WITH ACCOMPANYING
CD-ROM AT BACK OF
BOOK

Spatial structure and population dynamics of the British grey seal (*Halichoerus grypus*): inferences from genetic, photo-identification and abundance data

Philip John Harrison

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

University of St Andrews

January 2006

Spatial structure and population dynamics of the British grey seal (*Halichoerus grypus*): inferences from genetic, photo-identification and abundance data

Philip John Harrison

Abstract

Building models for wildlife populations is a vital part in the advancement of scientific understanding of the natural world we inhabit. However, most wildlife populations are not randomly distributed in space; models must account for spatial structure to avoid erroneous judgements and predictions. Furthermore, the dynamics of wildlife populations are generally determined by complex, non-linear processes that are strongly affected by stochasticity; our observations on these populations are seldom made without error. The state-space approach to wildlife modelling, which can incorporate all these facets, was utilised in this thesis.

This work examines the spatial population dynamics of the British grey seal (*Halichoerus grypus*) using genetic, photo-identification and abundance data. Microsatellite DNA data examined for this species revealed genetic differentiation between colonies from different regions around the British Isles, but approximate panmixia among the four main breeding colonies in the North Sea region. The remainder of this thesis was devoted to building spatially-explicit models for grey seals breeding at the four colonies in this region.

One of the most difficult problems in developing spatially-explicit models of population dynamics is the validation and parameterisation of the movement process. I demonstrate how movement models derived from a multisite capture-recapture analysis of photo-identification data can be improved by incorporating them into a set of spatially-explicit state-space metapopulation models, which are then fitted to a time series of abundance data.

The best fitting state-space model had only a single movement parameter. The variance associated with this parameter was greatly reduced compared to that obtained from the capture-recapture study. There was some support for a model that included the effect of distance between colonies on movement rates. Predictions of future colony sizes made using these models demonstrated that the incorporation of movement and the way in which it was modelled affected both local and regional dynamics.

Declarations

I, Philip John Harrison, hereby declare that this thesis, which is approximately 36,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

date 9th Jan 2006 signature of candidate

I was admitted as a research student in October 2002 and as a candidate for the degree of PhD in October 2003; the higher study for which this is a record was carried out in the University of St Andrews between 2002 and 2005.

date 9th Jan 2006 signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

date 9.1.06 signature of supervisor

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker.

date 9th Jan 2006 signature of candidate

Acknowledgements

First and foremost I owe a large measure of gratitude to my supervisors John Harwood and Steve Buckland for their endless help and encouragement. I would like to thank my advisor Bill Amos for his insights. I am also extremely grateful to Len Thomas and Ken Newman for imparting to me a great deal of knowledge concerning the finer details of state-space models and computer programming.

A special thanks to the Natural Environment Research Council (NERC) and the Engineering and Physical Sciences Research Council (EPSRC) for funding this project.

Thanks to all those at the Sea Mammal Research Unit involved in the collection and processing of the photo-ID and pup production data, especially Rob Harris for the former and Callan Duck for the latter. Thanks also to the staff of the National Trust and the Lincolnshire Wildlife Trust for conducting ground counts at the Farne Islands and Donna Nook respectively. Thanks to all those from St Andrews and Cambridge Universities involved in the collection and processing of the genetic data, especially Joe Hoffman for compiling the database I needed.

A very warm thank you to my fiancée Aino Inkinen for keeping me grounded and sane during the three years of this project and for her grey seal drawing on the second cover page of this thesis. Thank you to my parents too, for putting up with my one-track mindedness during the final writing up stage.

Last, but not least, I would like to thank the following people for their help in various ways: Christian Asseburg, Rémi Choquet, David Elston, Olivier Gimenez, Phil Hammond, Jean-Dominique Lebreton, Chris Lynam, Phil Le Feuvre, Mike Lonergan, Bernie McConnell, Tom Meagher, Paddy Pomeroy, Kevin Rideout, Rhona Rodger, Sophie Smout, Catriona Stephenson and John Watkins.

Contents

List of Tables

List of Figures

Chapter 1

General Introduction

The major cause of the worldwide decline in biodiversity is probably the anthropogenic destruction and fragmentation of natural landscapes (Gaggiotti *et al.* 2004; Hanski 2005). It is therefore vital for conservation of species that we understand their spatial dynamics. However, as it is not possible to collect data on all the worlds' species, we must select some for intensive study and use them as model systems with which to acquire a baseline of information (Ehrlich & Hanski 2004). The grey seal is one species that has been studied extensively and for which large datasets have been compiled. In this thesis I have used genetic, photo-identification (photo-ID) and abundance data to better understand the spatial structure of this species.

1.1 Pinnipeds and the grey seal

Pinnipeds are a group of marine mammals that descended from a common terrestrial carnivore ancestor (probably bear or otter-like) some 25 million years ago (Boness & Bowen 1996). They include 33 extant species of seals, sea-lions and walruses, making them the most numerous group of marine mammals within the order Carnivora. The most diverse pinniped family is the Phocidae, or true seals, with 18 extant species (Brownell, Ralls & Perrin 1995). It is to this family that the focal species of this thesis

– the grey seal, *Halichoerus grypus* (meaning "hooked-nose pig of the sea") (Fabricius 1791) – belongs.

The grey seal is conveniently divided into three distinct populations: the north-west Atlantic (the largest population); the north-east Atlantic; and a relatively small population in the Baltic. No mixing is believed to occur between these populations (Allen *et al.* 1995). Grey seals, like other pinnipeds, spend most of their time at sea foraging for food. During this time they disperse over a wide area, potentially travelling thousands of kilometres (McConnell *et al.* 1992; Hammond *et al.* 1993), but they require solid substrate (land or ice) on which to give birth. Females have spatially and temporally synchronized reproduction (Boyd 2002): an annual cycle which offers a predictable opportunity for males and females to interact (Brown, Beck & Austin 2002).

The grey seal has a world population of around 300,000 individuals, 40% of which breed around the British Isles (SCOS 2004). The total number of pups born at British colonies has grown steadily since the 1960s, when records began (Harwood & Prime 1978). The breeding season for British grey seals is during September-December when each breeding female gives birth to a single white coated pup. This white coat is known as the lanugo (Figure 1.1) and is shed after the pup has been suckled for approximately 18 days (Hall 2002). The lanugo makes pups very visible from the air and the Sea Mammal Research Unit (SMRU), based at St Andrews University, conducts annual aerial surveys of the major breeding colonies in Britain to determine the number of pups produced there (Duck 2004). This pup production data can be used in population dynamics modelling for this species (as in chapter 5 of this thesis). Furthermore, pups remain on the breeding colonies while they are nursed by their mothers and for a further 10 to 28 days post-weaning (Hall 2002), so that the collection of small tissue samples from them is relatively straightforward. DNA can then be extracted from this tissue and used in population genetics analyses for this species (as in chapter 3 of this thesis).

Figure 1.1 Grey seal pup showing partial moulting of its white coat, known as the lanugo, around the head and flippers (photograph taken by SWT/NTS Ranger, Kevin Rideout)

Females aggregate during the breeding season on small islands and coastal beaches. This favours the evolution of male mating strategies that involve competition for access to large numbers of oestrous females (Boness & Bowen 1996), potentially leading to high levels of polygyny. This provides an opportunity for sexual selection to act on male body size, because large males are likely to be better able to defend access to large numbers of females or to stay ashore for longer periods. Indeed, mature male grey seals (weighing 170-310kg) are much larger than mature females (weighing 100-190kg) (Hall 2002). Another difference between the sexes is the colour and markings on their pelage. Males have a relatively uniform dark pelage, whereas females have a light background pelage colouration with dark markings (Figure 1.2). The markings of females are often very distinctive and allow them to be individually recognised and "marked" for capture-recapture (CR) analysis using photo-ID (see chapter 4 of this thesis).

Figure 1.2 An adult female grey seal showing the typical light pelage colouration with individually identifiable dark markings (photograph taken by Philip Harrison).

1.2 Metapopulation dynamics

The colonial aggregations of female grey seals during the breeding season can be used in spatially-explicit modelling for this species. Ideally, for such modelling, we must take account both of processes that occur within aggregation (e.g. birth and death) and between them (e.g. movement and colonization) (Thrall, Burdon & Murray 2000). The metapopulation approach (Hanski & Simberloff 1997; Elmhagen & Angerbjörn 2001; Hanski & Gaggiotti 2004) provides a conceptual tool which allows both sets of processes to be modelled in a single framework. The term "metapopulation" was coined by Levins in 1969 and his classical metapopulation concept applies to a spatial structure in which: (1) individuals spend the majority of their time within their own aggregations; (2) individuals only occasionally move between aggregations; (3) the dynamics of aggregations are asynchronous: and (4) aggregations are prone to extinction, but the space they occupy can be recolonized (i.e. there is population turnover). In a metapopulation, the dynamics of any one aggregation cannot be completely understood without reference to the entire ensemble of aggregations (Stacey, Johnson & Taper 1997).

Real world metapopulations are unlikely to conform strictly to these four conditions, let alone Levins' original assumptions that all aggregations are equal in size and degree of isolation. However, the only critical requirement for a metapopulation is that aggregations are discrete but connected by migration. If this is not the case, then a metapopulation framework is inappropriate (Hanski & Simberloff 1997).

Grey seal colonies meet all the assumptions for a metapopulation: (1) studies of seals "marked" by hot-iron branding have shown high site fidelity in both sexes at one colony (Pomeroy *et al.* 1994, Twiss *et al.* 1994); (2) studies of adult female seals "marked" by photo-ID have shown that females occasionally move between colonies in the North Sea (see chapter 4); (3) the UK grey seal colonies show diverse dynamics (Gaggiotti *et al.* 2002); and (4) in the last 40 years, three of the 21 breeding colonies in the Orkney Isles (off northern Scotland) have gone extinct and two of these have been recolonized.

One of the key parameters in a metapopulation model is the probability of movement between aggregations (in this case, between colonies). In insect and small mammal populations this parameter can sometimes be estimated using direct observations of the movement of marked individuals (Elmhagen & Angerbjörn 2001). However, for species such as the grey seal, which spends more than 80% of its time at sea (McConnell *et al.* 1999) and 90% of this time below the surface (Thompson *et al.* 1991), direct observation of movement is impracticable. Indeed, Parker *et al.* (1998) state that estimating dispersal is "universally problematic" and according to Waser & Storbeck (1998) it remains "one of the most enigmatic parameters in population and conservation biology".

Various types of data can be used to make inferences about (meta)population structure and the nature and amount of movement between population units. I analysed three: (1) genetic data, where microsatellite markers were used to explore population differentiation between grey seal colonies (see section 1.3 and chapter 3); (2) photo-ID data, where multisite CR analysis was used to estimate movement probabilities (see section 1.4 and chapter 4); and (3) pup production estimates, which where used to parameterise a spatially-explicit model of the grey seal metapopulation in the North Sea (see section 1.5, 1.6 and chapter 5).

1.3 Microsatellite genetic markers

Genetics is essentially the study of heredity and inherited attributes. Most genetic investigations of natural populations – from both an evolutionary and ecological perspective – are concerned with relative differences in consanguinity (i.e. kinship or relatedness). However, the degree of relatedness we are interested in depends on the specifications of the study. For example, a study of population subdivision will be looking at a finer resolution in consanguinity than one attempting to construct a phylogenetic tree (Palsbøll 2002).

Genetic analyses have been widely used in the assessment of threatened species (Daugherty *et al.* 1990; May 1990; Moritz 1994). Traditionally, the emphasis has been on determining levels of inbreeding from surveys of gene and allelic diversity (Sherwin & Moritz 2000). Genetic information is less influenced by environmental factors than morphological characteristics and therefore provides a more representative measure of the degree of relatedness between individuals. In its most basic form this information comes from the nucleotide sequence of the genome itself. With the development of the polymerase chain reaction (PCR[1]) in 1987 it became possible to analyse DNA sequences directly. The more differences there are in the same sequence from different individuals (i.e. the more mutations that have occurred at the same locus) the less related they are. A mutation is an alteration in the nucleotide sequence of a DNA molecule (Miglani 2000): the most common forms of mutation are nucleotide substitutions, insertions and deletions (Palsbøll 2002).

Genetic diversity at the population level is measured using codominant molecular markers (such as allozymes, restriction fragment length polymorphisms (RFLPs) and microsatellites). RFLPs and microsatellites are neutral to the effects of selection and do not affect the individual's phenotype. These genetic analyses are principally

[1] PCR – a cyclic, in vitro enzymatic reaction by which a small fragment of DNA can be replicated exponentially (Baker & Lento 2002)

concerned with relative differences in allele frequencies[2]. Using these markers it is possible to analyse the mating system and social structure of populations, to infer parentage and to assess the degree of population differentiation and thus estimate the levels of genetic drift[3] and gene flow[4].

Microsatellites are tandemly repeated units of short nucleotide motifs 1-6 base-pairs (bp) long that are extensively distributed throughout the nuclear genomes of animals and plants. The most prevalent are di- tri- and tetranucleotide repeats (e.g. $(CA)_n$, $(AAT)_n$, and $(GATA)_n$ respectively), (Lindenmayer & Peakall 2000). Microsatellites are inherited in a Mendelian fashion – at each locus individuals inherit two alleles, one from each parent. Because they are highly variable, have a high mutation rate, are distributed in large numbers throughout the genome, and have codominant inheritance, microsatellite loci are now though of as the most appropriate nuclear markers for many population genetics applications (Gutiérrez-Espeleta *et al.* 2000).

1.3.1 Mating systems

Behavioural observations of grey seal mating systems have tended to imply that there are high levels of polygyny, with the majority of pups being fathered by a small number of dominant males. Microsatellite data, however, have shown that the mating system of the grey seal is rather more complex than was originally believed. For example, Amos *et al.* (1995) showed that both polygyny and mate fidelity occur within this species. It also appears that only a small handful of males enjoy elevated mating success at the heart of a group of aggregating females (Worthington Wilmer *et al.* 1999). The majority of pups are fathered by any of a large number of adult males, who all have low but approximately identical success rates (Worthington Wilmer *et al.* 2000). A sizeable proportion of these males, beyond the reaches of present

[2] Allele frequency is defined as the proportion of all alleles at a locus that are of the specified type, among a group of individuals (Hartl 2000).
[3] Genetic drift - a random sampling effect primarily effecting smaller subpopulations whereby alleles are fixed in the different subpopulations by chance (Campbell 1996)
[4] Gene flow – a consequence of migration, higher levels of which counteracts the effects of genetic drift and act to homogenise subpopulations (Campbell 1996)

sampling techniques, are thought to employ aquatic mating and may not even come ashore at the breeding colony (Worthington Wilmer *et al.* 1999). A further puzzling occurrence is that, on average, females "choose" partners in successive years that are less genetically related to the previous partner than would be expected from random mating (Amos *et al.* 2001b). In chapter 2 of this thesis I explore the effect of this mate choice mechanism on the genetic structure of a hypothetical grey seal colony using simulation models.

1.3.2 Population differentiation

Several genetic studies of marine mammals have been undertaken with the aim of discovering the degree of genetic heterogeneity among sub-populations (Palsbøll 2002). Microsatellite-based analyses of population structure of the grey seal have demonstrated significant differentiation between two colonies in Scotland (North Rona off the north-west coast and the Isle of May off the east coast), indicating restricted gene flow between these two colonies (Allen *et al.* 1995). More recently, Gaggiotti *et al.* (2002, 2004) studied the colonization process in a grey seal metapopulation in the Orkney Isles and demonstrated that new colonies are formed by seals moving from the nearest and largest of the surrounding established colonies.

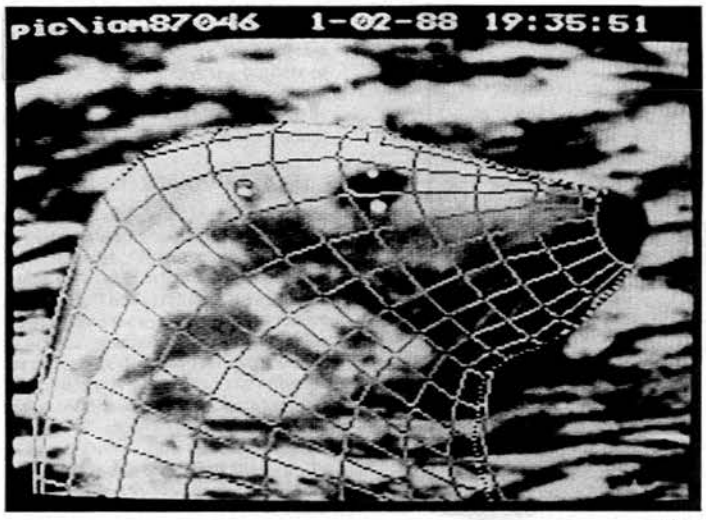
In chapter 3 of this thesis I explore the population structure of grey seal colonies in more detail using nine microsatellite markers. I compared the differentiation between the Canadian and UK populations, among regions in the UK population, and among colonies in one region: the North Sea. Four types of analyses were used on the microsatellite data: (1) exact tests, which use randomization procedures to test for differentiation between pairs of colonies (Goudet *et al.* 1996) ; (2) F_{ST} (Weir & Cockerham 1984), which measures the proportion of variance in allele frequencies that can be attributed to differences between pairs of colonies (Balloux & Lugon-Moulin 2002); (3) assignment-based tests (Cornuet *et al.* 1999), which assign each individual's genotype to the colony where its likelihood of occurrence is highest; and (4) Bayesian model-based clustering (Pritchard *et al.* 2000; Falush *et al.* 2003), which

calculates posterior probabilities for the number of distinct sub-populations represented in the data set analysed.

1.4 Capture-recapture analysis using photo-ID

Genetic data can help to establish the extent to which populations are subdivided but cannot easily be used to estimate the present levels of movement between population units. Multisite CR data, on the other hand, can be used to estimate such probabilities. Traditionally grey seals have been "marked" with tags or even hot-iron brands. A less intrusive method is to use the animals' own natural markings. For the UK grey seal a large photo-ID dataset has been compiled for identifiable females based on the unique marking patterns on their head and neck. Such an approach is not as feasible for males, however, as they are much less clearly marked.

There are often logistic and financial difficulties that prevent the systematic collection of samples for multisite CR analysis from all colonies that animals may frequent in a sequence of years (Spendelow *et al.* 1995). However, photographs of female grey seals breeding at the four main colonies in the North Sea (Isle of May, Fast Castle, Farne Islands and Donna Nook) were systematically collected between 1999 and 2001. Photographs of the same seal were matched using software developed by Hiby & Lovell (1990). The software fits a surface model to the photograph of the side of the head, which compensates for the viewpoint of the camera and the posture of the seal (Figure 1.3a). A pattern cell is then located in relation to morphological features on the seal's head and an identifier array is produced (Figure 1.3b). This identifier array is converted into a numerical description of the grey-scale intensities that can be compared to all others in the computer library. Those that score above a pre-set similarity threshold are presented for final comparison by eye.



(a)

(b)

Fig 1.3 (a) Surface model fitted to a digitised grey seal photograph.(b) The pattern cell and, to the right of the main photograph, the identifier array used by the image processing software of Hiby & Lovell (1990).

The output from this matching software was converted into a capture history for each animal. For example, a capture history of "102" would indicate that this female was photographed on colony 1 in 1999, not seen at any of the four North Sea colonies in 2000, and re-photographed on colony 2 in 2001. The data for all the females stored in this form were analysed using specifically designed software for multisite CR analysis known as *M-SURGE* (Choquet *et al.* 2004). Recent advances in approaches for statistically analysing multisite data, such as those utilised by *M-SURGE*, have greatly widened the scope for studying movement patterns in natural populations (Cam *et al.* 2004). Different models to describe these data were compared and the model(s)

giving the most parsimonious[5] description of the data were identified using standard model selection procedures (Johnson & Omland 2004) (see chapter 4).

1.5 Pup production estimates

Another potential source of information about movement within a grey seal metapopulation is the annual estimates of the number of pups produced at each breeding colony. However, these data are only useful in this context if there is some prior knowledge of the population dynamics of the species. For example, if the annual pup production increases faster than is possible as a result of internal recruitment of adult females into the breeding population, one may infer that net immigration must be taking place. This idea can be formalized by fitting the pup production data to different candidate models for the dynamics of the metapopulation. Model selection criteria can be used to evaluate the different models in much the same way as for the CR data. An ideal conceptual framework for this analysis is the state-space model (SSM) described in section 1.6.

Most of the major UK grey seal colonies have been surveyed annually from the air by SMRU each year since 1962, although some (such as the Farne Islands and Donna Nook) are surveyed by observers on the ground. However, aerial survey methods were changed somewhat after a fatal plane crash in 1983. These survey results are used to provide an annual estimate of pup production for each colony (Hiby & Duck, unpublished). In chapter 5 I use pup production estimates from 1984 to 2003 for the four North Sea colonies (Duck 2004) to estimate the parameters of a SSM of this region.

[5] Principle of parsimony – to make a trade off between bias and precision given the information content in the data. In general, more complicated models will fit the data better (because they are more flexible); they are penalised for the extra parameters that need to be estimated.

1.6 State-space models: a unifying framework

As a general rule, the dynamics of wildlife populations are determined by complex, non-linear processes that are strongly affected by stochasticity (Harwood & Stokes 2003); observations on these populations are seldom made without error. SSMs provide a framework that links measurements taken on a wildlife population to a stochastic population dynamics model (PDM). A SSM incorporates two parallel and simultaneous processes: the state process, which models the true but unknown state of the population; and the observation process, which links the state process to a time series of observational data. The advantage of this approach is that all major sources of uncertainty can be easily incorporated (Buckland *et al.* 2004). The explicit recognition of the various sources of uncertainty inherent in the population's dynamics, our observations of it, and our model formulations about it, provide a reliable and honest account of the present state of our understanding about the animal population in question.

The state process model is a modification of the familiar Leslie matrix. Individual processes (e.g. survival, birth, migration, etc.) are represented by separate sub-process matrices, either as expectations (with stochastic errors attached) or probability density functions (pdfs). These sub-process matrices are then chained together to form the complete PDM. This allows each of the sub-processes in a complex model with numerous interactions to be specified separately. It provides a convenient and accessible connection between mathematics, statistics and biology, because the model building process is done in a number of separate steps, each of which is easily manageable and models an intuitive aspect of the biology of the population in question (Buckland *et al.* 2004; Thomas *et al.* 2005). The SSM structure can also be extended to incorporate a wide range of biologically important interactions, such as competition, density dependence, predator-prey interactions and metapopulation dynamics. In chapter 5 of this thesis I develop SSMs for the female grey seals breeding at the four main colonies in the North Sea. These models imply that males do not influence the population's dynamics (Fujiwara & Caswell 2002). This assumption is valid for the grey seal system as the size of the male population is likely to be sufficient large at all times to ensure that female reproduction is not restricted by

a lack of mates. Figure 1.4 shows a schematic for these models indicating how the state-processes are broken down and the place of the observation process in this structure.

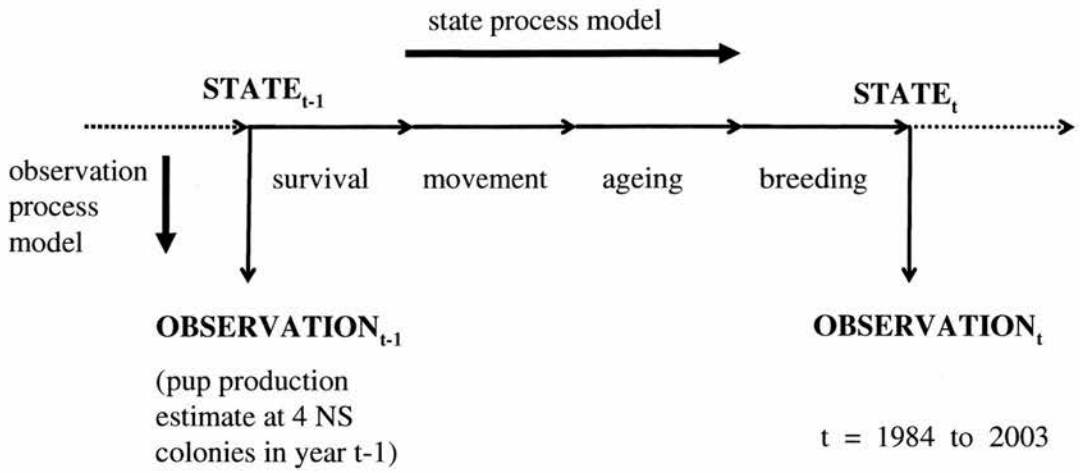


Figure 1.4 Schematic of the state-space model used in this thesis for the four North Sea (NS) grey seal colonies.

Fitting models to data necessitates the use of statistical methods. Which method to use is a controversial issue; on one side of the coin is the frequentist approach, which although imperfect, has a proven track record, and on the other side is the Bayesian approach. The frequentist statistician assumes that there exists a true underlying process with fixed parameter values. These can be estimated from the data, which are seen as random observations from this process. The Bayesian statistician, however, sees the data as fixed and the parameter values as being the random variables (Ellison 1996). The Bayesian approach: (a) provides a means by which uncertainty can be incorporated into the modelling framework; (b) can explicitly incorporate prior information from previous experiments – in the words of Hilborn & Mangel (1997): "If we fail to use what we have learnt in previous studies, we will learn very slowly indeed"; and (c) can make use of complicated models that would be intractable from a frequentist perspective (Beaumont & Rannala 2004).

Inference about which SSM provides the best fit to the available data is usually obtained using computer-intensive Bayesian methods (Millar & Meyer 2000; Doucet, Frietas & Gordon 2001; Liu 2001). Prior distributions for the model parameters and

initial states are specified using information from previous experiments and expert knowledge. A further advantage of the SSM approach for the research described in this thesis is that the results from the CR analysis (from chapter 4) could be used to provide informative priors for the population dynamics modelling, thus combining the strengths of both the CR and pup production data sets. Bayesian inference proceeds by interpreting these prior distributions with respect to the pup production data, and generating posterior distributions for the parameters and states of the model (Luikart & England 1999). These posterior distributions can then be used to project the models forward in time and make predictions about future animal abundances.

1.7 Thesis aims

Building models for wildlife populations is a vital part in the advancement of scientific understanding of the natural world we inhabit. Models help to clarify assumptions, combine knowledge from different sources, and oblige us to be unambiguous and precise in our logic (Burgman & Possingham 2000). Through model selection we can identify the important processes governing animal population dynamics. By projecting the models forward in time we can investigate the consequences of what we deem to be true. In this thesis I have used statistically defensible state-of-the-art modelling approaches to add to the scientific understanding of the spatial structure and population dynamics of the grey seal. It is also hoped that the work herein might advance the methods available to statistical ecologists studying wildlife populations in highly fragmented landscapes where spatial considerations are of utmost importance.

I start in chapter 2 by using individual-based simulation models to explore the effects of females "choosing" a mate in year t that is genetically dissimilar to their mate from year t-1. The effects on allele frequencies are monitored through time. These simulations are purely hypothetical; the models are not fitted to any data, but they are defined using the notation of SSMs, so this chapter also provides an introduction to the state-space framework used later in this thesis. In chapter 3 I explore the degree of population differentiation between grey seal colonies using microsatellite DNA data.

Samples from colonies around the UK are compared to one another and to a sample from Sable Island in Canada. I address the following questions. Are seals born in a colony on the east coast of the UK more differentiated from Canadian pups than those born on the west coast? Are colonies which are closer together around the UK less differentiated from one another than those separated by larger distances? In chapter 4 I investigate the photo-ID data for female grey seals breeding at colonies in the North Sea. The key questions in this section are as follows. What amount of movement is there between these colonies? Is the probability of moving between colonies related to the distance separating them? Is this probability related to the difference in abundance between the colonies? In chapter 5 I develop a set of SSMs for the females breeding in the North Sea using the results of chapter 4 to provide informative priors on the movement model parameters. These metapopulation models were fitted to a 20 year time series of pup production data using a Bayesian approach. I address the following questions. What movement model structure provides the best description of these abundance data? Is there information content in these data with which to refine the movement probability estimates obtained from the CR study? In chapter 6 I draw the thesis to an end with a general discussion of the findings from the previous chapters. In Appendix 1 I describe a simulation study, related to chapter 4, which investigated different photo-ID sampling strategies to determine the optimal sampling strategy to collect a large number of photographs of different animals within a single breeding season. In Appendix 2 I propose a suggestion for how to extend the SSMs developed in chapter 5 to simultaneously fit to both the pup production and photo-ID data. Finally, the CD that accompanies this thesis provides example source code and data, written using the statistical computing language R Version 2.0.1 (R Development Core Team, 2004), to implement the state-space modelling methods described in chapter 5.

Chapter 2

A Simulation-Based Exploration into the Mating System of the British Grey Seal

2.1 Introduction

Grey seals show sexual dimorphism in body size, with males being much larger than females (Hall 2002). Larger males are thought better able to compete for territory among aggregations of females during the breeding season, thus maximizing the number of copulations they achieve (Boness & Bowen 1996). Furthermore, as females invest much more in their pups than males, theory suggests that females should select males who have high genetic quality, as possibly indicated by body size (Amos *et al.* 2001b). The outcome of these factors is a classical polygynous mating system, with a relatively small number of dominant males fathering the majority of pups. Numerous behavioural studies have sustained this theory (Worthington Wilmer *et al.* 1999).

Recently, however, the degree of polygyny maintained on grey seal colonies around the UK has come under scrutiny based on the results of genetic studies. DNA fingerprints and single locus minisatellite analyses of grey seals on North Rona (Outer Hebrides, Scotland) have shown the presence of many pups sharing the same mother and father (full siblings), yet the majority of these fathers are not the dominant males

on the colony (Amos *et al.* 1995). This suggests that mate fidelity is in operation on this colony with animals coordinating their behaviours both within and between seasons.

An extensive microsatellite-based study of grey seals on both North Rona and the Isle of May (North Sea, Scotland) (Worthington Wilmer *et al.* 1999, 2000) which sampled a much more varied cross-section of males than had previously been done, found that behavioural dominance to increase fitness is a quality of only a handful of males at the core of a colony of aggregating females. Most pups were fathered by any of a large number of males who all shared a small but approximately equal success rate. Furthermore, between 50 and 70% of pups could not be allocated a father from those sampled, leading to the hypothesis that aquatic mating is important for this pinniped species.

Perhaps the most fascinating result to come from these genetic studies was the finding by Amos *et al.* (2001b) that, on average, pups sharing the same mother are less paternally related than would be expected by chance. This implies some mechanism of "choice" – where a female chooses partners in successive years that are genetically unrelated. (Possible mechanisms controlling this mate choice are discussed in section 2.4.1). It was also found that the signal in the data for this mate choice was stronger than could be explained by pups with the same mother being fathered by males from opposite ends of the Scottish metapopulation (North Rona and the Isle of May). Furthermore, this choice could not be explained by females selecting males based on genetic dissimilarity to themselves.

In this chapter I am primarily concerned with how this mate choice affects the genetic structure and dynamics of a hypothetical grey seal colony, which I have explored using computer simulation models. The simulations are initiated assuming that an island has been colonized by a group of breeding age male and female seals drawn at random from a hypothetical allele distribution. Following this colony founding event, for the sake of simplicity, I have assumed that there is no further immigration to or emigration from the island. From the simulations I wish to ascertain how the allele frequencies in the population change through time with different mating models. For these simulations I specifically explore how the "true" pup population allele

frequency distribution changes through time and how this is affected by changes in the mating model. Two mating models were investigated:

1. random mating – null model;
2. females "choose" a mate in year t genetically dissimilar to their mate from year $t-1$;

The simulation models are not fit to any data, but I have defined the model in the stochastic matrix formulation currently being utilized for state-space models (SSMs) (Buckland *et al.* 2004, Thomas *et al.* 2005, Newman *et al.* in press). SSMs were introduced in chapter 1 of this thesis and are also developed and fit to pup production data for the grey seal in chapter 5. However, this matrix formulation does not fully specify the model, and additional information is given when required. The simulations I have performed are akin to a sensitivity analysis: I wish to investigate how sensitive the pup allele frequencies are to non-random matings. The mate choice simulation necessitated the use of individual based models (IBMs).

2.2 Methods

2.2.1 Model formulation

Each seal is represented as a matrix that specifies its genotype. For instance, if we were interested in 3 microsatellite loci, then each individual seal would be represented by a 3 by 2 matrix (due to the fact that seals are diploid and have 2 alleles at each locus). These matrices are collected together according to sex and age and stored as 3D arrays, where the third dimension represents the number of seals present of a certain sex and a specific age. Despite the fact that the model is an individual-based one, the underlying process model is an age-based one. At each process step (e.g. survival, births, etc) probabilistic calculations are done on numbers of seals in each age and sex class (i.e. if the probabilistic survival function predicts x age 2 females will die between year $t-1$ and year t, then x of the age 2 female matrices are removed at random, from the array of age 2 females. The age 2 females remaining advance to

become age 3 females, and so forth). The formulation of the age-based processes can be expressed as an adaptation of the familiar Leslie matrix notation presently being used for state-space models of animal population dynamics (see section 2.3).

So let us consider the grey seals at the hypothetical colony from time $t-1$ to time t. The time step for the model is one year beginning just after the pups have been born in the autumn. The age-based state vector for the entire population is of length sixteen for the colony, representing female pups ($n_{0,f}$), age one females ($n_{1,f}$), through to age five females ($n_{5,f}$) and then age six and older females grouped into a single category ($n_{6+,f}$). In general $n_{a,f,t}$ represents the number of females of age a in the colony at time t. The following nine entries in the state vector represent the males in the population, however, for the males we go up to an 8+ age category, as males are thought to become sexually mature later than females. Thus, $n_{a,m,t}$ represents the number of males of age a in the colony at time t. Hence:

$$\mathbf{n}_t = \begin{bmatrix} n_{0,f,t} \\ n_{1,f,t} \\ \mathrm{M} \\ n_{5,f,t} \\ n_{6+,f,t} \\ n_{0,m,t} \\ n_{1,m,t} \\ \mathrm{M} \\ n_{7,m,t} \\ n_{8+,m,t} \end{bmatrix}$$

This state vector represents the state of the population at the end of the year after the state sub-processes, described below, have occurred.

2.2.2 State process: stochastic representation

The progression from $\mathbf{n}_{t-1}$ to $\mathbf{n}_t$ is characterised by a series of linked probability density functions (pdfs), which represent each stochastic sub-process, the input to each pdf being the output from the previous one. For the grey seal model we begin with survival (sub-process 1), followed by age advancement (sub-process 2), and finally births (which includes sexing), (sub-process 3).The distributions for these sub-processes at time t being expressed as:

$$\mathbf{u}_{1,t} \sim H_{1,t}(\mathbf{n}_{t-1})$$

$$\mathbf{u}_{2,t} \sim H_{2,t}(\mathbf{u}_{1,t})$$

$$\mathbf{n}_t \sim H_{3,t}(\mathbf{u}_{2,t})$$

where $H_{x,t}(\cdot)$ is the pdf for sub-process x and $\mathbf{u}_{x,t}$ is a realization of the state vector at time t after sub-process x has occurred. In more detail these sub-process pdfs are as follows. For the females (♀) and males (♂) in the population, where the females represent the first seven entries to the vector and the males the following nine, survival, $\mathbf{u}_{1,t} \sim H_{1,t}(\mathbf{n}_{t-1})$:

$$♀ = \begin{bmatrix} \mathrm{u}_{1,0,\mathrm{f},\mathrm{t}} \sim Binomial(n_{0,f,t-1}, \phi_{p,t}) \\ \mathrm{u}_{1,1,\mathrm{f},\mathrm{t}} \sim Binomial(n_{1,f,t-1}, \phi_f) \\ \mathrm{M} \\ \mathrm{u}_{1,5,\mathrm{f},\mathrm{t}} \sim Binomial(n_{5,f,t-1}, \phi_f) \\ \mathrm{u}_{1,6+,\mathrm{f},\mathrm{t}} \sim Binomial(n_{6+,f,t-1}, \phi_f) \end{bmatrix} \qquad ♂ = \begin{bmatrix} \mathrm{u}_{1,0,\mathrm{m},\mathrm{t}} \sim Binomial(n_{0,m,t-1}, \phi_{p,t}) \\ \mathrm{u}_{1,1,\mathrm{m},\mathrm{t}} \sim Binomial(n_{1,m,t-1}, \phi_m) \\ \mathrm{M} \\ \mathrm{u}_{1,7,\mathrm{m},\mathrm{t}} \sim Binomial(n_{7,m,t-1}, \phi_m) \\ \mathrm{u}_{1,8+,\mathrm{m},\mathrm{t}} \sim Binomial(n_{8+,m,t-1}, \phi_M) \end{bmatrix}$$

Where $\mathrm{u}_{1,0,\mathrm{f},\mathrm{t}}$ is the first element of the state vector for the female seals and it represents the number of female pups surviving. The probability of pup survival is a function of the number of pups (of both sexes) born in the previous year, $n_{pups,t-1}$, and is modelled as a density dependent binomial process, with survival probability given by:

$$\phi_{p,t} = \frac{\phi_{p\max}}{1+\beta \times n_{pups,t-1}} \qquad \text{(eqn 2.1)}$$

This is akin to the Beverton-Holt stock recruitment formulation frequently used in fisheries (Quinn & Deriso 1999) where $\phi_{p\max}$ represents the pup survival rate when there are few pups present and β is the main parameter invoking density dependent reductions in the numbers of pups surviving when their numbers increase. The larger the value of this parameter the earlier the onset of density-dependent reductions in pup survival (Thomas *et al.* 2005). Survival for females aged one and over is assumed to be density independent with survival probabilities ϕ_f .

Male pup survival is similarly density dependent, with the same parameter $\phi_{p,t}$. For males aged 1 to 7 $\phi_m = \phi_f$, however on reaching maturity, at the age of 8, the survival rate of males is thought to decrease, and this lower rate is given by ϕ_M .

Survival is then followed by age advancement, $\mathbf{u}_{2,t} \sim H_{2,t}(\mathbf{u}_{1,t})$:

$$\text{♀} = \begin{bmatrix} u_{2,0,f,t} = 0 \\ u_{2,1,f,t} = u_{1,0,f,t} \\ u_{2,2,f,t} = u_{1,1,f,t} \\ \text{M} \\ u_{2,5,f,t} = u_{1,4,f,t} \\ u_{2,6+,f,t} = u_{1,5,f,t} + u_{1,6+,f,t} \end{bmatrix} \qquad \text{♂} = \begin{bmatrix} u_{2,0,m,t} = 0 \\ u_{2,1,m,t} = u_{1,0,m,t} \\ u_{2,2,m,t} = u_{1,1,m,t} \\ \text{M} \\ u_{2,7,m,t} = u_{1,6,m,t} \\ u_{2,8+,m,t} = u_{1,7,m,t} + u_{1,8+,m,t} \end{bmatrix}$$

The age advancement sub-process is entirely deterministic; all females are aged by one year, except for age six and older females who remain in the same class. Males can again be described in the same manner but with the addition of two more age classes. (It should be noted that $H_{2,t}(\mathbf{u}_{1,t})$ is a degenerate distribution such that with $\mathbf{u}_{1,t}$ being given, $\mathbf{u}_{2,t}$ is known with certainty).

The final birth (and sexing of pups) process is given by:

$$(n_{0,f,t},\ n_{0,m,t},\ m_t) \sim Trinomial(u_{2,6+,f,t}, 0.5\gamma,\ 0.5\gamma,\ (1-\gamma))$$

where:

$n_{0,f,t}$ = the number of female pups born;

$n_{0,m,t}$ = the number of male pups born;

$u_{2,6+,f,t}$ = the number of mature (breeding age) females present;

$m_t = (u_{2,6+.f.t} - n_{0,f,t} - n_{0,m,t})$ = the number of mature females that have not successfully given birth, and;

γ = the probability that a female successfully gives birth to and weans a pup.

The trinomial distribution given above is equivalent to performing two binomial calculations in succession. The first calculation gives the number of pups produced, as a density independent function of the number of mature females present, $u_{2,6+,f,t}$, with binomial probability γ. The second binomial calculation partitions those pups born into sexes (i.e. the number of female pups produced, $n_{0,f,t}$, being a binomial random variable of the total number of pups produced with probability 0.5). The trinomial distribution thus allows us to perform both births and the sexing of pups in one step.

The entire population vector in year t is therefore given by:

$$\mathbf{n}_t = \begin{bmatrix} \mathbf{n}_{f,t} \\ \mathbf{n}_{m,t} \end{bmatrix}$$

where:

$$\mathbf{n}_{f,t} = \begin{bmatrix} n_{0,f,t} \sim as\ above \\ n_{1,f,t} = u_{2,1,f,t} \\ n_{2,f,t} = u_{2,2,f,t} \\ \mathrm{M} \\ n_{6+,f,t} = u_{2,6+,f,t} \end{bmatrix} \qquad \mathbf{n}_{m,t} = \begin{bmatrix} n_{0,m,t} \sim as\ above \\ n_{1,m,t} = u_{2,1,m,t} \\ n_{2,m,t} = u_{2,2,m,t} \\ \mathrm{M} \\ n_{8+,m,t} = u_{2,8+,m,t} \end{bmatrix}$$

The mating model that gives rise to the births process cannot easily be specified in the vector form as above. I shall, however, describe it presently out with this vector form. The mating model determines the genotypes of the pups born with respect to the genotypes of the mature seals that mate. If this were to be incorporated the states would have to reflect genotype as well as age and sex. The number of permutations of genotypes becomes rapidly very large as the number of loci, and the number of alleles possible at each locus, increases, and therefore a representation in this form has not been presented here.

For the random mating model each mature breeding age female (6+), that has been selected to breed, mates with one mature breeding age male (8+) chosen at random. This male is then replaced into the male population (i.e. one male can potentially father more than one pup). For the mate "choice" model the situation is somewhat more interesting.

I have attempted to emulate the mate "choice" mechanism probabilistically in the following way. If the female has not mated before then mating follows the random model. However, if the female has mated before, then she can copulate with an equal probability with between two and six males. Six males was chosen as an upper limit based on Amos *et al*'s (2001b) hypothesis that females mating both on land and aquatically could potentially mate with this number of males. These $2-6$ males are chosen at random from the 8+ male population. An index of relatedness (Queller & Goodnight 1989, Van de Casteele, Galbusera & Matthysen 2001), R_i, was calculated between each of these new males (subscripted *i*) and the female's previous partner (subscripted *old*) using the coefficient developed by Queller & Goodnight (1989), such that:

$$R_i = \frac{\sum_k \sum_a (P_{i,k,a} - \overline{P_{old,k,a}}) + \sum_k \sum_a (P_{old,k,a} - \overline{P_{i,k,a}})}{\sum_k \sum_a (P_{old,k,a} - \overline{P_{old,k,a}}) + \sum_k \sum_a (P_{i,k,a} - \overline{P_{i,k,a}})} \qquad \text{(eqn 2.2)}$$

The calculation is summed over loci (k) and allelic position (a) (seals are diploid and therefore the sum is over 2 allelic positions). In the first half of the equation (i.e. the numerator and denominator parts before the + signs) the female's previous partner is used as the reference individual such that $P_{old,k,a}$ can be 0.5 or 1, depending on whether the female's previous partner is heterozygous or homozygous at the locus under scrutiny. Whereas $\overline{P_{old,k,a}}$ is the frequency of the *old* partner's allele in the population at large. Finally, $P_{i,k,a}$ can be 0, 0.5, or 1 (0 if the new male does not have the allele under consideration, 0.5 if it has it and is heterozygous and 1 if it has it and is homozygous). The second half of the equation gives a similar calculation using the new male as the reference individual. This index of relatedness is only useful with multiallelic loci; with a diallelic locus the denominator is zero when the two compared individuals are both heterozygous (Van de Casteele *et al.* 2001). A positive R value means the individuals are related, and a zero or negative value indicates that the individuals are unrelated. The next task in the simulations is to calculate a probability of fertilization for each of the new males that the female has copulated with – to do this R_i was re-scaled so that it could only be positive and re-labelled as D_i. The probability of fertilization for male i, F_i, was then calculated as:

$$F_i = \frac{\frac{1}{D_i}}{\sum_{j=1}^{n} \frac{1}{D_j}} \qquad \text{(eqn 2.3)}$$

where the summation limit in the denominator, n, ranges from 2-6 depending on the number of copulations the female in question had. Of the new males selected for copulation, the one with the highest D_i value (the one most related to the females previous partner) will have the lowest F_i value. Now say, for example, we had 4 males (♂1 to ♂4) that had copulated with the female under consideration and found that:

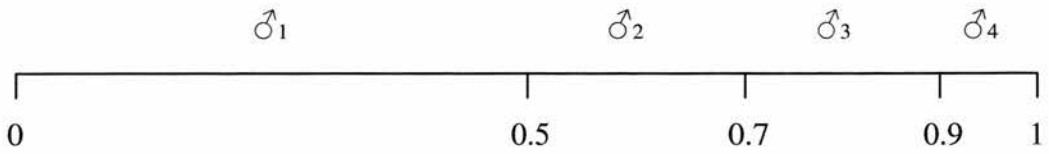
$$F_1 = 0.5$$

$$F_2 = 0.2$$

$$F_3 = 0.2$$

$$F_4 = 0.1$$

We then generate one deviate (i.e. one male for fertilization) from this multinomial distribution, for example by using one deviate from a Uniform distribution between zero and one:



The male that fertilized the female then replaces the female's previous partner (and is stored along with the female) and is used for comparison in the following year (assuming the female survives and is selected for mating in the following year). On mating, the pup produced inherits one allele from its mother and one from its father at each locus. The loci are assumed not to be linked. Furthermore, no affects of mutation are modelled to occur, as they are thought unlikely to be of great significance over the time scales I am investigating.

2.2.3 State process: matrix representation

For ease of interpretation it has become conventional to define the state process in terms of population projection matrices. The familiar Leslie matrix for population projection of the abundance of each age class to the next is defined (see Caswell 1989) as:

$$\mathbf{n}_t = \mathbf{A}\mathbf{n}_{t-1} \qquad \text{(eqn 2.4)}$$

where **A** is the population projection matrix. This formulation can be extended to account for a stochastic state process, such that:

$$E\left[\mathbf{n}_t \mid \mathbf{n}_{t-1}\right] = \mathbf{A}\mathbf{n}_{t-1} \qquad \text{(eqn 2.5)}$$

where $E\left[\mathbf{n}_t \mid \mathbf{n}_{t-1}\right]$ is the expected value of $\mathbf{n}_t$ given $\mathbf{n}_{t-1}$ and $\mathbf{A}$ now represents the average effect of a set of stochastic processes. It is useful to split $\mathbf{A}$ up into a series of sub-process matrices. For the age-structured definition of the seal model developed the key sub-processes are survival, advancement to the next age category and births (which includes the assignment of sex). These can be expressed respectively by the matrices $\mathbf{S}_t$, $\mathbf{A}$ and $\mathbf{B}$, such that:

$$E\left[\mathbf{n}_t \mid \mathbf{n}_{t-1}\right] \approx \mathbf{B}\mathbf{A}\mathbf{S}_t\mathbf{n}_{t-1} \qquad \text{(eqn 2.6)}$$

The expectation being an approximation due to the non-linearity in the pup survival function, as described earlier. Again, for the reasons that have been discussed, the mating process cannot be explicitly specified in the matrix formation. For the parts that can be specified, for the females and males in the population, we first have the survival matrix defined as:

$$\mathbf{S}_t = \begin{bmatrix} \mathbf{S}_{f,t} & 0 \\ 0 & \mathbf{S}_{m,t} \end{bmatrix}$$

where:

$$\mathbf{S}_{f,t} = \begin{bmatrix} \phi_{p,t} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_f & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_f & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_f & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_f & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \phi_f & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_f \end{bmatrix} \quad \mathbf{S}_{m,t} = \begin{bmatrix} \phi_{p,t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \phi_m & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \phi_m & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_M \end{bmatrix}$$

This is followed by age advancement:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_f & 0 \\ 0 & \mathbf{A}_m \end{bmatrix}$$

where:

$$\mathbf{A}_f = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad \mathbf{A}_m = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

And finally births (which includes the sexing of pups):

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0.5\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

In complex situations splitting up the process matrix into sub-processes is advantageous. For my model the single Leslie matrix is the product of the above three sub-process matrices and is:

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0.5\phi_f\alpha & 0.5\phi_f\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\phi_{p,t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \phi_f & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \phi_f & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \phi_f & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \phi_f & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \phi_f & \phi_f & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0.5\phi_f\alpha & 0.5\phi_f\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_{p,t} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi_m & \phi_M
\end{bmatrix}$$

2.2.4 Computer simulation description

The computer simulations were done using the computer package R: A Programming Environment for Data Analysis and Graphics Version 1.6.1. All model runs were given the parameter values shown in Table 2.1. All the parameters were given values that should be feasible based on the work of Thomas *et al.* (2005). The parameter related to the carrying capacity, β, was given a relatively high value to limit the population growth to a certain extent. The mature male survival parameter, ϕ_M, was given a value compatible with current opinion. One hundred age 8 males and one hundred age 6 females were created and given genotypes drawn at random from the initial allele frequency distributions. These seals having mated went through the survival process and new births occurred on the newly founded colony. As mature

males are given a lower survival rate than mature females it is likely that fewer males made it to the new colony than females. As a consequence of these differing survival rates there is a skew in the sex distribution for the grey seal population. The male population is believed to be approximately 60% of the female population (SCOS 2002).

Four models were explored, each one building on the previous ones. Models 1 and 2 each have 1 locus with 4 alleles possible at that locus (each locus having 2 alleles for each individual seal but 4 alleles possible from the gene pool). In model 3 the locus in model 1 is combined with the locus in model 2 to produce a 2 locus system (i.e. each individual seal having 2 loci with 2 alleles possible at each). In the fourth and final model two more loci are added, again each with 4 possible alleles, to the system of model 3. The initial gene pool allele frequencies for all the models are described in Table 2.2, where A stands for allele.

Table 2.1: Parameter values used in the simulation model runs

Parameter	Description	Value
ϕ_f	Age 1+ female survival rate	0.96
ϕ_m	Age 1 to age 7 male survival rate	0.96
ϕ_M	Mature (i.e. age 8+) male survival rate	0.8
$\phi_{p,\max}$	Pup survival rate when few are present	0.6
γ	Probability of female breeding successfully	0.92
β	Related to carrying capacity	0.005

Table 2.2: Initial allele frequencies for the four simulation models

Model	locus	Allele frequency															
		A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	A16
1	1	0.5	0.3	0.1	0.1												
2	1					0.4	0.3	0.2	0.1								
3	1	0.5	0.3	0.1	0.1												
	2					0.4	0.3	0.2	0.1								
4	1	0.5	0.3	0.1	0.1												
	2					0.4	0.3	0.2	0.1								
	3									0.6	0.2	0.1	0.1				
	4													0.6	0.25	0.1	0.05

All of the models were run with 100 "particles" for 50 years (where one particle represents one possible population with parameter values that are specific to it). The outputs of the models run with random mating and those run with mate choice were compared graphically. These outputs include the trajectory of the following through time: the total 1+ population size, the ratio of 1+ males to 1+ females, the rate of increase of the 1+ population and the "true" pup allele frequencies. As the models incorporated demographic stochasticity, the mean of the simulated allele frequency trajectories provides an estimate of the population's genetic structure through time, and the variation among trajectories indicates the anticipated demographic variability under the chosen model. Bootstrapping was also done to compare the mean differences between the pup allele frequencies for the random and mate choice models with 999 bootstrap resamples being taken. The bootstrapping done assumes that the particles are independently and identically distributed. As there is no interaction between the particles this seems a valid assumption. It is fair to also assume that the means of the bootstrap resamples are indicative of the variability we would observe in our sample means if we took repeated samples, thus circumventing the need for a larger number of particles to be run through the model. This is a benefit due to the extensive number of calculations being undertaken for each particle in the mate choice models. More details of the bootstrapping procedure are given below.

2.3 Results

Figure 2.1 shows the mean 1+ population size, the mean ratio of 1+ males to 1+ females, and the mean rate of increase for the population, with 95% confidence limits, for the simulation models.

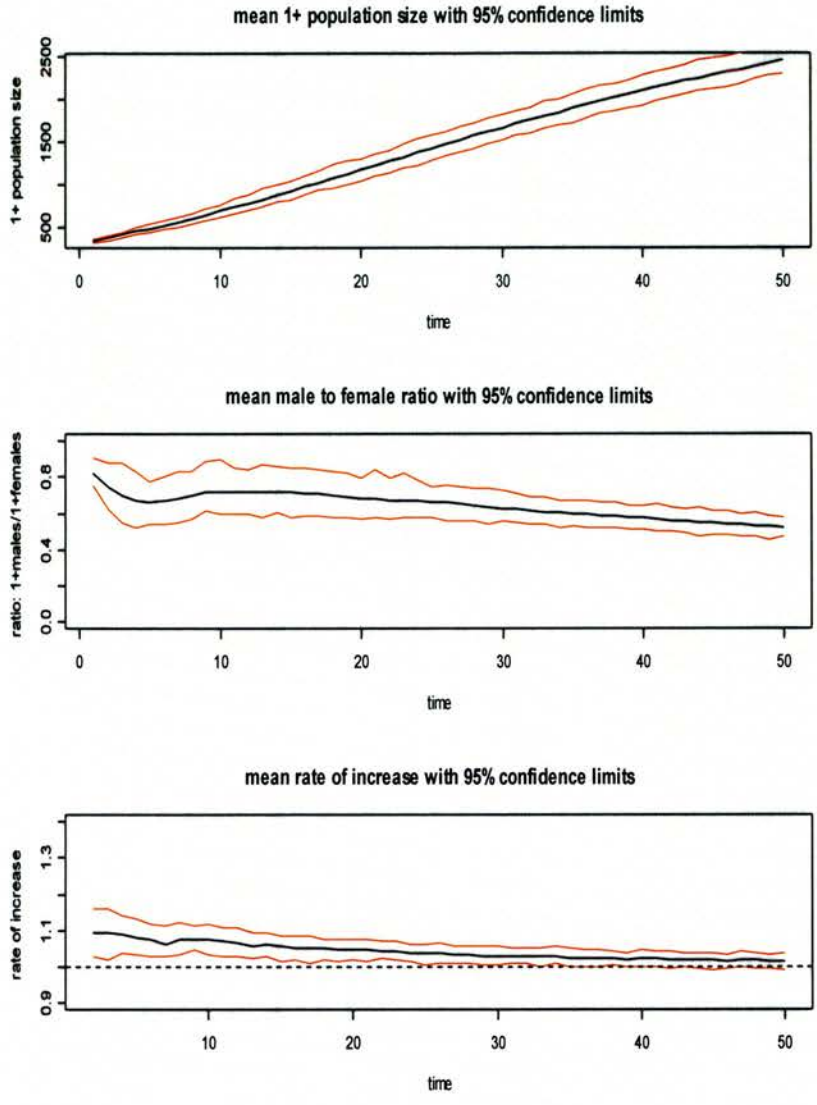


Figure 2.1: Graphs displaying demographic changes for the simulation model runs

These graphs show that the population growth is density dependent and the population size graph shows a slight sigmoidal shape in conjunction with a slowing down of the rate of increase as shown in the third graph. The sex ratio is initially quite unstable as a consequence of the population being started with only mature seals (see section 4 for a discussion of this artefact). This ratio appears to be stabilizing at between 0.5 and 0.6, which is in line with current opinion on the skew in this ratio.

I now focus on a graphical representation of the results for model 1 (i.e. which has a locus with initial allele frequencies given by 0.5, 0.3, 0.1 and 0.1). I follow the results given by this locus as the complexity of the system grows from a one locus system to a two locus system and finally a four locus system. The locus given in model 2

produced similar results to those I am about to describe for the locus in model 1. Two graphical representations are given for each of the systems. The first shows how the mean of the "true" pup allele frequency for the particles changes through time with 95% confidence limits. The mate choice simulation is shown in red and the random simulation in black. The second graphical representation for each model shows the bootstrap differences between the mean pup allele frequencies, for the mate choice simulations, for each particle at each time point, minus the initial allele frequencies. The difference was calculated in this manner (as opposed to subtracting the mean value for each particle in the random simulation) to reduce Monte Carlo error, as the random model runs in all situations did not show any significant deviation from the initial allele frequencies (as had been expected). In more detail these bootstraps run as follows: The initial allele frequency for allele 1 was subtracted from its frequency for particle 1 at time t =1 for the mate choice model. The same was done for the other 99 particles. 100 values were then resampled with replacement from these 100 differences and the mean recorded. This was repeated 999 times and a 95% confidence interval for the mean difference at time t=1 was produced. The same process was done for all of the time steps of the model. This bootstrapping procedure was done for each allele at each locus.

Figure 2.2 shows the trajectory through time for model 1 with one locus. It appears that for the mate choice simulations the most common allele in the population (allele 1) decreases quite markedly as time passes while the rarest alleles (3 and 4) increase. This is shown somewhat more clearly in the bootstraps of Figure 2.3. These differences are clearly significant as shown by the fact that the bootstrap confidence intervals for these three changing alleles do not contain zero by the end of the time horizon explored. Allele 2 does not change with the passing of time – this is perhaps because its initial frequency is close to 0.25 (1 divided by the number of alleles at the locus). This observation is consistent with the outcome expected from an outbreeding model.

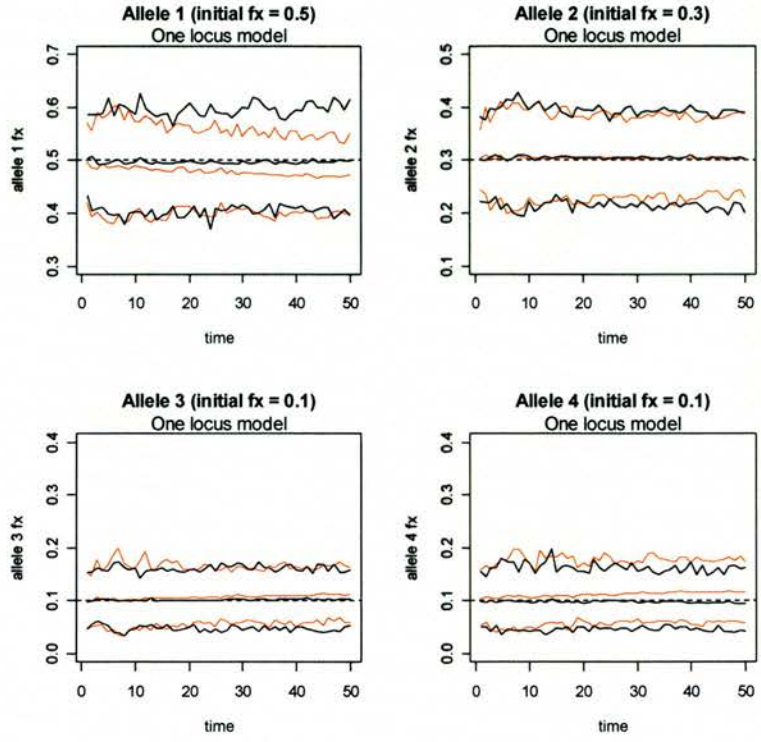


Figure 2.2: Trajectories for the pup allele frequencies through time for the random simulation (black) and the mate choice simulation (red) for model 1. The dotted line gives the initial allele frequency and the solid lines for each of the simulations show the mean, 2.5th and 97.5th percentiles.

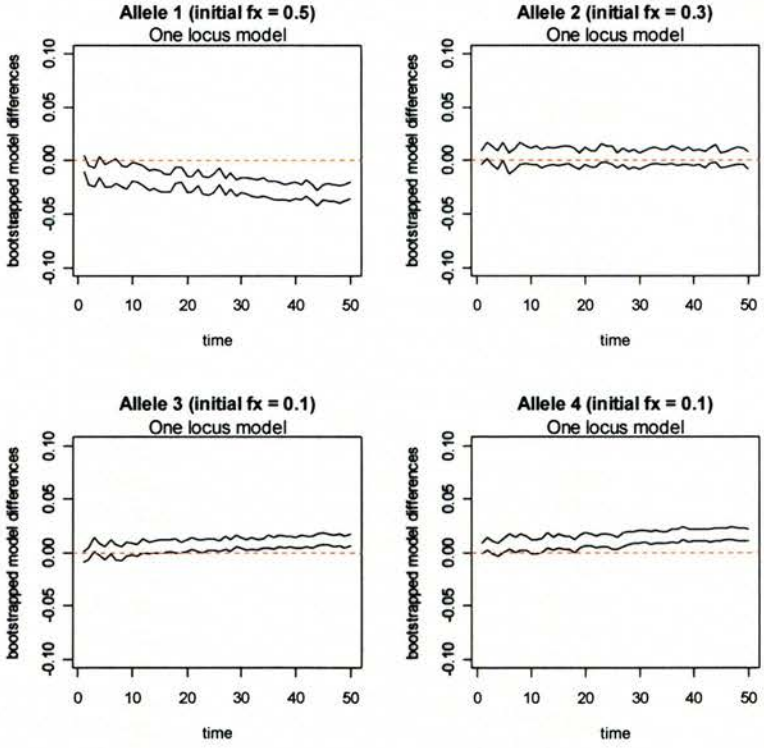


Figure 2.3: Bootstraps performed on the mate choice simulation differences from the initial allele frequencies for model 1. Solid lines show the 2.5th and 97.5th percentiles.

Figure 2.4 shows the allele frequency changes through time for the same locus used in model 1, in model 3 (where there are two loci for the population). These trajectories follow a similar pattern as the one observed for the single-locus model except that the effect is less marked. The bootstraps (shown in Figure 2.5) show that this difference is not immediately significant and emerges more gradually.

Figure 2.6 shows the pup allele frequency changes with time for our locus under scrutiny in the fourth and final model, wherein the seals have four loci. This time the pattern evident in the previous two simpler systems is diminishing somewhat. The most common allele at the locus shows a slight tendency to decrease as time passes; however, the bootstraps (Figure 2.7) show that this difference is not quite significant within the 50 year window. Furthermore, the rarer alleles are no longer showing a tendency to increase.

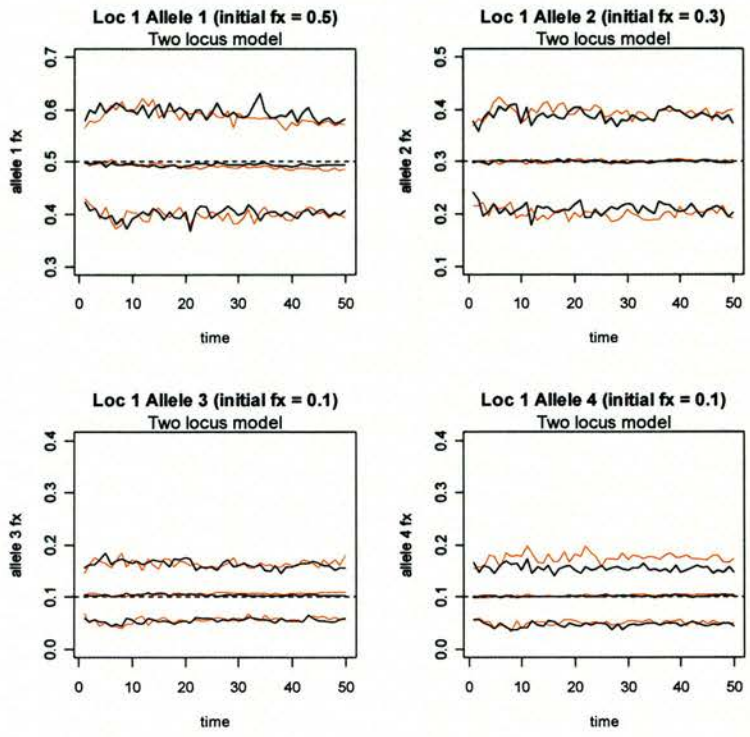


Figure 2.4: Trajectories for the pup allele frequencies through time for the random simulation (black) and the mate choice simulation (red) for model 3. The dotted line gives the initial allele frequency and the solid lines for each of the simulations show the mean, 2.5th and 97.5th percentiles.

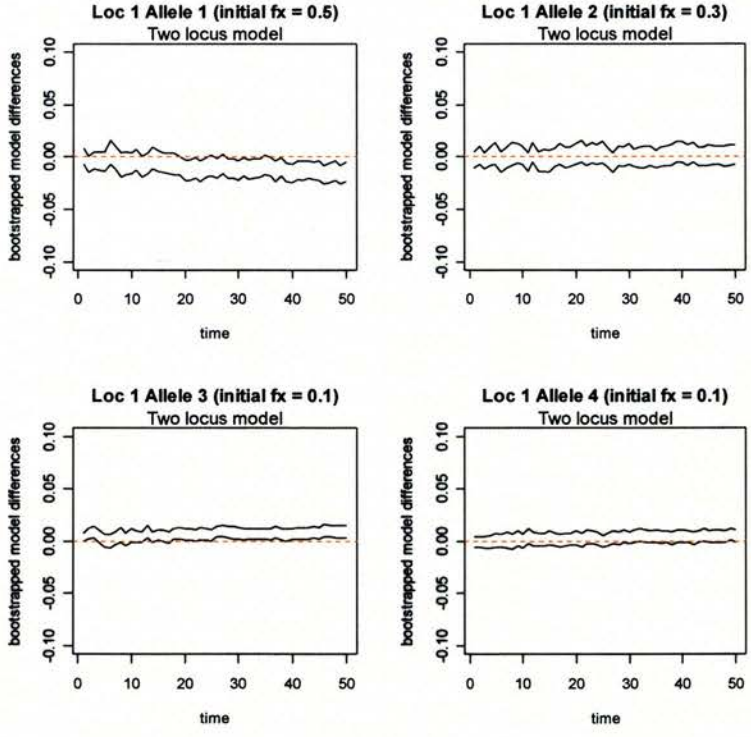


Figure 2.5: Bootstraps performed on the mate choice simulation differences from the initial allele frequencies for model 3. Solid lines show the 2.5th and 97.5th percentiles.

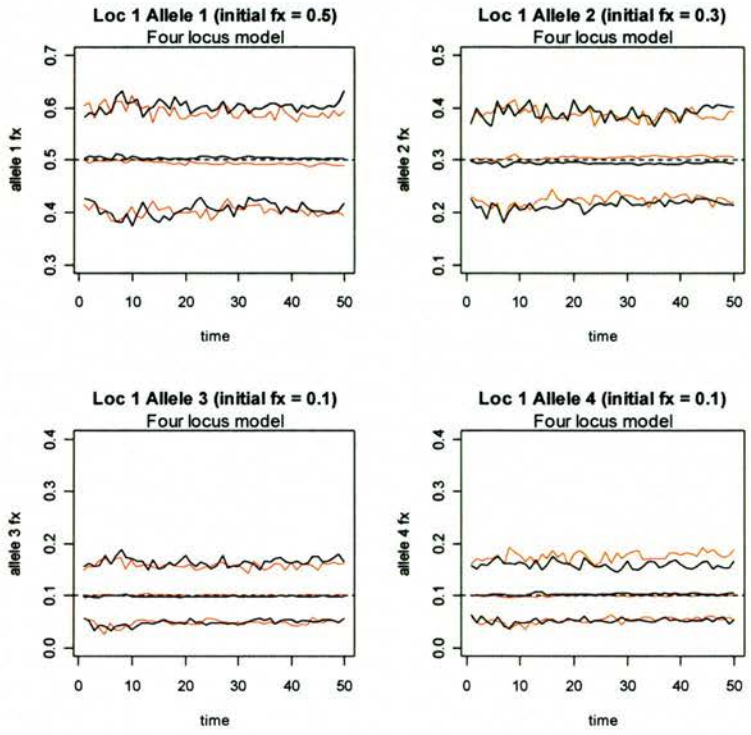


Figure 2.6: Trajectories for the pup allele frequencies through time for the random simulation (black) and the mate choice simulation (red) for model 4. The dotted line gives the initial allele frequency and the solid lines for each of the simulations show the mean, 2.5th and 97.5th percentiles.

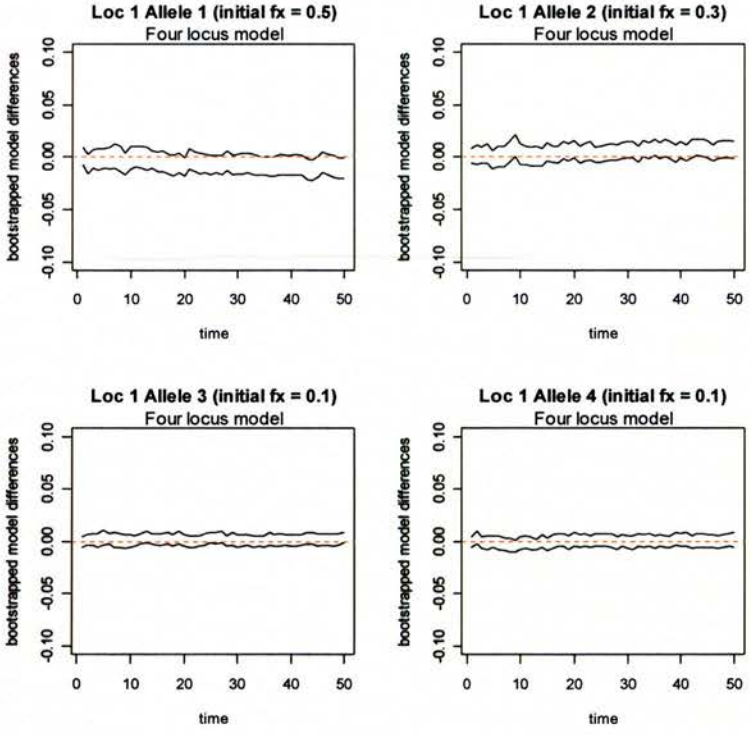


Figure 2.7: Bootstraps performed on the mate choice simulation differences from the initial allele frequencies for model 4. Solid lines show the 2.5th and 97.5th percentiles.

2.4 Discussion

The results hint at patterns, but also present some challenges to interpretation. It is feasible that the initial instability in the sex ratio may have affected the results of the mate choice model because a greater diversity of males is available to females for mating early in the time series. This discrepancy was, however, consistent across simulations. Furthermore it is unlikely that a newly founded colony would have a stable sex or age ratio.

Concerning the pup allele frequencies, it appears that at least in the simpler systems with fewer loci the mate choice mechanism as I have set it up in my simulations causes common alleles to decrease in frequency, rarer alleles to increase in frequency and those close to the average frequency (i.e. close to 0.25 when there are four alleles at the locus) not to change. Why would mate choice by females for a mate dissimilar to the previous one give rise to such observations? If a female chooses a dissimilar male to her last partner; then from those chosen to copulate the male with the rarest allele will more often be dissimilar to the previous, probably more common males, and will thus have a much higher probability of fertilization. On the other hand, a more common male that copulates with the female will often be more similar to the previous male, and will therefore have a lower probability of fertilization which would lead to the decreasing frequency of the commonest allele. One would assume that the endpoint of the simulations (i.e. the allele frequency distributions at equilibrium in this theoretical system) would give allele frequencies of $1/n$ for each allele (where n is the number of alleles at the locus under consideration).

These results are consistent with the expectation of an outbreeding model: rarer forms are favoured and more common forms are penalized. Such a mechanism would act to maintain genetic diversity in the population. However, this effect diminishes as the number of loci across which the comparison is made is increased.

In the real grey seal world the differences between individuals are likely to be affected by multiple interacting loci. Nevertheless, it is feasible that the mate choice mechanism may be driven by differences at only a handful of loci.

2.4.1 Inbreeding avoidance

The mechanism controlling the mate choice of female grey seals is presently somewhat elusive. The mechanism, however, is likely to be either behavioural or physiological. A behavioural mechanism could be driven by MHC olfactory cues, in which case it is plausible that diversity maintenance, where rarer forms are favoured (as represented by the simpler systems presented in this report), could be in operation. However, given that a female can copulate with as many as six males, she would need some way of "knowing" and "remembering" who the true father was.

A physiological mechanism is perhaps more plausible. Amos *et al.* (2001b) postulate that the mate choice may be a consequence of sperm competition – if females have antisperm antibodies, then weak immunointolerance could develop towards the sperm of the females' previous mates. In many mammals the cervical mucus contains such antibodies which develop due to exposure to sperm and may reduce fertility. However, Amos *et al.* (2001b) also point out that these antibodies would need to be raised after fertilization and that this hypothesis is presently without empirical support, so therefore MHC-related olfactory cues and other, at the moment unidentified, mechanisms cannot be dismissed.

Given the high levels of site fidelity of both male and female grey seals (Pomeroy *et al.* 1994; Twiss *et al.* 1994) under the classical polygynous mating system it is evident that inbreeding depression may result where the offspring of closely related parents show reduced fitness (Amos *et al.* 2001a). It therefore seems likely that mating systems, such as the mate choice system explored in this chapter, would evolve as a means of avoiding inbreeding. Inbreeding avoidance has also been put forward as a major determinant in the evolution of dispersal (Clobert, Ims & Rousset 2004). It is interesting to note that in the grey seal system a male arriving from another colony may accrue an above average success rate in fertilization of females given his dissimilarity to the other males at that colony.

2.4.2 Further modelling considerations

There are other complexities in the mating system that could be in operation that have not been explored here which may affect the genetic composition of grey seal colonies. These include: (i) some females having higher lifetime reproductive success than others; (ii) inbred pups having a reduced survival rate (Amos *et al.* 2001a) and; (iii) mate fidelity (Amos *et al.* 1995). These three scenarios could potentially be explored, separately or together, using individual-based simulation models similar to those used in this chapter. Furthermore, the mate choice mechanism could be set up to compare the males mated with to all of the females' previous partners as opposed to just the last partner.

The simulation models presented in this chapter utilised a relatedness coefficient developed by Queller & Goodnight (1989) to compare the copulating male seals to one another. There are other methods for calculating relatedness between pairs of individuals from microsatellite data that one can utilise (see examples in Van de Casteele *et al.* 2001). However, as the simulations explored in this chapter were purely hypothetical it would have been interesting to calculate the "true" relatedness between individuals based on the ancestry information that could potentially be saved and searched during the simulations. Models where all the founders of the colony are unrelated could be compared to those where some of the colony's founding seals are relatives. Such scenarios, however, would necessitate a lot more computational time and memory during the simulations. These possible extensions were not explored during this thesis due to time limitations.

Chapter 3

Use of microsatellite analysis to identify links between grey seal breeding colonies

3.1 Introduction

Reproduction in grey seals is highly synchronized, both spatially and temporally (Pomeroy *et al.* 1994; Boyd 2002), with females aggregating colonially during the breeding season to give birth, suckle their pups and to mate. These breeding colonies are grouped into three distinct populations between which no mixing is believed to occur: the north-west Atlantic (the largest population); the north-east Atlantic; and a relatively small population in the Baltic (Allen *et al.* 1995). The grey seal has a world population of around 300,000 individuals, 40% of which breed around the British Isles (SCOS 2004).

The grey seal has high dispersal capabilities: satellite tracking data have shown that these mammalian predators can travel long distances in short periods of time (McConnell *et al.* 1992; Hammond *et al.* 1993). Given the wide-ranging ability of these animals and their potential ability to visit many colonies during the breeding season, one might expect low genetic structuring (Jones *et al.* 2004) within the three recognized populations. However, Allen *et al.* (1995) found small but highly significant differences in microsatellite allele frequencies between two colonies,

North Rona and the Isle of May, in the north-east Atlantic population, indicating that there may be restricted gene flow between colonies. This is supported by studies that have shown high site fidelity by female grey seals on Sable Island in the north-west Atlantic population (Boness & James 1979) and by both males and females at North Rona (Pomeroy *et al.* 1994, Twiss *et al.* 1994). Both sexes have been shown to be philopatric (returning as adults to breed at their natal colony) at North Rona and the Isle of May (Pomeroy *et al.* 2000).

The advantages of site fidelity include familiarity with the terrain and with other animals visiting the colony – this could reduce energy expenditure on aggression between females and males, and between rival males (Twiss *et al.* 1994). The potential benefits of philopatry in females include access to prime breeding locations on a colony through matrilineal associations, and an increased likelihood of cross-sucking (Pomeroy *et al.* 2000). However, there is a striking disadvantage to these behaviours: the increased likelihood of mating between relatives, potentially leading to inbreeding depression (Amos *et al.* 2001a). Movement of animals between colonies could reduce the chances of this occurring. Indeed, inbreeding avoidance has been put forward as a major determinant in the evolution of dispersal (Clobert, Ims & Rousset 2004). The relative strength of these two forces (site fidelity and philopatry on the one hand and dispersal on the other) together with non-random mating patterns will determine the genetic structure of the population. High site fidelity accelerates genetic drift, causing colonies to become differentiated from one another, whereas dispersal will act to homogenise colonies (Slatkin 1987).

There is a growing body of evidence that movement does occur between grey seal breeding colonies, with anecdotal evidence of both male and female pups recruiting to non-natal colonies and adult females breeding on different colonies in different years (Harwood *et al.* 1975, Pomeroy *et al.* 2000). A recent multisite capture-recapture analysis of adult females breeding on the four main colonies in the North Sea region has shown that a small proportion of females do move between colonies in successive years and that the probability of movement is affected by the distances separating the colonies (see chapters 4 and 5). However, Thomas *et al.* (2005) found little evidence for the movement of recruiting females among regions in the UK.

The above evidence gives a direct indication of the present level of movement between colonies. Genetic data, on the other hand, provides a measure of the amount of movement between colonies averaged over a long time period (Slatkin 1987). If the present levels of exchange between colonies are indicative of past levels, one would expect colonies from different regions around the UK to be more differentiated than those from the same region. Of further interest is the level of differentiation between the three recognized populations of the grey seal, between which no present mixing is believed to occur. For instance, are animals from the north-east Atlantic population more differentiated from colonies on the east coast of the UK than from those on the west coast?

In this chapter I explore the population structure of grey seal colonies in more detail using nine microsatellite markers (see section 1.3 of chapter 1 for a description of these markers). I compared the differentiation between Sable Island and UK colonies, among colonies from different regions around the UK, and among colonies in one region: the North Sea. Four types of analyses were used: (1) pairwise exact tests, which use randomization procedures to test for differentiation between pairs of colonies (Goudet *et al.* 1996) ; (2) pairwise F_{ST} (Weir & Cockerham 1984), which measures the proportion of variance in allele frequencies that can be attributed to differences between pairs of colonies (Balloux & Lugon-Moulin 2002); (3) assignment-based tests (Cornuet *et al.* 1999), which assign each individual's genotype to the colony where its likelihood of occurrence is highest; and (4) Bayesian model-based clustering (Pritchard *et al.* 2000; Falush *et al* 2003), which calculates posterior probabilities for the number of distinct sub-populations represented in the data set analysed.

3.2 Methods

3.2.1 Study animals, sampling occasions and locations

Tissue samples were collected from a total of 541 grey seal pups from seven colonies. The colonies included one from the north-east Atlantic – Sable Island (Nova Scotia, *n*

= 28) and six from the north-west Atlantic: Faray (Orkney, n = 76); Monach Isles (Outer Hebrides, n = 94); Isle of May (North Sea, n = 134); Fast Castle (North Sea, n = 82); Farne Islands (North Sea, n = 94) and; Donna Nook (North Sea, n = 33). The locations of the six colonies around the UK are shown in Figure 1. All samples were collected in the late 1990s and early 2000s by staff of Cambridge University and St Andrews University.

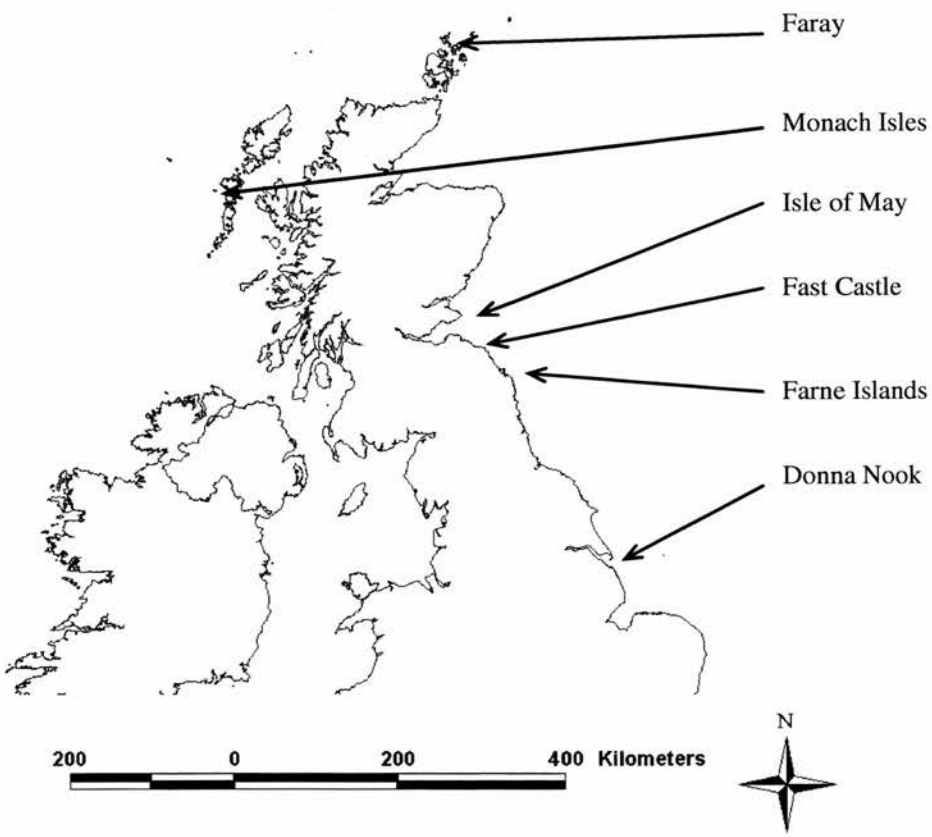


Figure 3.1 Locations of the six colonies around the UK from which tissue samples of grey seal pups were taken.

3.2.2 Microsatellite genotyping

DNA was extracted and purified from the tissue samples and PCR[6] was used for selective amplification of the nine microsatellite loci. The PCR products were then separated using gel electrophoresis (whereby the application of an electric current through the gel caused alleles to migrate to different positions along the gel). The gels were then exposed to X-ray film and the banding patterns evident on the film provided the required information on the alleles possessed by each pup at each locus (Joe Hoffman, pers. comm.). More details on the genotyping process can be found in Hoffman & Amos (2005). This genotyping was carried out by staff at Cambridge University.

3.2.3 Three levels of analysis

In the first analysis, the samples from the colonies at Sable Island and the Farne Islands, which were expected to be quite distinct, were compared using the four statistical methods outlined in section 3.2.5. This analysis was done to test these four methods on data where a pronounced difference is expected – given that no present exchange of individuals between the eastern and western Atlantic populations is thought to occur. In the second analysis, Sable Island was compared to colonies from three different regions around the UK – Faray, the Monach Isles and the Farne Islands. The level of structuring between the colonies from these different UK regions was also explored. In the third analysis, four colonies within the North Sea region – the Isle of May, Fast Castle, the Farne Islands and Donna Nook – were compared.

3.2.4 Preliminary data analysis

Test for departures from Hardy-Weinberg equilibrium (HWE) were undertaken for each locus at each colony using randomization tests. The tests report whether there is

[6] PCR – a cyclic, in vitro enzymatic reaction by which a small fragment of DNA can be replicated exponentially (Baker & Lento 2002)

a deficit or an excess of heterozygotes at any of the loci and overall. Departures from HWE at one locus may indicate the presence of null alleles[7]. Departures across all loci could be an indication that non-random mating is occurring (Allen *et al.* 1995).

Tests for the presence of genotypic disequilibrium (i.e. associations between loci) were performed for each pair of loci in each sample using randomization tests. Such associations between loci can be due to the loci being close to one another on the same chromosome or alternatively through the admixture of two or more populations with different gametic frequencies (Hartl 2000).

These preliminary tests on the data were carried out using the computer program *FSTAT* (Goudet 1995).

3.2.5 Colony differentiation and structuring

The degree of differentiation between the grey seal colonies was assessed using the four statistical methods detailed below.

Pairwise Exact tests

For each pair of colonies complete multilocus genotypes were randomised between the two samples. A multilocus log likelihood ratio statistic (known as the G-statistic) was then used to classify contingency tables. For each locus the contingency tables, excluding the marginal totals, were of size 2 (the number of colonies compared) by *A* (the number of alleles at the locus). For each table the individual locus G-statistic (Goudet *et al.* 1996, Petit, Balloux & Goudet 2001) was calculated as

$$G = -2\sum_{c=1}^{2}\sum_{a=1}^{A} n_{ca} \ln\left(\frac{n_{ca}}{n_c p_a}\right) \qquad \text{(eqn 3.1)}$$

[7] Null allele – an allele that fails to amplify during the genotyping process.

where n_{ca} is the number of occurrences of allele a (from a set of A alleles at the locus) at colony c, n_c is the number of alleles at colony c (twice the number of pups) and p_a is the number of occurrences of allele a in both colonies divided by the number of alleles sampled from both colonies at the locus under consideration. Hence, n_{ca} represent the observed frequencies and $n_c p_a$ the expected frequencies, for each randomized contingency table. The multilocus G-statistic is simply the summation over loci of the statistics for each locus. The proportion of randomizations with a multilocus G-statistic larger in magnitude than or equal to that of the original data set gave the required p-value of the test. This test statistic is powerful at detecting population differentiation even when sample sizes are unbalanced, and is valid even when there is non-random mating (Goudet *et al.* 1996; Petit, Balloux & Goudet 2001). This exact test was performed using the software *FSTAT*.

Pairwise F_{ST}

Weir & Cockerham's (1984) estimate of F_{ST} (θ) was calculated for each locus separately and over all loci, for each of the pairwise comparisons undertaken. *FSTAT* was utilized to make these comparisons. Pairwise θ values measure the proportion of variance in allele frequencies that is explained by differences between the two colonies (Weir 1996), such that

$$\theta = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_i^2 + \sigma_a^2} \qquad \text{(eqn 3.2)}$$

where the variance components correspond to between colonies, σ_c^2, between individuals within colonies, σ_i^2, and between alleles within individuals, σ_a^2. When the estimate of θ is close to unity there exists substantial differentiation, and conversely, a value close to zero is consistent with panmixia, or random mating between the two colonies (Balloux & Lugon-Moulin 2002).

The estimate of θ can be negative if the true value is positive but very close to zero, or when the parameter is in fact negative (a negative interclass correlation not a negative

variance). The latter case implies that alleles are less related within colonies than between them (Weir 1996).

A confidence interval was also calculated for the estimate of θ over all loci using nonparameteric bootstrapping of sampled loci (Rousset & Raymond 1997). This is advocated when the loci are unlinked and should therefore represent the result of replicated measures of the same evolutionary processes (Excoffier 2001). This resampling technique, however, is only asymptotically correct, converging to the correct confidence interval only when the number of loci sampled increases (for a discussion of this property see Raymond & Rousset 1995) – the program *FSTAT* only permits bootstrapping when more than four loci have been sampled.

An alternative measure of population subdivision for microsatellites is R_{ST} (Slatkin 1995) which uses information from the variance in allele sizes as opposed to variance in allele frequency (Balloux & Lugon-Moulin 2002). Gaggiotti *et al.* (1999) concluded from simulation studies that R_{ST} only outperforms F_{ST} under ideal conditions. Under the normal conditions faced by molecular ecologists, F_{ST} estimates of population differentiation may be more reliable than those from R_{ST}. For this reason only the F_{ST} estimates have been reported herein.

Traditionally, F_{ST} estimates have been translated into measures of gene flow (N_m = the number of migrants per generation). However, the mathematical model used to perform this translation makes biologically unrealistic assumptions, the violation of which can produce misleading results (Whitlock & McCauley 1999). For this reason, gene flow estimates have not been given in these analyses.

Assignment-based tests

Using assignment methods, each pup was assigned to the population where its likelihood of occurrence was highest. The likelihoods were calculated according to the criterion of Paetkau *et al.* (1995) and were executed using the computer package *GENECLASS2* (Piry *et al.* 2004). The assignments proceed as follows:

The frequency of allele a at colony c for locus l, p_{acl}, was calculated. If the colony considered was the one from which the pup was sampled, this frequency was computed with this pup excluded (Piry *et al.* 2004). Assuming HWE, the likelihood of the pups genotype, $X_a X_{a'}$, at locus l is proportional to $(p_{acl})^2$ if $a = a'$ and to $2p_{acl}p_{a'cl}$ if $a \neq a'$. Assuming that the loci are independent (i.e. not linked) the likelihood of the pup's multilocus genotype belonging to colony c is the product of the likelihoods across each locus (Cornuet *et al.* 1999). When an allele was found in the pup that was not in the colony considered, the frequency was set to 0.01 at that colony to avoid a zero likelihood being computed (given that the allele may be present but rare at the colony and was therefore not sampled). The value of 0.01 is reasonable given the sample sizes available in this study. According to Paetkau *et al.* (2004), the assignment results are not sensitive to the value set for missing alleles.

Bayesian model-based clustering

Bayesian model-based clustering, a method derived from assignment tests, was performed using the software *STRUCTURE* (Pritchard *et al.* 2000; Falush *et al.* 2003). When running the *STRUCTURE* program there are two primary modelling considerations: (i) whether individuals in the populations can be admixed (i.e. individuals may have recent ancestors in more than one of the populations); and (ii) whether allele frequencies are assumed to be correlated or independent (they could be correlated as a result of shared ancestry in the past).

The Bayesian approach to statistical inference is revolutionising the analysis of data on animal populations, including those from DNA samples (Beaumont & Rannala 2004), enabling scientists to build complex, realistic models which account for stochasticity and can incorporate prior information. Using the *STRUCTURE* program the primary inference task is to obtain samples from the posterior distribution, $\Pr(Z, P|X)$, given that:

$$\Pr(Z, P|X) \propto \Pr(Z)\Pr(P)\Pr(X|Z, P) \qquad \text{(eqn 3.3)}$$

where X represents the genotypes of the sampled pups, Z, the unknown population of origin of the pups and P the unknown allele frequencies in each population. $\Pr(Z)$ and $\Pr(P)$ give the independent prior distributions for P and Z and $\Pr(X|Z,P)$ gives the likelihood. In the admixture model the prior distribution for Z is modified to allow alleles within a pup to have originated from different populations. In the correlated allele frequencies model the prior distribution for P is modified to allow allele frequencies to be correlated across populations. Markov chain Monte Carlo (MCMC) methods[8] are used to obtain an approximate sample from the posterior distribution.

An important difference between the assignment tests undertaken above (section 3.2.5.3) and the analysis done by *STRUCTURE* is that the number of distinct populations is not known *a priori* when *STRUCTURE* is used (Jones *et al.* 2004). Instead this program assigns individuals to populations (without using any information about how many separate colonies were sampled) and gives posterior probabilities for the number of distinct populations (or clusters) represented in the data set analysed. However, Pritchard *et al.* (2000) stress the difficulties associated with inferring the true number of clusters, and propose that their solution (based on Bayesian deviance) should only be used as a guide.

3.3 Results

There was no evidence for departures from HWE or the presence of genotypic disequilibrium at any of the colonies explored in the three analyses using the tests described in section 3.2.4 after standard Bonferroni corrections for multiple tests had been applied.

The numbers of alleles sampled at each locus on each of the colonies explored are presented in Table 3.1.

[8] MCMC – a computer-intensive approach which simulates a Markov chain whose stationary distribution is the posterior of interest (Gamerman 1997).

Table 3.1 Numbers of alleles sampled at each locus on each of the colonies and the total number of alleles at each locus represented in the entire data set.

Locus	Sable Island	Faray	Monach Isles	Isle of May	Fast Castle	Farne Islands	Donna Nook	Total
Hg3.6	6	8	8	8	8	8	8	8
Hg4.2	5	8	5	6	6	7	5	8
Hg6.1	4	6	6	6	6	6	5	6
Hg6.3	6	6	6	6	6	6	6	6
Hg8.9	7	10	9	11	10	10	10	11
Hg8.10	5	9	10	10	9	9	7	10
Hgd.2	7	8	8	8	8	8	8	8
Pv9	4	7	7	7	7	7	7	7
Pv11	6	8	7	8	7	7	6	8

First analysis

Figure 3.2 compares the allele frequencies for the 9 loci genotyped at Sable Island and the Farne Islands. The multilocus G-test showed clear evidence of differentiation between these two colonies, with the results being significant at the 0.1% nominal level. Pairwise θ values are given in Table 3.2 for each locus separately and over all. A 95% bootstrap confidence interval is also provided for the estimate over all loci. There is clear differentiation between these two colonies based on this statistic, which is supported by the fact that the bootstrap confidence interval does not contain zero.

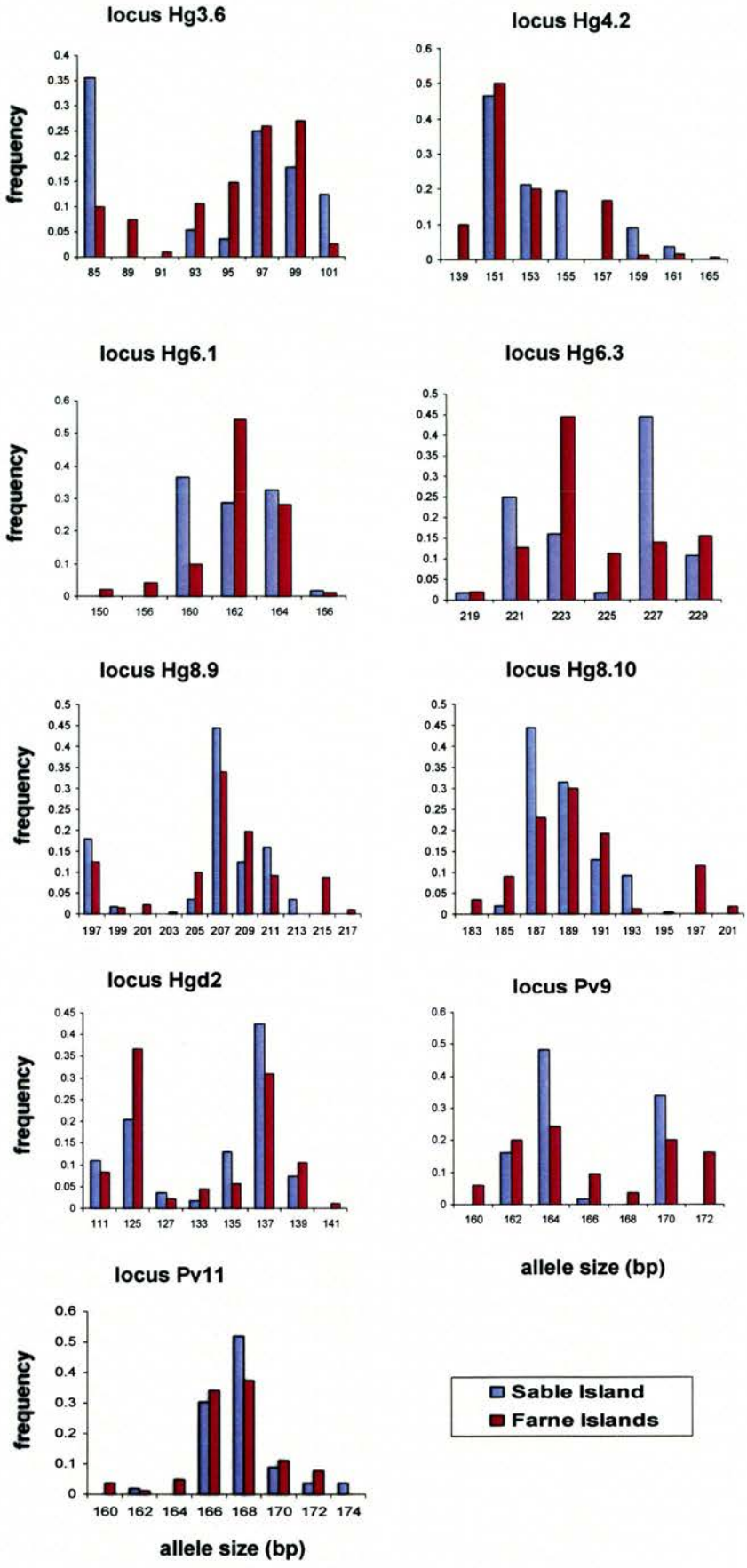


Figure 3.2 Allele frequency comparison plot for the 9 microsatellite loci genotyped for Sable Island and the Farne Islands.

Table 3.2 Weir & Cockerham's (1984) estimates of F_{ST} (θ) for each locus separately and over all in the comparison between Sable Island and the Farne Islands. A 95% bootstrap confidence interval is also provided for the estimate over all loci

locus	θ	95% CI
Hg3.6	0.052	–
Hg4.2	0.048	–
Hg6.1	0.085	–
Hg6.3	0.111	–
Hg8.9	0.012	–
Hg8.10	0.036	–
Hgd.2	0.019	–
Pv9	0.057	–
Pv11	0.009	–
All loci	0.048	(0.028, 0.070)

The results of the likelihood-based assignment methods are shown in Table 3.3. In this case 89.3% of animals were assigned to the colony where they were sampled.

Table 3.3 Proportion of pups assigned to each colony given their sampling locations

Sampling location	No. of pups sampled	Percentage of pups assigned to each colony	
		Sable Island	Farne Islands
Sable Island	28	0.893	0.107
Farne Islands	94	0.106	0.894

In the model-based clustering analysis the "no admixture" model was used based on the assumption that there is now very little exchange of individuals between these two widely separated colonies. The correlated allele frequencies model was also used based on the assumption that current allele frequencies could be correlated because of shared ancestry in the distant past. The models were run with K (the number of populations) ranging from 1 to 4. Each model (with a separate K) was run twice to assess whether or not the models had converged. A burn-in length of 50,000 and a run length of 100,000 were used.

There was little Monte Carlo error between runs of the same model, indicating that the burn-in and run lengths were sufficient. The posterior probabilities of K given in Table 3.4 indicate that there is strong support for two populations. The results from the model with $K = 2$ are shown in Table 3.5 and Figure 3.3. From these results it is

clear that the model is more successful at assigning the Sable Island animals to the correct population than the Farne Islands pups. Nevertheless, most animals were assigned to the correct clusters.

Table 3.4 Inferring the value of K (the number of distinct populations) from model-based clustering using data from Sable Island and the Farne Islands

K	log $P(X\|K)$	$P(K\|X)$
1	-3263.3	~0
2	-3225.9	~1
3	-3299.1	~0
4	-3324.7	~0

Table 3.5 Proportion of membership of each sampled colony in each of the two clusters

Sampling location	Inferred cluster		No. of pups sampled
	1	2	
Sable Island	0.914	0.086	28
Farne Islands	0.273	0.727	94

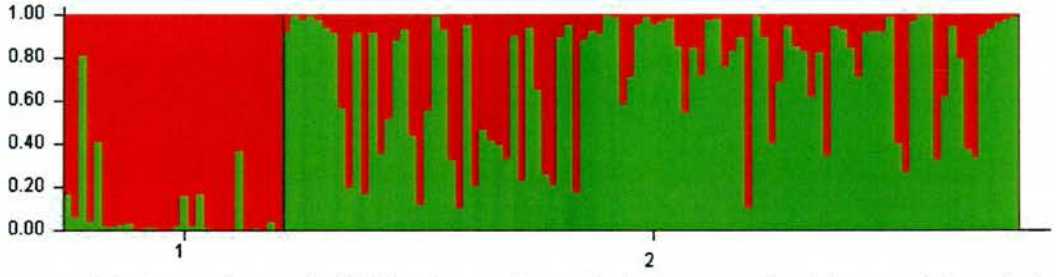


Figure 3.3 Posterior probabilities that each pup belongs to each of the two inferred clusters. Each pup is represented by a vertical coloured line. Red is the colour for inferred cluster 1 and green for cluster 2. The animals sampled in Sable Island are in section 1 along the horizontal axis and those from the Farne Islands are in section 2.

Second analysis

Figure 3.4 compares allele frequencies at the 9 loci genotyped at Sable Island, Faray, the Monach Isles and the Farne Islands. The multilocus G-tests showed clear evidence of differentiation for all pairwise comparisons made, with the results being significant at the 0.1% nominal level for all comparisons apart from Faray versus the Farne Islands, which was only significant at the 1% nominal level. Pairwise θ values are given in Table 3.6 for each locus separately and overall.

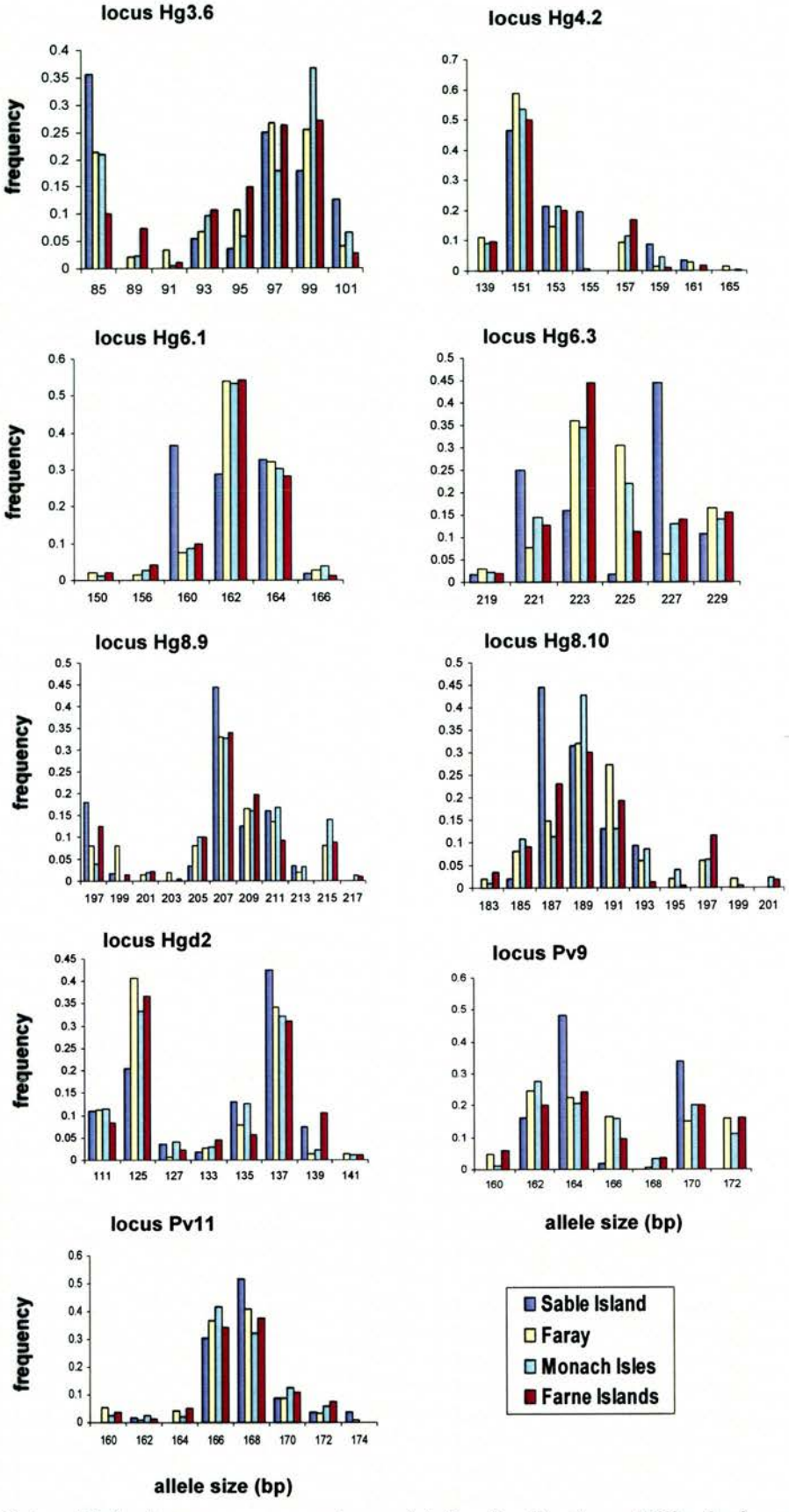


Figure 3.4 Allele frequency comparison plot for the 9 microsatellite loci genotyped for Sable Island, Faray, the Monach Isles and the Farne Islands.

Table 3.6 Weir & Cockerham's (1984) estimates of F_{ST} (θ) for each pairwise comparison at each locus separately and overall (1=Sable Island, 2=Faray, 3=Monach Isles, 4=Farne Islands). A 95% bootstrap confidence interval is also provided for the estimates over all loci

Locus	Pairwise θ					
	1↔2	1↔3	1↔4	2↔3	2↔4	3↔4
Hg3.6	0.012	0.031	0.052	0.009	0.006	0.020
Hg4.2	0.050	0.038	0.048	0.001	0.007	-0.001
Hg6.1	0.095	0.088	0.085	-0.006	-0.007	-0.007
Hg6.3	0.161	0.100	0.111	0.004	0.027	0.008
Hg8.9	0.011	0.024	0.012	0.001	0.000	0.006
Hg8.10	0.061	0.074	0.036	0.017	0.006	0.022
Hgd.2	0.025	0.009	0.019	0.001	0.003	0.004
Pv9	0.082	0.075	0.057	-0.001	-0.000	0.004
Pv11	0.005	0.029	0.009	0.004	-0.003	0.002
All loci	0.058	0.052	0.048	0.004	0.005	0.007
95% CI	(0.027, 0.093)	(0.016, 0.091)	(0.028, 0.070)	(-0.000, 0.008)	(-0.000, 0.013)	(0.001, 0.013)

These results show that the three UK populations are much more differentiated from the Sable Island population than they are among themselves, as had been expected. It had been hypothesised that the Sable Island pups may have been more closely related to seals on the west coast of the UK (Monach Isles) than to those on the east coast (Farne Islands) but, in fact, Faray was found to be the most differentiated colony from Sable Island, followed by the Monach Isles and lastly the Farne Islands. However, it should be noted that the sample sizes are unbalanced, with a rather small sample from Sable Island (28 pups).

The sample sizes for the three UK colonies were larger and more balanced. The most differentiated UK colonies are the two separated by the largest distance (the Monach Isles and the Farne Islands). Again, however, caution should be used in interpreting these results given that the θ values are quite variable between loci, with some negative values being estimated. It should be noted that the 95 % bootstrap confidence intervals for the comparisons between Faray and the Monach Isles, and between Faray and the Farne Islands have a lower limit on the zero borderline.

The results of the likelihood-based assignment methods are shown in Table 3.7. In this case, 52.1% of animals were assigned to the colony where they were sampled. When Sable Island was removed from the analysis, 51.9% of animals were assigned to the colony where they were sampled.

Table 3.7 Proportion of pups assigned to each colony given their sampling locations when Sable Island was retained in the analysis (a) and when it was removed (b)

(a)

Sampling location	No. of pups sampled	Proportion of pups assigned to each colony			
		Sable Island	Faray	Monach Isles	Farne Islands
Sable Island	28	0.893	0.036	0.036	0.036
Faray	76	0.026	0.434	0.316	0.224
Monach Isles	94	0.085	0.245	0.489	0.181
Farne Islands	94	0.096	0.213	0.181	0.511

(b)

Sampling location	No. of pups sampled	Proportion of pups assigned to each colony			
		Sable Island	Faray	Monach Isles	Farne Islands
Sable Island	-	-	-	-	-
Faray	76	-	0.434	0.316	0.250
Monach Isles	94	-	0.255	0.542	0.202
Farne Islands	94	-	0.223	0.213	0.564

Most of the results in Table 3.7 are consistent with the results based on F_{ST}. Only two animals sampled in Faray were assigned to Sable Island (Table 3.7a), most of the UK pups were assigned to their colony of birth (Tables 3.7a and b), but a large proportion were also assigned to the other two colonies. Most pups not assigned to their colony of birth were assigned to the closer of the other two colonies, although the difference in numbers was small. Furthermore, Faray, the colony mid-way between the other two colonies, had the lowest self-assignment rate.

Given that Sable Island is clearly quite differentiated from the other three colonies, it seems appropriate to exclude this colony from the *STRUCTURE* analysis to increase the power to detect the more subtle structure that exists among the UK colonies and also to justify the models used (e.g. it may be appropriate to use the admixture model for comparing the UK colonies to one another but not for comparing them to the north-west Atlantic population).

In this analysis, I used both the "no admixture" model and the "admixture" model. In both cases I assumed that the allele frequencies are correlated to some degree, due to the probable shared ancestry.

Neither of these models showed evidence of sub-structuring among the three UK colonies. Presumably the level of divergence between these colonies is insufficient to be detected using *STRUCTURE* and/or the sample sizes used and number of loci typed is too small.

Third analysis

In Figure 3.5 allele frequencies at the 9 loci genotyped are compared between the four colonies in the North Sea region. The results of the multilocus G-test in Table 3.8 show significant differentiation for the pairwise comparisons involving Donna Nook, but non-significant results for all other comparisons. However, after inspecting the allele frequency comparison plots in Figure 3.5 it appears the significance of the Donna Nook comparisons are driven almost entirely by Locus Hg8.10. If this locus is removed from the analysis, all the exact test comparisons are non-significant.

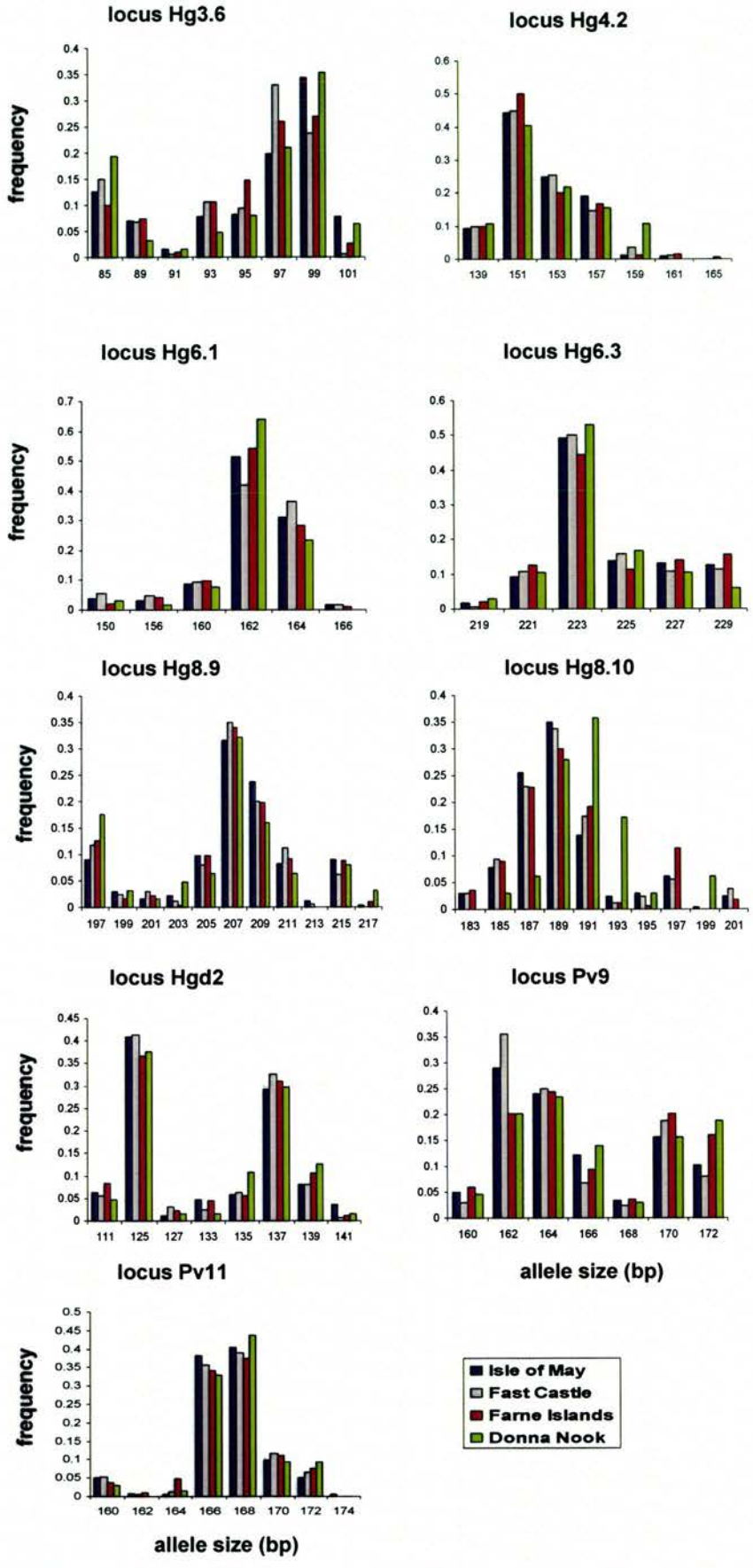


Figure 3.5 Allele frequency comparison plot for the 9 microsatellite loci genotyped for the Isle of May, Fast Castle, the Farne Islands and Donna Nook

Table 3.8 Pairwise multilocus G-test results for the four North Sea colonies. The results are based on standard Bonferroni corrections for multiple tests, where "***" corresponds to significance at the 0.1% nominal level, "**" significance at the 1% nominal level and "*" significance at the 5% nominal level. "NS" stands for non-significant.

	Fast Castle	Farne Islands	Donna Nook
Isle of May	NS	NS	**
Fast Castle	–	NS	***
Farne Islands	–	–	*

Pairwise θ values are given in Table 3.9 for each locus separately and over all. From these comparisons it is quite clear that there is very little differentiation between the four North Sea colonies (especially when one considers the number of negative values estimated and the fact that all the 95% bootstrap confidence intervals clearly contain zero). Again, the differences found between Donna Nook and the other three colonies can be almost entirely explained by differences at locus Hg8.10. The confidence intervals for the Donna Nook comparisons are also rather wide because of this locus.

Table 3.9 Weir & Cockerham's (1984) estimate of F_{ST} (θ) for each pairwise comparison at each locus separately and overall (1=Isle of May, 2=Fast Castle, 3=Farne Islands, 4=Donna Nook). A 95% bootstrap confidence interval is also provided for the estimates over all loci

Locus	Pairwise θ					
	1↔2	1↔3	1↔4	2↔3	2↔4	3↔4
Hg3.6	0.017	0.007	-0.006	0.002	0.013	0.009
Hg4.2	-0.003	-0.000	0.000	-0.001	-0.006	0.004
Hg6.1	0.005	-0.006	0.009	0.009	0.040	-0.002
Hg6.3	-0.004	-0.002	-0.004	-0.001	-0.009	0.003
Hg8.9	-0.001	-0.002	0.000	-0.004	-0.004	-0.005
Hg8.10	-0.003	0.001	0.065	-0.002	0.053	0.052
Hgd.2	-0.003	-0.002	-0.005	-0.003	-0.006	-0.008
Pv9	0.001	0.004	-0.000	0.014	0.016	-0.008
Pv11	-0.004	-0.000	-0.004	-0.005	-0.007	-0.006
All loci	0.001	0.000	0.007	0.001	0.010	0.005
95% CI	(-0.003, 0.005)	(-0.002, 0.002)	(-0.003, 0.024)	(-0.003, 0.005	(-0.002, 0.026)	(-0.005, 0.019)

The results of the likelihood-based assignment methods are shown in Table 3.10. In this case 33.5% of animals were assigned to the colony where they were sampled. When locus Hg8.10 was removed from the analysis, however, only 27.1% of animals were assigned to the colony where they were sampled: the animals sampled in each

colony were assigned in roughly equal amounts between the four possible colonies, indicating very little population structuring.

Table 3.10 Proportion of pups assigned to each colony given their sampling locations when locus Hg8.10 was retained in the analysis (a) and when it was removed (b)

(a)

Sampling location	No. of pups sampled	Proportion of pups assigned to each colony			
		Isle of May	Fast Castle	Farne Islands	Donna Nook
Isle of May	134	0.299	0.261	0.306	0.134
Fast Castle	82	0.341	0.256	0.305	0.098
Farne Islands	94	0.255	0.266	0.372	0.106
Donna Nook	33	0.182	0.091	0.152	0.576

(b)

Sampling location	No. of pups sampled	Proportion of pups assigned to each colony			
		Isle of May	Fast Castle	Farne Islands	Donna Nook
Isle of May	134	0.269	0.269	0.246	0.216
Fast Castle	82	0.256	0.244	0.280	0.220
Farne Islands	94	0.213	0.277	0.330	0.181
Donna Nook	33	0.303	0.242	0.273	0.182

No analysis was done of this North Sea data set using the *STRUCTURE* program because of the lack of genetic differentiation found between the colonies from this region.

3.3 Discussion

The grey seal pups sampled on Sable Island were clearly differentiated from those sampled at colonies around the British Isles. In the comparisons between Sable Island and the three widely separated UK colonies, Faray (the most northern of the UK colonies analysed) was the most strongly differentiated colony, based on the θ values and the assignment tests, and the Farne Islands (the most southern colony analysed), was the least differentiated. However, sample size at Sable Island was small and the confidence intervals for the θ values in these three comparisons show considerable overlap (see Table 3.6). A larger sample size from Sable Island would help to clarify this relationship. It would also be of interest to see how the animals from the Baltic population fit into this relationship.

In the comparison of colonies from different regions within the UK, there is some support for a correlation between degree of differentiation and distance, with the two most distant colonies being the most differentiated. The ordering of the average θ values across loci for these pairwise comparisons also supports a relationship with distance. In the assignment-based tests most pups were assigned to the colony where they were sampled. Those not assigned to their colony of sampling were assigned with a slight preference to the closer of the other two colonies.

It had been thought that *STRUCTURE*, would be able to detect the sub-structuring between the colonies from different regions around the UK. However, this program can have difficulties when allele frequency differences are small (Pritchard *et al* 2000). It is essentially solving a much more complicated problem than that addressed by the other tests used, where the colony labels are known in advance (Jonathan Pritchard, pers. comm.).

Allen *et al.* (1995) found a significant difference between microsatellite frequencies in samples taken a two UK colonies: North Rona (Outer Hebrides) and the Isle of May (North Sea). In spite of this differentiation, they proposed that there could still be major population subdivisions among the colonies around the UK. For example, there could be an east-west divide with approximate panmixia on either side. The data analysed here lends some support to this hypothesis: there does at least appear to be approximate panmixia among the colonies in the North Sea region.

However, for the North Sea comparison, it was surprising to see such strong differentiation at locus Hg8.10 for the pups sampled at Donna Nook. It was initially thought that this could represent a scoring error made whilst reading the gel for this sample. The gel was re-scored but this made no difference to the analysis results. The difference found at this locus could therefore be due to: sampling of families; founder effects; and/or small sample size.

Sampling of families could occur as a result of the high site fidelity and philopatry observed in grey seals (see section 3.1). However, Donna Nook is a vast, featureless beach. Spatial clustering of closely-related individuals is less likely to occur here than

at colonies like North Rona and the Isle of May, where such clustering has been observed. Founder effects could be important at Donna Nook, where the number of pups produced increase by three-fold between 1992 and 2000. A large proportion of this growth is believed to be attributable to immigration. However, this immigration would only lead to a founder effect if all the immigrants came from one part of one colony, or if most founders were from a few families.

A more intriguing possibility is that locus Hg8.10 is actually behaving differently to the other markers as a consequence of linkage to a gene that is under natural selection (Slatkin 1987), favouring different alleles at Donna Nook than at the other three North Sea colonies. Nevertheless, given the small sample size taken at Donna Nook, and the possibility of other factors contributing to this difference, this hypothesis can only be put forward very tentatively.

Overall, however, there appears to be little genetic differentiation among the four North Sea colonies. This is not surprising when one considers the history of colony formation in this region. The only long-established colony in the region is the Farne Islands, whereas only a few pups were born at the Isle of May prior to the mid-1970s, and none were born at Donna Nook. However, during the late 1970s the National Trust began a programme to deter seals from breeding on two of the most important islands in the Farne Islands group, in order to reduce damage to delicate vegetation (Prime 1981). Seals displaced from the Farne Islands are believed to have been responsible for the rapid growth at the Isle of May observed around this time. Fast Castle is the most recently established colony in the region: seals were first noticed there by local fishermen in 1990 (Kevin Rideout, SWT/NTS Ranger, pers. comm.), and most of the immigrants to this colony are thought to have come from the Isle of May.

In conclusion, there appears to be evidence for different levels of structuring of colonies around the British Isles with some evidence for a distance effect and of approximate panmixia in the North Sea region. The microsatellite DNA data compared in this chapter came from pups, so they represent the average effects of philopatry and site fidelity in both sexes. However, in many mammals, such colony faithful behaviours are thought to be stronger in females (Greenwood 1980), with

males being more likely to disperse. In grey seals it is difficult for a female to change breeding colonies because the timing of the peak in pupping varies among colonies. On the other hand, males could gain significant fitness benefits by tracking the shift in the timing of oestrus around the different colonies. However, this would require an ability to predict these differences in timing. Two other factors may make dispersal even more advantageous in males: (1) in the grey seal system a male arriving from another colony may accrue an above average success rate in fertilization of females given his dissimilarity to the other males at that colony (based on the mate "choice" mechanism uncovered by Amos *et al.* (2001b) and explored in chapter 2 of this thesis); and (2) given that a significant proportion of copulations are thought to happen aquatically (Worthington Wilmer *et al.* 1999, 2000), the survival cost to a dispersing male mating aquatically would be less than if he had to do battle with dominant males protecting females on shore. Genetic data collected from adult males and females could help uncover the level of bias in dispersal between the sexes. However, adult grey seals are much more difficult to sample than pups, especially those males that do not come ashore during the breeding season. An alternative, and logistically simpler method, would be to compare estimates of population differentiation from nuclear DNA (microsatellites) to those obtained from mitochondrial DNA (Petit & Mayer 1999; Petit, Balloux & Goudet 2001) for grey seal pups. If the latter indicates more differentiation than the former then a higher male-to-female dispersal rate is implied.

Chapter 4

A Multisite Capture-Recapture Analysis of Female Grey Seals Breeding at Colonies in the North Sea

4.1 Introduction

Grey seals in Britain breed in the autumn at around 50 breeding colonies (Gaggiotti *et al.* 2002), over 90% of which occur in Scotland, particularly in the Hebrides and in Orkney (SCOS 2004). At these colonies females give birth to and suckle a single offspring, and mate (Anderson, Burton & Summers 1975). At North Rona – an island off the north-west coast of Scotland – both males and females grey seals show a high level of breeding site fidelity and philopatry (Pomeroy *et al.* 1994; Twiss *et al.* 1994), suggesting that there is restricted movement between breeding colonies. Gaggiotti *et al* (2002) used the results of DNA analysis to show that grey seal colonies in Orkney have metapopulation structure: new colonies are formed by seals moving from the nearest and largest colonies to found new colonies. However, no assessment has yet been done to determine to what extent adult grey seals move among already established breeding colonies.

Grey seal colonies meet the assumptions for a metapopulation (as discussed in section 1.2 of chapter 1): the sites occupied by aggregations are discrete, individuals only

occasionally move between these sites, and there is turnover of occupied sites (i.e. previously occupied sites are recolonized following local extinction). However, metapopulation models are only useful if model parameters can be estimated reliably. One of the key parameters is the probability of movement between aggregations. For large mammals, such as the grey seal, which spends more than 80% of its time at sea (McConnell *et al.* 1999) and 90% of this time below the surface (Thompson *et al.* 1991), directly observing movement is practically impossible. However, movement probabilities can be estimated using multisite capture-recapture techniques (Brownie *et al.* 1993).

Individual female grey seals can be identified by unique patterns of black markings on their white background colouration. The Sea Mammal Research Unit (SMRU) has collected a large catalogue of photographs of UK grey seals. Photographs of the same individual are identified with the aid of a computer matching program that uses markings on a specific area on the side of the seal's neck that is usually boldly patterned (Hiby & Lovell 1990).Although photographs have not been collected from all grey seal breeding colonies in all years, systematic samples were collected at four colonies in the North Sea (Isle of May, Fast Castle, Farne Islands and Donna Nook) between 1999 and 2001.

For statistical modelling of multisite capture-recapture data the Arnason-Schwarz (AS) model (Arnason 1972, 1973; Schwarz, Schweigert & Arnason 1993) provides a good starting point. The AS model is the multisite equivalent of the familiar single site Cormack-Jolly-Seber (CJS) model (Cormack 1964, Jolly 1965, Seber 1965). Maximum likelihood estimates (MLEs) of the parameters for these models can be easily obtained using the iterative methods utilized in software programs such as *MARK* (White & Burnham 1999) and *M-SURGE* (Choquet *et al.* 2004).The AS model is highly parameterised, and it may not be possible to estimate all of these parameters given the information content in the data. However, reduced-parameter versions of this model are available (Brownie *et al.* 1993) and a suitable model can be chosen using standard model selection procedures (Lebreton *et al.* 1992).

For the single site case Lebreton *et al.* (1992) recommend the following procedure for model building and model selection for capture-recapture data: (1) select the most

general model and assess its fit; (2) select a more parsimonious model using Akaike's Information Criterion (AIC); (3) test biological questions (such as whether external factors affect the probabilities in the model); and (4) obtain MLEs for the model parameters. This procedure can also be applied to the multisite case, especially in light of recent advances in assessing the goodness-of-fit (GOF) of the AS model (Pradel, Wintrebert & Giminez 2003) using the program *U-CARE* (Choquet *et al.* 2003).

MARK and *M-SURGE* utilize a generalised linear model (GLM) framework, so that it is possible to relate variation in demographic rates to environmental variables (Buckland, Goudie & Borchers 2000). This has been referred to as ultrastructural modelling (Spendelow *et al.* 1995).

In this study, the model selection procedures described above were used to select the most parsimonious models to describe capture-recapture data collected from female grey seals breeding at the North Sea colonies between 1999 and 2001. I tested the hypotheses that the probability of movement between colonies was related to the distance between them and their population sizes.

4.2 Methods

4.2.1 Study animals, sampling occasions and locations

The locations of the four colonies are shown in Figure 4.1. The colonies are the Isle of May, Fast Castle, the Farne Islands and Donna Nook; these will occasionally be indexed 1 to 4 in what follows. Samples were taken at all four colonies between 1999 and 2001. The main visits to each colony occurred on one or two occasions each year between 26 October and 10 December. As many adult females as possible were photographed either on land or in the water using a Canon EOS 505si 35mm camera fitted with a 600mm lens and x1.4 converter xp2 film. The photographs were taken by members of the Sea Mammal Research Unit at St Andrews University, the majority by Rob Harris.

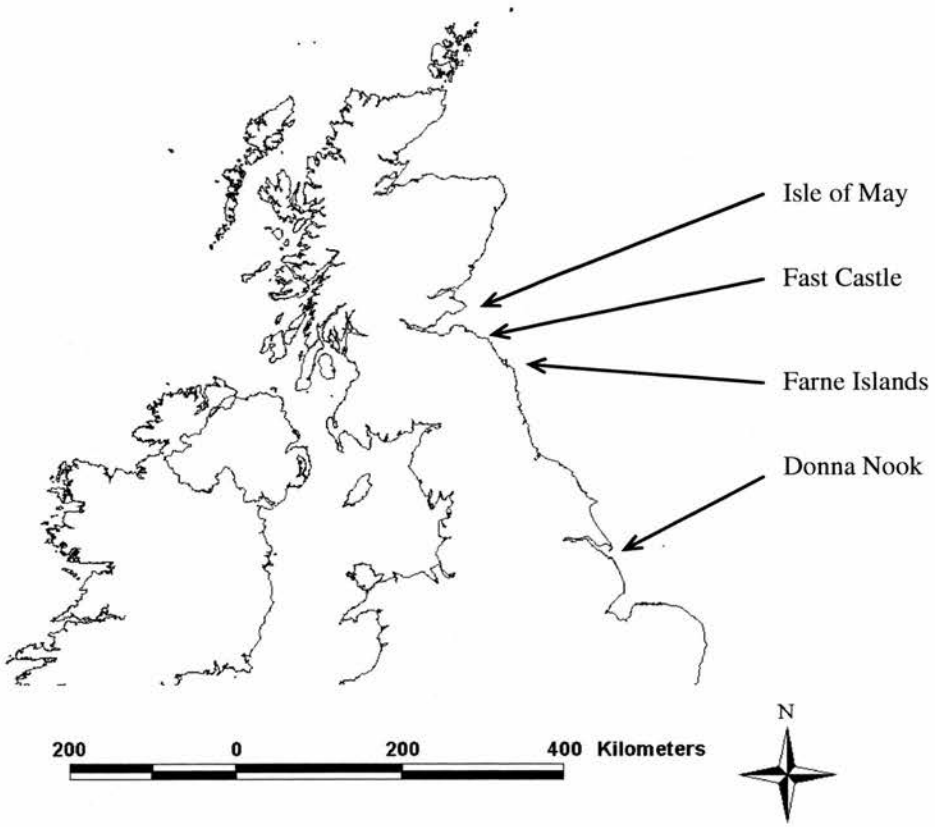


Figure 4.1. Map of the UK showing the four North Sea grey seal breeding colonies used in the capture-recapture study

Females were photographed from the left hand side (LHS) or the right hand side (RHS). In the data sets used for analysis there are 1122 female grey seals known by the LHS and 1181 known by the RHS. In the field it was sometimes possible to match the LHS and RHS photographs to the same animal – in the data analysed 358 animals are known by both sides. I assumed that all the photographed females were breeding at the colonies where they were photographed.

4.2.2 Recapture detection

The pelage on the side of an adult female grey seal's head and neck is boldly patterned and can act as a "fingerprint", thus enabling a seal to be recognised

whenever subsequently seen from that side. It is believed that these patterns are constant over time (Hiby & Lovell 1990).

Photographs were digitized and then analyzed using image-processing software developed by Hiby & Lovell (1990). Decisions on which photographs to put through the software were made based on the exposure, image size on the negative, the amount of blurring (due to camera or seal movement) and the distinctness of the pelage pattern (John Watkins, pers. comm.). This software locates a specific area on the side of the seal's neck and uses a three dimensional model of the seal's head to derive a numerical description of this area (see section 1.4 of chapter 1). This description is then compared to all others in the computer library and pairs of photographs that score above a pre-set similarity threshold are presented for final comparison by eye. This method is robust to changes in orientation and position of the seal, average brightness and contrast (Hiby & Lovell 1990; Hiby & Hammond 1998).

4.2.3 Capture histories

The LHS and RHS data were analyzed separately – to combine the two data sets would require the development of new models with non-standard parameters, such as the probability of matching the two sides of a seal's head whilst taking photographs at a colony. Parameter estimates from these two analyses are not, however, completely independent, because an unknown proportion of animals in one data set are also represented in the other. The input data for the capture-recapture software are coded as "capture histories", for example the capture history "302" represents a seal photographed at colony 3 in the first year, not seen at any colony in the second year and re-photographed at colony 2 in the third year.

4.2.4 Observation error

One of the assumptions of capture-recapture models is that marked animals are identified with certainty when they are recaptured. However, when photographic

information is used to uniquely identify individuals there is a chance that previously photographed animals may not be recognized when they are photographed again. That is, individual capture histories may contain false negatives. The user of the image processing software has some control of this process because final matches are made by the user. The conventional approach is to attempt to minimize the number of false positive matches (where two different animals are thought to represent the same animal). However, the more conservative this step, the higher the chances of making false negative errors (Stevick *et al.* 2001). In addition, the probability of generating a false negative record decreases over time as more photographs of the same animal are added to the library and are available for comparison.

4.2.5 Data screening

Seven seals (4 from the 1122 known by the LHS and 3 from the 1181 known by the RHS) were photographed at more than one colony in the same year, indicating that within-season movement does occur. Each of these animals was allocated to one breeding colony based on the relative timing of each photograph relative to the start and end of the pupping season and the duration of lactation. For instance one of the LHS seals seen at two colonies in one year was photographed at the Isle of May on 26th October 2001 and then again at Fast Castle on 3rd November 2001. This female was allocated to Fast Castle as the peak of pupping occurs at Fast Castle in November and at the Isle of May in late October. As lactation lasts approximately 3 weeks it is unlikely that she would have had time to have bred at the Isle of May and to have finished lactation by 3rd November when she was photographed at Fast Castle.

4.2.6 Models, parameters and assumptions

The full model considered for this analysis is the time-dependent conditional AS model (CAS - Choquet *et al.* 2004), which does not estimate population size. The model parameters in this analysis are:

$\Phi_{i,t}$ = the survival probability: The probability that a seal alive on colony i in year t–1 survives (and does not permanently leave the four North Sea colonies) until year t, where $i = 1,\ldots,4$ and $t = 2000$ or 2001.

$\Psi_{i\to j,t}$ = the movement probability: The probability that a seal (conditional on surviving) moves from colony i in year t–1 to colony j in year t.

$P_{j,t}$ = the recapture probability: The probability that a seal photographed in a previous year and alive on colony j at time t, is re-photographed there.

In this parameterisation I have used the time step t–1 to t, instead of t to t+1, which is conventionally used in the capture-recapture literature (e.g. Hestbeck, Nichols & Malecki 1991; Brownie *et al.* 1993; Nichols & Kendall 1995), for consistency with the state-space modelling described in chapter 5.

Using the parameterisation above; the probability that a seal photographed in 1999 at colony i, survives until 2000, moves to colony j and is identified there in 2000 is $\Phi_{i,2}\Psi_{i\to j,2}P_{j,2}$. Hence, the CAS model assumes that an animal's survival from t–1 to t depends only on the colony it occupies at time t–1. Movement between colonies is assumed to be a first-order Markov process (i.e. the probability of movement between t–1 and t depends only on the location at time t–1 and not before). Models that relax this assumption have been explored in the literature (see Hestbeck *et al.* 1991; Brownie *et al.* 1993) but, due to the increase in the number of parameters, require a much larger amount of data than was available for this analysis.

Further assumptions of the CAS model are: independence of individuals; identity of individuals (all share the same probabilities); independence of successive capture events (probabilities are independent of previous capture histories). The validity of some of these assumptions can be tested using GOF tests.

Female grey seals are long lived (35-40 years) so most animals were expected to survive the entire study period. It was therefore unlikely that it would be possible to obtain precise estimates of survival probabilities from this data set (Lebreton *et al.* 1992). The survival rate, Φ, was therefore fixed at 0.96 – an estimate based on an extended study of a group of branded and tagged adult females at the Isle of May

(Paddy Pomeroy, pers. comm.). The sensitivity of the analysis to uncertainty in this parameter estimate was checked by repeating the analysis using the lower and upper 95% confidence limits (0.896 and 0.985 respectively).

4.2.7 Maximum likelihood estimation

Summary information required for parameter estimation for the CAS model can be displayed in a multisite m-array (Brownie *et al.* 1993) such as Table 4.1, where:

$R_{i,t}$ = the number of seals uniquely identified on colony i at time t.

$m_{ij,tu}$ = the number of seals identified on colony i at time t first recaptured at time u on site j.

$r_{i,t} = \sum_j \sum_u m_{ij,tu}$ (i.e. the total number of first recaptures for the entire study at all colonies given release on colony i at time t).

Table 4.1. Multisite m-array representation of summary information required for parameter estimation for the Arnason-Schwarz model with three capture occasions and four colonies

		Number of first recaptures by year (and colony)								
Release year, colony	Number released	2				3				Not seen again
		(1	2	3	4)	(1	2	3	4)	
1, 1	$R_{1,1}$	$m_{11,12}$	$m_{12,12}$	$m_{13,12}$	$m_{14,12}$	$m_{11,13}$	$m_{12,13}$	$m_{13,13}$	$m_{14,13}$	$R_{1,1} - r_{1,1}$
1, 2	$R_{2,1}$	$m_{21,12}$	$m_{22,12}$	$m_{23,12}$	$m_{24,12}$	$m_{21,13}$	$m_{22,13}$	$m_{23,13}$	$m_{24,13}$	$R_{2,1} - r_{2,1}$
1, 3	$R_{3,1}$	$m_{31,12}$	$m_{32,12}$	$m_{33,12}$	$m_{34,12}$	$m_{31,13}$	$m_{32,13}$	$m_{33,13}$	$m_{34,13}$	$R_{3,1} - r_{3,1}$
1, 4	$R_{4,1}$	$m_{41,12}$	$m_{42,12}$	$m_{43,12}$	$m_{44,12}$	$m_{41,13}$	$m_{42,13}$	$m_{43,13}$	$m_{44,13}$	$R_{4,1} - r_{4,1}$
2, 1	$R_{1,2}$					$m_{11,23}$	$m_{12,23}$	$m_{13,23}$	$m_{14,23}$	$R_{1,2} - r_{1,2}$
2, 2	$R_{2,2}$					$m_{21,23}$	$m_{22,23}$	$m_{23,23}$	$m_{24,23}$	$R_{2,2} - r_{2,2}$
2, 3	$R_{3,2}$					$m_{31,23}$	$m_{32,23}$	$m_{33,23}$	$m_{34,23}$	$R_{3,2} - r_{3,2}$
2, 4	$R_{4,2}$					$m_{41,23}$	$m_{42,23}$	$m_{43,23}$	$m_{44,23}$	$R_{4,2} - r_{4,2}$

Conditional on the number of animals released, the values in each row of the m-array are mutually exclusive and are distributed multinomially. Programs like *M-SURGE* utilize iterative methods to calculate MLEs of the parameter values. The MLEs are the

parameter values that bring the observed and expected numbers in the m-array as close as possible given the flexibility in the model. This is known as minimizing the Deviance (which is calculated as minus two multiplied by the natural logarithm of the likelihood for the model). *M-SURGE* was used in the present analysis as it has been designed specifically for multisite capture-recapture data analysis.

Using the CAS model with survival fixed to 0.96, the individuals in the m-array that make up $m_{32,13}$ each have the capture history "302". These individuals may have visited any of the four colonies in the year where the zero was recorded. The probability of an individual visiting colony 1 on occasion 2, for instance, is given by $0.96 * \Psi_{3\to1,2} Q_{1,2} * 0.96 * \Psi_{1\to2,3} P_{2,3}$, where $Q_{1,2} = 1 - P_{1,2}$ (i.e. the seal released on colony 3 on occasion 1 survived the following time period, moved from colony 3 to colony 1, was not seen there, survived the next time period, moved from colony 1 to colony 2 and was seen on colony 2). Accounting for all the locations possible on occasion 2 the expected number in this cell of the m-array is thus given by:

$$E(m_{32,13}) = R_{3,1}[0.96^2\{(\Psi_{3\to1,2}Q_{1,2}\Psi_{1\to2,3}P_{2,3}) + (\Psi_{3\to2,2}Q_{2,2}\Psi_{2\to2,3}P_{2,3}) + (\Psi_{3\to3,2}Q_{3,2}\Psi_{3\to2,3}P_{2,3}) + (\Psi_{3\to4,2}Q_{4,2}\Psi_{4\to2,3}P_{2,3})\}] \qquad \text{(eqn 4.1)}$$

For the movement parameters in the above model the site fidelity parameter, $\Psi_{i\to i,t}$, i.e. the probability of returning to the colony that was last bred at, is post-processed as:

$$\Psi_{i\to i,t} = 1 - \sum_{j\neq i} \Psi_{i\to j,t} \qquad \text{(eqn 4.2)}$$

where i = 1,..,4.

Similar probabilistic statements can be written for all the other cells in the table. Note that those released in the m-array in the second year of the study are assumed to be a combination of those previously caught and those newly caught. This assumes that capture history does not affect capture probability, an assumption that can be validated with GOF tests.

4.2.8 Goodness of fit and variance inflation factors

Lebreton *et al.* (1992) and Burnham *et al.* (1987) recommend that the fit of the most general model being explored is evaluated first. Until recently, however, assessing the fit of multisite models was not straightforward and there was no standard method for doing this. However, Pradel *et al.* (2003) have developed methods for testing the assumption that all animals released at a given time behave the same irrespective of their past capture history and for comparing the fate of animals photographed at a particular time with those photographed at other times. These tests can be performed using the companion program to *M-SURGE* known as *U-CARE*. Pradel *et al.* (2003) recommend that these tests be utilised to obtain a variance inflation factor, $\hat{c}$, which should subsequently be used in model selection and variance estimation (Choquet *et al.* 2004).

As there are only three years of data in the present study only one of Pradel *et al*'s tests (test 3GSR) could be applied. This is a χ^2 test for transients, an age effect or a marking effect. It tests the assumption that all animals present at any given time on the same site behave the same regardless of their past capture history. Even this test could only be applied to the 2000 data because the animals must have been caught both before and after the occasion fixed on.

A reliable measure of lack of fit in this case is to use the ratio χ^2/df, known as $\hat{c}$, for the GOF tests that can be performed. A value of one for $\hat{c}$ is what is expected if the model assumptions have not been violated. When $\hat{c}$ is greater than one it can be an indication of over-dispersion in the data or that some of the structural assumptions of

the model are wrong (Lebreton *et al.* 1992). In this case $\hat{c}$ was calculated as $\chi^2_{3GSR} / df_{3GSR}$.

4.2.9 Model selection

Lebreton *et al.* (1992) recommend that the fit of all other models is compared to that of the general model using Akaike's information criterion (AIC) or quasi-likelihood adjusted AIC (QAIC), which is the AIC adjusted using $\hat{c}$ (Burnham & Anderson 2002), such that:

$$\mathrm{QAIC}_m = \frac{Dev_m}{\hat{c}} + 2p_m \qquad \text{(eqn 4.3)}$$

where the QAIC for model m depends on its deviance, Dev_m, the variance-inflation factor, $\hat{c}$, and the number of parameters it has, p_m. This model selection procedure makes a trade off between bias and precision based on the information content of the data. As more complicated models will fit the data better (because they are more flexible); they are penalised for the extra parameters that need to be estimated. Furthermore, when $\hat{c}$ is greater than unity, QAIC favours models with fewer parameters.

I also used the QAIC weights in model selection. This is essentially the Akaike weight that incorporates the variance inflation factor described above. This weight can be thought of as the probability for each model being the "best" approximating model from the models compared (Buckland *et al.* 1997; Burnham & Anderson 2002; Thomas & Harwood 2003; Cam *et al.* 2004; Johnson & Omland 2004) To calculate this weight first the difference between each model in the set and the "best" model (i.e. the model with the minimum QAIC value) are calculated

$$\Delta_m = \mathrm{QAIC}_m - \mathrm{QAIC}_{\min} \qquad \text{(eqn 4.4)}$$

Then the relative weight for each model, W_m, is calculated as

$$W_m = \frac{\exp(-\Delta_m / 2)}{\sum_q \exp(-\Delta_q / 2)} \qquad \text{(eqn 4.5)}$$

where $q = 1,\ldots, Q$ and Q represents the number of models being compared.

To limit the number of models explored, and to focus on the hypotheses of most interest, I followed this route through model space: 1) start from the CAS model with survival fixed to 0.96; 2) explore the effect of time on the movement and recapture probabilities; 3) reduce the number of recapture parameters and test biological hypotheses relating to recapture; 4) repeat step 3) for the movement models. I report only those models with a QAIC weight greater than 0.05.

4.2.10 Testing biological hypotheses

One of the aims of model selection is to test biological hypotheses such as the role of environmental covariates in explaining differences in parameter values (Buckland *et al.* 2000; Lebreton *et al.* 1992; Nichols & Kendall 1995; Spendelow *et al.* 1995). *M-SURGE* utilises a GLM framework, so such hypotheses can be easily tested (Choquet *et al.* 2004).

Often the recapture probabilities in capture-recapture analyses are thought of as nuisance parameters. They can, however, provide interesting insights into the processes that generated the data. The first hypothesis I tested was whether P was related to the number of marked seals divided by the known number of females breeding on a colony in a particular year ($Z_{i,t}$), where

$$Z_{i,t} = \frac{m_{6+,i,t}}{y_{0,i,t}} \qquad \text{(eqn 4.6)}$$

such that $m_{6+,i,t}$ represents the number of adult female seals (by assumption seals of six years and older) uniquely identified on colony i in year t, and $y_{0,i,t}$ is the pup production estimate on colony i in year t. Pup production estimates are obtained every year for all the major grey seal breeding colonies around Great Britain using counts of white-coated pups obtained from aerial photographs or direct census (SCOS 2004). The coefficient of variation of these estimates is generally low (Callan Duck, pers. comm.), so they should be a close approximation to the number of females present at the colony in each year as grey seals can only give birth to one pup.

I tested this hypothesis using a linear-logistic regression:

$$P_{i,t} = \frac{e^{\xi_0 + \xi_1 Z_{i,t}}}{\left(1 + e^{\xi_0 + \xi_1 Z_{i,t}}\right)} \qquad \text{(eqn 4.7)}$$

where ξ_0 and ξ_1 are intercept and slope parameters. The recapture parameters are replaced with the regression model in the multisite m-array and the ξ parameters are estimated directly (Nichols & Kendall 1995) by maximising the likelihood with respect to ξ_0 and ξ_1 . I expected that ξ_1 would be positive – as $Z_{i,t}$ increases, so too should $P_{i,t}$. For compactness the logistic models explored in this thesis will subsequently be described on the logit scale (Cam *et al.* 2004). The above model being described as:

$$\text{logit}\left(P_{i,t}\right) = \ln\left(\frac{P_{i,t}}{1 - P_{i,t}}\right) = \xi_0 + \xi_1 Z_{i,t} \qquad \text{(eqn 4.8)}$$

Then I tested the hypothesis that the probability of moving between colonies is related to the distance separating them using the linear-logistic model:

$$\text{logit}\left(\Psi_{i \to j,t}\right) = \alpha_0 + \frac{\alpha_1}{dist_{i \to j}} \qquad \text{(eqn 4.9)}$$

where $dist_{i\to j}$ is the distance between colony i and j in kms, calculated "as the seal swims". Taking the reciprocal with respect to the distance between colonies assumes that small changes in distance are more important between close colonies than between those that are further apart. For the above regression model, if the hypothesis of a decreased probability of movement with increased distance is correct the slope parameter α_1 should be positive. If α_1 equals one then one can infer that movement probabilities are independent of distance. Note that there is no time dependence in the regression equation and moves are assumed to be symmetrical (i.e. the probability of moving from colony i to j is identical to the probability of moving from colony j to colony i).

The distance model was then extended to include the difference in abundance between the two colonies such that:

$$\text{logit}\left(\Psi_{i\to j,t}\right) = \alpha_0 + \frac{\alpha_1}{dist_{i\to j}} + \alpha_2\left(y_{0,i,t-1} - y_{0,j,t-1}\right) \qquad \text{(eqn 4.10)}$$

The difference between the pup production estimates was used rather than their ratio, because it was thought to have better distributional properties for use on the logit scale. In the above model α_2 greater than zero indicates that seals avoid colonies larger than the one they last bred at, whereas α_2 less than zero indicates an attraction towards larger colonies. Note that there is time dependence in this regression equation and symmetric moves are no longer equivalent.

4.2.11 Numerical and statistical issues

As multisite capture-recapture models are relatively new, there are a number of statistical and numerical issues that can cause problems, some of which are not presently well understood. According to Lebreton & Pradel (2002), the three main and recurring problems are:

1. Parameter redundancy and identifiability issues. There are two types of parameter redundancy – "intrinsic" due to the structure of the model (such as the last survival and recapture parameters in the CJS model) and "extrinsic" due to limitations in the data (e.g. could be caused by a capture occasion on which no animals were recaptured);
2. Boundary parameters and the number of estimable parameters in a model. Boundary parameters (parameters estimated at either the zero or one boundary) cause irregularities in the likelihood surface, which are not well understood. Nevertheless it is advisable to include these as estimable parameters for model selection – this is preferred so that model selection by AIC will not favour models with several boundary parameters (Viallefont *et al.* 1998);
3. Local minima in the likelihood surface. This could mean that the estimates produced by the model are not the MLEs. This can be assessed automatically in *M-SURGE* by changing the initial values for the parameters (during the estimation stage) and doing several runs of the same model – if all of these model runs converge to the same deviance value, then there are no local minima in the likelihood surface.

4.3 Results

The raw capture data are presented as multisite m-arrays in Table 4.2. It is clear that the recapture rate is low and the site fidelity is high (the highest numbers within the m-array are for animals returning to the colonies where they were previously photographed). Furthermore, there are no movements to or from Donna Nook.

Table 4.2. Multisite m-arrays giving the number of animals photographed on their left-hand side (LHS) and right-hand side (RHS)

LHS		Number of first recaptures by year (and colony)								
Release year, colony	Number released	2 (1	2	3	4)	3 (1	2	3	4)	Not seen again
1, 1	230	7	1	1	0	10	1	0	0	211
1, 2	55	1	5	1	0	1	6	0	0	41
1, 3	192	1	1	21	0	1	2	9	0	157
1, 4	63	0	0	0	5	0	0	0	4	54
2, 1	252					22	0	0	0	230
2, 2	93					2	15	0	0	76
2, 3	180					0	0	12	0	168
2, 4	101					0	0	0	6	95

RHS		Number of first recaptures by year (and colony)								
Release year, colony	Number released	2 (1	2	3	4)	3 (1	2	3	4)	Not seen again
1, 1	259	13	0	0	0	7	2	0	0	237
1, 2	54	1	3	0	0	0	3	0	0	47
1, 3	228	0	0	28	0	0	3	8	0	189
1, 4	63	0	0	0	5	0	0	0	1	57
2, 1	272					28	1	0	0	243
2, 2	65					1	8	0	0	56
2, 3	179					1	0	12	0	166
2, 4	111					0	0	0	4	107

Test 3GSR showed some evidence of lack of fit to the data. $\hat{c}$ was therefore used as a precautionary measure to make the model selection and inference as robust as possible to potential failure of the assumptions. As it is difficult to conceive of reasons why the LHS data should be different from the RHS data, the average value of $\hat{c}$ from both analyses (2.79) was used. This value is not particularly high for capture-recapture models exhibiting over-dispersion (Lebreton *et al.* 1992). The tests results from the LHS and RHS data varied among colonies suggesting that there may be extra multinomial "noise", as opposed to structural error.

The results from the model selection procedure are shown in Table 4.3 for both LHS and RHS data. The estimated logistic regression model parameters are given in Table 4.4. The estimate of the slope parameters ξ_1 for the recapture rate regressions for the LHS models were positive, indicating that the recapture probability increased as the proportion of breeding seals that were photographed increased. Similarly, the estimate of the slope parameter α_1 was also positive for the models that included distance

effects on movement (L1, L2, R1 and R2), indicating that movement probabilities decreased with increasing distance between colonies. However, for the models that incorporated the effect of both distance and abundance on movement (L1 and R2) the estimate of the slope parameter α_2 was negative, indicating that seals were attracted towards colonies larger than the one they previously bred at. For both the LHS and RHS data there was also some support for a model with just a single parameter estimated for movement; model L3 and R3 both had a QAIC weight greater than 0.15.

Table 4.3 Number of parameters, QAIC and QAIC weight for the capture-recapture models for the left-hand side (LHS) and right-hand side (RHS) data. Only the models with a QAIC weight greater than 0.05 are shown. Model descriptions are given in the footnote. Models are indexed L1 to L3 for the LHS and R1 to R3 for the RHS and are ordered according to their QAIC weights.

LHS

Model	Model parameters	No. of params	QAIC	QAIC weight
L1	$\Phi_{(0.96)}\Psi_{(\text{dist \& abun})}P_{(Z)}$	5	374.687	0.586
L2	$\Phi_{(0.96)}\Psi_{(\text{dist})}P_{(Z)}$	4	376.391	0.250
L3	$\Phi_{(0.96)}\Psi_{(\text{int})}P_{(Z)}$	3	377.373	0.153

RHS

Model	Model parameters	No. of params	QAIC	QAIC weight
R1	$\Phi_{(0.96)}\Psi_{(\text{dist})}P_{(\text{int})}$	3	362.335	0.570
R2	$\Phi_{(0.96)}\Psi_{(\text{dist \& abun})}P_{(\text{int})}$	4	364.194	0.225
R3	$\Phi_{(0.96)}\Psi_{(\text{int})}P_{(\text{int})}$	2	364.530	0.190

QAIC: Quasi-likelihood adjusted Akaike's information criterion (see text)

Model parameters: Φ = survival probability; Ψ = movement probability; P = Recapture probability

Model specification: int = intercept (single parameter estimated); dist = movement probability constrained to be a function distance between colonies (see eqn 4.9); dist & abun = movement probability constrained to be a function of distance and abundance (see eqn 4.10); Z = recapture probability constrained to be a function of $Z_{i,t}$ (see eqn 4.8).

Table 4.4 Logistic regression model parameter estimates for the capture-recapture models given in Table 4.3. The mean, 2.5th and 97.5th percentiles are given for each parameter

	Logistic regression model parameters				
	Recapture parameters		Movement parameters		
Model	ξ_0	ξ_1	α_0	α_1	α_2
L1	-3.39	6.72	-5.63	114	-0.0008
	(-3.95, -2.84)	(3.28, 10.2)	(-7.24, -4.02)	(39.8, 188)	(-0.0013, -0.0003)
L2	-2.99	3.91	-5.38	96.3	not
	(-3.42, -2.55)	(1.49, 6.33)	(-6.92, -3.85)	(24.6, 168)	applicable
L3	-3.06	4.44	-3.78	not	not
	(-3.49, -2.64)	(2.10, 6.79)	(-4.34, -3.23)	applicable	applicable
R1	-2.49	not	-6.42	145	not
	(-2.67, -2.31)	applicable	(-8.63, -4.20)	(47.8, 242)	applicable
R2	-2.49	not	-6.44	149	-0.0002
	(-2.67, -2.31)	applicable	(-8.68, -4.20)	(50.5, 248)	(-0.0006, 0.0003)
R3	-2.49	not	-3.89	not	not
	(-2.67, -2.31)	applicable	(-4.54, -3.25)	applicable	applicable

Table 4.5 gives a comparison of the movement probabilities, Ψ, for the "best" models from the LHS and RHS analyses (i.e. the models with the highest QAIC weights). The site fidelity parameters are all very high. The movement probability estimates for the two models appear dissimilar, but their CI's overlap. For the LHS model L1 there is a great deal of uncertainty in the movement probabilities associated with Fast Castle.

Re-running the models with the upper and lower bounds on survival had a relatively small effect on the movement probabilities. Recapture probabilities, however, were sensitive. A comparison of the recapture probabilities, $P_{i,t}$, for the models L1 with the covariate $Z_{i,t}$ showed that the estimates of $P_{i,t}$ were significantly lower than the values of $Z_{i,t}$. This was an unexpected result; it was hypothesised that these two measures would be quite similar. The effect of re-running these models with the lower survival rate of 0.896 caused an increase in the $P_{i,t}$ estimates, but these estimates were still consistently lower than those predicted by the external covariate – see Table 4.6.

Table 4.5. A comparison of the movement probabilities from year $t-1$ to year t for the two "best" models selected by QAIC (models L1 and R1). The 95% CIs are given along with the parameter estimates. (Colony 1=Isle of May, colony 2 = Fast Castle, colony 3 = Farne Islands, colony 4 = Donna Nook):

Movement between colonies	**LHS** Probability with CI Model L1 t = 2000	Probability with CI Model L1 t = 2001	**RHS** Probability with CI Model R1 t = 2000 & 2001 (i.e. no time dependence)
1→1	0.971 (0.820, 0.995)	0.975 (0.827, 0.997)	0.916 (0.643, 0.979)
2→2	0.745 (0.257, 0.936)	0.701 (0.146, 0.932)	0.903 (0.630, 0.974)
3→3	0.951 (0.794, 0.988)	0.954 (0.797, 0.989)	0.967 (0.806, 0.993)
4→4	0.975 (0.754, 0.998)	0.970 (0.704, 0.997)	0.992 (0.818, 0.9997)
1→2	0.021 (0.003, 0.119)	0.017 (0.002, 0.115)	0.073 (0.019, 0.236)
1→3	0.006 (0.001, 0.035)	0.006 (0.001, 0.034)	0.009 (0.001, 0.060)
1→4	0.002 (0.0001, 0.026)	0.001 (0.0001, 0.024)	0.002 (0.0001, 0.061)
2→1	0.205 (0.047, 0.577)	0.242 (0.049, 0.663)	0.073 (0.019, 0.236)
2→3	0.043 (0.017, 0.107)	0.051 (0.018, 0.132)	0.022 (0.006, 0.073)
2→4	0.006 (0.0006, 0.059)	0.006 (0.0006, 0.059)	0.003 (0.0001, 0.061)
3→1	0.028 (0.006, 0.113)	0.029 (0.006, 0.118)	0.009 (0.001, 0.060)
3→2	0.017 (0.005, 0.052)	0.014 (0.004, 0.049)	0.022 (0.006, 0.073)
3→4	0.004 (0.0004, 0.041)	0.003 (0.0003, 0.036)	0.003 (0.0001, 0.060)
4→1	0.014 (0.001, 0.140)	0.017 (0.001, 0.178)	0.002 (0.0001, 0.061)
4→2	0.004 (0.0004, 0.043)	0.004 (0.0004, 0.043)	0.003 (0.0001, 0.061)
4→3	0.007 (0.0007, 0.064)	0.008 (0.0009, 0.075)	0.003 (0.0001, 0.060)

Table 4.6. A comparison of the recapture probabilities, $P_{i,t}$ (with their 95% CIs), for the model L1, and the covariate $Z_{i,t}$. The recapture estimates for the same model run with the lower survival rate of 0.896 are also given.

LHS: model L1

Colony and year	$Z_{i,t}$	P estimated with CI Φ=0.96	P estimated with CI Φ=0.896
Isle of May 2000	0.118	0.069 (0.049, 0.097)	0.076 (0.054, 0.106)
Isle of May 2001	0.134	0.076 (0.056, 0.103)	0.084 (0.062, 0.112)
Fast Castle 2000	0.244	0.148 (0.086, 0.242)	0.160 (0.094, 0.260)
Fast Castle 2001	0.324	0.229 (0.096, 0.455)	0.246 (0.104, 0.478)
Farne Islands 2000	0.154	0.086 (0.065, 0.114)	0.095 (0.071, 0.125)
Farne Islands 2001	0.076	0.053 (0.032, 0.087)	0.058 (0.035, 0.096)
Donna Nook 2000	0.163	0.091 (0.068, 0.121)	0.100 (0.075, 0.132)
Donna Nook 2001	0.161	0.090 (0.068, 0.119)	0.099 (0.074, 0.130)

4.4 Discussion

Concerning the numerical and statistical issues raised by Lebreton & Pradel (2002): There was no parameter redundancy or boundary parameters in the models shown in Table 4.3. However, there were boundary parameters for the more parameterised models that are not shown in this table, specifically for the movement probabilities associated with Donna Nook. This is due to the fact that no movements to or from this colony were present in the data sets analysed. There were no local minima detected in the likelihood surfaces for the models compared – multiple runs of each model with randomly selected initial values all converged to the same Deviance values.

Overall the results indicate that female grey seals showed a high degree of fidelity to the four North Sea breeding colonies between 1999 and 2001, supporting Pomeroy *et al*'s (1994) conclusions from the North Rona colony. The highest fidelity was for Donna Nook, which is relatively isolated from the other three colonies. In fact, no movements either into or out of Donna Nook were detected during the study period. However, such moves do occur: photographs were also taken at some colonies in 1998, incorporating these data into the analysis showed that one female moved from Donna Nook to the Farne Islands.

4.4.1 Covariates affecting movement

For both the LHS and RHS analyses, distance between colonies was an important covariate affecting the movement probability. It has long been held that distances between subpopulations are an important factor in animal movement models (MacArthur & Wilson 1967) and they have been said to influence movement probabilities in both theoretical and empirical studies (Serrano & Tella 2003).

For model L1 (the model with the highest QAIC weight for the LHS data analysis) there was also some evidence that colony size affected movement probabilities; where the movement probability was higher from the smaller colonies to the larger ones. This may have been a result of conspecific attraction (Smith & Peacock 1990; Serrano

& Tella 2003; Cam *et al.* 2004), or perhaps the larger colonies have more space and/or better conditions. One possible basis for conspecific attraction is avoidance of mate shortage (Hanski 1994). As grey seals mate after having lactated their pup, a small colony of breeding females may be less likely to attract a large number of prospecting males, and hence lower a female's fecundity for the following year.

However, the abundance effect observed could alternatively have been an artefact caused by the fact that Fast Castle, the smallest and most recently established colony, had a much lower fidelity rate than any other colony. This view is supported by the fact that the movement probabilities associated with Fast Castle had wide confidence limits and means of the movement estimates were highly sensitive to the choice of model.

It was initially thought that animals would leave the larger colonies in favour of the smaller ones due to intraspecific competition and/or habitat saturation (Serrano & Tella 2003). However, it could be that as all four North Sea colonies are still growing such saturation levels have not yet been reached.

4.4.2 Modelling considerations

Bayesian methods provide an alternative to the maximum likelihood framework used to analyse the multisite capture-recapture data in this chapter. Dupuis (1995) shows how the Arnason-Schwarz model can be implemented in a Bayesian context and King & Brooks (2002) show how this can be extended to perform model comparisons. The benefit of reframing the problem as a Bayesian one is that prior information can be brought into the models before fitting to the capture-recapture data. Use of a Bayesian approach in this particular analysis would have allowed a prior distribution on the survival probability to be specified rather than simply using a point estimate.

4.4.3 Within-season movements

There are some very important considerations to be taken into account when interpreting the estimated movement probabilities. The model assumes that all the females photographed at a specific colony during a specific year are breeding there. In fact, some females were seen at two colonies in one year, indicating that there is some within-season movements between colonies and that some females may not have bred at the colony to which they were attributed. Such females may have been photographed either when they were passing by a colony after breeding or when they were prospecting potential breeding sites. This effect is likely to result in the estimates of the movement rates being biased upwards, although the magnitude of that bias cannot be estimated at present.

4.4.4 False negative errors

There was support for the probability of recapture varying by colony and time as a function of the external covariate $Z_{i,t}$. However, the actual probabilities were significantly lower than those expected based on this covariate. Several factors may have contributed to this:

1. Within-season movements of breeding females could result in more females being present at a colony than those breeding there;
2. The population is growing and there is recruitment to the breeding population, thus some females that are breeding may not have been present to be photographed on a previous occasion;
3. The presence of non-breeding but recruited females at the colony. The pregnancy rate is curently estimated to be 90-95%, so the maximum bias resulting from this is likely to be small;
4. Juvenile males may have been mistaken for females;
5. The presence of false negatives (i.e. missing matches) in individual capture histories. The matching process is designed to minimize the probability of a

false positive match, so the probability of a false negative is relatively high and this would cause a downward bias on the recapture estimates;

6. Survival may have been lower in the years of the survey than the assumed value of 0.96. Using the lower 95% confidence limit for this estimate (0.896) did result in a small increase in the recapture estimates, but not enough to account for all the difference observed;
7. The pup production estimates are biased downwards. However, pup production estimates at the Farne Islands and Donna Nook are based on direct censuses, so this seems unlikely.

Of these seven possibilities, the fifth is the most likely: the current photograph matching protocol is very conservative and a high proportion of potential matches (possibly greater than 30%) are rejected.

One of the main objectives of this work was to obtain estimates of movement probabilities that could be used in a Bayesian state-space metapopulation model of the British grey seal population (Buckland *et al.* 2004; Thomas *et al.* 2005). This is described in chapter 5.

4.4.5 Future data collection

Collecting additional photo-ID data for the four North Sea colonies studied in this chapter would provide further information on movement rates between colonies and with a longer time series of data could allow the estimation of an independent survival rate for seals from these colonies. The fact that there will be a few years without any capture events is dealt with in the models by setting the recapture rate in these intervening years to zero.

However, there are a number of logistic and financial considerations that need to be considered before undertaking such an exercise. Appendix 1 details these considerations and compares different sampling protocols using simulation modelling. The main aim of these simulations was to determine the optimal sampling strategy to

collect a large number of photographs of different animals within a single breeding season. The models were parameterised with data on pup production at the Isle of May in 2003 and with other information on the breeding biology of this species.

Chapter 5

State-Space Modelling of Female Grey Seals Breeding at Colonies in the North Sea

5.1 Introduction

The grey seal (*Halichoerus grypus* Fabricius 1791) is a colonially breeding pinniped whose distribution is confined to the North Atlantic and its contiguous seas. It has a world population of around 300,000 individuals, 40% of whom breed around the British Isles (SCOS 2004). The total number of pups born at British colonies has grown steadily since the 1960s (Harwood & Prime 1978). Commercial fishermen are concerned about the impacts of seal predation on depleted fish stocks and conservation agencies about the effects of breeding seals on sensitive terrestrial habitats. It is therefore important to be able to predict the future size and spread of the British population. Individual female grey seals show high levels of fidelity to specific pupping locations at a single colony (Pomeroy *et al.* 1994). This may account for the highly significant differences in microsatellite allele frequency which have been observed between breeding colonies (Allen *et al.* 1995), and suggests that the British grey seal population cannot be modelled as if it is a single, homogenous unit. Instead, we require a spatially-explicit model which accounts for differences in colony dynamics, and the impact of movement between colonies on local and regional dynamics (Gaggiotti *et al.* 2004; Thomas *et al.* 2005; Matthiopoulos, Harwood &

Thomas 2005). Grey seal colonies meet all of the assumptions for a metapopulation (Hanski & Gaggiotti 2004): the sites occupied by aggregations are discrete, individuals only occasionally move between these sites, and there is turnover of occupied sites (i.e. previously occupied sites are recolonized following local extinction).

As with most spatially-explicit models, the most difficult problem experienced to date has been the validation and parameterisation of realistic models of movement between colonies. In the belief that much of the model and parameter uncertainty could be removed by bringing together the analysis of different data sets, I used input from a multisite capture-recapture (CR) study to parameterise spatially-explicit models of the dynamics of the same set of grey seal colonies.

It is clear that grey seal colonies meet the assumptions for a metapopulation (as discussed above). It is, however, unclear what level of realism should be incorporated into the movement process in metapopulation models. Hanski & Gaggiotti (2004) state that what is needed is a family of models that incorporate different levels of realism. In this chapter, I compare metapopulation models incorporating different movement processes in their ability to track a time series of pup production data.

It was thought (as discussed in section 1.5 of chapter 1) that the pup production data may contain some information on movement. For example, if the annual pup production increases faster than is possible as a result of internal recruitment of adult females into the breeding population, one may infer that net immigration must be taking place. In this chapter I formalize this idea by fitting different candidate state-space models (SSMs) for the dynamics of the metapopulation to the pup production data..

The state-space framework was introduced in section 1.6 of chapter 1. There are several benefits to using this modelling approach. For instance, one of the primary aims of state-space modelling is to explicitly model the observation process that is "always ... interposed between the ecological system and our notebooks" (Hilborn & Mangel 1997). SSMs can be applied to many ecological scenarios and wildlife populations. They can be used to assess the present status of an animal population or

its possible future status under a range of management options. They can potentially be fit to many different forms of data including count, capture-recapture, line transect and genetic data. They are very flexible and can incorporate realistic processes such as density dependence, competition, metapopulation dynamics and predator prey interactions. The fact that they take explicit account of all the major sources of uncertainty means that the results and inference they provide are much more reliable and give an honest account of the present state of understanding about the animal population in question.

SSMs are also fit to data using Bayesian inference (see section 1.6 of chapter 1) and thus results from previous analyses can be used to provide informative priors for the model parameters. In this chapter I have used parameters estimated from a CR study (see chapter 4) to provide the prior distributions for the movement process model parameters in this SSM analysis.

Thomas *et al.* (2005) developed a spatially-explicit SSM of the British grey seal population, which they divided into four regions: the North Sea (4 colonies), Inner Hebrides (19 colonies), Outer Hebrides (11 colonies) and the Orkneys (22 colonies). At this spatial scale, there was little evidence for movement of animals between local populations. Here, I develop colony-level SSMs of grey seal population dynamics in one of these regions (the North Sea) and use it to investigate the nature of movement at this scale. The SSMs were also projected into the future to investigate the consequences of the different model structures for the abundance of seals in the North Sea metapopulation at both the colony and regional level.

Bayesian modelling of metapopulation data has previously been done for patch occupancy data using Markov Chain Monte Carlo (MCMC) methods to estimate the model parameters (O'Hara *et al.* 2002; Ter Braak & Etienne 2003). Our method is different in two respects. Firstly, we use abundance data rather than patch occupancy data. Secondly, we fitted our models to the data using Sequential Importance Sampling (SIS) – an alternative Bayesian method to MCMC which is potentially better suited to analysing a time-series of abundance data (see section 5.2.6).

5.2 Methods

5.2.1 Data and model structure

The available data consisted of annual estimates of the pup production between 1984 and 2003 for four North Sea colonies derived from aerial surveys conducted at the Isle of May and Fast Castle by the Sea Mammal Research Unit, and ground censuses conducted at the Farne Islands and Donna Nook by staff of the National Trust and the Lincolnshire Wildlife Trust respectively (Duck 2004). These colonies are occasionally indexed 1-4 in what follows. There are no aerial survey results for Fast Castle prior to 1997. The locations of the four colonies are shown in Figure 4.1 in chapter 4.

Annual time steps began just after all pups had been born. We used individual age-classes for pups (age 0) and animals up to 5 year olds, and a composite class for adult seals ≥ 6 years. We assumed that only adult females move between colonies, because the capture-recapture data is for these animals. Allowing 5 year-olds (recruiting females) to move made no difference to the results. Age-class 0 included both male and female pups, but the other age-classes are for females only. Four processes (survival (s), movement (m), ageing (a) – which includes sex determination for pups – and breeding (b)) are modelled. The state vector $\mathbf{n}_t$ representing the state of the population at time t has 28 elements (4 colonies × 7 age classes). Only the elements for the first colony are shown below, the parts for the other three colonies being implied by the dots (…).

$$\mathbf{n}_t = \begin{bmatrix} n_{0,1,t} \\ n_{1,1,t} \\ n_{2,1,t} \\ n_{3,1,t} \\ n_{4,1,t} \\ n_{5,1,t} \\ n_{6+,1,t} \\ \mathrm{M} \end{bmatrix}$$

The first subscript is for age (0 to 6+), the second is for colony (1 to 4) and the third is for time. Uncertainty in the progression of animals from $\mathbf{n}_{t-1}$ to $\mathbf{n}_t$ is captured by a series of linked probability density functions (pdfs), in which the input to one pdf is the output from the previous pdf. This is akin to a "divide-and-conquer" strategy (Liu *et al.* 1998) where a complex problem is broken down into a series of connected and simpler structures:

$$\mathbf{u}_{s,t} \sim H_{s,t}\left(\mathbf{n}_{t-1}\right)$$
$$\mathbf{u}_{m,t} \sim H_{m,t}\left(\mathbf{u}_{s,t}\right)$$
$$\mathbf{u}_{a,t} \sim H_{a,t}\left(\mathbf{u}_{m,t}\right)$$
$$\mathbf{n}_t \sim H_{b,t}\left(\mathbf{u}_{a,t}\right)$$

where $H_{x,t}$ is the pdf for sub-process x and $\mathbf{u}_{x,t}$ is a realisation of the state vector at time t after sub-process x has occurred.

5.2.2 Full model: stochastic representation

The sub-process pdfs for the most fully-parameterised model (designated S1) used in this analysis, are given in detail below. Again only the elements for the first colony are shown, the other elements being implied by the dots (M).

Survival

$$\begin{bmatrix} u_{s,0,1,t} \sim \text{Binomial}\left(n_{0,1,t-1}, \phi_{p,1,t}\right) \\ u_{s,1,1,t} \sim \text{Binomial}\left(n_{1,1,t-1}, \phi_a\right) \\ u_{s,2,1,t} \sim \text{Binomial}\left(n_{2,1,t-1}, \phi_a\right) \\ u_{s,3,1,t} \sim \text{Binomial}\left(n_{3,1,t-1}, \phi_a\right) \\ u_{s,4,1,t} \sim \text{Binomial}\left(n_{4,1,t-1}, \phi_a\right) \\ u_{s,5,1,t} \sim \text{Binomial}\left(n_{5,1,t-1}, \phi_a\right) \\ u_{s,6+,1,t} \sim \text{Binomial}\left(n_{6+,1,t-1}, \phi_a\right) \\ \text{M} \end{bmatrix}$$

In this survival vector ϕ_a represents adult survival at colony i (animals of age 1 and older are assumed to have adult survival). Pup survival, $\phi_{p,i,t}$, is where density dependence is thought most likely to operate for the British grey seal population. I therefore model pup survival as a function of $n_{0,i,t-1}$, the number of pups (of both sexes) born on that colony in year $t-1$, using the Beverton-Holt function such that:

$$\phi_{p,i,t} = \frac{\phi_{pmax}}{1+\beta_i n_{0,i,t-1}} \qquad \text{(eqn 5.1)}$$

where $\phi_{p\max}$ represents the pup survival rate when there are few pups present. The equilibrium pup production at colony i is proportional to $1/\beta_i$. This function was originally used to describe variations in pup survival at the Farne Islands (Harwood 1981).

Movement

$$\begin{bmatrix} u_{m,0,1,t} = u_{s,0,1,t} \\ u_{m,1,1,t} = u_{s,1,1,t} \\ u_{m,2,1,t} = u_{s,2,1,t} \\ u_{m,3,1,t} = u_{s,3,1,t} \\ u_{m,4,1,t} = u_{s,4,1,t} \\ u_{m,5,1,t} = u_{s,5,1,t} \\ u_{m,6+,1,t} \sim \text{see below} \\ \text{M} \end{bmatrix}$$

Movement of age 6+ females is modelled as a multinomial random variable (Newman *et al.* in prep) such that:

$$\left(u_{m,6+,1,t},\ldots,u_{m,6+,4,t}\right) \sim \text{Multinomial}\left(\sum_{j=1}^{4}\left(u_{s,6+,j,t}\right), p_{6+,1,t},\ldots, p_{6+,4,t}\right)$$

where $p_{6+,j,t}$ is the probability that an age 6+ female will be in colony j following movement, given by

$$p_{6+,j,t} = \frac{\sum_{i=1}^{4}\left(u_{s,6+,i,t}\Psi_{6+,i\to j,t}\right)}{\sum_{i=1}^{4}\left(u_{s,6+,i,t}\right)} \qquad \text{(eqn 5.2)}$$

and $\Psi_{6+,i\to j,t}$ is the probability of movement for age 6+ females between colony i and colony j. The logit of this probability is given by:

$$\text{logit}\left(\Psi_{ij,t}\right) = \alpha_0 + \frac{\alpha_1}{dist_{ij}} + \alpha_2\left(n_{0,i,t-1} - n_{0,j,t-1}\right) \qquad \text{(eqn 5.3)}$$

The structure of this movement model and the prior distributions for the α parameters were derived from the CR analysis described in chapter 4. The CR study was (by assumption) of age 6+ females only, hence in the SSM only females in this age category are able to move between colonies. In the CR analysis the pup production estimates at each colony, $y_{0,i,t-1}$, were used in the movement model. In the SSM these pup production estimates were replaced by the pup numbers predicted by the model, $n_{0,i,t-1}$.

The site fidelity probability for the age 6+ females, $\Psi_{6+,i\to i,t}$, i.e. the probability of returning to breed at the same colony the female bred at the previous year, is post-processed as:

$$\Psi_{6+,i\to i,t} = 1 - \sum_{j\neq i} \Psi_{6+,i\to j,t} \qquad \text{(eqn 5.4)}$$

Ageing and sex determination

$$\begin{bmatrix} u_{a,0,1,t} = 0 \\ u_{a,1,1,t} \sim \text{Binomial}\left(u_{m,0,1,t}, 0.5\right) \\ u_{a,2,1,t} = u_{m,1,1,t} \\ u_{a,3,1,t} = u_{m,2,1,t} \\ u_{a,4,1,t} = u_{m,3,1,t} \\ u_{a,5,1,t} = u_{m,4,1,t} \\ u_{a,6+,1,t} = u_{m,5,1,t} + u_{m,6+,1,t} \\ \text{M} \end{bmatrix}$$

Hence, on average, half the pups are expected to be female.

Breeding

$$\begin{bmatrix} n_{0,1,t} \sim \text{Binomial}\left(u_{a,6+,1,t}, \gamma\right) \\ n_{1,1,t} = u_{a,1,1,t} \\ n_{2,1,t} = u_{a,2,1,t} \\ n_{3,1,t} = u_{a,3,1,t} \\ n_{4,1,t} = u_{a,4,1,t} \\ n_{5,1,t} = u_{a,5,1,t} \\ n_{6+,1,t} = u_{a,6+,1,t} \\ \text{M} \end{bmatrix}$$

In the breeding vector γ is the fecundity of 6+ females.

Observation model

The elements of the vector of pup production estimates at each of the colonies are assumed to be normally distributed. Their coefficient of variation (CV), δ, is assumed to be a linear function of the true pup production squared:

$$\begin{bmatrix} y_{0,1,t} \sim \text{Normal}\left(n_{0,1,t}, \delta^2 n_{0,1,t}^2\right) \\ y_{0,2,t} \sim \text{Normal}\left(n_{0,2,t}, \delta^2 n_{0,2,t}^2\right) \\ y_{0,3,t} \sim \text{Normal}\left(n_{0,3,t}, \delta^2 n_{0,3,t}^2\right) \\ y_{0,4,t} \sim \text{Normal}\left(n_{0,4,t}, \delta^2 n_{0,4,t}^2\right) \end{bmatrix}$$

This formulation implies that more emphasis is given to the fit of the model to the pup production data for the smaller colonies than the larger ones.

5.2.3 Full model: matrix representation

The process model can also be written as a product of the four sub-process models:

$$E\left[\mathbf{n}_t \mid \mathbf{n}_{t-1}\right] \approx \mathbf{BAM}_t\mathbf{S}_t\mathbf{n}_{t-1} \qquad \text{(eqn 5.5)}$$

As pup survival is a non-linear function of the number of pups born in the previous year this expectation is only approximate. The four sub-process matrices for model S1 are shown below, only the parts for the first and forth colonies are shown, the other two colonies being implied by dots (...). $\mathbf{S}_t$ is the matrix for survival, $\mathbf{M}_t$ is for movement between colonies, **A** is for ageing and sexing of pups and **B** is for breeding.

$$\mathbf{S}_t = \begin{bmatrix}
\phi_{p,1,t} & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \phi_a & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \phi_a & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \phi_a & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \phi_a & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \phi_a & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \phi_a & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & \phi_{p,4,t} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \phi_a & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \phi_a & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \phi_a & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \phi_a & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & \phi_a & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \phi_a
\end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0.5 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}$$

$$\mathbf{M}_t = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \Psi_{6+,1\to1,t} & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \Psi_{6+,4\to1,t} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \Psi_{6+,1\to4,t} & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \Psi_{6+,4\to4,t}
\end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & \gamma & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \gamma \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}$$

The full generalised Leslie matrix is:

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & \phi_a\gamma & \phi_a\Psi_{6+,1\to1,1}\gamma & \dots & 0 & 0 & 0 & 0 & 0 & \phi_a\gamma & \phi_a\Psi_{6+,4\to1,1}\gamma \\
0.5\phi_{p,1,1} & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \phi_a & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \phi_a & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \phi_a & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \phi_a & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \phi_a & \phi_a\Psi_{6+,1\to1,1} & \dots & 0 & 0 & 0 & 0 & 0 & \phi_a & \phi_a\Psi_{6+,4\to1,1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & 0 & 0 & 0 & \phi_a\gamma & \phi_a\Psi_{6+,4\to1,1}\gamma & \dots & 0 & 0 & 0 & 0 & 0 & \phi_a\gamma & \phi_a\Psi_{6+,4\to4,1}\gamma \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0.5\phi_{p,4,1} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & \phi_a & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & \phi_a & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & \phi_a & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \phi_a & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \phi_a & \phi_a\Psi_{6+,4\to1,1} & \dots & 0 & 0 & 0 & 0 & 0 & \phi_a & \phi_a\Psi_{6+,4\to4,1}
\end{bmatrix}$$

S1 has 11 parameters: adult survival, ϕ_a, maximum pup survival, ϕ_{pmax}, one density dependent parameter on pup survival for each colony, $\beta_{1,\dots,}\beta_4$, three movement model parameters, $\alpha_0, \alpha_1, \text{and} \alpha_2$, fecundity, γ, and the observation model CV, δ, and 28 states representing 7 age categories at 4 colonies.

5.2.4 Reduced models

Because SSMs have a modular structure, it is relatively easy to test hypotheses concerning the individual processes within the model. In this study, the primary focus was on the movement process model. Three movement models provided an adequate description of CR data for the four North Sea colonies. These were all logistic models. The full model had three parameters: an intercept, α_0, a regression slope for a distance covariate, α_1, and a regression slope for an abundance covariate, α_2. The other two models had $\alpha_2 = 0$, and $\alpha_1 = \alpha_2 = 0$. Three reduced models (S2 to S4), summarised in Table 5.1, were fitted to the data. In models S1, S2 and S3 age 6+

females could move and in model S4 there were no female movements (this model can be thought of as a control model).

Table 5.1: Details for the four state-space models (SSMs) fit to the pup production data. The models are ordered according to the number of parameters they have. The source of the α priors for each of the SSMs from the capture-recapture (CR) study is also given. The models are indexed S1 to S4 and for each the movement model form is given.

Model	nparams	Movement model form	Source of α priors
S1	11	$\text{logit}\left(\Psi_{ij,t}\right)=\alpha_0+\frac{\alpha_1}{dist_{ij}}+\alpha_2\left(n_{0,i,t-1}-n_{0,j,t-1}\right)$	CR model L1
S2	10	$\text{logit}\left(\Psi_{ij,t}\right)=\alpha_0+\frac{\alpha_1}{dist_{ij}}$	CR model R1
S3	9	$\text{logit}\left(\Psi_{ij,t}\right)=\alpha_0$	CR model R3
S4	8	no movement	not applicable

5.2.5 Prior distributions on model parameters

The prior distributions on all the model parameters are given in Table 5.2. The prior distribution on ϕ_a comes from a study of branded and tagged adult females on the Isle of May (Paddy Pomeroy, pers. comm.). The priors on ϕ_{pmax}, γ and δ are the same as those used by Thomas *et al.* (2005). The β_i parameters, which are related to the carrying capacity at each of the colonies are specified to give an expected pup carrying capacity $n_{0,cc,i}$, of 3000 for the Isle of May, 1000 for Fast Castle, 2500 for the Farne Islands, and 10000 for Donna Nook. The prior expectation for the β_i parameters were calculated for a model with no movement (S4) according to the equation given below (Len Thomas, pers. comm.):

$$\beta_i=\frac{\gamma\phi_{pmax}\phi_a^5-2(1-\phi_a)}{2(1-\phi_a)n_{0,cc,i}} \qquad \text{(eqn 5.6)}$$

The other parameters used in this calculation being set to their prior expectations. The values chosen for the pup carrying capacities, $n_{0,cc,i}$, were thought to represent a

reasonably informed guess for the expected pup carrying capacities based on the available space at each of the colonies. These values are, however, quite uncertain so the prior distributions on the β_i parameters were made quite dispersed to reflect this. The standard deviation on the β_i parameters is equivalent to a standard deviation on $n_{0,cc,i}$ of $0.5n_{0,cc,i}$.

The prior distributions on α_0, α_1 and α_2, and their variance-covariance (VC) matrices were obtained from the results of the CR analysis described in Chapter 4. These priors and VC matrices were used to generate correlated priors from a multivariate normal distribution. The movement probabilities estimated for the LHS and RHS CR analyses were very similar when the same movement model form was used. We therefore used the α priors from the CR analysis that gave the maximum variance in the movement probabilities, as it is more important that the priors encompass all likely values than have minimum variance. The capture-recapture estimates for the α parameters are given in Table 4.4 of Chapter 4. As the priors on ϕ_a, α_0, α_1 and α_2 came from previous data analyses they are known as "data-based priors" (Wade 2000).

Table 5.2: Prior distributions for the parameters used in the state-space models (SSMs)

Parameter	Prior distribution	Prior expectation
ϕ_a	logit(ϕ_a) ~ Norm(mean=3.178, sd=0.521)	0.96
ϕ_{pmax}	Beta(shape1=14.53, shape2=6.23)	0.70
γ	Beta(shape1=22.05, shape2=1.15)	0.95
δ	Gamma(shape=4, scale=0.025)	0.10
β_1	Gamma(shape=4, scale=0.000481)	0.00193
β_2	Gamma(shape=4, scale=0.00144)	0.00578
β_3	Gamma(shape=4, scale=0.000578)	0.00231
β_4	Gamma(shape=4, scale=0.000145)	0.000578
α_0, α_1 and α_2	Prior distributions and variance-covariance structure obtained from the CR analyses for the various movement model forms	-

5.2.6 Fitting the models

In its entirety, the state-space model is specified using four classes of pdfs:

$g_0(\Theta)$	Prior parameter distribution
$g_0(\mathbf{n}_0\|\Theta)$	Initial state distribution
$g_t(\mathbf{n}_t\|\mathbf{n}_{t-1},\Theta)$	State process distribution
$f_t(\mathbf{y}_t\|\mathbf{n}_t,\Theta)$	Observation process distribution

where Θ is a vector of the model parameters, $\mathbf{n}_t$ represents the state at time t and $\mathbf{y}_t$ represents the data observed at time t. Generally, it is the parameters Θ that we want to make inferences about, but we also need to estimate the unknown states, $\mathbf{n}_t$. Analytical approaches are rarely practicable in this case, due to the high-dimensionality of the integrals that would need to be solved, and methods based on simulated inference – such as likelihood based approaches and Bayesian analysis – are more appropriate (Buckland *et al.* 2004).

The approach used here was Bayesian. Bayesian analysis uses prior knowledge about the distribution of the model's parameters (these are set up as "prior distributions" on the parameters), and these prior parameters and their distributions are then interpreted with respect to the data. The resulting "posterior distributions" for the parameters are given in the modelling output (Luikart & England 1999). The high dimensionality of the parameter space means that Monte Carlo simulation procedures (see Gelman *et al.* 1995) are usually required to perform the integration necessary to find these posterior distributions. Two commonly used procedures are sequential importance sampling (SIS) and Markov chain Monte Carlo (McMC) algorithms. SIS was used here because it is well suited to the fitting to time series data (Trenkel *et al.* 2000).

The basic SIS algorithm

Step 1 Initializing: For the year 1984 ($t = 0$), K particles are drawn from the joint distribution of prior parameters and initial states. (see *Initializing the States*

below). A particle can be thought of as one possible population with parameter values that are specific to it. For the models investigated here, K = 350,000.

Step 2 Projection: The particle swarm is projected stochastically forward one year using the state process model.

Step 3 Correction: A weight, proportional to the likelihood for the particle given the current data, is calculated for each particle (see below).

Step 4 Resampling: Particles are resampled stochastically according to these weights. This is akin to "survival of the fittest" in a Darwinian analogy: particles with a high weight propagate themselves while those with a low weight are lost.

Step 5 Iteration: Steps 2 to 4 are repeated until the final time point (2003) is reached.

One benefit of the SIS algorithm is that only the observation model density, which gives the likelihoods, and not the state process density, which can be very complicated, needs to be evaluated. In the observation model used in the present state-space modelling the likelihood is the product of the normal densities for the colonies. As there is no pup production data for colony 2 (Fast Castle) prior to 1997 the likelihood $L_t^{[k]}$ for $t = \{1,\ldots,12\}$ for particle k is

$$L_t^{[k]} = f_t\left(\mathbf{y}_t \mid \mathbf{n}_t^{[k]}, \Theta^{[k]}\right) = \prod_i \left(\frac{1}{\sqrt{2\pi}\delta^{[k]} n_{0,i,t}^{[k]}} \exp\left(\frac{-\left(y_{0,i,t} - n_{0,i,t}^{[k]}\right)^2}{2\delta^{2[k]} n_{0,i,t}^{2[k]}} \right) \right)$$

where i = 1,3,4, the superscript k represents the particle number, k = 1,…,350,000, $y_{0,i,t}$ is the pup production estimate for colony i in year t, $n_{0,i,t}^{[k]}$ is the number of pups predicted for particle k and $\delta^{[k]}$ is the value of the observation model CV parameter for particle k. The likelihood for t =13,…,19, is as above except for this period there is data for all four colonies and hence i = 1,…,4.

The weight calculated in step 3 of the SIS algorithm for particle k, $w_t^{[k]}$, and used in step 4 for resampling the particles is derived as

$$w_t^{[k]} = \frac{L_t^{[k]}}{\sum_k L_t^{[k]}} \qquad \text{(eqn 5.7)}$$

where $k = 1,\ldots,350{,}000$.

Particle depletion

The main problem with this SIS algorithm is that the number of particles is rapidly depleted because the variance in the weights increases with time until a few particles have almost all the weight. In effect, Step 4 results in a loss of particle "diversity" (Doucet, Godsill & Andrieu 2000). The resulting posterior distributions show just a few point masses. This is not useful for inference and results in large Monte Carlo error. I used four techniques to combat particle depletion: rejection control (Liu, Chen & Wong 1998; Liu 2001); auxiliary particle filtering (Pitt & Shepherd 1999; Liu & West 2001; Thomas *et al.* 2005); residual resampling (Liu & Chen 1998; Liu 2001) and kernel smoothing (Trenkel *et al.* 2000; Liu & West 2001; Thomas *et al.* 2005; Newman *et al.* in press). Details of the four techniques and the augmented SIS algorithm incorporating them are given below.

1. Rejection Control

Rejection control is a technique to reduce the number of particles required to represent the distribution of the posterior at a given time point, without significantly increasing the Monte-Carlo error. It is described in detail by Liu (2001) and Liu, Chen & Wong (1998). I implemented rejection control in the second year of the SIS algorithm (the first year being used to initialize the states) because most of the particles generated from the priors are extremely unlikely given the data. In the second year particles with low likelihood were rejected probabilistically if their likelihood was less than a threshold of c (calculated as the 95th percentile of the likelihoods for the first run of the rejection controller). If the particle has a higher likelihood than the threshold it is accepted, if its likelihood is lower it is accepted with a probability equal to its likelihood divided by the threshold. Particles that survive,

even though their likelihood is less than c, have their likelihood inflated to c. The discarded particles are replaced with new ones drawn from the original priors and states.

In this analysis rejection control was only implemented in the second year. It is, however, possible to incorporate other checkpoints during the time-series. With a lot of checkpoints, however, the algorithm can become very demanding in terms of computational time.

2. Auxiliary Particle Filter

Auxiliary particle filtering works by looking ahead in the data and concentrating in parts of the time series where the model fit is expected to be particularly "good" (i.e. have high likelihood given the observation equation). The technique is described in detail by Liu & West (2001) and Pitt & Shepherd (1999).

The particles at time $t-1$ are projected forward to time t deterministically according to their parameter values. The particles are then resampled according to their weights at time t, (i.e. those that are expected to be in an area of high density are given a higher weight at this stage than the others, the idea being here to sample from particles with high predictive likelihoods). These are the first weights. The parameter values of this new set of "auxilliary" particles are kernel smoothed (see 3, below) and their states are taken back to time $t-1$ and projected forward stochastically. These particles are then resampled according to a second weight, which is proportional to the stochastic likelihood divided by the predictive (deterministic) likelihood for the particles – dividing by the deterministic likelihoods here corrects for having "looked ahead". This causes those particles that do better than expected from the first filtering to get a higher final weight.

3. Kernel Smoothing

Kernel smoothing involves perturbing the parameter values of the particles, to generate new parameter values in the vicinity of parameter space supported by the data (this is akin to "mutation" in a Darwinian analogy). In this analysis a multivariate normal kernel, which smoothes out the parameters with respect to their variance-covariance plane, was used. This form of kernel smoothing preserves the first and

second order moments of the parameters (i.e. their means and variances), and the correlations between them. Kernel smoothing of the states is not necessary, as these change stochastically as the particles are sent through the state process for each time step (Trenkel *et al.* 2000; Liu & West 2001; Thomas *et al.* 2005; Newman *et al.* in press).

Perturbing the parameters increases overall variation. One symptom is that those close to the edges of the variance-covariance plane will be pushed beyond its boundaries. To overcome this problem, the parameter values are shrunk back in towards their overall mean, which preserves the second order moments of the parameters. Liu & West (2001) suggest that auxiliary particle filters should be used in combination with kernel smoothing.

Many parameters are bounded; in order to deal with this, parameters are transformed before kernel smoothing (probabilities getting a logit transform and parameters that can only be positive getting a log transform). After smoothing the parameters are then back-transformed.

The amount of kernel smoothing done is governed by a smoothing parameter a. In the work described here a was set to 0.7 (a relatively moderate amount of smoothing), the same value used by Thomas *et al.* (2005) in their SSM for the British grey seal population and very close to the value used by Trenkel *et al.* (2000) in their SSM for red deer in Scotland.

Kernel smoothing makes the fitting algorithm more efficient and can compensate for poor choices of prior distributions and outliers in the data. However, as a result of the perterbations, some bias is introduced into the posterior estimates (Trenkel *et al.* 2000).

4. Residual Resampling

Residual resampling (Liu 2001) reduces the Monte-Carlo variance introduced during the resampling stage relative to simple random sampling with replacement. Samples are first drawn deterministically from the particles in the set and then the leftovers (the residuals) are drawn probabilistically. For example, in a set of 100 particles, a

particle with a weight of 0.501 will be sampled 50 times deterministically. At the end of the deterministic stage there may be 1 particle left to make up the new set of 100 particles. The residual for the aforementioned particle is 0.001. The residuals for all the particles are then normalised to sum to 1, and the particle set is made up by sampling stochastically from these residuals. Residual resampling gives the same expected distribution of particles as standard stochastic resampling (Thomas *et al.* 2005) but gives smaller Monte Carlo variance and uses less computational time (Liu 2001). It was therefore used when resampling particles as part of the auxiliary particle filter.

The augmented SIS algorithm

The time points in the model are $t = 0,\ldots,T$, where $T = 19$, $t = 0$ corresponds to 1984 and $t = T$ corresponds to 2003. Step 1 of the augmented algorithm builds up a swarm of relatively probably particles by implementing rejection control at time point $t = 1$. For all subsequent time points we use an auxiliary particle filter with kernel smoothing of parameter values as detailed in steps 2-11.

Step 1a *Initialising:* K particles $\left\{ n_0^{[k]}, \Theta^{[k]} \right\}$ are drawn from $g_0\left(n_0 \middle| \Theta\right)$ and $g_0(\Theta)$, where $k = 1,\ldots,K$. $K = 350{,}000$. Set $t = 1$.

Step 1b *Stochastic projection:* The particle swarm is projected forward stochastically through the state process model from time $t-1$ to time t to give $\left\{ n_t^{[k]}, \Theta^{[k]} \right\}$. A likelihood is calculated for each of the particles, $L_t^{[k]}$, given the data at time t and $\left\{ n_t^{[k]}, \Theta^{[k]} \right\}$.

Step 1c *Rejection Control:* Accept particle k with probability

$$r^{[k]} = \min\left(1, \frac{L_t^{[k]}}{c}\right)$$

where c is the 95$^{\text{th}}$ percentile of the likelihoods obtained on the first iteration of the rejection controller. Repeat Steps 1a to 1c until K particles have been accepted. Denote this new set of particles $\left\{ n_t^{[k]*}, \Theta^{[k]*} \right\}$ and set their likelihoods to

$$L_t^{[k]*} = \max\left(L_t^{[k]*}, c\right)$$

Step 1d *Correction:* Calculate a weight for each particle

$$w_t^{[k]*} = \frac{L_t^{[k]*}}{\sum_k L_t^{[k]*}}$$

Step 1e *Resampling:* Use these weights in residual resampling of the particles to obtain a new set of particles denoted $\left\{ n_t^{[k]**}, \Theta^{[k]**} \right\}$.

Step 1f For $k = 1,\ldots,K$ redefine $\left\{ n_t^{[k]**}, \Theta^{[k]**} \right\} = \left\{ n_t^{[k]}, \Theta^{[k]} \right\}$.

Step 1g Increment t to $t + 1$.

Step 2 *Deterministic projection:* The particle swarm is projected forward deterministically through the state process model from time t–1 to time t to give $\left\{ n_t^{[k]}, \Theta^{[k]} \right\}$.

Step 3 Calculate a kernel location, $m^{[k]}$, for the parameter vector, a weighted combination of the particle and the mean vector for the parameter values across all particles, $\overline{\Theta}$:

$$m^{[k]} = a\Theta^{[k]} + (1-a)\overline{\Theta}$$

herein $a = 0.7$. Then calculate a likelihood for each of the particles, $L_{\text{determ},t}^{[k]}$, given the data at time t and $\left\{ n_t^{[k]}, m^{[k]} \right\}$.

Step 4 *Correction:* Calculate a first weight for each particle

$$w_{1,t}^{[k]} = \frac{L_{\text{determ},t}^{[k]}}{\sum_k L_{\text{determ},t}^{[k]}}$$

Step 5 *Resampling:* Use the first weights in residual resampling of the particles to obtain a new set of "auxiliary" particles denoted $\left\{ n_t^{[k]*}, m^{[k]*} \right\}$.

Step 6 *Parameter mutation:* Kernel smooth the parameter values and denote the mutated parameters $\Theta^{[k]*}$:

$$\Theta^{[k]*} \sim \text{Multivariate Normal}\left(m^{[k]*}, h^2\mathbf{V} \right)$$

where $h^2 = 1 - a^2$ and $\mathbf{V}$ is the variance-covariance matrix of the parameter vector $\Theta^{[k]}$ prior to resampling.

Step 7 *Stochastic projection:* Trace the particle swarm back to time t–1 to give $\left\{ n_{t-1}^{[k]*}, \Theta^{[k]*} \right\}$. Project these particles forward stochastically through the state

process model to time t. A likelihood is then calculated for each of the particles, $L^{[k]*}_{\text{stoch},t}$, given the data at time t and $\left\{n_t^{[k]*}, \Theta^{[k]*}\right\}$

Step 8 *Correction:* Calculate a second weight for each particle

$$w_{2,t}^{[k]*} = \frac{\left(L^{[k]*}_{\text{stoch},t} \Big/ L^{[k]*}_{\text{determ},t} \right)}{\sum_k \left(L^{[k]*}_{\text{stoch},t} \Big/ L^{[k]*}_{\text{determ},t} \right)}$$

Step 9 *Resampling:* Use residual resampling on the particles according to the second weights to get the final set of K particles for time t denoted $\left\{n_t^{[k]**}, \Theta^{[k]**}\right\}$.

Step 10 For $k = 1,\ldots,K$ redefine $\left\{n_t^{[k]**}, \Theta^{[k]**}\right\} = \left\{n_t^{[k]}, \Theta^{[k]}\right\}$.

Step 11 *Iteration:* If $t < T$, increment to t+1, go to step 2, and repeat.

Initialising the states

The priors for the states were generated using the pup production estimate for 1984 together with the priors for the parameters (Thomas *et al.* 2005). The initial number of pups for each particle, $n_{0,i,0}^{[k]}$, was generated by "reversing" the observation equation, i.e. by sampling from

$$n_{0,i,0}^{[k]} \sim \text{Normal}\left(y_{0,i,0}, \delta^{2[k]} y_{0,i,0}^2 \right)$$

where $y_{0,i,0}$ is the estimated pup production at colony i in 1984 and $\delta^{[k]}$ is the value of the CV parameter δ sampled from the prior for particle k. These values were further dispersed by resampling them from a uniform distribution with bounds given by the sampled value divided by 1.3 and the sampled value multiplied by 1.3 and then rounded to the nearest integer. This was done to make sure all likely values for the initial states were contained in the sample (the value of 1.3 was chosen by Thomas *et al.* (2005) by trial and error). Initial values for the age 1 females were generated by assuming that half of the pups were female and sampling from

$$n_{1,i,0}^{[k]} \sim \text{Binomial}\left(n_{0,i,0}^{[k]}, 0.5\phi_{p,i,0}^{[k]}\right)$$

and ages 2 to 5 females from

$$n_{a,i,0}^{[k]} \sim \text{Binomial}\left(n_{a-1,i,0}^{[k]}, \phi_{a}^{[k]}\right)$$

where a = 2,…,5, for the subscript on n. Again these values were further dispersed using a uniform distribution. To generate the initial distribution of age 6+ females the breeding sub-process was "reversed" by sampling from

$$n_{6+,i,0}^{[k]} \sim \text{Negative Binomial}\left(n_{0,i,0}^{[k]}, \gamma^{[k]}\right) + n_{0,i,0}^{[k]}$$

and once again these values were further dispersed using a uniform distribution. Sampling from the Negative Binomial distribution gives the number of failures (i.e. the number of adult females not producing a pup) given the number of successes, $n_{0,i,0}^{[k]}$, for particle k (i.e. the number of females producing a pup).

For model S4 in which no movement occurred the number of female seals at Fast Castle was initialised in 1997 (the first year of data for this colony) using the same methods as above. For the other three models (S1 to S3) this was not necessary as seals were able to move to this colony from the established colonies from 1990 onwards according to the various movement process models used. Colonisation occurring at Fast Castle in 1990 as this was when seals were first noticed there by local fishermen (Kevin Rideout, SWT/NTS Ranger, pers. comm.).

5.2.7 Inference and model selection

Smoothed inference (Thomas & Harwood 2003; Thomas *et al.* 2005) was used for assessing the fit of the models to the data. Smoothed inference explores the past state of the population given all the data up to the current time. Smoothing was done by

tracing back through time to find the "ancestors" of the particles that were still present in the particle swarm in the final year. This provided the smoothed estimates for the number of pups at each of the four colonies in each of the years.

These smoothed estimates were also used for model selection by calculating the AIC (Akaike Information Criterion) values for the models (Buckland, Burnham & Augustin 1997; Burnham & Anderson 2002; Thomas & Harwood 2003; Cam *et al.* 2004; Johnson & Omland 2004). AIC works on the principle of parsimony – i.e. it makes a bias-precision trade-off given the information content in the data. More complicated models, which will fit the data better because they are more flexible, are penalised for the extra parameters that need to be estimated.

The algorithm for comparing the different models runs as follows.

1. Re-calculate the likelihoods at each time point for each of the smoothed particles for model m, $L_{m,t}^{[k]}$, using the posterior observation model CV, $\delta_{2003}^{[k]}$ (i.e. the CVs in the final time point for the particles), where $t = 0,\ldots,19$ and $k = 1,\ldots,350{,}000$. Note that δ has a time subscript because the mutations introduced by kernel smoothing perturb the parameter values slightly from year to year.

2. Calculate the likelihood for each particle in model m, $L_m^{[k]}$, as

$$L_m^{[k]} = \prod_t L_{m,t}^{[k]}$$

3. Calculate the AIC for each particle in model m, $\mathrm{AIC}_m^{[k]}$, using the likelihoods from step 2, such that

$$\mathrm{AIC}_m^{[k]} = -2\ln\left(L_m^{[k]}\right) + 2p_m$$

where p_m is the number of parameters in model m.

4. Calculate the mean posterior AIC for model m, MPAIC_m as

$$\text{MPAIC}_m = \frac{\sum_k \text{AIC}_m^{[k]}}{K}$$

where K (= 350,000) is the number of particles.

5. Repeat steps 1 to 4 for all models in the candidate set.

6. After all models have been run, calculate mean posterior Akaike weights (Thomas & Harwood 2003) W_m for each model. To calculate this weight first the difference between each model in the set and the "best" model (i.e. the model with the minimum MPAIC value) are calculated

$$\Delta_m = \text{MPAIC}_m - \text{MPAIC}_{\min}$$

Then the relative weight for each model, W_m, is calculated as

$$W_m = \frac{\exp(-\Delta_m / 2)}{\sum_q \exp(-\Delta_q / 2)}$$

where q = 1,…,4.

Due to particle depletion, however, using the data in the early part of the smoothed time series can cause substantial Monte Carlo error in the MPAIC values calculated. To minimise this problem, the Akaike weights were calculated for the MPAIC values calculated for the smoothed particles between 1990 and 2003 (i.e. t = 6,…,19).

5.2.8 Future abundance prediction

The future abundances of pups at each of the colonies were assessed by projecting the models forward using the posterior distributions of states and parameter values. In the

present analysis 1000 particles were randomly selected from the particles that survived until the final time point (2003). These were projected forward stochastically over a time horizon of 200 years, when the colonies had reached equilibrium.

All of the models were coded in R Version 2.0.1 (R Development Core Team, 2004). Source code is available in the CD that accompanies this thesis (see section 5.5 for details). Each model was run four times in order to assess the level of Monte Carlo variation in the posterior estimates and to assess whether or not the models had converged.

5.3 Results

The results from the model selection are given in Table 5.3. As there was some Monte Carlo error in the MPAIC values calculated (repeat runs of the models producing slightly different outputs), the average of the four MPAIC values was used in the calculation of the mean posterior Akaike weights. Model S3 has a clear advantage based on these weights. This model assumes a constant movement probability between colonies, with no effect of distance or abundance. The "no movement" model, S4, had effectively zero Akaike weight.

Table 5.3: State-space model (SSM) model comparison results

SSM	Average MPAIC over four runs	mean posterior Akaike weight (using average MPAIC over four runs)
S1	603.4	0.052
S2	600.6	0.211
S3	598.1	0.737
S4	643.1	0.000

MPAIC: Mean posterior Akaike's information criterion (see text)

Figure 5.1 shows the performance of model S3 using the run that had the smallest amount of particle depletion. Only two pup production estimates show a poor fit: the Farne Islands in 1999 and Fast Castle in 2003.

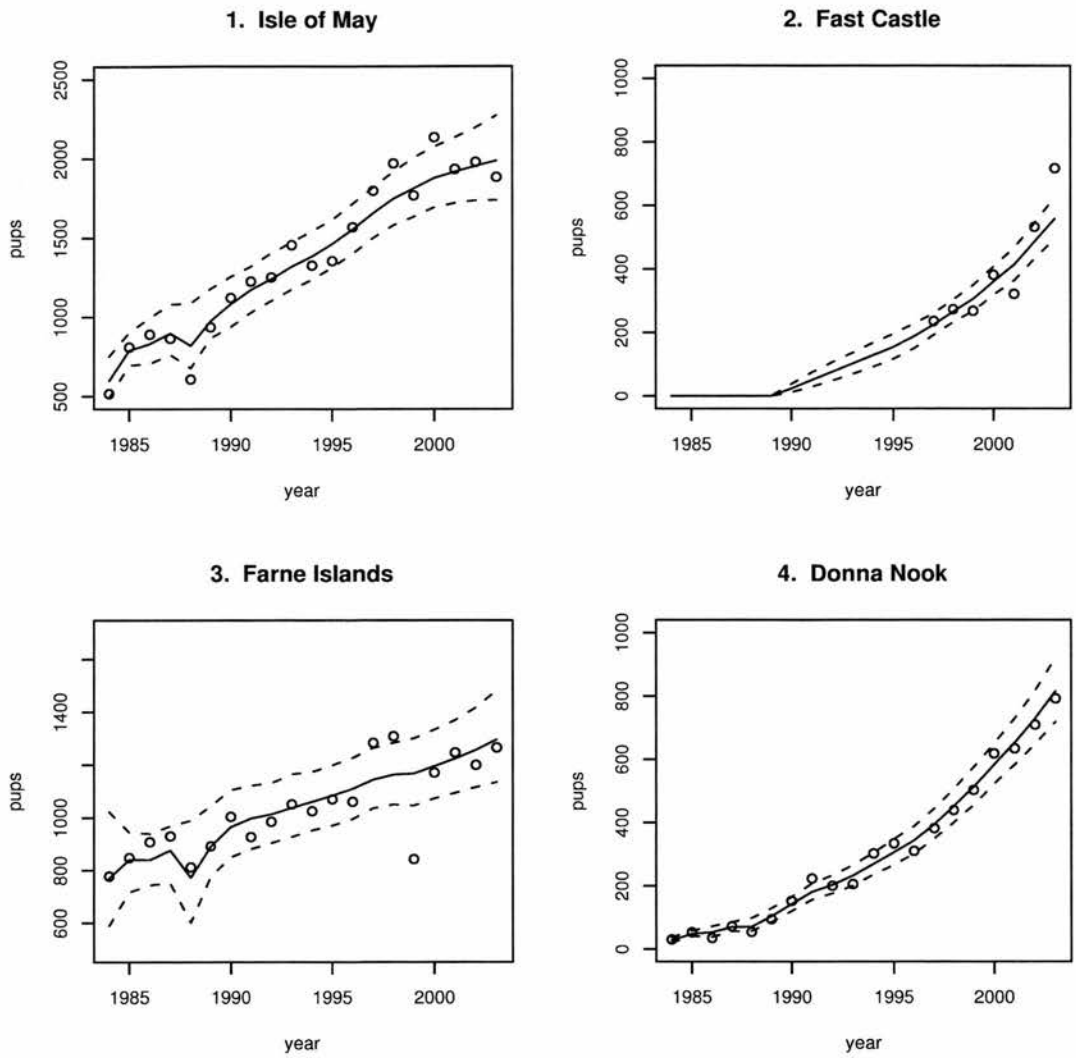


Figure 5.1: Pup production data and smoothed estimates for model S3. Input data are shown as circles, the smoothed mean of the particles is shown by the solid line and the dashed lines show the 2.5th and 97.5th percentiles (the posterior 95% Bayesian credibility intervals).

The prior and posterior distributions for the parameters in model S3 are shown in Figure 5.2. The posteriors on adult survival, ϕ_a, fecundity, γ, the observation model CV, δ, and the movement parameter, α_0, have all changed significantly from the priors, indicative of information content in the data on these parameter values.

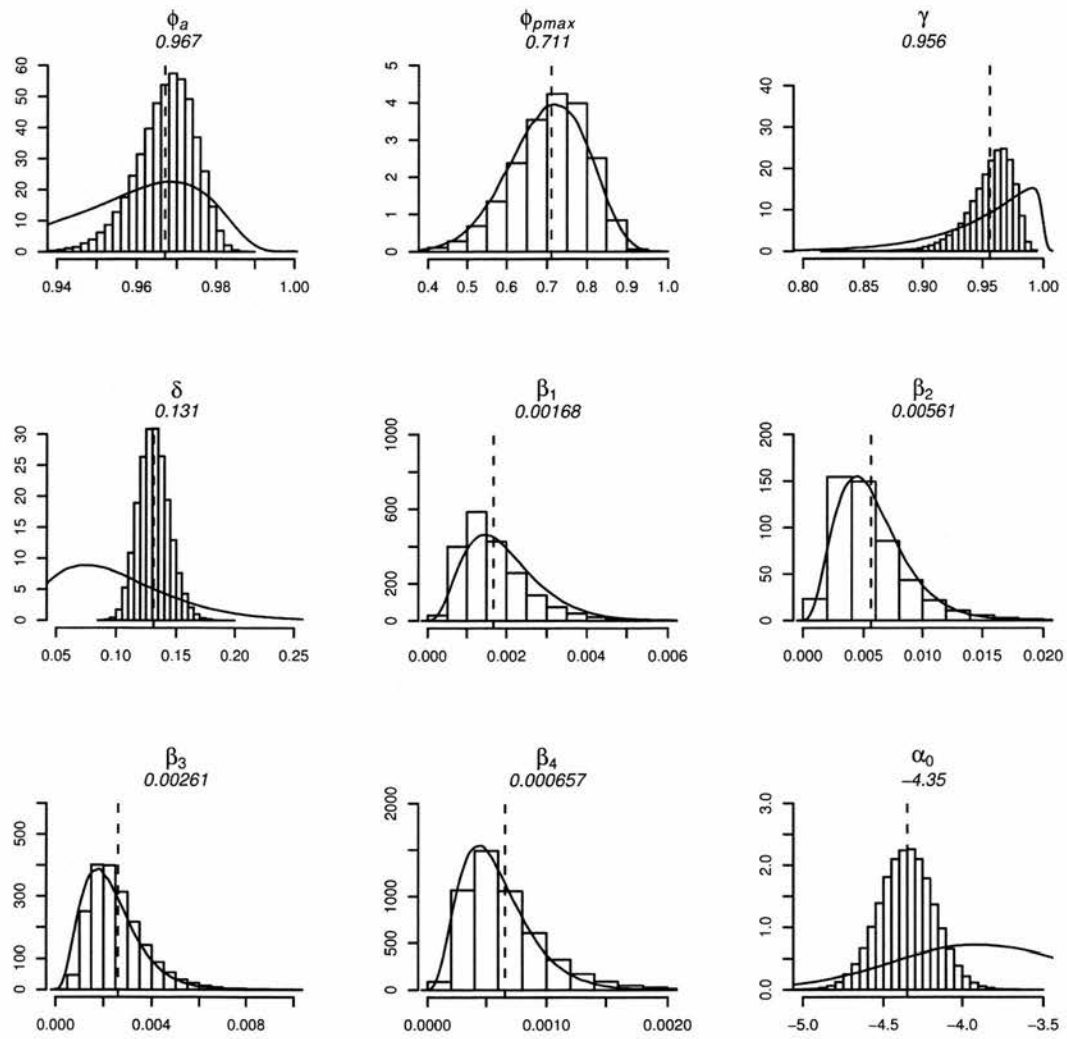


Figure 5.2: Priors (solid lines), posteriors (histograms) and posterior means (dotted vertical line) for model S3. The posterior mean is given underneath the name of the parameter

Table 5.4 compares the movement probabilities estimated in the initial CR analyses with those from model S3. The SSM analysis produces a lower estimate for movement that the CR analyses, and consequently the SSM estimate of the site fidelity rate (returning to the same colony from one year to the next) is higher. The width of the credibility interval for the movement probability, obtained from CR model R3 (from chapter 4), was reduced by 82% when the SSM S3 was fit to the pup production data.

Table 5.4: A comparison of the movement probability estimates from the capture-recapture (CR) models and the state-space model (SSM) S3 incorporating only a single movement parameter. The mean, 2.5th and 97.5th percentiles are given. The movement probability, Ψ, is given in terms of the source colony and the destination colony.

Movement probability Ψ	CR model L3	CR model R3 (α prior from this)	SSM S3 (run with smallest amount of particle depletion)
$i \leftrightarrow i$	0.933 (0.838, 0.973)	0.940 (0.830, 0.979)	0.961 (0.947, 0.973)
$i \leftrightarrow j$	0.022 (0.009, 0.054)	0.020 (0.007, 0.057)	0.013 (0.009, 0.018)

$i \leftrightarrow i$ gives the site fidelity probability and $i \leftrightarrow j$ gives the probability of moving from one colony to each of the other three colonies available (see eqn 5.4)

As there was also some support, based on Akaike weights, for model S2, Table 5.5 compares the movement rates estimated from the CR models with the same movement form as this SSM. The SSM estimates show a less pronounced distance effect than the CR estimates, again the level of movement picked up by the SSM is lower than the CR analysis suggested and the uncertainty in the movement probabilities is much reduced.

Table 5.5: A comparison of the movement probability estimates from the capture-recapture (CR) models and state-space model (SSM) S2 incorporating a distance effect for movement. The mean, 2.5th and 97.5th percentiles are given.

Ψ	CR model L2	CR model R1 (α priors from this)	SSM S2 (run with smallest amount of particle depletion)
$1 \leftrightarrow 1$	0.924 (0.716, 0.979)	0.916 (0.643, 0.979)	0.961 (0.946, 0.973)
$2 \leftrightarrow 2$	0.912 (0.703, 0.973)	0.903 (0.630, 0.974)	0.959 (0.943, 0.972)
$3 \leftrightarrow 3$	0.955 (0.828, 0.986)	0.967 (0.806, 0.993)	0.964 (0.949, 0.976)
$4 \leftrightarrow 4$	0.981 (0.838, 0.998)	0.992 (0.818, 0.9997)	0.970 (0.948, 0.984)
$1 \leftrightarrow 2$	0.057 (0.016, 0.177)	0.073 (0.019, 0.236)	0.017 (0.010, 0.028)
$1 \leftrightarrow 3$	0.014 (0.003, 0.053)	0.009 (0.001, 0.060)	0.012 (0.008, 0.017)
$1 \leftrightarrow 4$	0.006 (0.0006, 0.054)	0.002 (0.0001, 0.061)	0.010 (0.005, 0.017)
$2 \leftrightarrow 3$	0.025 (0.010, 0.066)	0.022 (0.006, 0.073)	0.014 (0.010, 0.019)
$2 \leftrightarrow 4$	0.006 (0.0007, 0.054)	0.003 (0.0001, 0.061)	0.010 (0.005, 0.017)
$3 \leftrightarrow 4$	0.006 (0.0007, 0.054)	0.003 (0.0001, 0.060)	0.010 (0.005, 0.017)

1 = the Isle of May, 2 = Fast Castle, 3 = the Farne Islands and 4 = Donna Nook.
$1 \leftrightarrow 1$ represents the probability of site fidelity at the Isle of May and $1 \leftrightarrow 2$ represents the probability of moving in either direction between the Isle of May and Fast Castle.

The predicted future pup abundances from projecting model S3 forward 200 year (towards equilibrium) using the posterior states and parameters are shown in Figure 5.3. There is wide variability in the equilibrium sizes predicted. These sizes, however, tend to be greater than our prior estimates for colonies 1, 2 and 3.

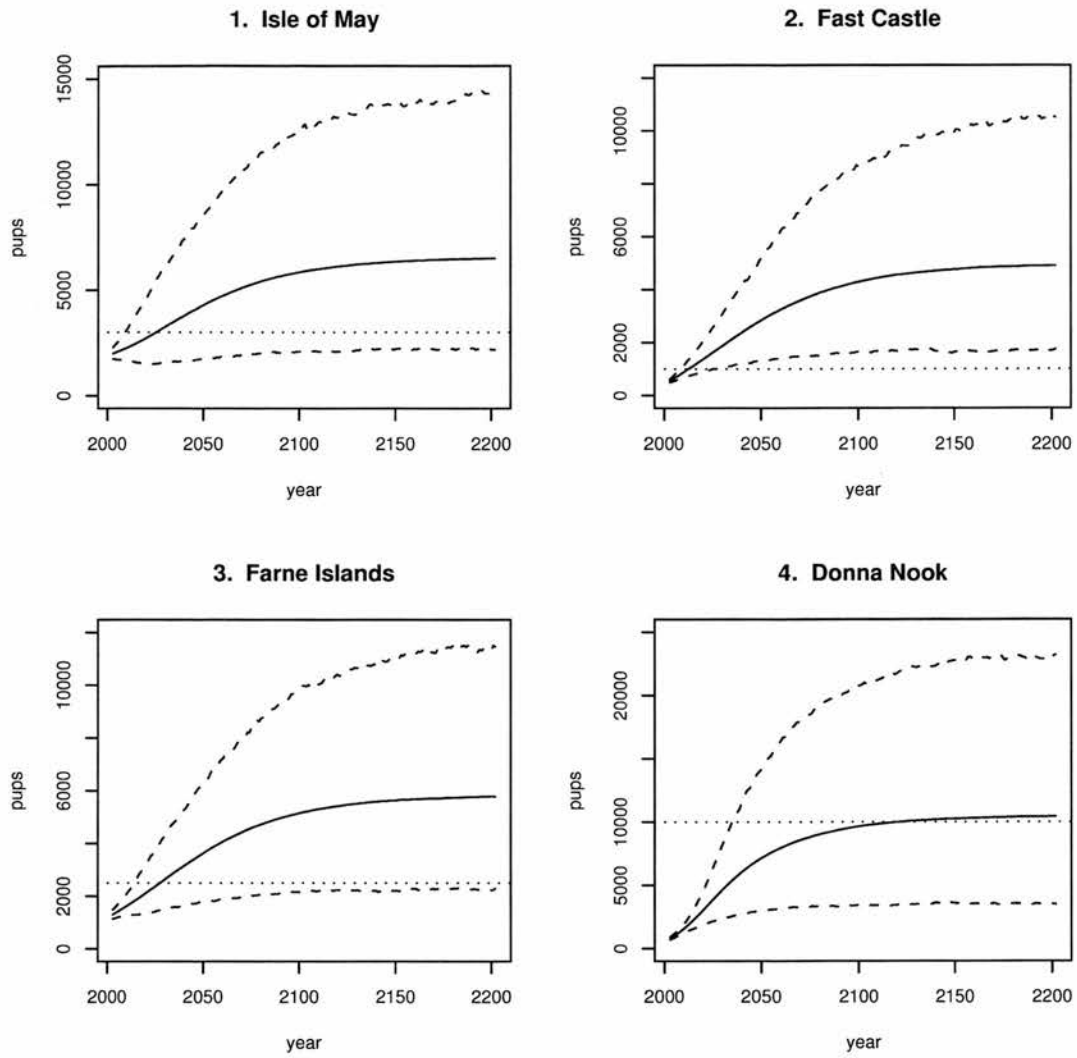


Figure 5.3: Predicted future pup abundances using posterior states and parameters from SSM S3. Mean of the particles is shown by the solid line, the dashed lines showing the 95% Bayesian credibility interval and the dotted horizontal line gives the prior estimates for pup abundance at carrying capacity at each colony.

A comparison of the mean predicted future pup abundances from projecting three of the four SSMs forward 200 years using the posterior states and parameters are shown in Figure 5.4. Model S1 produced unstable pathological behaviour where all animals were drawn from the smaller colonies into the larger ones; such instability is thought unlikely in reality, so this model has not been shown in this comparison. The no movement model, S4, despite getting a zero Akaike weight has been left in for comparison with the other two models. Incorporating movement into the SSMs results in a different equilibrium distribution of animals between the four colonies and also in a different equilibrium size for the whole system. The results from the two best fitting models, S2 and S3, are very similar in their predictions. For all three models,

however, there is considerable overlap in the 95 % Bayesian credibility intervals for the trajectories at all four colonies (not shown here). Again, all the models consistently reach higher equilibrium values for colonies 1, 2, and 3 than our prior estimates. The model-averaged carrying capacity for each colony (Buckland *et al.* 1997), calculated using the MPAIC weights were: Isle of May 6595 pups, Fast Castle 4809 pups, Farne Islands 5616 pups, and Donna Nook 10887 pups.

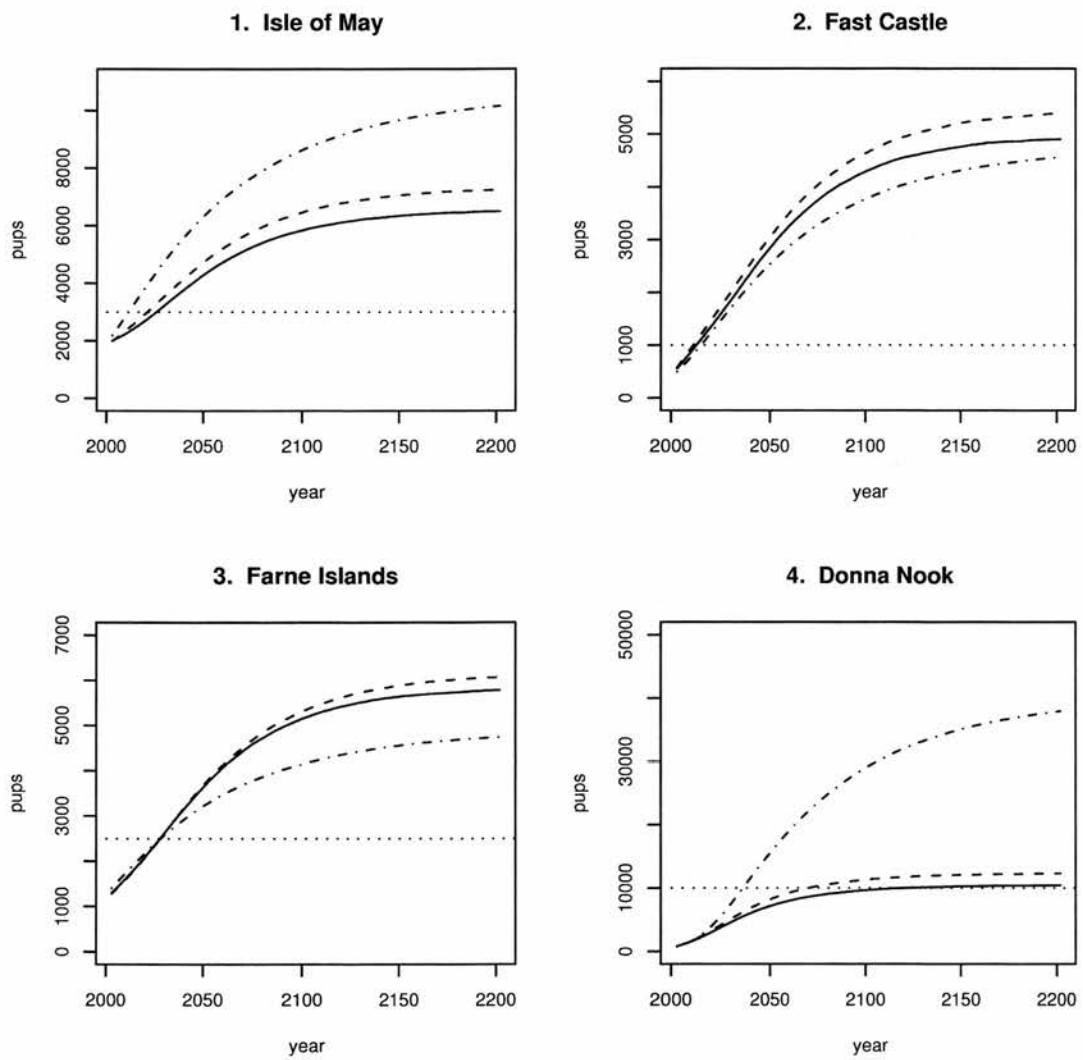


Figure 5.4: Mean predicted future pup abundances (from 2003 to 2203) using posterior states and parameters from models S2 (dashed line), S3 (solid line) and S4 (dotted and dashed line). The dotted horizontal line gives the prior estimates for pup abundance at carrying capacity at each colony.

5.4 Discussion

5.4.1 Posterior parameter estimates

We found that the pup production data did contain information on adult female survival, a similar result to that obtained by Thomas *et al.* (2005) for the entire UK grey seal data set. The posterior mean (0.967) was slightly higher than that of the prior (0.96), which came from a study of branded and tagged breeding females at the Isle of May (Paddy Pomeroy, pers. comm.). The prior estimate was based on the assumption that there was no emigration from this colony, and this may explain why the posterior mean was higher. There was also information in the data on the amount of movement by adult females between the four colonies (see next section). The pup production data provide additional information on the CV of the observation model. The posterior estimate of 13% was somewhat higher than expected. This may be because this parameter includes model misfit (Len Thomas, pers. comm.). There was also a small amount of information on the fecundity rate.

There was, however, little information to refine our estimates of the maximum pup survival rate and the beta parameters that affect the equilibrium values at the colonies. This is not surprising as all four colonies are still growing so there will be little if any information in this data on density dependence and the equilibrium values at the colonies.

5.4.2 Grey seal movement

There was strong evidence that incorporating movement between the colonies improved model performance. In fact, the model with no movement (S4) had effectively zero MPAIC weight. The estimated movement probability from the "best"

SSM, S3, was approximately 35% lower than that estimated from the CR data. This may be because the CR data are likely to have included some short-term within-season movements by females. Most of these movements are likely to be categorised as "true" between year moves between colonies, thus biasing the CR estimate upwards. The SSM estimate also had a much tighter confidence interval, indicating the additional information about movement contained in the count data.

Although this analysis provided strong evidence for movement between colonies, the estimated probability of not moving, from model S3, was high (0.961) so that most adult females remain faithful to the same breeding colony from one year to the next. Pomeroy *et al* (1994) obtained a similar result from a detailed study of permanently marked breeding female grey seals at North Rona. However, Thomas *et al.* (2005) found little evidence for movement when they modelled grey seal population dynamics at a regional level. This may be a result of the different spatial scale used by the two models: even relatively high levels of movement between adjacent colonies would not necessarily generate significant levels of movement between adjacent regions.

It has long been held that distance is an important factor in animal movement models (MacArthur & Wilson 1967) and there was some support for model S2 which incorporated distance effects into the movement process (Akaike weight of 0.211). The effect of distance on moving picked up by this model was not particularly strong however, which explains why this model was not preferred above model S3 based on Akaike weights. Again the stronger effect of distance picked up in the CR analysis, when compared to the SSM results, may have been a consequence of within-season movements of females affecting the CR data (within-season movements may be higher between closer colonies).

The CR analysis supported a model which included both distance and colony size as determinants of movement probability. However, this model predicted that seals are attracted from smaller colonies to larger ones, possibly as a result of conspecific attraction (Smith & Peacock 1990; Cam *et al.* 2004). An alternative explanation is that it is an artefact caused by the fact that Fast Castle, the smallest and most recently established colony, had a much lower fidelity rate than any other colony. This view is

supported by the fact that the movement probabilities associated with Fast Castle had wide confidence limits and were highly sensitive to the choice of movement model. When this movement process was incorporated into a SSM, there was little support from the pup production data (MPAIC weight of 0.052) and it produced pathological long-term dynamics, with the largest two colonies acting as sinks which drained all the animals out of the other two colonies.

5.4.3 Abundance prediction

Bowne and Bowers (2004) point out that relatively few studies of metapopulations document rates of inter-patch (in our case inter-colony) movements, and even those that do rarely determine population level consequences of these movements. The primary aim of the modelling done in this chapter was to explore the nature and amount of movement present in my study population, but I was also interested in the consequences of this movement for long-term dynamics. The best fitting model predicted equilibrium sizes for the individual colonies that were very different from the initial expectations. The results suggest that all four colonies are still growing and that there is little information on their equilibrium size in the time series that is currently available. The 95% Bayesian credible intervals in Fig. 5.3 indicate that the predictions range from virtually no increase to a continuation of the current growth trend.

Clearly, movement is necessary for the establishment of new colonies, such as Fast Castle, but it can also alter the dynamics of the individual colonies and the metapopulation as a whole (Hanski, Kuussaari & Nieminen 1994; Hanski, Alho & Moilanen 2000; Matthiopoulos *et al.* in press). Incorporating movement into the SSMs resulted in a different equilibrium distribution of animals between the four colonies and also in a different equilibrium size for the whole system.

The fact that the equilibrium densities at the Isle of May, Fast Castle and the Farne Islands predicted by model S3 (see Fig. 5.3) are quite different from the priors – in spite of the fact that the β posteriors did not change much from their priors – is partly

a consequence of the other parameters in the models changing from their priors. The β prior expectations were set relative to the prior expectations for the other model parameters. If there is no movement, the equilibrium size of each colony is determined by the β parameters. When movement is incorporated, however, equilibrium size at a particular colony is no longer determined only by the internal pup survival but it is also affected by emigration and immigration: at equilibrium some colonies will be net suppliers of migrants while others will be net receivers. The reason that the carrying capacity at Donna Nook does not overshoot its prior expectation (unlike the other three colonies) is because this colony (which is believed to have the potential to grow the largest) is a net supplier of migrants at equilibrium.

The most noticeable difference in the predictions under the different SSMs is for Donna Nook (see Fig. 5.4), where the no movement model (S4) predicts a much larger equilibrium density than the other models. This is likely to be because at equilibrium no animals were able to emigrate from Donna Nook in model S4 and also because this model was only able to explain the rapid early rise in pup numbers at this colony by having a very large equilibrium density. However, when migration into Donna Nook was allowed (as in models S3) the rapid early rise at this colony could be adequately charted (see Fig. 5.1).

5.4.4 Density dependence

The functional form for density dependent pup survival we used here implies that the largest changes in pup survival occur at relatively low pup production levels. It was initially developed using data collected at the Farne Islands before 1972 (Harwood & Prime 1978; Harwood 1981). However, it is generally believed that long-lived animals, like grey seals, show the strongest evidence of density-dependence when they are close to their carrying capacity (Fowler 1981); this could be modelled by incorporating a shape parameter into the pup survival function. There is also continuing debate about whether or not density dependence operates on pup survival (Twiss, Duck & Pomeroy 2003); it may instead operate on adult fecundity (Thomas & Harwood 2004a). Such an alteration to the model structure would be simple to

implement. Nevertheless, in a review of 160 studies of marine mammals and large terrestrial herbivores, Sinclair (1996) found that in 66 of these density dependence acted on first-year survival.

It is also possible that as the colonies start to level off in the future density may have an effect on female seals movement probabilities. A more appropriate movement model in this case may be one similar in form to that used by Thomas *et al* (2005) in their regional level SSM analysis of the British grey seal population – the form of their movement model was theoretically driven based on the hypothesis that fitness and distance factors determine movement (Ruxton & Rohani, 1999). In their movement process model a seal will only move to a new colony if the survival of its offspring is higher at that colony than at its original colony and when the new colony is not too far away from the original one.

These are just some of the factors that need to be considered before the future size and spread of the British grey seal population can be reliably predicted.

5.4.5 Fitting state-space models

A major problem with fitting SSMs using SIS is particle depletion, which gives rise to Monte Carlo (MC) error in the posterior estimates. In other words, particle depletion lowers precision. To lower the amount of particle depletion and reduce this variability one solution is to use more particles, but this requires more memory. The next generation of SSMs are being programmed using C as opposed to R or S-Plus to obtain this extra memory.

In the present state-space modelling exercise several techniques were used to combat the problem of particle depletion and hence reduce the level of MC error between model runs. However, there are problems with these techniques. For example, the optimum level for kernel smoothing is not well understood. If too much smoothing is done, then the relationship between the states and parameters is not preserved. Thomas and Harwood (2004a) explored a modified algorithm for fitting SSMs that

had no kernel smoothing at all. Although this algorithm was simple and reliable, it was inefficient, requiring a large amount of computer time.

Markov Chain Monte Carlo (MCMC) methods provide an alternative to SIS. However, there are significant difficulties in choosing an MCMC sampler (i.e. choosing the proposal distributions and updating schemes) when the models are non-linear and non-Gaussian. Furthermore, MCMC suffers from an analogous problem to particle depletion in SIS: the proposition of moves in parameter/state space that are rarely accepted, in which case the model becomes "stuck" in the parameter/state space and does not "mix" well (Newman *et al.* in prep).

5.4.6 Model selection

In this analysis, model selection was based on MPAIC. Each model was run independently, its MPAIC value was calculated, and then a model weight was calculated from the MPAIC value. An alternative approach with SIS is to run all the models together and to then resample particles at each time point according to their Akaike weight. Models that fit well will have more particles resampled from them than those that fit the data poorly. Posterior model probabilities are then determined by the proportion of surviving particles representing each model. This is analagous to the technique of Reversible-Jump MCMC[9] (Jamieson & Brooks 2003). In this case, posterior model probabilities are determined by the proportion of time spent in the parameter space for each model.

Clearly, the model selection criterion used will affect the weights given to each model in the candidate set. AIC and its "relatives" work on the principle of parsimony, such that there is a penalty term for the number of parameters that must be estimated. It is not clear, however, if it is appropriate to penalise the number of parameters in SSMs. We have also applied AIC, traditionally a frequentist tool, in a Bayesian framework to calculate the mean posterior AIC. Whether the mean was the right statistic to use is also uncertain. There are also other Bayesian approaches to model selection that are

[9] Reversible-Jump MCMC extends MCMC to search model space as well as parameter/state space.

based on somewhat different assumptions to those made by AIC (Burnham & Anderson 2002).

Another important issue with respect to model selection is the weighting of different models *a priori*. In the analysis done in this chapter all four models were essentially given equal weights to begin with. However, we could have perhaps capitalised more efficiently on the information present from the capture-recapture data by utilising the model weights given in Table 4.3 of chapter 4. However, as the ordering and size of the relative weights for the various movement model forms were different for the LHS and RHS, it would have been difficult to justify which weights to use. Furthermore, it would have been hard to determine a prior weight for the "no movement" SSM (S4).

5.4.7 Comparison with other modelling methods

Modelling procedures comparable to those used here have been used by Fujiwara and Caswell (2002). They demonstrate how multi-stage mark-recapture data can be used to estimate the transition probabilities in a stage-based population projection matrix. Our methodology is similar, although we used mark-recapture data on transitions between geographic locations rather than life stages to parameterise our age-based population projection matrix. We extend the use of such matrices, however, by including an observation model that allows us to fit our population projection matrix to further data on the species in question. This in turn enables us to further refine the parameters in such matrices as well as making future predictions using the models.

For population viability analysis of metapopulations the simulation software RAMAS/METAPOP (http://www.ramas.com/software.htm) is valuable. The models the user can build are very flexible and can incorporate environmental and demographic stochasticity and density dependence, and can be used to predict the risk of extinction due to human impacts. The results from CR experiments can be incorporated to provide the input movement parameters for the simulations and the effects of movement on future colony and regional population dynamics can be

explored (e.g. Inchausti & Weimerskirch 2002). However, the software presently does not explicitly incorporate uncertainty in the parameters used for input, nor can it be directly fit to abundance data to refine parameter estimates as can be done using SSMs.

5.5 Supplementary material

The CD that accompanies this thesis contains source code and data to implement model S3, written using the statistical computing language R Version 2.0.1 (R Development Core Team, 2004). The CD contains six text files for running the model and a ReadMe text file which lists the contents of the other files and how to use them.

Chapter 6

General Discussion

In the preceding chapters I have demonstrated a range of methods for making inferences about the population dynamics and spatial structure of wildlife populations. Specifically, I used a variety of statistical analysis tools to uncover patterns in genetic, photo-ID and abundance data for the British grey seal population.

6.1 Main findings

In chapter 2, I used individual-based simulation models to explore the genetic consequences of female grey seals "choosing" mates in year t that are genetically dissimilar to their mates from year t-1. When this choice was based on only one or two loci, a marked effect on the population allele frequencies was observed: rarer forms were favoured and more common forms were penalized. As the number of loci on which this choice was based increased, the effect on population allele frequencies (over a fifty year time horizon) was far less detectable. In chapter 3, I explored microsatellite DNA data for the grey seal and found that there is genetic differentiation between colonies from different regions around the British Isles, but approximate panmixia among the four main breeding colonies in the North Sea region. The remainder of this thesis explored the population dynamics and spatial structure of female grey seals breeding in this region. Multisite capture-recapture

modelling of photo-ID data for these colonies revealed that most females remained faithful to a specific colony, but that some of them did move between colonies in successive years. Model selection revealed that this movement was more likely between colonies that were close together and from smaller colonies to larger ones. This analysis was based on a rather short (3-year) dataset. Therefore, in chapter 5, I developed a set of metapopulation models for the North Sea system that were fitted to a 20 year time series of pup production estimates for these four colonies. A Bayesian state-space approach was used to fit these models to the data.

The results from the capture-recapture study provided informative priors for the movement process models in these metapopulation models. The precision in the movement parameter estimates were improved substantially by the use of this extended time series. Model selection in this instance favoured a movement model with only a single parameter and no effects of distance or abundance. However, there was also some support for a movement model that incorporated distance. There was little support in these data for an effect of abundance on movement probabilities. Predictions of future colony sizes made using these models demonstrated that the incorporation of movement, and the way in which it was modelled, affected both local and regional dynamics. The differences between model predictions were most evident as colonies approached their carrying capacities, suggesting that our ability to discriminate between models should improve as the length of the grey seal time series increases.

6.2 Grey seals and man in the North Sea

Most species on Earth are affected directly or indirectly, purposefully or accidentally, by the activities of man. The British grey seal is no exception to this rule. It has had a varied relationship with man over the years. In the past this species was exploited for the oil which could be extracted from seal carcases (Watt 1951). However, in 1914 the Grey Seal Protection Act was passed as it was believed that the population had been reduced to 500 individuals (Harwood 1984). This was replaced by the Grey Seal Protection Act 1932 which protected seals during the breeding season, from 1

September to 31 December (Watt 1951). As a result of this protection, the size of the population increased and, in 1963 and 1965, grey seal pups were killed at the Farne Islands (the only long-established grey seal breeding colony in the North Sea at that time) as part of an attempt to protect fisheries (Coulson & Hickling 1965, 1969). Howevere, the size of the population using the Farne Islands continued to increase, and these culls were abandoned. In the 1970s, the National Trust (which owns the Farne Islands) became concerned about the effects of seals breeding on Brownsman and Staple Island on areas of delicate vegetation, such as the sea campion, *Silene maritima*, and the subsequent erosion of the fragile soil cap (Bonner & Hickling 1971). It was also believed that further culling would result in better conditions for the seals, with less pup mortality and injury resulting from overcrowding. A management plan was adopted that involved the killing of adult seals and their pups (a more effective measure than only killing pups). This was carried out in 1972 (Bonner & Hickling 1974) and 1975 (Hickling, Hawkey & Harwood 1976). In 1977 a further measure to protect the two vegetated islands (Hickling, Hawkey & Harwood 1978), known as the "disturbance plan" (Prime 1981), was introduced. Seals tend to avoid islands continuously inhabited by humans, so wardens were housed on Brownsman and Staple during the breeding season (Hickling, Hawkey & Harwood 1978) to deter them from pupping there. This plan is still in operation today.

Intriguingly, the "disturbance plan" seems to have been responsible, at least in part, for the establishment and growth of two other colonies in the North Sea (the Isle of May and Donna Nook). The Isle of May, the closer of these two colonies to the Farne Islands, had an estimated pup production of 25 in 1977, but this rose to 300 in 1979 and 499 in 1980. In 1980 an adult female seal branded at the Farne Islands in 1970 was observed by the water's edge at the Isle of May (Prime 1981), supporting the hypothesis that at least some of the growth at the Isle of May was the result of immigration from the Farne Islands.

No grey seal pups were found at Donna Nook during a survey undertaken in 1973. No further counts were made until 1981, when 34 pups were present (SCOS 2002). The more dramatic initial rise in the numbers of seals at the Isle of May is perhaps due to the proximity of this island to the Farne Islands, and the fact that there were already a few seals present at the Isle of May in the mid-1970s. Dispersing individuals are

thought to use the presence of conspecifics as cues for habitat suitability when selecting a new colony (Clobert, Ims & Rousset 2004).

The state-space models (SSMs) developed in chapter 5 of this thesis were fitted to the pup production data for the North Sea colonies between 1984 and 2003. During this time another colony, Fast Castle, was established. The best fitting models predicted that most of the animals involved in this colonization came from the Isle of May, the closest and largest of the three colonies in the North Sea region.

Based on the history of the colonies in the North Sea outlined above it is not surprising that the microsatellite genetic analysis (chapter 3) uncovered little differentiation between them. This history also emphasizes the fact that human presence on a breeding site can strongly deter females from pupping there. Indeed, some of the apparent movements of females suggested by the photo-ID dataset may have been an artefact of the way in which the observations were collected. For example, a female looking for a place to pup on a particular colony might have been disturbed by the photography team and decided to go elsewhere to give birth. Furthermore, such a scenario could help to explain the within-season movements between colonies observed in this dataset. The preceding considerations bring to light two related complications when studying wildlife populations: (1) it is difficult to observe a system without also having an effect on it; (2) it is not easy to untangle "natural" from anthropogenic effects in wildlife population dynamics.

6.3 Embedding simulation models in inference

"*At the moment, the Bayesian revolution is in its earliest phase, and it will be some time yet before the dust has settled and we can judge which are the most promising avenues for exploration*" (Beaumont & Rannala 2004).

The culmination of this thesis is a set of SSMs describing the dynamics of female grey seals breeding at colonies in the North Sea. The state-space framework provides a very promising avenue for exploring ecological data from a Bayesian perspective.

SSMs use simulations from a realistic stochastic population dynamics model to create multiple realizations of the population (Buckland, Goudie & Borchers 2000). Complex simulation models are often criticised because the large number of parameters can rarely be rigorously estimated and structural assumptions cannot be tested (Hanski & Gaggiotti 2004). If no account is taken of the uncertainty associated with such models, it can be fatal to trust their predictions (Hanski 2004). However, the underlying process model of an SSM is embedded into an inference framework via the observation model, so that the models can be calibrated using survey data (Buckland et al 2004). The state-space approach can incorporate both process and measurement errors using probability density functions (Newman et al, in press). Structural assumptions can be explored using model selection techniques, such as those described in chapter 5, to compare different models of the movement process. Furthermore, parameters for which the data provides little information are easily identified because their posterior distributions are virtually identical to their priors (Harwood & Stokes 2003).

6.4 Future directions

In this thesis I have shown how two different data sources (in this case, photo-ID and pup production) can be combined to provide improved inference about levels of movement between grey seal colonies in the North Sea. In this case the photo-ID data provided informative priors for the SSMs that were fitted to the pup production data. A potentially more rigorous way to combine these data sources would have been to fit to both datasets simultaneously through the state-space framework. The SSM would require significant expansion to include a process and observation model for the capture-recapture analysis, but – provided the two data sets are independent – the weight for each simulated population (each particle) would be proportional to the product of its likelihoods for the two data sets.

The state-space approach could also provide a means to extract more information from the photo-ID data. Specifically, animals were photographed from both the left-hand side (LHS) and the right-hand side (RHS). In a substantial number of cases it

was possible to match the two sides of the head in the field, but this information was not included in the analysis. Instead, I analysed the LHS and RHS data separately. An additional difficulty with the photo-ID dataset is the fact that it contains a number of false negative records: previously photographed animals that were not recognized when they were photographed again. Accounting for such errors would reduce any bias potentially caused by them (Stevick *et al.* 2001). Appendix 2 describes how an SSM could be extended to incorporate the capture-recapture data directly, taking into consideration photographs of both sides of the seals head and the possibility of false negative errors.

SSMs that can be fitted to genetic data would also be valuable. Where a time series of genetic data for adult animals and pups exist, such as for the Isle of May, these SSMs could help to resolve complexities in the mating system. For example a model of the form given in chapter 2, which explores the consequences of mate choice, could, with appropriate modification, be fitted to the genetic data. However, the required SSM would probably need to be individual-based in order to replicate the complexities of the mating system. This would put extreme demands on computer power if fitting methods similar to those used in chapter 5 were adopted. Such computer-intensive models would probably require the fitting procedure to be recoded in a lower-level computing language, such as Fortran or C. The same modelling approach could also be used to estimate levels of polygyny and mate fidelity, and to examine the effects of inbreeding.

Genetic data can also be utilized to explore the colonization process if samples can be collected before pups born at newly established colonies have returned there to breed. Gaggiotti *et al* (2002, 2004) used this approach to determine the likely source of the animals that founded recently established grey seal colonies in Orkney. It would have been interesting to explore the colonization of Fast Castle in the North Sea region using such an approach. However, the sample sizes at the North Sea colonies, particularly Donna Nook, were probably too small to make this feasible, and those from Fast Castle were taken too long after that colony was established to be useful.

The state-space approach advocated in this thesis is still in its infancy. There are methodological issues that are currently being explored and refined, such as whether

models should be fit to data via sequential importance sampling (SIS) or Markov chain Monte Carlo (MCMC). SIS methods are simpler to program, and slightly more intuitive, but MCMC appears to give less Monte Carlo error in posterior estimates (Newman *et al*, in prep).

The Holy Grail for state-space modellers is the development of a spatiotemporal ecosystem model: a virtual reality on the computer (Buckland, Goudie & Borchers 2000). In such a model, each simulated "particle" would represent an entire ecosystem, rather than a single population. Such a model could be fitted simultaneously to data from many species in order to quantify key ecosystem level processes. Such modelling could, for example, help identify how best to slow down, or even halt, the present worldwide decline in biodiversity. Such modelling is clearly rather ambitious at present, but the foundations have been laid and, with the ever-increasing power of computers, there is great potential for such an integrated approach to wildlife conservation.

6.5 A concluding remark

Most of this thesis describes the building of models that were fitted to the various data sources. A model attempts to capture certain fundamental aspects of reality. However, the Earth's biota and environment are extraordinarily complex and due to our limited knowledge there is barely one aspect of their interactions that we can confidently describe with a mathematical equation (Lovelock, 1989). In the words of Cormack (1968): "Even the most general mathematical model is a plaything relative to the complexities of an animal population". It is likely that truth is high (effectively infinite) dimensional (Buckland, Burnham & Augustin 1997), such that, no matter how rich our data sources, we will only ever be able to approximate reality. Nevertheless, as more data are collected we should be able to build increasingly realistic models of biological systems. Model selection procedures enable us to assess which models are most appropriate, given the data at hand. Furthermore, the growing power of computers, along with the development of numerical model fitting

procedures, means that realistic models, such as SSMs, that are unsolvable analytically, can be used to gain a better approximation of reality.

References

Allen, PJ, Amos, W, Pomoroy, PP & Twiss, SD (1995) Microsatellite variation in grey seals (*Halichoerus grypus*) shows evidence of genetic differentiation between two British breeding colonies. *Molecular Ecology*, **4**, 653-662.

Amos, B, Twiss, S, Pomoroy, P & Anderson, S (1995) Evidence for mate fidelity in the grey seal. *Science*, **268**, 1897-1899.

Amos, W., Worthington Wilmer, J., Fullard, T., Burg, T.M., Croxall, J.P., Bloch, D. & Coulson, T. (2001a) The influence of parental relatedness on reproductive success. *Proceedings of the Royal Society of London Series B*, **268**, 2021-2027.

Amos, W., Worthington Wilmer, J., & Kokko, H. (2001b) Do female grey seals select genetically diverse mates? *Animal Behaviour*, **62**, 157-164.

Anderson, S.S., Burton, R.W., & Summers, C.F. (1975) Behaviour of grey seal (*Halichoerus grypus*) during the breeding season at North Rona. *Journal of Zoology, London*, **177**, 179-195.

Arnason, A.N. (1972) Parameter estimation for mark-recapture experiments on two populations subject to migration and death. *Researches in Population Ecology*, **13**, 97-113.

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

Baker, C.S. & Lento, G.M. (2002) Forensic genetic identification of marine mammals. *Encyclopedia of Marine Mammals* (eds W. Perrin, B. Wursig & J.G.M. Thewissen) pp. 450-453. Academic Press.

Balloux, F. & Lugon-Moulin, N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155-165.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5(4)**, 251-261.

Begon, M., Harper, J. L. & Townsend, C. R. (1996) *Ecology: Individuals, Populations and Communities 3rd Ed.*, Blackwell Science Ltd.

Boness, D.J. & Bowen, W.D. (1996) The evolution of maternal care in Pinnipeds: new findings raise questions about the evolution of maternal feeding strategies. *BioScience*, **46(9)**, 645-654.

Bonner, W.N. & Hickling, G. (1971) Grey seals at the Farne Islands. A management plan. *Natural Environment Research Council.*

Bonner, W.N. & Hickling, G. (1974) The grey seals of the Farne Islands 1971 to 1973. *Transactions of the Natural History Society of Northumbria*, **42**, 65-84.

Bowne, D.R. & Bowers, M.A. (2004) Interpatch movements in spatially structured populations: a literature review. *Landscape Ecology*, **19**, 1-20.

Boyd (2002) Pinniped life history. *Encyclopedia of Marine Mammals* (eds W. Perrin, B. Wursig & J.G.M. Thewissen) pp. 929-934. Academic Press.

Brown, W.D., Beck, C.A. & Austin, D.A. (2002) Pinniped ecology. *Encyclopedia of Marine Mammals* (eds W. Perrin, B. Wursig & J.G.M. Thewissen) pp. 911-921. Academic Press.

Brownell, R.L., Ralls, K. & Perrin, W.F. (1995) Marine mammal biodiversity. *Oceanus*, **38(2)**, 30-33.

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

Buckland, S.T., Burnham, K.P. & Augustin, N.H. (1997) Model selection: an integral part of inference. *Biometrics*, **53**, 603-618.

Buckland, S.T., Goudie, I.B.J. & Borchers, D.L. (2000) Wildlife population assessment: Past developments and future trends. *Biometrics*, **56**, 1-12.

Buckland, S.T., Newman, K.B., Thomas, L. & Kösters, N. (2004) State-space models for the dynamics of wild animal populations. *Ecological Modelling*, **171**, 157-175.

Burgman, M & Possingham, H (2000) Population viability for conservation: the good, the bad and the undescribed. *Genetics, Demography and Viability of Fragmented Populations* (eds G.M. Clark & A.G. Young), pp. 97-112. Cambridge University Press 2000.

Burnham, K.P. & Anderson, D.R (2002) *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach, 2nd Edition*, Springer-Verlag, New York.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C. & Pollock, K.H. (1987) Design and analysis methods for fish survival experiments based on release-recapture. *American Fisheries Society Monograph* **5**.

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

Campbell, NA (1996) *Biology, 4th Edition*, The Benjamin/Cummings Publishing Company, Inc.

Caswell, H. (1989), *Matrix Population Models: Construction, Analysis and Interpretation*, Sinauer Associates Inc.

Choquet, R., Reboulet, A.-M., Pradel, R., Giminez, O. & Lebreton J.-D. (2003) User's manual for U-CARE. Mimeographed document, CEFE/CNRS, Montpellier (ftp://ftp.cefe.cnrs-mop.fr/biom/Soft-CR/)

Choquet, R., Reboulet, A.-M., Pradel, R., Giminez, O. & Lebreton, J.-D. (2004) User's manual for M-SURGE 1.4. Mimeographed document, CEFE/CNRS, Montpellier (ftp://ftp.cefe.cnrs-mop.fr/biom/Soft-CR/)

Clobert, J., Ims, R.A. & Rousset, F. (2004) Causes, mechanisms and consequences of dispersal. *Ecology Genetics and Evolution of Metapopulations: Standard Methods for Inventory and Monitoring* (eds I. Hanski & O.E. Gaggiotti) pp. 307-335. Elsevier Academic Press, Inc.

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

Cormack, R.M. (1968) The statistics of capture-recapture methods. *Oceanography and Marine Biology*, **6**, 455-506.

Cornuet, J, Piry, S, Luikart, G, Estoup, A & Solignac, M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals, *Genetics*, **153**, 1989-2000.

Coulson, J.C. & Hickling, G. (1965) The grey seals of the Farne Islands. *Transactions of the Natural History Society of Northumbria*, **15**, 121-139.

Coulson, J.C. & Hickling, G. (1969) The grey seals of the Farne Islands. *Transactions of the Natural History Society of Northumbria*, **17**, 29-46.

Daugherty, C.H., Cree, A., Hay, J.M. & Thompson, M.B. (1990) Neglected taxonomy and continuing extinctions of tuatara (*Sphenodon*). *Nature*, **347**, 177-179.

Doucet, A., Frietas, N. & Gordon, N. (2001) *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, Berlin, Germany.

Doucet, A., Godsil, S.J. & Andrieu, C. (2000) On sequential Monte Carlo methods for Bayesian filtering. *Statistics and Computing*, **10(3)**, 197-208.

Duck, C.D. (2004) Pup production in the British grey seal population, *SCOS Briefing Paper 04/2*. In: Scientific advice on matters related to the management of seal populations: 2004. Reports of the UK Special Committee on Seals, available at http://smub.st-andrews.ac.uk

Dupuis (1995) Bayesian Estimation of Movement and Survival Probabilities from Capture-Recapture Data. *Biometrika*, **82(4)**, 761-772.

Ehrlich, P.R. & Hanski, I. (2004) Checkerspot research, background and origins. In: *On the wings of checkerspots: A model system for population biology*, pp. 3-16, Ehrlich, P.R. & Hanski, I (eds), Oxford University Press.

Ellison, A.M. (1996) An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecological Applications*, **6(4)**, 1036-1046.

Elmhagen, B & Angerbjörn, A (2001) The applicability of metapopulation theory to large mammals. *Oikos*, **94**, 89-100.

Excoffier, L (2001) Analysis of population subdivision. *Handbook of Statistical Genetics* (eds D.J. Balding, M. Bishop & C. Cannings) pp. 271-307.

Falush, D., Stephens, M. & Pritchard, J.K. (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, **164**, 1567-1587.

Fowler, C.W. (1981) Density dependence as related to life-history strategy. *Ecology*, **62(3)**, 602-610.

Fujiwara, M. & Caswell, H. (2002) Estimating population projection matrices from multi-stage mark-recapture data. *Ecology*, **83(12)**, 3257-3265.

Gaggiotti, OE, Jones, F, Lee, WM, Amos, W, Harwood, J & Nichols, RA (2002) Patterns of colonization in a metapopulation of grey seals. *Nature*, **416**, 424-427.

Gaggiotti, O.E., Brooks, W., Amos, W. & Harwood, J. (2004) Combining demographic, environmental and genetic data to test hypothesis about colonisation events in metapopulations. *Molecular Ecology*, **13(4)**, 811-825.

Gaggiotti, O.E., Lange, O., Rassmann, K., Gliddon, C. (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, **8**, 1513-1520.

Gamerman, D. (1997) *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*. Chapman & Hall.

Gelman, A, Carlin, J.B., Stern, H.S. & Rubin, D.B. (1995) *Bayesian Data Analysis*, Chapman & Hall.

Goudet, J. (1995) Fstat, Version 1.2: a computer program to calculate F-statistics. *Journal of Heredity*, **86**, 485-486.

Goudet, J. Raymond, M. Demeeus, T. & Rousset, F. (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933-1940.

Greenwood, P.J. (1980) Mating systems, philopatry and dispersal in birds and mammals. *Animal Behaviour*, **28**, 1140-1162.

Gutiérrez-Espeleta, GA, Kalinowski, ST & Hedrick, PW (2000) Genetic population structure in desert bighorn sheep: implications for conservation in Arizona. In: *Genetics, Demography and Viability of Fragmented Populations*, pp. 227-236, Clark, GM & Young, AG (eds.). Cambridge University Press 2000.

Hall, A. (2002) Grey seal (*Halichoerus grypus*). *Encyclopedia of Marine Mammals* (eds W. Perrin, B. Wursig & J.G.M. Thewissen) pp. 522-524. Academic Press.

Hammond, P.S., McConnell, B.J. & Fedak, M.A. (1993) Grey seals off the coast of Britain: distribution and movements at sea. *Symposium of the Zoological Society of London*, **66**, 211-224.

Hanski, I. (2004) Metapopulation theory, its use and misuse. *Basic and Applied Ecology*, **5**, 225-229.

Hanski, I. (2005) Landscape fragmentation, biodiversity loss and the societal response. The longterm consequences of our use of natural resources may be surprising and unpleasant. Embo reports, **6(5)**, 388-392.

Hanski, I. Alho, J. & Moilanen, A. (2000) Estimating the parameters of survival and migration of individuals in metapopulations. *Ecology*, **81(1)**, 239-251.

Hanski, I. & Gaggiotti, O.E. (2004) Metapopulation biology: past, present, and future. *Ecology Genetics and Evolution of Metapopulations: Standard Methods for Inventory and Monitoring* (eds I. Hanski & O.E. Gaggiotti) pp. 3-22. Elsevier Academic Press, Inc.

Hanski I., Kuussaari, M. & Nieminen, M. (1994) Metapopulation structure and migration in the butterfly *Melitaea cinxia*. *Ecology*, **75(3)**, 747-762.

Hanski, I & Simberloff, D (1997) The metapopulation approach, its history, conceptual domain, and application to conservation. *Metapopulation biology: ecology, genetics and evolution* (eds I. Hanski & M.E. Gilpin), pp. 5-26. Academic Press, Inc.

Hartl, D.L. (2000), *A Primer of Population Genetics, 3rd Edition*, Sinauer Associates Inc.

Harwood, J. (1981) Managing grey seal populations for optimum stability. *Dynamics of Large Mammal Populations* (eds C.W. Fowler & T.D. Smith), pp. 159-172. John Wiley & Sons, New York.

Harwood, J. (1984) Seals and fisheries. *Marine Pollution Bulletin*, **15**, 426-429.

Harwood, J., Anderson, S.S. & Curry, M.G. (1975) Branded grey seals at the Monach Isles, Outer Hebrides. *Journal of Zoology, London*, **180**, 506-508.

Harwood, J., & Prime, J.H. (1978) Some factors affecting the size of British grey seal populations. *Journal of Applied Ecology*, **15**, 401-411

Harwood, J. & Stokes, K. (2003) Coping with uncertainty in ecological advice: lessons from fisheries. *Trends in Ecology & Evolution*, **18(12)**, 617-622.

Hestbeck, J.B., Nichols, J.D. & Malecki, R.A. (1991) Estimates of movement and site fidelity using mark-resight data of wintering Canada geese. *Ecology*, **72(2)**, 523-533.

Hiby, L. & Duck, C. (unpublished). Estimates of the size of the British grey seal *Halichoerus grypus* population and levels of uncertainty.

Hiby, L. & Hammond, P.S. (1998) Monitoring grey seal populations using individual identification. Final report to MAFF under project MF0707.

Hiby, L & Lovell, P (1990) Computer aided matching of natural markings: A prototype system for grey seals. *Report of the International Whaling Commission* (Special Issue **12**), 57-61.

Hickling, G., Hawkey, P. & Harwood, L.H. (1976) The grey seals of the Farne Islands. The 1975 breeding season. *Transactions of the Natural History Society of Northumbria*, **42**, 107-114.

Hickling, G., Hawkey, P. & Harwood, L.H. (1978) The grey seals of the Farne Islands. The 1977 breeding season. *Transactions of the Natural History Society of Northumbria*, **43**, 25-34.

Hilborn, R. & Mangel, M. (1997) *The Ecological Detective: Confronting Models with Data*. Princeton University Press.

Hoffman, J.I. & Amos, W. (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, **14**, 599-612.

Inchausti, P. & Weimerskirch, H. (2002) Dispersal and metapopulation dynamics of an oceanic seabird, the wandering albatross, and its consequences for its responses to long-line fisheries. *Journal of Animal Ecology*, **71**, 765-770.

Jamieson, L. & Brooks, S. (2003) State space models for density dependence in population ecology, *Bayesian Statistics 7* (eds. M. Jose, M.J. Bernardo, A. Bayarri, P. Dawid, J.O. Berger, D. Heckerman, A.F.M. Smith & M. West), pp. 565-576. Oxford University Press.

Johnson J.B. & Omland, K.S. (2004) Model selection in ecology and evolution, *TRENDS in Ecology and Evolution*, **19(2)**, 101-108.

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

Jones, M.E., Paetkau, D., Geffen, E., & Moritz, C. (2004) Genetic diversity and population structure of Tasmanian devils, the largest marsupial carnivore. *Molecular Ecology*, **13**, 2197-2209.

King, R. and Brooks, S.P. (2002) Bayesian Model Discrimination for Multiple Strata Capture-Recapture Data. *Biometrika*, **89**, 785-806

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

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

Lindenmayer, D & Peakall, R (2000) The Tumult experiment – integrating demographic and genetic studies to unravel fragmentation effects: a case study of the native bush rat. In: *Genetics, Demography and Viability of Fragmented Populations*, pp. 173-202, Clark, GM & Young, AG (eds.). Cambridge University Press 2000.

Liu, J. & West, M. (2001) Combined parameter and state estimation in simulation-based filtering, *Sequential Monte Carlo Methods in Practice* (eds A. Doucet, N. de Freitas & N. Gordon), pp. 197-223. Springer-Verlag, Berlin, Germany.

Liu, J.S. (2001) *Monte Carlo Strategies in Scientific Computing*, Springer-Verlag, Berlin, Germany.

Liu, J.S., Chen, R. & Wong, W.H. (1998) Rejection Control and Sequential Importance Sampling, *Journal of the American Statistical Association*, **93**, 1022-1031.

Lovelock, J (1989) *The Ages of Gaia: A Biography of our Living Earth.* Oxford University Press.

Luikart, G & England, PR (1999) Statistical analyses of microsatellite DNA data, *Trends in Ecology & Evolution*, **14(7)**, 253-256.

MacArthur, R.H. & Wilson E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton, New Jersey, USA.

May, R.M. (1990) Taxonomy as destiny. Nature, **347**, 129-130.

McConnell, B.J., Chambers, C., Nicholas, K.S. & Fedak, M.A. (1992) Satellite tracking of grey seals (*Halichoerus grypus*). *Journal of Zoology*, **226**, 271-282

McConnell, BJ, Fedak, MA, Lovell, P & Hammond, PS (1999) Movements and foraging areas of grey seals in the North Sea, *Journal of Applied Ecology*, **36(4)**, 573-590.

Miglani, G.S. (2000) *Basic genetics*. Narosa Publishing House, London.

Millar R.B. & Meyer, R. (2000) Bayesian state-space modeling of age-structured data: fitting a model is just the beginning, *Canadian Journal of Fisheries & Aquatic Sciences*, **57(1)**, 43-50.

Moritz, C. (1994) Defining evolutionarily significant units for conservation, *Trends in Ecology & Evolution*, **9**, 373-375.

Newman, K.B., Buckland, S.T., Lindsey, S.T., Thomas, L., Fernández, C. (in press) Hidden process models for animal population dynamics, *Ecological Applications*

Newman, K.B., Fernández, C., Buckland, S.T. & Thomas, L. (in prep) Monte Carlo inference for state-space models for wild animal populations

Nichols, J.D. & Kendall W.L. (1995) The use of multi-state capture-recapture models to address questions in evolutionary ecology. *Journal of Applied Statistics*, **22**, 835-846.

O' Hara, B., Arjas, E., Toivonen, H. & Hanski, I. (2002) Bayesian analysis of metapopulation data. *Ecology*, **83(9)**, 2408-2415.

Palsbøll, PM (2002) Genetics, Overview. In: *Encyclopaedia of Marine Mammals*, pp. 495-504, Perrin, WF, Würsig, B & Thewissen, JGM (eds.). Academic Press.

Parker, PG, Snow, AA, Schug, MD, Booton, GC & Fuerst, PA (1998) What molecules can tell us about populations: choosing and using a molecular marker, *Ecology*, **79(2)**, 361-382.

Paetkau, D., Calvert, W., Stirling, I. & Strobeck, C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, **4**, 347-354.

Paetkau, D., Slade, R., Burden, M. & Estoup, A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55-65.

Petit, E., Balloux, F. & Goudet, J. (2001) Sex-biased dispersal in a migratory bat: a characterisation using sex-specific demographic parameters. *Evolution*, **55(3)**, 635-640.

Petit, E. & Mayer, F. (1999) Male dispersal in the noctule bat (*Nyctalus noctula*): where are the limits? *Proceedings of the Royal Society of London, Series B*, **266**, 1717-1722.

Piry, S., Alapetite, A., Cornuet, J.M., Paetkau, D., Baudouin, L. & Estoup, A. (2004) GENECLASS2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity*, **95**, 536-539.

Pitt, M.K. & Shepherd, N. (1999) Filtering via simulation: auxiliary particle filters, *Journal of the American Statistical Association*, **94**, 590-599.

Pomeroy, PP, Anderson, SS, Twiss, SD & McConnell, BJ (1994) Dispersion and site fidelity of breeding female grey seals (*Halichoerus grypus*) on North Rona, Scotland, *Journal of the Zoology, London*, **233**, 429-447.

Pomeroy, P.P., Twiss, S.D. & Redman, P. (2000) Philopatry, site fidelity and local kin associations within grey seal breeding colonies. *Ethology*, **106**, 899-919.

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

Prime, J.H. (1981) Breeding grey seals on the Isle of May, 1980. *Transactions of the Natural History Society of Northumbria*, **47**, 13-16.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Queller, D.C. & Goodnight K.F. (1989) Estimating Relatedness Using Genetic Markers, *Evolution*, **48(2)**, 258-275.

Quinn, T.J. II & Deriso, R.B. (1999) *Quantitative Fish Dynamics*, Oxford University Press, New York.

R Development Core Team (2004) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org.

Raymond, M. & Rouset, F. (1995) An exact test for population disfferentiation. *Evolution*, **49**, 1280-1283.

Rousset, F. & Raymond, M. (1997) Statistical analysis of population genetic data: new tools, old concepts. *Trends in Ecology & Evolution*, **12**, 313-317.

Ruxton, G.D. & Rohani, P. (1998) Fitness-dependent dispersal in metapopulations and its consequences for persistence and synchrony, *Journal of Animal Ecology*, **67**, 530-539.

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

SCOS (2002), Scientific advice on matters related to the management of seal populations: 2004. Reports of the UK Special Committee on Seals, available at http://smub.st-andrews.ac.uk

SCOS (2004) Scientific advice on matters related to the management of seal populations: 2004. Reports of the UK Special Committee on Seals, available at http://smub.st-andrews.ac.uk

Seber, J.A.F. (1965) A note on the multiple recapture census. *Biometrika*, **52**, 249-259.

Serrano, D. & Tella, J.L. (2003) Dispersal within a spatially structured population of lesser kestrels: the role of spatial isolation and conspecific attraction. *Journal of Animal Ecology*, **72**, 400-410.

Sherwin, WB & Moritz, C (2000) Managing and monitoring genetic erosion. In: *Genetics, Demography and Viability of Fragmented Populations*, pp. 9-34, Clark, GM & Young, AG (eds.). Cambridge University Press 2000.

Sinclair, A. (1996) Mammal populations: fluctuation, regulation, life history theory and their implications for conservation. *Frontiers of Population Ecology* (eds R.B. Floyd, W. Sheppard & P.J. De Barro), pp 127-154. CSIRO, Melbourne, Australia.

Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787-792.

Smith A.T. & Peacock M.M. (1990) Conspecific attraction and the determination of metapopulation colonization rates. *Conservation Biology*, **4(3)**, 320-323.

Spendelow, J.A., Nichols J.D., Nisbet, I.C.T., Hayes, H. & Cormons, G.D. (1995) Estimating annual survival and movement rates of adults within a metapopulation of roseate turns. *Ecology*, **76(8)**, 2415-2428.

Stacy, P.B., Johnson, V.A. & Taper, M.L. (1997) Migration within metapopulations: the impact upon local populations. *Metapopulation biology: ecology, genetics and evolution* (eds I. Hanski & M.E. Gilpin), pp. 267-292. Academic Press, Inc

Stevick, P.T., Palsbøll, P.J., Smith, T.D., Bravington, M.V. & Hammond, P.S. (2001) Errors in identification using natural markings: rates, sources, and effects on capture-recapture estimates of abundance. *Canadian Journal of Fisheries & Aquatic Sciences*, **58**, 1861-1870.

Ter Braak, C.J.F. & Etienne, R.S. (2003) Improved Bayesian analysis of metapopulation data with an application to a tree frog metapopulation. *Ecology*, **84(1)**, 231-241.

Thomas, L., Buckland, S.T., Newman, K.B. & Harwood, J. (2005) A Unified Framework for Modelling Wildlife Population Dynamics. *Australian & New Zealand Journal of Statistics*, **47(1)**, 19-34.

Thomas, L. & Harwood, J. (2003) Estimating grey seal population size using a Bayesian state-space model, *SCOS Briefing Paper 03/3*. In: Scientific advice on matters related to the management of seal populations: 2003. Reports of the UK Special Committee on Seals, available at http://smub.st-andrews.ac.uk

Thomas, L. & Harwood, J. (2004a) A comparison of grey seal population models incorporating density dependent pup survival and fecundity, SCOS Briefing Paper 04/6. In: Scientific advice on matters related to the management of seal populations: 2004. Reports of the UK Special Committee on Seals, available at http://smub.st-andrews.ac.uk

Thomas, L. & Harwood, J. (2004b) Possible impacts on the British grey seal population of deliberate killing related to salmon farming, SCOS Briefing Paper

04/7. In: Scientific advice on matters related to the management of seal populations: 2004. Reports of the UK Special Committee on Seals, available at http://smub.st-andrews.ac.uk

Thompson, D, Hammond, PS, Nicholas, KS & Fedak, MA (1991) Movements, diving and foraging behaviour of grey seals (*Halichoerus grypus*). *Journal of Zoology*, **224**, 223-232.

Thrall P.H., Burdon, J.J. & Murray, M.J. (2000) The metapopulation paradigm: a fragmented view of conservation biology. In: *Genetics, Demography and Viability of Fragmented Populations*, pp. 77-95, Clark, GM & Young, AG (eds.). Cambridge University Press 2000.

Trenkel, V.M., Elston, D.A. & Buckland S.T. (2000) Fitting Population Dynamics Models to Count and Cull Data Using Sequential Importance Sampling. *Journal of the American Statistical Association,* **95(450)**, 363-374.

Twiss, S.D., Duck, C. & Pomeroy, P.P. (2003) Grey seal (*Halichoerus grypus*) pup mortality not explained by local breeding density on North Rona, Scotland. *Journal of Zoology*, **259**, 83-91.

Twiss, D.S., Pomeroy, P.P. & Anderson, S.S. (1994) Dispersion and site fidelity of breeding male grey seals (*Halichoerus grypus*) on North Rona, Scotland. *Journal of the Zoology, London*, **233**, 683-693.

Van de Casteele, T., Galbusera, P. & Matthysen, E. (2001) A comparison of microsatellite-based pairwise relatedness estimators. *Molecular Ecology*, **10**, 1539-1549.

Viallefont A., Lebreton, J.-D., Reboulet, A,-M & Gory, G. (1998) Parameter identifiability and model selection in capture-recapture models: a numerical approach. *Biometrical Journal*, **40(3)**, 313-325.

Wade, P.R. (2000) Bayesian methods in conservation biology. *Conservation Biology*, **14(5)**, 1308-1316.

Waser, PM & Strobeck, C (1998) Genetic signatures of interpopulation dispersal, *Trends in Ecology & Evolution*, **13**, 43-44.

Watt, G. (1951) *The Farne Islands: Their history and wildlife*. Country Life Ltd.

Weir, B.S. (1996) *Genetic Data Analysis II*, Sinauer Associates, Inc.

Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358-1370.

White, G.C. & Burnham K.P. (1999) Program MARK: Survival estimation from populations of marked animals. *Bird Study*, **46**, 120-138.

Whitlock, M.C.& McCauley, D.E (1999) Indirect measures of gene flow and migration: FST≠1/(4Nm+1). *Heredity*, **82**, 117-125.

Worthington Wilmer, J, Allen, PJ, Pomeroy, PP, Twiss, SD & Amos, W (1999) Where have all the fathers gone? An extensive microsatellite analysis of paternity in the grey seal (*Halichoerus grypus*). *Molecular Ecology*, **8**, 1417-1429.

Worthington Wilmer, J, Overall, AJ, Pomeroy, PP, Twiss, SD & Amos, W (2000) Patterns of paternal relatedness in British grey seal colonies, *Molecular Ecology*, **9**, 283-292.

Appendix 1

Comparing photo-ID sampling strategies

Additional photo-ID data for the four North Sea colonies studied in Chapter 4 would provide further information on movement rates between colonies and could allow the estimation of an independent survival rate for seals from these colonies. However, there are a number of logistic and financial considerations that need to be considered before undertaking such an exercise. As more photographs are taken at a colony the likelihood of re-photographing the same animal increases. Since there is a fixed cost associated with processing each photograph, any additional photography increases total cost.

The option of making more than one visit to a colony separated by a reasonable length of time may reduce the number of times the same animal is photographed, but this must be set against the cost of the additional visits. Furthermore, if the colony is only visited during the peak of pupping, females who consistently breed either early or late in the season will not be available for capture, thus adding a potential source of bias to the capture-recapture parameter estimates.

In this appendix I compare different sampling protocols using simulation modelling. The main aim was to determine the optimal sampling strategy to collect a large number of photographs of different animals within a single breeding season. The

models were parameterised with data on pup production at the Isle of May in 2003 and with other information on the breeding biology of this species.

A1.1 Simulation model input

Female grey seals are thought to arrive at a colony 1-3 days before they give birth and they tend to stay for approximately 25 days, with a standard deviation of 5 days. On the Isle of May in 2003 the pup production estimate was 1882. For simplicity in the simulations it is assumed that 1882 females used the colony in this year. The estimated birth curve parameters (Duck, pers. comm.) from the model used to estimate pup production for the Isle of May in 2003 (Duck 2004), were:

day 1 = 1st October
start date = 20 August
mean pupping date = 28 October
standard deviation = 7.92

The birth curve is assumed to have a Lognormal distribution, so it is necessary to convert the mean and standard deviation for use with this distribution (Newman, pers. comm.). Let us label the mean and variance on the log scale a and b respectively, and the equivalent values on the linear scale as m and v. Then:

$$b = \log(\frac{v}{m^2} + 1)$$

$$a = \log(m) - \frac{b}{2}$$

By substituting $m = 69.2$ and $v = 7.92^2$ into these equations we can calculate a and b.

A1.2 Simulation model structure

Combining the above information I simulated 400 populations. For each population I calculated an arrival date and a departure date for each of the 1882 females in the populations as follows:

$$d_{arrive,i} = x_{1,i} - x_{2,i}$$
$$d_{depart,i} = d_{arrive,i} + x_{3,i}$$

where $d_{arrive,i}$ is the arrival date for female i, where i = 1,...,1882, and $d_{depart,i}$ is the departure date for female i. For the three x values we sample one deviate respectively from each of the following distributions:

$$x_{1,i} \sim \text{Lognormal}(a, b)$$
$$x_{2,i} \sim \text{Uniform}(1, 3)$$
$$x_{3,i} \sim \text{Normal}(25, 5)$$

As each female's head has two sides that can be photographed, the total number of unique photographs possible during the simulations are 3764 (= 2*1882). Female seals are photographed with replacement during the simulations.

In the simulations I explored four different sampling protocols: (1). Peak day sampling, done on the day when most females were present (this value was calculated from the simulations); (2). Two samples were taken 26 days apart centred on the peak day; (3). Two samples were taken 30 days apart; (4). Two samples were taken 40 days apart. Under each protocol 500, 750 and 1000 photographs were taken of randomly chosen animals with replacement. For protocols 2-4, half of the photographs were taken on the first date and half on the second. The number of unique photographs taken under each scenario was recorded, as was the total number of breeding females that had not been photographed on either side. For protocols 2-4, the number of animals that arrived and left between the two sampling events and the number that were present on both sampling events were recorded.

A1.3 Results

The results from the simulations are shown in Table A1.1. There was little Monte Carlo error in these results (i.e. repeat runs of the simulation model produced practically identical results). The date when the largest numbers of females were present was on 7 November (ten days after the date when most pups were born). The 95% confidence interval on this value was ± 1 day.

Table A1.1 Statistics estimated from the simulation model comparing different photo-identification sampling protocols.

Statistic†	Mean and 95% CI
1. PEAK SAMPLING	
unique500	463 (452, 474)
unique700	671 (655, 686)
unique1000	862 (844, 881)
already left	119 (83, 163)
not yet arrived	134 (93, 172)
2. TWO SAMPLES 26 DAYS APART	
unique500	463 (452, 473)
unique750	669 (654, 683)
unique1000	859 (838, 878)
already left	2 (0, 5)
missed in middle	168 (142, 196)
not yet arrived	6 (1, 12)
overlappers	170 (145, 198)
3. TWO SAMPLES 30 DAYS APART	
unique500	459 (448, 470)
unique750	660 (645, 676)
unique1000	845 (825, 864)
alreadyleft	1 (0, 3)
missed in middle	394 (354, 428)
not yet arrived	3 (0, 7)
overlappers	48 (32, 63)
4. TWO SAMPLES 40 DAYS APART	
unique500	423 (407, 438)
unique750	588 (564, 609)
unique1000	725 (695, 753)
alreadyleft	0 (0, 1)
missed in middle	1108 (1068, 1148)
not yet arrived	1 (0, 3)
overlappers	0 (0, 1)

† All the statistics starting with the word "unique" are out of 3764 (i.e. the total number of unique photographs possible), whereas the other statistics are out of 1882 (i.e. the total number of females breeding on the Isle of May). The number following the word "unique" gives the total number of photographs taken. (e.g. unique500 when peak sampling was done represents the statistic for the number of unique photographs achieved when 500 photographs were taken on the peak day).

From Table A1.1 it appears that protocol 1 provides the greatest proportion of unique photographs. For this protocol both the proportion of females that bred early in the season and have left already (~ 0.06) and the proportion that bred late and had not yet arrived (~ 0.07) were relatively small. Protocol 2 reduces these proportions but results in some animals (~ 0.09) that arrive in the middle of the season being missed. Approximately 9% of the females present on the first visit were also present on the second. Increasing the number of days between the sampling events greatly reduces the proportion of seals that overlap the two sampling events but substantially increases the proportion of the seals from the middle of the season that are missed.

Protocols 1 and 2 yielded rather similar numbers of unique photographs; protocols 3 and 4 were less effective. According to the simulation results, taking 1000 photographs on the peak day provides 862 unique photographs, equivalent to a capture probability of ~ 0.23 for each side of the head.

A1.4 Discussion

The higher the capture probability in capture-recapture models the greater the precision with which all the model parameters can be estimated. Thus it is important to maximize the number of unique animals identified on each sampling occasion.

There appear to be significant benefits to only visiting each colony once. Two visits will be twice as expensive as one, and weather conditions may not allow two visits within a season. Some animals that bred early and late in the season are not available for sampling when visiting on the peak day, but these are a relatively small proportion of the population. However, females that consistently pup early or late in the season will be missed. If migrant animals are more likely to pup at these times, then migration rates may be underestimated. This potential bias could be avoided by making two visits to the colony, but the additional extra cost and logistic difficulties involved in making more than one visit are hard to justify because there would be no increase in the number of unique photographs collected.

Appendix 2

Extending the State-Space Modelling to Incorporate the Photo-ID Data

This appendix details a suggestion for how to extend the state-space models (SSMs) developed in chapter 5 to simultaneously fit to both the pup production and the capture-recapture photo-ID data for the grey seal (as discussed in section 6.4 of chapter 6), taking into consideration photographs of both sides of the seals head and the possibility of false negative errors.

For simplicity I only consider the formulation specifically necessary for the capture-recapture process and observation models of the SSM. I give suggested formulations for how to incorporate photographs of both sides of the seal's head, thus accounting for instances where the two sides of the head are matched in the field. I also consider a method for dealing with false negative errors (whereby previously photographed animals are not recognized when they are photographed again). I assume that error of the opposite nature – false positives (whereby two different animals are thought to represent the same animal) – are negligible due to the conservative nature of the matching process, as detailed in section 4.2.4 of chapter 4. I only present the single-site case and just the first two time points are considered, thus keeping the formulations in a presentable form. I also assume 100% survival and no recruitment into the breeding population between the first and second time period for ease of explanation.

The length of the state and observation vectors increase with time due to the increasing number of permutations of capture histories possible. The observation vector is a subset of the state vector and in parts is a convolution of more than one element from the state vector as detailed below. The state vectors for the first two time periods are as follows:

$$\mathbf{n}_1 = \begin{bmatrix} n_{L,1} \\ n_{R,1} \\ n_{B,1} \\ n_{0,1} \end{bmatrix} \rightarrow \mathbf{n}_2 = \begin{bmatrix} n_{LL,2} \\ n_{LR,2} \\ n_{LB,2} \\ n_{L0,2} \\ n_{RL,2} \\ n_{RR,2} \\ n_{RB,2} \\ n_{R0,2} \\ n_{BL,2} \\ n_{BR,2} \\ n_{BB,2} \\ n_{B0,2} \\ n_{0L,2} \\ n_{0R,2} \\ n_{0B,2} \\ n_{00,2} \end{bmatrix}$$

Where the states for the first time period, $n_{L,1}, n_{R,1}$ and $n_{B,1}$, are the numbers of females photographed from the LHS only, the RHS only and both sides during the first capture event, and the number of females not captured is given by:

$$n_{0,1} = N_1 - n_{L,1} - n_{R,1} - n_{B,1}$$

where N_1 is the "true" number of females present in the population when the first capture event occurs. There is pup production data available before the first capture event, so fitting to both pup production and photo-ID data sets simultaneously means

a value for N_1 will be estimated for each simulated population. The increase in the number of elements in the state vector for the second time point is apparent where, for example, $n_{LL,2}$ represents the number of animals seen by the LHS on both the first and second capture occasions. The number of distinct elements of $\mathbf{n}_t$ at each time point is 4^t.

The states for the second time period are computed through the capture-recapture process model given below. Only the elements for the first four and last four states are shown, the other elements being implied by the dots (M).

$$\begin{gathered}\left(n_{LL,2}, n_{LR,2}, n_{LB,2}, n_{L0,2}\right) \sim \text{Multinomial}\left(n_{L,1}, p_L, p_R, p_B, (1-p_L-p_R-p_B)\right) \\ \text{M} \\ \left(n_{0L,2}, n_{0R,2}, n_{0B,2}, n_{00,2}\right) \sim \text{Multinomial}\left(n_{0,1}, p_L, p_R, p_B, (1-p_L-p_R-p_B)\right)\end{gathered}$$

The recapture probabilities, p_L, p_R and p_B, can be constant or time varying and/or a function of covariates (as in chapter 4). The hypothesis that $p_L = p_R$ can also be tested.

The elements in the observation vector in the first time period are simply:

$$\mathbf{y}_1 = \begin{bmatrix} n_{L,1} \\ n_{R,1} \\ n_{B,1} \end{bmatrix}$$

In the second time period, however, the elements in the observation vector are somewhat more complex and convoluted: (1) there are false negative errors to be incorporated and (2) individual animals seen by one side in the first year, and the other side in the second year, are indistinguishable from those animals seen in the second year by one of the sides but missed altogether in the first year and those animals whose true matches were missed. The elements in the observation vector in the second time period are therefore given by:

$$\mathbf{y}_2 = \begin{bmatrix} y_{LL,2} \sim \text{Binomial}\left(n_{LL,2}, (1-f)\right) \\ y_{LB,2} \sim \text{Binomial}\left(n_{LB,2}, (1-f)\right) \\ y_{RR,2} \sim \text{Binomial}\left(n_{RR,2}, (1-f)\right) \\ y_{RB,2} \sim \text{Binomial}\left(n_{RB,2}, (1-f)\right) \\ y_{BL,2} \sim \text{Binomial}\left(n_{BL,2}, (1-f)\right) \\ y_{BR,2} \sim \text{Binomial}\left(n_{BR,2}, (1-f)\right) \\ y_{BB,2} \sim \text{Binomial}\left(n_{BB,2}, (1-f^2)\right) \\ y_{0L,2} = n_{0L,2} + n_{RL,2} + \left(n_{LL,2} - y_{LL,2}\right) + \left(n_{BL,2} - y_{BL,2}\right) \\ y_{0R,2} = n_{0R,2} + n_{LR,2} + \left(n_{RR,2} - y_{RR,2}\right) + \left(n_{BR,2} - y_{BR,2}\right) \\ y_{0B,2} = n_{0B,2} + \left(n_{LB,2} - y_{LB,2}\right) + \left(n_{RB,2} - y_{RB,2}\right) + \left(n_{BB,2} - y_{BB,2}\right) \end{bmatrix}$$

where f represents the probability of making a false negative error. f^2 is used in the calculation for $y_{BB,2}$ as to make a false negative error in this instance requires that both the LHS match and the RHS match are missed. An independent estimate of f would be extremely useful for defining a prior distribution on this parameter. For the final three rows in the above vector the parts in brackets represent those animals whose true matches were missed.

SSMs of the kind detailed above, fitting to the pup production and capture-recapture data simultaneously, were not explored in this thesis due to time limitations. Incorporating photographs of both sides of the head and false negative errors is clearly quite complicated – a recommendation for exploring such models would be to first fit to the pup production and capture-recapture data for one of the North Sea colonies, ignoring the other three in the system. When such a model has been suitably developed and explored, extending to the multisite case, although significantly more convoluted, should be possible.