UNIVERSITAT DE VALÈNCIA

Programa de Doctorat en Estadística i Optimització

New insights in Bayesian survival analysis in Ecology

by

Blanca Sarzo Carles

PhD Thesis
in Statistics and Optimization

Supervised by
David Valentín Conesa Guillén,
Carmen Armero Cervera
and Jonas Hentati Sundberg

in the
Faculty of Mathematics
Department of Statistics and Operations Research

March 2020

This thesis has been supported by grant BES-2014-070766 supported by MTM2013-42323-P from the Spanish Ministry of Economy and Competitiveness. The last part of this research has been carried out during the two visits to Dr. Ruth King at the University of Edinburgh.

Acknowledgements

Cuando empecé la tesis nunca me hubiera imaginado lo que iban a implicar estos cinco años, y ya no solo a nivel de formación, sino también (y principalmente) a nivel personal. La tesis es un aprendizaje de vida, un proceso en muchas ocasiones muy muy duro que, o realmente te apasiona, o es muy fácil pensar que no vale la pena tanto esfuerzo. En mi caso pasé de estar todo el día trabajando en el campo a estar sentada delante de un ordenador. El cambio fue muy difícil la verdad, pero ver que todos los 'números', verosimilitudes y probabilidades que escribía iban a servir para la conservación de las aves marinas era lo que me motivaba (y la verdad me picaba) a seguir. Cuando algo te gusta es muy peligroso...

Aún así ha habido momentos muy difíciles a lo largo de estos años, y si no hubiera sido por la gente que me ha apoyado en este proceso, a veces sin ni siquiera entenderlo del todo, tengo muy claro que no estaría aquí. Así que con esto empiezo mi larga lista de agradecimientos.

Primero de todo, a los miembros de la Sociedad Española de Biometría (SEB) y de la red Bioestatnet, tanto seniors como estudiantes. Desde el principio me acogieron, me hicieron sentir que encajaba allí y que formaba parte de aquello, y la verdad, eso es algo que les agradeceré siempre. Por supuesto, gracias por esas magníficas Jornadas de Estudiantes, gracias a ellas he podido conocer a gente especial como Paco, Leire, Amaia, Lore, Irantzu, Natalia (vale, estas dos últimas ya son mayores, pero...).

Of course I have to mention my supervisors who have taught me several important lessons along this process and who have helped me to perform this thesis. You have instructed me how to be a reasearcher and you have motivated me to improve at each step of this thesis. Without you I would not be here, thank you.

Also, thank you for the oportunity to be in Stora Karlsö, an amazing island and experience, and which gave me the opportunity to meet Aron, a great naturalist, photograph, musician and a great person.

Y cómo no, tengo que hacer especial mención a mis compañeros de tesis. Lo bueno de las cosas duras es que creas vínculos especiales, ese es el caso de Joaquín y Elena, mis dos grandes apoyos a lo largo de esta tesis. Gracias de corazón, gracias a vosotros los días eran menos difíciles, gracias por vuestras bromas, y por compartir conmigo vuestras preocupaciones y alegrías, ha sido un placer trabajar con vosotros, y espero que ésto no acabe aquí. Por supuesto, también he de mencionar a Rufo, Miguel, Paula, Consu, Danilo, Jessica, Gabi, Irene y Sabela, aunque con unos he coincidido más que con otros, ha sido un placer estar con vosotros.

To Ruth, thanks for being so kind to me, for remembering me why I started the thesis, for teaching me kindly and always smiling. Thanks for answering all my questions (even when it was not the first time asked) and without complaints. But specially for the way you combine work and humor, I learned a lot with you but more importantly, I also enjoyed a lot... A pleasure to know you and to work with you.

Por supuesto a María y Ainoha, que creo todavía no entienden en lo que estoy trabajando, pero siempre me han escuchado, me han apoyado y han intentado entender todo lo que les contaba. Sois dos grandes amigas y dos grandes personas. Un placer formar parte de vuestra vida.

A mi familia cómo no, a mis padres, Concha y Emilio, gracias a ellos estoy donde estoy. Me habéis apoyado siempre, en todas las decisiones que he tomado, aunque no estuviérais de acuerdo. Gracias por darme la oportunidad de hacer lo que me gustaba, gracias por quererme tanto y gracias por respetarme tal y como soy. Por supuesto gracias por no dejar que pasara frío en Edimburgo, ni hambre, por poner mi felicidad muchas veces por encima de la vuestra. El trabajo de los padres no está pagado... Gracias. Por supuesto también agradecer a mi hermana Belén y a Alberto, por ayudarme cuando lo necesitaba, y a los 'enanos', Naiara y Xabier, aunque digáis que ya tengo edad para encontrar un trabajo. Estos jóvenes de hoy en día...

Y finalmente, quiero hacer mención especial a Mar. Hemos pasado mucho ya juntas, unas oposiciones y una tesis (entre otras cosas por el camino) creo que ya son demasiado, pero aquí estamos. Gracias por creer en mi, por pensar que podía ser mejor persona, por pensar que podía conseguir lo que quisiera. Gracias por tu paciencia, por tu comprensión, por entender que igual nos va a tocar viajar un poquito, y respetarlo porque sabes que me hace feliz. Gracias por querer que yo sea feliz, que crezca a nivel personal y profesional y por quererme. Espero que esta nueva etapa sea buena para las dos porque tengo muchas ganas de disfrutar la vida contigo.

En resumen, si me hubieran dicho a priori lo que iba a implicar una tesis no sé si me hubiera atrevido a empezarla, pero una vez hecha me he dado cuenta de que con trabajo e ilusión, casi todo se consigue, y ahora me siento muchísimo más fuerte que cuando empecé. Ese ha sido mi aprendizaje.

"Strength does not come from physical capacity.
It comes from indomitable will."

Mahatma Gandhi

"It is not the strongest or the most intelligent
species who will survive but those
who can best manage change."

Charles Darwin

A la Ape y a Oma.

UNIVERSITAT DE VALÈNCIA

Resumen

Facultad de Ciencias Matemáticas
Departamento de Estadística e Investigación Operativa

Programa de Doctorado en Estadística y Optimización

La fauna silvestre está asediada. Y ésta no es solo una frase impactante con la que empezar una tesis, tristemente, es una realidad. En el último siglo, muchas especies han disminuido drásticamente, mientras que otras afrontan su extinción debido, principalmente, a los rápidos cambios (y a gran escala) ocurridos tanto en hábitats como en ecosistemas. El cambio climático, las especies invasoras, la caza ilegal y la sobrepesca son sólo algunas de las principales amenazas que afectan a las poblaciones de animales silvestres en la actualidad. Para abordar este problema, se requiere de un compromiso a todos los niveles, desde las comunidades locales hasta los gobiernos, pasando por los expertos, la educación y la investigación. De hecho, la investigación es una herramienta fundamental para la conservación de las poblaciones silvestres. Entender los factores que afectan a las poblaciones nos permite mejorar su gestión así como su seguimiento, y por lo tanto, su conservación.

En las últimas décadas ha habido un gran aumento en la cantidad (y la variedad) de datos recogidos en los sistemas ecológicos, lo que ha conducido al desarrollo de modelos estadísticos más complejos. Esta complejidad hace que el proceso inferencial sea difícil de llevar a cabo. La perspectiva de la estadística Bayesiana aparece como una buena alternativa para realizar dicho proceso inferencial debido a los avances computacionales ocurridos en las últimas décadas. Además, permite incorporar de manera sencilla la información inicial (si existe y está diponible) así como tiene en cuenta la incertidumbre relativa tanto a los modelos como a los parámetros.

En este trabajo investigamos las probabilidades de supervivencia, de recaptura, de recuperación y las probabilidades de migración en el contexto de los modelos de captura-recaptura(-recuperación). Estos modelos tienen en cuenta la detección imperfecta de los individuos, algo muy habitual en los sistemas ecológicos. De hecho, si la detección imperfecta no se tiene en consideración en el modelo puede causar sesgos en las estimaciones de los parámetros.

Este trabajo se enmarca en un contexto real, en concreto, el estudio de una colonia de aves marina, el Arao común (*Uria aalge*). Las aves marinas se consideran centinelas del mar. Los cambios en sus poblaciones reflejan los cambios producidos en el mar, por lo que se consideran bioindicadoras de cambios medioambientales. Asimismo, el cálculo de las probabilidades de supervivencia juvenil en aves marinas es difícil debido a la ecología de las especies. Por todo ello, esta tesis no solo supone un reto a nivel estadístico sino también a nivel ecológico.

Con todo esto en mente, la estructura de la tesis es la siguiente. El Capítulo 1 ofrece la motivación así como una revisión general de los métodos de captura-recaptura(-recuperación) y los modelos estadísticos asociados. En este Capítulo presentamos en detalle los dos modelos utilizados: los modelos de Cormack-Jolly-Seber (CJS) y los modelos de marcaje-recaptura-recuperación (modelos integrados). Además, este capítulo ofrece una introducción a la estadística Bayesiana, así como una descripción de uno de los principales métodos para llevarla a cabo, los métodos Markov chain Monte Carlo (MCMC). Por último, concluimos este Capítulo con una breve explicación de los dos algoritmos de simulación básicos basados en Métodos Monte Carlo de cadenas de Markov más conocidos (Metropolis-Hastings y Gibbs sampler), junto con una relación de programas y páginas web para la implementación de una gran variedad de modelos de captura-recaptura(-recuperación).

El Capítulo 2 ofrece el contexto ecológico de esta tesis. Para ello, presentamos la especie objeto de estudio, el Arao común, así como la colonia de estudio, Stora Karlsö (Gotland, Suecia). En este capítulo destacamos algunas de sus principales características ecológicas, características que necesitaremos conocer para el desarrollo de modelos estadísticos biológicamente motivados. La parte principal de este Capítulo consiste en una descripción de las dos bases de datos que han motivado el desarrollo metodológico realizado en esta tesis: una base de datos de captura-recaptura y otra de marcaje-recaptura-recuperación.

En los Capítulos 3, 4 y 5 presentamos los estudios desarrollados a lo largo de esta tesis. En particular, el Capítulo 3 ofrece estimaciones fiables de supervivencia juvenil para el Arao común. La novedad de este trabajo radica en que los individuos inmaduros de esta especie (y en general, de las aves marinas) no pueden ser observados debido a que pasan largas temporadas en el mar. Sin embargo, esta colonia tiene una característica especial: una elevada proporción de individuos inmaduros son observados, lo que permite obtener dichas estimaciones fiables de supervivencia juvenil. Además, este trabajo representa una primera aproximación al problema del monitoreo parcial, el cual causa sesgos en las estimaciones de los parámetros de interés. Para ello, adoptamos una aproximación subjetiva

Bayesiana, de manera que incorporamos información previa acerca de las áreas donde el problema de monitoreo parcial está presente.

El Capítulo 4 incorpora una metodología comúnmente utilizada en los estudios médicos de supervivencia en el contexto de los modelos de captura-recaptura en ecología. En particular, mostramos cómo las historias de captura se presentan de manera diferente dependiendo de la escala temporal escogida. Además, en este Capítulo presentamos cómo el uso de la escala temporal alternativa presentada (la edad) permite una mejor interpretación de los parámetros del modelo cuando la edad es el principal interés en el estudio.

Finalmente, en el Capítulo 5 ofrecemos un contexto integrado de marcaje-recaptura-recuperación para estudios parcialmente monitorizados. La información suministrada por los datos de anillamiento-recuperación permite corregir los sesgos en las estimaciones de supervivencia obtenidos con datos (solo) de captura-recaptura debido al monitoreo parcial. Además, debido al gran tamaño de la base de datos, el modelo lo presentamos en formulación multinomial, de manera que ofrecemos la expresión explícita y eficiente de la verosimilitud así como los estadísticos suficientes asociados del modelo integrado propuesto. Tanto la corrección del problema de monitoreo parcial (muy extendido en especies coloniales), como la construcción de las m-arrays integradas y la eficiente verosimilitud suponen un paso adelante en este área, tanto desde un punto de vista aplicado como metodológico.

Esta tesis concluye con el Capítulo 6, donde se presentan las conclusiones y algunas líneas futuras de investigación, seguido de la bibliografía utilizada.

Contents

List of Figures

List of Tables

Introduction

Wildlife is under siege. And this is not only a fancy sentence to start a thesis, sadly, it is a fact. Over the last century, many wildlife species have seriously declined and many others face their extinction due to rapid and large-scale changes in habitats and ecosystems. Climate change, invasive species, illegal hunting and overfishing are only some of the main threats affecting animal populations nowadays. To address these concerns will require commitment at all levels, from local communities to governments, as well as experts, education and research. Indeed, research is a fundamental tool for conservation. Understanding the factors affecting wildlife populations allow us to improve the management of animal populations and therefore, their conservation.

The recently increase in the amount (and variety) of data collected on ecological systems has led to the development of more complex statistical models. This complexity has made the inferential process challenging to perform. The Bayesian approach arises as an alternative to address these issues due to the computational advances occurred in the last decades. Further, prior information (if available) can be easily incorporated as well as this approach takes fuller account of the uncertainties related to models and parameters.

In this work we investigate survival, recapture, recovery and migration probabilities in the context of capture-recapture(-recovery) models. These models account for imperfect detection, a common issue in ecological systems. Indeed, if imperfect detection is not taken into account, it may cause biases in estimated demographic parameters of interest.

The context of this work is a real one, in particular, a seabird species, the common guillemot (*Uria aalge*). Seabirds are sentinels of the sea. Their populations reflect conditions over large spatial and long term scales, making them bioindicators of environmental change. Further, the estimation of juvenile survival probabilities in seabirds is far from simple, mainly due to their ecological characteristics. Therefore, this thesis supposes not only a challenge from statistical perspective but also from ecological one.

With all this in mind, this thesis is structured as follows. Chapter 1 is devoted to provide the motivation along with an overview of the capture-recapture methods and associated statistical models. After that, we explain in detail the two models used: Cormack-Jolly-Seber (CJS) models and mark-recapture-recovery (MRR) models. Furthermore, we give an introduction of Bayesian inference, before describing the Markov chain Monte Carlo (MCMC) methods. Lastly, we conclude with a brief explanation of the two basic Markov chain simulation algorithms (Metropolis-Hastings and Gibbs sampler), as well as we provide useful software packages and web pages to implement a large variety of capture-recapture(-recovery) models.

Chapter 2 takes a brief look at the ecological context of this thesis. To do so, we introduce the species focus of this research, the common guillemot, as well as the study colony, Stora Karlsö (Gotland, Sweden). We briefly provide some of its main ecological characteristics that will be necessary to know in order to fit statistical models biologically motivated. The main part of this Chapter consists in a detailed description of the two data sets that motivated the methodological developments performed in this thesis, a capture-recapture and a mark-recapture-recovery database.

Chapters 3, 4 and 5 are dedicated to display the studies performed along this thesis. In particular, Chapter 3 is devoted to provide reliable juvenile survival estimates for common guillemots. The novelty of this work lies in the difficulty on the assessment of juvenile survival due to in this species (and in general, in most seabird species) young birds spend large periods at sea, remaining hence unobservable. However, the study colony has an special feature: a big proportion of immature birds are resighted allowing so to provide those reliable juvenile survival estimates. Further, this work represents a first approximation to the problem of partial monitoring that causes bias in parameter estimates. In particular, we adopt a subjective Bayesian approach so that we incorporate prior information corresponding to the areas where the partial monitoring is affecting.

Chapter 4 incorporates a methodology commonly performed in medical survival studies in the context of ecological capture-recapture framework. In particular, we show how differently the capture histories are presented depending on the selected temporal scale. Further, the use of the alternative temporal scale presented (the age) may allow a better interpretation of model parameters when age is the primary interest.

Finally, in Chapter 5 we provide an integrated mark-recapture-recovery framework for partially monitored studies. The information gathered by the ring-recovery data allows to correct the bias in survival estimates obtained with only capture-recapture data due to partial monitoring. Moreover, due to the (big) size of the database, we present it in multinomial formulation, so that we provide the explicit efficient likelihood expression along with the associated sufficient statistics of the integrated model proposed. Both the correction of partial monitoring problem (widespread in colonial species) and the construction of the integrated m-arrays along with the efficient likelihood suppose an step forward on this area, either from a practical or methodological perspective.

Chapter 6 provides some conclusions and future lines of research, and finally a generic bibliography used along this work is presented.

CHAPTER 1

SURVIVAL ANALYSIS IN ECOLOGY

1.1 Motivation

There are only few places left on the planet where the impact of human people has not been felt. Major threats to wildlife come from many different kinds of human activities such as habitat destruction, degradation or fragmentation, overexploitation, hunting, illegal wildlife trade, overfishing, invasive alien species and climate change.

The recent fires in the Amazon and Australia rightly drew attention to how fragile the most important ecosystems on the planet are. Half of the world's original forests are gone, and what remains is being cut down ten times faster than it can be replaced (Keenan et al., 2015). Animals which need large areas to live, are being squeezed in small areas, and one of the consequences is the increasing of human-wildlife conflicts. For example, in Kenya, elephants feed on crops on their migratory routes; in Nepal, tigers and leopards turn to livestock for an easy meal; in Spain, grizzly bears also feed on crops due to habitat destruction.

A recent United Nations scientific report documents the ‘unprecedented’ decline in global biodiversity that has alarming implications for human health, prosperity and long-term survival (Díaz et al., 2019). Indeed, the IUCN (International Union for Conservation of Nature) Red List of Threatened Species (`http://www.iucnredlist.org`) estimates that 28,000 species are at risk of extinction. In particular, 40% of amphibian species, almost 33% of reef corals, 25% of mammal species, 14% of bird species and 30% of sharks and rays.

All these threats are causing unnatural and fast environmental changes which require a rapid adaptation from wild animals. In this unsteady scenario, the study of populations turns out of crucial importance for wildlife conservation and statistics, an essential tool to perform it.

The main objective of population ecology is the analysis of the factors that affect population dynamics through the study of key demographic characteristics (such as survival, movement and productivity). The knowledge of these characteristics which allow us to predict how the populations will develop in the future. Indeed, survival is one of the fundamental processes governing animal population dynamics.

Although the interest is often to estimate survival probabilities, there has been an increased interest also in describing the reasons underneath those estimations. Indeed, several studies have analysed how survival can vary with age (Crespin et al., 2006; Lee et al., 2008; Reynolds et al., 2009; Sanz-Aguilar et al., 2016; Meade et al., 2013), between individuals (Harris et al., 1992; Cam et al., 2002; Royle, 2008; Giménez and Choquet, 2010), between sexes (Aebischer and Coulson, 1990; Owens, 2002; Hastings et al., 2012), in relation to environmental covariates (Schaub et al., 2005; Votier et al., 2008; Hegg et al., 2012; Ramos et al., 2012), to density-dependence (Milner et al., 1999; Frederiksen and Bregnballe, 2000; Barker et al., 2002) and time (Hastings et al., 1999; Harris et al., 2007).

Among other factors, longevity in wild animals can vary with body mass (large species tend to live longer than smaller species), but it can be also associated with the extrinsic mortality, i.e. the risk of death due to external causes. Clearly, species with low extrinsic mortality are generally long-lived. This is due to the fact that, when premature death is not likely, investment in long-term maintenance and survival is favoured in contrast to an early and frequent reproduction. That is tipically the case for seabirds.

Seabirds comprise several different families of birds inhabiting marine ecosystems, but most of them share several life history traits: a high adult survival, a progressive access to reproduction at relatively old ages, a low reproductive rate and a low number of predators. Adult survival, but also juvenile survival, are considered important traits that influence population dynamics in seabird species. However, juvenile survival is difficult to estimate due to this group of species have long periods of immaturity spending several years at sea before returning to the colonies. Indeed, less than 2% of seabird species have juvenile survival estimates (Lewison et al., 2012).

The survival estimation in ecology is far for simple. Ideally, to analyse survival probabilities in the wild, individuals must be followed from birth to death over a study period, but clearly, this is rarely possible in wild populations. The problem arises since we cannot always make a direct measurement of the state variable of interest (i.e. alive or dead). Hence in the wild, if an individual is not observed two options are possible: (i) it is dead, or (ii) it is still alive (but it has not been observed). Nevertheless, this common problem is resolved using the methods of capture-mark-recapture (CMR).

Capture-mark-recapture methods (also referred to as mark-recapture methods) were originally first used in s.XVIII (see Seber and Schofield, 2019, and references therein) and they were developed to accommodate the chance that a not observed individual may be not dead but has simply been overlooked. Indeed, CMR models model the probability of observing each individual as a product of its survival rate and the probability of observing the individual given that it has survived.

Next Section is devoted to present the CMR methods as well as to provide a general review of the main CMR models and to describe the two models implemented along this thesis, the Cormack-Jolly-Seber (CJS) models and the integrated models.

1.2 Capture-mark-recapture models (and methods)

CMR models are a large class of models that have become increasingly used in ecological applications of statistical modeling. Over the past decades, a body of theory and methods have been developed for the analysis of capture-recapture(-recovery) data, and the area remains in active research (Lebreton et al., 1992; Royle and Dorazio, 2008; King et al., 2010; McCrea and Morgan, 2015; Seber and Schofield, 2019). The key concept behind the CMR models relays on the fact that they account for imperfect detection through the estimation of detection probabilities, which provide the link between what we observe and the true population parameters. Indeed, if imperfect detection is not accounted for in the modeling of ecological systems, this leads to biased parameter estimates (Kéry et al., 2009; Kéry and Schaub, 2011).

CMR methods use data collection protocols where observers go into the field at a series of capture events, denoted $t = 1, \ldots, T$. At each capture event, the observer marks the new individuals, records all observed individuals (already marked) and releases them back into the population. It is mandatory to ensure that animals can be identified uniquely, therefore, animals are uniquely marked and hence, uniquely identifiable. When individuals are only recaptured alive, this is referred to capture-recapture data. However, if individuals are recovered dead, it is called ring-recovery data and finally, if individuals are both recaptured (alive) and recovered (dead), it is referred to mark-recapture-recovery (MRR, capture-recapture-recovery or integrated) data.

Historically, recapture meant physical capture of the animals, but in the last decades new methods have been developed that do not require physical capture (e.g. traps). Further, the type of tag or mark used depends very much on the species of interest, although some methods can be used across different species. In ecology is widespread the use of rings (metallic and/or plastic), collars, ear rings, camera traps, acoustic recording devices, etc. (Royle et al., 2014, and references therein).

Standard capture-recapture data (also referred as mark-recapture data) are typically displayed in the form of individual capture histories for each individual observed along the study period. Each individual capture history is a sequence of 1s and 0s, denoting if an individual was encountered (or not) at each occasion over the study period. For example, the encounter history:

$$1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1$$

indicates that the individual is first marked at occasion $t = 1$, observed again at occasions $t = 4, 6$ and not observed at occasions $t = 2, 3, 5$.

However, in the case of ring-recovery data (also referred to as mark-recovery), the corresponding capture histories are a combination of the first 1 (marking), 0s (not recovered) and 2s (recovered dead). For example:

$$1 \quad 0 \quad 0 \quad 0 \quad 2 \quad 0$$

indicates the individual is marked at occasion $t = 1$, and recovered dead at occasion $t = 5$.

Finally, if both recaptures and recoveries are available, then we work with integrated data. In this case, the encounter histories are a combination of 0s, 1s and 2s. For example, the encounter history:

$$0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 2$$

indicates that the individual is first marked at occasion $t = 2$, observed (alive) again at occasions $t = 3, 4$, recovered (dead) at $t = 6$, and not observed at occasion $t = 5$.

The associated models for capture-recapture(-recovery) data and their corresponding statistical analysis clearly depend on the exact form of the data collection process. There are two general classes of models depending on the type of populations (closed or open). A population is called closed if, within the study period, there are not births, deaths or migration. One can say that in reality, the populations are far from closed, and he/she would be right. However, if population changes related to births, deaths or migration are small enough, we may consider a population as closed (over that study period). Additionally, if studies are performed over a short time period (e.g. weeks), the population can also be considered as closed (Borchers et al., 2002). Alternatively, open populations do change due to births, deaths or migration (over the study period). As everyone can imagine, this is often the case for wildlife populations, and that is why the development of capture-recapture models for open populations has had a exponential increase from the first models presented (Amstrup et al., 2005; King et al., 2010; King, 2014).

The different CMR models can also be divided into those primarily associated with estimating demographic parameters (survival, recapture and recovery probabilities, migration probability, recruitment, etc.) and those that estimate abundance. Here we present a review of the main open CMR models before presenting in detail the ones used in this thesis.

Let us start with the Cormack-Jolly-Seber (CJS) models (Cormack, 1964; Jolly, 1965; Seber, 1965) that are considered a general framework for several number of extensions. They are conditional on the initial capture of each individual, however, if we do not condition on first capture (and hence, we can use the complete capture histories), we then have the Jolly-Seber (JS) model (Jolly, 1965; Seber, 1965). JS model assumes that individuals either marked or unmarked have the same probability of being caught. In other words, newly captured individuals must be a random sample of all unmarked individuals of the population. Pledger et al. (2009) extended the JS model by specifying the survival probability of an individual as dependent on its time of arrival to the study (recruitment). These models, also called stop-over models, allow the estimation of the survival, the arrival times and the retention probabilities (stop-over).

In the last two decades CJS models have been also extended to allow for additional data calling them integrated models (Burnham, 1993; Barker, 1997; Catchpole et al., 1998; King and Brooks, 2003; Abadi et al., 2010). Additional data can vary from ring-recovery, populations counts, reproductive data, etc. An additional extension to CJS models are the so called Multistate models (Arnason, 1973; Schwarz et al., 1993) where individuals can move between states (i.e. geographical sites, behavioral or physiological conditions) according to the transition probabilities between those states. Multistate models assume first-order Markov process for movement between states, however, King and Langrock (2016) extended them to allow for more flexible transitions via Semi-Markov models. Additionally, Multistate models can also be extended to Multievent models (Pradel, 2005) that allow accommodating uncertainty in state assignment.

A recent development in capture-mark-recapture models are the spatial explicit capture-recapture (SECR) models (Efford, 2004; Borchers and Efford, 2008; Royle and Young, 2008) in which multiple traps are arranged in an array within a given area for each capture occasion. SECR models extend standard capture-mark-recapture models incorporating the spatial information. Therefore, the data comprise the encounter histories and, given that an individual is observed, the corresponding trap(s) over each capture occasion.

Finally, it is worth to mention the occupancy models (MacKenzie et al., 2002, 2003; Tyre et al., 2003). Occupancy data can be seen as a form of capture-recapture data. In particular, these data comprise the presence-absence of the species in several sites over a number of visits. Occupancy models have a similar structure to CMR models, allowing to jointly model the probability of occupancy and the detection in animals (or plants).

Table 1.1 presents a summary of the main CMR models reviewed as well as the population parameters of interest, the kind of data and type of population regarding each model.

Parameter(s) of interest	**Data**	**Population**	**Model**
Apparent Survival	CR	Open	Cormack-Jolly-Seber
True Survival	RR	Open	Cormack-Jolly-Seber
True Survival, migration, recovery prob.	MRR	Open	Integrated models
True Survival, migration, recovery prob.	MRR	Open	Multistate/Multievent
Apparent survival and movement	CR	Open	Multistate/Multievent
Apparent survival, recruitment, population size	CR	Open	Jolly-Seber
Apparent survival, recruitment, retention probability	CR	Open	Stop over
Abundance, movement, resource selection	SCR	Open/Closed	Spatial Explicit CR
Occupancy, species distribution	Presence/Absence	Open	Occupancy models

TABLE 1.1: Summary of the main capture-mark-recapture models indicating the parameters of interest, type of data and type of population. CR: capture-recapture; RR: ring-recovery; MRR: mark-recapture-recovery; and SCR: spatial capture-recapture.

1.3 Cormack-Jolly-Seber models

Cormack–Jolly–Seber models consider capture-recapture(-recovery) data and are formulated conditional on the initial capture of each individual. These models can be interpreted as partially observed hidden Markov models and can be expressed using either multinomial or state-space formulation (King, 2012; McCrea and Morgan, 2015). CJS models make important assumptions which violation may bias parameter estimates: (i) marks must not be lost, (ii) identity of the individuals must always be recorded without errors, and (iii) captured and recaptured individuals are regarded as a random sample from the study population.

The estimated parameters differ depending on the kind of data used (i.e. capture-recapture or ring-recovery). When working with capture-recapture data, CJS models allow us to estimate the *apparent* survival probability (denoted by ϕ) and the recapture probability (denoted by p). The adjective *apparent* comes from the fact that it is the product of the probabilities of the *true* survival (denoted by S) and the fidelity to the study area (denoted by F). Note that, assuming that migration is permanent, this implies that dead and migration are confounded. It is also worth mentioning that survival probability is commonly reported as annual due to wild animals experience a natural yearly life cycle and data are collected annually.

Alternatively, when we work with ring-recovery data the estimated parameters are: the *true* survival probability (S) and the recovery probability (denoted by λ). The recovery probability (instead of the recapture probability) represents the probability that an individual is found dead. Alternatively, the *true* survival probability is the probability that an individual survives from time t to $t + 1$.

We next introduce the general notation as well as we describe in detail both state-space and multinomial formulations for CJS models to the case of capture-recapture and ring-recovery data. For illustration, here we only present the case of time-dependent parameters (allowing for individual heterogeneity). However, further extensions of dependence structure can easily be applied. For the full background of the CJS models see for example Lebreton et al. (1992), Brooks et al. (2000), Pledger et al. (2003), Giménez et al. (2007), Royle (2008), Giménez et al. (2009), King et al. (2010), Kéry and Schaub (2011), King (2012), King (2014), McCrea and Morgan (2015) and Seber and Schofield (2019).

1.3.1 State Space formulation

State-space models are models for two discrete time processes running in parallel: the state and the observation process. This is a common scenario for ecological studies where there are two linked stochastic processes, one corresponding to the underlying biological state (that it is normally latent), and the other being the observation process (observed with error) (Kéry et al., 2009; Kéry and Schaub, 2011;

King, 2012). Although for an ecologist the primary interest is often based on the state model, the observation model provides the essential link with the underlying state model.

State-space models are first-order hidden Markov models, hence, the probability of a particular state depends only on the previous state (Figure 1.1). From now on, we denote by $i = 1, \dots, N$ the marked individuals, $t = 1, \dots, T$ represents the capture occasions and $\boldsymbol{f} \in \{f_1, \dots, f_N\}$ is the vector that contains the year in which each individual is marked, so that $f_i \in \{1, \dots, T-1\}$. Let $z_{i,t}$ denote the latent variable that describes the true state of individual i at time $t = f_i, \dots, T$. The possible states that the variable $z_{i,t}$ can take (indicating the true state of individual i at time t) are equivalent for both capture-recapture and ring-recovery data:

$$z_{i,t} = \begin{cases} 0 & \text{individual } i \text{ is dead at time } t; \\ 1 & \text{individual } i \text{ is alive at time } t. \end{cases}$$

Similarly, let $y_{i,t}$ denote the binary variable that describes whether individual i is observed or not at time $t = f_i + 1, \dots, T$; being the possible states that this binary variable can take:

$$y_{i,t} = \begin{cases} 0 & \text{individual } i \text{ is not observed at time } t; \\ 1 & \text{individual } i \text{ is observed at time } t. \end{cases}$$

Finally, $x_{i,t}$ stands for the binary variable that describes whether individual i is recovered (dead) or not at time $t = f_i + 1, \dots, T$. The possible states are:

$$x_{i,t} = \begin{cases} 0 & \text{individual } i \text{ is not recovered dead at time } t; \\ 1 & \text{individual } i \text{ is recovered dead at time } t. \end{cases}$$

We note that for both capture-recapture and ring-recovery data, $y_{i,f_i} = x_{i,f_i} = 1$ for all individuals due to CJS models are conditional on the first time an individual is observed alive (i.e. first capture).

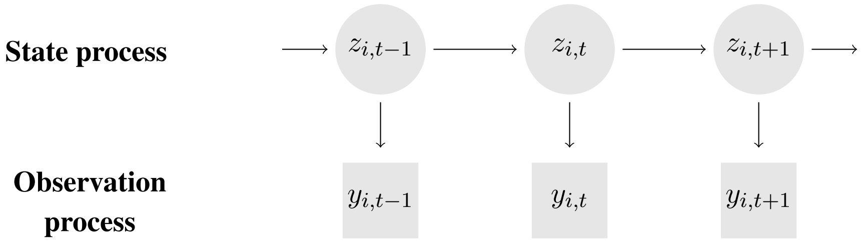

FIGURE 1.1: Diagram of a state-space model (SSM).

We initially describe the state and observation processes regarding capture-recapture data. The state and observation equations can be written as:

$$z_{i,t+1}|z_{i,t} = z_{i,t} \sim \text{Bernoulli}(z_{i,t}\phi_{i,t}), \text{ for } t = f_i, \ldots, T-1, \tag{1.1}$$

$$y_{i,t}|z_{i,t} = z_{i,t} \sim \text{Bernoulli}(z_{i,t}p_{i,t}), \text{ for } t = f_i + 1, \ldots, T, \tag{1.2}$$

for $i = 1, \ldots, N$ and being $\phi_{i,t}$ the *apparent* survival probability of individual i from time t to $t+1$, and $p_{i,t}$ the recapture probability of individual i at time t (given alive).

Alternatively, the state and observation equations regarding ring-recovery data are given by:

$$z_{i,t+1}|z_{i,t} = z_{i,t} \sim \text{Bernoulli}(z_{i,t}S_{i,t}), \text{ for } t = f_i, \ldots, T-1, \tag{1.3}$$

$$x_{i,t}|z_{i,t} = z_{i,t} \sim \text{Bernoulli}(\lambda_{i,t}(1 - z_{i,t})z_{i,t-1}), \text{ for } t = f_i + 1, \ldots, T, \tag{1.4}$$

for $i = 1, \ldots, N$ and being $S_{i,t}$ the *true* survival probability of individual i from time t to $t+1$. Further, if an individual is recovered dead at time t it is often assumed that the individual died in the interval $(t-1, t]$ due to marks degradation (although it is possible to relax this assumption, see Catchpole et al., 2001). The product $(1 - z_{i,t})z_{i,t-1}$ ensures this assumption. Finally, $\lambda_{i,t}$ denotes the recovery probability of individual i at time t given dies in $(t-1, t]$.

For both kind of data, assuming independence of individuals and conditional on the first capture, the resulting complete data likelihood is given by the product of all individual likelihood components,

$$\ell(\mathbf{y},\mathbf{z}|\boldsymbol{\theta}) \propto \prod_{i=1}^{N} \left\{ \underbrace{\left(p_{z_{f_i}}(z_{i,f_i}|\boldsymbol{\theta}) \prod_{t=f_i}^{T-1} p_z(z_{i,t+1}|z_{i,t},\boldsymbol{\theta}) \right)}_{\ell_{state}(\boldsymbol{z}|\boldsymbol{\theta})} \times \underbrace{\left(\prod_{t=f_i+1}^{T} p_y(y_{i,t}|z_{i,t},\boldsymbol{\theta}) \right)}_{\ell_{obs}(\boldsymbol{y}|\boldsymbol{z},\boldsymbol{\theta})} \right\}, \tag{1.5}$$

where $\boldsymbol{\theta}$ represents the vector of all the parameters of the model. Also, the term $p_{z_{f_i}}$ denotes the probability function for the initial state (marking time); while p_z and p_y (or alternatively p_x for ring-recovery data) are the (Bernoulli) probability functions of the state and observation processes, respectively. Finally, ℓ_{state} and ℓ_{obs} represent the likelihood terms related to the state and observation process, respectively. Note that the observed data likelihood, $\ell(\boldsymbol{y}|\boldsymbol{\theta})$, is obtained by summing out over the $\boldsymbol{z}$ values.

1.3.2 Multinomial formulation

Although state-space formulation is easy to interpret as well as flexible, the likelihood becomes large and complex, requiring significant computational effort when proceeding to the inferential process, especially when dealing with large databases. The underneath reason is that all the capture histories are analysed individually, further, every unknown latent state needs to be estimated. Nevertheless, capture-recapture(-recovery) data can be efficiently condensed in the so-called reduced *m*-array (Lebreton et al., 1992) which summarizes the number of individuals released at occasion $j = 1, \ldots, T-1$ (denoted by R_j), and subsequently next observed (or recovered dead) at occasion $t = 2, \ldots, T$. The *m*-arrays reduce importantly the computational cost in the inferential process, however, an important disadvantage is that models with individual effects can no longer be fitted due to full individual histories are not retained.

Next paragraphs are devoted to explain in detail the *m*-arrays, cell probabilities and the complete data likelihood for capture-recapture and ring-recovery data separately, considering full time dependence on model parameters.

1.3.2.1 Capture-recapture data

For capture-recapture data under consideration, an *m*-array is an upper triangular matrix where rows correspond to release occasions (denoted by j), and columns represent first recapture occasions (denoted by t). There is an additional column $(T+1)$ corresponding to individuals never recaptured. Let denote $m_{j,t}$ the number of individuals released at time j and next captured in time $t{+}1$, for $j = 1, \ldots, T{-}1$ and $t = 1, \ldots, T-1$, then by definition, $m_{j,t} = 0$ for $t < j$. In other words, the lower triangular elements of the *m*-array are structural zeros (i.e. impossible events). χ_j denotes the probability of an individual released at time j is not observed again during the study, and $m_{j,T+1}$ represents the corresponding number of individuals.

Let R_j denote the number of individuals released at time j. This number represents either the individuals newly marked and released at time j or those already marked and first recaptured at time t and then 'released' at occasion $j = t$. In other words, once an individual is first recaptured at time t, it is then released as a member of a new cohort at release occasion $j = t$. Therefore, under this model formulation individuals cannot be recaptured more than once in a given release occasion j.

The expected values of the entries of the *m*-array are given on the underlying model parameters and the number of released individuals. All of them define the cell probabilities of the multinomial distributions, one for each release occasion. As an example, the probability associated with the $(j = 1, t = 3)$ cell (denoted $q_{j,t}$) corresponds to the probability of surviving from occasion $t = 1$ to $t = 2$ (ϕ_1), and from occasion $t = 2$ to $t = 3$ (ϕ_2), multiplied by the probability of not being recaptured at occasion $t = 2$ $(1 - p_2)$, and recaptured at occasion $t = 3$ (p_3):

$$q_{1,3} = \phi_1(1 - p_2)\phi_2 p_3.$$

Additionally, the probability of never observing an individual is defined: $\chi_j = 1 - \sum_{t=j+1}^{T} q_{j,t}$.

To better understand, in Table 1.2 we present the expected and observed values considering the duration of the study is $T = 4$. For the sake of simplicity, $\tilde{p}_t$ denotes the probability of not being observed, that is, $\tilde{p}_t = (1 - p_t)$.

			First Recapture occasions (t)			
			2	**3**	**4**	**Never recaptured**
Release occ. (j)	**1**	R_1	$R_1\phi_1p_2$ $(m_{1,2})$	$R_1\phi_1\tilde{p_2}\phi_2p_3$ $(m_{1,3})$	$R_1\phi_1\tilde{p_2}\phi_2\tilde{p_3}\phi_3p_4$ $(m_{1,4})$	$R_1\chi_1$ $(m_{1,5})$
	2	R_2	-	$R_2\phi_2p_3$ $(m_{2,3})$	$R_2\phi_2\tilde{p_3}\phi_3p_4$ $(m_{2,4})$	$R_2\chi_2$ $(m_{2,5})$
	3	R_3	-	-	$R_3\phi_3p_4$ $(m_{3,4})$	$R_3\chi_3$ $(m_{3,5})$

TABLE 1.2: Example of m-array for capture-recapture data. Individuals released (R_j), expected and observed (parenthesis) individuals in an study with $T = 4$.

Conditional on the numbers released and assuming independence between individuals, the CJS model likelihood can be written as a product of multinomial probability distributions corresponding to each row of the m-arrays. The general likelihood of CJS model for capture-recapture data takes the form:

$$\ell(\boldsymbol{\theta}) \propto \prod_{j=1}^{T-1} \left\{ \prod_{t=j+1}^{T} \left(\phi_{t-1}p_t \prod_{k=j}^{t-2} \phi_k(1-p_{k+1}) \right)^{m_{j,t}} \chi_j^{m_{j,T+1}} \right\}, \tag{1.6}$$

where $\boldsymbol{\theta}$ represents the vector of all the parameters involved in the model and

$$\chi_j = 1 - \sum_{t=j+1}^{T} \left(\phi_{t-1}p_t \prod_{k=j}^{t-2} \phi_k(1-p_{k+1}) \right).$$

1.3.2.2 Ring-recovery data

Alternatively for ring-recovery data, the rows of the m-array correspond to release occasions (denoted by j), and columns represent recovery occasions (denoted by t). There is also an additional column ($T+1$) corresponding to never recovered individuals. Let denote $m_{j,t}$ the number of individuals released at time j and recovered dead in time $t+1$, for $j = 1, \dots, T-1$ and $t = 1, \dots, T-1$, again by definition, $m_{j,t} = 0$ for $t < j$. The lower triangular elements of the m-array are also structural zeros (i.e. impossible events). χ_j denotes the probability of an individual released at time j is not recovered during the study, and $m_{j,T+1}$ represents the corresponding number of individuals.

R_j denotes the number of individuals released at time j. For ring-recovery data, this number comprises only the released individuals at time j, due to just one further observation is possible, i.e. the dead recovery.

Following the example presented above, the probability associated with the $(j = 1, t = 3)$ cell (denoted $q_{j,t}$) for the ring-recovery *m*-array corresponds to the probability of surviving from occasion $t = 1$ to $t = 2$ (S_1), and dying from occasion $t = 2$ to $t = 3$ $(1 - S_2)$, multiplied by the probability of being recovered at occasion $t = 3$ (λ_3):

$$q_{1,3} = S_1(1 - S_2)\lambda_3.$$

Similarly, the probability of never observing an individual in the ring-recovery *m*-array is defined: $\chi_j = 1 - \sum_{t=j+1}^{T} q_{j,t}$. To better understand, in Table 1.3 we present the expected and observed values considering the duration of the study is $T = 4$ for ring-recovery data.

			Recovery occasions (t)		
Release occ. (j)		**2**	**3**	**4**	**Never recovered**
1	R_1	$R_1(1-S_1)\lambda_2$ $(m_{1,2})$	$R_1S_1(1-S_2)\lambda_3$ $(m_{1,3})$	$R_1S_1S_2(1-S_3)\lambda_4$ $(m_{1,4})$	$R_1\chi_1$ $(m_{1,5})$
2	R_2	-	$R_2(1-S_2)\lambda_3$ $(m_{2,3})$	$R_2S_2(1-S_3)\lambda_4$ $(m_{2,4})$	$R_2\chi_2$ $(m_{2,5})$
3	R_3	-	-	$R_3(1-S_3)\lambda_4$ $(m_{3,4})$	$R_3\chi_3$ $(m_{3,5})$

TABLE 1.3: Example of *m*-array for ring-recovery data. Individuals released (R_j), expected and observed (parenthesis) individuals in an study with $T = 4$.

It is worth pointing out (and as it can be appreciated in Tables 1.2 and 1.3) the different information gathered for each kind of data. For example, the main diagonal of the *m*-arrays represents, for capture-recapture data, the probability of surviving and being recaptured, $\phi_t p_{t+1}$, while for ring-recovery data the main diagonal is indicating the probability of dying and being recovered dead, $(1 - S_t)\lambda_{t+1}$.

Conditional on the numbers released and assuming independence between individuals, the CJS model likelihood for ring-recovery data can be written as a product of multinomial probability distributions corresponding to each row of the m-arrays. The general likelihood takes the form:

$$\ell(\boldsymbol{\theta}) \propto \prod_{j=1}^{T-1} \left\{ \prod_{t=j+1}^{T} \left((1 - S_{t-1})\lambda_t \prod_{k=j}^{t-2} S_k \right)^{m_{j,t}} \chi_j^{m_{j,T+1}} \right\}, \tag{1.7}$$

where $\boldsymbol{\theta}$ represents the vector of all the parameters involved in the model and

$$\chi_j = 1 - \sum_{t=j+1}^{T} \left((1 - S_{t-1})\lambda_t \prod_{k=j}^{t-2} S_k \right).$$

1.4 Integrated models

Integrated models are designed to analyze multiple data sources simultaneously within a single robust analysis. By doing so, the information can be shared across data sets. In the ecological context this feature is particularly relevant, mainly due to ecological data are generally incomplete, or there may be biases because of the study design. In this scenario, integrated models allow to fill the gaps through the combination of the information gathered by the different data sets. Some clear advantages of these models are that the precision of parameter estimates usually increases, but more importantly, that they allow to estimate other demographic parameters not possible to be estimated otherwise due to parameter redundancy issues (Besbeas et al., 2005; McCrea and Morgan, 2015).

Two types of integrated models can be distinguished depending on whether independent or nonindependent data sets are used. Integrated models for nonindependent data sets are typically more complex (see McCrea and Morgan, 2015 for more details). However, integrated models for independent data sets can be easily constructed by decomposing the joint likelihood of the different data sets into a product of the corresponding likelihoods for each individual data set (given the relevant parameter values). A key element of this modelling is that one (or several) parameters are common in several components of the likelihood. If we let $\{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$ denote the different data sets, as stated in King (2014), the integrated likelihood is then:

$$f(\mathcal{D}_1, \ldots, \mathcal{D}_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(\mathcal{D}_i|\boldsymbol{\theta}).$$

The integrated approach has been applied to a wide range of data combinations: capture-recapture and ring-recovery data (nonindependent data sets); population counts, productivity and capture-recapture data; population counts and ring-recovery data; and distance sampling and capture-recapture data; among many others (Lebreton et al., 1995; Barker et al., 2004; McCrea et al., 2010; Kéry and Schaub, 2011; Schaub and Abadi, 2011; King, 2014; McCrea and Morgan, 2015).

In this thesis (Chapter 5), we have implemented a particular integrated model where capture-recapture and ring-recovery data from the same individuals were analysed simultaneously. The following paragraphs give a brief introduction of the mark-recapture-recovery (MRR) models but in Chapter 5 we present them in detail to the particular case of a colony of common guillemot.

Mark-recapture-recovery models (MRR)

The joint analysis of live recaptures and ring-recovery data was first presented by Burnham (1993). This study was focused on time-dependent model parameters, considering both random and permanent migration. This model was lately generalised by Barker (1997), allowing for the joint analysis of live-recapture, dead recovery and live resighting data in the special case of random temporary migration, which was further extended to age-dependence and temporary marking effect in another work by Barker (1999). Further, Catchpole et al. (1998) extended Burnham's model to incorporate age dependent parameters.

Recent developments have extended the model of Catchpole et al. (1998), and have included a state-space formulation (Bonner et al., 2010; Colchero and Clark, 2012; King, 2012), multi-site/multi-state MRR models (Lebreton et al., 1999; King and Brooks, 2003; McCrea et al., 2012; King, 2014; McCrea and Morgan, 2015) and incorporated individual time-varying continuous covariates (Cathpole et al., 2008; King et al., 2008; Bonner et al., 2010; Langrock and King, 2013; Worthington et al., 2015; King et al., 2010).

The data collection protocol for integrated mark-recapture-recovery (MRR) studies consists of a repeated population sampling at a series of capture occasions, denoted by $t = 1, \dots, T$. At each sampling occasion t, new individuals are marked, and recaptures and recoveries are performed. Thus, mark-recapture-recovery data are commonly presented in the form of individual encounter histories describing whether the individual has been observed alive or dead, or it has not been observed for each capture occasion t. We let $i = 1, \dots, N$ denote the marked individuals and $t = 1, \dots, T$ the associated capture occasions. We let $\boldsymbol{y}$ denote the $N \times T$ matrix such that,

$$y_{i,t} = \begin{cases} 0 & \text{individual } i \text{ is not resighted at time } t; \\ 1 & \text{individual } i \text{ is resighted alive at time } t; \\ 2 & \text{individual } i \text{ is recovered dead in the interval } (t-1, t]. \end{cases}$$

As for standard ring-recovery models the distinction of individuals recovered dead in the interval $(t-1, t]$ is a general assumption to MRR models due to the degeneration of the marks once the animal has died (Langrock and King, 2013). Hence, only individuals dead at time t but alive at time $t-1$ (recently dead individuals) can be recovered; conversely to those individuals recovered dead at time t but which were already dead at time $t-1$ (long dead individuals).

Capture-recapture data alone allow the estimation of *apparent* survival probabilities, denoted by ϕ, and recapture probabilities, denoted by p. It is called *apparent* due to *true* survival is confounded with migration (denoted by γ) from the study site. However, ring-recovery data alone contain information about *true* survival probabilities, denoted by S, and recovery probabilities, denoted by λ. Integrated models allow the estimation of parameters not estimable otherwise. This is the case for migration probability for MRR data. In what follows we consider the case of full time dependence on model parameters.

In ecological studies there are three options when dealing with migration: (i) migration is permanent, (ii) migration is random, (iii) there is no migration. We consider the case where migration is permanent, so that if an individual alive at time t, migrates with probability γ_t and it is not at risk of capture in subsequent times $t+1,\dots,T$. However, when migration is random, individuals alive at time t can migrate at any time and they are available for capture at subsequent times $t+1,\dots,T$ (i.e. they can return). Therefore, when migration is random, the probability that an animal is at risk of capture at time t does not depend on whether it was at risk of capture on earlier occasions (Barker and White, 2001). Furthermore, when random or permanent migration is present, we can reparameterize the *apparent* survival probability as $\phi_t = (1-\gamma_t)S_t$, or equivalently, $\phi_t = F_t \times S_t$, denoting by F_t the fidelity to the study site at time t. In case there is no migration, the *apparent* and the *true* survival are equal, $\phi_t = S_t$, namely, complete site fidelity. Finally, we also need to account for the probability of not observing an individual again (either alive or dead), this χ_t term can be defined recursively as,

$$1-\chi_t = (1-S_t)\lambda_t + S_t 1 - (1-p_{t+1})\chi_{t+1},$$

for $t < T$, and being $\chi_T = 1$.

Following McCrea and Morgan (2015), in Table 1.4 we present an example of the corresponding probabilities to four capture histories in an study of $T = 4$, considering full time-dependence in model parameters and assuming no migration in the study (i.e. $\phi_t = S_t$).

$t=1$	$t=2$	$t=3$	$t=4$	Probabilities of encounter histories
1	1	2	0	$S_1 p_2 (1-S_2)\lambda_2$
1	0	0	2	$S_1(1-p_2)S_2(1-p_3)(1-S_3)\lambda_3$
1	1	0	1	$S_1 p_2 S_2 (1-p_3) S_3 p_4$
1	0	0	0	χ_1

TABLE 1.4: Example of four encounter histories and their associated probabilities in an MRR study assuming no migration (i.e. $\phi_t = S_t$).

1.5 Inference

Once the model is selected the next step is the inference. CMR models can be fitted using either a Bayesian or a classical approach. Classical inference assumes that parameters are fixed and data are random observations. Parameters can be estimated using, for example, maximum likelihood, i.e. the value of the parameters that maximize the associated likelihood function of the given model. Alternatively, the Bayesian reasoning is based on the assumption that parameters are random variables. Moreover, in the Bayesian paradigm all types of uncertainty are always expressed in terms of probability distributions.

Although both perspectives are equally valid to make inference, in this thesis the inference has been performed within the Bayesian framework mainly due to two reasons. Firstly, the hierarchical structure of the Bayesian approach provides a natural framework to combine data from different sources into a single analysis. And secondly, the Bayesian paradigm allows us to explicitly incorporate all the available prior knowledge about the unknown parameters into the inferential process, an important feature for ecological studies (King et al., 2010; Christensen et al., 2011; Martínez-Abraín et al., 2014).

In what follows we introduce the Bayesian paradigm, before describing Markov chain Monte Carlo (MCMC) methods (Robert and Casella, 2011) which allow us to apply such methods in those cases in which posterior distributions are not analytical (which is the case with the models here described). To conclude, we also provide computing programs, `R` packages and web pages that we consider relevant in the capture-recapture framework. Finally, for a detailed explanation of classical inference in the capture-recapture framework the reader can refer to McCrea and Morgan (2015).

1.5.1 Bayesian inference

The Bayesian approach to inference dates from s.XVIII, when Thomas Bayes presented a mathematical treatment of statistical data analysis named *the inverse probability* and that we know as Bayes' Theorem:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})}{f(y)} \propto f(\boldsymbol{y}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}), \tag{1.8}$$

where $\boldsymbol{\theta}$ denotes the parameters on which we want to make inference and $\boldsymbol{y}$ denotes the observations from a known probability density function (data).

In the above expression of Bayes theorem, $f(\boldsymbol{y}|\boldsymbol{\theta})$ represents the likelihood, containing the information given by the data under the model parameters. Also, the current knowledge about the parameters is expressed by the prior distribution, p($\boldsymbol{\theta}$). Note that if there is no previous knowledge about them, prior distribution should be as less informative as possible. The term $f(y)$ is known as the normalisation constant that assures that the distribution integrates to one. Finally, the term $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ refers to the posterior distribution of model parameters which reflects the updated knowledge, balancing prior knowledge with observed data. In other words, the key of the Bayesian paradigm is the 'learning process', namely, initial beliefs could be evaluated, updated and modified with new information. Therefore, given a set of observed data and a model that relates them with the parameters, the knowledge about these parameters can be updated through Bayes' Theorem.

Historically, Bayesian methods were restricted by the need to perform integrations analytically to obtain the marginal posterior distributions of parameters of interest. Hence, although the basis of the Bayesian methodology is simple and intuitive, its application to complex real problems does not have an analytic closed expression.

In the late 80s and beginning of the 90s of the last century all this changed due to the development of different algorithms that allowed to simulate or approximate (instead of analytically integrating) the posterior distributions. The most popular ways to approximate (via simulation) those posterior distributions are the Markov chain Monte Carlo (MCMC) methods (Gelfand and Smith, 1990; Smith and Gelfand, 1992), the sequential Monte Carlo methods (Del Moral, 1996; Liu and Cheng, 1998) and importance sampling. But there are other good ways to approximate posterior distributions such as the integrated nested Laplace approximation methodology (Rue et al., 2009, `www.r-inla.org`) and the Approximate Bayesian Computation methods (Tavaré et al., 1997).

In our case, we have used the MCMC approach to approximate the posterior distributions. Next Section is devoted to explain it, as well as to present the two basic Markov chain simulation algorithms: Metropolis-Hastings sampling (Metropolis et al., 1953; Hastings, 1970) and Gibbs sampler (Gelfand and Smith, 1990).

1.5.2 Markov chain Monte Carlo (MCMC) methods

MCMC methods perform Monte Carlo integration using a Markov chain to generate observations from the posterior distribution of interest, $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. A Markov chain is a sequence of random numbers ($\theta^t, \theta^{t+1}, \ldots$) for which, for any t, the distribution of θ^t given all the values of θ depends only on the previous value θ^{t-1}. MCMC methods construct a Markov chain whose values are updated at each iteration t such that, if the chain is running for long enough, the distribution of the chain converges to the posterior distribution of interest. In other words, the stationary distribution of the Markov chain is then the posterior distribution of interest. Therefore, when the posterior distribution is analytically intractable, MCMC methods allow us to sample iteratively so that in each step of the process we expect to draw from a distribution that becomes closer to the posterior distribution of interest.

Once the chain has converged to the stationary distribution (our posterior distribution of interest), we can obtain empirical (Monte Carlo) estimates of any posterior summaries of interest. However, it is necessary to check the convergence of the simulated sequences (i.e. ensure that the Markov chain has reached the stationary distribution). To that end, we need to use only the realisations (simulations) once the chain has converged, discarding then the early ones (period called *burn-in*). There are many issues regarding the application of these MCMC methods that we do not discuss here, such as the length of the *burn-in* period, the number of iterations of the chains, the *thinning* period and the number of chains. For more information to this respect see for example Christensen et al. (2011) and Gelman et al. (2014).

Metropolis-Hastings algorithm

The way of constructing MCMC samplers is similar to the acceptance-rejection sampling algorithm: values are drawn from a proposal distribution and are corrected so that they behave as random observations from the target distribution. One of the most populars MCMC algorithms is the Metropolis Hastings,in which the draws are conditional only upon the last observation (Markov chain property).

The Metropolis Hastings algorithm starts with an initial value $\boldsymbol{\theta^0}$ from some arbitrary distribution. We then generate a new value $\boldsymbol{\theta^{t+1}}$ dependent only upon $\boldsymbol{\theta^t}$:

$$\boldsymbol{\theta}^{t+1} | \boldsymbol{\theta}^{t} \sim \mathcal{K}(\boldsymbol{\theta}^{t}, \boldsymbol{\theta}),$$

where $\mathcal{K}$ is the transition kernel of the Markov chain and only describes the dynamics of the chain. The algorithm can be described as:

STEP 0. Initialization. Generate an arbitrary point $\boldsymbol{\theta^0}$ and choose a proposal density $q(\boldsymbol{\theta^*}, \boldsymbol{\theta^0})$.

STEP 1. At each iteration t generate a candidate $\boldsymbol{\theta^*}$ from the proposal density and a random number $u \sim U[0, 1]$.

STEP 2. Calculate the acceptance ratio $\alpha(\boldsymbol{\theta^t}, \boldsymbol{\theta^*})$:

- If $u \leq \alpha(\boldsymbol{\theta^t}, \boldsymbol{\theta^*})$, $\boldsymbol{\theta^*}$ is accepted → $\boldsymbol{\theta^{t+1}} = \boldsymbol{\theta^*}$.
- If $u > \alpha(\boldsymbol{\theta^t}, \boldsymbol{\theta^*})$, $\boldsymbol{\theta^*}$ is rejected (there is no move) → $\boldsymbol{\theta^{t+1}} = \boldsymbol{\theta^t}$.

The standard choice of $\alpha(\boldsymbol{\theta^t}, \boldsymbol{\theta^*})$ which minimizes the probability of rejection is:

$$\alpha(\boldsymbol{\theta^t}, \boldsymbol{\theta^*}) = \min\left\{ \frac{\pi(\boldsymbol{\theta^{t+1}}|\boldsymbol{y})q(\boldsymbol{\theta^{t+1}}, \boldsymbol{\theta^t})}{\pi(\boldsymbol{\theta^t}|\boldsymbol{y})q(\boldsymbol{\theta^t}, \boldsymbol{\theta^{t+1}})} \right\}.$$

STEP 3. Return to *STEP 1.*

It is worth noting that Metropolis Hastings algorithm updates all the parameters at the same time, however, parameters can be updated one at a time, it is called single-update Metropolis Hastings (for more details see King et al., 2010).

Gibbs sampler

The Gibbs sampler is a special case of Metropolis Hastings algorithm, where the acceptance probabilities are equal to 1, so that the acceptance-rejection step can be omitted. In the Gibbs sampler algorithm the proposal distribution for any parameter is the conditional posterior distribution of that parameter given the current value of the others. The sampling is done successively from each conditional distribution, in order.

Given the parameters $\boldsymbol{\theta} = \{\theta_1, \ldots, \theta_p\}$, with distribution $\pi(\boldsymbol{\theta})$, the Gibbs sampler uses the set of full conditional distributions of π to sample from the marginal distributions $\pi(\theta_i|\boldsymbol{\theta_{(i)}})$, being $\boldsymbol{\theta_{(i)}} = \{\theta_i, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_p\}$. The algorithm can be described as:

STEP 0. We start with an arbitrary vector $\boldsymbol{\theta^0} = \{\theta_1^0, \ldots, \theta_p^0\}$.

STEP 1. At each iteration t simulate:

- θ_1^{t+1} from sample $\pi(\theta_1|\boldsymbol{\theta^t_{(1)}})$;
- θ_2^{t+1} from sample $\pi(\theta_2|\boldsymbol{\theta^t_{(2)}})$;

$$\vdots$$

- θ_p^{t+1} from sample $\pi(\theta_p|\boldsymbol{\theta^t_{(p)}})$.

STEP 3. Return to *STEP 1.*

Finally, note that conditional distributions may (or may not) have an standard form. If they do have it, the implementation of MCMC methods are straightforward. However, when conditional distributions are non-standard it can be used the Metropolis Hastings algorithm. It is called Metropolis-within-Gibbs (for more details see Tierney, 1994).

1.5.3 Computing

The software most frequently used to implement MCMC algorithms in Bayesian framework are `WinBUGS` and `OpenBUGS` (Lunn et al., 2000), `JAGS` (Plummer, 2003), `STAN` (Carpenter et al., 2017), `NIMBLE` (Valpine et al., 2017), and `BayesX` (Umlauf et al., 2015). All the inference presented in this thesis was performed with `JAGS` software, although `WinBUGS`, `OpenBUGS`, `STAN` and `NIMBLE` have also been used. Further, MCMC methods can also be programmed in computational languages such as `R` (R Core Team, 2019) or `C++` (Stroustrup, 2013).

Additionally, the capture-recapture(-recovery) models reviewed in this Chapter (and many others) can be implemented in comprehensive software packages such as the program `Mark` (White and Burnham, 1999), `Presence` (that estimates patch occupancy rates), `M-SURGE` (for Multistate models), `E-SURGE` (Mutievent models) and `Density` (for spatial explicit capture-recapture models). There are also available `R` packages as: `RMark` (Laake, 2013), a convenient `R` package with interface to program `Mark`; `marked` (Laake et al., 2013), an `R` package for classical and Bayesian analyses of capture-recapture data; `secr`, an `R` package for spatially explicit capture-recapture; `Rcaptured`, that estimates demographic parameters using log-linear models; `BaSTA` (Colchero et al., 2012), that performs age-specific survival analysis from incomplete capture-recapture(-recovery) data in Bayesian framework; `unmarked` (Fiske and Chandler, 2011), that models variation in abundance for capture-recapture data; and `multimark` (McClintock, 2015), that analyses capture-recapture data using non-invasive marks.

Finally, some web pages provide useful `R` code examples to fit capture-recapture(-recovery) models such as: `www.capturerecapture.co.uk` (based on the book by McCrea and Morgan, 2015); the web page: `www.vogelwarte.ch/de/projekte/publikationen/bpa/`, is based on the book by Kéry and Schaub, 2011 and provides the necessary `R` code to implement these models both in `WinBUGS` and `JAGS` languages; and lastly, `www.montana.edu/rotella/502/Schedule.html`, gathers several and comprehensive examples of `RMark` code.

CHAPTER 2

ANALYSING THE SURVIVAL OF COMMON GUILLEMOT: THE CASE OF STORA KARLSÖ

The data used along this thesis come from the largest colony of a seabird species, the common guillemot (*Uria aalge*). When we first met the researchers from the Baltic Seabird Project (`http://www.balticseabird.com`), they reported us their interest on the analysis of juvenile survival of the common guillemot colony in Stora Karlsö (Sweden). Previous studies in this colony had already performed analysis of adult survival probabilities but more importantly, they told us about an special characteristic of this colony. In contrast with other colonies, in Stora Karlsö a 'big' proportion of young birds (1 and 2 years old individuals) were resighted. Furthermore, they had an extensive ringing protocol that started at 1913, but with an annual ringing from the 40s, with recaptures taking place from year 2002. The size of the resulting capture-recapture database (uncommonly big) as well as the topic of interest (only less than 2% of seabird species have juvenile survival estimates) encouraged us to start this project.

Nevertheless, we found several problems regarding the size of the database and its study design. In particular, after the first analysis was performed (Chapter 3), we realised that parameters estimated from the capture-recapture database were biased in older age classes due to a partial monitoring problem at breeding ledges (explained in detail in Chapter 5). In ecological studies it is often the case that only a fraction of the colony is monitored, and these monitored areas are assumed to be representative of the whole population. However, this assumption does not held true when partial monitoring is present. To overcome this problem we needed information about the ‘invisible’ individuals, i.e. individuals settled in non-monitored areas. Hence, we incorporated into the analysis a second data set, in particular, a ring-recovery data set from the same individuals. This ring-recovery database provided information about individuals recovered dead all over the island, and so, that information shed light to the individuals settled in non-monitored ledges.

The main objective of this Chapter is twofold. First, we introduce the species of interest (the common guillemot) as well as we explain in detail the colony of Stora Karlsö. The knowledge of some of the main ecological characteristics of this species is relevant in order to to implement the statistical models. Second, we present a detailed description of the two data sets above mentioned.

2.1 The species

Seabirds are top predators in marine ecosystems. Their populations tend to reflect conditions over large spatial and long term scales, making them bioindicators of environmental change. Adult survival is an important life-history trait in long-lived species and determinant for the sustainability of the populations. Juvenile survival is also a very important but unknown characteristic due to seabirds have long periods of immaturity spending several years at sea before returning to colonies when they reach the sexual maturity (Croxall and Rothery, 1991). Consequently, understanding survival probabilities is crucial for modelling population dynamics. Indeed, changes in survival over time could provide insight into the possible effects of changes in climate and anthropogenic perturbations.

The common guillemot (also called common murre) is a long-lived, pursuit-diving and widespread in the Northern hemisphere, being one of the few auk species that breed in both Atlantic and Pacific oceans (Nettleship and Birkhead, 1985). This species is one of the most widely studied seabirds and its monitoring is part of many national monitoring programmes.

Historically, this species has been considered as an important source of eggs and meat for humans. Due to this fact, the species was almost extinct at the end of the 19th century in the Baltic Sea (only $\approx$20 pairs remained, Olsson et al., 2000). At present, the main threats affecting guillemots vary from legal hunting in Canada and Greenland (Merkel and Barry, 2008), oil pollution (Hatchwell and Birkhead, 1991; Votier et al., 2008), incidental captures in fishing gear (bycatch) (Österblom et al., 2002), overfishing (Österblom et al., 2006) and climate change (Sandvik et al., 2005, 2012).

The common guillemot (hereafter guillemots) forage for food by swimming underwater using its wings for propulsion, usually diving to depths of 20-50m (although depths up to 150 m have been recorded). It is a highly specialized feeder and a single-prey loader. During the chick rearing, only a single prey item is brought to the chick by parent. In the Baltic Sea, its main prey is sprat (*Spratus spratus*) (Hedgren and Linnman, 1979; Österblom et al., 2006).

Seabirds, and in particular the guillemots, are philopatric to both the breeding colony and the sub-colony where they were born (Halley et al., 1995; Harris et al., 1996a,b). Further, it is a social species and usually breeds in large colonies at very high densities, being the main advantage of this high density nesting an anti-predator strategy. There have been recorded a variety of breeding areas: low-lying islands and stacks, broad and narrow cliff ledges and occasionally under boulders and in caves. However, maximum densities are achieved on broad, flat areas (Parslow, 1966; Birkhead and Hudson, 1977a). This species has a complex life history with a long prospecting behavior that caused that its population modelling is challenging.

The guillemots do not make nest and lay only one egg that incubate for $\approx$32 days. The fledging age is around 20 days (Hedgren and Linnman, 1979; Nettleship and Birkhead, 1985). They are sexually monomorphic and both sexes invest heavily in reproduction, being after fledging when the roles of both sexes are remarkable different. Indeed, the male parent takes the chick to sea and feeds it for many weeks until it completes its development; conversely the female continues to visit the breeding site (Wanless and Harris, 1986).

FIGURE 2.1: Guillemot chick at fledging time (left) and adult incubating its egg (right).

The island of Stora Karlsö, west of the island of Gotland (Sweden), holds the largest guillemot colony on the Baltic Sea, with 15,700 pairs in 2014, i.e. $\approx$2/3 of the Baltic Sea population (Olsson and Hentati-Sundberg, 2017). The main threats that affect guillemots in this area are bycatch, overfishing and climate change. In fact, large-scale and long-term changes in the ecosystem have been recorded in the Baltic Sea which have affected the abundance of sprat, the main prey for guillemots in this area (Hedgren, 1976; Lyngs and Durinck, 1998; Österblom et al., 2006).

To sum up, seabirds have been proposed as valuable indicators of changes in marine ecosystems. Nowadays when climate change arises as a crude reality, the species will have to show their ability to adapt in order to face the currently changing environmental conditions. Furthermore, specialist species (as the guillemots) are particularly challenged by rapid and directional changes. Therefore, the study of this species turns out of special relevance in order to understand how these changes are affecting specialist species keeping always in mind an ultimate goal: to conserve them.

2.2 Available information

This Section is devoted to present a detailed description of the two data sets used along this thesis: a capture-recapture database that comprises years 2006 to 2016, and a mark-recapture-recovery (MRR) database regarding years 1992 to 2018.

2.2.1 Capture-recapture database (2006-2016)

Within the period 2006 to 2016, a total of $28,930$ chicks were ringed (with metallic and plastic rings) in late June and early July after jumping from the ledges at an age of 20 days. Mean ringing effort was $2,893$ by year, with a maximum of $4,956$ individuals ringed (related to year 2015) and a minimum of $1,965$ individuals ringed (related to year 2007). Along the study period, a total of $7,625$ resightings were made in the reproductive period (May to July) with minimal disturbance. Resightings were performed in five different areas of the island, three breeding ledges and two clubs (aggregations of birds on rocks in the water beneath the breeding ledges) (Figure 2.2). The mean resigting effort was 82.80 hours by year, with a maximum of $1,284$ individuals resighted (related to year 2016) and a minimum of 174 individuals resighted (year 2007). The total number of individuals resighted corresponds to $5,493$ different individuals (19% of the ringed birds).

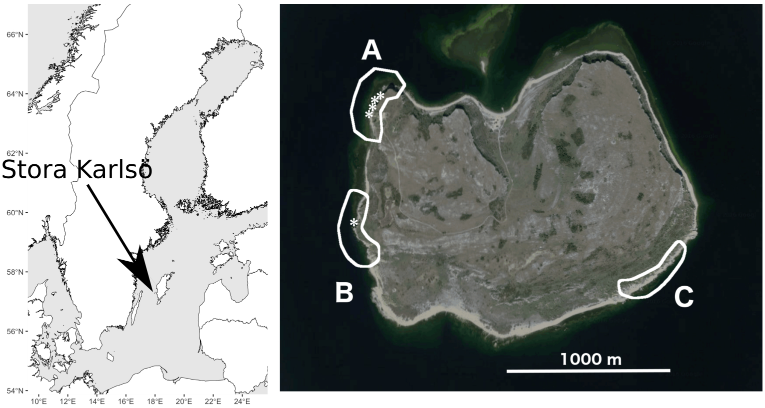

FIGURE 2.2: Location of the large guillemot (*Uria aalge*) breeding colony on the island of Stora Karlsö in the Baltic Sea (left), where chicks were captured, banded, and resighted from 2006-2016. Map of the island (right) shows the location of the main breeding ledges used by guillemots (areas A, B and C) and locations that were monitored to resight marked birds (indicated with asterisks). Chicks were captured and banded in area A.

Figure 2.3 on the left shows the annual number of resightings, while Figure 2.3 on the right displays the annual number of ringed individuals within the study period 2006-2016. As it can be appreciated, there has been an increasing effort along the study period both in the number of resightings and the number of ringed birds.

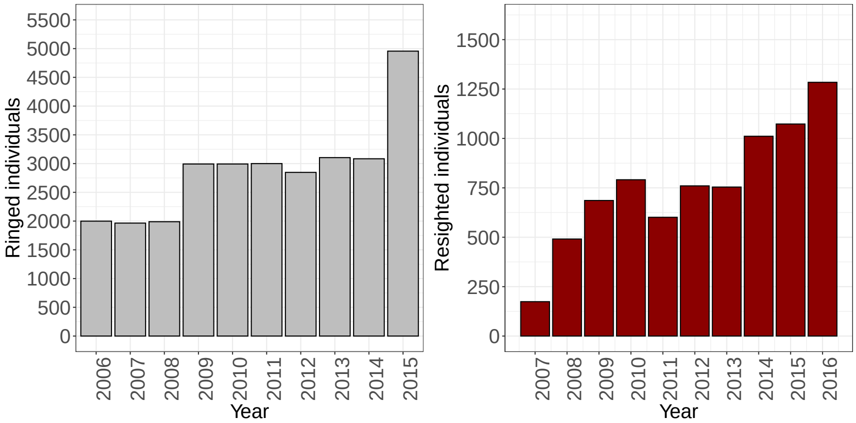

FIGURE 2.3: Annual number of ringed (left) and resighted individuals (right) in the capture-recapture database (2006-2016).

Furthermore, Table 2.1 presents the annual number, the cumulative number of the ringed individuals and the resighted individuals in this capture-recapture database. The proportion of marked individuals observed one or more times over the study period was $(5,493/28,930)= 0.1898$, and consequently about the 81.10% of the ringed individuals have never been seen again. Note that this high value of never seen individuals is common for capture-recapture studies.

Year	**Ringed**	**Cumulative ringed**	**Resighted**
2006	1,999	1,999	–
2007	1,965	3,964	174
2008	1,989	5,953	491
2009	2,993	8,946	686
2010	2,993	11,939	791
2011	3,000	14,939	601
2012	2,847	17,786	760
2013	3,104	20,890	754
2014	3,084	23,974	1,011
2015	4,956	28,930	1,073
2016	–	–	1,284

TABLE 2.1: Ringed individuals, cumulative number of ringed individuals and resighted individuals from 2006 to 2016.

As it can be appreciated in Figure 2.4, the most resighted age is two years old (36% of the resighted individuals), followed by three years old (25%), four years old (14%), five-ten years old (13%) and one year old (12%). It is worth noting that in this case an individual can be included at different ages. The wide-spread colony attendance of 1 and 2 year old birds in Stora Karlsö colony is an aspect which appears unique among guillemot colonies.

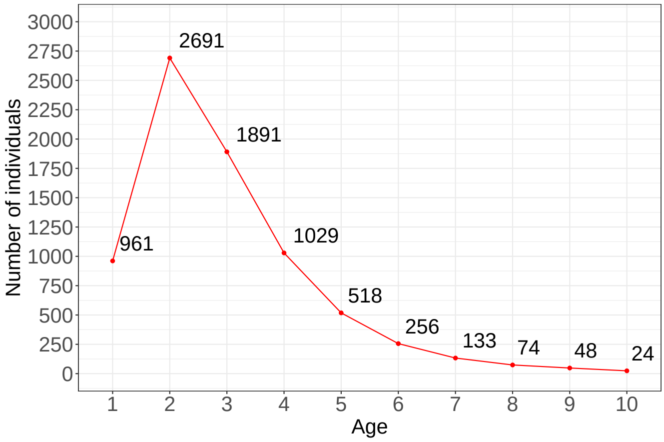

FIGURE 2.4: Absolute number of resighted individuals by age in the capture-recapture database (2006-2016).

Note that older ages (from 5 to 10 years old) do not seem to be represented in this capture-recapture database, the reason underneath being that the study was designed for the analysis of juvenile survival probabilities. As previously mentioned, the study was performed on five different areas of the colony (Figure 2.2), where resightings with telescopes and binoculars were made from above. Along that interval of time, observers tried to read as many plastic or metal rings as possible, resighting mostly juveniles (until age 4). This is due to experienced breeders (5+ aged individuals) are usually settled in the 'best areas', i.e. the central areas of the ledges, making their resighting difficult. On the contrary, the youngest birds are usually at peripheral areas which are unsuitable for breeding but are really helpful for resightings.

The maximum of times that an individual has been observed in this capture-recapture database is ten times. However, as it can be appreciated in Figure 2.5, the major part of the individuals have been never resighted, followed by once, two times, three, . . ., up to those ten times. Further, each individual was observed an average of 1.3881 times.

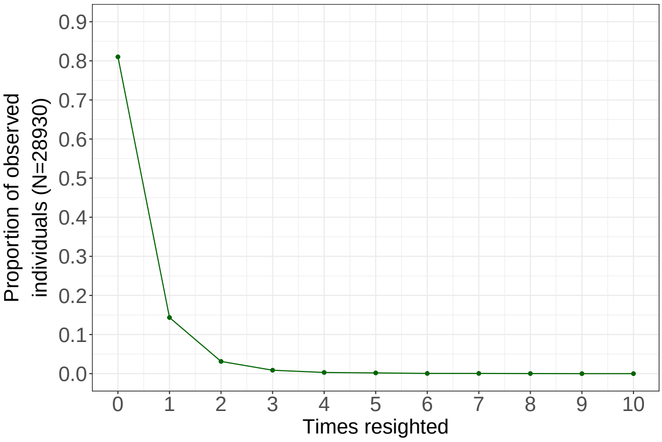

FIGURE 2.5: Proportion of individuals resighted from zero to ten times in the capture-recapture database (2006-2016).

In order to get insight about their age when animals were last resighted, Figure 2.6 displays the absolute number of individuals seen at their last resighting occasion by age, while Table 2.2 shows the proportion of individuals seen at their last resighting occasion by age. Again, age two is the most observed at the time of their last resighting occasion, followed by three, four and one. Moreover, the 87% of the individuals were last resighted at an age of 1-4 years. These results are in line with those presented in Figure 2.4.

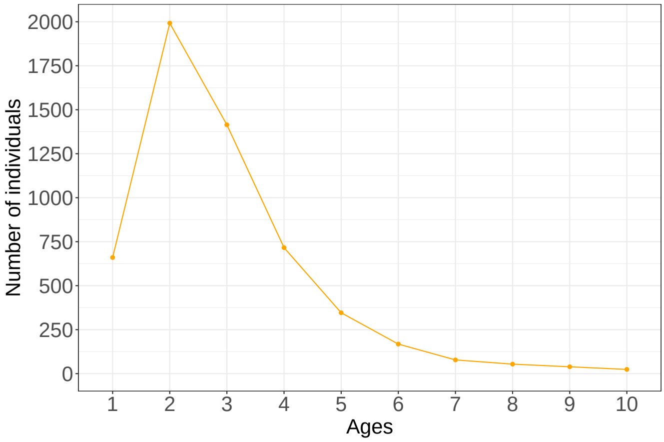

FIGURE 2.6: Age of the individuals at their last resighting occasion in the capture-recapture database (2006-2016).

Age	**Proportion of individuals** ($N = 5493$)
1	0.1202
2	0.3627
3	0.2575
4	0.1304
5	0.0630
6	0.0305
7	0.0142
8	0.0098
9	0.0071
10	0.0043

TABLE 2.2: Age and proportions of individuals at their last resighting occasion in the capture-recapture database (2006-2016).

To sum up, this capture-recapture database contains information about $28,930$ individuals ringed as chicks and $5,493$ resighted individuals along the period 2006-2016. Further, this guillemot colony from Stora Karlsö has the unique feature that many 1 year old individuals return to the colony (12%), being the most resighted age two years old. Moreover, this database gathers scarce information about the adult classes (from 5 to 10 years old) mainly due to study design. To conclude, nowadays databases are becoming more complex in terms of the volume and variety of the data, and this is the case for this capture-recapture database, whose unusual (big) dimensions made the inference challenging.

2.2.2 Mark-recapture-recovery (MRR) database (1992-2018)

Integrated models are based in the joint analysis of different databases. In this thesis, we present an integrated model where an extended version of the previous capture-recapture database is jointly analyzed with a ring-recovery database from the same individuals.

The ringing program in Stora Karlsö has a long history, with the first ringing campaign performed in 1913 (Olsson et al., 2000). However, over the first years a low number of birds were ringed as well as at some years no ringing was performed. Therefore, and due to the above exposed, our MRR database comprise the period 1992-2015 and contains information about $39,681$ individuals ringed as chicks (Figure 2.7 on the right).

In ring-recovery studies, it is often the case that a small proportion of the individuals are found dead. Indeed, dead animals may be recovered physically but, most of the times, is the ring itself which is recovered. Further, far from all recoveries are reported, specially those from animals caught in fishing gears, one of the main threats for this species in the Baltic Sea (Žydelis et al., 2009). All the information about dead individuals is frequently stored in national recovery data sets. Indeed, for this colony, the recovery information has been provided by the Bird Ringing Centre (Swedish Museum of Natural History).

For this integrated database, the ring-recovery data gathers information about 957 individuals recovered dead along all the study period 1993-2018 (Figure 2.7 on the left), while the capture-recapture data includes recaptures events between 2002-2016, with a total of $6,477$ unique individuals observed (Figure 2.7 on the left). Therefore, within the period 1993 to 2001 and 2017-18 no resightings were performed. Finally, the number of individuals seen at both databases (either resighted alive or recovered dead) were 47.

As it can be appreciated in Figure 2.7 on the left, during the capture-recapture study period there were two different observational regimes. From 2002-6 resighting effort was generally low (mean seasonal resighting effort of 35.80 hours by year), and significantly lower than the effort in later years between 2007-16 (mean seasonal resighting effort of 82.80 hours by year).

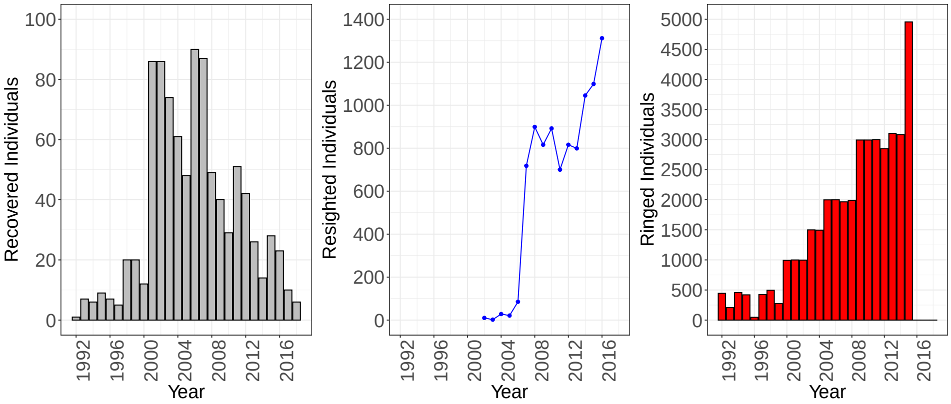

FIGURE 2.7: Annual number of recovered individuals (grey bars, on the left), resighted individuals (blue line, on the centre) and ringed individuals (red bars, on the right) in the MRR database.

Figure 2.8 displays the number of individuals recaptured by age in this capture-recapture database. Note that, although this data set contains ‘extra’ information due to years 2002-2006, the pattern is equivalent to that observed in the previously presented capture-recapture database (Figure 2.4). Indeed, the age most resighted is two years old (34% of the resighted individuals), followed by three years old (27%), four years old (14%), five-ten years old (13%) and one year old (12%) (Figure 2.8). Similarly, the maximum of times that an individual has been observed is again ten times and the proportion of never seen individuals is the biggest one (corresponding to $33,204$ individuals), followed by once, two, $\ldots$, up to those ten (Figure 2.9). Further, the number of times that one individual is observed is slightly bigger, 1.4269 times (compared to 1.3881).

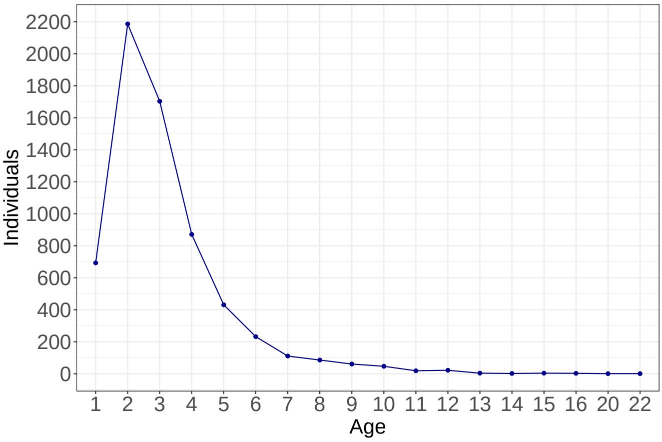

FIGURE 2.8: Number of individuals resighted by age in the MRR database.

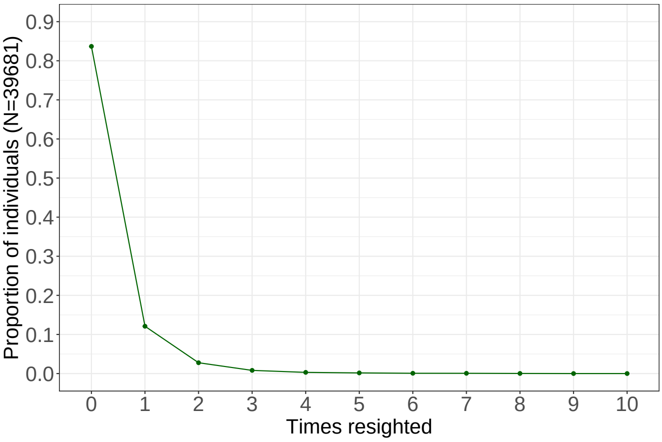

FIGURE 2.9: Proportion of individuals resighted from zero to ten times in the MRR database.

In the ring-recovery database (as expected) only a small proportion of individuals were recovered dead (2.41%). Further, as it can be appreciated in Figure 2.10, most of the recovered individuals were 1 year old birds (61%), followed by 2 years old (19%), 3 years old (6%), . . ., up to 19 years old. These results are in line with previous studies that state that in many species, as seabirds, survival at first year of life is lower than later (Harris et al., 2007).

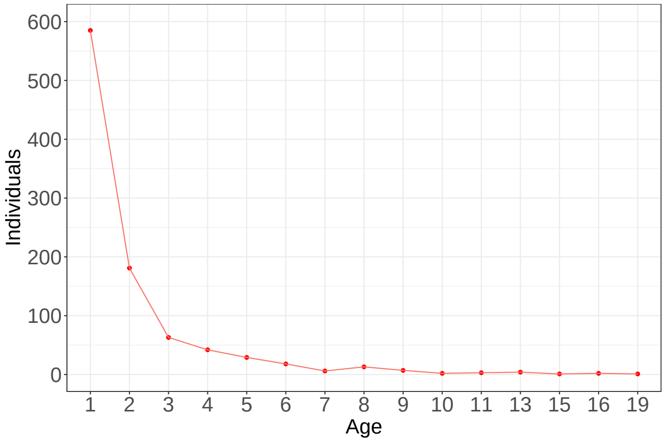

FIGURE 2.10: Absolute number of recovered individuals by age.

Additionally, and similarly to what happened in the capture-recapture data set presented in previous Section, this ring-recovery database gathers scarce information about adult individuals (5+) (Figure 2.10). Indeed, only the 13% of the individuals recovered dead were 5-19 years old. The underneath reason is not now related to the study design, it is much simpler: guillemots are long-lived species. The oldest guillemot seen alive was resighted in this colony and it was 47 years old (Swedish Museum of Natural History). For long-lived species, adult survival is considered an important trait for the maintenance of the populations. Indeed, in natural circumstances, the probability that an adult guillemot dies is very low (adult survival estimates vary between $[0.9, 1]$, Harris et al., 2000). Further, the guillemot colony from Stora Karlsö has almost tripled from 1980 until present (Hentati-Sundberg and Olsson, 2016; Olsson and Hentati-Sundberg, 2017). Therefore the fact that adult ages are not represented in the ring-recovery database is what we expected in these circumstances of colony increase.

In conclusion, capture-recapture and ring-recovery databases gather different temporal information. Whereas ring-recovery data set contains information about all the study period (1992-2018), capture-recapture information is limited to years

2002-2016, with a different recapture effort between years 2002-06 and 2007-16. Further, we have shown that this capture-recapture database has a similar pattern that the corresponding to years 2006-2016. Finally, it is worth pointing out that while recaptures are performed in the monitored areas (presented in Figure 2.2), recoveries (for this database) take place at colony level (monitored and not monitored areas), the Baltic Sea and Atlantic Ocean.

CHAPTER 3

SUBJECTIVE BAYESIAN JUVENILE SURVIVAL ANALYSIS

Up to this point, all the tools required to develop this thesis have been described. Next Chapters are devoted to give a more detailed explanation of the studies performed, giving an introduction to the problem, describing concisely the methodology employed, and finally depicting the results.

In particular, in this Chapter we present our first analysis performed in the guillemot colony from Stora Karlsö. To that end, we used the capture-recapture database explained in Chapter 2 (Section 2.2.1). The main objective of this work was to provide reliable juvenile survival and recapture probabilities estimates by means of Cormack-Jolly-Seber models. The novelty of this work lies in the use of a subjective Bayesian approach to try to correct the partial monitoring problem detected at breeding ledges. Indeed, we incorporate relevant information (expert opinion) about recapture probabilities for the oldest age class via subjective prior distribution.

This study depicts a first approximation to the problem of partial monitoring. Nevertheless, in Chapter 5 we go further, and we present an integrated model that reduces the corresponding bias obtained in the estimated survival probabilities from capture-recapture data only.

3.1 Introduction

It is well known that seabirds are of particular conservation concern due to their vulnerability to human activities, including climate change, bycatch, invasive species and oil spills (Sandvik et al., 2005; Martínez-Abraín et al., 2006; Lewison et al., 2012). According to the IUCN Red List of Threatened Species, seabirds are the group of birds more threatened, having the 43% of the seabird species some category of threat (Crowall et al., 2012).

In long lived species, such as seabirds, population dynamics are sensitive to changes in adult survival, reproductive success and juvenile survival (Lewison et al., 2012). For seabirds, estimation of age or stage- specific demographic parameters is particularly important because decline in breeding populations can be influenced by non-breeders and juvenile birds (Votier et al., 2008). However, seabirds often have long periods of immaturity, remaining unobservable at sea for several years until they come back to the colony (Croxall and Rothery, 1991; Sandvik et al., 2008), making estimation of the juvenile stage difficult (Lee et al., 2008; Votier et al., 2008). In particular, less than 2% of seabird species have estimates of juvenile survival (Lewison et al., 2012).

The guillemot colony from Stora Karlsö (Baltic Sea, Sweden), in contrast with other colonies, has the unique feature that many one-year-old birds return to the colony (12%). This colony shows an important behavioural heterogeneity, involving several phases of club attendance, ledge attendance, pair bonding and ledge establishment. Indeed, in this colony, one year old individuals are almost exclusively resighted at clubs (flat boulders 2-5 m in diameter beneath the breeding cliffs in the water 10-20 m off the shore, holding aggregations of non-breeding birds), while at the following ages they gradually shift over to almost only being observed at breeding ledges. Other studies have reported this behaviour heterogeneity in several colonies (Nettleship and Birkhead, 1985; Halley et al., 1995) but so far few studies have considered it as a factor in survival models (Crespin et al., 2008; Lee et al., 2008; Votier et al., 2008).

Additionally, age is an important factor affecting survival and recapture probabilities in animal populations (Crespin et al., 2006; Harris et al., 2007; Morrison et al., 2009) and commonly, in most species, the first-year survival is lower than in later ages (Harris et al., 2007). In order to incorporate any possible differences in survival along age classes, in this work we used Cormack-Jolly-Seber (CJS) model with age effects to jointly estimate survival and recapture probabilities of juvenile guillemots, always taking into account that CJS models estimate *apparent* (instead of *true*) survival probabilities, as dead and permanent emigration is confounded (King et al., 2010; Kéry and Schaub, 2011). Finally, we modelled recapture probabilities as age and time dependent in order to take into account different detectabilities caused by the different rings used along the study.

A strong assumption of capture-recapture models is that animals remain faithful to the study area (Newman et al., 2014). Seabirds, and in particular the guillemots, are philopatric to the breeding colony, but also to the sub-colony where they were born (Halley et al., 1995; Harris et al., 1996a,b). As juvenile birds get older, they are increasingly present on breeding ledges, reflecting the prospecting behaviour in preparation for breeding (Oro and Pradel, 2000; Kadin, 2007). Hence, individuals hatched on the monitored breeding ledges will have a higher recapture probability than birds hatched outside them.

To deal with the partial monitoring problem, in this Chapter we used a subjective Bayesian approach in birds older than 2 years old incorporating former knowledge about the proportion of breeding birds in monitored ledges in the area of study (Olsson and Hentati-Sundberg, 2017). The elicitation of informative prior distributions has been (and still is) a controversial issue in Bayesian inference (King et al., 2010; Golchi, 2016). However, when relevant information is available, informative prior distributions are the appropriate way of introducing expert opinion or information related to the study (McCarthy and Masters, 2005; King et al., 2010; Christensen et al., 2011; Martínez-Abraín et al., 2014; Golchi, 2016).

3.2 Database

Between 2006 and 2015, $28,930$ chicks were marked in late June and early July on a beach under the cliffs from the largest sub-colony of the island. Throughout the study, it was used the official triangular stainless steel ring with a unique ID code on one leg, provided by the Ringing Centre at the Swedish Museum of Natural History and manufactured by IÖ Mekaniska, Bankeryd, Sweden. Additionally, were used the following complementary rings with IDs and colour codes on the other leg: round plastic rings (Protouch Engraving and Signage, Saskatoon, Saskatchewan, Canada) 2006-2008 and 2013-2015, round aluminium rings (IÖ Mekaniska, Bankeryd, Sweden) 2009 and large triangular steel rings 2010-2012 (IÖ Mekaniska, Bankeryd, Sweden). All types of rings were possible to read with a telescope at a distance at all monitoring sites.

Resightings were made in 10 consecutive breeding seasons (from 2007 to 2016) and in five different locations of the colony: two clubs and three breeding ledges (Figure 3.1). During the breeding period (May to July) resightings were performed with telescopes from above, resulting in minimal disturbance. We determined the birds' ID from both the official stainless steel rings and the complementary rings. Mean seasonal resighting effort was 82.80 h/yr, with a maximum of 138 h/yr (year 2015), and a minimum of 57 h/yr (year 2008).

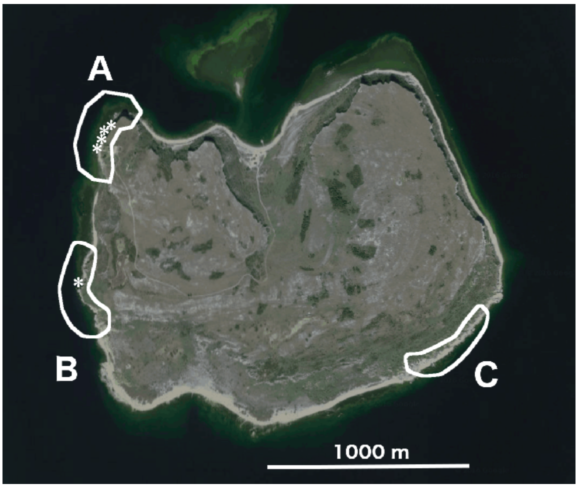

FIGURE 3.1: Island of Stora Karlsö with the location of the main breeding ledges used by guillemots (areas A, B and C) and locations that were monitored to resight marked birds (indicated with asterisks). Chicks were captured and banded in area A.

3.3 Survival Modelling

We estimated annual survival and annual resighting probabilities using the well-known multinomial formulation for computational reasons due to the big size of the database (Lebreton et al., 1992; Kéry and Schaub, 2011; King et al., 2010). Our main aim was to establish robust, stage-specific demographic parameter estimates for the colony. Although survival can be year dependent (due to variation in food availability or extreme weather events; Votier et al., 2008; Harris et al., 2007), in this study we modelled survival probability age dependent due to previous studies performed in this colony suggested that the population trajectory is both stable and positive (Hentati-Sundberg and Olsson, 2016; Olsson and Hentati-Sundberg, 2017). However, the use of different quality of the complementary rings (plastic/aluminium/steel) during the study could affect annual resighting probabilities, therefore, we modelled recapture probability as age and time dependent (in logistic scale).

We initially describe the model parameters and notation before explaining the inferential process. Let be i the marked individuals, $i = 1, \ldots, N$, and t the capture occasions, $t = 1, \ldots, T$, being each one of these values related to the study years (2006-2016). Let $\boldsymbol{f}$ be the vector that indicates the year in which an individual enters into the study (marking occasion), being $f_i \in \{1, \ldots, T-1\}$, and $\boldsymbol{f} \in \{f_1, \ldots, f_N\}$. $\phi_{a,t}$ represents the *apparent* survival probability, and it is the probability that an individual i survives at age a from time t to $t+1$. In this study we consider survival probability is only age dependent, not considering differences between individuals or years. Let $p_{a,t}$ be the recapture probability of individual i at time t and at age a, given alive at time t. Due to the different rings used along the study, we model recapture probabilities age and time dependent (in logit scale). Therefore, if we denote by $a = 1, \ldots, A$, the age classes, the survival and recapture probabilities can be expressed as:

$$\begin{aligned} \phi_{a,t} &= \phi_a, \quad &&\text{for } a = 1, \ldots, A; t = f_i, \ldots, T-1; \\ \text{logit}(p_{a,t}) &= \alpha_a + \beta_t, \quad &&\text{for } a = 2, \ldots, A; t = f_i + 1, \ldots, T. \end{aligned}$$

In this study we establish four age classes, $a = 1, 2, 3, 4+$, being the underneath reasons mainly two. First, the main focus of this study is to analyze juvenile survival and recapture probabilities. And second, in this colony individuals aged $3+$

became less resighted due to their movement to unobserved (i.e. not monitored) areas of the colony.

The full set of parameters regarding to the recapture probabilities are: $\boldsymbol{\alpha_a} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_{4+}\}$, corresponding to one year old, two years old, three years old and 4+ years old individuals; and $\boldsymbol{\beta_t} = \{\beta_1, \beta_2, \ldots, \beta_{10}\}$, corresponding to the study years 2007-2016. However, due to in this colony individuals 3+ are primary observed at breeding ledges, we set $\alpha_3 = \alpha_{4+}$, which we denote by α_{3+}. Further, time dependent CJS models have identifiability problems in the last two parameters (King et al., 2010). Therefore, in order to avoid this lack of identifiability we set $\alpha_1 = 0$. Alternatively, the full set of parameters regarding the survival probabilities are $\boldsymbol{\phi} = \{\phi_1; \phi_2; \phi_3; \phi_{4+}\}$, corresponding to one, two, three, and 4+ years old individuals.

3.4 Posterior Inference

Bayesian statistics allows to incorporate not only the data but also all the available prior knowledge about the unknown parameters into the inferential process. This information needs to be expressed in probabilistic language in the so-called prior distributions. Bayes' theorem combines both types of information and provides the posterior distribution, which contains all the relevant knowledge about the parameters of interest (McCarthy and Masters, 2005; King et al., 2010; Christensen et al., 2011; Kruschke, 2011).

Most of the time, Bayesian inferential processes are based in vague or non-informative prior distributions, the reason being that no information is available. Nevertheless, as stated in Martínez-Abraín et al. (2014), "ecologists should make a greater effort to make use of available prior information because this is their most legitimate contribution to the inferential process". In other words, when information about the parameters is available, we should try to incorporate it.

In our case, we incorporated expert prior information about the partial monitoring of this colony in the CJS model. Based on a recent census of the colony, Olsson and Hentati-Sundberg (2017) stated that the size of the monitored breeding ledges represented 19% of breeding pairs in the A area (Figure 3.1). Furthermore, individuals of age class three were mainly observed at breeding ledges, not at the clubs, due to the gradual settlement of the prospecting individuals on breeding ledges (Kadin (2007); this study). Therefore, we used informative prior distributions but only for the recapture probabilities for individuals three-ten- years old (p_{3+}), as we used non-informative prior distributions for the remaining parameters of the model, in particular, Uniform $(0, 1)$.

The probability of seeing an individual that is alive, $\mathrm{P}(p_{i,t} = 1|z_{i,t} = 1)$, can be expressed as:

$$\begin{aligned}\mathrm{P}(p_{i,t} = 1|z_{i,t} = 1) = {} & \mathrm{P}(p_{i,t} = 1, M|z_{i,t} = 1) + \ \mathrm{P}(p_{i,t} = 1, \overline{M}|z_{i,t} = 1) = \\ & \mathrm{P}(p_{i,t} = 1|M, z_{i,t} = 1) \times \ \mathrm{P}(M|z_{i,t} = 1) + \\ & \mathrm{P}(p_{i,t} = 1|\overline{M}, z_{i,t} = 1) \times \ \mathrm{P}(\overline{M}|z_{i,t} = 1) = \\ & \mathrm{P}(p_{i,t} = 1|M, z_{i,t} = 1) \times 0.19.\end{aligned}$$

Where the event M indicates that the individual is in a monitored breeding ledge and $\overline{M}$ indicates the opposite. Additionally, even though a bird is present at a ledge, it is not certain that its ring could be read, mainly because individuals are standing close to each other and not always showing their rings. Anyway, in order to be the less informative, we considered that we could see all the individuals, which is far from reality, and so, the recapture probabilities should not be greater than 0.19. In other words, we used informative prior distributions that constrained the parametric space for the recapture probabilities in age class 3+ in order not to exceed that 0.19 value.

As usual in this context, the resulting posterior distributions containing all the information about the system have no closed expression, so we need numerical approaches to approximate them. A feasible possibility is to use Markov chain Monte Carlo (MCMC) methods (Robert and Casella, 2011). In this study, we used MCMC through the jags software (Plummer, 2003) implemented in the R software (R Core Team, 2019). We obtained the final approximate random sample using three chains, $200,000$ iterations, discarding the first $20,000$ and saving 1 of every 300 of the remaining for the memory storage purposes and to reduce auto correlation.

3.5 Results

One year old birds were almost exclusively resighted at the clubs and only a few (5%) on the ledges (Table 3.1). In the following age categories there was a clear trend that share of birds only seen at the breeding ledges were increasing drastically with age, observing 89% of 3-10 years old birds only at breeding ledges (Table 3.1).

Age	Clubs	Breeding ledges	Both
1	704 (95%)	32 (4.31%)	5 (0.68%)
2	1183 (52.77%)	930 (41.48%)	129 (5.75%)
3	285 (17.71%)	1306 (81.17%)	18 (1.12%)
4	57 (6.38%)	830 (92.95%)	6 (0.67%)
5	8 (1.65%)	477 (98.35%)	0 (0.00%)
6	4 (1.62%)	243 (98.38%)	0 (0.00%)
7	1 (0.78%)	127 (99.22%)	0 (0.00%)
8	1 (1.39%)	71 (98.61%)	0 (0.00%)
9	1 (2.17%)	45 (97.83%)	0 (0.00%)
10	1 (4.76%)	20 (95.24%)	0 (0.00%)

TABLE 3.1: Age and number (with percentage) of banded guillemots (*Uria aalge*) resighted at clubs, breeding ledges, and both areas in the large breeding colony on the island of Stora Karlsö in the Baltic Sea, Sweden.

In Figure 3.2 we present the simulated values of the marginal posterior distribution of the annual survival probabilities with their corresponding 95% symmetric credible intervals. Survival probabilities differed among age classes, being highest for age classes 2 and 3. As expected, survival at first year of life was the minimum estimate. Recapture probability showed interannual variations among years, but it is important to remark that all the age classes show the same variability pattern between years (Fig. 3.3). Resighting probability for 1 year old birds ($\boldsymbol{p_{1,t}}$: $t = 2, \ldots, T$) was extremely low, peaking at 2 years old ($\boldsymbol{p_{2,t}}$: $t = 2, \ldots, T$), and being around to the informative value provided for the older age class ($\boldsymbol{p_{3+,t}}$: $t = 2, \ldots, T$).

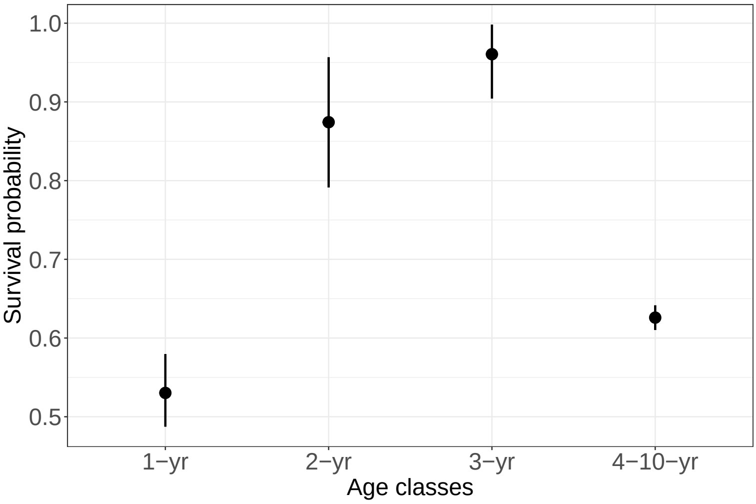

FIGURE 3.2: Mean and 95% symmetric credible intervals of the marginal posterior distribution of the *apparent* survival probabilities.

FIGURE 3.3: Mean and 95% symmetric credible intervals of the marginal posterior distribution of the recapture probabilities by age classes along the study period.

Furthermore, to illustrate the effect of the use of the informative prior distribution to recapture probabilities corresponding to $3+$ individuals, Figure 3.4 displays the simulated values of the three Markov chains regarding those parameters. A mere look at those chains shows that most of those parameters are in the upper limit of the informative prior distribution specified. With the use of informative prior distributions we are constraining the parametric space for recapture probabilities regarding 3+ years old individuals so that they can not be greater than 0.19 (i.e. our prior knowledge about the colony).

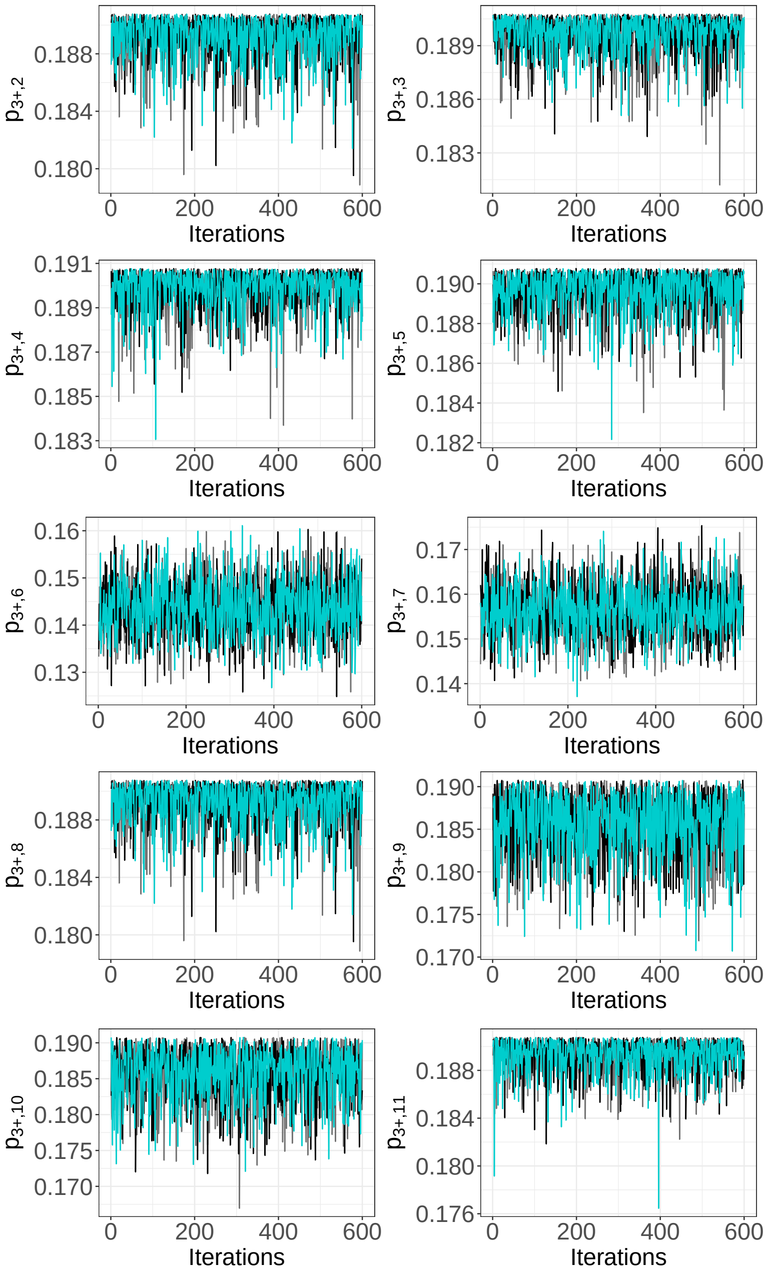

FIGURE 3.4: Trace plots of the simulated values of the three Markov chains corresponding to the posterior distributions of the recapture probabilities related to age class $a = 3+$ along the study period ($\boldsymbol{p_{3+,t}}$: $t = 2, \ldots, T$).

3.6 Discussion

Through the use of Bayesian CJS models, we provide new estimations of annual survival and resighting probabilities of young birds at the largest common guillemot colony in the Baltic Sea. The large number of ringed and resighted birds offered a robust database to obtain relevant inferences.

Unique among guillemot colonies, a number of birds returned to this colony in their first summer, i.e. when 1 year old (Kadin, 2007). Our results showed that 1 year old individuals were almost exclusively observed at clubs, but the following age classes gradually shift over to almost only being observed at breeding ledges. The clubs help youngest birds to learn social and sexual behaviours without the risk of being attacked by breeding birds as well as to gain experience in location of fishing grounds near the colony (Halley et al., 1995). We considered that resightings in clubs were not biased due to the fact that birds used these boulders randomly, that is, all the individuals that belonged to this sub-colony where we ringed them had the same probability of being resighted there. This assumption did, however, not held true in monitored breeding ledges. Birds that are more than one years old begin to prospect the breeding ledges in order to learn the competences for breeding (Halley et al., 1995; Harris et al., 1996b). Seabirds, and in particular the common guillemot, are philopatric to natal colony, but also to the sub-colony where they hatched (Halley et al., 1995; Harris et al., 1996a,b). Hence, individuals that settle in monitored breeding ledges will have a higher resighting probability compared to individuals settling in other ledges in the same sub-colony - risking underestimation of survival estimates (Lebreton et al., 1992; Sanz-Aguilar et al., 2016).

Keeping in mind differences with regard to data sources and statistical methodology, in Table 3.2 we compare our annual survival estimates with the ones obtained in other studies at the Isle of May (Scotland; Crespin et al., 2006; Harris et al., 2007; Reynolds et al., 2009), Skomer island (Wales; Birkhead and Hudson, 1977a; Votier et al., 2008; Meade et al., 2013) and Farallon island (California; Lee et al., 2008). As we can observe, our estimate of survival at first year of life is the lowest value reported for guillemots, whereas for 2 years old birds is the highest value reported. This may be caused by the fact that in the other studies no one year old individuals were seen alive, and so, inference in first year survival probability was done without that information. In our study, however, 12% and 35% of 1 year old and 2 years old individuals respectively were resighted. The fact that birds in our study had higher mortality during their first winter than during the second, confirms earlier studies that highlight the first winter as a critical period for common guillemots, as indicated e.g. as an over representation in bycatch recoveries (Olsson et al., 2000). Other comparable outputs are those for the survival up to 2 years old, that we can easily calculate combining survival of the 1 year and 2 years age classes. In particular, this calculated survival probability up to 2 years old is 0.34 (0.58^2) in Crespin et al. (2006), while it is 0.49 (0.70^2) in Votier et al. (2008), and it is 0.41 (0.54×0.76) in Reynolds et al. (2009). In our study, the posterior mean and 95% credible interval of surviving until 2 years old was 0.46 and [0.44,0.49], i.e. close to those estimations.

Colonies, Studies, Time period	Survival estimates	Age class	Banded chicks
Farallon Is., Lee et al. (2008). 1986-2004	0.59	1-yr	375
	0.67	2-yrs	
	0.75	3-yrs	
	0.80	4-yrs	
May Is., Crespin et al. (2006). 1983-2001	$0.58[0.44, 0.71]$	1-2-yrs	–
	$0.86[0.81, 0.91]$	3-4-yrs	
	$0.70[0.65, 0.73]$	5-18-yrs	
May Is., Harris et al. (2007). 1983-2002	$0.56[0.54, 0.59]$	1-yr	96-237 year^{-1}
	$0.79[0.72, 0.85]$	2-yrs	
	$0.92[0.87, 0.95]$	3-yrs	
	$0.94[0.87, 0.98]$	4-yrs	
	$0.97[0.94, 0.98]$	5+ yrs	
May Is., Reynolds et al. (2009). 1983-2003	0.54	1-yr	5,594
	$0.76[0.72, 0.81]$	2-yrs	
	$0.93[0.89, 0.96]$	3-yrs	
	$0.91[0.87, 0.94]$	4-yrs	
Skomer Is., Birkhead and Hudson (1977a). 1972-1977	0.26**	1-3-yrs	279
	0.21**	1-5-yrs	
Skomer Is., Votier et al. (2008). 1985-2005	0.70 ± 0.04	1-2-yrs	1,522
	0.95 ± 0.08	3-yrs	
	0.87 ± 0.06	4-yrs	
	0.78 ± 0.05	5-20-yrs	
Skoner Is., Meade et al. (2013). 1985-2004	0.43**	1-3-yrs	5,943
Stora Karlsö, this study. 2006-2016	$0.53[0.49, 0.58]$	1-yr	28,930
	$0.87[0.79, 0.96]$	2-yrs	
	$0.96[0.90, 0.99]$	3-yrs	
	$0.63[0.61, 0.64]$	4-10-yrs	

TABLE 3.2: Previously reported survival probability estimates for guillemots from various colonies (expressed in terms of the mean, the mean and 95% credible intervals, or the mean and SD; estimates are for annual survival probability except Skomer Island studies, which report cumulative survival probabilities).

With respect to 3 years old survival, our estimate was also in line with other studies. In particular, the posterior mean was close to the estimations obtained by Votier et al. (2008) and Reynolds et al. (2009). Also, and previously mentioned, in order to compare our results with those by Birkhead and Hudson (1977a) and Meade et al. (2013) (in which the estimations were for individuals until three years old), we have to consider the calculated posterior probability of the cumulative survival up to three years. Concretely, in our case the posterior mean and 95% credible interval of surviving until three years in Stora Karlsö colony was 0.44 [0.43,0.46], which was a similar result to the one from Meade et al. (2013).

Our final age class correspond to young adult birds (4-10 years). Adult survival in seabirds is the life-history characteristic of greatest importance for population growth and it is crucial for the colony maintenance (Sandvik et al., 2012). The study by Österblom et al. (2004) suggested high adult survival estimates in this colony (0.937 excluding one year with a cholera outbreak). Additionally, the guillemot colony of Stora Karlsö has almost tripled from 1980 until present (Hentati-Sundberg and Olsson, 2016; Olsson and Hentati-Sundberg, 2017).

Given the relatively low annual survival estimation for birds aged 4-10 years old, and the current breeding success, which is in the range between 0.6 and 0.7 (Berglund, 2016), there would not be a strong population increase without a strong immigration to the colony. But, as Stora Karlsö is by far the largest colony in the Baltic Sea (Olsson and Hentati-Sundberg, 2017), and due to the fact that no ringed birds from other colonies have been observed at Stora Karlsö, we can rule out the possibility of a strong immigration. There is though evidence of emigration, as birds ringed at Stora Karlsö are breeding at several other locations, mainly in the Baltic Sea but also in the Atlantic Ocean.

Rather, we consider that our annual survival estimation was underestimated despite our correction for limited resighting through the use of prior information about the colony. A potential reason for this situation could be that this database provided very little information about adult survival (indeed, it was designed to analyse juvenile survival). Interestingly, Votier et al. (2008) and Crespin et al. (2006) reported a similar survival decline with age both in annual survival from ages 4 to 5, and from age classes 3-4 to 5-18 (respectively). Both studies suggested an over representation of non-breeders with low individual quality as a possible explanation to this survival decline.

We however think that the main explanation may be the clearly observable age variation in prospecting behaviour: individuals at a later stage of prospecting become increasingly difficult to be observed when they recruit at non-observable ledges, and thus they become decreasingly likely to show up at monitored ones. Another complementary explanation could be that emigration takes place after initial prospection in the natal colony (Harris et al., 2007). Future studies involving resighting data from other colonies could help to clarify these hypotheses.

Although the official triangular stainless steel ring has remained the same throughout the study, wear of the complementary rings has been a difficult practical issue in this study and none of the rings types used has been without drawbacks. The coloured plastic rings used for most years (2006-2008, 2013-2015) were initially clearly readable, even at large distances, but tended to wear heavily after a few seasons in the colony, and eventually become unreadable. Hence those rings are best suited for observing young birds at clubs and less optimal for longer-term studies. The ID-codes of the aluminium rings (2009) were worn out very quickly, and those rings can therefore not be recommended. The large triangular steel rings (2010-2012) had a good durability but were less easy to read at longer distances. The fact that a large proportion of the resightings were based on reading the official stainless steel rings lessen the problem with different ring types. Our estimate of the recapture probabilities revealed that the 2011-2013 had the lowest recapture probabilities. These years corresponds to one year after the use of the large steel rings (2010-2012) suggesting a link between recapture probabilities and ring readability.

In conclusion, in this study we present the first results from an extensive study carried out in the largest guillemot colony from the Baltic Sea on Stora Karlsö. The wide-spread colony attendance of one and two years old birds reported in this study is an aspect which appears unique among common guillemot colonies. Continued research is needed to investigate the drivers behind the strong positive population trends in Stora Karlsö. In addition, in this work we provide a first approximation to the partial monitoring problem by incorporating expert knowledge to the inferential process through the use of prior distributions. One of the factors that can surely affect recapture probabilities is behavioral heterogeneity. An exciting way forward to obtain more accurate pre-breeding survival estimates would be to implement models that better account for the complex prospecting behavior in common guillemot.

The study presented in this Chapter has been published in the following paper:

- **Sarzo, B., Armero, C., Conesa, D., Hentati-Sundberg, J. and Olsson, O. (2019). Bayesian immature survival analysis of the largest colony of common murre *Uria aalge* in the Baltic Sea. *Waterbirds,* 42(3): 304-313.**

CHAPTER 4

CORMACK-JOLLY-SEBER MODELS: TIME AND AGE PERSPECTIVES

This Chapter is devoted to compare two different temporal scales, (i) calendar (or capture occasion) and (ii) age (or time within study), in terms of the way the data may be represented and in relation to the ecological Cormack-Jolly-Seber-type model.

The elicitation of the temporal scale is a methodology commonly performed in medical survival studies. However, in ecological framework the temporal scale is often primarily specified at the capture occasion level. In this study we present how considering the different temporal scales provides insights into the different underlying structures, which can then be combined into a joint dependence model. Finally, in this Chapter we provide an example of a data representation, model and associated results for a capture-recapture study relating to guillemots.

4.1 Introduction

Survival analysis is an area of research of special relevance in many fields such as agronomy, ecology, engineering and medicine (among others). Although in all these fields the final objective consists of assessing time to the occurrence of events, differences on the way data are gathered often result in different models and methods. Taking into account that the final objective is primarily the same, it should be the case that ideas would move from one field to the others. But this not always happens.

In the context of medical survival studies, one of the main requirements which is commonly performed at the beginning of the experiment is that the origin time (time zero) must be unambiguously defined for each individual. The time scale of the study needs also to be determined (see Table 4.1 for different choices of origin time and time scales). Indeed, the use of different temporal scales may result in different inferences and interpretations (Kom et al., 1997; Lamarca et al., 1998; Cheung et al., 2003; Westreich et al., 2010). In particular, Lamarca et al. (1998) showed that in the study of the aging process, the results obtained for some parameters were different when working with age scale or with the standard time-on-study scale.

Time zero	**Time scale**
Birth	Age
Diagnosis of disease	Duration
Entry into state	Waiting time
Bleeding	Duration of pregnancy
Start of treatment	Length of treatment
Baseline measurement	Calendar time

TABLE 4.1: Examples of possible choices of origin time and time scales in medical studies extracted from Hougaard (2000).

In ecological studies, time to death is often the main interest and (although other options are possible), monitoring data are frequently obtained using a capture-recapture protocol. This involves a series of capture occasions whereby at each occasion individuals are observed, uniquely identified (possibly by artificial means such as tag or ring) and released. For simplicity, we assume a single capture occasion each calendar year. The data correspond to the capture histories of each individual observed in the study. In the associated model, survival is commonly expressed in terms of annual survival (or survival between capture occasions) due to in general, wild animals experience a natural yearly life-cycle. But more importantly, the elicitation of the temporal scale is not always addressed in capture-recapture studies, the standard choice often being the capture-occasion (or calendar) scale. In this context, the time zero naturally corresponds to the calendar year in which each individual enters into the study at a given capture occasion.

In line with this, here we present a comparison between two different temporal scales (age and calendar) in the context of ecological capture-recapture models. We illustrate the usefulness of the not-so-commonly age scale in this context as well as emphasize both the differences and similarities in the notation and interpretation of parameters depending on the scale used. Among other capture-recapture models, our choice to perform this comparison are one of the most common statistical models used to jointly estimate annual survival and resighting probabilities in ecology in the presence of imperfect detection, the well-known Cormack-Jolly-Seber (CJS) models (Cormack, 1964; Jolly, 1965; Seber, 1965). In particular, we present a CJS model in age scale that allows incorporating both age and time dependence (both in state-space and multinomial formulation), and then we use it to analyse the juvenile survival of the largest colony of common guillemot (*Uria aalge*) in the Baltic Sea (Stora Karlsö, Sweden).

After this introduction, the remaining of the Chapter is as follows. Section 4.2 presents the notation associated with capture-recapture studies and the different parameterisations depending on the scales, while in Section 4.3 we extend the dependence structure to age and time and discuss the general expression of the CJS models in both state-space and multinomial formulations. Section 5.1 presents the application of the model in a real example, and finally Section 4.5 concludes.

4.2 Notation

In this section we present the notation associated with capture-recapture studies, and in particular focus on the different parameterisations based on the different temporal scales of the model parameters (i.e. capture occasion or age). Further we discuss the associated representation of the data for the different temporal scales.

Let $i = 1, \ldots, N$ denote the individuals who are observed within the study period; and $t = 1, \ldots, T$ the associated capture occasions. For notational simplicity we assume that the capture occasions correspond to an occasion within a calendar year, so that we refer to the temporal scale of the capture occasion to be (calendar) year (but this is clearly more general). Further, we let $a = 0, \ldots, A$ correspond to the possible ages of an individual within the study period. Note that in practice (as for our case study) the upper limit A will often correspond to an individual of at least age A (for example when it becomes an adult). If all individuals are observed at age $a = 0$ in the study, then a corresponds to length of time an individual is in the study, akin to the temporal scale of medical studies of time (assuming $A = T - 1$).

4.2.1 Data representation

The data corresponds to the observed capture histories of every individual observed within the study. This is typically expressed in terms of the capture occasion as the temporal scale. We let:

$$y_{it} = \begin{cases} 0 & \text{individual } i \text{ is not observed at capture occasion } t; \\ 1 & \text{individual } i \text{ is observed at capture occasion } t. \end{cases}$$

The capture history of individual i is then denoted $\mathbf{y}_i = \{y_{it} : t = 1, \ldots, T\}$; and full set of capture histories by $\mathbf{y}$. In addition, if individuals can be observed at different ages, this is also recorded as an observed (discrete) covariate. For example, we let a_{i0} denote the age of individual i at first capture; and $\boldsymbol{a_0} = \{a_{i0} : i = 1, \ldots, N\}$ the set of age covariate values. Further we let f_i denote the capture occasion on which individual i is first observed; and set $\mathbf{f} = \{f_1, \ldots, f_N\}$. Finally we introduce the idea of cohort $w = 1, \ldots, T - 1$ such that individual i belongs to cohort w if $f_i = w$ (i.e. individual i is observed for the first time on capture occasion w). Thus, cohort is defined on the capture occasion (or calendar year) scale. We consider the general case where individuals may enter into the study at different ages so that individuals belonging to the same cohort may be of different ages; if all individuals are observed at the same age at initial capture, then cohort and age are interchangeable.

Alternatively, and akin to how such studies are recorded within medical studies, we can present the data in terms of the age of an individual (which in turn can be sub-divided into the initial age of an individual combined with length of time in the study). This represents a focus on the age of an individual as a primary factor within the study, for example, due to the life cycle of the given species. We note that the transforming of the data to the age scale, if this is the relevant temporal scale within the study, provides a more useful representation. For example, the capture histories can be regarded as simply the observations following initial capture, corresponding to $\mathbf{y}_i = \{y_{it} : t = f_i, \dots, T : i = 1, \dots, N\}$, combined with the initial age a_{i0}. We note that the capture histories will then be of different lengths, dependent on time of first capture. The difference in representations between the calendar year and age temporal scales of the capture histories is illustrated in the toy example presented in Figures 4.1 and 4.2.

Figure 4.1 on the left shows capture-histories in the standard calendar scale, and on the right the same individual histories but in age scale. In order to highlight the differences between the temporal scales, in this simple case we assume that all individuals are age 0 at their initial capture. If individuals differ in age at initial capture then there will be additional "steps" observed in the right hand plot of Figure 4.1 with age and "time in study" not equivalent to each other. Figure 4.2 shows cohorts 1, 2 and 3 in age (top) and calendar scale (bottom) corresponding to the same individuals in Figure 4.1. Note that the later the cohort, the smaller the number of occasions or the maximum "time in study" value (in calendar and age scale, respectively). This can be explained since as the cohort number increases, the number of possible capture occasions after initial capture decreases. Moreover, the individuals represented in cohort 1 (Figure 4.2) correspond to those individuals in the first rows (first "step") of Figure 4.1. Similarly, cohorts 2 and 3 correspond to those individuals in the second and third "step" of Figure 4.1, respectively.

4.2.2 Parameter dependence

The commonly used Cormack-Jolly-Seber model has two sets of parameters corresponding to the survival probabilities and observation (or capture) probabilities. These parameters are typically assumed to be dependent on the capture occasion. In particular the model parameters (allowing for additional individual heterogeneity) are given by:

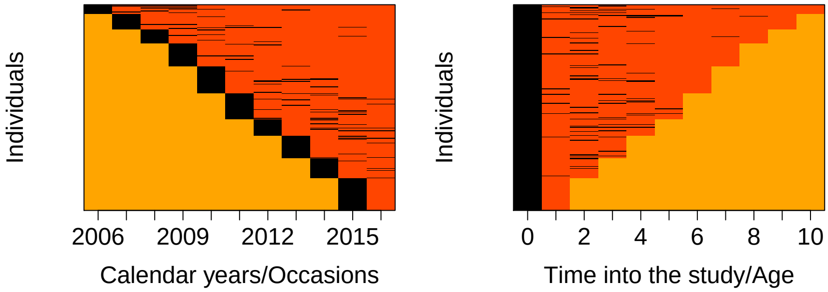

FIGURE 4.1: Toy example representing the different structure of the same individual capture-recapture histories in calendar (on the left) and age scale (on the right) assuming all individuals have age zero at the beginning of the study. Rows represent individuals (n=300), orange lines represent non information, red lines indicate that individual has not been seen, and black lines represent that the individual has been seen.

$$\phi_{i,t} = \mathbb{P}(\text{individual } i \text{ is alive at time } t+1 \mid \text{alive at time } t); \text{ and}$$

$$p_{i,t+1} = \mathbb{P}(\text{individual } i \text{ observed at time } t+1 \mid \text{alive at time } t+1),$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T-1$. Survival from one capture occasion to the next may change temporally due to, for example, weather conditions, food availability, etc. Similarly, the capture probability may be dependent on the effort at the given occasion, or local conditions for observing individuals. For a more detailed explanation of the CJS models in calendar scale, see for example, Lebreton et al. (1992), Brooks et al. (2000), Giménez et al. (2007), King et al. (2010), McCrea and Morgan (2015) and Seber and Schofield (2019), among many others.

Alternatively, the age (or time in study) scale suggests the analogous parameters given by:

$$\psi_{i,a} = \mathbb{P}(\text{individual } i \text{ is alive at age } a+1 \mid \text{alive at age } a); \text{ and}$$

$$\eta_{i,a+1} = \mathbb{P}(\text{individual } i \text{ observed at age } a+1 \mid \text{alive at age } a+1),$$

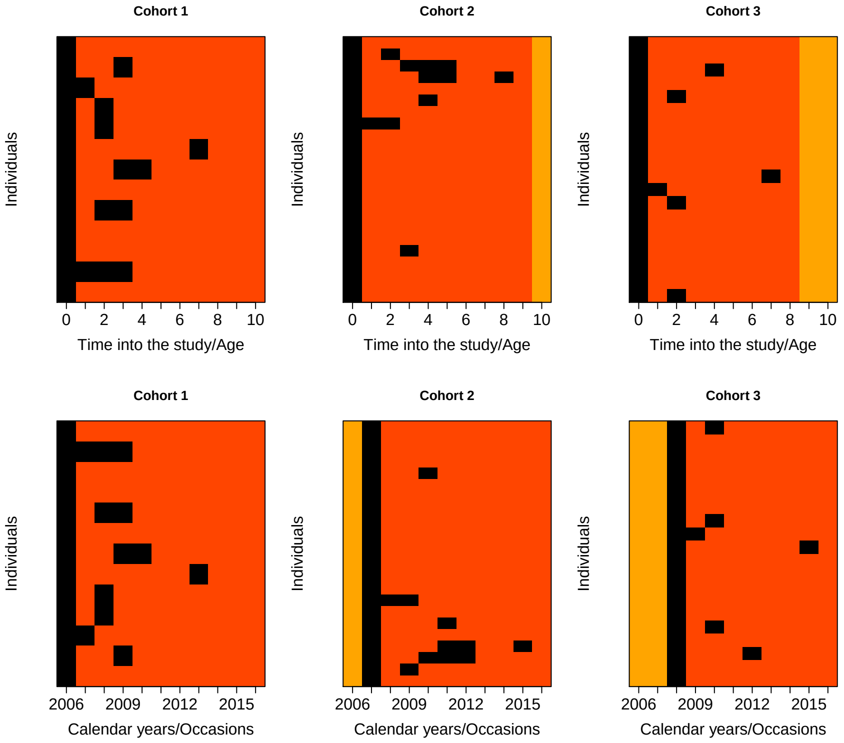

FIGURE 4.2: Capture-histories of cohorts 1, 2 and 3 in age (top) and calendar scale (bottom) of individuals of the toy example in Figure 4.1. Rows represent individuals, orange lines represent non information, red lines indicate that individual has not been seen, and black lines represent that the individual has been seen.

for $i = 1, \ldots, N$ and $a = 0, \ldots, A-1$. The parameters typically reflect the life cycle of an individual, for example, from juvenile to breeding adult where the different life stages (i.e. ages) may affect both their survival and capture probabilities based on age-related behaviour. Age has been incorporated into capture-recapture analyses, with the first-year survival probability often a key component for realistic models (Freeman and Morgan, 1992; Catchpole and Morgan, 1996).

The differences between both scales can be easily visualised in Table 4.2, where we present a particular example of an individual marked as a chick in 2009 (i.e., $a_{i0} = 0$), observed in years 2011, 2012 and 2014 (red = observed, orange =

not observed), that dies in 2016 (green = alive, black = dead). Table 4.2 includes both scales and the associated model parameters along with the different values that t and a take. As mentioned above, the age scale provides a more intuitive interpretation of the parameters when age is the main focus, as they are directly indicating the age of the individual: the probability that an individual aged a survives one year (instead of the probability that an individual in year t survives until the next year) and the probability that an alive individual aged a is seen (instead of the probability that an alive individual is seen on occasion t).

Clearly parameterising the model parameters in terms of the different temporal scales leads to very different interpretations (as it can be appreciated at Table 4.2). Both forms of dependence may be appropriate dependent on the system; and in particular both temporal scales may be important. We describe how we can incorporate both temporal scales by extending the dependence structure of the parameters (and also consider the implications for the data representation).

4.3 CJS models with year and age temporal scales

We present the general formulation of the CJS models with combined year and age as the reference temporal scales. The year-only or age-only models can then be seen as special cases of this general case. We present two (equivalent) formulations: state-space and multinomial. The state-space formulation provides an intuitive model formulation (Royle, 2008; King and McCrea, 2014); whereas the multinomial formulation is mathematically efficient and permits additional absolute goodness-of-fit tests to be applied (McCrea and Morgan, 2015).

State-Space formulation

Assuming that capture occasions are annual, let $z_{i,a,t}$ be the latent variable that describes the true state of individual i at age a and time t, for $a = a_{i0}, \dots, T - f_i + a_{i0}$, for $t = 1, \dots, T$ and all i. The possible states are alive and available for capture, $z_{i,a,t} = 1$, and dead and not available for capture (i.e. dead or migrated

Temporal Scales	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Calendar	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$	$t = 7$	$t = 8$	$t = 9$	$t = 10$	$T = 11$
Age				$a = 0$	$a = 1$	$a = 2$	$a = 3$	$a = 4$	$a = 5$	$a = 6$	$a = 7$

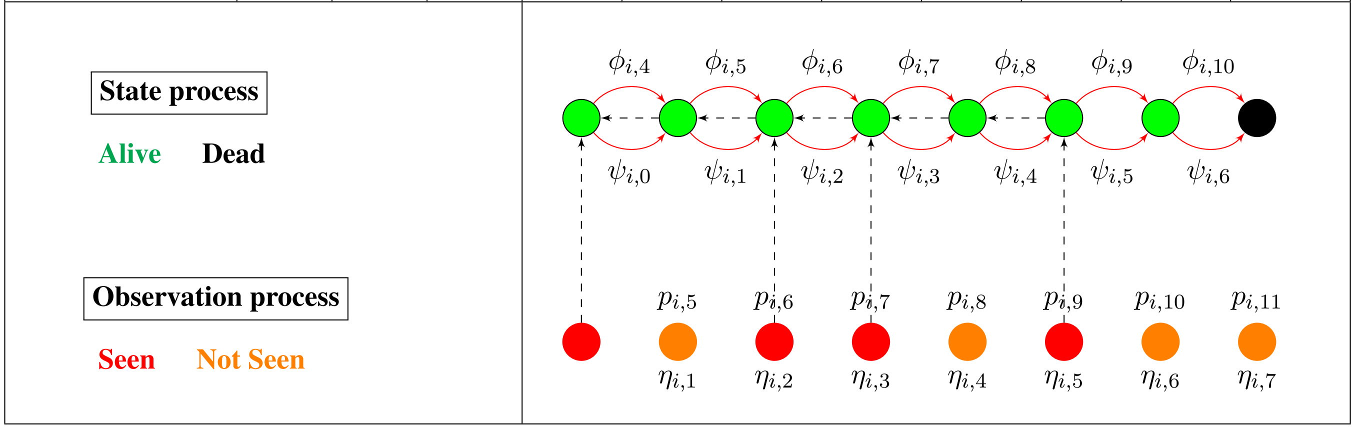

TABLE 4.2: Annual survival and resigthing probabilities, calendar years t and a values for an individual ringed as chick in 2009 in both temporal scales.

from study), $z_{i,a,t} = 0$. The corresponding survival process is given by,

$$z_{i,a+1,t+1}|z_{i,a,t} \sim \text{Bernoulli}(z_{i,a,t}\psi_{i,a,t}), \tag{4.1}$$

for $i = 1, \ldots, N$, $t = f_i, \ldots, T-1$ and $a = a_{i0}, \ldots, T - f_i + a_{i0} - 1$, where $\psi_{i,a,t}$ denotes the annual survival probability of individual i aged a at time t to time $t+1$. When the interest is focused on the survival at different ages, this parameter has an intuitive interpretation, as it represents the probability that an individual aged a at time t survives one year.

Let $y_{i,a,t}$ denote a binary variable that describes whether individual i is observed or not at age a and time t, given it is alive and available for capture, for $a = a_{i0} + 1, \ldots, T - f_i + a_{i0}$, $t = f_i + 1, \ldots, T$ and all i. The possible observations are seen ($Y_{i,a,t} = 1$), and not seen ($Y_{i,a,t} = 0$). The observation process is then given by,

$$Y_{i,a,t}|z_{i,a,t} \sim \text{Bernoulli}(z_{i,a,t}\eta_{i,a,t}), \tag{4.2}$$

for $i = 1, \ldots, N$, $t = f_i + 1, \ldots, T$ and $a = a_{i0} + 1, \ldots, T - f_i + a_{i0}$, where $\eta_{i,a,t}$ denotes the recapture probability of individual i at age a and time t given it is alive (i.e. available for capture).

Assuming independence between individuals and conditional on the first capture, the resulting likelihood is given by the product over each individual likelihood component (state and observation):

$$\begin{aligned}\ell(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta}) = \prod_{i=1}^{N}\Bigg[\prod_{a_{i0}=0}^{A}\Bigg\{&\underbrace{\Bigg(\prod_{t=f_i}^{T-1}\ \prod_{a=a_{i0}}^{T-f_i+a_{i0}+1} p_z(z_{i,a+1,t+1}|z_{i,a,t}, \boldsymbol{\theta})\Bigg)}_{\ell_{state}} \\ &\times \underbrace{\Bigg(\prod_{t=f_i+1}^{T}\ \prod_{a=a_{i0}+1}^{T-f_i+a_{i0}} p_y(y_{i,a,t}|z_{i,a,t}, \boldsymbol{\theta})\Bigg)}_{\ell_{obs}}\Bigg\}\Bigg],\end{aligned} \tag{4.3}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\psi}, \boldsymbol{\eta}\}$ represents the vector of all the parameters of the model, p_z and p_y are the (Bernoulli) probability functions of the state and observation processes respectively, and ℓ_{state} and ℓ_{obs} the likelihood terms related to the state and observation process for each individual, respectively. The observed data likelihood, $\ell(\boldsymbol{y}|\boldsymbol{\theta})$, is obtained by summing out over the $\boldsymbol{z}$ values.

Alternatively in calendar scale, if we extend the dependence structure to incorporate both temporal scales (age and time), model parameters have the same interpretation as in age scale:

$$\begin{aligned}\phi_{i,t,a} &= \mathbb{P}(\text{individual } i \text{ is alive at time } t+1 \text{ at age } a+1 \mid \text{alive at time } t \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{at age } a); \\ p_{i,t+1,a+1} &= \mathbb{P}(\text{individual } i \text{ observed at time } t+1 \text{ at age } a+1 \mid \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{alive at time } t+1 \text{ at age } a+1),\end{aligned}$$

for $i = 1, \dots, N$, $t = f_i, \dots, T-1$ and $a = a_0, \dots, T - f_i + a_0 - 1$. Therefore, in contrast to what happens when parameters are year-only or age-only dependent, when we extend the parameter dependence to time and age, model parameters are equivalent for both temporal scales.

Multinomial formulation

Capture-recapture data are often summarised in the form of *m*- arrays (McCrea and Morgan, 2015). The *m*-array summarizes the number of individuals released and subsequently observed again leading to efficient multinomial rather than Bernoulli likelihoods, with a significant reduction in computing time (Lebreton et al., 1992). However, an important disadvantage of this formulation is that models with individual effects can no longer be fitted due to the individual capture histories are broken up.

An *m*-array in age scale summarizes, in the form of sufficient statistics, the number of individuals released at certain age that are next captured at each subsequent age. The corresponding summary statistics correspond to an upper triangular matrix in which rows (denoted by a) correspond to age of release and columns (denoted by c) to age at next recapture.

When age and time dependence are included, model parameters are equivalent for both temporal scales. However, if we work in multinomial formulation one of the main differences between scales appears: the number of *m*-arrays needed when they are specified on the natural scale. In age scale, the number of *m*-arrays needed will differ depending on whether individuals enter into the study at different ages (general scenario) or they have a common initial age (simplest case), independently of the model used. In calendar scale, only one m-array per age is required. For the alternative age scale and considering the general scenario where individuals may have different ages at the beginning of the study, we need to construct one age m-array for each initial age per cohort. Therefore, each cohort will have as many m-arrays as there are initial ages observed in the study. Nevertheless, if we consider the simplest scenario where all the individuals have a common initial age, the number of m-arrays needed and their construction is simpler, namely, we only need to construct one m-array for each cohort.

The number of rows and columns of each cohort age m-array in age scale (or each cohort m-array in the simplest scenario where a common initial age for all individuals is considered) depend on the cohort we are working with. Therefore, for cohort $w = 1, \ldots, T-1$, the rows can vary in $a = a_0, \ldots, T-w+a_0-1$, while the number of columns varies in $c = a_0+1, \ldots, T-w+a_0$. Hence, the later the cohort the smaller the number of rows and columns of the associated m-array. As usual, each m-array contains an additional column corresponding to individuals never recaptured (column number $T-w+a_0+1$ in age scale).

By contrast, in calendar scale the rows (denoted by j) and the columns (denoted by t) of the m-array correspond to release occasions and first recapture occasions, respectively. Furthermore, in calendar scale their limits are $j = 1, \ldots, T-1$ and $t = 2, \ldots, T$; representing the number of rows and columns, respectively. Finally, there is also an additional column, $(T+1)$, corresponding to never seen individuals. If age dependence is included, we need one m-array for each age a and hence, the number of rows and columns of each age a m-array will be $j = a, \ldots, T-1$ and $t = a+1, \ldots, T$.

In age scale, the element (a, c) of the age m-array of cohort w and initial age a_0, denoted m_{a,c,w,a_0}, represents the number of individuals with initial age a_0 that belong to cohort w and which were released at age a and next captured at age $c+1$, for $a = a_0, \ldots, T-w-1+a_0$ and $c = a_0, \ldots, T-w-1+a_0$. By construction, $m_{a,c,w,a_0} = 0$ for $a \leq c$ (the matrix is upper triangular). Further, $m_{a,(T-w+1+a_0),w,a_0}$ denotes the number of individuals that enter the study at initial age a_0 that belong to cohort w that were released at age a and subsequently not observed again during the study (final column of the m-array). Further $R_{a,w,a_0} = \sum_c m_{a,c,w,a_0}$ represents the number of individuals with initial age a_0 that belong to cohort w, which were released at age a (i.e. R_{a,w,a_0} correspond to the sum of the elements in row a from cohort w and initial age a_0). Clearly, $m_{a,(T-w+1+a_0),w,a_0} = R_{a,w,a_0} - \sum_{c=a_0+1}^{T-w+a_0} m_{a,c,w,a_0}$, for $a = a_0, \ldots, T-w-1+a_0$ and for $c = a_0+1, \ldots, T-w+a_0$.

Denoting q_{a,c,w,a_0} the cell probabilities of the *m*-arrays, the expected values of the entries of the *m*-array ($R_{a,w,a_0} q_{a,c,w,a_0}$) are a function of the model parameters and the number of released individuals (R_{a,w,a_0}). Finally, the probability that an individual with initial age a_0 that belong to cohort w which was released at age a is not observed again during the study can be obtained as, $\chi_{a,w,a_0} = 1 - \sum_c q_{a,c,w,a_0}$.

In order to clarify concepts, Tables 4.3 and 4.4 present the *m*-arrays corresponding to individuals with initial age $a_0 = 0$ of cohorts 1 and 2, respectively, in a toy example where individuals can have different initial ages, and the number of study years is $T = 6$. For comparison, Tables 4.5 and 4.6 present *m*-arrays corresponding to individuals of cohort 1 in an example where all the individuals have a common initial age of 1 ($a_0 = 1$) or 2 ($a_0 = 2$) years, respectively, and the same study duration. Finally, in Table 4.7 we present the corresponding *m*-array in calendar scale.

The multinomial cell probabilities differ depending on the model considered. Here we present the likelihood related to the CJS model with age and time dependent model parameters. Conditional on the numbers released and assuming independence between individuals the likelihood is,

$$\ell(\boldsymbol{y}|\boldsymbol{\theta}) \propto \prod_{w=1}^{T-1} \left[\prod_{a_0=0}^{T-w+a_0} \left[\prod_{a=a_0}^{T-w+a_0-1} \left\{ \left(q_{a,c,w,a_o} \right)^{m_{a,c,w,a_0}} \chi_{a,w,a_0}^{m_{a,T-w+1+a_0,w,a_0}} \right\} \right] \right], \tag{4.4}$$

where,

$$q_{a,c,w,a_o} = \begin{cases} \prod_{c=a+1}^{T-w+a_0} \Big(& \psi_{c-a+a_0-1,c-a+w-1} \eta_{c-a+a_0,c-a+w} \\ & \times \prod_{k=a_0}^{c-a+a_0-2} \psi_{k,k+w-a_0} \left(1 - \eta_{k+1,k+w-a_0+1}\right) \Big); \\ 0 & \text{for } a \leq c. \end{cases}$$

The probability of an individual released at age a that belong to cohort w with initial age a_0 is not observed again can be expressed as,

$$\chi_{a,w,a_0} = 1 - \sum_{c=a+1}^{T-w+a_0} (q_{a,c,w,a_o}).$$

We note that, in equation (4.4) when $k = a_0$ and $c = a + 1$, the product is over the null set and $\prod_{k=a_0}^{c+a_0-a-2} \equiv 1$. Further, we note that in equation (4.4), the t terms (calendar time) can be expressed as a function of the initial age (a_0), the age of release (rows of the *m*-array, a) (or the age at first recapture, i.e. columns of the *m*-arrays, c) and cohort (w). Finally we note that, further restrictions may be specified on these parameters to represent age classes, denoted by a.

Therefore, in capture-recapture framework this model can be fitted in both calendar and/or age scales. However, working with this model and under multinomial formulation, the differences between scales are more remarkable regarding the presentation of the data on the natural age or time scales (i.e. number of *m*-arrays needed).

We note that when time dependence is included, these models have an intrinsic identifiability problem in the last two parameters (King et al., 2010; Giménez et al., 2009). This lack of identifiability results in estimating the product of both parameters but not each one separately (i.e. they are confounded). Nevertheless, it is important to mention that although a model may be unidentifiable, it can be still adequate to describe the data via its identifiable parameters (Garrett and Zeger, 2000).

Cohort 1 ($a_0 = 0$)								
Age at first recapture								
			1	**2**	**3**	**4**	**5**	**Never seen**
Age of release	**0**	$R_{0,1,0}$	$m_{0,1,1,0}$	$m_{0,2,1,0}$	$m_{0,3,1,0}$	$m_{0,4,1,0}$	$m_{0,5,1,0}$	$m_{0,6,1,0}$
			$(q_{0,1,1,0})$	$(q_{0,2,1,0})$	$(q_{0,3,1,0})$	$(q_{0,4,1,0})$	$(q_{0,5,1,0})$	$(\chi_{0,1,0})$
	1	$R_{1,1,0}$	–	$m_{1,2,1,0}$	$m_{1,3,1,0}$	$m_{1,4,1,0}$	$m_{1,5,1,0}$	$m_{1,6,1,0}$
				$(q_{1,2,1,0})$	$(q_{1,3,1,0})$	$(q_{1,4,1,0})$	$(q_{1,5,1,0})$	$(\chi_{1,1,0})$
	$\vdots$				$\vdots$			$\vdots$
	4	$R_{4,1,0}$	–	–	–	–	$m_{4,5,1,0}$	$m_{4,6,1,0}$
							$(q_{4,5,1,0})$	$(\chi_{4,1,0})$

TABLE 4.3: *M*-array for initial age $a_0 = 1$ and cohort 1 for an study where individuals can have different initial ages. Released individuals (R_{a,w,a_0}), observed individuals (m_{a,c,w,a_0}) and cell probabilities (q_{a,c,w,a_0} and χ_{a,w,a_0}).

Cohort 2 ($a_0 = 0$)								
Age at first recapture								
			1	**2**	**3**	**4**	–	**Never seen**
Age of release	**0**	$R_{0,2,0}$	$m_{0,1,2,0}$	$m_{0,2,2,0}$	$m_{0,3,2,0}$	$m_{0,4,2,0}$	–	$m_{0,5,2,0}$
			$(q_{0,1,2,0})$	$(q_{0,2,2,0})$	$(q_{0,3,2,0})$	$(q_{0,4,2,0})$	–	$(\chi_{0,2,0})$
	1	$R_{1,2,0}$	–	$m_{1,2,2,0}$	$m_{1,3,2,0}$	$m_{1,4,2,0}$	–	$m_{1,5,2,0}$
				$(q_{1,2,2,0})$	$(q_{1,3,2,0})$	$(q_{1,4,2,0})$	–	$(\chi_{1,2,0})$
	$\vdots$				$\vdots$			$\vdots$
	3	$R_{3,2,0}$	–	–	–	$m_{3,4,2,0}$	–	$m_{3,5,2,0}$
						$(q_{3,4,2,0})$	–	$(\chi_{3,2,0})$
	–	–	–	–	–	–	–	–

TABLE 4.4: *M*-array for initial age $a_0 = 1$ and cohort 2 for an study where individuals can have different initial ages. Released individuals (R_{a,w,a_0}), observed individuals (m_{a,c,w,a_0}) and cell probabilities (q_{a,c,w,a_0} and χ_{a,w,a_0}).

			Cohort 1 (common $a_0 = 1$)					
			Age at first recapture					
			2	**3**	**4**	**5**	**6**	**Never seen**
Age of release	**1**	$R_{1,1,1}$	$m_{1,2,1,1}$ $(q_{1,2,1,1})$	$m_{1,3,1,1}$ $(q_{1,3,1,1})$	$m_{1,4,1,1}$ $(q_{1,4,1,1})$	$m_{1,5,1,1}$ $(q_{1,5,1,1})$	$m_{1,6,1,1}$ $(q_{1,6,1,1})$	$m_{1,7,1,1}$ $(\chi_{1,1,1})$
	2	$R_{2,1,1}$	–	$m_{2,3,1,1}$ $(q_{2,3,1,1})$	$m_{2,4,1,1}$ $(q_{2,4,1,1})$	$m_{2,5,1,1}$ $(q_{2,5,1,1})$	$m_{2,6,1,1}$ $(q_{2,6,1,1})$	$m_{2,7,1,1}$ $(\chi_{2,1,1})$
	$\vdots$				$\vdots$			$\vdots$
	5	$R_{5,1,1}$	–	–	–	–	$m_{5,6,1,1}$ $(q_{5,6,1,1})$	$m_{5,7,1,1}$ $(q_{5,7,1,1})$

TABLE 4.5: M-array for initial age $a_0 = 1$ and cohort 1 in an study where all the individuals have a common initial age $a_0 = 1$. Released individuals (R_{a,w,a_0}), observed individuals (m_{a,c,w,a_0}) and cell probabilities (q_{a,c,w,a_0} and χ_{a,w,a_0}).

			Cohort 1 (common $a_0 = 2$)					
			Age at first recapture					
			3	**4**	**5**	**6**	**7**	**Never seen**
Age of release	**2**	$R_{2,1,2}$	$m_{2,3,1,2}$ $(q_{2,3,1,2})$	$m_{2,4,1,2}$ $(q_{2,4,1,2})$	$m_{2,5,1,2}$ $(q_{2,5,1,2})$	$m_{2,6,1,2}$ $(q_{2,6,1,2})$	$m_{2,7,1,2}$ $(q_{2,7,1,2})$	$m_{2,8,1,2}$ $(\chi_{2,1,2})$
	3	$R_{3,1,2}$	–	$m_{3,4,1,2}$ $(q_{3,4,1,2})$	$m_{3,5,1,2}$ $(q_{3,5,1,2})$	$m_{3,6,1,2}$ $(q_{3,6,1,2})$	$m_{3,7,1,2}$ $(q_{3,7,1,2})$	$m_{3,8,1,2}$ $(\chi_{3,1,2})$
	$\vdots$				$\vdots$			$\vdots$
	6	$R_{6,1,2}$	–	–	$m_{6,5,1,2}$ $(q_{6,5,1,2})$	$m_{6,6,1,2}$ $(q_{6,6,1,2})$	$m_{6,7,1,2}$ $(q_{6,7,1,2})$	$m_{6,8,1,2}$ $(\chi_{6,1,2})$

TABLE 4.6: M-array for initial age $a_0 = 2$ and cohort 1 in an study where all the individuals have a common initial age $a_0 = 2$. Released individuals (R_{a,w,a_0}), observed individuals (m_{a,c,w,a_0}) and cell probabilities (q_{a,c,w,a_0} and χ_{a,w,a_0}).

4.4 Survival analysis of a colony of common guillemot

In order to illustrate differences and similarities between both temporal scales, in what follows we implement two models in a real data set: (i) the year-only and age-only model in calendar and age scale, respectively, and (ii) the time and age dependent model in age scale (due to in calendar scale both models are equivalent).

In particular, we present the analysis of a database obtained from monitoring programs carried out by the Baltic Seabird Project from 2006 until 2016 in the

		First Recapture occasion					
Release occ.		**2**	**3**	**4**	**5**	**6**	**Never seen**
1	R_1	$m_{1,2}$ $(q_{1,2})$	$m_{1,3}$ $(q_{1,3})$	$m_{1,4}$ $(q_{1,4})$	$m_{1,5}$ $(q_{1,5})$	$m_{1,6}$ $(q_{1,6})$	$m_{1,7}$ (χ_1)
2	R_2	–	$m_{2,3}$ $(q_{2,3})$	$m_{2,4}$ $(q_{2,4})$	$m_{2,5}$ $(q_{2,5})$	$m_{2,6}$ $(q_{2,6})$	$m_{2,7}$ (χ_2)
$\vdots$				$\vdots$			$\vdots$
5	R_5	–	–	–	–	$m_{3,6}$ $(q_{3,6})$	$m_{3,7}$ (χ_5)

TABLE 4.7: General m-array in calendar scale: released (R_t), observed individuals ($m_{t,j}$) and cell probabilities ($q_{t,j}$ and χ_t)

largest colony of common guillemot in the Baltic Sea (Stora Karlsö, Sweden). Each year chicks were captured and ringed individually (with metallic and plastic rings) after jumping from the ledges at an age of about 20 days. The total number of individuals ringed is $N = 28\ 930$. During their reproductive period (May to July), resightings with telescopes and binoculars were made from above the breeding ledges with minimal disturbance in order to identify marked individuals through the reading of the metallic and/or plastic ring codes (recaptures). Hence, all the individuals were ringed at a common initial age of zero years old.

It is worth noting that in this real example we work with the simplest scenario where we construct only one m-array per cohort (working with age scale). To emphasize differences in data presentation when using both temporal scales and for reproducibility issues, in Appendix 4.6 we present the cohort m-arrays and the age m-arrays related to age and calendar scales, respectively.

In order to highlight differences between temporal scales, we will not establish age classes in the age-only or time-only CJS models and therefore, we will have as many parameters as possible ages within the study (in age scale) or alternatively, as many parameters as study years minus one (in calendar scale). Nevertheless, we do establish age classes in the CJS model where both temporal scales are considered. Indeed, we let denote $a = 0, 1, 2, 3+$ the age classes, being those related to individuals in their first year of life, second, third and finally, individuals which are four or more years old (adults), respectively. Age and time effects are additive (on the logit scale) for both survival and recapture probabilities. We set logit($\psi_{a,t}$) $= \alpha_a + \beta_t$, for $a = 0, \ldots, 3+$ and $t = 1, \ldots, T-1$, and alternatively, logit($\eta_{a,t}$) $= \alpha'_a + \beta'_t$, for $a = 1, \ldots, 3+$ and $t = 2, \ldots, T$.

Therefore, the full set of parameters will differ for each model proposed and the temporal scale considered. The model parameters of the full time (or age) dependent CJS models are, $\boldsymbol{\psi} = \{\psi_0, \psi_1, \ldots, \psi_9\}$ and $\boldsymbol{\eta} = \{\eta_1, \eta_2, \ldots, \eta_{10}\}$, in age scale; and $\boldsymbol{\phi} = \{\phi_1, \phi_2, \ldots, \phi_{10}\}$ and $\boldsymbol{p} = \{p_2, p_3, \ldots, p_{11}\}$, in calendar time scale. However, parameters ψ_9 and η_{10} in age scale; and similarly, ϕ_{10} and p_{11} in calendar scale, are not identifiable. Alternatively, the corresponding set of model parameters for the CJS model with age and time dependence are, $\boldsymbol{\alpha} = \{\alpha_0, \ldots, \alpha_{3+}\}$ and $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_{10}\}$ (regarding survival probabilities) and $\boldsymbol{\alpha'} = \{\alpha'_1, \ldots, \alpha'_{3+}\}$ and $\boldsymbol{\beta'} = \{\beta'_1, \ldots, \beta'_{10}\}$ (regarding capture probabilities). We have to point out that α'_a for $a = 1, 2, 3+$ is interpreted as the age effect relative to age 1, 2, and age 3+. Finally, further comments on the identifiability of these parameters are given in next Section.

Bayesian inference

Once the models are fully specified, the next step is to make inference on their corresponding parameters. Maximum likelihood estimation or Bayesian methods can be used to perform inference in capture-recapture data (King et al., 2010; Newman et al., 2014; McCrea and Morgan, 2015; Seber and Schofield, 2019, and references therein).

In our case, inference on these parameters is performed within the Bayesian framework, as we can benefit from the advantages of this approach: probabilistic statements about the veracity of hypotheses or relationships can be made given the data, and it is possible to explicitly incorporate all the available prior knowledge about the unknown parameters into the inferential process (McCarthy and Masters, 2005; King et al., 2010; Christensen et al., 2011; Golchi, 2016).

As usual in this context, the resulting posterior distributions containing all the information about the system have no closed expression, so we need numerical approaches to approximate them. We consider Markov chain Monte Carlo (MCMC) methods (Robert and Casella, 2011). In this work, MCMC methods have been implemented via the `jags` software (Plummer, 2003) with the R program (R Core Team, 2019). As no prior knowledge about the parameters was available, we elicited non informative prior distributions for all the parameters of the model. In particular, Uniform(0,1) distributions for all survival and recapture probabilities in the CJS model with only-time or only-age dependence, $\boldsymbol{\phi_t}$ and $\boldsymbol{p_t}$). Further, we specify Normal(0,10) distributions for age ($\boldsymbol{\alpha_a}$ and $\boldsymbol{\alpha'_a}$) and time ($\boldsymbol{\beta_t}$ and $\boldsymbol{\beta'_t}$) parameters in CJS model with age and time dependence. Finally, to avoid identifiability issues, we set $\alpha_1 = \alpha'_1 = 0$.

The final approximate random sample was obtained using three chains, 200,000 iterations, discarding the first 20,000 (to ensure that the Markov chain has reached the stationary distribution) and thinning every 300 of the iterations for the memory storage purposes and to reduce auto correlation.

Results

Figure 4.3 displays a graphical representation of posterior distribution of survival (cyan) and recapture (blue) probabilities for both temporal scales considering full time dependence (age or time). Due to the identifiability problem mentioned, in this Figure we only present the estimable parameters.

As it can be appreciated in Figure 4.3 (and highlighted in Table 4.2) if we only incorporate time or age dependence the model parameters differ between temporal scales. In particular, while in age scale model parameters represent differences in survival and recapture probabilities regarding the age of individuals, in calendar scale they indicate interannual variations in survival and resighting probabilities. Note that credible intervals for older ages (5+) are wider than those corresponding to later times (from time $t = 6$) as there are fewer older individuals observed within the study.

However, if we extend the dependence to age and time, model parameters are equivalent in both temporal scales. Hence, Figure 4.4 displays the simulated values for each posterior survival distributions (on the left) and posterior recapture distributions (on the right) along the study period by age classes. Note that the lowest survival estimate corresponds to one year old individuals. These results are in line with previous studies that stated first year survival is lower than later

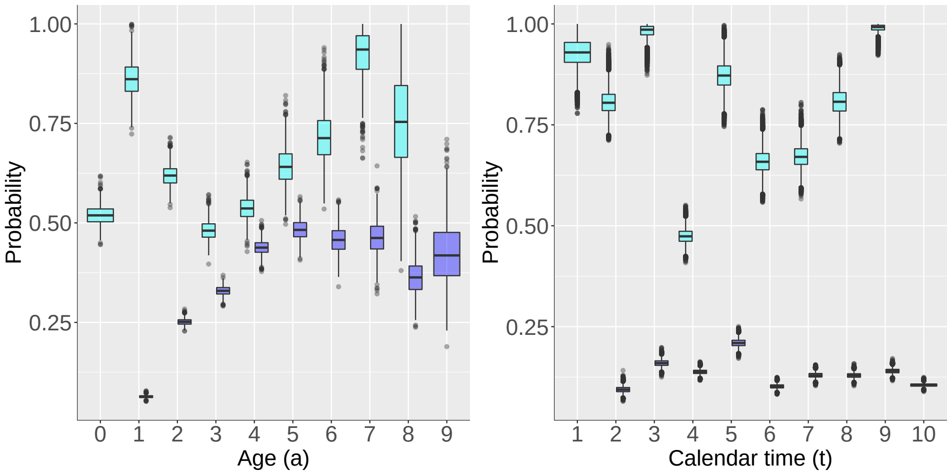

FIGURE 4.3: Graphical representation of the approximated posterior marginal distributions associated to annual survival (cyan) and resighting probabilities (blue) probabilities in the full time dependent CJS model in age scale (left) and calendar time scale (right).

(Harris et al., 2007). Additionally, survival estimate for 2 years old individuals is extremely high (and unrealistic for this species). Finally, recapture probabilities show a strong age-dependence, increasing with age. The lowest estimate corresponds to one year old individuals due to the limited number of individuals that come back to the colony at this age.

4.5 Conclusions

Measuring time to the occurrence of events, also called survival analysis, is a major focus on statistics. In this study we try to connect two distant areas in the context of survival analysis: human demography and population ecology. The development of this work was motivated for the different emphasis given to the elicitation of the temporal scale in survival studies performed in the context of ecology or public health. Indeed, in medical studies the origin time and consequently, the temporal scale, must be defined at the beginning of the study. In this context, there are several time scales defined in relation to the possible times zero (Hougaard, 2000).

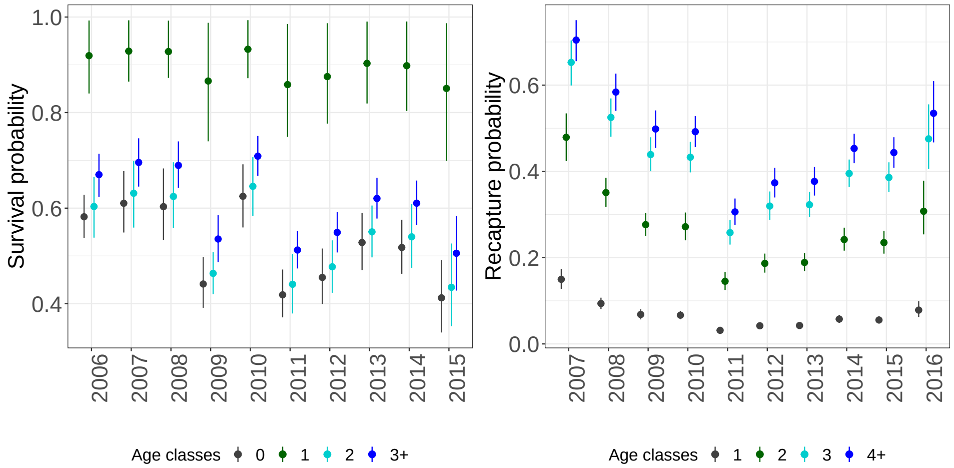

FIGURE 4.4: Mean and 95% symmetric credible intervals of the approximated posterior marginal distributions associated to annual survival probabilities (left) and recapture probabilities (right) along the study period by age classes.

However, in ecology these procedures are not performed, being the standard choice the capture-occasion (or calendar) scale. In this work, we adopt the approach and data presentation used in medical studies to ecological capture-recapture framework. By doing so we present the individual capture-histories in an alternative temporal scale, the age. Age is an important factor affecting survival probabilities in animal populations and consequently, several studies have been focused in estimating age dependent survival and recapture probabilities. Additionally, in several capture-recapture protocols (as the real example presented in this work) the individuals are marked days after they were born being the time zero the birth time. Thus, from medical framework perspective, if time zero is birth time the 'natural' temporal scale is then the age.

In ecology, different models can be used to estimate survival and recapture probabilities from capture-recapture data. Although it can be easily extended to other models, we have used the well known Cormack-Jolly-Seber (CJS) models. We present the general notation of the CJS models allowing for age and time dependence on model parameters in both formulations (state-space and multinomial) using the age as the temporal scale. Additionally, we present the results obtained through two different CJS models implemented in a real database, the largest colony of common guillemot in the Baltic Sea: (i) CJS model with full time dependence (being age or time, depending on the temporal scale used), and (ii) CJS model with age and time dependence in model parameters. Furthermore in this work we also provide the general notation of the CJS models for two common scenarios. First, considering the individuals may enter into the study at different ages; and second, the simplest one, where all the individuals have a common initial age at the beginning of the study.

One of the more remarkable differences between temporal scales is the number of m-arrays needed in multinomial formulation. In particular in calendar scale, when no age dependence is present in any of the parameters of the model, we only need to construct one general m-array to summarize the whole data set; alternatively, if age-dependence is present, we need a m-array for each age class. However, working with the age scale, the number of m-arrays depends on the initial age of the individuals, irrespective of whether or not the parameters are age dependent. Particularly, if the individuals have different ages at the beginning of the study, this natural scale leads to one m-array for each initial age per cohort, therefore, we have as many m-arrays as possible initial ages (per cohort). Nevertheless, if all the individuals have a common initial age, only one m-array per cohort is needed.

In this Chapter we present how changing the temporal scale, the same individual capture-histories are presented differently. Indeed, whereas the length of the capture-histories depends on the cohort of the individual in both scales, its meaning is slightly different. In calendar scale, the later the cohort the smaller the number of capture occasions (years) after initial capture. However, in age scale, as cohort increases, the number of possible ages that the individual can take along the study, decreases.

In the medical framework, it is well known that changes in temporal scales may lead to different inferences and parameter interpretations (Lamarca et al., 1998; Cheung et al., 2003; Westreich et al., 2010). In the ecological capture-recapture framework, differences appear in the data representation. Further, when model parameters are only time or age dependent they have a different interpretation depending on the temporal scale used. However, if we generalize this model, extending the temporal structure to age and time, the model can be fitted in both temporal scales and it does not result in different inferences since model parameters are equivalent. This special case of age or time only dependence is obtained by specifying given restrictions on the parameters.

The alternative temporal scale presented in this work directly takes into account the age effect in survival and resighting probabilities, providing a more understandable interpretation of the age-dependent model parameters. Indeed, the data are more naturally associated with the model on the same temporal scale and hence, we propose the selection of one temporal scale or another depending on the main interest of the study. Therefore, this aspect may be particularly useful in studies where age is one of the main concerns or whether inter-annual variations in survival are not considered (e.g. in cases where the population trajectory is stable). Finally, although the methodology presented in this study is general, it can be easily extended to any particular model within capture-recapture framework.

To conclude, medical and ecological survival studies are often assumed very distant areas particularly in terms of temporal scales but in this work we have tried to bring both together. Particularly, we want to remark that the similarities found may help to borrow ideas in order to develop models further in both frameworks and thereby, providing the possibility of a room of improvement in both areas.

4.6 M-arrays

In order to highlight differences in data presentation when using both temporal scales and for reproducibility issues, here we present the ten cohort m-arrays in age scale (Tables 4.8 to 4.17) and the four age m-arrays in calendar scale (Tables 4.18 to 4.21), corresponding to the study database.

The study presented in this Chapter has been submitted to an indexed journal with the title: "Cormack-Jolly-Seber models: time and age perspectives".

		Cohort 1										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	174	361	100	62	21	7	3	5	1	0	1265
	1	0	80	20	4	2	0	0	1	0	0	67
	2	0	0	163	35	6	2	1	2	1	1	230
	3	0	0	0	96	23	6	2	0	0	2	154
	4	0	0	0	0	61	8	4	2	1	0	121
	5	0	0	0	0	0	37	11	4	2	1	58
	6	0	0	0	0	0	0	18	5	3	3	31
	7	0	0	0	0	0	0	0	16	7	3	13
	8	0	0	0	0	0	0	0	0	9	5	21
	9	0	0	0	0	0	0	0	0	0	9	15

TABLE 4.8: *M*-array cohort 1 in age scale.

		Cohort 2										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	50	290	175	51	18	8	5	2	4	-	1362
	1	0	12	4	2	0	0	0	0	0	-	32
	2	0	0	82	16	5	1	1	0	0	-	197
	3	0	0	0	63	10	1	0	1	2	-	184
	4	0	0	0	0	43	4	1	0	1	-	83
	5	0	0	0	0	0	34	4	2	1	-	35
	6	0	0	0	0	0	0	23	3	3	-	19
	7	0	0	0	0	0	0	0	12	7	-	15
	8	0	0	0	0	0	0	0	0	6	-	14
	9	-	-	-	-	-	-	-	-	-	-	0

TABLE 4.9: *M*-array cohort 2 in age scale.

		Cohort 3										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	101	246	123	49	13	3	12	2	-	-	1440
	1	0	30	10	4	0	0	1	0	-	-	56
	2	0	0	64	13	2	1	0	0	-	-	196
	3	0	0	0	42	5	4	0	0	-	-	146
	4	0	0	0	0	26	5	2	2	-	-	73
	5	0	0	0	0	0	17	2	2	-	-	25
	6	0	0	0	0	0	0	11	3	-	-	16
	7	0	0	0	0	0	0	0	10	-	-	18
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.10: *M*-array cohort 3 in age scale.

		Cohort 4										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	57	110	126	65	29	11	9	-	-	-	2586
	1	0	5	3	0	0	1	0	-	-	-	48
	2	0	0	19	5	2	1	1	-	-	-	87
	3	0	0	0	32	3	0	0	-	-	-	113
	4	0	0	0	0	31	9	1	-	-	-	61
	5	0	0	0	0	0	19	5	-	-	-	41
	6	0	0	0	0	0	0	16	-	-	-	25
	7	-	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.11: *M*-array cohort 4 in age scale.

		Cohort 5										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	44	297	212	111	42	24	-	-	-	-	2263
	1	0	9	3	0	1	0	-	-	-	-	31
	2	0	0	51	16	9	7	-	-	-	-	223
	3	0	0	0	52	13	6	-	-	-	-	195
	4	0	0	0	0	51	9	-	-	-	-	119
	5	0	0	0	0	0	31	-	-	-	-	85
	6	-	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.12: M-array cohort 5 in age scale.

		Cohort 6										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	61	200	186	99	31	-	-	-	-	-	2423
	1	0	12	5	3	3	-	-	-	-	-	38
	2	0	0	47	6	7	-	-	-	-	-	152
	3	0	0	0	50	9	-	-	-	-	-	179
	4	0	0	0	0	52	-	-	-	-	-	106
	5	-	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.13: M-array cohort 6 in age scale.

		Cohort 7										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	41	272	174	80	-	-	-	-	-	-	2280
	1	0	8	5	0	-	-	-	-	-	-	28
	2	0	0	48	19	-	-	-	-	-	-	213
	3	0	0	0	54	-	-	-	-	-	-	173
	4	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.14: M-array cohort 7 in age scale.

		Cohort 8										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	149	310	192	-	-	-	-	-	-	-	2453
	1	0	38	14	-	-	-	-	-	-	-	97
	2	0	0	64	-	-	-	-	-	-	-	284
	3	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.15: M-array cohort 8 in age scale.

		Cohort 9										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	111	389	-	-	-	-	-	-	-	-	2584
	1	0	21	-	-	-	-	-	-	-	-	90
	2	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.16: *M*-array cohort 9 in age scale.

		Cohort 10										
		Age at first recapture										
		1	2	3	4	5	6	7	8	9	10	Never seen
Age of release	0	173	-	-	-	-	-	-	-	-	-	4783
	1	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	-	-	-	-	-	-	-
	6	-	-	-	-	-	-	-	-	-	-	-
	7	-	-	-	-	-	-	-	-	-	-	-
	8	-	-	-	-	-	-	-	-	-	-	-
	9	-	-	-	-	-	-	-	-	-	-	-

TABLE 4.17: *M*-array cohort 10 in age scale.

		First Recapture occasion										
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Never seen**
Release occ.	1	174	361	100	62	21	7	3	5	1	0	1265
	2	0	50	290	175	51	18	8	5	2	4	1362
	3	0	0	101	246	123	49	13	3	12	2	1440
	4	0	0	0	57	110	126	65	29	11	9	2586
	5	0	0	0	0	44	297	212	111	42	24	2263
	6	0	0	0	0	0	61	200	186	99	31	2423
	7	0	0	0	0	0	0	41	272	174	80	2280
	8	0	0	0	0	0	0	0	149	310	192	2453
	9	0	0	0	0	0	0	0	0	111	389	2584
	10	0	0	0	0	0	0	0	0	0	173	4783

TABLE 4.18: Age 1 *m*-array in calendar scale.

		First Recapture occasion										
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Never seen**
Release occ.	1	-	-	-	-	-	-	-	-	-	-	-
	2	0	80	20	4	2	0	0	1	0	0	67
	3	0	0	12	4	2	0	0	0	0	0	32
	4	0	0	0	30	10	4	0	0	1	0	56
	5	0	0	0	0	5	3	0	0	1	0	48
	6	0	0	0	0	0	9	3	0	1	0	31
	7	0	0	0	0	0	0	12	5	3	3	38
	8	0	0	0	0	0	0	0	8	5	0	28
	9	0	0	0	0	0	0	0	0	38	14	97
	10	0	0	0	0	0	0	0	0	0	21	90

TABLE 4.19: Age 2 *m*-array in calendar scale.

		First Recapture occasion										
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Never seen**
Release occ.	1	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-
	3	0	0	163	35	6	2	1	2	1	1	230
	4	0	0	0	82	16	5	1	1	0	0	197
	5	0	0	0	0	64	13	2	1	0	0	196
	6	0	0	0	0	0	19	5	2	1	1	87
	7	0	0	0	0	0	0	51	16	9	7	223
	8	0	0	0	0	0	0	0	47	6	7	152
	9	0	0	0	0	0	0	0	0	48	19	213
	10	0	0	0	0	0	0	0	0	0	64	284

TABLE 4.20: Age 3 *m*-array in calendar scale.

		First Recapture occasion										
		2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Never seen**
Release occ.	1	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-	-
	4	0	0	0	96	23	6	2	0	0	2	154
	5	0	0	0	0	124	18	5	2	2	2	305
	6	0	0	0	0	0	122	20	9	2	2	287
	7	0	0	0	0	0	0	110	17	7	6	252
	8	0	0	0	0	0	0	0	139	34	15	313
	9	0	0	0	0	0	0	0	0	152	38	391
	10	0	0	0	0	0	0	0	0	0	178	436

TABLE 4.21: Age 4+ *m*-array in calendar scale.

CHAPTER 5

BEYOND INTEGRATED MODELS: CORRECTING BIAS IN PARTIALLY MONITORED POPULATIONS

This Chapter is devoted to present the last study performed within this Thesis. This work arose with a challenging goal: to correct the bias in parameter estimates due to the partial monitoring problem (already mentioned in Chapter 3). To that end, two research visits were made with Professor Ruth King (University of Edinburgh).

In this study we provide an integrated mark-recapture-recovery framework for partially monitored recapture studies. In these studies live resightings are only observable at a set of monitored locations, so that if an individual leaves these locations they become unavailable for capture. However, the additional ring-recovery data reduces the corresponding bias obtained in the survival probability estimates obtained from capture-recapture data only due to the confounding with colony dispersal. We derive an explicit efficient likelihood expression, and state the associated sufficient statistics. In this Chapter we demonstrate the significant improvements in the estimation of the survival probabilities for a colony of guillemots, where we additionally specify a hierarchical approach to deal with low sample size over the early period of the study.

5.1 Introduction

Capture-recapture studies are frequently used to study wild animal populations, often with a particular focus on estimating survival probabilities. These involve biologists going into the field at a series of capture occasions to observe live resightings of uniquely marked individuals. For an overview of capture-recapture studies and associated statistical models, see for example, King et al. (2010); McCrea and Morgan (2015); Seber and Schofield (2019). However, such studies are generally limited with mortality confounded with dispersal from the study area. The associated *apparent* survival estimates are negatively biased with regard to *true* survival probabilities. The greater the dispersal from the study area, the greater the level of bias.

We consider partially monitored populations where only a limited number of locations are monitored at the study site. The problem is motivated by a long-term study of a colony of common guillemots (*Uria aalge*), hereafter guillemots. Individuals are marked as young at the monitored sites, but due to the life cycle of the species they may disperse (at natal or breeding stage) to unmonitored sites and become unavailable at subsequent capture occasions. Due to this dispersal, fitting standard capture-recapture models leads to significantly biased survival estimates (Sarzo et al., 2019). Previous attempts to correct for the partial monitoring have included the use of external estimates of the probability that individuals recruit into the monitored sites (Sanz-Aguilar et al., 2016). Alternatively, integrating capture-recapture data (of live resightings) with ring-recovery data (of dead recoveries) also permits the disentangling of survival from dispersal, assuming that all individuals remain available to be recovered dead within the study (irrespective of whether or not they are available for resighting). The joint analysis of live resightings and ring-recovery (MRR) data was first presented by Burnham (1993), allowing for time-dependent model parameters, and extended by Barker (1999) to include age-dependent parameters. Catchpole et al. (1998) included both time and age dependent model parameters, and deduced a set of sufficient statistics (though we note that these cannot be used for goodness-of-fit as they are non-independent).

Recent developments have focused on extending the model of Catchpole et al. (1998), and have included developing a state-space formulation (Bonner et al., 2010; King, 2012), and incorporating time-varying individual covariates, which are either discrete (King and Brooks, 2003; McCrea et al., 2012; McCrea and Morgan, 2015) or continuous (Cathpole et al., 2008; Bonner et al., 2010; Langrock and King, 2013).

We consider the particular case of MRR data where all individuals are able to be recovered dead, but not all individuals are available to be resighted, if they move to an unmonitored site. We provide an efficient likelihood expression, and associated (independent) sufficient statistics, for MRR data for partially monitored populations in the presence of both age and time parameter dependence. We develop the model further by specifying a hierarchical model formulation to permit additional borrowing of information over time leading to increased precision of the model parameters, particularly for early years of the study where sample size is small. We fit the model to the guillemot data using a Bayesian approach, permitting prior information to be specified on relative resighting probabilities over time given the study design.

In Section 5.2 we present a description of the data. Section 5.3 provides the model formulation of the integrated likelihood for partial monitored populations, including the efficient m-array formulation and associated summary statistics. In Section 5.4 we describe the hierarchical model and fit the model to the guillemot data, comparing the results to the capture-recapture-only case. Finally we conclude with a discussion in Section 5.5.

5.2 Data

We consider data relating to the largest guillemot colony in the Baltic Sea, on the island of Stora Karlsö (Sweden), with a recorded breeding population of 15,700 pairs in 2014, corresponding to $\approx$2/3 of the Baltic Sea population (Olsson and Hentati-Sundberg, 2017). Over the study period a total of 39,681 chicks are ringed between late June and early July after jumping from the ledges on to a beach under the cliffs in the largest sub-colony of the island. We consider resightings of individuals carried out at five different monitored locations: two clubs (aggregations of

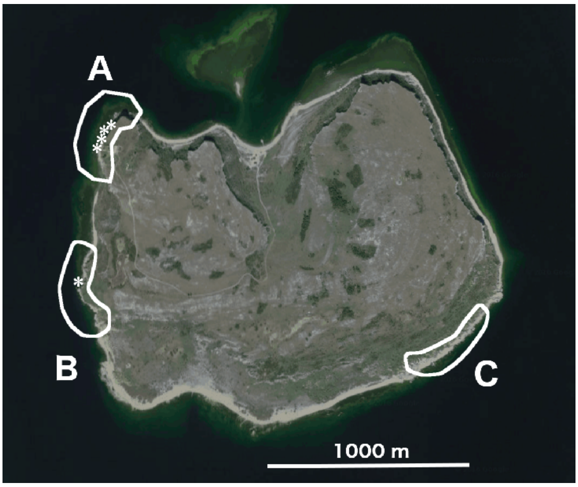

FIGURE 5.1: Location of the Stora Karlsö guillemot colony in the Baltic Sea (left). Map of the island (right) shows the location of the main breeding ledges used by guillemots (areas A, B and C) and monitored areas (indicated with asterisks). Chicks were captured and banded in area A. Recoveries take place over the whole colony and elsewhere (mainly in Baltic Sea but also in the Atlantic Ocean).

birds at rocks beneath the breeding ledges) and three breeding ledges. Long-ranged telescopes are used to record the unique ring identifiers of birds at the monitored locations to minimise disturbance. Recoveries of dead individuals occur at both the colony level and further afield. The capture-recapture data are collected over the period 2002 to 2016, while the ring-recovery database spans 1992 to 2018. A total of 6,477 unique individuals are resighted within the study period and 937 individuals recovered dead. The data (in form of m-arrays) are provided at the end of this Chapter.

Previous studies have shown that one and two year old birds are mainly resighted at rocks while individuals aged more than two years old are mainly resighted at the monitored breeding ledges ($\approx$89% of the $>$2 year old individuals observed; Kadin, 2007; Sarzo et al., 2019). The location of the birds is related to their life cycle. Guillemots are philopatric to both their breeding colony, and also to the sub-colony where they are born (Halley et al., 1995). Younger birds generally move around the colony and hence are observable, but as immature birds get older, they are increasingly present on breeding ledges, reflecting the prospecting behaviour in preparation for breeding (Oro and Pradel, 2000), and due to their philopatric behaviour remain at their chosen ledge. This latter behaviour leads to capture heterogeneity dependent on whether or not an individual settles at a monitored breeding ledge or not. Sarzo et al. (2019) showed that this behaviour combined with the given monitoring regime consequently leads to biased capture-recapture parameter estimates, with reduced survival probabilities for older ages (related to those resighted at the monitored breeding ledges) and increased capture probabilities at the monitored breeding ledges for the older ages. Additional ring-recovery data of (dead) individuals are collected by the Bird Ringing Centre mainly at the colony level and in the Baltic Sea so that recoveries are not affected by the partial monitoring as for the live resightings.

5.3 Methods

We initially describe the notation and model parameters before providing a description of the associated m-array summary statistics for MRR data and (efficient) likelihood function.

5.3.1 Notation

MRR data are typically displayed in the form of individual capture histories for each individual observed (alive or dead) within the study period. We let $i = 1, \dots, N$ denote the individuals marked in the study and $t = 1, \dots, T$ the associated capture occasions. We let $\boldsymbol{y}$ denote the $N \times T$ observed data matrix such that,

$$y_{i,t} = \begin{cases} 0 & \text{individual } i \text{ is not resighted at time } t; \\ 1 & \text{individual } i \text{ is resighted alive at time } t; \\ 2 & \text{individual } i \text{ is recovered dead in the interval } (t-1, t]. \end{cases}$$

Therefore, $\boldsymbol{y}_i = \{y_{i,t} : t = 1, \dots, T\}$ (i.e. the ith row of $\boldsymbol{y}$) corresponds to the capture history of individual i. We note that when an individual is recovered dead in the interval $(t-1, t]$, it is recorded as recovered at time t. For example consider the following two capture histories: $\boldsymbol{y_1} = \{1,\ 0,\ 0,\ 1,\ 0,\ 2\}$ and $\boldsymbol{y_2} = \{0,\ 1,\ 1,\ 0,\ 1,\ 0\}$.

The first capture history ($\boldsymbol{y_1}$) represents an individual first marked at occasion 1, subsequently observed at occasion 4 (they are unobserved at occasions 2, 3 and 5) before being recovered dead at occasion 6. The second history ($\boldsymbol{y_2}$) corresponds to an individual first marked at time 2, subsequently observed at times 3 and 5, (they are unobserved at occasions 4 and 6). For our case study, all the resighted individuals are observed only in monitored areas, however, individuals may be recovered dead (if they die) irrespective of whether or not they are in the monitored areas.

5.3.2 Model parameters

We initially describe the model parameters associated with the standard capture-recapture model and ring-recovery model in turn, before considering the integrated case. For capture-recapture data we define:

$$\begin{aligned}\phi_{a,t} &= \mathbb{P}(\text{an individual aged } a \text{ at time } t \text{ is alive and available for capture at} \\ &\qquad\qquad \text{time } t+1 \mid \text{alive and available for capture at time } t); \\ p_{a,t+1} &= \mathbb{P}(\text{an individual aged } a \text{ at time } t+1 \text{ is observed} \mid \text{alive and available} \\ &\qquad\qquad \text{for capture at time } t+1),\end{aligned}$$

for $t = 1, \dots, T-1$ and $a = 1, \dots, A$. We note that available for capture equates to the individual being present at the monitored colony locations. We note that $\phi_{a,t}$ is often referred to as the survival probability - however this is the *apparent* survival probability (death and availability/dispersal are confounded).

For ring-recovery data, we define the associated model parameters:

$$\begin{aligned}S_{a,t} &= \mathbb{P}(\text{an individual aged } a \text{ at time } t \text{ is alive at time } t+1 \mid \text{alive at} \\ &\qquad\qquad \text{time } t); \\ \lambda_{a,t+1} &= \mathbb{P}(\text{an individual aged } a \text{ is recovered dead in the interval } (t, t+1] \mid \\ &\qquad\qquad \text{dies in the interval } (t, t+1]),\end{aligned}$$

for $t = 1, \ldots, T-1$ and $a = 1, \ldots, A$. Again $S_{a,t}$ is often referred to as the survival probability, but in this case recovered individuals are known to be dead (rather than unavailable, which may include dispersal), and so this corresponds to the *true* survival probability.

Combining the capture-recapture data with the ring-recovery data we note the difference in interpretation between $\phi_{a,t}$ and $S_{a,t}$ in terms of apparent (including dispersal) and true survival, respectively. This directly permits us to estimate the dispersal probability, i.e. the probability an individual moves from a monitored site to an unmonitored site (where they are unavailable for resighting but available to recover, if they die). In particular we have the relationship $\phi_{a,t} = S_{a,t}(1-\gamma_{a,t})$, such that,

$$\begin{aligned}\gamma_{a,t} \quad = \quad & \mathbb{P}(\text{an individual aged } a \text{ is at an unmonitored site at time } t+1 \mid \text{alive} \\ & \qquad \text{at time } t+1 \text{ and in a monitored site at time } t),\end{aligned}$$

for $t = 1, \ldots, T-1$ and $a = 1, \ldots, A$. For notational convenience we let $\boldsymbol{S} = \{S_{a,t} : a = 1, \ldots, A; t = 1, \ldots, T-1\}$, and analogously for $\boldsymbol{p}, \boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$. Finally the set of all model parameters is denoted by $\boldsymbol{\theta} = \{\boldsymbol{S}, \boldsymbol{p}, \boldsymbol{\lambda}, \boldsymbol{\gamma}\}$. This ability to estimate additional parameters from integrated models (in our case $\boldsymbol{\gamma}$), which are inestimable in the individual models due to confounding, has been noted many times within the literature for different types of data (see for example, Burnham, 1993; Reynolds et al., 2009; McCrea et al., 2010; King and McCrea, 2014; McCrea and Morgan, 2015). Finally we note that a full age and time dependence on the model parameters may lead to parameter redundancy (Hubbard et al., 2014). We discuss in Section 5.4.1 the particular dependence structure we impose on the model parameters for our case study, based on biological knowledge, for which no such issues are identified.

5.3.3 Sufficient statistics

Capture-recapture data and ring-recovery data can each be usefully summarised in terms of *m*-arrays (Lebreton et al., 1992; McCrea and Morgan, 2015). The *m*-arrays summarize the number of individuals released at occasion $j = 1, \ldots, T-1$ and subsequently next observed alive (for capture-recapture data), or recorded to have been recovered dead (for ring-recovery data) at capture occasion $t =$

$2, \ldots, T$. Often, instead of recording the number of individuals released in a given year, the m-arrays are given with the (equivalent) number of individuals never observed again. We use this latter approach in the presentation of the m-arrays. We present the combined, or extended, m-array formulation for the combined MRR data before using these to form the efficient likelihood specification. With the additional age dependence this extends the data to a three-dimensional m-array, or alternatively, where we specify a separate extended m-array for each age that an individual may be released.

We define an $A \times (T-1) \times (2T-1)$ extended (or integrated) m-array with cell entries denoted by $m_{a,j,t}$ such that the cells are indexed by $a = 1, \ldots, A$, $j = 1, \ldots, T-1$ and $t = 2, \ldots, 2T$. For simplicity we condition on age, a, and consider the associated $(T-1) \times (2T-1)$ array. The first $T-1$ columns (labelled $t = 2, \ldots, T$) correspond to the capture-recapture components (i.e. live resightings), with cell entries, $m_{a,j,t}$, equal to the number of individuals that are observed at time $j = 1, \ldots, T-1$ (aged a) that are *next* observed alive at time $t = 2, \ldots, T$. The following $T-1$ columns (labelled $t = T+1, \ldots, 2T-1$) correspond to the ring-recovery components (i.e. dead recoveries) with cell entries, $m_{a,j,t}$, equal to the number of individuals that are observed at the time $j = 1, \ldots, T-1$ (aged a) that are *next* observed as a dead recovery at time $t^* = t - (T-1)$, for $t = T+1, \ldots, 2T-1$ (so that $t^* = 2, \ldots, T$). The final column (labelled $t = 2T$) has cell entries, $m_{a,j,2T}$, corresponding to the number of individuals last observed at time $j = 1, \ldots, T-1$ (aged a) that are never observed again within the study (either alive or dead). Finally, we let $q_{a,j,t}$ denote the associated cell probabilities of the integrated *m*-array, for $a = 1, \ldots, A$, $j = 1, \ldots, T-1$ and $t = 2, \ldots, 2T$. Finally, for notational purposes we let the set of sufficient statistics be denoted by $\boldsymbol{m} = \{m_{a,j,t} : a = 1, \ldots, A; j = 1, \ldots, T; t = 2, \ldots, 2T\}$, and similarly for $\boldsymbol{q}$ (the set of cell probabilities).

For illustration, we provide a toy numerical example where $T = 5$ and there is a single age for all individuals (e.g. all individuals are ringed and observed as adults), so that $A = 1$. Table 5.1 provides the capture histories of a set of individuals; while Table 5.2 the corresponding set of m-arrays associated with only capture-recapture data (no dead recoveries); only ring-recovery data (no live resightings); and the combined MRR data.

Capture occasion						Capture occasion					
1	2	3	4	5	Number	1	2	3	4	5	Number
1	0	0	0	0	1466	0	1	1	1	0	25
1	1	0	0	0	101	0	1	1	1	2	2
1	2	0	0	0	13	0	1	0	1	0	43
1	1	1	0	0	15	0	1	0	2	0	10
1	1	2	0	0	1	0	1	0	0	1	5
1	1	0	1	0	13	0	0	1	0	0	1251
1	1	0	0	2	1	0	0	1	1	0	98
1	0	1	0	0	76	0	0	1	2	0	17
1	0	2	0	0	9	0	0	1	1	1	23
1	0	1	1	0	8	0	0	1	1	2	2
1	0	0	1	0	42	0	0	1	0	1	15
1	0	0	0	1	12	0	0	1	0	2	5
0	1	0	0	0	1001	0	0	0	1	0	1235
0	1	1	0	0	94	0	0	0	1	1	68
0	1	2	0	0	18	0	0	0	1	2	10

TABLE 5.1: Toy example capture histories, where 0 denotes unobserved; 1 observed alive; and 2 recovered dead; and associated number of individuals with given history.

In general we note that the observed m-array cell entries associated with the resighting of individuals do not change between the capture-recapture only m-array and the integrated MRR m-array (corresponding to columns labelled $t = 2, \dots, T$). However, for the m-array cells corresponding to the recovery of individuals, these do differ between the ring-recovery only data and the integrated MRR data (corresponding to columns labelled $t = T+1, \dots, 2T-1$), as this is dependent on whether an individual *is* resighted prior to recovery. If a bird is resighted, then the year of release for the individual will differ between the ring-recovery only m-array and integrated m-array; alternatively if the individual is not resighted alive within the study the contribution to the ring-recovery component of the integrated m-arrays remains the same. Similarly the number of individuals not observed again following release will differ in the integrated MMR m-array to the individual m-arrays to take into account the additional resightings or recoveries not observed within the individual ring-recovery and capture-recapture only m-arrays, respectively (columns headed “Never seen” in the example m-arrays in Table 5.2).

Capture-recapture only m-array

	Capture occasion				
	2	**3**	**4**	**5**	**Never seen**
1	131	84	42	12	1488
2	–	136	56	5	1132
3	–	–	158	15	1458
4	–	–	–	91	1478

Ring recovery only m-array

	Capture occasion				
	2	**3**	**4**	**5**	**Never seen**
1	13	10	0	1	1733
2	–	18	10	2	1168
3	–	–	17	7	1387
4	–	–	–	10	1303

Integrated m-array

	Capture-recapture component				**Ring recovery component**				
	Capture occasion								
	2	**3**	**4**	**5**	**2**	**3**	**4**	**5**	**Never seen**
1	131 $(q_{1,1,2})$	84 $(q_{1,1,3})$	42 $(q_{1,1,4})$	12 $(q_{1,1,5})$	13 $(q_{1,1,6})$	9 $(q_{1,1,7})$	0 $(q_{1,1,8})$	0 $(q_{1,1,9})$	1466 $(\chi_{1,1})$
2	–	136 $(q_{1,2,3})$	56 $(q_{1,2,4})$	5 $(q_{1,2,5})$	–	19 $(q_{1,2,7})$	10 $(q_{1,2,8})$	1 $(q_{1,2,9})$	1102 $(\chi_{1,2})$
3	–	–	158 $(q_{1,3,4})$	15 $(q_{1,3,5})$	–	–	17 $(q_{1,3,8})$	5 $(q_{1,3,9})$	1436 $(\chi_{1,3})$
4	–	–	–	91 $(q_{1,4,5})$	–	–	–	12 $(q_{1,4,9})$	1464 $(\chi_{1,4})$
	Column label (t)								
	2	3	4	5	6	7	8	9	10

TABLE 5.2: Example of m-arrays for capture-recapture (top left), ring-recovery (top right) data and the resulting integrated m-array (bottom) with the observed individuals and cell probabilities (parenthesis) for $T = 5$ and a single adult age category. A '–' denotes a structural zero entry (and hence cell probability is equal to 0).

Finally we also note that the number of individuals released each year (equal to the row sums) differs between the capture-recapture only m-array and the ring-recovery only m-array as birds that are resighted in year t, say, are then counted within the number of birds released in year t. However, for the ring-recovery only m-array there are no resightings, and so all releases correspond to initial captures. The row sums for the capture-recapture only and integrated m-arrays are however equal, with the difference between these m-arrays corresponding to what happens following release in terms of whether dead recoveries are recorded or not.

5.3.4 Likelihood

The likelihood is calculated as a product over multinomial terms. Each row, $j = 2, \dots, T$ of the integrated m-array (for a given age, $a = 1, \dots, A$) has a multinomial distribution, with sample size equal to the number of individuals released at time j, denoted $R_{a,j}$, and associated cell probabilities $\boldsymbol{q}_{a,j} = \{q_{a,j,t} : t = 2, \dots, 2T\}$. We note that $R_{a,j} = \sum_{t=2}^{2T} m_{a,j,t}$. Thus, the likelihood can be expressed in the form:

$$f(\boldsymbol{m}; \boldsymbol{\theta}) = \prod_{a=1}^{A} \prod_{j=1}^{T-1} \frac{R_{a,j}!}{\left(\prod_{t=2}^{2T} m_{a,j,t}!\right)} \left(\prod_{t=j+1}^{2T-1} q_{a,j,t}^{m_{a,j,t}} \right) \chi_{a,j}^{m_{a,j,2T}}. \tag{5.1}$$

To specify the cell probabilities we initially consider two separate cases corresponding to the resightings and recoveries, before we consider the final column corresponding to those individuals not observed again following their release (the final term in Equation (5.1)).

Case I: Columns $t = 2, \dots, T$ (live resightings)

We initially consider the probability an individual released at time $j = 1, \dots, T-1$, aged $a = 1, \dots, A$ is next observed alive again at time $t = 2, \dots, T$, denoted by $q_{a,j,t}$. In this case the individual must survive each time between release to resighting, remain available for capture, not be observed at times $j+1, \dots, t-1$, before being resighted again at time t, so that,

$$q_{a,j,t} = \prod_{k=j}^{t-1} \{S_{a+k-j,k}(1 - \gamma_{a+k-j,k})\} \times \prod_{k=j}^{t-2} (1 - p_{a+k-j+1,k+1}) \times p_{a+t-j,t},$$

where we use the convention throughout that the null product $\prod_{k=j}^{j-1} \equiv 1$. The first product corresponds to individuals surviving and remaining available for capture within the monitored areas from time j to t; the second product corresponds to the observation process (not being observed between times $j+1$ to $t-1$ before being observed at time t). Recall that, apparent survival, $\phi_{a,t}$, can be expressed as a simple function of the true survival probability, $S_{a,t}$, and dispersal probability,

$\gamma_{a,t}$, such that $\phi_{a,t} = S_{a,t}(1 - \gamma_{a,t})$. Substituting this relationship into the expression for $q_{a,j,t}$ gives the "standard" capture-recapture only m-array cell probability. Finally, we note that as is common in capture-recapture studies, we assume that there is no temporary dispersal, so that once an individual leaves the monitored sites (monitored rocks or breeding ledges) they are not able to return. This seems reasonable for the given case study with the philopatric nature of individuals once they settle at a given location within their life cycle.

Case II: Columns $t = T + 1, \ldots, 2T - 1$ (dead recoveries)

We now consider the dead recovery component. We note that the column labels $t = T + 1, \ldots, 2T - 1$ correspond to the capture occasions $2, \ldots, T$. Thus, we define t^* to be the capture occasion associated with column label t, such that $t^* = t - T + 1$, for $t = T + 1, \ldots, 2T - 1$. Now, consider the probability an individual observed at time $j = 1, \ldots, T - 1$ at age $a = 1, \ldots, A$ is unseen until recovered dead in the interval $(t^* - 1, t^*]$. This means that an individual survives between times j to $t^* - 1$ before dying within the interval $(t^* - 1, t^*]$ and is recovered at this time (we make the standard ring-recovery assumption that an animal that dies in the interval $(t^* - 1, t^*]$ can only be recovered in the same interval, due to mark loss after this period). To account for not observing the individual at times $j + 1, \ldots, t^* - 1$ we also need to take into account that an individual may move away from the monitored locations, where they are not observable via the capture-recapture study, but still available to be recovered dead if they die. Accounting for these different processes, the cell probabilities can be expressed in the form,

$$\begin{aligned} q_{a,j,t} &= (1 - S_{a+t^*-1-j,t^*-1})\lambda_{a+t^*-j,t^*} \times \left(\prod_{k=j}^{t^*-2} S_{a+k-j,k} \right) \\ &\quad \times \left\{ \sum_{k=j}^{t^*-2} \left((\gamma_{a+k-j,k}) \prod_{n=j}^{k-1} (1 - \gamma_{a+n-j,n})(1 - p_{a+n-j+1,n+1}) \right) \right. \\ &\quad \left. + \left(\prod_{n=j}^{t^*-2} (1 - \gamma_{a+n-j,n})(1 - p_{a+n-j+1,n+1}) \right) \right\}, \end{aligned}$$

where $t^* = j - T + 1$. The first line corresponds to the probability that an individual released at time j aged a survives until time $t^* - 1$, before subsequently dying in the interval $(t^* - 1, t^*]$ and being recovered. The second and third lines correspond to the probability an individual released at time j aged a is not observed alive prior to being recovered dead at time t^*. However, for such an individual that is not resighted we need to account for the partial monitoring - an individual may move to an unmonitored site at each capture occasion. While they remain at the monitored sites they are available for capture (but are not observed); once they disperse to an unmomitored site they are no longer available for capture. Thus the associated probability of the resighting process is calculated by summing over all possible times an individual may disperse to an unmonitored site following release (line 2), or remains throughout at the monitored site (line 3), until its death.

Case III: Columns $t = 2T$ (not observed again)

We first note that the lower triangular elements of the resighting and recovery components are structural zeros (i.e. impossible events - an individual cannot be resighted or recovered again prior to or at its time of release). This means that for each $j = 1, \dots, T-1$, and $t = 1, \dots, j, T, \dots, T-1+j$, we have that, $q_{a,j,t} = 0$. Then, since an individual released at time j must either be observed again (either alive or dead), or not observed, the row cell probabilities must sum to unity. Thus, the probability of an individual released at time j, aged a, is not observed again can be expressed as,

$$\chi_{a,j} = 1 - \sum_{t=2}^{2T-1} q_{a,j,t}.$$

Substituting the $q_{a,j,t}$ and $\chi_{a,j}$ terms into Equation (5.1) provides an explicit expression for the likelihood function. We note that setting $\gamma_{a,j,t} = 0$ for all $a = 1, \dots, A$, $j = 1, \dots, T-1$ and $t = 2, \dots, T$ reduces the likelihood to the standard (non partial monitoring) MRR likelihood where the survival probabilities are assumed to be equal for the capture-recapture and ring-recovery components (see for example, Catchpole et al., 1998, though the likelihood specified is not expressed using the multinomial form).

5.4 Case Study

We consider the data relating to guillemots described in Section 5.2. The MRR data are collected from 1992-2018 (so that $T = 27$), where the live resightings only occur within the period 2002-2016; and dead recoveries are available throughout the whole study duration, 1993-2018.

5.4.1 Model

We initially consider the dependence of the model parameters on age and time, based on biological knowledge and study protocols. Due to the life cycle of guillemots we consider the maximum set of distinct ages, $1, 2, 3, 4, 5+$, where age corresponds to "year of life" and 5+ corresponds to (adult) individuals of at least age 5.

Firstly we consider the observation processes. The capture-recapture database includes resighting events between 2002-2016. Thus for years where there are no resightings possible, we set the resighting probability equal to 0, i.e. $p_{a,t} = 0$ for $t = 2, \ldots, 10, 26, 27$, corresponding to years $1993, \ldots, 2001; 2017, 2018$, for $a = 2, \ldots, 5+$. During the capture-recapture study period resightings occurred during the breeding season (May to July) but under two different intensity regimes of low and high effort. The low effort regime lasted from 2002-6 (mean resighting effort of 35.80 hours per year); while from 2007-16 a higher effort was used (mean resighting effort of 82.8 hours per year). Thus we consider two distinct resighting probability regimes relating to the different effort periods. Further, the behaviour of individuals is such that most juveniles (ages 1 and 2) are generally observed on rocks (48$\%$ of resightings), while ages 3+ are increasingly present on the breeding ledges, first to learn competencies for breeding and when they reach sexual maturity, to breed. Thus to allow for the differing proportions of individuals who may use the rocks/ledges in early adulthood, the resighting probabilities are assumed to be fully age dependent. Thus we specify the capture probabilities such that $p_{a,t} = p_a^*$ for $t = 11, \ldots, 15$ (i.e. years 2002-6) and $p_{a,t} = p_a$ for $t = 16, \ldots, 25$ (i.e. years 2007-16); for $a = 2, \ldots, 5+$. In contrast, the recovery effort was constant throughout the period, and with no rationale to suggest the age of the death of an individual would influence whether or not they are recovered, we assume a constant recovery probability across all ages and time, such that $\lambda_{a,t} = \lambda$ for all a and t. For the survival probabilities, it is well understood that these typically vary inter-annually due to variation in food availability and environmental conditions (Votier et al., 2008). Similarly, age is generally a very important factor affecting survival (see for example, Crespin et al., 2006; Harris et al., 2007). We assume the presence of both age and time effects, where they are additive (on the logit scale) and set $\text{logit}(S_{a,t}) = \alpha_a + \beta_t$ for $a = 1, \ldots, 5+$ and $t = 1, \ldots, T-1$. However, we note that sexual maturity of guillemots occurs at ages 4-6 (Birkhead and Hudson, 1977b), and so we consider the reduced age structure and set $\alpha_4 = \alpha_{5+}$, which we denote by α_{4+}.

Finally, for the dispersal probabilities it is expected that the majority of these will occur as individuals begin to increase their use of breeding ledges (i.e. breeding dispersal), although there will also be some additional natal dispersal in younger birds. For the given colony, individuals generally start to visit breeding ledges from 3 years of age, so that for the dispersal probabilities we consider the ages structure, $1, 2, 3+$ (where ages 1 and 2 we refer to as natal dispersal; and for age 3+ as breeding dispersal). We note that we assume dispersal from the monitored locations are permanent, as individuals once settled are philopatric.

Thus, the full set of model parameters in the proposed integrated model is: $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_{4+}\}$, $\boldsymbol{\beta} = \{\beta_1, \ldots, \beta_{T-1}\}$, $\boldsymbol{p} = \{p_2^*, \ldots, p_{5+}^*, p_2, \ldots, p_{5+}\}$, $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2, \gamma_{3+}\}$ and λ. Note that we comment further on the given dependence structure in Section 5.4.3.

5.4.2 Bayesian modeling

We consider a Bayesian modelling approach, which permits us to formally incorporate some (weak) prior information on the structure of the capture probabilities. We have no prior information on the resighting probabilities for the period 2007-16, or the recovery probabilities so that we specify $p_a \sim U[0, 1]$ for $a = 2, \ldots, 5+$ and $\lambda \sim U[0, 1]$, respectively. However, for the earlier period 2002-6, with the significantly lower number of resighting hours, we expect the associated resighting probabilities to be less than the later study years, and so specify the conditional prior distribution, $p_a^* | p_a \sim U[0, p_a]$, for $a = 2, \ldots, 5+$.

For the survival parameters, there are several years with relatively limited amount of data. For birds ringed prior to 2006, there are very few individuals resighted (a total of 986 individuals, corresponding to an average of 66 individuals per year), and few recoveries in future years. Thus, we consider a hierarchical structure to model the time effects, so that information can be borrowed across the whole study period and specify, $\beta_t \sim N(\mu, \sigma^2)$, where $\mu \sim N(0, 10)$ and $\sigma^2 \sim \Gamma^{-1}(0.001, 0.001)$. For the age effects, for identifiability, we set $\alpha_1 = 0$, and specify non-informative priors on the remaining terms, such that $\alpha_a \sim N(0, 5)$ for $a = 2, \ldots, 4+$. Thus we note that α_a is interpreted as the age effect relative to age, for $a = 2, \ldots, 4+$. A summary of the priors specified on the parameters is provided in Table 5.3

Bayesian inference was performed using Markov chain Monte Carlo (MCMC) (Robert and Casella, 2011) using the `jags` software (Plummer, 2003) within `R` (R Core Team, 2019). The simulations were run for 400,000 iterations using three chains, and over-dispersed starting values. A conservative burn-in of 40,000 iterations were discarded. Essentially identical results were obtained from each chain and no convergence issues identified using the BGR statistic (Gelman et al., 2014).

Parameter	**Prior distribution**
p_a	$U[0, 1]$
$p_a^* \vert p_a$	$U[0, p_a]$
λ	$U[0, 1]$
α_a	$N(0, 5)$
$\beta_t \vert \mu, \sigma^2$	$N(\mu, \sigma^2)$
γ_a	$U[0, 1]$
Hyperparameter	
μ	$N(0, 10)$
σ^2	$\Gamma^{-1}(0.001, 0.001)$

TABLE 5.3: Prior distributions specified on the model parameters and associated hyperparameters.

5.4.3 Results

We consider both the integrated MRR data and the capture-recapture only data for comparison. Figure 5.2 provides the posterior mean and associated 95% symmetric credible intervals for the survival probabilities, $S_{a,t}$, for $a = 1, \dots, 4+$ and $t = 1, \dots, T-1$. Table 5.4 provides the associated posterior mean and 95% symmetric credible intervals (CIs) for the α_a (for $a = 2, \dots, 4+$) and β_t (for $t = 1, \dots, T-1$) parameters. We note that for the capture-recapture analysis we only use the data from 2006-2016, due to very few individuals released prior to 2006. Further the estimates for the capture-recapture only data refer to *apparent* survival, whereas for the integrated data these correspond to *true* survival, correcting the biases present in the capture-recapture data due to the partial monitoring of the study site. We observe that the survival probabilities are very similar for ages 3 and 4+ for the integrated analysis. Further, the difference between the apparent survival and true survival estimates are most marked for ages 3+. This is unsurprising as this is the age where the partial monitoring has most impact, when individuals are more likely to move to breeding ledges away from the monitored locations (i.e. move from the resighting monitored sites). We note that for the capture-recapture-only analysis the *apparent* survival probabilities for age 2 individuals are unrealistically high (and higher than the true survival probability estimates in the integrated analysis). For the capture-recapture data relatively few individuals are observed at the first year of life (i.e. p_2); while a significantly greater number of individuals are observed aged 2 (i.e. p_3). This leads to a reduced (apparent) first-year survival probability (and hence expected number of birds at age 2), which to then account for the large number of birds observed at age 2 leads to the very large (and unrealistic) second-year survival probabilities. However, including the additional dead recoveries provides additional information in relation to the survival probabilities, leading to an increased first-year survival probability which in turn leads to more realistic age 2 survival probabilities to fit the large number of observed individuals at age 2. Further we note that for the integrated analysis, as would be expected biologically, $\mathbb{E}_\pi(\alpha_i) < \mathbb{E}_\pi(\alpha_j)$ for $i < j$, so that survival is increasing with age for younger birds until adulthood. Finally, we observe that all the survival probabilities are significantly reduced in 1999. This is likely due to in 1998 there was an avian cholera outbreak in this colony (Österblom et al., 2004).

Parameters	**Mean**	**95% CI**
α_2	1.3229	[1.0159, 1.6301]
α_3	1.5851	[1.2645, 1.9234]
α_{4+}	1.9186	[1.6415, 2.2036]
β_1	1.4285	[0.2798, 2.7788]
β_2	0.7833	[-0.3613, 2.0526]
β_3	1.8220	[0.5949, 3.3563]
β_4	0.8539	[-0.0682, 1.8581]
β_5	-0.6996	[-1.5265, 0.1897]
β_6	0.3576	[-0.5183, 1.3265]
β_7	0.5759	[-0.1960, 1.4469]
β_8	-1.5185	[-2.1863, -0.8935]
β_9	0.4803	[-0.0374, 1.0392]
β_{10}	-0.1640	[-0.5435, 0.2349]
β_{11}	-0.1342	[-0.5579, 0.3236]
β_{12}	0.0278	[-0.3023, 0.3688]
β_{13}	-0.0850	[-0.3432, 0.1936]
β_{14}	0.4477	[0.2123, 0.7101]
β_{15}	1.7642	[1.3515, 2.2043]
β_{16}	1.1945	[0.8500, 1.5895]
β_{17}	1.4474	[1.0163, 1.9279]
β_{18}	-0.3267	[-0.5151, -0.1226]
β_{19}	0.6430	[0.3839, 0.9215]
β_{20}	0.2781	[0.0460, 0.5254]
β_{21}	0.3693	[0.1213, 0.6277]
β_{22}	0.6018	[0.3418, 0.8869]
β_{23}	0.9734	[0.6005, 1.3947]
β_{24}	1.3904	[0.8441, 1.9976]
β_{25}	0.8114	[0.1824, 1.4674]
β_{26}	2.5583	[1.4207, 4.0125]

TABLE 5.4: Mean and 95% symmetric credible intervals (CIs) for α_a ($a = 2, \ldots, 4+$; $\alpha_1 = 0$) and β_t ($t = 1, \ldots, T-1$) parameters obtained through the integrated model.

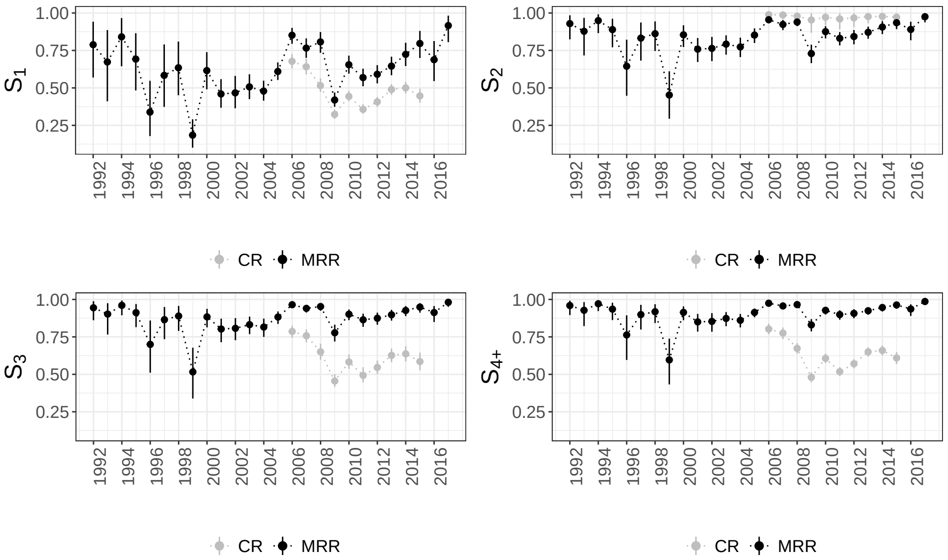

FIGURE 5.2: Posterior mean and 95% symmetric credible intervals (CIs) for the survival probabilities by age classes ($a = 1, \dots, 4+$) for the study period, obtained through capture-recapture (CR) data only (grey) and integrated MRR data (black).

We note that we can compare the estimated adult (*true*) survival probabilities with previous estimates for guillemot colonies. For example, for this colony, Österblom et al. (2004) suggested high adult survival estimates, of 0.937 (0.91, 0.96). Alternatively, for the Isle of May guillemot colony, adult (age 4+) survival probabilities were estimated to have a posterior mean of 0.912 with 95% posterior credible interval (0.872, 0.941) (Reynolds et al., 2009). Thus, the survival estimates obtained from this integrated MRR study are comparable; however, as expected (and noted above), the capture-recapture only estimates are significantly lower (except for two year old individuals) as these are *apparent* survival probabilities and are confounded with dispersal.

The posterior mean and 95% posterior credible intervals for the resighting, recovery and dispersal parameters are given in Table 5.5. The resighting probabilities for the early years of the study (with low resighting effort) are all estimated to be very small, as expected. For the later years (with increased resighting effort) there is clear, strong, age dependence across the resighting probabilities in the later years. Age 1 individuals are significantly less likely to be resighted within the study. The resighting probability increases for age 2 birds and again for individuals of ages 3 onwards, for the integrated analysis. There is very little difference in the resighting probabilities for those aged 3 and 4+, suggesting that individuals are equally observable from the age of 3, if they remain in the monitored locations. This is most likely due to young guillemots (mainly 1 and 2 years old) spending time at sea, so that, the proportion of individuals resighted at these ages is commonly low (specially for first year old individuals, Harris et al., 2007). Comparing the resighting probabilities between the integrated model and capture-recapture only model it is noticeable that the estimates are generally smaller in the capture-recapture only model (the exception is for age 1). In particular, we note that the largest difference is for adult birds (aged 4+), where the credible intervals between the integrated analysis, and capture-recapture only analysis, are non-overlapping. For capture-recapture(-recovery) studies the survival probabilities are negatively correlated with the resighting probabilities: lower resighting probabilities lead to higher survival probabilities (if the probability of seeing an individual lower, this means that more individuals are alive in the study area but not seen, hence leading to an increased survival probability). Conversely, higher resighting probabilities (as for the capture-recapture only analysis) generally leads to lower survival estimates. Finally we note that the recovery probability is relatively small, as is often the case with ring-recovery studies.

As noted previously, one common advantage of integrated models is that they permit the estimation of parameters not otherwise estimable from the data. In our case, combining the capture-recapture data with the ring-recovery data permits the estimation of the dispersal probabilities, γ_a, to an unmonitored areas. Harris et al. (2007) and Crespin et al. (2006) suggested that prospecting for breeding sites is biased towards the natal colony; and prebreeders who decide to disperse do so if conditions are poor at their natal site or if they are unable to secure a site. From Table 5.5 it is clear that dispersal is most pronounced for individuals aged age 3+, as expected, given their general behaviour. In general, individuals aged 1 and 2 temporarily visit rocks; individuals aged 2 and 3 will start to visit breeding ledges; while individuals aged 3 or more are more likely to be at breeding ledges and start pair-bonding after which they are philopatric to the given ledge. While individuals at the monitored rocks are more likely to be observable, not all breeding ledges are monitored, and so if an individual settles at one of these they are no longer

	Integrated		**Capture-recapture only**	
Parameter	Mean	95% CI	Mean	95% CI
p_2^*	0.0191	[0.0143, 0.0246]	–	–
p_3^*	0.0328	[0.0245, 0.0424]	–	–
p_4^*	0.0290	[0.0179, 0.0434]	–	–
p_{5+}^*	0.0209	[0.0110, 0.0337]	–	–
p_2	0.0612	[0.0552, 0.0673]	0.0717	[0.0662, 0.0774]
p_3	0.2699	[0.2569, 0.2831]	0.2471	[0.2332, 0.2623]
p_4	0.3787	[0.3616, 0.3965]	0.3520	[0.3303, 0.3740]
p_{5+}	0.3658	[0.3483, 0.3837]	0.4070	[0.3840, 0.4300]
λ	0.0153	[0.0132, 0.0176]	–	–
γ_1	0.1958	[0.1057, 0.2725]	–	–
γ_2	0.0717	[0.0066, 0.1526]	–	–
γ_{3+}	0.3187	[0.2967, 0.3397]	–	–

TABLE 5.5: Posterior mean and 95% symmetric credible intervals (CIs) for the resighting probabilities (p_a^* and p_a, for $a = 2, \dots, 5+$), recovery probability (λ) and dispersal probabilities (γ_a, for $a = 1, \dots, 3+$) for the integrated model; and capture-recapture only model.

available for resighting (but are available for recovery if they die). Extending the model with separate age-dependent dispersal probabilities, 3, 4+ led to similar probabilities for the specific ages.

Finally, we note that we conducted a prior sensitivity analysis. The corresponding results were generally robust to different uninformative priors, with the interpretation remaining consistent throughout.

5.5 Discussion

Reliable estimates of demographic parameters are essential to understand dynamics of wildlife populations. Capture-recapture (and/or ring-recovery) are commonly used across a wide range of species to study populations, often with the aim of estimating survival probabilities, which can be important for conservation and/or management. Within such studies there can be many different practical issues within the data collection process. We focus on the particular issue of partial monitoring populations, where live resightings are only observed at a small subset of the given colony. The approach developed is motivated by a colony of guillemots on the island of Stora Karlsö, the largest colony in the Baltic Sea (Olsson and Hentati-Sundberg, 2017). For this colony, the birds are ringed as chicks and are highly philopatric to their ledges, once they reach maturity. Thus, individuals may permanently move to locations that are outside the monitored locations, at which they point they are no longer able to be observed. This typically leads to biased estimates of survival (Sanz-Aguilar et al., 2016; Sarzo et al., 2019), with true survival and movement away from the monitored locations confounded.

To address the issue, we use additional ring-recovery data within an integrated modelling framework to permit the estimation of both dispersal and *true* survival probabilities, where the majority of dispersal is assumed to be within-colony with few birds recovered long-distances from the colony. We derive the associated efficient integrated MMR m-array likelihood expression in the presence of dispersal, corresponding to a product of multinomial distributions where the model parameters are both age and time dependent. This likelihood formulation also permits the use of standard goodness-of-fit tests to be applied (McCrea et al., 2010).

Further, for these data, additional modelling considerations are specified, due to the particular structure of the guillemot study and resulting small sample size of the data within time periods of the study. In particular, for the early years of the study (1993-2001) there are no live resightings, and the number of birds released (and recovered) are generally low. Thus, we use a hierarchical structure for the year effects for the survival probabilities which permits the borrowing of information across years, significantly improving the precision of the estimated survival probabilities for the earlier years. The corresponding adult survival estimates for the integrated MRR analysis are significantly higher than for the capture-recapture only data, reflecting the different between *true* survival, compared to *apparent* survival, and correcting for the additional dispersal from the monitored locations. In addition the integrated analysis led to biologically plausible survival estimates for age 2 individuals, whose estimates are unrealistically high for the capture-recapture only data (as discussed in Section 5.4.3). The newly derived true survival estimates are comparable to other studies of guillemot colonies (for example, Reynolds et al., 2009; Österblom et al., 2004). Further, the survival probabilities vary over years, which may be related to additional factors, such as prey availability and/or environmental conditions. Investigating such relationships is the focus of current research. Finally we note that the age-specific dispersal probabilities of interest which largely distinguishes between natal and breeding dispersal, with the largest dispersal estimated to be for ages 3+ corresponding to when individuals start to reach maturity (breeding dispersal).

Long-term studies are essential for assessing trends over time and/or permitting the identification of relationships with possible factors. Combining studies of differing lengths can extend the possible inferences, and a hierarchical modelling structure permits borrowing of information where less information is available to improve the precision of the estimates over these periods of time. We use a hierarchical structure to model year effects due to the heterogeneity in the amount of information contained in the data over the years due to periods of low sample size. Individual heterogeneity can also be an important biological modelling component, and hierarchical structures are often implemented to model these. However, for these data (with approximately 30,000 individuals), fitting these individual heterogeneity models using standard techniques, such as Bayesian data augmentation (often referred to as a complete data likelihood approach), leads to significant, and potentially infeasible, computational expense. Consequently, further computational techniques are needed for such analyses, which is an area of current focus.

5.6 Integrated *m*-arrays

In Tables 5.6 to 5.9 we present the four integrated age *m*-arrays for the mark-recapture-recovery database, relating to guillemots.

The study presented in this Chapter has been submitted to an indexed journal with title: "Correcting bias in survival probabilities for partially monitored populations via integrated models".

	Age 1																																																	
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	4	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	435
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	204
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	2	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	449
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	2	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	409
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	2	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	414
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	7	1	2	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	482
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	261
9	0	0	0	0	0	0	0	0	0	5	1	0	0	0	11	4	4	2	1	2	1	0	1	2	0	0	0	0	0	0	0	0	9	8	2	1	4	1	0	2	0	0	0	1	0	0	1	0	0	0	932
10	0	0	0	0	0	0	0	0	0	5	0	3	1	7	18	3	1	0	6	2	2	0	1	0	0	0	0	0	0	0	0	0	0	15	1	6	4	1	0	0	0	0	1	0	0	0	0	0	0	0	922
11	0	0	0	0	0	0	0	0	0	0	1	10	1	5	11	5	0	2	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	5	1	1	7	0	0	0	0	0	0	0	0	0	0	0	0	946
12	0	0	0	0	0	0	0	0	0	0	0	15	8	15	61	16	2	2	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	4	7	0	0	0	1	1	0	0	0	0	0	0	0	1355
13	0	0	0	0	0	0	0	0	0	0	0	0	11	33	119	25	3	4	2	2	1	0	3	5	0	0	0	0	0	0	0	0	0	0	0	0	13	10	0	0	0	0	1	0	0	0	0	0	0	0	1263
14	0	0	0	0	0	0	0	0	0	0	0	0	0	23	281	159	27	14	18	3	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	2	1	1	0	0	0	0	0	0	0	2	0	1442
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	174	361	100	62	21	7	3	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	4	2	1	0	0	0	0	1	0	0	0	1240
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	50	290	175	51	18	8	5	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	3	3	1	0	1	0	0	0	1	0	1340
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	101	246	123	49	13	3	12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	6	2	1	0	0	1	0	0	0	1424
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	57	110	126	65	29	11	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	22	1	4	0	5	0	0	1	0	2553
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	44	297	212	111	42	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	1	0	7	1	0	0	0	2243
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	61	200	186	99	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	1	1	1	0	0	2418
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	272	174	80	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	0	2	0	0	2273
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	149	310	192	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	1	0	0	0	2448
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	111	389	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	1	0	2578
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	173	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	4779

TABLE 5.6: Age 1 integrated *m*-array.

	Age 2																																																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	6
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	80	20	4	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	67
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	32
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	10	4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	31
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	5	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	90

TABLE 5.7: Age 2 integrated *m*-array.

	Age 3																																																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
14	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	121	11	3	1	2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	151
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	163	35	6	2	1	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	227
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	82	16	5	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	195
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	64	13	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	195
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	5	2	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	16	9	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	222
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	6	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	152
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	213
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	64	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	284

TABLE 5.8: Age 3 integrated *m*-array.

	Age 4+																																																		
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	67	1	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	179
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	21	21	4	6	3	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	269
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	147	28	8	3	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	223
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	164	26	8	4	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	351
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	153	25	12	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	341
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	130	24	10	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	276
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	152	35	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	340
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	165	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	410
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	189	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	450

TABLE 5.9: Age 4+ integrated *m*-array.

CHAPTER 6

FINAL REMARKS AND FUTURE WORK

In this thesis we have intended to offer three specific contributions to the field of the statistical methods for capture-recapture(-recovery) data from a Bayesian perspective. Firstly, we have provided reliable juvenile survival estimates for common guillemots, a difficult issue to assess in seabird species. Further, we have detected a problem of partial monitoring in this colony which leads to biased survival parameter estimates. Partial monitoring is widespread in colonial species but up to now, there were not statistical models to correct it. We have presented a first approximation to this problem adopting a Bayesian subjective approach so that prior information about the colony is incorporated. The results obtained improve the parameter estimation, however, adult survival probabilities are still underestimated despite our correction.

The second big aim of this thesis came naturally, and it was to correct the problem of partial monitoring in colonial species. We went a step further and we analysed jointly two data sets from the same individuals via integrated models. The information gathered by the ring-recovery database have shed light to individuals settled both at monitored areas and not monitored. Therefore, this additional information about dead individuals reduces the corresponding bias obtained in the survival probability estimates obtained from capture-recapture data only due to the confounding with migration. Further, we have derived an explicit efficient likelihood expression to the integrated model proposed and have stated the associated sufficient statistics.

Finally, the last aim of this thesis was to bring together two (a priori) different areas with a common primary goal, the survival analysis. In these two areas (human demography and population ecology) it is often the case that data collection protocols are different, resulting in different models and methods of analysis. In this thesis we have adopted a methodology commonly performed in medical survival studies in the ecological capture-recapture framework, the elicitation of the temporal scale. Hence, we have presented the Cormack-Jolly-Seber model (although it can be extended to other models) in an alternative temporal scale, the age. Moreover, we have also discussed differences and similarities regarding data presentation, model parameters, model formulation and results.

Once again we point out that using Bayesian inference has allowed us to make inference with relative ease on these models. The results in Chapter 3 also underline the usefulness of Bayesian methods when incorporating prior information. Further, the Bayesian hierarchical modelling performed in Chapter 5 has allowed to borrow the different temporal information gathered by two databases. To conclude, the posterior probabilities obtained through Bayesian paradigm provide a full picture of what is known about each parameter based on the model and the data, together with any prior information in form of probability distributions.

In overall, in this PhD, we have proposed a number of models that have quite effectively tackled some challenges in population ecology. However, the scope of research is still extensive. In what follows we provide a list of topics that we consider of special interest in this context:

1. **Relating survival with prey abundance**. We want to investigate factors affecting the temporal variation in survival probabilities observed in the integrated model presented in this thesis. In particular, our main focus is to associate those changes in relation to the abundance of sprat, the main prey of common guillemots in the Baltic Sea.

2. **Individual heterogeneity**. The size of the databases has been a constant 'issue' along this thesis. For these kind of data (big), the inclusion of individual

heterogeneity using standard techniques, such as Bayesian data augmentation, leads to significant, and potentially infeasible, computational expense. Therefore, one of our proposals (and an area of current research), is to develop computationally efficient approaches for such models via importance sampling approach (a joint work with Professor Ruth King, University of Edinburgh).

3. **Modelling communities**. Conservation cannot be focused in one species at a time, however, it requires the adoption of multispecies or entire-community approaches. Community-level modelling offers an opportunity to move beyond species-level predictions and to predict broader impacts of environmental changes. This is a recent research area that we consider must become an important focus to further research.

References

Abadi, F., Giménez, O., Ullrich, B., Arlettaz, R., and Schaub, M. (2010). Estimation of immigration rate using integrated population models. *Journal of Applied Ecology*, 47(2):393–400.

Aebischer, N. J. and Coulson, J. C. (1990). Survival of the kittiwake in relation to sex, year, breeding experience and position in the colony. *Journal of Animal Ecology*, 59(3):1063–1071.

Amstrup, S. C., McDonald, T. L., and Manly, B. F. J. (2005). *Handbook of Capture-Recapture Analysis*. Princetown University Press.

Arnason, A. N. (1973). The estimation of population size, migration rates and survival in a stratified population. *Researches on Population Ecology*, 15:1–8.

Barker, R., Fletcher, D., and Scofield, P. (2002). Measuring density dependence in survival from mark-recapture data. *Journal of Applied Statistics*, 29:305–313.

Barker, R. J. (1997). Joint modeling of Live-Recapture, Tag-Resight, and Tag-Recovery data. *Biometrics*, 53(2):666–677.

Barker, R. J. (1999). Joint analysis of mark-recapture, resighting and ring-recovery data with age-dependence and marking-effect. *Bird Study*, 46:82–91.

Barker, R. J., Burnham, K. P., and White, G. (2004). Encounter history modelling of joint mark-recapture, tag-resighting and tag-recovery data under temporary emigration. *Statistica Sinica*, 14:1037–1055.

Barker, R. J. and White, G. C. (2001). Joint analysis of live and dead encounters of marked animals. In *International Wildlife Management Congress*, pages 361–367.

Berglund, P. A. (2016). Evaluating ten years of ecological seabird research in the baltic Sea. B.S. Thesis, Department of Systems Ecology, Stockholm University, Stockholm.

Besbeas, P., Freeman, S., and Morgan, B. (2005). The potential of integrated population modelling†. *Australian & New Zealand Journal of Statistics*, 47(1):35–48.

Birkhead, T. R. and Hudson, P. J. (1977a). Population parameters Common Guillemot *Uria aalge*. *Ornis Scandinavica*, 8:145–154.

Birkhead, T. R. and Hudson, P. J. (1977b). Population parameters for the common murre *Uria aalge*. *Ornis Scandinavica*, pages 145–154.

Bonner, S. J., Byron, J. T., and King, R. (2010). Continuous covariates in mark-recapture-recovery analysis: A comparison of methods. *Biometrics*, 66:1256–1265.

Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance: Closed Populations*. Statistics for Biology and Health. Springer-Verlag London, 1 edition.

Borchers, D. L. and Efford, M. G. (2008). Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics*, 64:377–385.

Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2000). Bayesian animal survival estimation. *Statistical Science*, 56:357–376.

Burnham, K. (1993). A theory for combined analyis of ring recovery and recapture data. In Lebreton, J. D. and North, P. M., editors, *Marked Individuals in the Study of Bird Population*, pages 199–213. Birkhäuser Verlag Basel, Switzerland.

Cam, E., Link, W. A., Cooch, E. G., Monnat, J., and Danchin, E. (2002). Individual covariation in lifehistory traits: Seeing the trees despite the forest. *The American Naturalist*, 159(1):96–105.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Bentancourt, M., Brubaker, M., Guo, J., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Catchpole, E. A., Freeman, S. N., Morgan, B. J. T., and Harris, M. P. (1998). Integrated recovery/recapture data analysis. *Biometrics*, 54(1):33–46.

Catchpole, E. A., Kgosi, P. M., Morgan, B. J. T., and Harris, M. P. (2001). On the near-singularity of models for animal recovery data. *Biometrics*, 57:720–726.

Catchpole, E. A. and Morgan, B. J. T. (1996). Model selection in ring-recovery models using score tests. *Biometrics*, 52(2):664–672.

Cathpole, E. A., Morgan, B. J. T., and Tavecchia, J. (2008). A new method for analysing discrete life history data with missing covariate values. *Journal of the Royal Statistical Society*, 70(Part 2):445–460.

Cheung, Y. B., Gao, F., and Khoo, K. S. (2003). Age at diagnosis and the choice of survival analysis methods in cancer epidemiology. *Journal of Clinical Epidemiology*, 56(1):38–43.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2011). *Bayesian ideas and Data analysis: An introduction for scientists and statisticians*. CRC Press.

Colchero, F. and Clark, J. S. (2012). Bayesian inference on age-specific survival for censored and truncated data. *Journal of Animal Ecology*, 81:139–149.

Colchero, F., Jones, O. R., and Rebke, M. (2012). BaSTA: An R package for Bayesian estimation of age-specific survival from incomplete mark–recapture/recovery data with covariates. *Methods in Ecology and Evolution*, 3:466–470.

Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, 51:429–438.

Crespin, L., Choquet, R., Lima, M., Merritt, J., and Pradel, R. (2008). Is heterogeneity of catchability in capture-recapture studies a mere sampling artifact or a biologically relevant feature of the population? *Population Ecology*, 50:247–256.

Crespin, L., Harris, M. P., Lebreton, J. D., Frederiksen, M., and Wanless, S. (2006). Recruitment to a seabird population depends on environmental factors and on population size. *Journal of Animal Ecology*, 75(1):228–238.

Crowall, J., Butchart, S. H. M., Lascelles, B., Stattersfield, A. J.and Sullivan, B., and Symes, A. (2012). Seabird conservation status, threats and priority actions: A global assessment. *Bird Conservation International*, 22:1–34.

Croxall, J. P. and Rothery, P. (1991). Population regulation of seabirds: Implications of their demography for conservation. *Bird population studies: relevance to conservation and management. Oxford University Press, Oxford*, pages 272–296.

Del Moral, P. (1996). Nonlinear filtering: Interacting particle resolution. *Markov Processes and Related Fields*, 2(4):555–580.

Díaz, S., Settele, J., Brondízio, E. S., Ngo, H. T., Guèze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K. A., Butchart, S. H. M., Chan, K. M. A., Garibaldi, L. A., Ichii, K., Liu, J., Subramanian, S. M., Midgley, G. F., Miloslavich, P., Molnár, Z., Obura, D., Pfaff, A., Polasky, S., Purvis, A., Razzaque, J., Reyers, B., R. Roy Chowdhury, Y. J. S., Visseren-Hamakers, I. J., Willis, K. J., and Zayas, C. N. (2019). Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Technical report, IPBES secretariat, Bonn, Germany.

Efford, M. G. (2004). Density estimation in live-trapping studies. *Oikos*, 106:598–610.

Fiske, I. J. and Chandler, R. B. (2011). An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*, 43(10):1–20.

Frederiksen, M. and Bregnballe, T. (2000). Evidence for density-dependent survival in adult cormorants from a combined analysis of recoveries and resightings. *Journal of Animal Ecology*, 69(5):737–752.

Freeman, S. N. and Morgan, B. J. T. (1992). A modelling strategy for recovery data from birds ringed as nestlings. *Biometrics*, 48(1):217–.235.

Garrett, E. S. and Zeger, S. L. (2000). Latent class model diagnosis. *Biometrics*, 56(4):1055–1067.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85:398–409.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. Chapman & Hall/CRC Boca Raton, FL, USA.

Giménez, O. and Choquet, R. (2010). Individual heterogeneity in studies on marked animals using numerical integration: Capture-recapture mixed models. *Ecology*, 91(4):951–957.

Giménez, O., Morgan, B. J. T., and Brooks, S. P. (2009). Weak identifiability in models for mark-recapture-recovery data. In *Modeling demographic processes in marked populations*, pages 1055–1067. Springer.

Giménez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J. P., and Pradel, R. (2007). State-space modelling of data on marked individuals. *Ecological Modelling*, 206(3):431–438.

Golchi, S. (2016). Informative priors and Bayesian computation. In *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference*, pages 782–789.

Halley, D. J., Harris, M. P., and Wanless, S. (1995). Colony attendance patterns and recruitment in inmature Common murres (*Uria aalge*). *The Auk*, 112(4):647–957.

Harris, M., Halley, D., and Wanless, S. (1992). The postfledging survival of young guillemots *Uria aalge* in relation to hatching date and growth. *Ibis*, 134(4):335–339.

Harris, M. P., Frederiksen, M., and Wanless, S. (2007). Within-and between-year variation in the juvenile survival of Common Guillemots *Uria aalge*. *Ibis*, 149(3):472–481.

Harris, M. P., Halley, D. J., and Wanless, S. (1996a). Philopatry in the Common Murre. *Bird Study*, 43:134–137.

Harris, M. P., Wanless, S., and Barton, T. R. (1996b). Site use and fidelity in the Common Murre *Uria aalge*. *Ibis*, 138:399–404.

Harris, M. P., Wanless, S., Rothery, P., Swann, R. L., and Jardine, D. (2000). Survival of adult Common Guillemots *Uria aalge* at three Scottish colonies. *Bird Study*, 47(1):1–7.

Hastings, K. K., Small, R. J., and Pendleton, G. W. (2012). Sex- and age-specific survival of harbor seals (*Phoca vitulina*) from Tugidak island, Alaska. *Journal of Mammalogy*, 93(5):1368–1379.

Hastings, K. K., Testa, J. W., and Rexstad, E. A. (1999). Interannual variation in survival of juvenile weddell seals (*Leptonychotes weddellii*) from McMurdo Sound, Antarctica: Effects of cohort, sex and age. *Journal of Zoology*, 248(3):307–323.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Hatchwell, B. J. and Birkhead, T. R. (1991). Population dynamics of Common Murres *Uria aalge* on Skomer Island, Wales. *Ornis Scandinavica*, pages 55–59.

Hedgren, S. (1976). Om sillgrisslans uria aalgeföda vidstora karlsö. *Vår Fågelv*, 35:287–290.

Hedgren, S. and Linnman, A. (1979). Growth of Murre *Uria aalge* chicks in relation to time of hatching. *Ornis Scandinavica*, 10:29–36.

Hegg, D., Greaves, G., Maxwell, J. M., MacKenzie, D. I., and Jamieson, I. G. (2012). Demography of takahe (*Porphyrio hochstetteri*) in fiordland: Environmental factors and management affect survival and breeding success. *New Zealand Journal of Ecology*, 36(1):75–89.

Hentati-Sundberg, J. and Olsson, O. (2016). Amateur photographs reveal population history of a colonial seabird. *Current Biology*, 26:226–228.

Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer.

Hubbard, B. A., Cole, D. J., and Morgan, B. J. T. (2014). Parameter redundancy in capture-recapture-recovery models. *Statistical Methodology*, 17:17–29.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration–stochastic model. *Biometrika*, 52:225–247.

Kadin, M. (2007). Post-fledging survival and sexual maturation in Common guillemots *Uria aalge*–the influence of biological and ecological factors. Master's thesis, Department of Systems Ecology, Stockholm University, Stockholm.

Keenan, R., Reams, G. A., Frédéric, A., Freitas, J., Grainger, A., and Lindquist, E. (2015). Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *Forest Ecology Management*, (352):9–20.

Kéry, M., Dorazio, R. M., Soldaat, L., van Strien, A., Zuiderwijk, A., and Royle, J. A. (2009). Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, 46:1163–1172.

Kéry, M. and Schaub, M. (2011). *Bayesian Population Analysis using WinBUGS: A hierarchical perspective*. Academic Press.

King, R. (2012). A review of Bayesian state-space modelling of capture–recapture–recovery data. *Interface Focus*, 2:190–204.

King, R. (2014). Statistical Ecology. *The Annual Review of Statistics and Its Application*, 1:401–426.

King, R. and Brooks, S. P. (2003). Closed-form likelihoods for Arnason-Schwarz models. *Biometrika*, 90(2):435–444.

King, R., Brooks, S. P., Mazzetta, C., Freeman, S. N., and Morgan, B. J. T. (2008). Identifying and diagnosing population declines: A Bayesian assessment of lapwings in the UK. *Applied Statistics*, 57(5):609–632.

King, R. and Langrock, R. (2016). Semi-Markov Arnason-Schwarz models. *Biometrics*, 72(2):619–628.

King, R. and McCrea, R. (2014). A generalised likelihood framework for partially observed capture-recapture-recovery models. *Statistical Methodology*, 17:30–45.

King, R., Morgan, B. J. T., Giménez, O., and Brooks, S. (2010). *Bayesian Analysis for Population Ecology*. CRC Press.

Kom, E., Graubard, B., and Midthune, D. (1997). Time-to-Event Analysis of Longitudinal Follow-up of a Survey: Choice of the Time-scale. *American Journal of Epidemiology*, 145(1):72–80.

Kruschke, J. K. (2011). *Doing Bayesian data analysis: A tutorial with R and BUGS*. Academic Press.

Laake, J. (2013). RMark: An R interface for analysis of capture-recapture data with MARK. AFSC Processed Rep. 2013-01, Alaska Fisheries Science Centre, NOAA, National Marine Fisheries Service, Seattle, WA.

Laake, J. L., Johnson, D. S., and Conn, P. B. (2013). marked: An R package for maximum likelihood and Markov chain Monte Carlo analysis of capture-recapture data. *Methods in Ecology and Evolution*, 4(9):885–890.

Lamarca, R., Alonso, J., Gómez, G., and Muñoz, A. (1998). Left-truncated data with age as time scale: An alternative for survival analysis in the elderly population. *The Journals of Gerontology: Series A*, 53A(5):337–343.

Langrock, R. and King, R. (2013). Maximum likelihood estimation of Mark-Recapture-Recovery models in the presence of continuous covariates. *Annals of Applied Statistics*, 7:1709–1732.

Lebreton, J. D., Almeras, T., and Pradel, R. (1999). Competing events, mixtures of information and multistratum recapture models. *Bird Study*, 46:39–46.

Lebreton, J. D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological monographs*, 62(1):67–118.

Lebreton, J. D., Morgan, B. J. T., Pradel, R., and N., F. S. (1995). A simultaneous survival rate analysis of dead recovery and live recapture data. *Biometrics*, 51(4):1418–1428.

Lee, D., Abraham, C. L., Warzybok, P. M., Bradley, R. W., and Sydeman, W. J. (2008). Age-specific Survival, Breeding Success, and Recruitment in Common Murres (*Uria aalge*) of the California Current System. *The Auk*, 125(2):316–325.

Lewison, R., Oro, D., Godley, B. J., Underhill, L., Bearhop, S., Wilson, R. P., Ainley, D., Arcos, J. M., Boersma, P. D., Borboroglu, P. G., Boulinier, T., Frederiksen, M., Genovart, M., González-Solís, J., Green, J. A., Grémillet, D., Hamer, K. C., Hilton, G. M., Hyrenbach, K. D., Martínez-Abraín, A., Montevecchi, W. A., Phillips, R. A., Ryan, P. G., Sagar, P., Sydeman, W. J., Wanless, S., Watanuki, Y., Weimerskirch, H., and Yorio, P. (2012). Research priorities for seabirds: Improving conservation and management in the 21st century. *Endangered Species Research*, 78:93–121.

Liu, J. S. and Cheng, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93(443):1032–1044.

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS: A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and computing*, 10(4):325–337.

Lyngs, P. and Durinck, J. (1998). Diet of guillemots *Uria aalge* in the central Baltic Sea. *Dansk Orn. Foren. Tidsskr*, 35:197–200.

MacKenzie, D. I., Nicholes, J. D., Lachman, G. B., S., D., Royle, J. A., and Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255.

MacKenzie, D. I., Nichols, J. D., Hines, J. E., Knutson, M. G., and Franklin, A. B. (2003). Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84:2200–2207.

Martínez-Abraín, A., Conesa, D., and Forte, A. (2014). Subjectivism as an unavoidable feature of ecological statistics. *Animal Biodiversity and Conservation*, 37(2):141–143.

Martínez-Abraín, A., Velando, A., Oro, D., Genovart, M., Gerique, C., Bartolomé, M. A., Villuendas, E., and Sarzo, B. (2006). Sex-specific mortality of European shags after the Prestige oil spill: Demographic implications for the recovery of colonies. *Marine Ecology Progress Series*, 318:271–276.

McCarthy, M. A. and Masters, P. (2005). Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology*, 42:1012–1019.

McClintock, B. T. (2015). Multimark: An R package for analysis of capture–recapture data consisting of multiple “noninvasive” marks. *Ecology and Evolution*, 5(21):4920–4931.

McCrea, R. and Morgan, B. J. T. (2015). *Analysis of Capture-recapture Data*. CRC Press.

McCrea, R., Morgan, B. J. T., Giménez, O., and Besbeas, P. B. (2010). Multi-site Integrated Population Modelling. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(4):539–561.

McCrea, R., Morgan, B. J. T., Giménez, O., and Bregnballe, T. (2012). Model comparison and assessment for multi-state capture-recapture-recovery models. *Journal of Ornithology*, 152(Suppl.2):293–303.

Meade, J., Hatchwell, B. J., Blanchard, J. L., and Birkhead, T. R. (2013). The population increase of common guillemots *Uria aalge* on skomer island is explained by intrinsic demographic properties. *Journal of Avian Biology*, 44:055–061.

Merkel, F. and Barry, T. (2008). Seabird harvest in the Arctic. CAFF Technical Report 16, CAFF International Secretariat, Circumpolar Seabird Group (CBird), Seattle, WA.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Milner, J. M., Elston, D. A., and Albon, S. D. (1999). Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of soay sheep. *Journal of Animal Ecology*, 68(6):1235–1247.

Morrison, K. W., Hipfner, J. M., Gjerdrum, C., and Green, D. J. (2009). Wing length and mass at fledging predict local juvenile survival and age at first return in tufted puffins. *The Condor*, 111(3):433–441.

Nettleship, D. N. and Birkhead, T. R. (1985). The Atlantic Alcidae: the Evolution, Distribution and Biology of the Auks Inhabiting the Atlantic Ocean and Adjacent Water Areas. *Academic Press*.

Newman, K., Buckland, S., Morgan, B. J. T., King, R., Borchers, D., Cole, D., Besbeas, P., Giménez, O., and Thomas, L. (2014). *Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-Space Methods*. Springer.

Olsson, O. and Hentati-Sundberg, J. (2017). Population trends and status of four seabird species (*Uria aalge*, *Alca torda*, *Larus fuscus*, *Larus argentatus*) at Stora Karlsö in the Baltic Sea. *Ornys Svecica*, 27:64–93.

Olsson, O., Nilsson, T., and Fransson, T. (2000). Long-term study of mortality in the common guillemot in the Baltic Sea. Technical Report 5057, Swedish Environmental Protection Agency, Stockholm, Sweden.

Oro, D. and Pradel, R. (2000). Determinants of local recruitment in a growing colony of Audouin's gull. *Journal of Animal Ecology*, 69(1):119–132.

Owens, I. P. F. (2002). Sex differences in mortality rate. *Science*, 297(5589):2008–2009.

Parslow, J. L. F. (1966). Hole-nesting by Guillemots. *British Birds*, 59:120–121.

Pledger, S., Efford, M., Pollock, K., Collazo, J., and Lyons, J. (2009). Stopover duration analysis with departure probability dependent on unknown time since arrival. In *Modeling Demographic Processes In Marked Populations*, pages 349–363. Springer.

Pledger, S., Pollock, K. H., and Norris, J. L. (2003). Open capture-recapture models with heterogenity: I. Cormack-Jolly-Seber model. *Biometrics*, 59:786–794.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, pages 1–9.

Pradel, R. (2005). Multievent: An extension of Multistate capture-recapture models to uncertain states. *Biometrics*, 61:442–447.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramos, R., Granadeiro, J., Nevoux, M., Mougin, J. L., Dias, M., and Catry, P. (2012). Combined spatio-temporal impacts of climate and longline fisheries on the survival of a trans-equatorial marine migrant. *PLoS ONE*, 7(7):1–12.

Reynolds, T. J., King, R., Harwood, J., Frederiksen, M., Harris, M. P., and Wanless, S. (2009). Integrated data analysis in the presence of emigration and mark loss. *Journal of Agricultural, Biological, and Environmental Statistics*, 14:411–431.

Robert, C. and Casella, G. (2011). A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, 26(1):102–115.

Royle, J. A. (2008). Modeling individual effects in the Cormack–Jolly–Seber model: A state–space formulation. *Biometrics*, 64(2):364–370.

Royle, J. A., Chandler, R. B., Sollmann, R., and Gardner, B. (2014). *Spatial Capture-Recapture*. Elsevier.

Royle, J. A. and Dorazio, R. M. (2008). *Hierarchical modelling and inference in Ecology*. Elsevier.

Royle, J. A. and Young, K. V. (2008). A hierarchical model for spatial capture-recapture data. *Ecology*, 89:2281–2289.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(2):319–392.

Sandvik, H., Erikstad, K., Barrett, R., and Yoccoz, N. G. (2005). The effect of climate on adult survival of five species of North Atlantic seabirds. *Journal of Animal Ecology*, 74:817–831.

Sandvik, H., Erikstad, K. E., Fauchald, P., and Tveraa, T. (2008). High survival of immatures in a long-lived seabird: Insights from a long-term study of the Atlantic puffin (*Fratercula arctica*). *The Auk*, 125(3):723–730.

Sandvik, H., Erikstad, K. E., and Saether, B. E. (2012). Climate affects seabird population dynamics both via reproduction and adult survival. *Marine Ecology Progress Series*, 454:273–284.

Sanz-Aguilar, A., Igual, J. M., Oro, D., Genovart, M., and Tavecchia, G. (2016). Estimating recruitment and survival in partially monitored populations. *Journal of Applied Ecology*, 53(1):73–82.

Sarzo, B., Armero, C., Conesa, D., Hentati-Sundberg, J., and Olsson, O. (2019). Bayesian immature survival analysis of the largest colony of Common murre *Uria aalge* in the Baltic Sea. *Waterbirds*, 42(3):304–313.

Schaub, M. and Abadi, F. (2011). Integrated population models: A novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology*, 152(1):227–237.

Schaub, M., Wojciech, K., and Köppen, U. (2005). Variation of primary production during winter induces synchrony in survival rates in migratory white storks *Ciconia ciconia*. *Journal of Animal Ecology*, 74(4):656–666.

Schwarz, C. J., Schweigert, J. F., and Arnason, A. N. (1993). Estimating migration rates using tag-recovery data. *Biometrics*, 49:177–193.

Seber, G. A. F. (1965). A note on the multiple recapture census. *Biometrika*, 52:249–259.

Seber, G. A. F. and Schofield, M. R. (2019). *Capture-Recapture: Parameter Estimation for Open Animal Populations*. Springer.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: A sampling-resampling perspective. *The American Statistician*, 46:84–88.

Stroustrup, B. (2013). *The* `C++` *programming language (4th edition)*. Addison-Wesley.

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997). Inferring coalescence times from DNA sequence data. *Genetics*, 145(2):505–518.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1762.

Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., and Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: Estimating false-negative error rates. *Ecological applications*, 13:1790–1801.

Umlauf, N., Adler, D., Kneib, T., Lang, S., and Zeileis, A. (2015). Structured additive regression models: An R interface to BayesX. *Journal of Statistical Software*, 63(21):1–46.

Valpine, P. D., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2):403–413.

Votier, S. C., Birkhead, T. R., Oro, D., Trinder, M., Grantham, M. J., Clark, J. A., McCleery, R. H., and Hatchwell, B. J. (2008). Recruitment and survival of immature seabirds in relation to oil spills and climate variability. *Journal of Animal Ecology*, 77(5):974–983.

Wanless, S. and Harris, M. P. (1986). Time spent at the colony by male and female Guillemots *Uria aalge* and Razorbills *Alca torda*. *Bird Study*, 33:186–176.

Westreich, D., Cole, S. R., Tien, P. C., Chmiel, J. S., Kingsley, L., Funk, M. J., Anastos, K., and Jacobson, L. P. (2010). Time scale and adjusted survival curves for marginal structural Cox models. *American Journal of Epidemiology*, 171(6):691–700.

White, G. C. and Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study*, 46(sup1):120–139.

Worthington, H., King, R., and Buckland, S. T. (2015). Analysing Mark-recapture-recovery Data in the Presence of Missing Covariate data via Multiple Imputation. *Journal of Agricultural, Biological, and Environmental Statistics*, 20:28–46.

Österblom, H., Casini, M., Olsson, O., and Bignert, A. (2006). Fish, seabirds and trophic cascades in the Baltic Sea. *Ecology Progress Series*, 323:233–238.

Österblom, H., Fransson, T., and Olsson, O. (2002). Bycatches of common guillemot (*Uria aalge*) in the Baltic Sea gillnet fishery. *Biological Conservation*, 105:309–319.

Österblom, H., Van der Jeugd, H. P., and Olsson, O. (2004). Adult survival and avian cholera in Common Guillemots *Uria aalge* in the Baltic Sea. *Ibis*, 146:1–4.

Žydelis, R., Bellebaum, J., Österblom, H., Vetemaa, M., Schirmeister, B., Stipniece, R. A., Dagys, M., Eerden, M., and Garthe, S. (2009). Bycatch in gillnet fisheries – An overlooked threat to waterbird populations. *Biological Conservation*, 142:1269–1281.