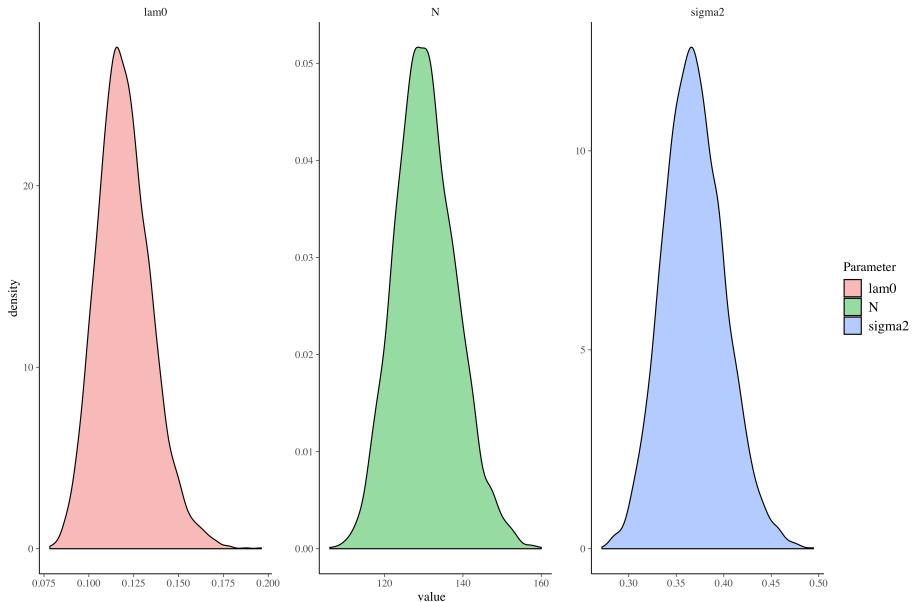


Figure 5: The map of density based on the estimates of the positions of sampled individuals. We have applied the mask that filters the positions and selects only those that belong to the favorable area.

The individual activity centers were grouped with the Ward's method using  $h = 40,000^2$ . We found that the number of packs varied between 8 and 14 with a posterior mean at 11.2 and standard deviation at 1.0 (Figure 6). The number of individuals per pack ranged from 6.7 to 11.8 with a posterior mean of 8.5 and standard deviation of 0.8.

We obtained 4,371 possible dyads total, from these, 50 dyads are present more than 90% of times, forming 12 groups (Figure 7).

### 3.2 Simulations

We obtained 100 simulated datasets, which characteristics are summarized in Table 2.

We obtained 100 simulated datasets, which characteristics are summarized in .

The bias in the number of groups and population size estimators was negligible (see Table 3, and it was of similar magnitude whatever the interval for the number of groups.

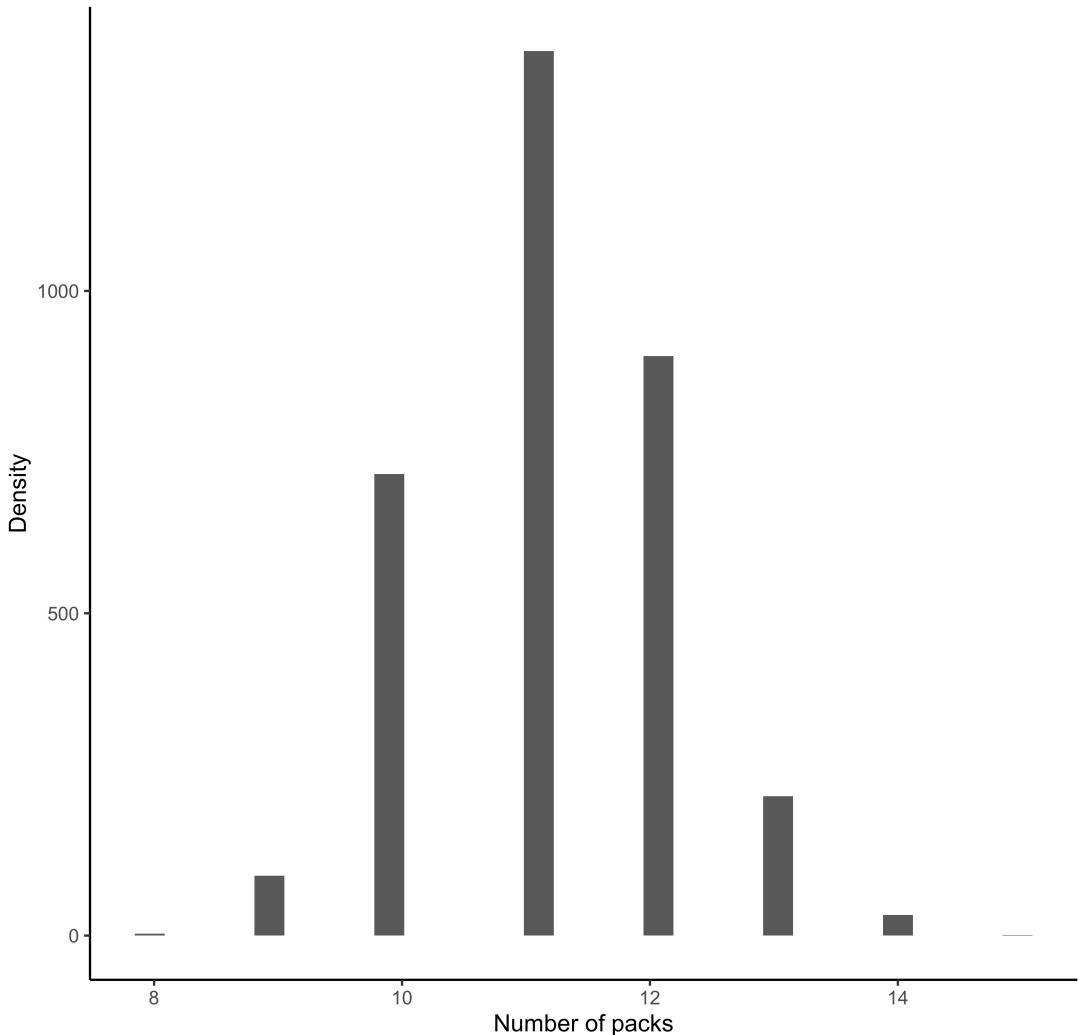


Figure 6: Histogram of the number of wolf packs.

## <sup>187</sup> 4 Discussion

<sup>188</sup> In the ecological literature, the complex structure of social species is at best acknowledged but often  
<sup>189</sup> ignored when estimating abundance and density (Hacke and Sollmann 2018). Several attempts have been  
<sup>190</sup> made to accommodate dependence of individuals when inferring population dynamics. However, to date,  
<sup>191</sup> no method exists to reliably estimate the number of groups in a population of a social species. We proposed  
<sup>192</sup> a two-step approach combining SECR models and hierarchical clustering to estimate the number of packs  
<sup>193</sup> in a wolf population. Our proposal has several appealing advantages. First, ignoring imperfect and  
<sup>194</sup> possibly heterogeneous detection may lead to biased inference about population abundance and density.  
<sup>195</sup> Our approach provides a robust method to estimate the number of clusters in a population. Second, our

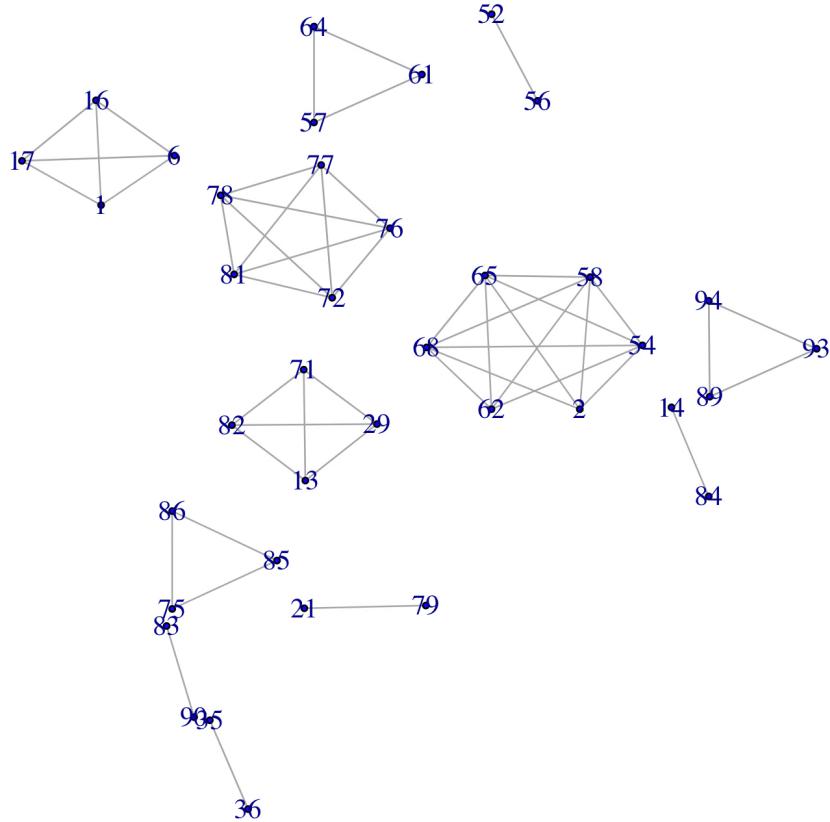


Figure 7: Network of associations between wolves with proportion of associations > 90%.

approach allows the quantification and propagation of uncertainty associated with the unknown latent position of activity centers into the estimation of the number of clusters. If we think of clusters as packs, this is particularly important when the number of packs quantity is converted into population size beyond the study area using some conversion factor (e.g. Chapron et al. 2016). Last, the position of the clusters can be visualised in space. Again, if we think of the position of packs in the wolf case study, our approach provides a tool for assessing potential conflicts like depredation on livestock.

Regarding the wolf case study, we found 11 packs on average. This figure underestimates the 17 packs known to be present in the study area on the period we considered. Despite our efforts, we cannot explain this discrepancy. Reassuringly, the results of the simulation study which characteristics were chosen to mimic the wolf case study suggest little bias in the number of groups estimator.

	Mean	SD of mean	2.5%	50%	97.5%
N Simulated	92.68	33.10	-	-	-
$\tilde{N}$	94.77	20.70	82.31	96.14	100.37
$\tilde{\lambda}_0$	0.016	0.027	0.006	0.015	0.033
$\tilde{\sigma}_2$	2.68	1.94	0.77	2.26	6.40
$\tilde{\psi}$	0.930	0.055	0.785	0.945	0.998
Number of simulated groups	10.230	3.623	-	-	-
Number of estimated groups	13.194	6.885	10.83	13.125	15.691
Average pack size	4.864	2.552	-	-	-
Distance between simulated packs	1764.988	267.925	-	-	-
Distance between estimated packs	1522.778	722.035	-	-	-
Distance between activity centers	759.425	225.355	-	-	-
Distance $h$	1081.736	353.798	-	-	-

Table 2: Main characteristics of the simulation scenario.

	$\forall a$	$a \leq 8$	$8 < a \leq 14$	$15 < a$
Bias ( $\hat{a}, a$ )	1.29	0.29	0.29	0.30
RMSE ( $\hat{a}$ )	6.23	5.88	6.56	5.20
Bias ( $\hat{N}, N$ )	1.02	0.16	0.04	-0.29
RMSE ( $\hat{N}$ )	14.65	16.94	8.84	28.21

Table 3: Performance of the estimation method for various intervals of number of groups. Bias and root mean square error are provided for both the number of groups  $a$  and population size  $N$ .

Our proposal relies on assumptions that need to be discussed. The main assumption we made is that of closed populations. However, birth and death or immigration and emigration events obviously occur in animal populations over long periods of time. The extension of our model to open populations is feasible by using open SECR models (Gardner et al., 2010). This extension would allow accounting for the fate of individuals and, as a by-product, the dynamics of the number of groups or clusters. A challenge will be to track individuals changing groups (e.g., Cubaynes et al. 2010), which would permit studying fission-fusion societies like cetaceans or primates in natural conditions. Another extension would be to handle the dependence in the individual activity centers directly in the model instead of inferring clusters of estimated quantities. A promising avenue would be to adopt the framework proposed by Reich and Gardner (2014) and use an inhomogeneous point process for the individual activity centers instead of a homogeneous Poisson point process. A possible candidate is the Neyman-Scott point process that allows clustering (Baddeley et al. 2016).

## 5 References

BADDELEY A., E. RUBAK E. AND R. TURNER (2016). Spatial Point Patterns, Methodology and applications with R. CRC Press.

- 221 CANIGLIA R., E. FABRI, M. GALAVERNI, P. MILANESI AND E. RANDI (2014). Noninvasive sampling  
222 and genetic variability, pack, structure, and dynamics in an expanding wolf population. *Journal of Mam-*  
223 *mology* 95:41-59.
- 224 CHAPRON, G., C. WIKENROS, O. LIBERG, P. WABAKKEN, Ø. FLAGSTAD, C. MILLERET, J. MÅNSSON,  
225 L. SVENSSON, B. ZIMMERMANN, M. ÅKESSON AND H. SAND 2016. Estimating wolf (*Canis lupus*) pop-  
226 ulation size from number of packs and an individual based model. *Ecological Modelling* 339: 33-44.
- 227 CUBAYNES, S. PRADEL, R. CHOQUET, R. DUCHAMP, C. GAILLARD, J.-M., LEBRETON, J.-D., MAR-  
228 BOUTIN, E., MIQUEL, C., REBOULET, A.-M., POILLOT, C., TABERLET, P. AND O. GIMENEZ (2010).  
229 Importance of accounting for detection heterogeneity when estimating abundance: the case of French  
230 wolves. *Conservation Biology* 24: 621-626.
- 231 GARDNER B., J. REPPUCCI, M. LUCHERINI AND J.A. ROYLE (2010). Spatially explicit inference for  
232 open populations: Estimating demographic parameters from camera-trap studies. *Ecology* 91: 3376-3383.
- 233 GIMENEZ, O., S. GATTI, C. DUCHAMP, E. GERMAIN, A. LAURENT, F. ZIMMERMANN AND E. MAR-  
234 BOUTIN (2019). Spatial density estimates of Eurasian lynx (*Lynx lynx*) in the French Jura and Vosges  
235 Mountains. bioRxiv 600015.
- 236 HICKEY J.R. AND R. SOLLMANN (2018). A new mark-recapture approach for abundance estimation of  
237 social species. *PLoS ONE* 13(12): e0208726.
- 238 LÓPEZ-BAO, J.V., R. GODINHO, C. PACHECO, F.J. LEMA, E. GARCÍA, L. LLANEZA, V. PALACIOS  
239 AND J. JIMÉNEZ (2018) Toward reliable population estimates of wolves by combining spatial capture-  
240 recapture models and non-invasive DNA monitoring. *Scientific Reports* 8(1):2177.
- 241 MARTIN, J., J.A. ROYLE, D.I. MACKENZIE, H.H. EDWARDS, M. KÉRY, AND B. GARDNER (2011).  
242 Accounting for non-independent detection when estimating abundance of organisms with a Bayesian ap-  
243 proach. *Methods in Ecology and Evolution* 2: 595-601.
- 244 MARUCCO, F. AND E.J. MCINTIRE (2010). Predicting spatio-temporal recolonization of large carnivore  
245 populations and livestock depredation risk: wolves in the Italian Alps. *Journal of Applied Ecology* 47:  
246 789-798.
- 247 MATTIOLI, L., A. CANU, D. PASSILONGO, M. SCANDURA, AND M. APOLLONIO 2018. Estimation of  
248 pack density in grey wolf (*Canis lupus*) by applying spatially explicit capture-recapture models to camera  
249 trap data supported by genetic monitoring. *Frontiers in Zoology* 15:38.
- 250 REICH, B.J. AND B. GARDNER (2014). A spatial capture-recapture model for territorial species. *Envi-*  
251 *ronmetrics* 25: 630-637.
- 252 ROYLE, J.A., M. KÉRY AND J. GUÉLAT 2011. Spatial capture-recapture models for search-encounter  
253 data. *Methods in Ecology and Evolution*, 2, 602–611.

- 254 ROYLE, J.A., R. CHANDLER, R. SOLLMANN AND B. GARDNER (2014). Spatial Capture-Recapture,  
255 Elsevier, USA.
- 256 RUSSELL, R.E., J.A. ROYLE, R. DESIMONE, M.K. SCHWARTZ, V.L. EDWARDS, K.P. PILGRIM AND  
257 K.S. MCKELVEY (2012). Estimating Abundance of Mountain Lions From Unstructured Spatial Sampling.  
258 *Journal of Wildlife Management* 76: 1551-1561.
- 259 SARMENTO, P., J. CRUZ, C. EIRA AND C. FONSECA (2014). A spatially explicit approach for estimating  
260 space use and density of common genets. *Animal Biodiversity and Conservation* 37.1: 23–33.
- 261 SOLLMANN, R., B. GARDNER AND J. BELANT (2012). How does spatial study design influence density  
262 estimates from spatial capture-recapture models? *PLoS ONE* 7(4):e34575.
- 263 SUN, C.C., A.K. FULLER AND J.A. ROYLE (2014). Trap Configuration and Spacing Influences Param-  
264 eter Estimates in Spatial Capture-Recapture Models. *PLoS ONE* 9(2): e88025.
- 265 WHITTINGTON, J. AND M.A. SAWAYA (2015). A Comparison of Grizzly Bear Demographic Parameters  
266 Estimated from Non-Spatial and Spatial Open Population Capture-Recapture Models. *PLoS ONE* 10(7):  
267 e0134446.
- 268 WILLIAMS, B.K., J.D. NICHOLS J.D., AND M.J. CONROY (2002). Analysis and management of animal  
269 populations, San Diego: Academic Press.

Modèles intégrés de population pour  
des espèces sociales – Un cas d'étude  
avec les éléphants de mer

### 3.1 ARTICLE 2. Integrated population modeling for social species: a case study with elephants seals

#### Résumé

L'organisation sociale, la structure sociale et le système de reproduction sont des composants clés des systèmes sociaux qui influencent fortement la dynamique de la population. Ignorer le système social dans la modélisation démographique peut conduire à une fausse inférence sur la dynamique de la population et, à son tour, à des actions de gestion erronées. Ici, nous nous appuyons sur les travaux d'autres auteurs et proposons un cadre statistique pour évaluer formellement l'importance de la structure sociale dans une population de phoques de l'éléphant du sud (*Mirounga leonina*) pour en déduire la dynamique de sa population. Nous avons utilisé des modèles intégrés de population (IPM) en combinant des données à long terme de capture-recapture sur la reproduction des femelles et le dénombrement des éléphanteaux sur l'île de Marion, dans le sud de l'océan Indien, au cours de la période 1978-2016.

En partageant des informations sur les paramètres communs entre les deux sources de données, nous avons mis en évidence un effet de socialité sur la dynamique de la population de phoques éléphants, qui ne pouvait pas être démontré lorsque les dénombrements étaient analysés séparément.

Bien que ces modèles soient de plus en plus utilisés en écologie de population, les IPM n'ont été étendus que récemment aux espèces sociales. Ainsi, nous espérons que notre proposition encouragera l'application des IPMs pour une meilleure compréhension de la démographie des espèces sociales.

**Mots-clés:** Détection imparfaite, modèles intégrés de population, capture-recapture, modèles matriciels, espèce sociale.



# INTEGRATED POPULATION MODELING FOR SOCIAL SPECIES: A CASE STUDY WITH ELEPHANT SEALS

3

Lorena Mansilla<sup>1,2</sup>, W. Chris Oosthuizen<sup>3,4</sup>, Roger Pradel<sup>1,5</sup>, P. J. Nico de Bruyn<sup>3,6</sup>, Marthán  
N. Bester<sup>3,7</sup>, and Olivier Gimenez<sup>1,8</sup>

6

*7      <sup>1</sup>Centre d'Ecologie Fonctionnelle et Evolutive (UMR 5175), CNRS – Université de  
8      Montpellier – Université Paul-Valéry Montpellier – EPHE, campus CNRS, Montpellier cedex  
9      5, France*

<sup>10</sup>E-mail: lorena.rociomansilla@gmail.com

11     <sup>3</sup> Mammal Research Institute, Department of Zoology and Entomology, University of  
12     Pretoria, Private Bag X20, Hatfield 0028,  
13     South Africa

14 <sup>4</sup>E-mail: wcoosthuizen@zoology.up.ac.za

<sup>15</sup>E-mail: roger.pradel@cefe.cnrs.fr

16 <sup>6</sup>E-mail: pjndebruyn@zoology.up.ac.za

<sup>17</sup>E-mail: mnbester@zoology.up.ac.za

19     *Abstract.*— Social organization, social structure and mating system are key components of  
20     social systems that strongly influence population dynamics. Ignoring the social system in  
21     demographic modelling can lead to false inference about population dynamics and, in turn, to  
22     erroneous management actions. Here, we build on previous work and propose a statistical  
23     framework to formally assess the importance of social structure in inferring population  
24     dynamics of social species. We used integrated population models (IPMs) that are  
25     increasingly used in population ecology but have only recently been extended to social  
26     species. As a case study, we estimated and tested the effect of the social structure on  
27     population dynamics of southern elephant seals (*Mirounga leonina*) by combining long-term  
28     capture-recapture data on females breeding and pup counts at Marion Island, in the southern  
29     Indian Ocean, during the period 1978-2016. By sharing information on parameters that are in  
30     common between the two sources of data, we demonstrated an effect of sociality on  
31     population dynamics of elephant seals that could not be showed when the counts were  
32     analyzed in isolation. Overall, we hope that our proposal will foster the application of  
33     integrated population models for a better understanding of the demography of social species.

34     *Key words.*— imperfect detection, integrated population models, mark-recapture, matrix  
35     models, *Mirounga leonina*, social species

36

37

## INTRODUCTION

38     Social species occur in functional units like colonies, groups or harems. The social system of  
39     these species is usually described by their social organization, social structure and mating  
40     system (e.g., Kappeler and van Schaik 2002). Several studies showed that these key  
41     components of social systems strongly influence the dynamics of populations (Gerber 2006;  
42     González-Suárez and Gerber 2008; Ferrari et al. 2009; Hickey and Sollmann 2018). Besides,

43 ignoring the social system in demographic modelling can lead to false inference about  
44 population dynamics and, in turn, to erroneous management actions (Vucetich et al. 1997;  
45 Zeigler and Walters 2014; Gerber and White 2014).

46 To infer population dynamics, integrated population models (IPMs) are increasingly  
47 used in population ecology (Schaub and Abadi 2011; Zipkin and Saunders 2018). These  
48 models allow the integration of several types of information, such as capture-recapture, age-  
49 at-harvest, telemetry, and count data into a single statistical analysis (Besbeas et al. 2002).  
50 IPMs combine a matrix population model at their core to capture the species demography  
51 with relevant statistical models for the different sources of data. By sharing information on  
52 parameters that are in common between these models, IPMs yield estimates that are more  
53 precise than when the different sources of data are analyzed separately (Besbeas et al. 2002).

54 Until recently, IPMs have been restricted to monogamous species, which limited their  
55 application to social species. Tenan et al. (2016) extended classical IPMs to nonmonogamous  
56 species by using a two-sex matrix population model (e.g., (Jenouvrier et al. 2010)). The  
57 authors illustrated their approach using a brown bear (*Ursus arctos*) population. Here, we  
58 build on the work of Tenan et al. (2016) and develop a proper statistical framework to  
59 formally assess the importance of social structure in inferring population dynamics of social  
60 species.

61 In this paper, we develop an IPM to extend the population model that Ferrari et al.  
62 (2009) used to quantify the effect of the social structure on population dynamics of southern  
63 elephant seals (*Mirounga leonina*). As a case study we used long-term data on female  
64 southern elephant seals breeding at Marion Island, in the southern Indian Ocean (Bester et al.  
65 2011). Southern elephant seals are the largest living pinnipeds, and among the most sexually  
66 dimorphic and polygynous extant mammals (Le Boeuf and Laws 1994). Females aggregate in

67 harems during the synchronous breeding season with territorial males fighting for dominance  
68 status. Aggregation and territoriality make southern elephant seals a relevant candidate model  
69 for testing IPMs for social species.

70 **METHODS**

71 **Elephant seal biology**

72 Southern elephant seals are wide-ranging marine predators that breed annually on numerous  
73 Southern Ocean islands as well as at Peninsula Valdés in Argentina. Southern elephant seals  
74 are extreme capital breeders and both males and females fast while hauled out at breeding  
75 colonies, relying on catabolism of blubber lipids and body protein for metabolic energy (Boyd  
76 2000). Socially mature males arrive ashore first at the start of breeding seasons (in mid-  
77 August) and establish territories before females arrive (Galimberti et al. 2002). Male  
78 dominance hierarchies are initially formed through antagonistic interactions (fighting), and  
79 the social hierarchy is subsequently maintained through the breeding season by way of  
80 acoustic cues (Casey et al. 2015). The mating system is extreme polygyny (Fabiani et al.  
81 2004; de Bruyn et al. 2011), and males compete for access to females which aggregate in  
82 groups known as harems. Females arrive at breeding colonies in September and October.  
83 Nearly all females ashore during a breeding season give birth to a single pup, and females  
84 remain ashore for the entire lactation period of approximately 23 days. Females are in estrus  
85 during the last few days of lactation, at which time they will mate.

86 At breeding sites such as Marion Island, where harems are relatively small (< 60  
87 females) and spatially segregated, most harems are under the near exclusive control of a  
88 single dominant male, the alpha male (Fabiani et al. 2004). Subordinate males typically linger  
89 at the periphery of harems and will only mate opportunistically with females on the harem-  
90 boundary, or with those females which moved away from harems, if at all (Wilkinson and van

91 Aarde 1999). The large size of elephant seal harems, up to one order of magnitude larger than  
92 harems of other polygynous mammals (Clutton-Brock 1989), and the persistent attempts of  
93 subordinate males to interact with females, produce a complex social network (Galimberti et  
94 al. 2002).

95 **Data collection**

96 *Population counts data*

97 The number of female elephant seals ashore over the course of a breeding season closely  
98 approximates a normal distribution with a peak around 15 October each year (Authier et al.  
99 2011, Fig. S1). Every year from 1973 to 2016, direct counts of elephant seals were made  
100 during the peak of the breeding season at beaches along the north-eastern coastline of Marion  
101 Island.

102 Approximately 25% of the population breed within the area between Ship's Cove and  
103 Archway Bay, and trends observed here are similar to that of the total population. We used 15  
104 October counts of breeding females and breeding males (alpha males and subordinate males)  
105 as an index of abundance and the operational sex ratio (OSR) of the adult population. The  
106 annual mean harem size, defined as the average number of females per harem, was also  
107 estimated using 15 October counts. Here, a harem refers to a group of two or more females,  
108 with or without a male, formed on the six beaches within the study area. Accessibility of  
109 breeding beaches and the relative ease with which breeding elephant seals can be approached  
110 means that direct counts contain minimal observation error. Counts are nonetheless subject to  
111 a degree of spatial and temporal sampling variance, as only a portion of the breeding  
112 population was counted, and because the staggered arrival and departure date of females mean  
113 that even the peak breeding season census can miss females that may have come ashore to  
114 breed (Authier et al. 2011). The total number of pups born over the course of the breeding

115 season was estimated as the sum of the number of breeding females, weaned pups and dead  
116 pups counted on 15 October (e.g., Lewis et al. 2004; Ferrari et al. 2009).

117

118 *Capture-recapture data*

119 From 1983 to 2009, 6,439 recently weaned female elephant seal pups born at Marion Island  
120 were marked with two unique hind-flipper tags (Pistorius et al. 2011). Capture-recapture  
121 resight effort occurred in all years on a systematic 7- or 10-day cycle (Oosthuizen et al.  
122 2019b) on all beaches on Marion Island where female elephant seals hauled out for breeding,  
123 moulting and resting (Mulaudzi et al. 2008). A total of 65,602 resightings of marked female  
124 elephant seals were made between 1983 and 2014.

125

126 **Integrated population model**

127 In this section, we describe the two components of the IPM we developed based on the  
128 Southern elephant seal life cycle (Figure 1). Our IPM is built as the product of the likelihood  
129 for the population counts model and the likelihood for the capture-recapture model (Figure 2).  
130 We describe these two components below.

131

132 [FIGURE 1 AROUND HERE]

133

134 *Population dynamics model*

135 We follow (Ferrari et al. 2009) and define  $N_t$  the number of pups in breeding season  $t$  as a  
136 function of fertility  $F$  and  $N_f$  the number of adult females. We have:

$$N(t) = \alpha F(t - 1) N_f(t) \quad (1)$$

137 where  $\alpha$  is a fertility constant and the function  $F$  is defined over the breeding season  $t - 1$   
 138 (whenever they were pregnant) as:

$$F(t) = (1 + R(t)^a)^{1/a} \quad (2)$$

139 where  $R$  represents the social structure and is the mean harem size times the number of adult  
 140 males per adult female. Parameter  $a$  determines the strength of this variable on the function  
 141 value (Ferrari et al. 2009). More precisely, whenever  $a \rightarrow +/\infty$ , we have  $F(t) = 1$  and there  
 142 is no influence of the social structure whereas whenever  $a \rightarrow 0$ , we have  $F(t) = \infty$  and there  
 143 is a strong influence of social structure.

144 Assuming breeding starts not before two years,, and that most females reproduce at 3 and 4  
 145 years old, we also have:

$$N_f(t) = \phi N_f(t - 1) + \pi \rho r N(t - 3) + (1 - \pi) \rho r \phi_{a3} N(t - 4) \quad (3)$$

146 where  $\phi$  is the adult female survival,  $\pi$  is the proportion of females breeding at age 3 (and  
 147  $1 - \pi$  the proportion of females breeding at age 4),  $\rho$  is a constant corresponding to the mean  
 148 sex ratio (proportion of females) at birth recorded from 1983 to 2009 (0.506),  $r$  the  
 149 recruitment rate of adult females and  $\phi_{a3}$  the female survival from age 3 to 4. The first term  
 150 in equation (3) represents surviving breeders of  $t - 1$ , the second term is a proportion  $\pi$  of 3-  
 151 year old females and third term is a proportion  $1 - \pi$  of 4-year females that did not breed at  
 152 age 3.

153 Replacing  $N_f(t)$  of equation (3) in equation (1) and writing  $N_f(t - 1) = N(t - 1)/\alpha F(t -$   
 154 2), we get:

$$N(t) = \phi \frac{F(t)}{F(t-1)} N(t-1) + qF(t)[(1-\pi)\phi_{a3}N(t-4) + \pi N(t-3)] \quad (4)$$

155 where  $q = \alpha\pi r$ . We assume a multiplicative sampling error  $V_t$ , so that observed pup counts  $y_t$   
 156 in breeding season  $t$  is:

$$y_t = N_t V_t \quad (5)$$

157 With  $V_t \sim Lognormal(0, \sigma^2)$ , we get the likelihood for the pup counts  $L_{counts}$  as:

$$-2\ln L_{counts} = \exp(-0.5\ln(2\pi) - \ln(\sigma) - \frac{1}{2\sigma^2} \sum (\ln y_t - \ln N_t)^2) \quad (6)$$

158

159 *Capture-recapture model*

160 We built a multievent capture-recapture model (Pradel 2005) to estimate demographic  
 161 parameters with likelihood  $L_{capture-recapture}$ . Female elephant seals were assumed to occupy  
 162 one of the following states each year: pre-breeder  $PB$  (has not previously pupped), breeder  $B$   
 163 (pupped in the current year); non-breeder  $NB$  (pupped previously, but not in the current year);  
 164 and  $D$  dead. All individuals entered the marked population as weaned pups (thus in the pre-  
 165 breeder state). Encounter histories summarised multiple sightings of an individual between  
 166 two consecutive breeding seasons (September in year  $t$  to August in year  $t+1$ ) as a single  
 167 encounter: 0 for non-observed, 1 for non-breeding and 2 for breeding. Pre-breeding and non-  
 168 breeding elephant seals typically do not attend breeding colonies in the breeding season but  
 169 are observed at other times of the year (e.g., during the moult and winter haulout) (Oosthuizen  
 170 et al. 2019b).

171           The transition probabilities between states correspond to apparent annual survival  
172           (hereafter survival) and breeding probability. We assumed age-specific pre-breeder survival  
173           (Pistorius and Bester 2002) and constant adult survival. We denoted  $\phi_{a0}$  as the survival  
174           probability of newborn females,  $\phi_{a1}, \phi_{a2}, \phi_{a3}, \phi_{a4}$  the survival probability of 1-year old, 2-  
175           year old, 3-year old and 4-year old females and  $\phi$  the survival probability of older adult  
176           females.

177           In elephant seals, there is no recruitment to the breeding population on the beachers  
178           prior to age 3 and in our dataset, most surviving females recruited at age 3 or age 4 years  
179           (Oosthuizen et al. In review). The transition probability from pre-breeder to breeder was  
180           therefore assumed to depend on age, with  $\psi_{B2}, \psi_{B3}$  and  $\psi_{B4}$  the probability of breeding in the  
181           next year for 2-year old, 3-year old and 4-year old females,  $\psi_B$  the probability of a  
182           reproductive female to remain a reproductive female.

183           Because attendance behaviour differs between breeding states, we allowed capture  
184           probabilities to vary with state, with  $p_{PB}$  the probability of observing a pre-breeding female  
185           outside the breeding season,  $p_B$  the probability of observing a breeding female in the breeding  
186           season, and  $p_{NB}$  the probability of observing a non-breeding female outside of the breeding  
187           season.

188

189           [FIGURE 2 AROUND HERE]

190

## 191           **Parameter estimation**

192           We used maximum likelihood theory to estimate the parameters of the IPM model and to  
193           derive confidence intervals. To maximize the IPM likelihood, we used the quasi-Newton

194 algorithm as implemented in the optim function from program R (R Code Team 2019). We  
195 expected a better precision for parameters that were in common between the two likelihoods  
196  $L_{counts}$  and  $L_{capture-recapture}$ , namely  $\phi$  and  $\phi_{a3}$ . We also obtained confidence intervals for  
197 the number of pups observed using the parametric bootstrap method (Davison and Hinkley  
198 1997).

199

200 **Test of the influence of sociality**

201 To formally test the effect of sociality on population dynamics, we considered the null  
202 hypothesis of no influence of social structure where  $a$  reached infinity ( $a = 1,000$ ) in the IPM  
203 likelihood and tested it with a likelihood ratio test (LRT) against the alternative hypothesis in  
204 which  $a$  was estimated (representing nonlinear fertility dependence on social structure). We  
205 also considered this comparison using models with pup counts only to determine whether  
206 combining sources of data in an IPM increased our ability to detect the influence of social  
207 structure.

208

209 **RESULTS**

210 When comparing models in which parameter  $a$  was fixed vs. estimated with count data only,  
211 the null hypothesis of no effect of sociality on demography could not be rejected (Table 1). In  
212 contrast, when count and capture-recapture data were combined in the IPM, the LRT was  
213 significant ( $p = 0.05$ ). The effect of sociality was confirmed by the estimate of parameter  $a$   
214 which was far from infinity in both the IPM (-0.99 [-1.18; -0.83]) and the model for count  
215 data only (-2.08 [-5.47; -0.88]).

216

217

[TABLE 1 AROUND HERE]

218

219       The survival of newborn females was low (0.60 [0.59; 0.61]) while that of pre-  
220 breeders and adults varied between 0.71 ([0.68; 0.74]) for 4-year old females and 0.79 ([0.77;  
221 0.81]) for 3-year old females (Table 1). The probability of breeding for the first time varied  
222 substantially with age with a peak for 4-year old females. Regarding the detection process, the  
223 probability of observing a breeding female in the breeding season was higher than that of  
224 observing a pre-breeder or non-breeder female outside the breeding season (Table 1). The  
225 demographic parameter estimates were indistinguishable whether we let parameter  $a$  be  
226 estimated or we fixed it.

227       When inspecting the predictions of number of pups born annually from the IPM, the  
228 fit of the model to the observed counts was satisfactory from 1998 onwards, while it captured  
229 poorly the observed decreasing trend from 1981 to 1995 (Figure 3).

230

231

[FIGURE 3 AROUND HERE]

232

233

234

## DISCUSSION

235       The inclusion of social structure (e.g., operational sex ratios) in population models can  
236 produce different dynamics to female-only models that excludes behavioral data, even in  
237 polygynous species where female numbers are thought to limit reproduction (Gerber 2006;  
238 Gerber and White 2014). We illustrated the relevance of integrated population models (IPMs)  
239 for modeling the demography of polygynous species and provide a statistical framework to  
240 formally assess the effect of sociality on population dynamics in such species.

241 Identifying whether behavioral attributes influence demographic parameters can  
242 enhance our understanding of population dynamics. A shortage of breeding males, resulting in  
243 low insemination rates, is one mechanism previously suggested to contribute to the southern  
244 elephant seal population decline at Marion Island (Skinner and van Aarde 1983). This  
245 hypothesis was later refuted as dominant bulls are capable of mating multiple times with all  
246 females in harems (Wilkinson and van Aarde 1999). However, the role of males in the  
247 dynamics of populations is not limited to active mechanisms, such as sexual activity, by  
248 which they may influence the breeding rates of females and so population dynamics  
249 (Mysterud et al. 2002). Males also have a more indirect influence on population dynamics and  
250 female demographic parameters simply by being a component of population density  
251 (Mysterud et al. 2002), through sexual harassment of females, or by modifying female  
252 behavior (e.g., Milner-Gulland et al. 2003; Galliard et al. 2005). Single-sex population matrix  
253 models incorporating sex-ratios may therefore account partly for male influence on  
254 population dynamics, although they do not incorporate sex-specific vital rates. Two-sex  
255 matrix population models are becoming more common in demographic studies of species with  
256 unbalanced sex ratios or unique social structures (e.g., Tongen et al. 2016; Shyu and Caswell  
257 2018) and their use in the IPM framework is encouraged (see Tenan et al. 2016 for an  
258 illustration of two-sex IPMs).

259 In IPMs, the single model likelihoods of several datasets (population count data and  
260 demographic data in this study) are combined to create a joint likelihood model upon which  
261 inference is based (Schaub and Abadi 2011). Typically, this approach enhances statistical  
262 power to increase parameter estimate precision. We expected a better precision for parameters  
263 that were in common between the two likelihoods  $L_{counts}$  and  $L_{capture-recapture}$ , namely  $\phi$   
264 and  $\phi_{a3}$ . However, demographic parameters of the combined capture-recapture and count  
265 data IPM did not have better precision in this study, probably because the capture-recapture

266 data are informative enough (recapture probabilities are high) so that these parameters are  
267 already very well estimated. The IPM and count model differed, however, in their estimates  
268 of parameter  $a$ , which determines the strength of social structure on the fertility function. The  
269 IPM estimate of  $a$  was -0.99, which approximates the harmonic mean fertility function  
270 (Gerber 2006). The effect of social structure was weaker in the count model.

271 Ferrari et al. (2009) previously used a deterministic population model, similar to our  
272 count model, to investigate the relationship between social structure and population dynamics  
273 of southern elephant seals breeding at Peninsula Valdés in Argentina. Their results suggested  
274 a better fit of predicted values to observed data when social structure was included in the  
275 fertility function of the model. One advantage of formulating this model as an IPM is that we  
276 could independently estimate parameters  $\rho$ ,  $r$  and  $\alpha$  (Equations 2 and 3) which was part of a  
277 composite recruitment parameter in Ferrari et al. (2009). Also, by combining with capture-  
278 recapture data, we could relax the assumption of known demographic parameters (survival  
279 and transitions between reproductive states) that was made in Ferrari et al. (2009) and  
280 estimate them directly. Overall, our estimates of survival and breeding probabilities were  
281 comparable to previous estimates obtained for this population (e.g., Pistorius et al. 1999a,  
282 2011; Pistorius and Bester 2002, Oosthuizen et al. 2018).

283 Although testing it directly would be difficult, the most plausible hypothesis  
284 explaining the decline in pup counts we observed and predicted (Figure 2) is a reduction in  
285 food resources and/or quality through environmental change (McMahon et al. 2005). An  
286 alternative hypothesis based on historical fish extraction (Ainley and Blight 2009), again  
287 suggests that alteration in prey availability was key to population declines. Investigating the  
288 effect of environmental stochasticity in southern elephant seals breeding at Marion Island is  
289 the object of ongoing research. Life table response experiment analyses (e.g., Koons et al.  
290 2016) and their extension to IPMs (Koons et al. 2017) will aid our ecological understanding

291 of the observed changes in the population trajectory by revealing the contribution of different  
292 demographic parameters to variation in realized population growth rates.

293 Overall, we hope that our proposal will foster the application of integrated population  
294 models to polygynous species for a better understanding of the demography of social species.

295

296 ACKNOWLEDGMENTS

297 This research is part of a doctoral thesis of the first author, funded by CONICYT,  
298 PFCHA/DOCTORADO BECAS CHILE/2016-72170563. We thank Marion Island Marine  
299 Mammal Programme field personnel and the South African Department of Environmental  
300 Affairs for field support. The Department of Science and Technology of South Africa  
301 provided funding for elephant seal research at Marion Island through the National Research  
302 Foundation (NRF). Opinions expressed and conclusions arrived at, are those of the authors  
303 and are not necessarily to be attributed to the NRF.

304

305 LITERATURE CITED

306 Ainley DG, Blight LK (2009) Ecological repercussions of historical fish extraction from the  
307 Southern Ocean. *Fish and Fisheries* 10:13–38. doi: 10.1111/j.1467-2979.2008.00293.x

308 Authier M, Delord K, Guinet C (2011) Population trends of female Elephant Seals breeding  
309 on the Courbet Peninsula, îles Kerguelen. *Polar Biology* 34:319–328. doi: 10.1007/s00300-  
310 010-0881-1

311 Besbeas P, Freeman SN, Morgan BJT, Catchpole EA (2002) Integrating mark-recapture-  
312 recovery and census data to estimate animal abundance and demographic parameters.  
313 *Biometrics* 58:540–547

- 314 Bester M, de Bruyn P, Oosthuizen W, et al (2011) The Marine Mammal Programme at the  
315 Prince Edward Islands: 38 years of research. *African Journal of Marine Science* 33:511–521.  
316 doi: 10.2989/1814232X.2011.637356
- 317 Boyd IL (2000) State-dependent fertility in pinnipeds: contrasting capital and income  
318 breeders. *Functional Ecology* 14:623–630. doi: 10.1046/j.1365-2435.2000.t01-1-00463.x
- 319 Casey C, Charrier I, Mathevon N, Reichmuth C (2015) Rival assessment among northern  
320 elephant seals: evidence of associative learning during male–male contests. *Royal Society*  
321 *Open Science* 2:150228. doi: 10.1098/rsos.150228
- 322 Clutton-Brock TH (1989) Review Lecture: Mammalian mating systems. *Proceedings of the*  
323 *Royal Society of London B Biological Sciences* 236:339–372. doi: 10.1098/rspb.1989.0027
- 324 Davison AC, Hinkley DV (1997) *Bootstrap Methods and Their Application*. Cambridge  
325 University Press
- 326 de Bruyn PJN, Tosh CA, Bester MN, et al (2011) Sex at sea: alternative mating system in an  
327 extremely polygynous mammal. *Animal Behaviour* 82:445–451. doi:  
328 10.1016/j.anbehav.2011.06.006
- 329 Fabiani A, Galimberti F, Sanvito S, Hoelzel AR (2004) Extreme polygyny among southern  
330 elephant seals on Sea Lion Island, Falkland Islands. *Behav Ecol* 15:961–969. doi:  
331 10.1093/beheco/arh112
- 332 Ferrari MA, Lewis MN, Pascual MA, Campagna C (2009) Interdependence of social structure  
333 and demography in the southern elephant seal colony of Península Valdés, Argentina. *Marine*  
334 *Mammal Science* 25:681–692. doi: 10.1111/j.1748-7692.2008.00268.x
- 335 Galimberti F, Fabiani A, Sanvito S (2002) Measures of breeding inequality: a case study in

- 336 southern elephant seals. *Can J Zool* 80:1240–1249. doi: 10.1139/z02-117
- 337 Galliard J-FL, Fitze PS, Ferrière R, Clobert J (2005) Sex ratio bias, male aggression, and  
338 population collapse in lizards. *PNAS* 102:18231–18236. doi: 10.1073/pnas.0505172102
- 339 Gerber LR (2006) Including behavioral data in demographic models improves estimates of  
340 population viability. *Frontiers in Ecology and the Environment* 4:419–427. doi:  
341 10.1890/1540-9295(2006)4[419:IBDIDM]2.0.CO;2
- 342 Gerber LR, White ER (2014) Two-sex matrix models in assessing population viability: when  
343 do male dynamics matter? *Journal of Applied Ecology* 51:270–278. doi: 10.1111/1365-  
344 2664.12177
- 345 González-Suárez M, Gerber LR (2008) A Behaviorally Explicit Demographic Model  
346 Integrating Habitat Selection and Population Dynamics in California Sea Lions. *Conservation  
347 Biology* 22:1608–1618. doi: 10.1111/j.1523-1739.2008.00995.x
- 348 Hickey JR, Sollmann R (2018) A new mark-recapture approach for abundance estimation of  
349 social species. *PLOS ONE* 13:e0208726. doi: 10.1371/journal.pone.0208726
- 350 Jenouvrier S, Caswell H, Barbraud C, Weimerskirch H (2010) Mating Behavior, Population  
351 Growth, and the Operational Sex Ratio: A Periodic Two-Sex Model Approach. *The American  
352 Naturalist* 175:739–752. doi: 10.1086/652436
- 353 Kappeler PM, van Schaik CP (2002) Evolution of Primate Social Systems. *International  
354 Journal of Primatology* 23:707–740. doi: 10.1023/A:1015520830318
- 355 Koons DN, Arnold TW, Schaub M (2017) Understanding the demographic drivers of realized  
356 population growth rates. *Ecological Applications* 27:2102–2115. doi: 10.1002/eap.1594
- 357 Koons DN, Iles DT, Schaub M, Caswell H (2016) A life-history perspective on the

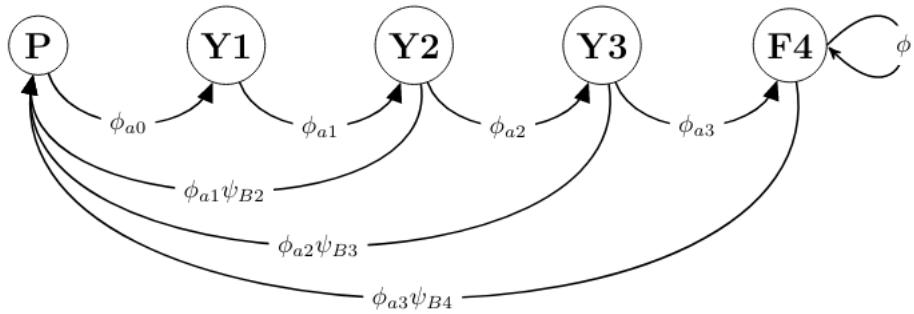
- 358 demographic drivers of structured population dynamics in changing environments. *Ecology*  
359 Letters 19:1023–1031. doi: 10.1111/ele.12628
- 360 Le Boeuf BJ, Laws RM (eds) (1994) Elephant seals: population ecology, behavior, and  
361 physiology. University of California Press, Berkeley
- 362 Lewis M, Campagna C, Zavatti J (2004) Annual cycle and inter-annual variation in the haul-  
363 out pattern of an increasing southern elephant seal colony. *Antarctic Science* 16:219–226. doi:  
364 10.1017/S0954102004002020
- 365 McMahon CR, Bester MN, Hindell MA, et al (2009) Shifting trends: detecting  
366 environmentally mediated regulation in long-lived marine vertebrates using time-series data.  
367 *Oecologia* 159:69–82. doi: 10.1007/s00442-008-1205-9
- 368 Milner-Gulland EJ, Bukreeva OM, Coulson T, et al (2003) Reproductive collapse in saiga  
369 antelope harems. *Nature* 422:135. doi: 10.1038/422135a
- 370 Mulaudzi TW, Hofmeyr GJG, Bester MN, et al (2008) Haulout site selection by southern  
371 elephant seals at Marion Island. *African Zoology* 43:25–33. doi:  
372 10.1080/15627020.2008.11407403
- 373 Mysterud A, Coulson T, Stenseth NC (2002) The role of males in the dynamics of ungulate  
374 populations. *Journal of Animal Ecology* 71:907–915. doi: 10.1046/j.1365-2656.2002.00655.x
- 375 Oosthuizen WC, Postma M, Altwegg R, et al (2019a) Individual heterogeneity in life-history  
376 trade-offs with age at first reproduction in capital breeding elephant seals. In review in  
377 *Population Ecology*.
- 378 Oosthuizen WC, Pradel R, Bester MN, de Bruyn PJN (2019b) Making use of multiple  
379 surveys: Estimating breeding probability using a multievent-robust design capture–recapture

- 380 model. *Ecology and Evolution* 9:836–848. doi: 10.1002/ece3.4828
- 381 Pistorius PA, Bester MN (2002) Juvenile survival and population regulation in southern  
382 elephant seals at Marion Island. *African Zoology* 37:35–41. doi:  
383 10.1080/15627020.2002.11657152
- 384 Pistorius PA, Bester MN, Kirkman SP (1999) Survivorship of a declining population of  
385 southern elephant seals, *Mirounga leonina*, in relation to age, sex and cohort. *Oecologia*  
386 121:201–211. doi: 10.1007/s004420050922
- 387 Pistorius PA, Bruyn P de, Bester MN (2011) Population dynamics of southern elephant seals:  
388 a synthesis of three decades of demographic research at Marion Island. *African Journal of  
389 Marine Science* 33:523–534. doi: 10.2989/1814232X.2011.637357
- 390 Pradel R (2005) Multievent: an extension of multistate capture-recapture models to uncertain  
391 states. *Biometrics* 61:442–7. doi: 10.1111/j.1541-0420.2005.00318.x
- 392 Schaub M, Abadi F (2011) Integrated population models: a novel analysis framework for  
393 deeper insights into population dynamics. *J Ornithol* 152:227–237. doi: 10.1007/s10336-010-  
394 0632-7
- 395 Shyu E, Caswell H (2018) Mating, births, and transitions: a flexible two-sex matrix model for  
396 evolutionary demography. *Popul Ecol* 60:21–36. doi: 10.1007/s10144-018-0615-8
- 397 Skinner JD, van Aarde RJ (1983) Observations on the Trend of the Breeding Population of  
398 Southern Elephant Seals, *Mirounga leonina*, at Marion Island. *Journal of Applied Ecology*  
399 20:707–712. doi: 10.2307/2403121
- 400 Tenan S, Iemma A, Bragalanti N, et al (2016) Evaluating mortality rates with a novel  
401 integrated framework for nonmonogamous species: Mortality in Nonmonogamous Species.

- 402 Conservation Biology 30:1307–1319. doi: 10.1111/cobi.12736
- 403 Tongen A, Zubillaga M, Rabinovich J (2016) A two-sex matrix population model to represent  
404 harem structure. Mathematical Biosciences and Engineering 13:1077–1092. doi:  
405 10.3934/mbe.2016031
- 406 Vucetich JA, Peterson RO, Waite TA (1997) Effects of Social Structure and Prey Dynamics  
407 on Extinction Risk in Gray Wolves. Conservation Biology 11:957–965. doi: 10.1046/j.1523-  
408 1739.1997.95366.x
- 409 Wilkinson IS, van Aarde RJ (1999) Marion Island elephant seals: the paucity-of-males  
410 hypothesis tested. Can J Zool 77:1547–1554. doi: 10.1139/z99-127
- 411 Zeigler SL, Walters JR (2014) Population models for social species: lessons learned from  
412 models of Red-cockaded Woodpeckers (*Picoides borealis*). Ecological Applications 24:2144–  
413 2154. doi: 10.1890/13-1275.1
- 414 Zipkin EF, Saunders SP (2018) Synthesizing multiple data types for biological conservation  
415 using integrated population models. Biological Conservation 217:240–250. doi:  
416 10.1016/j.biocon.2017.10.017
- 417

418 TABLE 1: Model comparison and parameter estimates for the southern elephant seal analysis.  
 419 We provide estimates with 95% confidence intervals for the parameters of the integrated  
 420 population model (IPM) and a model with count data only in which parameter  $a$  is either fixed  
 421 to infinity ( $a = 1,000$ ) or estimated. The p-value of a likelihood-ratio test (LRT) for the null  
 422 hypothesis that  $a$  is infinite is given.

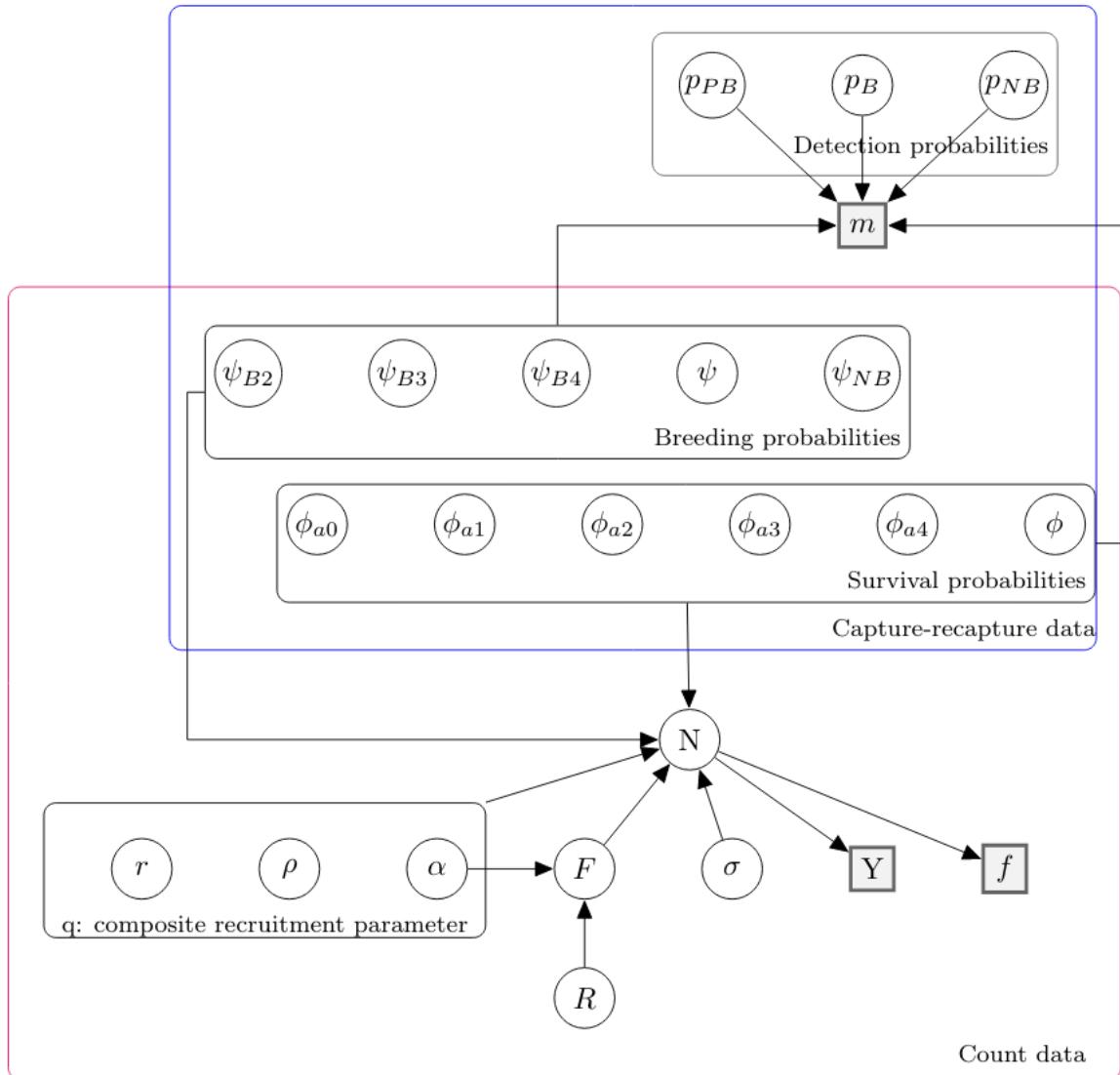
	<b>IPM</b>	<b>IPM <math>a</math> fixed</b>	<b>Count model</b>	<b>Count model <math>a</math> fixed</b>
deviance	43166.99	43170.68	1.96	2.04
LRT p-value	0.05		0.79	
$a$	-0.99 (-1.18; -0.83)	-	-2.08 (-5.47; -0.88)	-
$\sigma$	1.63 (0.61; 4.35)	10.29 (3.85; 28.50)	1.62 (0.61; 4.31)	1.61 (0.63; 4.48)
$\phi_{a0}$	0.60 (0.59; 0.61)	0.60 (0.59; 0.61)	-	-
$\phi_{a1}$	0.76 (0.75; 0.78)	0.76 (0.75; 0.78)	-	-
$\phi_{a2}$	0.78 (0.77; 0.79)	0.78 (0.77; 0.79)	-	-
$\phi_{a3}$	0.79 (0.77; 0.81)	0.79 (0.77; 0.81)	-	-
$\phi_{a4}$	0.71 (0.68; 0.74)	0.71 (0.68; 0.74)	-	-
$\phi$	0.76 (0.75; 0.77)	0.76 (0.75; 0.77)	-	-
$\psi_{B2}$	0.31 (0.29; 0.32)	0.31 (0.29; 0.32)	-	-
$\psi_{B3}$	0.68 (0.65; 0.70)	0.67 (0.65; 0.70)	-	-
$\psi_{B4}$	0.56 (0.52; 0.61)	0.56 (0.52; 0.61)	-	-
$\psi_B$	0.82 (0.80; 0.83)	0.82 (0.80; 0.83)	-	-
$\psi_{NB}$	0.73 (0.68; 0.76)	0.73 (0.68; 0.76)	-	-
$p_{PB}$	0.78 (0.77; 0.78)	0.78 (0.77; 0.78)	-	-
$p_B$	0.91 (0.89; 0.93)	0.91 (0.89; 0.93)	-	-
$p_{NB}$	0.87 (0.77; 0.93)	0.87 (0.77; 0.93)	-	-



423

424 FIGURE 1: Southern elephant seal life cycle diagram. The states of development of females are  
 425 represented by P: pups, Y1: 1-year old females, Y2: 2-year old females, Y3: 3-year old  
 426 females and F4: females  $\geq$  4-years old.

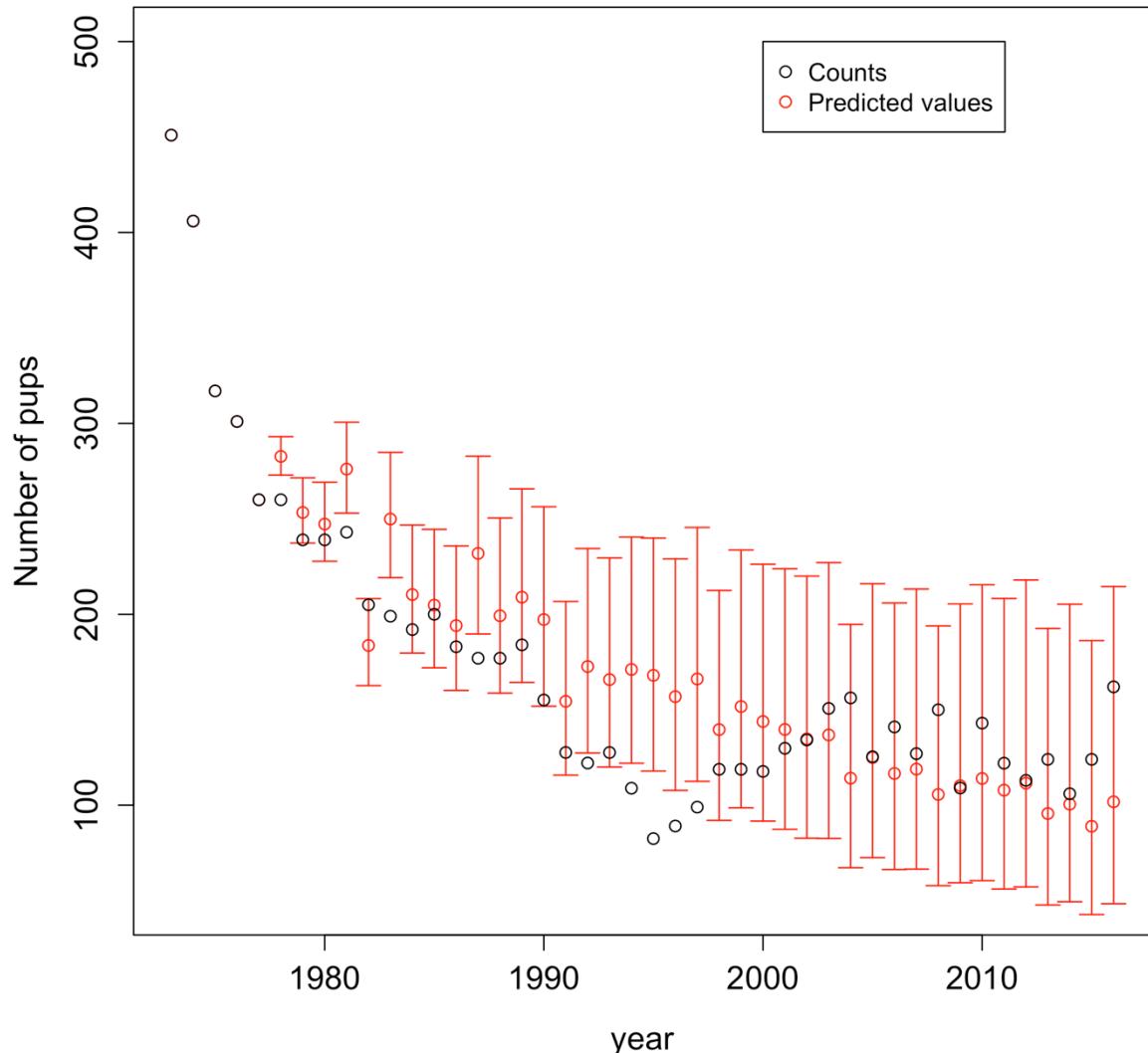
427



428

429 FIGURE 2: Directed Acyclic Graph (DAG) displaying the parameters and data used in the  
 430 IPM.  $m$  stands for the capture-recapture data,  $Y$  is for the pup counts data,  $f$  is for female  
 431 counts data. The notation for the model parameters are given in the main text.

432



433

434 FIGURE 3: The number of southern elephants seal pups born at Marion Island from 1973 to  
 435 2016. We compare the observed counts (in black) vs. the predicted counts (in red) with 95%  
 436 confidence intervals from the IPM combining both count and capture-recapture data. Note  
 437 that there are no estimates for the first five years – see equation (4).



Décrire et quantifier les réseaux sociaux animaux quand la détectabilité est imparfaite

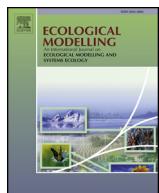
## 4.1 ARTICLE 3. Inferring animal social networks with imperfect detection

### Résumé

L'analyse de réseau social fournit un outil puissant pour comprendre l'organisation sociale des animaux. Cependant, dans les populations en liberté, il est presque impossible de surveiller de manière exhaustive les individus d'une population et de suivre leurs associations. Ignorer le problème de la détection individuelle imparfaite et éventuellement hétérogène peut conduire à un biais substantiel dans les mesures standards des caractéristiques d'un réseau social. Ici nous proposons un modèle à espace d'états pour analyser des données de capture-recapture et décrire un réseau social. Nous réalisons une étude de simulation pour valider notre approche. En outre, nous montrons comment la visualisation des réseaux et le calcul de métriques standards peuvent bénéficier de la prise en compte des probabilités de détection. La méthode est illustrée par les données d'une population de dauphins de Commerson (*Cephalorynchus commersonii*) en Patagonie Argentine. Notre approche constitue un pas en avant vers un cadre statistique général pour l'analyse des réseaux sociaux de populations d'animaux sauvages.

**Mots-clés:** Inférence bayésienne, Capture-recapture, modèles multiétats, réseau social.





## Inferring animal social networks with imperfect detection

Olivier Gimenez<sup>a,\*</sup>, Lorena Mansilla<sup>a</sup>, M. Javier Klaich<sup>b</sup>, Mariano A. Coscarella<sup>b,c</sup>, Susana N. Pedraza<sup>c</sup>, Enrique A. Crespo<sup>b,c</sup>

<sup>a</sup> CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHÉ, IRD, Montpellier, France

<sup>b</sup> Facultad de Ciencias Naturales y de la Salud, Universidad Nacional de la Patagonia San Juan Bosco, Blvd. Brown 3051, 9120, Puerto Madryn, Chubut, Argentina

<sup>c</sup> Laboratorio de Mamíferos Marinos, CESIMAR, CONICET, Blvd. Brown 2915, 9120 Puerto Madryn, Chubut, Argentina



### ARTICLE INFO

**Keywords:**  
Bayesian inference  
Capture-recapture  
Multistate models  
Social networks

### ABSTRACT

Social network analysis provides a powerful tool for understanding social organisation of animals. However, in free-ranging populations, it is almost impossible to monitor exhaustively the individuals of a population and to track their associations. Ignoring the issue of imperfect and possibly heterogeneous individual detection can lead to substantial bias in standard network measures. Here, we develop capture-recapture models to analyse network data while accounting for imperfect and heterogeneous detection. We carry out a simulation study to validate our approach. In addition, we show how the visualisation of networks and the calculation of standard metrics can account for detection probabilities. The method is illustrated with data from a population of Commerson's dolphin (*Cephalorhynchus commersonii*) in Patagonia Argentina. Our approach provides a step towards a general statistical framework for the analysis of social networks of wild animal populations.

### 1. Introduction

Knowledge of the social organisation of animal populations is essential to develop sound conservation and management strategies as social structure affects habitat use, information diffusion, as well as the genetic composition and the spread of information and diseases within these populations (Krause and Ruxton, 2002).

Social network analysis (SNA; Croft et al., 2008; Whitehead, 2008) has recently known an increasing number of applications to characterize in particular the social structure of animal populations. SNA allows the study of social networks through their visualisation and the calculation of several descriptive statistics, with important applications in ecology, evolution, epidemiology and behavioural ecology (Craft and Caillaud, 2011; Farine and Whitehead, 2015; Krause et al., 2007; Sih et al., 2009; Wey et al., 2008).

In free-ranging populations however, individuals may or may not be seen (or recaptured) at various times over a study period. This raises the issue of detectability less than one that makes it difficult to track associations between individuals. In other words, when one or two individuals of a dyad are missed, were they associated or not? Besides being imperfect, detection is often heterogeneous due to variation in individual traits such as, e.g., sex (Tavecchia et al., 2001), social status (Cubaynes et al., 2010; Hickey and Sollmann, 2019), infection status (Marescot et al., 2018) or pair-bond status (Choquet and Gimenez,

2012; Culina et al., 2013). Overall, ignoring the issue of imperfect and heterogeneous individual detection can lead to substantial bias in estimating the probability of association between individuals (Hoppitt and Farine, 2018; Lusseau et al., 2008; Weko, 2018).

To address these issues, Klaich et al. (2011) developed a capture-recapture model where detection probabilities of individuals in dyads varied between individuals that are associated and those that are not. Their approach requires complex probabilistic calculations that make it specific to their case study, and therefore difficult to extend to other situations. Here, we use a state-space modelling (SSM) approach (e.g., Buckland et al., 2004) to acknowledge that data on associations between individuals derived from field studies are imperfect observations of the underlying social structure. Specifically, the SSM approach makes the two-component process underlying network structure explicit: i) the temporal dynamic of associations between individuals and ii) the observations generated from the underlying process in i).

We apply the SSM framework to capture-recapture (CR) data (Gimenez et al., 2007) to analyse network data while accounting for imperfect and heterogeneous detection of individuals. We estimate dyad association probability and distinguish the dynamic of associated vs. non-associated states from their partial observation. We carry out a simulation study to assess bias in the association probability. Last, we show how the visualisation and the calculation of standard network metrics can account for detection probabilities. The approach is

\* Corresponding author at: CEFE, 1919 route de Mende, 34293, Montpellier, France.

E-mail address: [olivier.gimenez@cefe.cnrs.fr](mailto:olivier.gimenez@cefe.cnrs.fr) (O. Gimenez).

illustrated with data from a population of Commerson's dolphin (*Cephalorhynchus commersonii*) in Patagonia Argentina.

## 2. Model development

### 2.1. State-space modelling of capture-recapture data

Following Klaich et al. (2011), we derived dyad association histories from individual captures and non-captures. For example, let us assume a 4-occasion CR experiment in which two individuals have capture histories '1011' and '1001' where a '1' stands for an individual detection and '0' for a non-detection. We considered that behavioural interactions between individuals occurred within groups ('gambit of the group' *sensu* Whitehead and Dufault, 1999). Let us assume that these two individuals were both detected in the same group at the first occasion but in a different group at the last one, then the association history for this particular dyad is '2013' where '0' stands for none of the two individuals of a dyad are seen, '1' for one individual only of the dyad is seen, '2' for the two individuals of a dyad are seen associated and '3' for the two individuals of a dyad are seen non-associated.

To analyse these dyadic data, we implemented a SSM formulation (Gimenez et al., 2007) of multistate CR models (Lebreton et al., 2009) for closed populations. We considered two states A and B for 'dyad associated' and 'dyad non-associated' respectively. We denoted  $x_t^i$ , a multinomial trial taking values (1,0) or (0,1) if, at time  $t$ , dyad  $i$  is in state A or B respectively. Given the underlying states, a dyad may be recaptured in the observations 0, 1, 2 or 3 defined above considering imperfect detection. We denoted  $y_t^i$ , a multinomial trial taking values (1,0,0,0), (0,1,0,0), (0,0,1,0), (0,0,0,1) if, at time  $t$ , dyad  $i$  is observed as a 0, 1, 2 or 3. The state-space model relies on a combination of two equations. First, the state equation specifies the state of dyad  $i$  at time  $t$  given its state at time  $t - 1$ :

$$x_t^i \sim \text{Multinomial}(1, \Psi x_{t-1}^i)$$

where  $\Psi$  gathers the probabilities for a dyad of staying associated and non-associated between two successive occasions (Table 1a). We also defined the probability  $\pi$  for a dyad of being in initial state associated. Second, the observation equation specifies the observation of dyad  $i$  at time  $t$  given its state at time  $t$ :

$$y_t^i \sim \text{Multinomial}(1, P x_t^i)$$

where  $P$  gathers the detection probabilities and of an individual being associated and non-associated in a dyad (Table 1b).

**Table 1**

Transition matrices used in the state and observation equations of the state-space CR network model. States A and B are for associated and non-associated. Parameters  $p$  and  $\psi$  are the detection and transition probabilities.

a) State matrix

Previous occasion	Current occasion	
	A	B
A	$\psi^{AA}$	$1 - \psi^{AA}$
B	$1 - \psi^{BB}$	$\psi^{BB}$

b) Observation matrix  $P$

Current occasion	Current occasion			
	0	1	2	3
A	$(1 - p^A)(1 - p^A)$	$2p^A(1 - p^A)$	$p^A p^A$	0
B	$(1 - p^B)(1 - p^B)$	$2p^B(1 - p^B)$	0	$p^B p^B$

**Table 2**

Bias in parameter estimates for the homogeneous scenarios.

scenario	p	$\pi$	$\psi^{AA}$	$\psi^{BB}$	bias p	bias $\pi$	bias $\psi^{AA}$	bias $\psi^{BB}$
1	0.3	0.2	0.1	0.1	0.50	26.49	120.98	58.77
2	0.8	0.2	0.1	0.1	0.08	9.46	4.37	1.36
3	0.3	0.7	0.1	0.1	-0.33	-1.23	142.04	22.83
4	0.8	0.7	0.1	0.1	-0.21	-3.23	8.91	-0.04
5	0.3	0.2	0.4	0.1	0.07	14.73	27.30	53.03
6	0.8	0.2	0.4	0.1	-0.04	1.02	-1.65	4.74
7	0.3	0.7	0.4	0.1	0.60	-10.96	65.19	26.16
8	0.8	0.7	0.4	0.1	-0.04	-0.37	-8.88	0.79
9	0.3	0.2	0.9	0.1	0.29	4.46	-23.10	37.30
10	0.8	0.2	0.9	0.1	0.11	2.29	-5.57	7.26
11	0.3	0.7	0.9	0.1	-0.25	0.30	-14.44	28.57
12	0.8	0.7	0.9	0.1	0.07	-0.55	-7.99	3.91
13	0.3	0.2	0.1	0.4	-0.74	54.58	45.20	24.95
14	0.8	0.2	0.1	0.4	-0.08	6.23	2.19	4.71
15	0.3	0.7	0.1	0.4	0.27	-25.83	29.36	7.66
16	0.8	0.7	0.1	0.4	-0.11	-11.71	3.05	1.72
17	0.3	0.2	0.4	0.4	0.45	14.96	10.59	21.80
18	0.8	0.2	0.4	0.4	-0.09	3.22	-1.45	-0.27
19	0.3	0.7	0.4	0.4	0.64	-13.37	5.67	5.96
20	0.8	0.7	0.4	0.4	0.02	0.24	-1.44	-0.71
21	0.3	0.2	0.9	0.4	-0.26	8.35	-17.84	-28.74
22	0.8	0.2	0.9	0.4	0.01	1.28	-1.62	-1.72
23	0.3	0.7	0.9	0.4	0.45	-1.59	-10.12	-5.75
24	0.8	0.7	0.9	0.4	-0.08	-0.52	-2.47	-0.54
25	0.3	0.2	0.1	0.9	0.94	38.86	21.21	-1.08
26	0.8	0.2	0.1	0.9	0.08	8.48	2.90	0.87
27	0.3	0.7	0.1	0.9	0.11	-47.67	10.35	-2.45
28	0.8	0.7	0.1	0.9	-0.34	2.48	1.29	-0.87
29	0.3	0.2	0.4	0.9	-0.46	11.66	-4.68	-16.83
30	0.8	0.2	0.4	0.9	-0.22	2.55	-0.36	-1.68
31	0.3	0.7	0.4	0.9	-0.27	-6.82	-7.96	-7.75
32	0.8	0.7	0.4	0.9	0.04	-1.00	-0.86	-1.29
33	0.3	0.2	0.9	0.9	1.18	3.33	-30.90	-55.94
34	0.8	0.2	0.9	0.9	0.12	0.47	-1.45	-2.53
35	0.3	0.7	0.9	0.9	-0.74	-3.30	-20.09	-38.39
36	0.8	0.7	0.9	0.9	-0.16	-1.19	-0.85	-1.26

### 2.2. Bayesian fitting using MCMC methods

We used Bayesian theory in conjunction with Markov Chain Monte Carlo (MCMC) methods to carry out inference. Inference was based on empirical medians and credible intervals. As a by-product of the MCMC simulations, we also obtained numerical summaries for any function of the parameters, in particular the metrics describing the network structure.

### 2.3. Calculating network measures while accounting for imperfect detection

In SNA, a wide range of descriptive statistics can be used to characterize the properties of the structure of a network. Here, we focused on four of them. We used for each animal in the network the number of other animals with which it was associated – *degree* – and the number of shortest paths between pairs of animals that passed through it – *betweenness*. In addition, we quantified the degree to which an animal's immediate neighbours were associated – *cluster coefficient* – and the average of all path lengths between all pairs of animals in the network – *average path length* (Croft et al., 2008). These measures are useful to characterize the properties of a network regarding the spread of disease or information (Craft and Caillaud, 2011; Watts and Strogatz, 1998).

A feature of MCMC algorithms is that the dyad states  $x_t^i$ 's are treated as parameters to be estimated, just like the transition and detection probabilities. We generated values from the posterior distributions of the dyads' states, which, in turn, were used to visualize the network and characterize its structure over time. Specifically, for each MCMC iteration, we calculated the degree and betweenness for each individual (R package sna; Butts, 2008), as well as the clustering coefficient and the average path length (R package igraph; Csardi and Nepusz, 2006),

**Table 3**  
Bias in parameter estimates for the heterogeneous scenarios.

scenario	p <sup>A</sup>	p <sup>B</sup>	π	ψ <sup>AA</sup>	ψ <sup>BB</sup>	bias p <sup>A</sup>	bias p <sub>B</sub>	bias π	bias ψ <sup>AA</sup>	bias ψ <sup>BB</sup>
1	0.3	0.8	0.2	0.1	0.1	0.59	-0.24	4.67	5.86	5.61
2	0.3	0.8	0.7	0.1	0.1	0.18	-0.39	-6.01	1.17	12.27
3	0.3	0.8	0.2	0.4	0.1	-0.16	-0.65	-2.13	-6.94	2.94
4	0.3	0.8	0.7	0.4	0.1	0.84	-0.49	-25.65	-2.16	4.77
5	0.3	0.8	0.2	0.9	0.1	-0.02	-2.53	-15.23	-6.76	1.87
6	0.3	0.8	0.7	0.9	0.1	0.67	-14.33	-131.51	-8.99	8.40
7	0.3	0.8	0.2	0.1	0.4	-0.11	0.34	2.82	9.96	6.86
8	0.3	0.8	0.7	0.1	0.4	1.55	0.30	-0.32	2.34	0.53
9	0.3	0.8	0.2	0.4	0.4	-0.65	-0.69	0.88	-2.29	2.30
10	0.3	0.8	0.7	0.4	0.4	1.41	-1.75	-3.50	-2.06	7.67
11	0.3	0.8	0.2	0.9	0.4	0.49	-0.06	2.86	-3.88	-0.25
12	0.3	0.8	0.7	0.9	0.4	0.81	-4.39	-11.31	-1.57	2.71
13	0.3	0.8	0.2	0.1	0.9	4.24	-0.21	0.26	27.14	-1.08
14	0.3	0.8	0.7	0.1	0.9	1.63	-0.22	-0.49	1.94	-2.15
15	0.3	0.8	0.2	0.4	0.9	7.33	-0.96	-1.02	-1.77	0.34
16	0.3	0.8	0.7	0.4	0.9	-0.05	-0.38	-1.43	-1.41	-0.78
17	0.3	0.8	0.2	0.9	0.9	-1.53	-0.24	0.43	-8.22	-0.36
18	0.3	0.8	0.7	0.9	0.9	0.84	-0.48	0.53	-1.08	-1.45
19	0.8	0.3	0.2	0.1	0.1	-0.39	-0.01	11.75	4.63	3.82
20	0.8	0.3	0.7	0.1	0.1	0.03	-0.63	4.72	1.48	7.67
21	0.8	0.3	0.2	0.4	0.1	0.47	0.73	16.29	3.33	3.73
22	0.8	0.3	0.7	0.4	0.1	0.52	-1.16	-0.14	1.50	5.51
23	0.8	0.3	0.2	0.9	0.1	0.00	0.04	17.21	-2.35	2.46
24	0.8	0.3	0.7	0.9	0.1	-0.13	1.98	-8.68	0.17	4.71
25	0.8	0.3	0.2	0.1	0.4	-0.94	0.38	2.54	11.96	-26.02
26	0.8	0.3	0.7	0.1	0.4	-0.79	-0.64	-0.89	1.91	-11.83
27	0.8	0.3	0.2	0.4	0.4	-0.91	-0.14	5.53	8.95	-3.96
28	0.8	0.3	0.7	0.4	0.4	-0.52	0.57	-0.87	1.32	-3.13
29	0.8	0.3	0.2	0.9	0.4	-0.48	1.01	2.13	-3.45	0.13
30	0.8	0.3	0.7	0.9	0.4	-0.48	0.58	-1.17	0.17	-1.46
31	0.8	0.3	0.2	0.1	0.9	-30.02	21.00	27.86	112.23	-196.71
32	0.8	0.3	0.7	0.1	0.9	-1.90	0.23	2.38	1.93	-8.34
33	0.8	0.3	0.2	0.4	0.9	-9.87	2.22	7.98	27.76	-6.73
34	0.8	0.3	0.7	0.4	0.9	-1.40	0.28	0.70	1.65	-1.85
35	0.8	0.3	0.2	0.9	0.9	0.14	-0.29	0.86	-8.69	-1.06
36	0.8	0.3	0.7	0.9	0.9	0.22	0.16	-0.41	-0.32	-1.45

**Table 4**  
Parameters estimates (posterior medians) with 95% credible intervals for the Commerson's dolphin case study.

Parameter	Estimate with 95% credible interval				
	Occasion 1	Occasion 2	Occasion 3	Occasion 4	Occasion 5
Average path length	1.31 [1.25; 1.38]	1.65 [1.54; 1.79]	1.61 [1.57; 1.66]	1.60 [1.55; 1.65]	1.61 [1.56; 1.66]
Clustering coefficient	0.68 [0.61; 0.74]	0.36 [0.27; 0.45]	0.42 [0.39; 0.45]	0.39 [0.35; 0.43]	0.40 [0.36; 0.43]
Individual detection	0.27 [0.26; 0.28]	0.11 [0.10; 0.12]	0.44 [0.42; 0.45]	0.17 [0.16; 0.18]	0.20 [0.19; 0.21]
Staying associated			0.33 [0.17; 0.50]		
Staying non-associated			0.57 [0.48; 0.69]		

hence obtaining the posterior distribution for each of these metrics. Data and codes are available on GitHub [https://github.com/oliviergimenez/social\\_networks\\_capture\\_recapture](https://github.com/oliviergimenez/social_networks_capture_recapture).

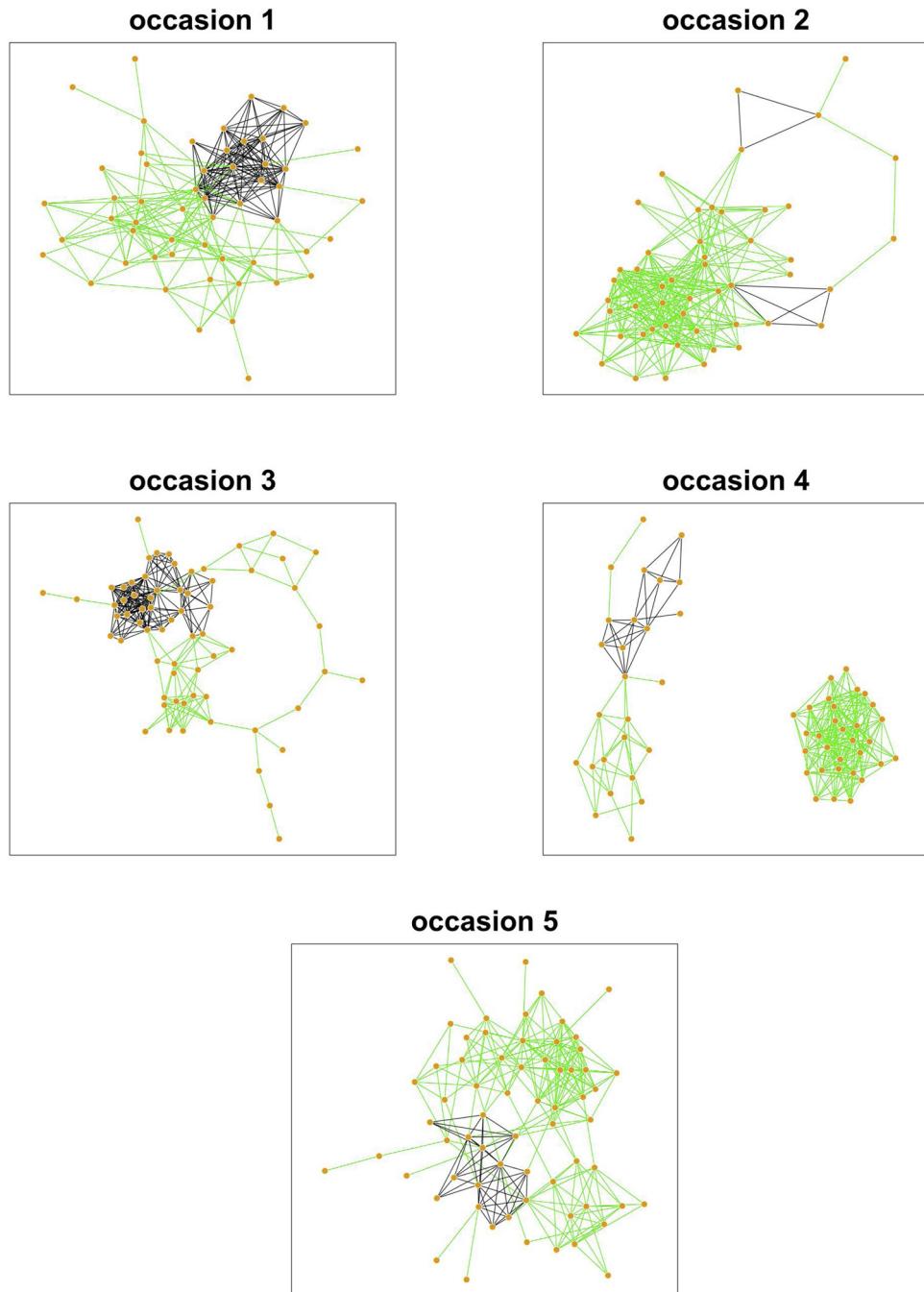
### 3. Simulation study

We conducted a simulation study to assess the bias in parameter estimates. We considered a scenario where detection probabilities were homogeneous. We simulated 100 CR datasets with  $\pi = 0.2, 0.7, \psi^{AA} = 0.1, 0.4, 0.9$  and  $\psi^{BB} = 0.1, 0.4, 0.9$  and  $p^A = p^B = 0.3, 0.8$  (in total, 36 different configurations) and to each simulated dataset we fitted a CR model with homogeneous detection probabilities. We also considered a heterogeneous scenario where all parameters were set to the same values as in the homogeneous scenarios, except the detection probabilities which we set to  $p^A = 0.3, p^B = 0.8$  and  $p^A = 0.8, p^B = 0.3$  (in total, 36 different configurations). We fitted a model with heterogeneous detection probabilities to these simulated datasets. For both the homogeneous and the heterogeneous scenarios, we calculated the relative bias of all parameters.

For the homogeneous scenarios, the bias decreased when detection increased (Table 2). Bias was negligible on detection, around +5% on the transition probabilities and around -13% on  $\pi$  in scenario 19 with  $\psi^{BB} = 0.4$ . When  $\psi^{BB} = 0.9$  in scenario 31, the bias in  $\pi$  decreased by a factor 2. For the heterogeneous scenarios, the bias was negligible, except for scenario 31 in which the proportion of associated dyads was low and all dyads tended to remain non-associated (Table 3).

### 4. Case study

To illustrate our methodological approach, we used a real-world example as a case study. We used photo-identification data on a population of Commerson's dolphin (*C. commersonii*) that was monitored in the coastal waters near the Chubut River mouth ( $43^{\circ}20' S$ ,  $65^{\circ}00' W$ ) in the Patagonian coast (Coscarella et al., 2003). Commerson's dolphins are particularly abundant in the area during the austral spring (Coscarella et al., 2010). The mean residence time in the sampling area was 15 days (SE = 6.4), therefore we sampled 5 times in October 2007 to unravel which individual was associated with which,



**Fig. 1.** Visualisation of the network for the Commerson's dolphin population, over five sampling occasions, for the year 2007, showing associations (lines) between individuals (orange circles). For each edge, we calculated the average number of times the corresponding dyad was estimated as being associated ( $x = 1$ ) over the total number of MCMC simulations. Then, we displayed only the edges for which this number was larger than the 0.90 quantile of the distribution of  $x$ . Black edges are for observed dyads (also corresponding to  $x = 1$  for simulations) while green edges are for dyads that are estimated to be associated (with probability 0.69, 0.39, 0.42, 0.42 and 0.40 for capture occasion 1, 2, 3, 4 and 5 respectively) but for which one or the two individuals were not detected. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

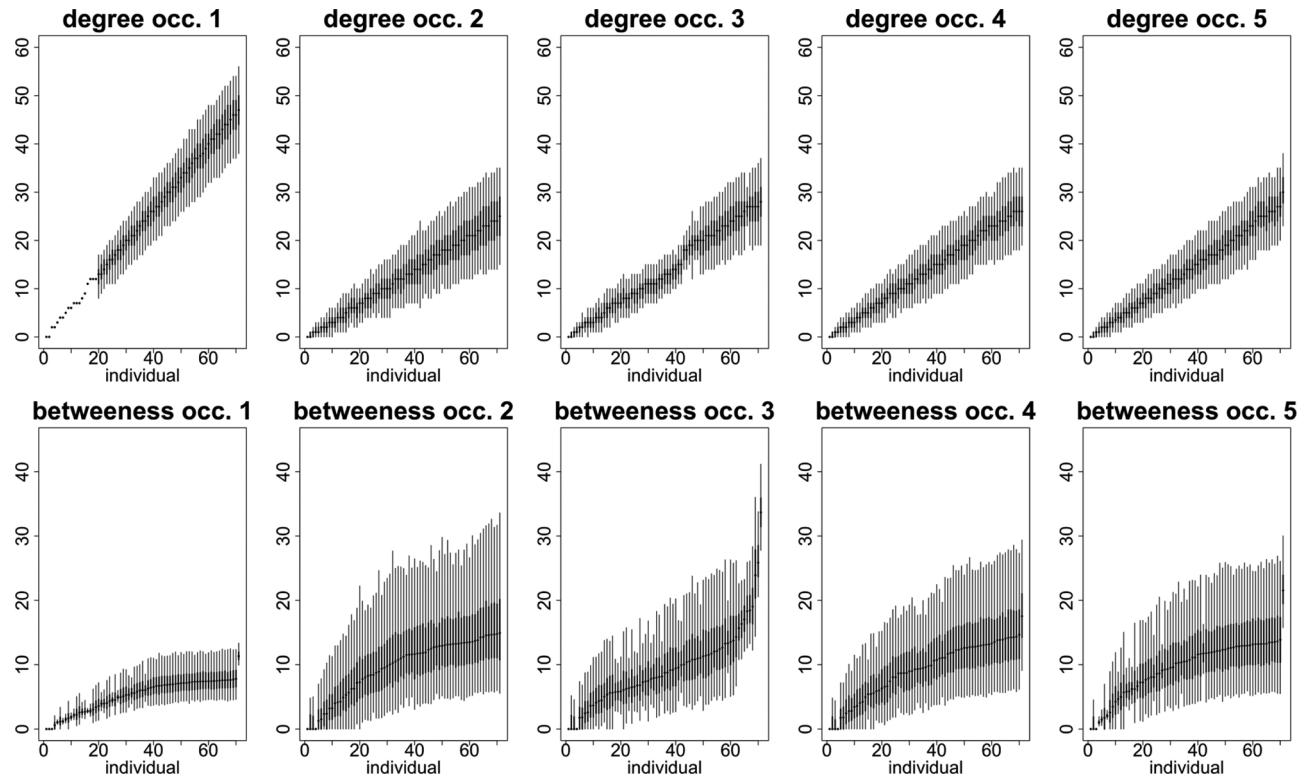
while arriving and leaving the area together (Coscarella et al., 2011). Two individuals were considered associated when they were photo-identified during the same encounter, while they were considered not associated otherwise (Coscarella et al., 2011).

Over the study, a total of 71 dolphins were detected which led to  $71*(71-1)/2 = 2485$  association histories. Based on previous analyses (Klaich et al., 2011), we considered time-dependent state-independent individual detection probabilities. Individual detections varied between 11% and 44% (Table 4). The probability of staying associated was 33% while that of staying non-associated was 57% with very little overlap in the credible intervals (Table 4), suggesting a high turnover in the dynamic of associations and a fission-fusion social organization.

Along the five sampling occasions, the estimated network showed changes in its structure (Fig. 1). At occasions 1, 2, 3 and 5, the estimated network had a single component with a higher number of associated dyads at occasion 1 than at occasions 2, 3 and 5. Although the

number of dyads was higher at occasion 1, all networks were fully connected (i.e. none individual or group of individuals were isolated from other individuals). At occasion 4, the network estimated had two components, isolated from each other (i.e. none of the individuals from one component was associated with any of the individuals in the other component). This suggests that at least two groups might exist having preferential associations between individuals inside each group.

Average path length was lower on the first sampling occasion than in the subsequent ones, while the reverse pattern was observed for the clustering coefficient (Table 4). These estimated values also suggest high individual connectivity and that the estimated social network has features related to a small-world type network. At the individual level, degree was heterogeneous (Fig. 2), with individuals spreading all over the range of its distribution (Fig. 2). In contrast, betweenness appeared relatively homogeneous, despite some dolphins with low betweenness and a single animal with very high betweenness (Fig. 2).



**Fig. 2.** Local properties of the Commerson's dolphin network. For each individual and for each of the 5 capture occasions, degree (top panels) and betweenness (bottom panels) are summarized with the posterior mean (circle), the 50% (thick line) and 95% (thin line) credible intervals.

## 5. Discussion

We have proposed a new statistical approach combining network analyses with CR models formulated as state-space models. Our framework has several appealing advantages. First and most importantly, ignoring imperfect and possibly heterogeneous detection may lead to biased results about the structure and dynamics of associations (see Fig. 1). Our CR model provides a robust method to estimate social networks. Second, in addition to social status, our model can easily incorporate individual-level traits such as age or sex through regression-like functions. This opens an avenue towards investigating the relationships between the phenotype and social position of individuals. Third, our method provides unbiased and precise estimates of relevant metrics to characterise the properties of social networks (see the Simulation study section), the whole process being controlled for imperfect and heterogeneous detection. Another appealing feature of our approach is the quantification of uncertainty associated to network measures under the form of Bayesian credible intervals (Table 2 and Fig. 2). Last, the social organisation can be visualised over time while accounting for imperfect detection, providing the opportunity for testing socio-ecological hypotheses in free-ranging animal populations. For example, the rapid turnover of the free ranging Commerson's dolphin groups has been previously proposed (Coscarella et al., 2011), and here we could identify this turnover within the fission-fusion society model.

When inspecting the results of the dolphin case study, there are advantages in adopting a CR approach to infer social networks. First, when it comes to visualizing the network, we illustrate in Fig. 1 what we would obtain with a standard approach with in black edges, while the green edges correspond to the dyads that are estimated to be associated with the new approach by correcting for imperfect detection. Clearly, the structure and dynamics of the network are different depending on whether we ignore imperfect detection (black edges only) or we consider the model-based estimated network (edges of both

colors). Second, regarding network metrics, the only way to estimate degree and betweenness for all occasions when non-detections occur (Fig. 2) is to resort to a CR approach to account for missing values.

Our CR model requires data on individuals that can be uniquely identifiable. Identifying individuals can be achieved using non-invasive marking (such as coat patterns, body scars, or genetic profiling for mammals; e.g., Cubaynes et al., 2010; Marescot et al., 2018; Santostasi et al., 2016) or invasive marking (such as rings for birds, colouring for insects or passive integrated transponders for fishes; e.g., Băncilă et al., 2018; Buoro et al., 2010; Lagrange et al., 2014). The model also needs data on interactions or associations. Here, we rely on the 'gambit of the group' method which states that all individuals within a group of animals observed at a point in time are associated (Farine and Whitehead, 2015).

Our model relies on several assumptions. First, we have considered closed populations while demographic process might occur in animal populations. The extension of our model to open populations is feasible (Lebreton et al., 2009) to incorporate survival and dispersal, therefore allowing to assess the influence of social structure on fitness. Second, we assumed that association states were correctly assigned while some uncertainty might occur due to incomplete information. In the SSM framework, incorporating uncertainty in state assignment is relatively straightforward (Gimenez et al., 2012; Pradel, 2005). Third, we assumed independence of the association histories to form the SSM likelihood. To account for an individual effect, random effects can be incorporated in CR models (Choquet et al., 2013; Choquet and Gimenez, 2012; Gimenez and Choquet, 2010), which opens a promising avenue towards a general statistical framework for the analysis of animal social networks (Cross et al., 2012; Van Duijn et al., 2004).

Overall, we hope our proposal will foster applications of social network analysis to free-ranging animal population in behavioural ecology to describe social behaviour and social dynamics, in evolution ecology to explore the fitness consequences of the social positions of individuals and in epidemiological ecology to determine the

implications of network structure and dynamics in the spread of diseases.

## Competing interests

The authors declare no competing interests.

## Acknowledgements

This paper is dedicated to our colleague Dr. Susana N. Pedraza, who passed away while this work was being conducted. Logistic support was provided by Centro Nacional Patagónico (CONICET). This study was partially funded by the National Research Council of Argentina (CONICET-PIP 0742/98), Agencia Nacional de Promoción Científica y Tecnológica (01-04030A, 11679, and 33934), Fundación BBVA (BIOCON 04) and GEF Project CNP-BB 27. This work is part of the second author's doctoral thesis funded by CONICYT, PFCHA/DOCTORADO BECAS CHILE/2016-72170563. Fieldwork was authorised by the Dirección de Áreas Protegidas de la Provincia del Chubut.

## References

- Băncilă, R.I., Pradel, R., Choquet, R., Plăiașu, R., Giménez, O., 2018. Using temporary emigration to inform movement behaviour of cave-dwelling invertebrates: a case study of a cave harvestman species. *Ecol. Entomol.* 43, 551–559. <https://doi.org/10.1111/een.12645>.
- Buckland, S.T., Newman, K.B., Thomas, L., Koesters, N.B., 2004. State-space models for the dynamics of wild animal populations. *Ecol. Model.* 171, 157–175. <https://doi.org/10.1016/j.ecolmodel.2003.08.002>.
- Buoro, M., Prévost, E., Giménez, O., 2010. Investigating evolutionary trade-offs in wild populations of atlantic salmon (*salmo salar*): incorporating detection probabilities and individual heterogeneity. *Evolution* 64, 2629–2642. <https://doi.org/10.1111/j.1558-5646.2010.01029.x>.
- Butts, C.T., 2008. Social network analysis with sna. *J. Stat. Softw.* 24. <https://doi.org/10.18637/jss.v024.i06>.
- Choquet, R., Giménez, O., 2012. Towards built-in capture-recapture mixed models in program E-SURGE. *J. Ornithol.* 152, 625–639.
- Choquet, R., Sanz-Aguilar, A., Doligez, B., Nogué, E., Pradel, R., Gustafsson, L., Giménez, O., 2013. Estimating demographic parameters from capture-recapture data with dependence among individuals within clusters. *Methods Ecol. Evol.* 4, 474–482. <https://doi.org/10.1111/2041-210X.12030>.
- Coscarella, M.A., Dans, S.L., Crespo, E.A., Pedraza, S.N., 2003. Potential impact of unregulated dolphin watching activities in Patagonia. *J. Cetacean Res. Manag.* 77–84.
- Coscarella, M.A., Pedraza, S.N., Crespo, E.A., 2010. Behavior and seasonal variation in the relative abundance of Commerson's dolphin (*Cephalorhynchus commersonii*) in northern Patagonia, Argentina. *J. Ethol.* 28, 463–470. <https://doi.org/10.1007/s10164-010-0206-4>.
- 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. *J. Mammal.* 92, 544–551. <https://doi.org/10.1644/10-MAMM-A-029.1>.
- Craft, M.E., Caillaud, D., 2011. network models: an underutilized tool in wildlife epidemiology? *Interdiscip. Perspect. Infect. Dis.* 2011, 1–12. <https://doi.org/10.1155/2011/676949>.
- Croft, D.P., James, R., Krause, J., 2008. *Exploring Animal Social Networks*. Princeton University Press, Princeton, NJ.
- Cross, P.C., Creech, T.G., Ebinger, M.R., Heisey, D.M., Irvine, K.M., Creel, S., 2012. Wildlife contact analysis: emerging methods, questions, and challenges. *Behav. Ecol. Sociobiol.* 66, 1437–1447. <https://doi.org/10.1007/s00265-012-1376-6>.
- Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research. *InterJ. Complex Syst.* 9.
- 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., Giménez, O., 2010. Importance of accounting for detection heterogeneity when estimating abundance: the case of French wolves. *Conserv. Biol.* 24, 621–626. <https://doi.org/10.1111/j.1523-1739.2009.01431.x>.
- Culina, A., Lachish, S., Pradel, R., Choquet, R., Sheldon, B.C., 2013. A multievent approach to estimating pair fidelity and heterogeneity in state transitions. *Ecol. Evol.* 3, 4326–4338. <https://doi.org/10.1002/ee3.729>.
- Farine, D.R., Whitehead, H., 2015. Constructing, conducting and interpreting animal social network analysis. *J. Anim. Ecol.* 84, 1144–1163. <https://doi.org/10.1111/1365-2656.12418>.
- Giménez, O., Choquet, R., 2010. Incorporating individual heterogeneity in studies on marked animals using numerical integration: capture-recapture mixed models. *Ecology* 91, 951–957. <https://doi.org/10.1007/s10336-010-0613-x>.
- Giménez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J.-P., Pradel, R., 2007. State-space modelling of data on marked individuals. *Ecol. Model.* 206, 431–438. <https://doi.org/10.1016/j.ecolmodel.2007.03.040>.
- Giménez, O., Lebreton, J.-D., Gaillard, J.-M., Choquet, R., Pradel, R., 2012. Estimating demographic parameters using hidden process dynamic models. *Theor. Popul. Biol.* 82, 307–316.
- Hickey, J.R., Sollmann, R., 2019. A new mark-recapture approach for abundance estimation of social species. *PLoS One* 14.
- Hoppitt, W.J.E., Farine, D.R., 2018. Association indices for quantifying social relationships: how to deal with missing observations of individuals or groups. *Anim. Behav.* 136, 227–238. <https://doi.org/10.1016/j.anbehav.2017.08.029>.
- Klaich, M.J., Kinas, P.G., Pedraza, S.N., Coscarella, M.A., Crespo, E.A., 2011. Estimating dyad association probability under imperfect and heterogeneous detection. *Ecol. Model.* 222, 2642–2650. <https://doi.org/10.1016/j.ecolmodel.2011.03.027>.
- Krause, J., Ruxton, G.D., 2002. *Living in Groups*. Oxford University Press, Oxford, UK.
- Krause, J., Croft, D.P., James, R., 2007. Social network theory in the behavioural sciences: potential applications. *Behav. Ecol. Sociobiol.* 62, 15–27. <https://doi.org/10.1007/s00265-007-0445-8>.
- Lagrange, P., Pradel, R., B??lise, M., Giménez, O., 2014. Estimating dispersal among numerous sites using capture-recapture data. *Ecology* 95, 2316–2323.
- Lebreton, J.-D., Nichols, J.D., Barker, R.J., Pradel, R., Spendelow, J.A., 2009. Modeling individual animal histories with multistate capture-recapture models. *Adv. Ecol. Res.* 41, 87–173. [https://doi.org/10.1016/S0065-2504\(09\)00403-6](https://doi.org/10.1016/S0065-2504(09)00403-6).
- Lousseau, D., Whitehead, H., Gero, S., 2008. Incorporating uncertainty into the study of animal social networks. *Anim. Behav.* 75, 1809–1815. <https://doi.org/10.1016/j.anbehav.2007.10.029>.
- Marescot, L., Benhaiem, S., Giménez, O., Hofer, H., Lebreton, J.-D., Olarte-Castillo, X.A., Kramer-Schadt, S., East, M.L., 2018. Social status mediates the fitness costs of infection with canine distemper virus in Serengeti spotted hyenas. *Funct. Ecol.* 32, 1237–1250. <https://doi.org/10.1111/1365-2435.13059>.
- Pradel, R., 2005. Multievent: an extension of multistate capture-recapture models to uncertain states. *Biometrics* 61, 442–447. <https://doi.org/10.1111/j.1541-0420.2005.00318.x>.
- Santostasi, N.L., Bonizzoni, S., Bearzi, G., Eddy, L., Giménez, O., 2016. A robust design capture-recapture analysis of abundance, survival and temporary emigration of three odontocete species in the gulf of Corinth, Greece. *PLoS One* 11, e0166650. <https://doi.org/10.1371/journal.pone.0166650>.
- Sih, A., Hanser, S.F., McHugh, K.A., 2009. Social network theory: new insights and issues for behavioral ecologists. *Behav. Ecol. Sociobiol.* 63, 975–988. <https://doi.org/10.1007/s00265-009-0725-6>.
- Tavecchia, G., Pradel, R., Boy, V., Johnson, A.R., Zilly, F.C., 2001. Sex- and age-related variation in survival and cost of first reproduction in Greater Flamingos. *Ecology* 82, 165–174.
- Van Duijn, M.A.J., Snijders, T.A.B., Zijlstra, B.J.H., 2004. p2: a random effects model with covariates for directed graphs. *Stat. Neerl.* 234–254.
- Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 3.
- Weko, C.W., 2018. Isolating bias in association indices. *Anim. Behav.* 139, 147–159. <https://doi.org/10.1016/j.anbehav.2018.03.011>.
- Wey, T., Blumstein, D.T., Shen, W., Jordán, F., 2008. Social network analysis of animal behaviour: a promising tool for the study of sociality. *Anim. Behav.* 75, 333–344. <https://doi.org/10.1016/j.anbehav.2007.06.020>.
- Whitehead, H., 2008. *Analyzing Animal Societies: Quantitative Methods for Vertebrate Social Analysis*. University of Chicago Press.
- Whitehead, H., Dufault, S., 1999. Techniques for analyzing vertebrate social structure using identified individuals: review and recommendations. *Advances in the Study of Behavior*. Elsevier, pp. 33–74. [https://doi.org/10.1016/S0065-3454\(08\)60215-6](https://doi.org/10.1016/S0065-3454(08)60215-6).

Le comportement des individus est lié à leur survie et constitue une composante essentielle à la compréhension de la persistance des espèces. L'on s'intéresse entre autres à savoir si un individu peut survivre jusqu'à assurer la survie de ses descendants, quelle est son espérance de vie et comment les caractéristiques de socialité mesurées à l'échelle individuelle impactent les populations. En effet, l'individu est lié à la population dont il fait partie au travers de son cycle de vie (Caswell 2001). Souvent, le cycle de vie d'un individu est structuré par l'âge, mais l'âge ne détermine pas complètement le devenir d'une population, et d'autres facteurs comme la socialité y contribuent. Dans cette thèse, nous avons présenté trois modèles démographiques pour trois espèces sociales, en utilisant des données de capture-recapture, avec pour objectif de quantifier les caractéristiques sociales de ces espèces.

Dans le premier chapitre, nous avons combiné un modèle de capture-recapture spatialement explicite et une méthode de partitionnement pour estimer le nombre et la taille des meutes. Il y a peu de travaux dans la littérature qui se sont intéressés au problème (Caniglia et al. 2014 ; López-Bao et al. 2018; Mattioli et al. 2018). Notre approche permet, à partir des estimations de centres d'activité individuels, d'extraire l'information contenue dans les données de capture-recapture sur certaines composantes de la structure sociale des loups, y compris les tailles et emplacements probables des groupes. Cette information pourrait être utile pour déterminer les caractéristiques structurelles qui incitent les groupes à se déplacer, les individus à se disperser ou ce qui fait que les groupes disparaissent. Les modèles de capture-recapture spatialement explicites permettent de lier le type et l'amplitude des espaces que certains groupes occupent en fonction des membres qui composent ces groupes, leurs schémas de déplacement, les types d'habitats requis et leur succès et patrons de dispersion (Royle et al. 2017). Par conséquent, au travers de notre méthodologie, on pourrait aussi considérer l'hétérogénéité dans la détection (Pledger et al. 2003, Pledger et al. 2010, Royle 2008) en fonction de changements potentiels des états sociaux pendant le cycle de vie des individus (subordonné ou dominant par exemple) ce qui pourrait nous aider à mieux comprendre la dynamique de ces populations.

Dans notre deuxième chapitre, on développe un modèle intégré de population, en combinant des données de capture-recapture de femelles reproductrices d'éléphants de mer avec des données de comptage des petits, pour expliquer les variations temporelles dans les effectifs de petits. On estime aussi les probabilités de survie selon les états de cycle de vie et les probabilités de transition entre les états reproductifs. Nous montrons comment incorporer la structure sociale (sex-ratio) dans un modèle intégré, et comment la combinaison de données permet d'estimer des paramètres autrement non estimables (Ferrari et al. 2009). Notre approche généralise le travail de Ferrari et al. (2009) en proposant la modélisation intégrée et celui de Tenan et al. (2016) en permettant de tester formellement l'influence de la structure sociale sur la dynamique de la population. Une extension possible de notre approche consiste à intégrer explicitement les mâles, car pour certaines populations, leur comportement agressif via la compétition pour le territoire et les événements d'accouplement affecte la fécondité des femelles (Kvarnemo & Ahnesjö 1996). Techniquement, cela consisterait à incorporer des proportions de chaque sexe (Ferrari et al. 2009; Gerber et al.

---

2010) ou de considérer des matrices de projection de population à deux sexes (Gonzalez-Suarez & Gerber 2008, Gonzalez-Suarez 2014). Ces développements permettraient de déterminer l'impact des mâles sur la dynamique de la population.

Dans le troisième chapitre, nous proposons une nouvelle approche pour quantifier les réseaux sociaux grâce à des données de capture-recapture et l'illustrent sur une population de dauphins. Notre modèle d'inférence peut être facilement appliqué à d'autres espèces pour peu que l'on dispose des bases de données de capture-recapture et d'association des individus. L'approche permet d'améliorer notre connaissance des structures et dynamiques sociales, et du temps que les individus investissent dans ces relations sociales. Les relations possibles entre la structure sociale et la démographie des espèces ont été peu étudiées, les animaux marins ne faisant pas exception.

Dans les groupes sociaux d'épaulards, le rôle de certains individus dans les réseaux a été étudié (Williams & Lusseau 2006) et il a été constaté que, pour la même espèce, de fortes associations sont corrélées positivement à la survie apparente : les individus en groupes perturbés par plusieurs événements de mortalité ont des associations plus faibles avec une diminution conséquente des performances individuelles (Busson et al. 2018). Des réseaux sociaux ont également été développés pour étudier les comportements d'association des hyènes tachetées entre parents et non parents en liant ces comportements à la disponibilité des ressources (Holekamp et al. 2011), ou encore pour étudier la dynamique de contagion des agents pathogènes chez les macaques (Lehmann et al. 2016). Notre approche permet d'aborder ces questions en conditions naturelles en prenant en compte le défaut de détection. Une possibilité d'extension de l'approche consisterait à différentier les raisons de l'agrégation des individus; en effet, le réseau social peut changer de caractéristiques si l'on considère le fait que les individus sont en activité d'alimentation, de jeu, d'accouplement, etc.

L'inclusion du statut social a été récemment incorporé dans les analyses de viabilité des populations. Par exemple, Marescot et al. (2016) et Benhaiem et al. (2018) ont étudié les effets de la structure sociale sur l'épidémie de la maladie de Carré dans une population de hyènes tachetées en discrétilisant les états sociaux. Une piste de recherche serait d'inclure le réseau social directement dans l'analyse de viabilité.

En conclusion, la méthodologie développée dans cette thèse repose sur l'analyse de trois modèles de population, chacun incluant différents aspects de la structure sociale des espèces, à savoir le domaine vital, le cycle de vie individuel et les associations dyadiques. L'originalité du travail consiste à avoir combiné plusieurs types de modèles pour analyser des bases de données imparfaites, avec l'objectif de prendre en compte tous les individus, y compris ceux qui ne peuvent pas être observés de manière exhaustive. Nous espérons que ces développements encourageront l'étude de la démographie des espèces sociales dans leur milieu naturel.

## References

- [1] APOLLONIO M, MATTIOLI L, SCANDURA M, MAURI L, GAZZOLA A, AVANZINELLI E. 2004. Wolves in the Casentinesi Forests: insights for wolf conservation in Italy from a protected area with a rich wild prey community. *Biol Conserv* 120:249–260
- [2] BENHAIEM S, MARESCOT, L, EAST M., KRAMER-SCHADT, S, GIMENEZ O., LEBRETON J-D., AND HOFER HERIBERT. 2018. Slow recovery from a disease epidemic in the spotted hyena, a keystone social carnivore. *Communications biology*. 1:201
- [3] BUCKLAND, S.T., NEWMAN, K.B., THOMAS, L., KOESTERS, N.B. 2004. State-space models for the dynamics of wild animal populations. *Ecological Modelling* 171:157–175.
- [4] BUSSON, M., AUTHIER, M., BARBRAUD, C., TIXIER, P., REISINGER, R., JANC, A., ET GUINET C. 2019. Role of sociality in the response of killer whales to an additive mortality event. *PNAS*. 116(24): 11812-11817.
- [5] CANIGLIA R., FABBRI E., CUBAYNES S., GIMENEZ O., LEBRETON J.D., RANDI E. 2012. An improved procedure to estimate wolf abundance using non-invasive genetic sampling and capture–recapture mixture models. *Conservation Genetics*. 13(1):53-64.
- [6] CANIGLIA, R., FABBRI, E., GALAVERNI, M., MILANESI, P., AND RANDI, E. 2014. Noninvasive sampling and genetic variability, pack structure, and dynamics in an expanding wolf population. *J. Mamm.* 95(1): 41-59.
- [7] CARO, T. 1998. *Behavioral Ecology and conservation*. Oxford University Press.
- [8] DANERI, G., R. CARLINI. A AND RODHOUSE, P. 2000. Cephalopod diet of the southern elephant seal, *Mirounga leonina*, at King George Island, South Shetland Islands. *Antarctic Science*. 12(1):16-19.
- [9] DANERI, G. AND CARLINI, A. 2002. Fish prey of southern elephant seals, *Mirounga leonina*, at King George Island. *Polar Biol* 25(10):739–743.
- [10] FERRARI M., LEWIS M., PASCUAL M., AND CAMPAGNA C. 2009. Interdependence of social structure and demography in the southern elephant seal colony of Península Valdés, Argentina. *Marine mammal science*. 25(3): 681–692.
- [11] GERBER L., GONZÁLEZ-SUÁREZ M., HERNANDEZ- C. AND YOUNG J., SABO J. 2010. The cost of male aggression and polygyny in California Sea Lions (*Zalophus californianus*). *Plos One*. 5(8):1-8.
- [12] GONZALEZ-SUAREZ M. 2014. ¿Es relevante la ecología del comportamiento para entender y predecir la dinámica de las poblaciones? *Ecosistemas*. 23(4):93-97.
- [13] GERBER, L. 2006. Including behavioral data in demographic models improves estimates of population viability. *Front Ecol Environ*. 4(8): 419-427.

- [14] GIMENEZ, O., LEBRETON, J-D., GAILLARD, J-M., CHOQUET, R., AND PRADEL, R. 2012. Estimating demographic parameters using hidden process dynamic models. *Theoretical Population Biology.* 82:307–316.
- [15] GOODALL, R.N.P., GALEAZZI, A.R., LEATHERWOOD, S., MILLER, K.W., CAMERON, I.S., KASTELEIN, R., AND SOBRAL, A.P. 1988. Studies of Commerson's dolphins, *Cephalorhynchus commersonii*, off Tierra del Fuego, 1976-1984, with a review of information on the species in the South Atlantic. *Biology of the Genus Cephalorhynchus.* 9: 3-70.
- [16] GONZÁLEZ-SUÁREZ, M., AND GERBER, L. 2008. A Behaviorally Explicit Demographic Model Integrating Habitat Selection and Population Dynamics in California Sea Lions. *Conservation Biology* 22(6):1608-18.
- [17] GONZÁLEZ-SUÁREZ, M. 2014. ¿Es relevante la ecología del comportamiento para entender y predecir la dinámica de las poblaciones? *Ecosistemas.* 23(4):93-97.
- [18] HINDELL, M. A., BRADSHAW, C. J. A., SUMNER, M.D., KELVIN, J.M., AND BURTON, H.R. 2003. Dispersal of female southern elephant seals and their prey consumption during the austral summer: relevance to management and oceanographic zones. *Journal of Applied Ecology.* 40: 703–715.
- [19] HOEKAMP K., SMITH J., STRELIOFF C, VAN HORN R., AND WATTS H. 2011. Society, demography and genetic structure in the spotted hyena. *Molecular Ecology.* 1-20.
- [20] IÑÍGUEZ, M.A. AND TOSSENBERGER, V.P. 2010. Commerson's Dolphins (*Cephalorhynchus commersonii*) off Ría Deseado, Patagonia, Argentina. *Aquatic Mammals Journal.* 33(3): 276-285.
- [21] KAPPELER, P., AND VAN SCHAIK, C. 2002. Evolution of Primate Social systems. *International Journal of Primatology.* 23(4):707-740.
- [22] KRUSCHKE, J. 2010. Doing Bayesian Data Analysis: A Introduction with R and BUGS. Academic Press.
- [23] KVARNEMO C. AND AHNESJÖ I. 1996. The dynamics of operational sex ratios and competition for mates. *Tree.* 11(10):404-408.
- [24] LAWS, R.M. 1953. The Elephant Seal (*Mirounga leonina*, Linn.): I. Growth and age. London, HMSO, 62pp. (Falkland Islands Dependencies Survey Scientific Reports, 8).
- [25] LAWS R.M. 1994. History and present status of elephant seal populations. In: *Elephant seals: population ecology, behavior, and physiology*, Le Boeuf BJ, Laws RM (eds), pp. 49-65, University of California Press, Berkeley. USA.
- [26] LE BOEUF, B.J., AND LAWS, R.M. 1994. Elephant seals:an introduction to the genus. In LE boeuf, B.J. and R.M. Laws (Eds.) *Elephant seals: population ecology, behavior, and physiology.* pp. 1-26, University of California Press, Berkeley, USA.
- [27] LEE, P.M. 2012. *Bayesian statistics an introduction.* Wiley and Sons, Ltd. Fourth edition. Pp: 462.

## REFERENCES

---

- [28] LEHMANN, J., MAJOLO, B., AND MCFARLAND, R. 2016. The effects of social network position on the survival of wild Barbary macaques, *Macaca sylvanus*. *Behavioral Ecology*, 27 (1): 20-28.
- [29] LÓPEZ-BAO J.V., GODINHO R., PACHECO C., LEMA F. J., GARCÍA E., LLANEZA L., PALACIOS V. AND JIMÉNEZ J. 2018. Toward reliable population estimates of wolves by combining spatial capture-recapture models and non-invasive DNA monitoring. *Scientific Reports*.8: 2177.
- [30] MARESCOT, L., BENHAIEM, S., GIMENEZ O., HOFER, H., LEBRETON, J-D., OLARTE-CASTILLO, X., KRAMER-SCHADT, S., EAST, M. 2018. Social status mediates the fitness costs of infection with canine distemper virus in Serengeti spotted hyenas. *Functional Ecolog.* 32(5):1237-1250.
- [31] MATTIOLI L, CANU A, PASSILONGO D, SCANDURA M, AND APOLLONIO M. 2018. Estimation of pack density in grey wolf (*Canis lupus*) by applying spatially explicit capture-recapture models to camera trap data supported by genetic monitoring. *Frontiers in Zoology*.15:38.
- [32] MECH, L.D. 1999. Alpha status, dominance, and division of labor in wolf packs. *Can. J. Zool.* 77:1196-1203.
- [33] MECH, L.D., ET BOITANI, L. 2003. *Wolves: behavior, ecology, and conservation*. University of Chicago Press. Pp. 472.
- [34] OOSTHUIZEN C., POSTMA M., ALTWEGG R., NEVOUX M., PRADEL R., BESTER M., DE BRUYN P.J.N. 2019. Individual heterogeneity in life-history trade-offs with age at first reproduction in capital breeding elephant seals. *Population Ecology*. 1-15.
- [35] OOSTHUIZEN, C. W., PRADEL, R., BESTER, M., AND DE BRUYN, P.J.N. 2019. Making use of multiple surveys: Estimating breeding probability using a multievent-robust design capture–recapture model. *Ecology and Evolution*, 9(2):836-848.
- [36] PISTORIUS, P.A., DE BRUYN, P. AND BESTER M.N. 2011. Population dynamics of southern elephant seals: a synthesis of three decades of demographic research at Marion Island. *African Journal of Marine Science* 33:523–534.
- [37] PLEDGER S, POLLOCK K, NORRIS J. 2003. Open Capture-Recapture Models with Heterogeneity: I. Cormack-Jolly-Seber Model. *Biometrics*. 59(4):786-794.
- [38] PLEDGER, S., POLLOCK, K.H. AND NORRIS, J.L. 2010. Open Capture–Recapture Models with Heterogeneity: II. Jolly–Seber Model. *Biometrics* 66, 883–890.
- [39] SUTHERLAND, W. 1998. The importance of behavioural studies in conservation biology. *Animal Behaviour*. 56:801-809.
- [40] RABINER, L.R. 1989. A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEE*, 77(2):257-286.

- [41] ROYLE J.A., KULLER A, AND SUTHERLAND C. 2017. Unifying population and landscape ecology with spatial capture-recapture. *Ecography* 40:1-12.
- [42] ROYLE, J.A. 2008. Modeling Individual Effects in the Cormack–Jolly–Seber Model: A State–Space Formulation. *Biometrics* 64, 364–370.
- [43] UICN. Page WEB: <https://www.iucnredlist.org/species> (visité: 22/08/2019).
- [44] TENAN S., IEMMA A., BRAGALANTI N., PEDRINI P, DE BARBA M., RANDI E., GROFF C. AND GENOVART. 2016. Evaluating mortality rates with a novel integrated framework for nonmonogamous species. *Conservation Biology*, 30(6):1307-1319.
- [45] VUCETICH, J., PETERSON, R., AND WAITE, T. 1997. Effects of social structure and prey dynamics on extinction risk in gray wolves. *Conservation Biology*. 11(4): 957-965.
- [46] WILSON, E. O. 1998. *Sociobiology: the new synthesis*. Cambridge, MA: Belknap Press. 8° edition. Schaub, M. and F. Abadi. 2011. Integrated population models: a novel analysis framework for deeper insights into population dynamics. *J Ornithol.* 152(1):S227-S237.
- [47] WILIAMS R., AND LUSSEAU D. 2006. A killer whale social network is vulnerable to targeted removals. *Biol. Lett.* 2(4):497–500.
- [48] Withehead, H. 2008. *Analyzing Animal Societies, Quantitative Methods for Vertebrate Social Analysis*. The University of Chicago Press.
- [49] ZEIGLER, S. AND WALTERS, J. 2014. Population models for social species: lessons learned from models of Red-cockaded Woodpeckers (*Picoides borealis*). *Ecological Applications*. 24(8): 2144-2154.

**Annexe A: A propos du Chapitre 2**







```

# Parameters to estimate
params = c("lam0","sigma2","N","psi","SX","SY")

# Initial values
sigma2 = 5
psi = .6
SX = as.vector(runif(M,bb[1,1], bb[1,2])/scale_par)
SY = as.vector(runif(M,bb[2,1], bb[2,2])/scale_par)
init1 = list(psi=psi, lam0=.2, SX=SX, SY=SY, sigma2=sigma2)
init2 = list(psi=psi, lam0=.2, SX=SX, SY=SY, sigma2=sigma2)
inits<-list(init1,init2)

# call JAGS from R
library(jagsUI) # jagsUI is supposed to be more verbose than R2jags.

# mcmc options
ni <- 6000
na <- 1000
nc <- 2
nb <- 1000
nt <- 3

start<-as.POSIXlt(Sys.time())
out1_caniglia_6mesesless <- jagsUI::jags(data = dataless, # 9 min 19 sec to
run
    inits = inits,
    parameters.to.save = params,
    model.file = "SECRmodel.txt",
    n.chains = nc,
    n.adapt = na,
    n.iter = ni,
    n.burnin = nb,
    n.thin = nt,
    DIC = TRUE,
    parallel=TRUE,
    n.cores=2)
end <-as.POSIXlt(Sys.time())
duration = end-start
print(out1_caniglia_6mesesless)
save(out1_caniglia_6mesesless, file = 'reswolf.RData')

# numerical summaries
library(MCMCvis)
round(MCMCsummary(out1_caniglia_6mesesless,
    params = c('lam0','sigma2','N'),
    Rhat = TRUE,
    n.eff = TRUE),2)

# posterior distributions
library(ggmcmc)
mcmc_df <- as.data.frame(as.matrix(out1_caniglia_6mesesless $samples))
par_mcmc <- as.matrix(mcmc_df[,c('lam0','sigma2','N')])
par_mcmc <- as.mcmc(par_mcmc)
par <- ggs(par_mcmc)
p <- ggplot(par, aes(x = value)) +
  geom_density(aes(fill = Parameter), alpha = 0.5) +
#  geom_line(aes(y = density), data = par, colour = "red") +
  facet_wrap(~ Parameter, scales='free')
ggsave('dens.jpg',dpi=600, width = 30, height = 20, units = "cm")

```



```

# pack.number
# 8      9     10     11     12     13     14
# 4    104   647  1371   920   249    37

pn <- enframe(pack.number)
ggplot(pn, aes(x = value)) + geom_histogram() + scale_color_grey() +
scale_fill_grey() + labs(title="",x="Number of packs", y="Density") +
theme_classic()ggsave('histpacks.jpg',dpi=600)
# list of pack sizes per iteration (note: the biggest pack over the # whole
sample has 27 individuals)
pack.size <- lapply(pack.composition, function(cluster)
as.numeric(lapply(cluster,length)))
# mean pack size per iteration
mean.pack.size <- as.numeric(lapply(pack.size,mean)) # ranges from
# 6.71 to 11.75
mean(mean.pack.size) # 8.46
sd(mean.pack.size) # 0.75

library(reshape2)
library(igraph)

# matrix of frequencies of dyads
freq_matrix <- interact
freq_matrix[upper.tri(freq_matrix)] <- prod(dim(freq_matrix))
freq_df <- melt(freq_matrix)
# filter out the upper matrix values
freq_df <- filter(freq_df, value != nrow(freq_df)) %>% filter(Var1 != Var2)
hist(freq_df$value)

# create adjacency list with frequencies > 0.9
adj_list <- freq_df %>% filter(value > 0.9)
names(adj_list) <- c('from', 'to', 'weight') # Frequency of dyads

dim(adj_list)

# create igraph S3 object
net <- graph.data.frame(adj_list, directed = FALSE)

# store original margins
orig_mar <- par()$mar

plot(net, layout = layout_components(net), edge.width = E(net)$weight,
vertex.size=1.5, vertex.color="blue")

# how many packs in the wolf real data
library(lubridate)
read_csv2('DB_ISPRA_23032017_filtered_Olivier.csv') %>%
select(month,Year,Pack) %>%
filter(Year %in% c(2006,2007)) %>%
mutate(date = paste0(month,'/',Year)) %>%
mutate(date = zoo::as.yearmon(date, "%m/%Y")) %>%
filter(date > 'sep 2006', date < 'avr 2007') %>%
filter(!is.na(Pack)) %>%
pull(Pack) %>%
unique()

```













```

# y coord
Sy <- out[, (M+5) :(2*M+5-1) ]
dim(Sy)
# ind variable: z is 1 if ind belongs to the pop, 0 otherwise
z <- out[, (2*M+5):(3*M+5-1) ]
dim(z)

# get rid of individuals that are not in the pop
Sx[z==0]<-NA
Sy[z==0]<-NA

# array of estimated activity centers coordinates
AC <- array(NA,c(nrow(Sx),ncol(Sx),2))
AC[,,1] <- Sx
AC[,,2] <- Sy

# list of everything
obj <- list(Sx=Sx,Sy=Sy,z=z)

# individuals that do belong to the pop
mask = which(apply(z,2,mean)==1) # I keep only the individuals with z = 1 for all iterations
mask # It have the labels of the individuals we keep here

# estimated activity centers
plot(S)
plot(sg_study_area, add=TRUE)
points(centers, col="blue", pch=19,cex=0.7)

matrix_AC = NULL
for (i in mask) {
    current_AC <- AC[, i, ]
    average_AC <- apply(current_AC,2,mean)*scale_par
    matrix_AC <- rbind(matrix_AC,average_AC)
    points(average_AC[1],average_AC[2], pch=19,col='red')
    name = paste("Ind", i, sep="")
    text(average_AC[1],average_AC[2], name,cex = 0.8)
}

distance<-c(dist(matrix_AC, method="euclidean"),replace=TRUE)
print(mean(distance))
obs_75qt <- quantile(distance,probs=75/100);print(obs_75qt)

load("out_sim_11groups_12092019.RData")
# retrieve activity centers for each individual and iteration
ACA <- simplify2array(out$sims.list[c("SX","SY")])*scale_par # an array (descaled)
ACA <- ACA[,1:nrow(histoires2),] # retain only observed individuals
nsample <- dim(ACA)[1]

h_wolf=obs_75qt
## pack.composition will save the details of identified packs at each iteration
pack.composition <- vector("list", nsample)
for (i in 1:nsample){ # loop over MCMC samples
    clust <- hclust(dist(ACA[i,,])^2 , method = "ward.D") # clustering for current sample
    contar<- cutree(clust,h = (h_wolf)^2) # application of cutting distance
    current_nb_cluster <- length(unique(contar)) # nb of packs identified in current MCMC sample
    individuos.clus<- lapply(1:current_nb_cluster,function(eso) which(contar==eso)) # packs
composition
    pack.composition[[i]] <- individuos.clus ## current pack composition saved
}

# estimated number of packs
pack.number <- as.numeric(lapply(pack.composition,length))
print(mean(pack.number)); print(min(pack.number)); print(max(pack.number))#
print(sd(pack.number))
qt.pack.number <- as.numeric(quantile(pack.number,probs=c(2.5/100,50/100,97.5/100)))
print(qt.pack.number)

# list of pack sizes per iteration (note: the biggest pack over the whole sample has 27
# individuals)
pack.size <- lapply(pack.composition, function(cluster) as.numeric(lapply(cluster,length)))
# mean pack size per iteration
mean.pack.size <- as.numeric(lapply(pack.size,mean)); print(mean(mean.pack.size));
print(sd(mean.pack.size))
min.pack.size <- as.numeric(lapply(pack.size,min)); print(mean(min.pack.size))
max.pack.size <- as.numeric(lapply(pack.size,max)); print(mean(max.pack.size))
grupos_kmean <- kmeans(current_AC, mean(pack.number)); print(grupos_kmean)

```

```
#positions of packs
contar_kmean <- grupos_kmean$cluster; print(contar_kmean)
print(table(contar_kmean))# sizes of groups
print(mean(dist(grupos_kmean$centers, method="euclidean")))# mean distance of estimated packs
return(mean(pack.number))
}

set.seed(50)
nb_sim=100
r <- rpois(nb_sim,11)
allres<-matrix(NA, nrow=nb_sim, ncol=2)
colnames(allres)=c('r','mean.pack.number')
allres[,1] <- r
for(i in 51:length(allres[,1])){
  allres[i,2]=sim_CRgroups2(allres[i,1])
}
allres
save.image(file = "out_end_sim_11groups_12092019.RData")
```



**Annexe B: A propos du Chapitre 3**



```

Ft1 <- (1 + dat$R2[dat$year==(t_start)]^aa)^(1/aa)
Ft2 <- (1 + dat$R2[dat$year==(t_start-1)]^aa)^(1/aa)
# RP: modified to pi starting at 3 and 1-pi starting at 4
rho <- dat$rho[1]
dat$Npredict[dat$year==t_start] <- qq * rho * Ft1 * (pp*(1-
pi)*dat$Npredict[dat$year==(t_start-4)] +
pi*dat$Npredict[dat$year==(t_start-3)]) + pp * Ft1/Ft2 *
dat$Npredict[dat$year==(t_start-1)]
# then subsequent years
for (t in (t_start+1):2016){
  Ft1 <- (1 + dat$R2[dat$year==(t)]^aa)^(1/aa)
  Ft2 <- (1 + dat$R2[dat$year==(t-1)]^aa)^(1/aa)
  dat$Npredict[dat$year==t] <- qq * rho * Ft1 * (pp*(1-
pi)*dat$Npredict[dat$year==(t-4)] + pi*dat$Npredict[dat$year==(t-3)]) + pp
  * Ft1/Ft2 * dat$Npredict[dat$year==(t-1)]
}
sum((log(dat$N[dat$year>=t_start])- log(dat$Npredict[dat$year>=t_start]))^2)/(sigma*sigma) + 2 * log(sigma)
}

# initial values
theta_init <- c(runif(1,0,1),runif(1,-5,0),runif(1,0,1)) # qq, a, sigma
#[1] 0.5775979 -2.7058086 0.8354707

# evaluate deviance
dev_seals(theta_init,dat)
#[1] 255.7823

# minimization
tmpmin <- optim(par = theta_init, fn = dev_seals, gr = NULL, dat = dat,
hessian=TRUE, method="BFGS", control=list(trace=1, REPORT=1, maxit =
10000))

# get parameter estimates
(qq <- 1/(1+exp(-tmpmin$par[1]))) # 0.5336822
(aa <- -exp(tmpmin$par[2])) # -2.076718
(sigma <- exp(tmpmin$par[3])) # 1.620012

# standard errors on the link scale
SE_1 <- sqrt(diag(solve(tmpmin$hessian)))
# 2.2247467 3.9986898 0.5000032

# confidence intervals
plogis(c(tmpmin$par[1] - 1.96 * SE_1[1],tmpmin$par[1] + 1.96 * SE_1[1]))
# 0.01440642 0.98896336
-exp(c(tmpmin$par[2] - 1.96 * SE_1[2],tmpmin$par[2] + 1.96 * SE_1[2]))
# -8.196419e-04 -5.261760e+03
exp(c(tmpmin$par[3] - 1.96 * SE_1[3],tmpmin$par[3] + 1.96 * SE_1[3]))
# 0.6080047 4.3164791

# deviance is
tmpmin$value # 1.964853

# AIC is
tmpmin$value + 2*3 # 7.964853

# now predict pups abundance using estimated parameters
t_start <- 1978
dat$Npredict <- dat$N
pp <- 0.76
pi <- 0.3

```

```

rho <- dat$rho[1] #0.5062727
a <- -exp(tmpmin$par[2]) # -2.076718
Ft1 <- (1 + dat$R2[dat$year==(t_start)]^a)^(1/a)
Ft2 <- (1 + dat$R2[dat$year==(t_start-1)]^a)^(1/a)
dat$Npredict[dat$year==t_start] <- qq * rho * Ft1 * (pp*(1-
pi)*dat$Npredict[dat$year==(t_start-4)] +
pi*dat$Npredict[dat$year==(t_start-3)]) + pp * Ft1/Ft2 *
dat$Npredict[dat$year==(t_start-1)]
# then subsequent years
for (t in (t_start+1):2016){
  Ft1 <- (1 + dat$R2[dat$year==(t)]^aa)^(1/aa)
  Ft2 <- (1 + dat$R2[dat$year==(t-1)]^aa)^(1/aa)
  dat$Npredict[dat$year==t] <- qq * rho * Ft1 * (pp*(1-
pi)*dat$Npredict[dat$year==(t-4)] + pi*dat$Npredict[dat$year==(t-3)]) + pp
  * Ft1/Ft2 * dat$Npredict[dat$year==(t-1)]
}
dat
mean(dat$Npredict)
#[1] 184.1097

# compare predictions to observed values
plot(1973:2016,dat$N,xlab='year',ylab='Number of pups',ylim=c(50,800))
points(1973:2016,dat$Npredict,col='red')
legend(1990,800,c('Counts','Predicted values'), pch = c(1,1),
col=c('black','red'))

```





```

# model integrating capture-recapture data and pup counts
# for elephant seals in R with a maximum likelihood approach
# october 2018
# Authors: Gimenez O., Oosthuizen C., Pradel R., Mansilla L.
#-----#
#----- DEVIANCE OF THE INTEGRATED MODEL -----#
#-----#
dev_integrated_model <- function(b,dat,data,eff,e,garb,nh,kml) {

#----- 1. PARAMETERS
# b[1:14] =
[phiPB0,phiPB1,phiPB2,phiPB3,phiPB4p,phi,psiPB_B2,psiPB_B3,psiPB_B4,psiB_B,
psiNB_B,p1_PB,p2_B,p1_NB]
# b[15] = a in F = (1+R2^a)^(1/a)
# b[16] = sigma observation error on counts

#----- 2. DATA INPUTS
# dat is a data.frame containing N = number of pups, R2 = social structure.
# variable used to calculate fertility function.
# data contains the encounter histories.
# eff counts.
# e vector of dates of first captures.
# garb vector of initial states.
# kml nb of recapture occasions (nb of capture occ - 1).
# nh nb ind.

#----- 3. CAPTURE-RECAPTURE LIKELIHOOD
# OBSERVATIONS (+1)
# 0 : not seen
# 1 : seen outside of breeding
# 2 : seen breeding

# STATES
# PB : pre-breeder
# B : breeder
# NB : non-breeder
# D : dead

# logit link for all parameters
lb <- 1/(1+exp(-b))
phiPB0 <- lb[1]
phiPB1 <- lb[2]
phiPB2 <- lb[3]
phiPB3 <- lb[4]
phiPB4p <- lb[5]
phi <- lb[6]
psiPB_B2 <- lb[7]
psiPB_B3 <- lb[8]
psiPB_B4 <- lb[9]
psiB_B <- lb[10]
psiNB_B <- lb[11]
p1_PB <- lb[12]
p2_B <- lb[13]
p1_NB <- lb[14]

##-- prob of obs (rows) cond on states (col)

# capture
B <- matrix(c(
1-p1_PB,p1_PB,0,
1-p2_B,0,p2_B,

```

```

1-p1_NB,p1_NB,0,
1,0,0),nrow=4,ncol=3,byrow=T)
B <- t(B)

## first encounter

BE <- matrix(c(
0,1,0,
0,0,1,
0,1,0,
1,0,0),nrow=4,ncol=3,byrow=T)
BE <- t(BE)

## prob of states at t+1 given states at t

# survival
A1_age0 <- matrix(c(
phiPB0,0,0,1-phiPB0,
0,0,0,1,
0,0,0,1,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age1 <- matrix(c(
phiPB1,0,0,1-phiPB1,
0,0,0,1,
0,0,0,1,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age2 <- matrix(c(
phiPB2,0,0,1-phiPB2,
0,0,0,1,
0,0,0,1,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age3 <- matrix(c(
phiPB3,0,0,1-phiPB3,
0,phi,0,1-phi,
0,0,phi,1-phi,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age4p <- matrix(c(
phiPB4p,0,0,1-phiPB4p,
0,phi,0,1-phi,
0,0,phi,1-phi,
0,0,0,1),nrow=4,ncol=4,byrow=T)

# breeding
A2_age0 <- matrix(c(
1,0,0,0,
0,0,1,0,
0,0,1,0,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A2_age1 <- matrix(c(
1,0,0,0,
0,0,1,0,
0,0,1,0,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A2_age2 <- matrix(c(
1-psiPB_B2,psiPB_B2,0,0,

```

```

0,0,1,0,
0,0,1,0,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A2_age3 <- matrix(c(
1-psiPB_B3,psiPB_B3,0,0,
0,psiB_B,1-psiB_B,0,
0,0,1,0,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A2_age4p <- matrix(c(
1-psiPB_B4,psiPB_B4,0,0,
0,psiB_B,1-psiB_B,0,
0,psiNB_B,1-psiNB_B,0,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A_age0 <- A1_age0 %*% A2_age0
A_age1 <- A1_age1 %*% A2_age1
A_age2 <- A1_age2 %*% A2_age2
A_age3 <- A1_age3 %*% A2_age3
A_age4p <- A1_age4p %*% A2_age4p

# init states
PI <- c(1,0,0,0)

# likelihood
l <- 0
for (i in 1:nh) # loop on ind
{
  ei <- e[i] # date of first det
  oe <- garb[i] + 1 # init obs
  evennt <- data[,i] + 1 # add 1 to obs to avoid 0s in indexing
  ALPHA <- PI*BE[oe,]
  for (j in (ei+1):(km1+1)) # cond on first capture
  {
    if (j == (ei+1)) ALPHA <- (ALPHA %*% A_age0)*B[evennt[j],]
    if (j == (ei+2)) ALPHA <- (ALPHA %*% A_age1)*B[evennt[j],]
    if (j == (ei+3)) ALPHA <- (ALPHA %*% A_age2)*B[evennt[j],]
    if (j == (ei+4)) ALPHA <- (ALPHA %*% A_age3)*B[evennt[j],]
    if (j > (ei+4)) ALPHA <- (ALPHA %*% A_age4p)*B[evennt[j],]
  }
  l <- l + log(sum(ALPHA))*eff[i]
}
dev_capturerecapture <- -2*l

#----- 4. PUPS COUNTS LIKELIHOOD
# build the deviance function to minimize, see Eq 7 in Ferrari's paper
## we use data from 1973 to 1977 to initialize
# we predict pop size from 1978-2016

# get parameters
t_start <- 1978
aa <- -exp(b[15]) # a in F(t)
sigma <- exp(b[16]) # observation error
pp <- phi
rho <- dat$rho[1]
qq <- 2 * rho * phiPB0 * (phiPB1 * phiPB2) # alpha is what Ferrari calls
# fertility constant basically litter size, 1 in our case
dat$Npredict <- dat$N
Ft1 <- (1 + dat$R2[dat$year==(t_start)]^aa)^(1/aa)
Ft2 <- (1 + dat$R2[dat$year==(t_start-1)]^aa)^(1/aa)

```

```

dat$Npredict[dat$year==t_start] <- qq * Ft1 * (phiPB3 * (1-psiPB_B2) *
dat$Npredict[dat$year==(t_start-4)] + psiPB_B2 *
dat$Npredict[dat$year==(t_start-3)]) + pp * Ft1/Ft2 *
dat$Npredict[dat$year==(t_start-1)]
for (t in (t_start+1):2016){
  Ft1 <- (1 + dat$R2[dat$year==(t)]^aa)^(1/aa)
  Ft2 <- (1 + dat$R2[dat$year==(t-1)]^aa)^(1/aa)
  dat$Npredict[dat$year==t] <- qq * Ft1 * (phiPB3 * (1-psiPB_B2) *
dat$Npredict[dat$year==(t-4)] + psiPB_B2 * dat$Npredict[dat$year==(t-3)]) +
pp * Ft1/Ft2 * dat$Npredict[dat$year==(t-1)]
}
dev_pupcounts <- sum((log(dat$N[dat$year>=t_start])-log(dat$Npredict[dat$year>=t_start]))^2)/(sigma*sigma) + 2 * log(sigma)
dev <- dev_capturerecapture + dev_pupcounts
dev
}

#-----#
#----- MODEL FITTING / PARAMETER ESTIMATION -----#
#-----#
# read in capture-recapture data
data <- read.table('C:/Users/mansilla/Desktop/Des elephantes de la
mer/ipm/ES EHM ESURGE headed format.txt')
head(data)
data <- R2ucare:::group_data(data[,1:(ncol(data)-1)],rep(1,nrow(data)))
eff <- data[,ncol(data)]
data <- data[,1:(ncol(data)-1)]
# define various quantities
nh <- dim(data)[1]
k <- dim(data)[2]
km1 <- k-1
# compute the date of first capture fc, and state at initial capture
init.state
fc <- NULL
init.state <- NULL
for (i in 1:nh){
temp <- 1:k
fc <- c(fc,min(which(data[i,]!=0)))
init.state <- c(init.state,data[i,fc[i]])
}
# transpose data
data <- t(data)

# read in pups counts data
nb_females_all <- read.table('C:/Users/mansilla/Desktop/Des elephantes de
la mer/ipm/nb_females_all_MSA.txt',header=T)
# Npups
nb_pups <- read.table('C:/Users/mansilla/Desktop/Des elephantes de la
mer/ipm/nb_pups_MSA.txt',header=T)
N <- nb_pups$number_pups
# Nb of females per males
nb_females_per_males <-
nb_females_all$nb_females[nb_females_all$component=='females_per_adult_male
']
# Adult sex ratio (males per female)
adult_sexratio <- 1 / nb_females_per_males
# Harem size
harem_size <-
nb_females_all$nb_females[nb_females_all$component=='harem_size']
# R2 variable
R2 <- harem_size * adult_sexratio

```

```

# birth sex ratio
bsr <- read.table('C:/Users/mansilla/Desktop/Des elephantes de la
mer/ipm/birth_sexratio_MSA.txt',header=T)
bsr <- bsr / 2
# altogether
dat <- data.frame(year=1973:2016,N,R2,rho=mean(as.matrix(bsr)))

# init values
binit <- c(rep(0.5,14), -1, 1.5)

# evaluate the integrated pop model deviance at the initial values, just to
check
dev_integrated_model(binit,dat,data,eff,fc,init.state,nh,kml)

# fit model
deb=Sys.time()
tmpmin <-
optim(binit,dev_integrated_model,NULL,hessian=TRUE,dat,data,eff,fc,init.state,
nh,kml,method="BFGS",control=list(trace=1, REPORT=1, maxit = 10000))
fin=Sys.time()
fin-deb

# get estimates and back-transform
b <- tmpmin$par

data.frame(param=c('phiPB0',
'phiPB1',
'phiPB2',
'phiPB3',
'phiPB4p',
'phi',
'psiPB_B2',
'psiPB_B3',
'psiPB_B4',
'psiB_B',
'psiNB_B',
'p1_PB',
'p2_B',
'p1_NB',
'a',
'sigma'), integrated =
c(1/(1+exp(-b[1:14])), -exp(b[15]), exp(b[16])))

#      param integrated
1    phiPB0  0.5984056
2    phiPB1  0.7647213
3    phiPB2  0.7806313
4    phiPB3  0.7910896
5    phiPB4p 0.7113818
6      phi  0.7611661
7  psiPB_B2  0.3077905
8  psiPB_B3  0.6751706
9  psiPB_B4  0.5641837
10   psiB_B  0.8181329
11  psiNB_B  0.7270977
12    p1_PB  0.7764373
13    p2_B   0.9090268
14    p1_NB  0.8731530
15      a   -0.9874015
16    sigma  1.6316564

```

```

# deviance is
tmpmin$value # 43166.99

# AIC is
tmpmin$value + 2*16 # 43198.99

#-----#
#----- PUPS COUNTS PREDICTION -----#
#-----#

t_start <- 1978

# now predict pups abundance using estimated parameters
pp <- 1/(1+exp(-b[6]))
a <- -exp(b[15])
phiPB0 <- 1/(1+exp(-b[1]))
phiPB1 <- 1/(1+exp(-b[2]))
phiPB2 <- 1/(1+exp(-b[3]))
phiPB3 <- 1/(1+exp(-b[4]))
psiPB_B2 <- 1/(1+exp(-b[7]))

qq <- 2 * mean(as.matrix(bsr)) * phiPB0 * (phiPB1 * phiPB2)

dat$Npredict <- dat$N
Ft1 <- (1 + dat$R2[dat$year==(t_start)]^a)^(1/a)
Ft2 <- (1 + dat$R2[dat$year==(t_start-1)]^a)^(1/a)
dat$Npredict[dat$year==t_start] <- qq * Ft1 * (phiPB3 * (1-psiPB_B2) *
dat$Npredict[dat$year==(t_start-4)] + psiPB_B2 *
dat$Npredict[dat$year==(t_start-3)]) + pp * Ft1/Ft2 *
dat$Npredict[dat$year==(t_start-1)]
for (t in (t_start+1):2016) {
  Ft1 <- (1 + dat$R2[dat$year==(t)]^a)^(1/a)
  Ft2 <- (1 + dat$R2[dat$year==(t-1)]^a)^(1/a)
  dat$Npredict[dat$year==t] <- qq * Ft1 * (phiPB3 * (1-psiPB_B2) *
dat$Npredict[dat$year==(t-4)] + psiPB_B2 * dat$Npredict[dat$year==(t-3)]) +
pp * Ft1/Ft2 * dat$Npredict[dat$year==(t-1)]
}
dat
Npredict_point <- dat$Npredict

# compare predictions to observed values
plot(1973:2016,dat$N,xlab='year',ylab='Number of pups',ylim=c(50,800))
points(1973:2016,dat$Npredict,col='red')
legend(1990,800,c('Counts','Predicted values'), pch = c(1,1),
col=c('black','red'))

#-----#
#----- PUPS COUNTS PREDICTION -----#
#----- WITH CONF INTERVALS -----#
#-----#

# get confidence intervals
SE_link <- sqrt(diag(solve(tmpmin$hessian)))

# generate bootstrap values in conf intervals (use of beta distributions
would be more elegant)
nbboot <- 250
pseudo_p <- 1/(1+exp(-rnorm(nbboot,b[6],SE_link[6])))
pseudo_a <- -exp(rnorm(nbboot,b[15],SE_link[15]))
pseudo_phiPB0 <- 1/(1+exp(-rnorm(nbboot,b[1],SE_link[1])))
pseudo_phiPB1 <- 1/(1+exp(-rnorm(nbboot,b[2],SE_link[2])))

```

```

pseudo_phiPB2 <- 1/(1+exp(-rnorm(nbboot,b[3],SE_link[3])))
pseudo_phiPB3 <- 1/(1+exp(-rnorm(nbboot,b[4],SE_link[4])))
pseudo_psiPB2 <- 1/(1+exp(-rnorm(nbboot,b[7],SE_link[7])))

res <- NULL
for (i in 1:nbboot){
# predict pups abundance using estimated parameters
t_start <- 1978
qq <- 2 * mean(as.matrix(bsr)) * pseudo_phiPB0[i] * (pseudo_phiPB1[i] *
pseudo_phiPB2[i])
dat$Npredict <- dat$N
Ft1 <- (1 + dat$R2[dat$year==(t_start)]^pseudo_a[i])^(1/pseudo_a[i])
Ft2 <- (1 + dat$R2[dat$year==(t_start-1)]^pseudo_a[i])^(1/pseudo_a[i])
dat$Npredict[dat$year==t_start] <- qq * Ft1 * (pseudo_phiPB3[i] * (1-
pseudo_psiPB2[i]) *
dat$Npredict[dat$year==(t_start-4)] + pseudo_psiPB2[i] *
dat$Npredict[dat$year==(t_start-3)] +
pseudo_p[i] * Ft1/Ft2 * dat$Npredict[dat$year==(t_start-1)]
for (t in (t_start+1):2016){
    Ft1 <- (1 + dat$R2[dat$year==(t)]^pseudo_a[i])^(1/pseudo_a[i])
    Ft2 <- (1 + dat$R2[dat$year==(t-1)]^pseudo_a[i])^(1/pseudo_a[i])
dat$Npredict[dat$year==t] <- qq * Ft1 * (pseudo_phiPB3[i] * (1-
pseudo_psiPB2[i]) *
dat$Npredict[dat$year==(t-4)] + pseudo_psiPB2[i] *
dat$Npredict[dat$year==(t-3)] +
pseudo_p[i] * Ft1/Ft2 * dat$Npredict[dat$year==(t-1)]
}
res <- cbind(res,dat$Npredict)
}

dim(res)
L <- apply(res,1,quantile,probs=2.5/100) # lower bound bootstrap conf
interval on Npredict
U <- apply(res,1,quantile,probs=97.5/100) # upper bound bootstrap conf
interval on Npredict

# compare predictions to observed values
plotrix:::plotCI(1973:2016, Npredict_point, ui=U,
li=L,col='red',ylim=c(50,500))
points(1973:2016,dat$N,xlab='year',ylab='Number of pups')
legend(1990,400,c('Counts','Predicted values'), pch = c(1,1),
col=c('black','red'))

```

```

# model integrating capture-recapture data and pup counts
# for elephant seals in R with a maximum likelihood approach
# (Fixed social structure parameter "a")
# october 2018
# Authors: Gimenez O., Oosthuizen C., Pradel R., Mansilla L.
#-----#
#----- DEVIANCE OF THE INTEGRATED MODEL -----#
#-----#
dev_integrated_model <- function(b,dat,data,eff,e,garb,nh,kml) {

#----- 1. PARAMETERS
# b[1:14] = [phiPB0,phiPB1,phiPB2,phiPB3,phiPB4p,phi,psiPB_B2,
#             psiPB_B3,psiPB_B4,psiB_B,psiNB_B,p1_PB,p2_B,p1_NB]
# b[15] = sigma observation error on counts

#----- 2. DATA INPUTS
# dat is a data.frame containing N = number of pups, R2 = social
# structure
# variable used to calculate fertility function
# data contains the encounter histories
# eff counts
# e vector of dates of first captures
# garb vector of initial states
# kml nb of recapture occasions (nb of capture occ - 1)
# nh nb ind

#----- 3. CAPTURE-RECAPTURE LIKELIHOOD
# OBSERVATIONS (+1)
# 0 : not seen
# 1 : seen outside of breeding
# 2 : seen breeding

# STATES
# PB : pre-breeder
# B : breeder
# NB : non-breeder
# D : dead

# logit link for all parameters
lb <- 1/(1+exp(-b))
phiPB0 <- lb[1]
phiPB1 <- lb[2]
phiPB2 <- lb[3]
phiPB3 <- lb[4]
phiPB4p <- lb[5]
phi <- lb[6]
psiPB_B2 <- lb[7]
psiPB_B3 <- lb[8]
psiPB_B4 <- lb[9]
psiB_B <- lb[10]
psiNB_B <- lb[11]
p1_PB <- lb[12]
p2_B <- lb[13]
p1_NB <- lb[14]

## prob of obs (rows) cond on states (col)
# capture
B <- matrix(c(
1-p1_PB,p1_PB,0,

```

```

1-p2_B,0,p2_B,
1-p1_NB,p1_NB,0,
1,0,0),nrow=4,ncol=3,byrow=T)
B <- t(B)

#-- first encounter
BE <- matrix(c(
0,1,0,
0,0,1,
0,1,0,
1,0,0),nrow=4,ncol=3,byrow=T)
BE <- t(BE)

#-- prob of states at t+1 given states at t
# survival
A1_age0 <- matrix(c(
phiPB0,0,0,1-phiPB0,
0,0,0,1,
0,0,0,1,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age1 <- matrix(c(
phiPB1,0,0,1-phiPB1,
0,0,0,1,
0,0,0,1,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age2 <- matrix(c(
phiPB2,0,0,1-phiPB2,
0,0,0,1,
0,0,0,1,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age3 <- matrix(c(
phiPB3,0,0,1-phiPB3,
0,phi,0,1-phi,
0,0,phi,1-phi,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A1_age4p <- matrix(c(
phiPB4p,0,0,1-phiPB4p,
0,phi,0,1-phi,
0,0,phi,1-phi,
0,0,0,1),nrow=4,ncol=4,byrow=T)

# breeding
A2_age0 <- matrix(c(
1,0,0,0,
0,0,1,0,
0,0,1,0,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A2_age1 <- matrix(c(
1,0,0,0,
0,0,1,0,
0,0,1,0,
0,0,0,1),nrow=4,ncol=4,byrow=T)

A2_age2 <- matrix(c(
1-psiPB_B2,psiPB_B2,0,0,
0,0,1,0,

```

## Appendix: Details on the multievent capture-recapture model

### ***I. Notation***

- $\phi_{a0}$  survival probability of newborn females  
 $\phi_{a1}$  survival probability of 1-year old females  
 $\phi_{a2}$  survival probability of 2-year old females  
 $\phi_{a3}$  survival probability of 3-year old females  
 $\phi_{a4}$  survival probability of 4-year old females  
 $\phi$  survival probability of adult females  
 $\psi_{B2}$  probability of breeding in the next year for 2-year old females  
 $\psi_{B3}$  probability of breeding in the next year for 3-year old females  
 $\psi_{B4}$  probability of breeding in the next year for 4-year old females  
 $\psi_B$  probability of a reproductive female to remain a reproductive female  
 $\psi_{NB}$  probability of a non-breeding female to remain a non-breeding female  
 $p_{PB}$  probability of observing a pre-breeding female outside the breeding season  
 $p_B$  probability of observing a breeding female in the breeding season  
 $p_{NB}$  probability of observing a non-breeding female outside of the breeding season  
 $a$  Parameter that determines the strength of the influence of the social structure on the function  $F$ ,  
 $\sigma$  standard deviation of the distribution for sampling error  
 $\pi$  is the proportion of adult females population which breed at 3 years old  
 $1 - \pi$  proportion of female which are breeders at 4 years of age,  
 $r$  is the recruitment rate of adult females,  $\rho$  is a constant corresponding to the mean sex ratio (proportion of females) at birth  
 $q$  is a composite recruitment parameter ( $q = \alpha pr$ ).  
We defined four states: PB for Pre-Breeder, B for Breeder, NB for Non-Breeder and D for Dead. We also had three types of observations: 0 for not seen, 1 for seen outside of breeding and 2 for seen breeding.

***II. Matrix of observations or events.***

		States			
Obs		PB	B	NB	D
Not seen	0	$1-p_{PB}$	$p_B$	$p_{NB}$	1
Seen outside breeding	1	$p_{PB}$	0	$p_{NB}$	0
Seen breeding	2	0	$p_B$	0	0

Table 1: The matrix B of observations is defined given the states.

		States			
Obs		PB	B	NB	D
Not seen	0	0	0	0	1
Seen outside breeding	1	1	0	1	0
Seen breeding	2	0	1	0	0

Table 2: We fix the matrix B of probability of the first encounter.

### ***III. Matrix of states***

#### ***IIIa. Survival***

States $t = 0$	States $t = 1$			
	PB	B	NB	D
PB	$\phi_{a0}$	0	0	$1 - \phi_{a0}$
B	0	0	0	1
NB	0	0	0	1
D	0	0	0	1

Table 3: Matrix of survival of individual of newborn to 1 year old.

States $t = 1$	States $t = 2$			
	PB	B	NB	D
PB	$\phi_{a1}$	0	0	$1 - \phi_{a1}$
B	0	0	0	1
NB	0	0	0	1
D	0	0	0	1

Table 4: Matrix of survival of individual of 1 to 2 years old.

States $t = 2$	States $t = 3$			
	PB	B	NB	D
PB	$\phi_{a2}$	0	0	$1 - \phi_{a2}$
B	0	0	0	1
NB	0	0	0	1
D	0	0	0	1

Table 5: Matrix of survival of individual of 2 to 3 years old.

States $t = 3$	States $t = 4$			
	PB	B	NB	D
PB	$\phi_{a3}$	0	0	$1 - \phi_{a3}$
B	0	$\phi$	0	$1 - \phi$
NB	0	0	$\phi$	$1 - \phi$
D	0	0	0	1

Table 6: Matrix of survival of individual of 3 to 4 years old.

States $t = 4$	States $t = +4$			
	PB	B	NB	D
PB	$\phi_{a4}$	0	0	$1 - \phi_{a4}$
B	0	$\phi$	0	$1 - \phi$
NB	0	0	$\phi$	$1 - \phi$
D	0	0	0	1

Table 7: Matrix of survival of individual of 4 to more years old.

