**Colonization, gene flow, and range
expansion in a pioneering seabird species**

A Dissertation Presented

by

Rachael W. Herman

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Ecology and Evolution

Stony Brook University

September 2023

Stony Brook University

The Graduate School

Rachael W. Herman

We, the dissertation committee for the above candidate for the
Doctor of Philosophy degree, hereby recommend
acceptance of this dissertation.

Heather Lynch – Dissertation Advisor
IACS Endowed Chair of Ecology and Evolution
Department of Ecology and Evolution / IACS

Krishna Veeramah – Chairperson of Defense
Associate Professor
Department of Ecology and Evolution

Liliana Davalos – Internal Committee Member
Professor
Department of Ecology and Evolution

Sushma Reddy – External Committee Member
Associate Professor
Bell Museum of Natural History
University of Minnesota

This dissertation is accepted by the Graduate School

Celia Marshik
Dean of the Graduate School

ii

Abstract of the Dissertation

**Colonization, gene flow, and range
expansion in a pioneering seabird species**

by

Rachael W. Herman

Doctor of Philosophy

in

Ecology and Evolution

Stony Brook University

2023

Species' range shifts and expansions are ongoing and accelerated by rapid environmental changes. While many shifts are attributed to the last deglaciation period approximately 10,000 years ago, current climate disruptions are intensifying these patterns. Warming climates tent to push species towards higher latitudes, where new habitat become available. Gentoo penguin (*Pygoscelis papua*) populations on the Western Antarctic Peninsula (WAP) have exploded under climate change and their range has expanded southward 60 km over just 13 years, an incredible feat for an historically sub-Antarctic, colonial species considered highly site faithful. These recently colonized gentoo penguin breeding locations have exhibited rapid population growth that could not be sustained by local recruitment alone. Using a combination of age-structured population models and robust time series, we implemented a rejection-based Approximate Bayesian Computation (ABC) modelling approach to estimate the amount of immigration required to sustain the rapid population growth exhibited by recently established colonies. Results indicated that a continued and an even increasing number of immigrants is required for many years, suggesting a paradoxical and intricate pattern of dispersal and migration in a species historically known to be site faithful.

Continuing our investigation, we conducted a fine-scale population genomics study to explore historical and recent gentoo penguin dispersal along the WAP. We inferred a stepping-stone pattern from the South Shetland Islands led to colonization of the Palmer Archipelago and

then the mainland WAP. Recent southward expansion mirrored this dispersal pattern, with some post-divergence gene flow from colonies on the Palmer Archipelago. Genetic diversity appeared to be maintained across colonies during the historical dispersal process, and range edge populations are still growing. This suggests continued gene flow and high numbers of migrants provided a buffer against founder effects typically expected in the classic stepping-stone model.

Finally, we built an agent-based model to forecast continued gentoo penguin colonization on the WAP. We informed this custom model using remote sensing assets to explore terrestrial and oceanographic characteristics, results from our ABC model, and the fine-scale mapping of gentoo penguin genomics described above. We present a 20-year forecast of potential future colonization events and continued range expansion.

Dedication Page

For my family - Mary Alice, Franklin, and Jared.

Frontispiece

Gentoo penguin colony – Neko Harbor, Western Antarctic Peninsula, December 2014.

<div align="center">

Table of Contents

</div>

Introduction

Chapter 1: Update on the global abundance and distribution of breeding gentoo penguins

Chapter 2: Age-structured model reveals prolonged immigration is key for colony establishment in gentoo penguins

Chapter 3: Whole genome sequencing reveals stepping-stone dispersal buffered against founder effects in a range expanding seabird

Table of Contents

Chapter 4: Forecasting colonization events and range expansion of gentoo penguins along the Western Antarctic Peninsula using agent-based modeling

List of Figures

List of Figures

Acknowledgments

This dissertation is the result of many years of adventure, teamwork, support, and mentorship. First and foremost, I would like to thank my advisor, Heather Lynch, who I have learned so much from over the past six years. Heather convinced me to continue on from my MS work and apply to do a PhD in her lab even though I was apprehensive about my lack of quantitative skills. She told me, and I quote, "I can teach anybody statistics, but I can't teach them how to be field biologists". I never forgot that.

I want to thank all of my lab mates, past and present, for your unyielding support and comradery over the years. I would also like to thank my PhD cohort, who became my close friends and academic family for life. Julia, José, Anna, and Ken - you four were my foundation and kept me going, especially during those early years. Casey, Shyamie, and Bilgecan – you three were also huge parts of that foundation, both in academia and friendship. I want to thank my committee members, Krishna Veeramah, Liliana Davalos, and Sushma Reddy. The mentorship and support I have received from you over these years has been priceless. Krishna, thank you for teaching me how to be a population geneticist, inspiring me to continue to pursue penguin pop gen, and providing me with your unwavering humor over Slack messages. Liliana, thank you for supporting me with words of encouragement and understanding during my crippling back injury. Sushma, thank you for giving me one of the best field seasons in Antarctica and being a strong mentor and inspiration for me. I would like to thank Tom Hart for his tireless collaboration and enthusiasm on all things penguin related over the years.

I gratefully acknowledge Quark Expeditions, National Geographic Lindblad Expeditions, Hurtigruten Expeditions, Skip Novak and Pelagic Expeditions, and the crew of the Vinson of Antarctica (Alec Hazel, Kenneth Perigón, and Jose Gritti) for providing logistical support for field work in Antarctica. I want to thank Kerry Reid for providing wet lab training and guidance in the Veeramah lab. I want to thank Jane Younger, Gemma Clucas, Bryce Robinson, Katie O'Brien, Irby Lovette and Jacob Cooper for helping to collect so many genomic samples in the field. I want to thank John Bates for believing in me as a researcher and making me feel like I belong among the world of ornithologists. I want to thank Ron Naveen for providing permitting for sample collection during the 2019/2020 field season (Antarctic Conservation Act Permit No. 2019-0001) and logistical support over the years. Sampling during the 2021/2022 season was permitted through the UK Foreign and Commonwealth Office (Permit No. 20-2021-22). This work was funded by the IACS STRIDE Fellowship, NASA FINESST Grant Award No. 80NSSC20K1605, the Stony Brook University Faculty-Staff Dissertation Fellowship, the Institute for Advanced Computational Sciences Endowment for Ecology & Evolution, the National Geographic Society, and the Field Museum's Robert A. Pritzker Center of Meteoritics and Polar Studies. I want to thank Melissa Cohen, Martha Nolan, Donna DiGiovanni, Tamara Gregorian, and Chelsey Dollinger for being the best administrators I have ever had the pleasure of working with at any university or organization. Thank you all for keeping me on track and navigating the bureaucratic tape for me.

I want to thank all of my friends near and far who have continued to support me and provide me with laughs over these last six years. Thank you for creating a home here for me in NYC and allowing me space to let loose while I worked on my dissertation. Bryan, Josephine, Pamela, M&L, Simone, and Claire – you are my rocks. Lastly, I want to thank my mom, dad, and brother for always supporting me and never letting me give up on my dreams. None of this would have been possible without you. I hope I can bring you all to Antarctica with me some day. I love you so much.

Introduction

0.1 Background

Gene flow, the exchange of alleles among populations through migration, maintains genetic variation within populations of a species (Lenormand 2002). Range expansion is the process by which species spread to new locations, and this can influence population genetic structure within a species. Range expansion and gene flow both allow for organisms to respond and adapt to climate change by permitting colonization of new locations and maintaining genetic diversity (Davis et al. 2001; Chen et al. 2011; Kremer et al. 2012), while gene flow can also facilitate genetic rescue for populations at risk of extirpation (Frankham 2015; Whiteley et al. 2015). However, it remains unclear how the propensity for range expansion and gene flow varies among organisms, and whether in most groups it will be sufficient to respond to current environmental change.

Seabirds are long-lived, colony-nesting bird species that forage primarily in marine environments. Seabirds are commonly considered to include Procellariiformes (tubenoses), Sphenisciformes (penguins), Pelecaniformes, and some of Charadriiformes (skuas, gulls, terns, and auks), a polyphyletic assemblage of lineages united by their similar life histories (Hackett et al. 2008, Prum et al. 2015). Most pelagic seabirds are characterized by strong long-distance flight abilities (or in the case of penguins, long distance swimming abilities), resulting in wide foraging distributions (Pinaud and Weimerskirch 2007; Thaxter et al. 2012). Their dispersive natures may reduce opportunities for population isolation and allow for rapid range expansion or dynamic distributional changes. However, many seabird species are highly philopatric (Coulson 2002, 2016). The tendency to breed at their natal nesting grounds may restrict seabirds' dispersal and ability to expand. Philopatry may also reduce gene flow between populations (Abbott and Double 2003a, 2003b), which can enable genetic drift to facilitate the loss of genetic variation and lead to differentiation. Physical barriers like landmasses (Morris-Pocock et al. 2011, 2016), ocean currents and the distribution of food sources (Gómez-Díaz et al. 2009), and information barriers such as unwillingness to colonize islands that lack conspecifics, may further limit seabird dispersal and range expansion (Forbes and Kaiser 1994). Given the opposing impacts of dispersal and

philopatry, it is unclear whether seabirds might be particularly prone or averse to both range expansion and gene flow among populations.

Mechanisms of Range Expansion

Although there are many barriers to dispersal and gene flow in seabirds, dispersal events do occur, resulting in new breeding colonies and range expansions. One cause for dispersal and expansion is density-dependent decline in habitat quality of current breeding locations that results in individuals relocating to new, unused habitat and establishing new colonies. A long-term study on the formation and growth of new black-legged kittiwake (*Rissa tridactyla*) colonies by Kildaw et al. (2005) found that kittiwakes formed new colonies during periods of low habitat quality on established colonies characterized by low productivity, large colony size, and declining population trends. Results from a population growth model indicated that some established breeders most likely relocated from old to new colonies despite high site-fidelity due to philopatry. In addition, modeling of population growth in Audouin's gulls (*Larus audouinii*) revealed large-scale immigration as a major source of individuals to newly established colonies, suggesting that established breeders were seeking out more suitable breeding habitat relative to their natal sites (Oro and Ruxton 2001). This mechanism is likely to be an important mediator of seabird range dynamics as climate change reduces the quality of established breeding areas.

Another possible reason for dispersal and range expansion in seabirds is the availability of newly suitable habitat due to environmental change. For example, many colonial seabirds have successfully colonized artificial islands created recently by humans (Leberg et al. 1995; Erwin et al. 2003). Competitive release from a conspecific in breeding or foraging habitat might also open up new breeding areas. Historic and modern samples of Yellow-eyed penguins (*Megadyptes antipodes*) revealed that this species expanded its range to the New Zealand mainland from Campbell and Auckland Islands only in the last few hundred years following the extinction of its previously unrecognized sister species, *M. waitaha*, after Polynesian settlement in New Zealand (Boessenkool et al. 2009). Although this is only one documented case of seabird range expansion due to loss of a competitor in the literature, it is possible that this has occurred among other species.

Post colonization: what sustains a new colony?

In many species, a single pair of birds typically pioneers colonization of a new breeding area, and a baseline number of pairs are required for establishment (Coulson 2002). Once colonization occurs, the recruitment mechanism by which the new populations increase can vary. In some species natal recruitment appears to take over after the pioneering individuals have established. Australasian gannets (*Morus serrator*) in Port Phillip Bay, Victoria, colonized an artificial structure triggered by dispersing individuals from a nearby colony, but switched predominantly to natal recruits within five years of establishment, with more than 80% of the 426 adults consisting of natal recruits (Pyk et al. 2012).

In contrast, other species show a high dependency on continued migrants in order to sustain a population, similar to a sink population dynamic. For example, Audouin's gulls (*Ichthyaetus audouinii*) colonized a new breeding location in the Mediterranean and the growth of the colony over the course of 15 years required large-scale immigration of individuals from other breeding locations (Oro and Ruxton 2001). Audouin's gulls immigrated to the new colony during times of high breeding failure at other established colonies, suggesting that adult breeders are relocating. Wedge-tailed Shearwaters (*Ardenna pacifica*) demonstrated a similar pattern of range expansion off the coast of Western Australia in the 1990s (Dunlop 2009). Populations nesting on islands off the NW coast colonized new islands further south than their previously southern range limit. Rapid growth of this population coincided with a period of poor breeding performance in the northern colonies attributed to declines in prey availability in this region due to El Niño. Common noddies (*Anous stolidus*) exhibited the same pattern of relocation off the coast of Western Australia in response to El Niño effects, although net immigration stopped approximately thirteen years after colonization, during which natal recruitment took over. (Dunlop 2009). More data are needed, however, on the degree to which new populations are self-sustaining versus dependent on continued immigration.

Metapopulations and Range Expansion

A metapopulation can be defined as an assemblage of local populations that disperse within a network of habitat patches, or subpopulations (Levins 1969). The breeding distributions of seabird species inherently mirror this patchy network because they nest on multiple islands and discrete colonies separated by the ocean. Indeed, metapopulation dynamics have been used to

explain connectivity among colonies in numerous seabird species (Buckley and Downer 1992; Inchausti and Wimmerskirch, 2002 Akçakaya et al. 2003; Kildaw et al. 2008; Oro 2003).

Group adherence and mate fidelity may limit the level of dispersal among subpopulations, as individuals will be less inclined to disperse to other habitats. In addition, in colonial breeders, colonies or patches only exist during the breeding season and unsuitable habitat for breeding becomes highly suitable during other times of the year for seabirds, which does not reflect the classic models of metapopulation theory (Matthiopoulos et al. 2005). Philopatry in seabirds can also confound the classic models of metapopulations because the propensity of veterans and recruits to return to their natal breeding colonies can affect the level of dispersal that is assumed in a metapopulation model (Buckley and Downer 1991). Range expansion of seabird metapopulations via colonization may therefore be slowed and even inhibited by the level of philopatry within a species. Modeling of seabird metapopulations suggests that when site-fidelity is strong, metapopulations follow a step-wise pattern, and population growth of new colonies occurs very rarely since individuals must overcome strong philopatric tendencies (Matthiopoulos et al. (2005)). Therefore, seabird species with lower levels of philopatry may be more successful at range expansion than others.

Significance

The majority of seabirds currently face environmental and anthropogenic threats at both regional and global scales that are leading to declines and extinctions of populations, colonies, and even species (Croxall et al. 2012). For example, chinstrap penguins (*Pygoscelis antarctica*) and Adélie penguins (*P. adeliae*), have declined markedly on the Antarctic Peninsula (Lynch et al. 2012), a change attributed to decreases in sea ice and burgeoning commercial krill fisheries (Trivelpiece et al. 2011). Surprisingly, the closely related gentoo penguin (*P. papua*) has both increased in population size and expanded its breeding range further south along the Antarctic Peninsula, resulting in numerous new breeding colonies in recent years. The dynamics of gentoo penguin proliferation and range expansion provide a unique model system to explore environmental changes that mediate the success and failure of populations in a metapopulation framework (Trathan et al. 2015), as well as identify habitat characteristics that enable dispersal and act as barriers to gene flow in a pioneering species (Munroe and Burg 2017).

0.2 Study System

Penguins (Sphenisciformes) diverged from their sister group, the tube-nosed seabirds (Procellariiformes), between 60 and 77 million years ago (Slack et al. 2006, Baker et al. 2006). Originating in New Zealand, the fossil record indicates that penguins dispersed to Antarctica, Australia, and South America by the Mid Eocene (Clarke et al. 2007; Ksepka et al. 2006; Gavryushkina et al. 2016). The first penguins that arrived in Antarctica were distantly related to the crown-group, Spheniscidae, with the *Pygoscelis* lineage originating in the late Miocene (12-10 million years ago) (Gavryushkina et al. 2016). Extant penguins currently comprise 18 recognized species (Borboroglu and Boersma 2015) found only in the southern hemisphere.

Gentoo penguins are one of the most widespread penguin species, with a circumpolar breeding distribution and a wide latitudinal range stretching from 46°00' S in the Crozet Islands south to 65°16' S on the Antarctic Peninsula (Lynch et al. 2013). Sub-Antarctic populations in South Georgia and the Falkland Islands have been relatively stable (Baylis et al. 2013), while populations in the South Sandwich Islands, South Orkneys, and South Shetland Islands have undergone increases in population size (Lynch et al. 2012; Forcada and Trathan 2009; Convey et al. 1999). However, gentoo penguins have exhibited both increases in population size and a southward expansion of their breeding range along the Western Antarctic Peninsula (WAP), a region that has undergone substantial warming, while populations of Adélie and chinstrap penguins are drastically declining in these areas (Lynch et al. 2012). Current hypotheses for the "climate winning" trends of gentoo penguins include plasticity in their breeding phenology (Lynch et al. 2013) a more generalized diet (Polito et al. 2015; Herman et al. 2017), and breeding habitat preferences that correspond to warming temperatures and declines of sea ice on the WAP (Cimino et al. 2013).

Coulson (2002) defines a seabird colony as "a group of breeding individuals which associate together and maintain the association to an extent that is greater (often much greater) than that expected by chance." *Pygoscelis* penguins in the Antarctic breed in such groups and exhibit dense aggregations of nests within colonies that are sometimes referred to as subcolonies (Williams 1995). The size and spatial structure of colonies, however, varies across species and location. Gentoo penguins tend to nest in much smaller and patchier colonies compared to their sister species. The patchiness of gentoo penguins begs the question of the importance of sub-colonies, and whether it is ecologically relevant to consider, as interactions between such sub-

colonies has not been researched.

The relatively small size of gentoo penguin colonies may be due to central place foraging theory. Both Adélie and chinstrap penguins have wide, pelagic foraging ranges, while gentoo penguins tend to forage in benthic habitat that is closer to shore (Miller et al. 2010; Polito et al. 2015). This foraging habitat may result in more competition for prey items among conspecifics, resulting in a density-dependent response in the number breeding pairs able to nest within the same colony (Oppel et al. 2015). However, the discrepancy in colony size among pygoscelids may be an artifact of the total number of Adélie and chinstrap penguin nests estimated on the Antarctic Peninsula (approximately 4.6 million and 1.3 million, respectively) compared to the far smaller gentoo penguin nest estimate (approximately 143,000) (from MAPPPD, Che-Castaldo et al. 2023).

Gentoo penguins typically have a localized foraging range and year-round presence at their breeding locations (Tanton et al. 2004; Miller et al. 2010), minimizing the propensity for dispersal (Levy et al. 2016) and Clucus et al. (2018) theorize that these behavioral traits likely result in the significant population differentiation observed among colonies throughout the Atlantic sector, and even colonies on the WAP within 50km of one another. Detection of such fine-scale differentiation makes it possible to explore population structure and colonization events at such scales.

0.3 Scope

The overarching goal of this dissertation is to investigate what is driving gentoo penguin southern range expansion, and what characteristics allow them to be successful colonizers during a time of rapid climate change and environmental variability. I explore this in four chapters.

The first chapter is a comprehensive update on the global distributions and population sizes of gentoo penguins, accounting for many new colonies and recent population growth trends. This chapter was published in *Polar Biology* in October 2020 (Herman et al. 2020).

The second chapter uses population modeling within an approximate Bayesian computation framework to investigate the contribution of immigration to rapid population growth observed in recently colonized gentoo penguin breeding locations. This chapter was published in *Ornithological Applications* in April 2022 (Herman and Lynch 2022).

The third chapter uses high-resolution genomic methods to explore the population genetic and demographic history of gentoo penguins on the WAP. It provides insight into the patterns of

dispersal and range expansion, along with identifying the source populations of recently colonized areas. This chapter is currently in review.

The fourth chapter applies findings from chapters two and three to an agent-based model that forecasts continued colonization events and southern range expansion of gentoo penguins along the WAP.

Chapter 1: Update on the global abundance and distribution of breeding gentoo penguins

1.1 Introduction

Climate change is widely known to affect the distribution and abundance of species, with some taxa experiencing range retractions and extinctions and others experiencing latitudinal shifts in response to warming conditions (Thomas et al. 2006; Hickling et al. 2006; Chen et al. 2011; Pecl et al. 2017). While attention has focused largely on species at risk due to the impacts of climate change, less attention has focused on species that may benefit. One such species is the gentoo penguin, whose populations along the rapidly warming Antarctic Peninsula (Vaughan et al. 2003) have not only undergone population growth since the early 2000s but have expanded their breeding range southwards with the establishment of many new breeding colonies (Lynch et al. 2013).

These trends stand in stark contrast to their sister species, the Adélie penguin and the chinstrap penguin, which have experienced considerable population declines on the Western Antarctic Peninsula (WAP), the principal drivers of which are still being debated (Forcada and Trathan 2009; Trivelpiece et al. 2011; Lynch et al. 2012). Current hypotheses for population increases and range expansion of gentoo penguins in this region include plasticity in their breeding phenology (Lynch et al. 2011), generalist foraging strategies and a flexible trophic niche (Polito et al. 2015; Herman et al. 2017; McMahon et al. 2019), and breeding habitat preferences (Cimino et al. 2013) vis a vis warming temperatures and declines of sea ice (Stammerjohn et al. 2008) throughout the waters around the WAP.

Since the last global assessment of approximately 387,000 breeding pairs (Lynch 2013), many new data have been collected across the gentoo penguin's distributional range. While the WAP population has been generally increasing, these trends are not homogenous, as some colonies have experienced recent declines (Petry et al. 2018; Dunn et al. 2019). In addition, the population in the Falkland Islands, where approximately one third of all gentoo penguins nest, increased overall by 105% between 2005 and 2010 (Baylis et al. 2013), though a subset of annually monitored colonies have since declined (Crofts and Stanworth 2019). In contrast, the majority of Indian Ocean colonies (approximately 16% of the global population) are rarely surveyed and represent areas of significant uncertainty for both abundance and trend.

To update our understanding of gentoo penguin abundance and distribution, we compiled all census data available to estimate the current global abundance and distribution of breeding gentoo penguins including new ground-count survey data from previously unknown colonies along the northwestern section of the WAP and on the Danger Islands. This updated population assessment allows us to identify gaps and associated priorities for future research and forms the basis for our forecasts of continued range expansion along the WAP.

1.2 Methods

The majority of the gentoo penguin distribution falls within the area managed by the Convention for the Conservation of Antarctic Marine Living Resources (CCAMLR), which is divided into three statistical areas: Area 48 (Atlantic Ocean sector), Area 58 (Indian Ocean sector) and Area 88 (Pacific Ocean sector). Gentoo penguins are found throughout Area 48 (including the Antarctic Peninsula, the South Shetland Islands, the Danger Islands, the South Sandwich Islands, the South Orkney Islands, and South Georgia) and Area 58 (including Marion-Prince Edward Islands, Crozet Island, Heard Island, and Kerguelen Islands) (Figure 1.1). Outside the CCAMLR area, gentoo penguins have a large breeding population on the Falkland Islands. Smaller populations also exist on Martillo Island in Tierra Del Fuego, Argentina, and Macquarie Island in the southwestern Pacific Ocean.

Much of the census data from the Antarctic Peninsula and South Shetland Islands are collected via opportunistic vessel-based field surveys (Lynch et al. 2013), though several additional surveys near permanent stations have been published (Petry et al. 2018; Smagol et al. 2018: Dunn et al. 2019). Most surveys were conducted via manual ground counts of individual nests, counting individual nests in panoramic photos taken from the ground or from an offshore vessel, or counting individual penguins from aerial photographs captured by unmanned aerial vehicles (Borowicz et al. 2018). The majority of surveys used in this assessment are precise to within 5% (i.e., they are Accuracy N1 counts, following Croxall and Kirkwood 1979).

The South Sandwich Island census data were either collected via direct ground counts or estimated from the aerial extent of the colony derived from satellite imagery (as described in Lynch et al. 2016). The Falkland Island data are from the most recent island-wide census from Baylis et al. (2013) and from an annual census of a subset of colonies by the Falkland Islands Seabird Monitoring Program (Crofts and Stanworth 2019). Although some sites on South Georgia have

been recently surveyed during short landings at popular tourist landing spots, the majority of gentoo penguin colonies have not been surveyed recently. The last published comprehensive census data for many South Georgia colonies derives from Trathan et al. (1996), which were based on field data collected between 1985 and 1987 by P.A. Prince and S. Poncet (Unpublished). Many CCAMLR Area 58 populations (Indian Ocean) have not been surveyed comprehensively since the 1980s and 1990s (Weimerskirch et al. 1988; Woehler 1993; Jouventin 1994) with the exception of Marion Island and Prince Edward Island, which were last surveyed in 2008 (Crawford et al. 2009). We include a subset of site-specific census data from Lescroël and Bost (2006) and Weimerskirch (*pers. comm.*) in Online Resource 1 but chose to use the most recent comprehensive survey data for the Kerguelen and Crozet Island Archipelagos for our global population estimates (Weimerskirch et al. 1988; Jouventin 1994). We also collated the most recent count data for Heard Island (Woehler 1993), Marion Island and Prince Edward Island (Crawford et al. 2009), Macquarie Island (Parks and Wildlife Service 2006), and Martillo Island (Ray et al. 2014).

We used a method for denoting census accuracy used in many other penguin assessments, denoting five levels of accuracy: 1) accurate to better than ± 5%; 2) accurate to 5-10%; 3) accurate to 10-15%; 4) accurate to 25-50%; 5) accurate to the nearest order of magnitude (Croxall and Kirkwood 1979; Naveen et al. 2012; Lynch et al. 2016; Borowicz et al. 2018). We assumed a truncated [0,∞) Normal distribution for census errors and propagated observation error to total and regional abundances by drawing from the corresponding Normal distribution (n=1000) centered on each individual count with a standard deviation appropriate for count accuracy (See Online Resource 1). We grouped sites into regions, and attained population estimates for that region by summing across draws for each site within that region. This allows us to extract a population estimate (the mean of the distribution for the sum) and 95% confidence intervals for the total abundance within that region of interest. Recognizing that our uncertainty on current abundance reflects both the uncertainty of the original survey and the time elapsed since the most recent survey of a colony, we have downgraded (for the purposes of summing up abundances at the regional scale and propagating our uncertainty in those regional abundances) the precision of counts older than 2015 by either one step (e.g., from accuracy=2 to accuracy=3) for counts 2005-2014, by two steps for counts 1995-2004, by three steps for counts from 1985-1994, and by four steps for counts prior to 1985, noting that the accuracy code saturates at 5. All analyses were performed in the R computing environment (R Development Core Team 2017).

To explore deglaciation as a possible mechanism for range expansion and colonization in gentoo penguins (Cole et al. 2019), we analyzed high-resolution satellite imagery provided through Google Earth for eight of 14 recently colonized areas and compared it with digitized aerial photography taken in December 1956 by the Falkland Islands and Dependencies Aerial Survey (Mott and Wiggins 1965).

1.3 Results

We estimate the global population of gentoo penguins to be 432,144 (95th CI: 338,059 – 534,114) (Table 1). Colony-specific census data, where available, are detailed in Online Resource 1. Approximately 364,359 (95th CI: 324,052 – 405,132) breeding pairs (85% of the population) live in the Atlantic Sector. We report on 14 colonies in the Danger Islands and along the northern coast of Joinville Island unknown at the time of the last global review of gentoo penguin abundance and distribution (Lynch 2013; see Online Resource 1 for full details). Including these new and previously unreported populations, the total abundance of gentoo penguin breeding pairs on the Antarctic Peninsula is 127,320 (95th CI: 118,199 -137,208). In addition to widespread population growth on the WAP, gentoo penguins have colonized at least nine new locations since 1994, with six of those locations contributing to a recent southward expansion of this species' range by approximately 60 km (Figure 1.2). One of these newly established colonies is on Green Island, part of the Berthelot Islands, as identified as an area for range expansion by Lynch et al. (2012). There has been rapid population growth at these sites (Figure 1.3) and there are currently 1,789 breeding pairs in this expansion zone as of 2018. These colonization events have occurred in a cascading southward pattern such that the most recent colonies are at the southernmost end of the species current range.

Comparing our assessment of the global gentoo penguin breeding population with Lynch (2013), we estimate that the global gentoo penguin breeding population has increased by 11% since 2013. The WAP population (excluding the Joinville and Danger Island colonies) estimated to be 116,521 (95th CI: 107,692 – 126,331) has increased by approximately 23% since 2013. The current Falkland Islands breeding population estimate is 131,059 pairs (95th CI: 129,036 – 133,123). This estimate combines new colony census data and the most recent comprehensive survey in 2010.

Of the eight newly established colonies in the Danger Islands and Joinville Island region for which historic aerial imagery were available, two sites appear to have recently experienced glacial retreat, exposing land that is now occupied by breeding penguins (Figure 1.4).

1.4 Discussion

Population growth and range expansion on the Antarctic Peninsula

The estimated 11% increase in the total global gentoo penguin population – to 432,144 breeding pairs since the last global estimate published in 2013 – can be attributed largely to increases on the Antarctic Peninsula, which has experienced warming air temperatures; increasing precipitation; and declines in the extent, seasonal duration, and thickness of sea ice (Trivelpiece et al. 2011; Turner et al. 2017; Shepherd et al. 2018). These trends are interesting in the context of gentoo penguin population growth and range expansion on the Antarctic Peninsula as they are in direct contrast to population trends of Adélie and chinstrap penguins in the same region. Since the 1980s, both Adélie and chinstrap penguin populations have been declining throughout the Antarctic Peninsula, even as gentoo penguin populations have grown steadily. While the drivers of decline for Adélie and chinstrap penguins are debated (Forcada and Trathan 2009; Trivelpiece et al. 2011; Lynch et al. 2012) there is considerable evidence that gentoo penguin life history provides some relative advantages. In particular, gentoo penguins have more flexibility in diet, foraging behavior, and breeding phenology, which suggests they may be more resilient to environmental change and better able to colonize newly exposed territory (Lynch et al. 2011; Masello et al. 2017; McMahon 2019). Moreover, gentoo penguins, unlike their congeners, can relay if the brood is lost early during incubation (Bost and Clobert 1992) adding considerable benefit in the ability to buffer population growth from early season events.

In some colonial-nesting bird species, the major contribution to rapid population growth at newly established breeding colonies has been the large-scale immigration of individuals (Oro and Ruxton 2001; Dunlop 2009; Santoro et al. 2016). Given their fast growth rates, it is unlikely that the growth of newly established gentoo penguin colonies can be explained by reproductive success and subsequent offspring recruitment alone; continued immigration from other locations over a period of several years would also be required.

Intraspecific competition is a possible explanation for the range expansion exhibited by gentoo penguins. A density-dependent decline in the habitat quality of current breeding locations

could result in individuals relocating to new, unused habitat and establishing new colonies. Newly-available habitat – the result of increased snow melt or glacial retreat (Trathan et al. 2013)– could drive emigration from established colonies or provide habitat to individuals previously unable to recruit into established colonies due to density-dependent processes. It is possible that sites beyond the former southern range margin were previously inaccessible due to heavy winter sea ice conditions, which would preclude foraging in the overwinter period. Declines in the extent or concentration of winter sea ice might make these sites available for breeding and overwinter residency and explain the gentoo penguin's southern range expansion. Historically, deglaciation and terrestrial ice sheet retreat has led to the expansion and proliferation of Antarctic and sub-Antarctic penguin populations since the last glacial maximum (LaRue et al. 2013; Clucas et al. 2014; Younger et al. 2015; Cole et al. 2019). The WAP has continued to experience rapid thinning of its ice shelves and their tributaries since the 1990s, which in part is attributed to increases in upper ocean heat content (Pritchard et al. 2012; Paulo et al. 2015). While gentoo penguins are expanding their range further south, they have also colonized two areas that appear to have become available due to very recent deglaciation. Aerial photographs of Noble Rocks in the Neumayer Channel and Moot Point in the Penola Strait, both taken in 1956 (Mott and Wiggins 1965), show these regions as previously covered by terrestrial ice sheets (Figure 1.4). For Noble Rocks, the advent of Landsat satellite imagery corresponds to the period in time in which the ice covering this small island began disintegrating, resulting in its full exposure probably early in the 1990s. These sites now have exposed bare rock and are physically separated from the glaciers by ocean, and both have been colonized by gentoo penguins.

While we do not have data to determine the pace of colonization at these sites, the imagery for Noble Rocks indicates that a colony of 40 nests can establish in as little as 25 years post-exposure. Moreover, this colony, situated on a space-limited small rock, seems likely to be at carrying capacity and may be serving as a source population for other bare rock patches in the vicinity. As glacial retreat continues to occur on the Antarctic Peninsula, and perhaps especially on its ice sheet-covered islands whose ice masses are no longer connected to the continental ice sheets, it is likely that expanding gentoo penguins may continue to seek out newly exposed areas of the rocky coastline as potential breeding habitat.

Population changes in Indian Ocean Sector

Though the last comprehensive survey of the breeding gentoo penguins on the Kerguelen Archipelago was conducted in 1985, the colonies on Courbet Peninsula, Penn Island, and Longue Island, have been regularly monitored since the 1980s. Lescroël and Bost (2006) reported an approximately 30% drop in breeding numbers from the mid 1980s to the mid-2000s on the Courbet Peninsula, which they suggested was the result of inshore food shortages where gentoo penguins typically forage (Lescroël and Bost 2005). However, a 2018 survey found that the colony had recovered, surpassing its 1985 breeding population size by 9% (Weimerskirch pers. comm.). In contrast, Possession Island, part of the Crozet Archipelago, was also surveyed in 2018 and found to have declined in breeding population size by 32% (Weimerskirch pers. comm). Similarly, the breeding population at Marion Island has declined by 52% since 1994 (Crawford et al. 2014). With the exception of Courbet Peninsula, these population trends suggest an overall decline among colonies found in the Indian Ocean sector, likely related to declines in food availability due to environmental changes.

Notable data gaps and future work

South Georgia's last population estimate was in the mid-1980s (P.A. Prince and S. Poncet, Unpublished) and estimated at 98,867 breeding pairs (Trathan et al. 1996), which constitutes over 25% of the last global gentoo penguin census. As this population is in close proximity to the Antarctic Peninsula, where populations have increased, and the Falkland Islands, where they appear in decline, it is essential that future efforts be made to fully survey this region. Similarly, the Indian Ocean gentoo penguin breeding colonies found on Kerguelen and Crozet Island Archipelagos and those on Heard Island, approximately 14% of the global population, have not been comprehensively surveyed since the early 1970s to late 1980s. Given recent genetic evidence for at least one distinct subspecies in the Indian Ocean (Levy et al. 2016; Vianna et al. 2017; Clucas et al. 2018), updated census data for these regions is particularly urgent.

Monitoring of changes in marine predators is critical to our understanding of marine ecosystem responses to climate change and other environmental perturbations. Predators are sensitive to shifts in prey resources and foraging habitat and can therefore act as proxies for identifying environmental perturbations (Weimerskirch et al. 2003; Sandvick et al. 2005; Cherel et al. 2007; Baylis et al. 2015). While much research is focused on species undergoing declines,

few studies focus on species that are robust to – or even benefit from – environmental change (Somero et al. 2010; Fulton et al. 2011; Clucas et al. 2014). gentoo penguins are one of the few examples of a marine predator that appears to be adaptive and resilient to environmental change, resulting in continued population growth and range expansion. Future work should focus on the mechanisms of population growth and range expansion in gentoo penguins, as well as the creation of habitat suitability models to identify future potential habitat. Updated census data for large populations on South Georgia and the colonies of the sub-Antarctic islands in the Indian Ocean should be considered of high priority.

1.5 Tables and Figures

Table 1.1: Estimated abundance (in breeding pairs) of gentoo penguins (*Pygoscelis papua*) for each major region and sub-regions. *Note that we did not include the 2018 Possession Island and Courbet Peninsula counts in the global estimate because we cannot directly differentiate them from the historical overall population counts of the Crozet and Kerguelen Island groups.

CCAMLR Sector	Region/Island Group	Abundance (95th Percentile CI)	Average Year of Survey
- -	Falkland Islands	131,059 (129,036 – 133,123)	2010
48.1	Antarctic Peninsula	127,320 (118,199 – 137,208)	2007
	WAP	116,521 (107,692 – 126,331)	2007
	Danger Islands	4,523 (4,398 – 4,655)	2015
	Joinville Island	6,022 (5,857 – 6,189)	2015
48.2	South Orkney Islands	5,364 (3,081 – 7,810)	1990
48.3	South Georgia	98,867	1985-1987
48.4	South Sandwich Islands	1,902 (1,790 – 2,022)	2010
- -	Martillo Island, Tierra del Fuego	31	2013
58.7	Marion Island	550	2012
58.7	Prince Edward Island	40	2008
58.6	Crozet Islands	9,000	1970-1982
	Possession Island	577	2018*
58.5	Kerguelen Islands	35,000	1985
	Courbet Peninsula	8392	2018*
58.4	Heard Island	16,574	1987
- -	Macquarie Island	3,800	2006
Total Estimate		**432,144** (338,059 – 534,114)	

Figure 1.1: Global gentoo penguin (*Pygoscelis papua*) breeding distribution (triangles) and CCAMLR statistical sections (bold numbers). Figure created using QGIS Development Team, (2018).
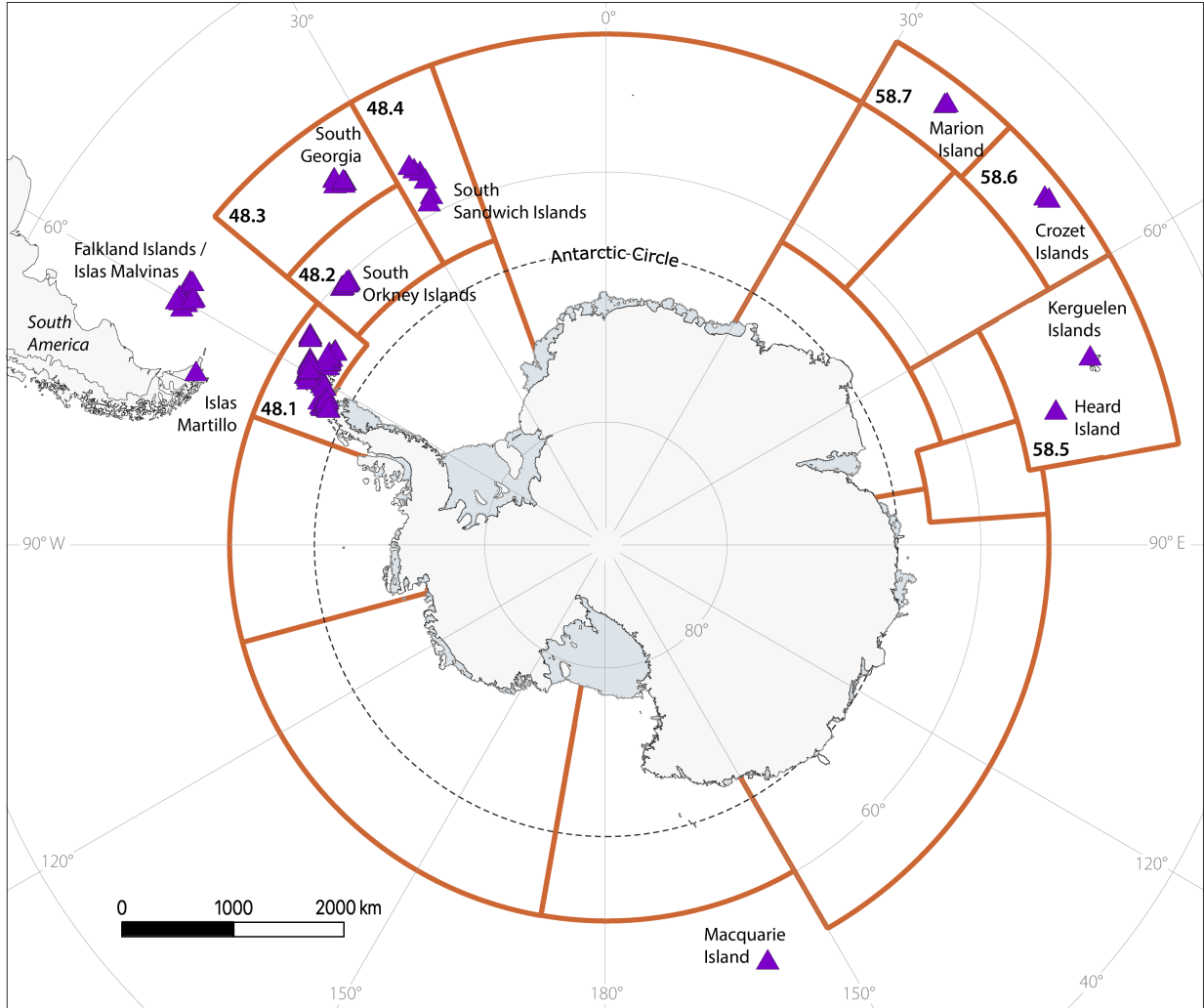
Figure 1.2: Locations of all gentoo penguin (*Pygoscelis papua*) breeding colonies on the Antarctic Peninsula (purple circles). Newest colonies formed after 1994 are in orange triangles. Figure created using QGIS Development Team, (2018).
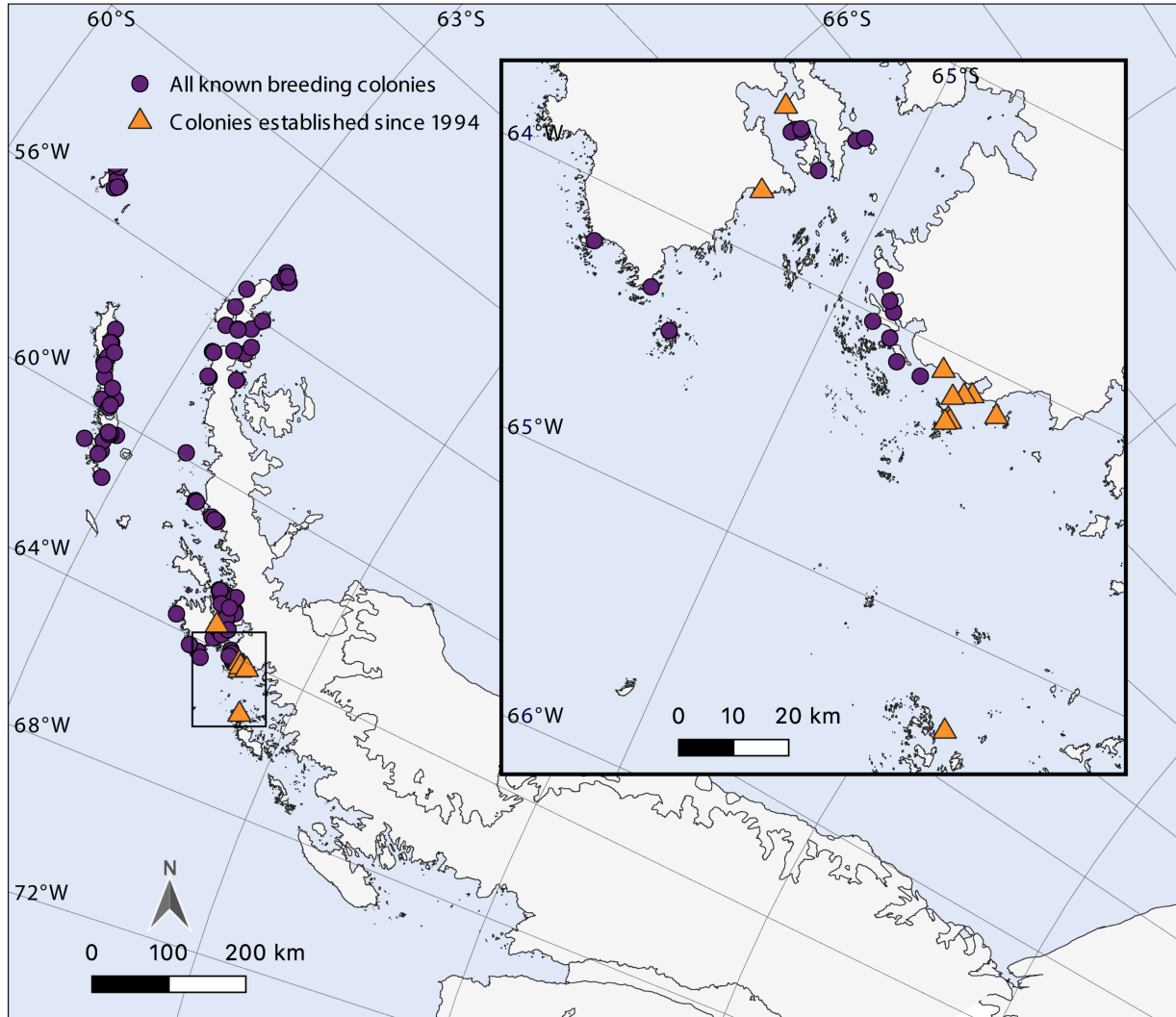
Figure 1.3: Population trends for six new gentoo penguin (*Pygoscelis papua*) breeding colonies of interest (data from MAPPPD; Humphries et al. 2017; see Online Resource 1).
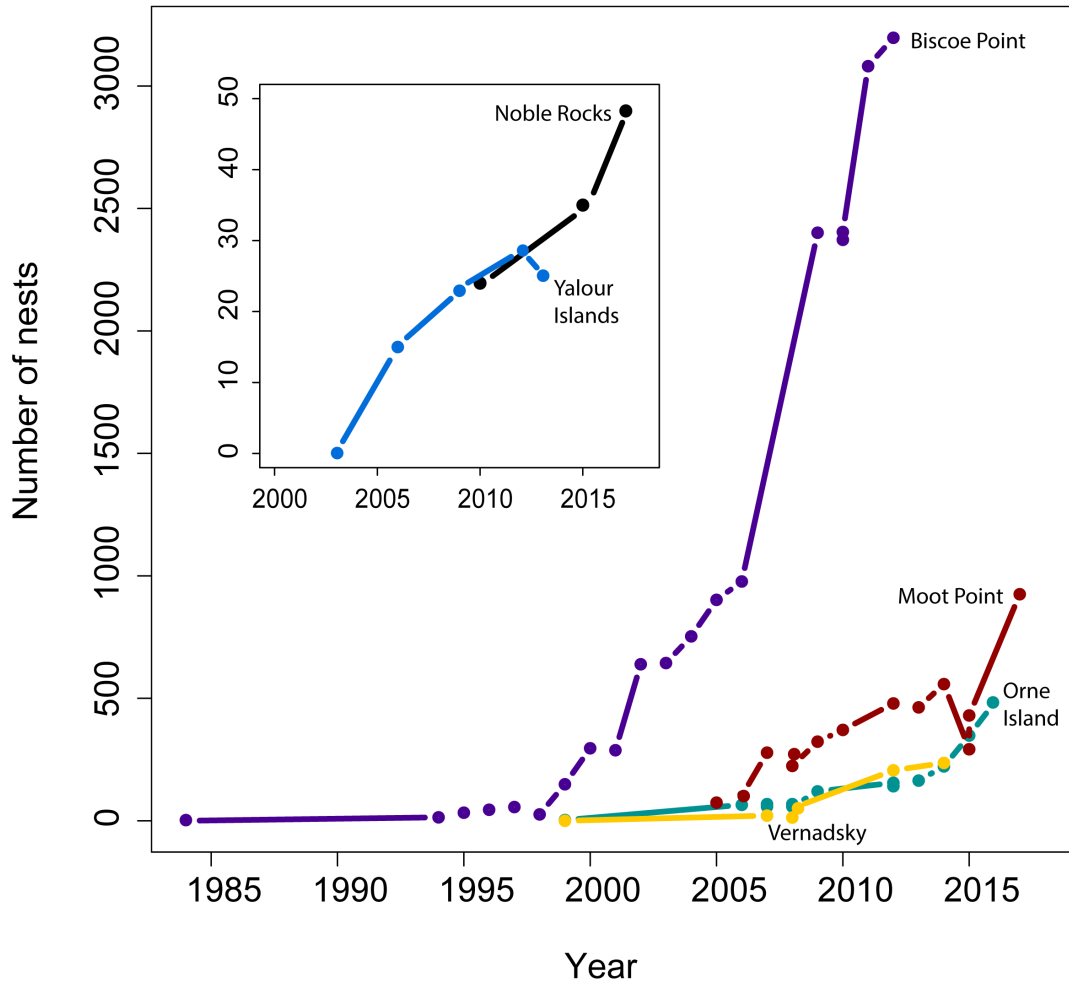
Figure 1.4: Glacial retreat at newly colonized gentoo penguin (*Pygoscelis papua*) breeding sites: A – Falkland Islands and Dependencies Aerial Survey Expedition (FIDASE) of Moot Point (Mott and Wiggins 1965); B - Recent high-resolution satellite imagery taken of Moot Point (2016; Google, CNES/Airbus); C – Falkland Islands and Dependencies Aerial Survey Expedition (FIDASE) of Noble Rocks (Mott and Wiggins 1965); D - Recent satellite imagery taken of Noble Rocks (2012; Google Earth, Maxar Technologies).
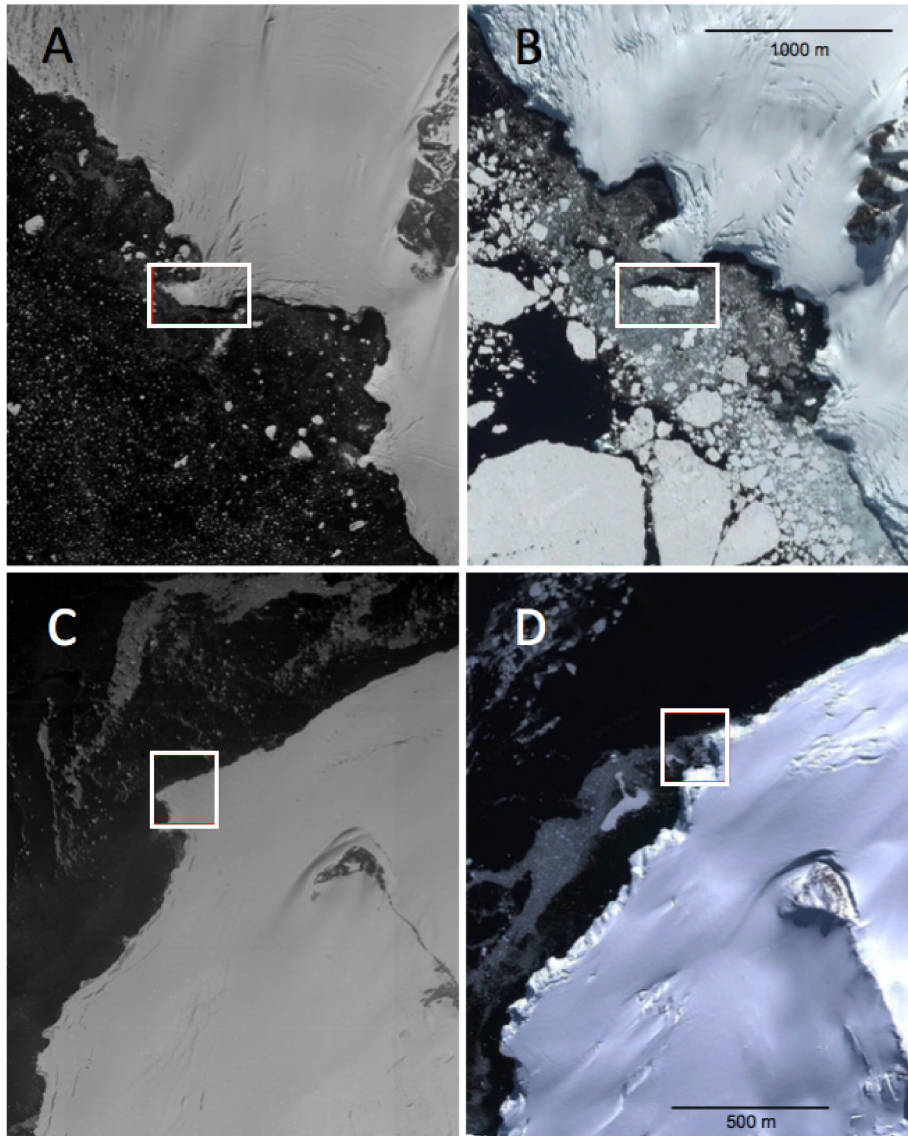
Chapter 2: Age-structured model reveals prolonged immigration is key for colony establishment in gentoo penguins

2.1 Introduction

Identifying the demographic processes behind colonization events and range expansion is important for understanding the population dynamics of a species, particularly when such species have patchy distributions and exhibit strong site fidelity (Sweanor et al. 2000; Lesage et al. 2000; Matthiopoulos et al. 2005; Secor et al. 2009; Millsap 2018). This is especially true for seabirds, as colonization events can be rare, and successful establishment of new colonies is difficult in long-lived species with delayed sexual maturity, as local recruitment alone is typically not enough to sustain population growth (Oro and Ruxton 2001, Coulson 2002, Dunlop 2009). Estimating the demographic processes underpinning colonization events in seabirds is critical to their conservation and management, as this group is vulnerable to changes in both their breeding and foraging habitat due to changing climate (Grémillet et al. 2009; Barbraud et al. 2012). By identifying networks of dispersal and migration among established seabird colonies, and between established colonies and new ones, conservation efforts can be focused on protecting islands and marine regions that facilitate range shifts. However, the ability to study these processes in seabirds is challenging due to the difficulty of obtaining data on demographic traits such as adult and juvenile survival, which are often estimated from long-term mark-recapture datasets (e.g. capture-mark-recapture, sight-resight, band-recovery) of individuals (Santaro et al. 2014; Szostek et al. 2014). Even more difficult to obtain are direct observations pertaining to the immigration of individuals among sites. While these issues challenge all seabird research, they are exacerbated in very remote regions where access to colonies is sporadic or unreliable.

When seabird mark-recapture programs are successful and the datasets are available, one can use methods such as integrated population models (IPMs) to infer latent demographic parameters that are related to species movement (Abadi et al. 2017). For example, a previously collected capture-mark-recapture dataset documenting individual histories (from which vital rates alone can be estimated) integrated with a time series of abundance via a joint likelihood function provides not only more robust estimates of the vital rates but also estimates for the number of immigrants dispersing to a nascent colony of a species (Abadi et al. 2010; Schaub and Abadi 2011).

Unfortunately, while such studies yield critical estimates of vital rates, the original mark-recapture data sets are not always available so IPMs cannot be developed. This is particularly an issue with many species of penguins (see Williams 1995). Furthermore, evidence that banding or tagging may impact survival or behavior of penguins (Dugger et al. 2006) has limited opportunities to conduct additional banding studies.

Gentoo penguins (*Pygoscelis papua*) along the Western Antarctic Peninsula have recently undergone widespread population growth while simultaneously expanding their range further south (Lynch et al. 2013; Herman et al. 2020). Census data reveal rapid population growth at these sites (Lynch 2012, Herman et al. 2020) suggesting immigration may be an important factor in the observed growth. Many studies have examined site-specific or regional population dynamics of gentoo penguins (e.g., Hinke et al. 2007; Forcada and Trathan 2009; Pistorius et al. 2010; Lynch et al. 2010; Trivelpiece et al. 2011), however no previous study has focused on the growth of new colonies due to range expansion. Genetic data from other portions of the gentoo range show strong genetic differentiation consistent with high site-fidelity (Clucas et al. 2014; Levy et al. 2016). In this context, the observed range expansion along the Western Antarctic Peninsula represents a paradox, since range expansion requires *ipso facto* the movement of individuals away from their natal colony.

We investigated the contribution of immigration on observed early population growth of gentoo penguins at four recently colonized sites (Figure 2.1). While an extensive long term data set of breeding censuses exists for gentoo penguin colonies on the Western Antarctic Peninsula, there is only one publication that includes apparent adult and juvenile survival, and that publication does not include the accompanying mark-recapture data from which these vital rates were estimated (Williams 1995). Because of the data limitations of our study system, we sought an alternative to an IPM framework approaches that require multiple datasets to infer immigration (Abadi et al. 2010; Schaub and Abadi 2011). We applied a rejection-based ABC sampling approach to a combination of an age-structured matrix model and published vital rates and breeding success. We also included a simulation study to test the validity of our approach. To our knowledge, this is the first application of ABC to estimate rates of cross-colony movement in a seabird species. Our use of rejection-based ABC is easily extended to study colonization events for other species where census data and some prior knowledge of life history characteristics are available.

2.2 Methods

We selected four recently colonized gentoo penguin breeding locations on the Western Antarctic Peninsula (Biscoe Point, Orne Island, Moot Point, and Vernadsky Station; Figure 1) with 12-20 years of abundance data. Population counts (specifically, counts of active nests during the breeding season) are available at the Mapping Application for penguin Populations and Projected Dynamics (MAPPPD; http://www.penguinmap.com; Humphries et al. 2017). Population census counts were conducted via manual ground counts of individual nests. All surveys were carried out following Croxall and Kirkwood (1979) and are precise to within 5%.

Age-structured Matrix Model

To simulate time series of abundance at newly established colonies, we used a female-based post-breeding age-structured matrix model (Eq. 1; Figure 2.2) where the vectors $\vec{n}_{t+1}$ and $\vec{n}_t$ represent the abundance of females in each age group from age=0 (chicks) to age=6+. We define immigrants as female immigrants (φ) and add φ as a vector to $\vec{n}_t$ at each time step.

(1)

$$\vec{n}_{t+1} = \begin{bmatrix} 0 & S_{1(t)}m_{(t+1)} & S_{2(t)}m_{(t+1)} & S_{3(t)}m_{(t+1)} & S_{4(t)}m_{(t+1)} & S_{5(t)}m_{(t+1)} & S_{5+(t)}m_{(t+1)} \\ S_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & S_{1(t)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & S_{2(t)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{3(t)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{4(t)} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{5(t)} & S_{5+(t)} \end{bmatrix} \times \left(\vec{n}_t + \begin{bmatrix} 0 \\ 0 \\ \varphi \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right)$$

S_i is the probability of survival at age class $i = 0$ (chick) through 5+ (adults), and m is breeding success (number of chicks produced per female). We relied on Williams (1995), the only publication of which we are aware that provides apparent survival rates for gentoo penguins: annual adult survival 0.75-0.89 and first-year survival 0.27-0.59. We assume that these vital rates were estimated from band resight data. For adult apparent survival, we used a truncated normal distribution with the mean equal to the midpoint of 0.75-0.89 and set sigma (σ) such that 0.75-0.89 is equal to $\pm 2\sigma$. We did the same for first-year apparent survival such that the range of 0.27-0.59 was equal to $\pm 2\sigma$. We used chick productivity rates $0.62 - 0.72$ females/nest (assuming a 1:1 sex ratio) from Lynch et al. (2010), in which reproductive data was collected via daily passive

observations of nest contents to detect egg or chick mortality. When chicks reached the créche stage, chicks were counted daily to get an estimate of chick survival for the colony. We again used a truncated Normal distribution with the mean equal to the midpoint of $0.62 – 0.72$ and set σ such that the range of $0.62 – 0.72$ is equal to $\pm 2\sigma$.

The number of immigrants in each year (φ) is drawn from a Poisson distribution with parameter λ (the expected value and variance of the distribution). Our methods for modelling λ are described in detail in the following section. We assume that immigrants show up after breeding in year t; this assumption means that immigrants arriving in year t are not counted in the year t census and that immigrants have to survive the winter before returning to breed for the first time at the new location in year $t + 1$ (see Appendix 2.1). We chose to model immigrants as being in age class 2, because seabirds who disperse from their natal breeding grounds tend to be young, subadults, or inexperienced breeders (Barbraud and Dalord 2021).

A preliminary analysis showed that a population growth model where lambda is constant did not capture the curved shape of the observed time series for Biscoe Point, Orne Islands, and Vernadsky Station (Appendix 2.2). We therefore selected two dynamic population growth models: in the first model, λ varied linearly with abundance (Eq. 2), as the literature suggests that seabird immigrants are more inclined to disperse to colonies with a relatively high presence of conspecifics (Forbes and Kaiser 1994):

$$\varphi_{ij} \sim Poisson\left(\lambda_{ij}\right) \qquad (2)$$
$$\lambda_{ij} = intercept_j + slope_j \times abundance_{ij}$$

We compared this model to a second model in which λ varied linearly with year since the initial colonization:

$$\varphi_{ij} \sim Poisson\left(\lambda_{ij}\right) \qquad (3)$$
$$\lambda_{ij} = intercept_j + slope_j \times year_{ij}$$

While it is common to use the log-link function when modelling a Poisson distribution due to the requirement that the parameter remain nonnegative, the log-link function is not suitable in this case

because we want the number of immigrants to have a linear, rather than a multiplicative or exponential, relationship to the covariate of interest (either abundance or year). We therefore use the identity link function (as shown in Eqs. 2 and 3) and impose a nonnegativity constraint on the resulting λ (Marschner et al. 2010). To assess the minimum period over which immigration must have occurred, we also fit versions of these two models in which immigration was restricted to a fixed number of years before stopping. We used the number of nests in the first non-zero population count for year 1 for our population simulations.

Application of rejection-based Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) is a likelihood-free statistical technique used to infer parameters in complicated systems (Beaumont 2010). Parameters are sampled from a probability distribution and accepted or rejected by comparing simulated data generated from sampled parameter values to observed data (Csilléry et al. 2010). While ABC techniques were developed for, and are used throughout, the field of genetics and evolutionary biology (Beaumont 2010; Csilléry et al. 2010), very few applications are found in the field of population biology (exceptions include van der Vaart et al. 2015; Chapron et al. 2016; Foley and Lynch 2020). Scranton et al. (2014) evaluated ABC as an effective tool to estimate demographic parameters of simulated time series generated from an artificial stage-structured population model.

Because forward simulation of dynamics is straightforward conditional on the model parameters, we used ABC to fit this model in the vein of Minter and Retkute (2019). ABC provides flexibility to infer parameters in complex model systems where constructing a likelihood function is difficult (Beaumont et al. 2010). Our rejection-ABC approach proceeds as follows:

1. Calculate λ based on parameters (intercept and slope) in the above linear models.
2. Sample $\varphi \sim Pois(\lambda)$.
3. Simulate abundance time series using φ and the dynamics described by Eq. 1.
4. Accept or reject simulations conditional on a threshold comparing the simulated data to the true data using a summary statistic (described in detail below). Values of λ (and the associated slope and intercept parameters) associated with accepted simulations are retained in the posterior.
5. Repeat until N approximate posterior samples are obtained.

Code for the ABC model can be found in Appendix 2.2. We used uniform prior distributions for the intercept and slope parameters with bounds that varied by site. Since ABC is a retroactive modelling framework, we determined the prior bounds by running simulations until the parameter values used were fully retained within the posterior distributions. Following a comparison with two other candidate summary statistics (See Appendix 2.3), we chose the mean absolute percentage error (MAPE), which is a measure of prediction accuracy commonly used in forecasting (Armstrong and Collopy 1992), applying the following equation:

$$MAPE = \frac{100}{n} \sum_{t=1}^{n} \left| \frac{A_t - S_t}{A_t} \right|$$

where A_t is the true nest count at a given year and S_t is the simulated nest count at that same year. MAPE also acts as a measure of goodness-of-fit in itself as it evaluates the fit of simulated time series to the observed time series. Simulated time series with MAPE values less than or equal to a set MAPE threshold value were considered similar enough to the empirical time series that the parameters used for that simulation were retained in the posterior distribution. The MAPE threshold values were tuned independently across each site to permit a maximum acceptance rate of approximately 0.1%. Increasing the MAPE threshold would accept more trajectories but with worse fit and would widen but not shift the posterior distributions of slope and intercept. We first verified the recovery of the true parameters in a simulation study (Supplemental Material S.2). We then applied our ABC model to the time series of the four gentoo penguin colonies of interest. We continued with our ABC sampling procedure until the number of accepted time series for each site was 1000. We also tested the sensitivity of the age-structured ABC model to the age of immigrants by running simulations in which φ was shifted across all age classes within the immigrant vector of the age-structured matrix model. Finally, we aimed to identify the minimum number of years of immigration required to sustain observed population growth at all four sites by turning off immigration at each time step and running the ABC simulations until MAPE values reached a threshold for which the acceptance rate remained reasonable (~0.1%). We implemented our ABC fitting routine in R version 4.0.2 (R Development Core Team 2017).

2.3 Results

The posterior distributions were contained within, and narrower than, the prior distributions for all model parameters, indicating that our prior distributions were not having an undue impact on our

final inference (Figure 2.3). The posterior distribution means and standard deviation are listed in Table 1 along with the site-specific MAPE thresholds and MAPE mean and standard deviations of accepted simulations. Both the abundance-dependent (Figure 2.3) and time-dependent models (Appendix 2.2) produced similar results for accepted simulations. We therefore used the abundance-dependent model for further analyses as it represented the most realistic depiction of dispersal and immigration. Immigration contributed to population growth for all sites and suggests immigration was required for several years at the four sites using both the abundance-dependent model (Figure 2.3). The number of immigrants increased over time at Biscoe Point, Orne Islands, and Vernadsky Station, while they decreased over time at Moot Point. MAPE values plateaued at year 15 for Biscoe Point and year 7 for Moot Point, suggesting a finite period of immigration required to generate the observed dynamics. In contrast, MAPE values never plateaued for Orne Island or Vernadsky Station (Figures 2.4 and 2.5), consistent with ongoing immigration at this site.

The minimum cumulative number of female immigrants over the time period of interest for each colony was 6529 [4558 - 8831] for Biscoe Island, 721 [374 - 1224] for Orne Island, 1039 [558 - 1666] for Moot Point, and 573 [260-1004] for Vernadsky Station (Table 2.2; Figure 2.6), which in total represent approximately 7.6% of the total gentoo penguin population on the western Antarctic Peninsula. The age-structured ABC model was insensitive to the demographics of immigrants if they were between age classes 2-5 (Appendix 2.4), but required higher immigration if immigrants were to arrive at age class 1 because of the lower survival of this younger age class. Therefore, our assumption that immigrants arrived at age class 2 yielded a reasonable and, if anything, conservative estimate of total immigration. Overall, the accepted simulations fit the observed time series data of Biscoe Point, Orne Islands, Moot Point well, with both abundance and time-dependent models capturing the curve of population growth (Figure 2.3). The model's fit to Vernadsky Station was poorer overall (Figure 2.4).

2.4 Discussion

To our knowledge, this is the first study to apply ABC methods to the use of census data of unmarked individuals and an age-structured population model combined with published vital rates and breeding success rates to estimate the relative contributions of local recruitment versus immigration in birds. The method provides an alternative and simple approach to estimating the number of migrating individuals among colonies when mark-recapture datasets are not available

for more integrated approaches such as IPMs. Rather than a single immigration "event", we found evidence for a process of continued immigration over a period of several years at the four gentoo penguin colonies considered. Moreover, rapid population growth observed at three of the four colonies suggests that immigration rates (numbers of new arrivals each year) actually increased over time (Figure 2.5). These results agree with a number of studies on seabirds for which mark-recapture datasets were available that found that sustained immigration contributes considerably to population growth at new colonies (Kildaw et al. 2005; Dunlop 2009; Pyk et al. 2012; Szostek et al. 2014; Santaro et al. 2016).

Such large influxes of immigrants and the resulting range expansion of gentoo penguins suggests ecological changes happening at local population levels. It is possible that the continued migration to new colonies suggested by our model is driven by breeding failures at nearby colonies, and individuals are relocating to more suitable habitat (Kildaw et al. 2005; Dunlop 2009; Pyk et al. 2012; Szostek et al. 2014; Santaro et al. 2016). However, gentoo penguins in this region have been increasing in population numbers overall (Herman et al. 2020), and it is unlikely that this influx of immigrants to new colonies is a result of poor breeding success. For example, Petermann Island is a large gentoo penguin colony (3516 nests as of 2017, Herman et al. 2020) within 10 km of both Moot Point and Vernadsky Station that could, in principle, represent a potential source of immigrants to these two sites. However, breeding success at Petermann Island (ranging 1.23 – 1.52 chicks créched per nest) was high in this period compared with Port Lockroy (0.86 chicks créched per nest), another gentoo penguin breeding colony close by (Lynch et al. 2010) and experienced rapid population growth at the same time Moot Point and Vernadsky Station were colonized, so there is no evidence to suggest that penguins would be abandoning this colony in high numbers due to poor breeding conditions. While an alternative hypothesis would be that Petermann Island or other nearby colonies had reached carrying capacity, forcing younger more inexperienced birds to seek out other breeding locations (Barbraud and Dalord 2021), all the surrounding colonies kept growing during and after the period in which individuals would have emigrated from source colonies. It's worth noting that Georges Point, a large and growing colony adjacent to Orne Islands, exhibited population fluctuations during the period in which the Orne Island colony was established and could have represented an immigration source (see Appendix 2.5). However, the census data for Georges Point do not suggest that the site had reached the carrying capacity and it is unclear if and why individuals from this colony would have sought out new breeding locations.

Unlike the other three colonies examined, Biscoe Point has no large (>2000 pairs) gentoo penguin colony within 50 km from which immigrants might be derived. This is particularly noteworthy because we estimate that 6529 individuals would have had to immigrate to Biscoe Point to sustain the observed population growth. However, our model also demonstrated that gentoo penguin population growth is sensitive to survival in age classes 0 and 1 (Appendix 2.5), and it is technically possible that Biscoe Point may have experienced higher sub-adult survival than has been reported in the literature. Though there is no *a priori* reason to suspect Biscoe Point would have a higher sub-adult survival than neighboring colonies, that scenario is the only possible alternative explanation for Biscoe Point's rapid growth beyond the massive and sustained immigration suggested by our model.

Though our model uses informative priors to estimate juvenile and adult survival in each year, it is still possible that these vital rates could be affected by density-dependent processes. For example, juvenile survival rate has been shown to decrease as populations grow in Wandering Albatross (*Diomedia exulans*) and African penguins (*Spheniscus demerscus*), suggesting that these trends were driven by adults having to compete to provision chicks and, in addition, fledged juveniles having to compete with adults for food (Sherley et al. 2014; Fay et al. 2015). It is therefore possible that our model underestimated the number of immigrants coming into these colonies due to a negative density-dependent relationship between juvenile survival rate and colony size. Our estimates of immigration are, for that reason, likely to be conservative and immigration may play an even larger role in the population growth of new gentoo penguin colonies that suggested by our estimates.

In contrast, there is evidence of positive density-dependence due to Allee effects in seabirds (Schipper et al. 2011; Ashbrook et al. 2011; Henson 2018), and therefore the rapid population growth in new gentoo penguin colonies could be due to increasing survival and reproductive success as the colony grows. However, our prior distribution for reproductive success was informed by a study located very close to these new colonies at a colony with very high productivity (larger, in fact, than other published values; see Williams 1995 and Lynch et al. 2010). Furthermore, Pozzi et al. (2015) found that although smaller, new Magellanic penguin colonies grew faster than older established ones due to positive density-dependence, yet this growth rate could not be sustained by internal productivity alone and must be driven at least in part by substantial immigration.

Studies on vagrancy in birds, the phenomenon in which an individual appears far outside of its normal range, as a driver of range expansion suggests that this behavior may be advantageous and adaptive rather than erroneous, particularly during the era of climate change (Veit 1990; Veit 1997; Veit et al. 2016). Population growth and reproduction have been found to correlate with increased vagrant numbers, particularly at small spatial scales (Veit 2000) and we hypothesize that the sustained and widespread population growth of gentoo penguins across the Western Antarctic Peninsula may be producing higher numbers of "vagrants" or dispersing individuals. This may explain why we found evidence for widespread movement among colonies even for a species otherwise considered to have strong natal philopatry.

Rapid population growth driven at newly established colonies may also be facilitated by a release on conspecific informational barriers. Forbes and Kaiser (1994) proposed the information barrier hypothesis of seabird range dynamics in which individual seabirds use the presence or absence of other birds at a prospective breeding location as cues about the quality of breeding locations. Depending on these cues, individuals may not disperse to a new breeding site, even though high-quality breeding sites may be available. Forbes and Kaiser (1994) suggest that established breeding colonies are most likely in areas of good quality habitat that result in long-term occupancy, which translates to greater opportunities for individuals to find mates and reproduce successfully. Therefore, it may be risky for individual birds to "pioneer" new colonies because they face many unknowns such as the presence of predators and potentially unfavorable abiotic conditions. The failure to recruit individuals until colonies have reached a certain size may slow the advancement of range expansion, though our model suggests that once the initial colonization event occurs, early arrivals may help to "recruit" future immigrants, and the information barrier of dispersal may diminish in the presence of larger numbers of conspecifics.

The use of ABC for population dynamics

We have demonstrated that ABC can successfully estimate immigration as a driver of population growth in gentoo penguin colonies. While ABC has been evaluated for its ability to estimate demographic parameters in population models using simulated data (Scranton et al. 2014), we are among the first to apply this method to empirical data and an age-structured population model.

Priors for an ABC model need to be carefully chosen to reflect the demographic parameters of interest. Because we modelled immigration as a function of abundance, it was important to set

prior bounds to include all possible intercept and slope values that would produce the population growth observed in the true time series. Our slope and intercept priors for modeling immigration were relatively uninformative, as more narrow priors did not fully capture the posterior distributions. Since the determination of priors is a modelling choice, wider bounds can be used but doing so will be computationally inefficient.

We chose the mean absolute percentage error (MAPE) as our summary statistic because it is a good measure of prediction accuracy in forecasting models (Armstrong and Collopy 1992) and performed the best compared to the other candidate summary statistics (Appendiz 2.3). It also simultaneously acts as a measure of goodness-of-fit, since the MAPE values are representative of how well simulations fit the observed data. One caveat is that MAPE appears to be sensitive to missing data. For example, the model had trouble fitting simulations to the Vernadsky Station time series, where 6 of the 12 year-time steps are missing data (Figure 2.3). The model's poor fit is likely due to the drop in nest numbers from 2008 to 2009, and then a relatively large increase in nests in 2012 (Figure 2.3). MAPE is likely more sensitive to deviations early in the time series and this constrains the model's ability to capture the shape of the observed time series.

While other well-developed analytical frameworks such as IPMs can leverage the integration of mark-recapture datasets and census datasets to estimate demographic parameters such as immigration (Abadi et al. 2010; Schaub and Abadi 2011), these methods are difficult to apply to many study systems such as penguin species where there are published vital rates but no available mark-recapture datasets, and future efforts to collect such datasets through widespread marking of penguins is infeasible and possibly unethical (Dugger et al. 2006). We have demonstrated that census data, previously published vital rates, and an age-structured model fit using ABC methods can reveal patterns of immigration and their impact on population growth in recently colonized gentoo penguin breeding locations and, in doing so, represents an accessible complementary approach that can be applied in situations where the source populations are unknown. Having an alternative method for estimating demographic processes in seabird species such as penguins with limited available demographic datasets will be critically helpful to their conservation and management. The Western Antarctic Peninsula is undergoing some of the most drastic warming associated with climate change, with increasing air temperatures and declines in sea ice, and species such as Adelie and chinstrap penguins are exhibiting significant regional population declines (Forcada and Trathan 2009; Trivelpiece et al. 2011; Naveen et al. 2012).

Understanding the demographic processes of these species will aid in future efforts to monitor how they may be affected by and respond to habitat perturbations due to changing climate conditions.

2.5 Tables and Figures

Table 2.1 Summary of prior and posterior distributions for slope and intercept and MAPE thresholds. Includes length of true time series and the number of available census counts used for each site.

	Intercept Prior Bounds	Intercept Posterior (mean + std)	Slope Prior Bounds	Slope Posterior (mean + std)	MAPE threshold	MAPE Mean + std)
Biscoe Point	*Unif* (0, 40)	14.97 ± 4.74	*Unif* (0, 1)	0.43 ± 0.04	35	33.19 ± 1.44
Orne Islands	*Unif* (0, 40)	1.14 ± 2.50	*Unif* (0, 1)	0.34 ± 0.07	20	16.21 ± 2.48
Moot Point	*Unif* (0, 250)	112.13 ± 30.03	*Unif* (-1, 1)	-0.07 ± 0.11	18	16.33 ± 1.19
Vernadsky Station	*Unif* (-50, 50)	-8.38 ± 7.32	*Unif* (-1, 2)	0.75 ± 0.24	45	42.52 ± 1.85

Table 2.2 Length of time series, number of observed counts, most recent nest counts, and total number of immigrants and CIs estimated over length of time series for each colony.

	Length of time series in years	Number of nest counts	Most recent nest count	Total # of immigrants and CIs estimated over length of time series
Biscoe Point	19	17	3197	6529 [4558 - 8831]
Orne Islands	20	12	401	721 [374 - 1224]
Moot Point	13	11	925	1039 [558 - 1666]
Vernadsky Station	12	6	379	573 [260-1004]

Figure 2.1 Map of all known gentoo penguin colonies at the southern breeding range edge along the Western Antarctic Peninsula (blue circles). Yellow triangles indicate colonies established since 1994. Additional panels a-d illustrate the observed time series data of nest counts for the four colonies of interest.
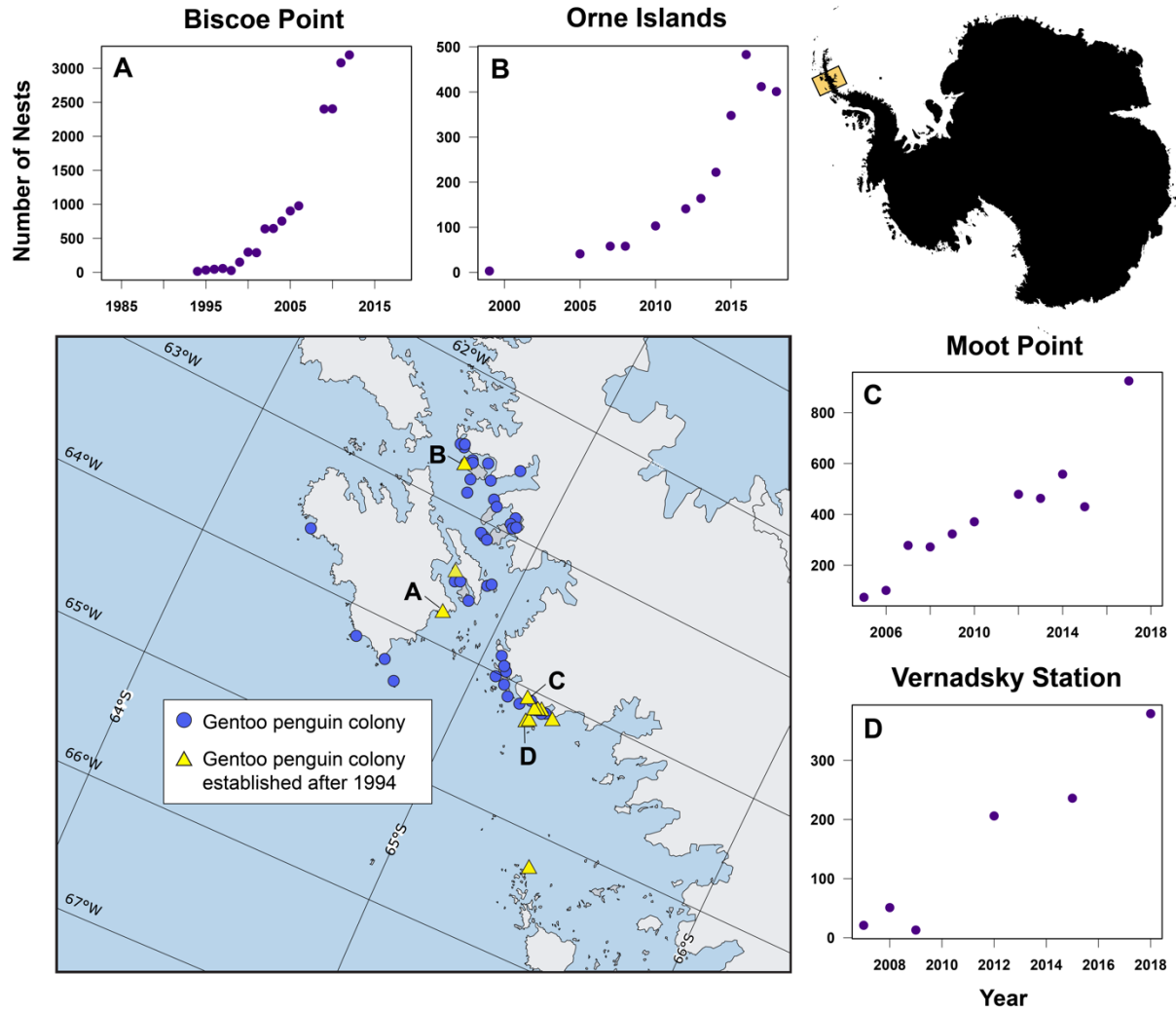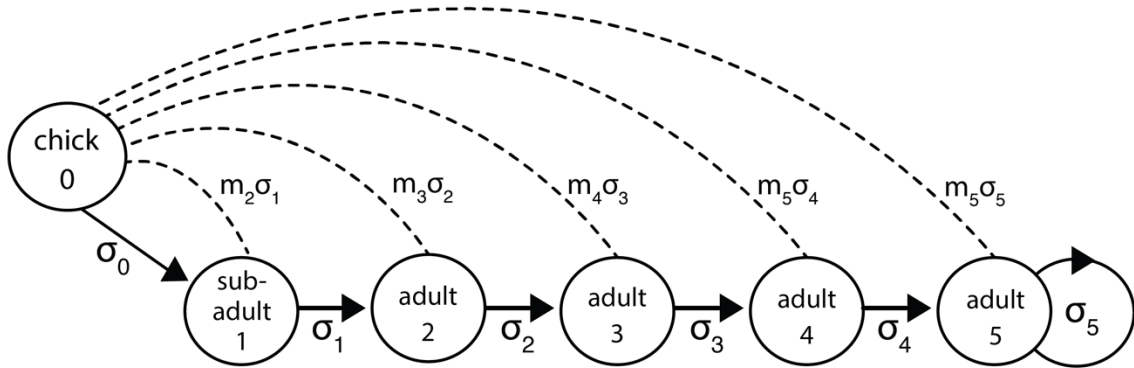
Figure 2.2 Post-breeding life-cycle schematic for female gentoo penguins.

σ_i = probability of survival at age class i

m_i = number of chicks produced per female at age class i

Figure 2.3 Results of the Approximate Bayesian Computation (ABC) model fitting in which immigration is a linear function of abundance. The model successfully captured the population growth curves for Biscoe Point, Orne Island and Moot Point, but struggled to fit the mid-series counts for Vernadsky Station. Grey lines are the rejected simulated time series, and blue lines are the 100 accepted time series for each site. The observed time series are represented by black circles. Insets illustrate the approximate posterior distributions for the intercept and slope parameters.
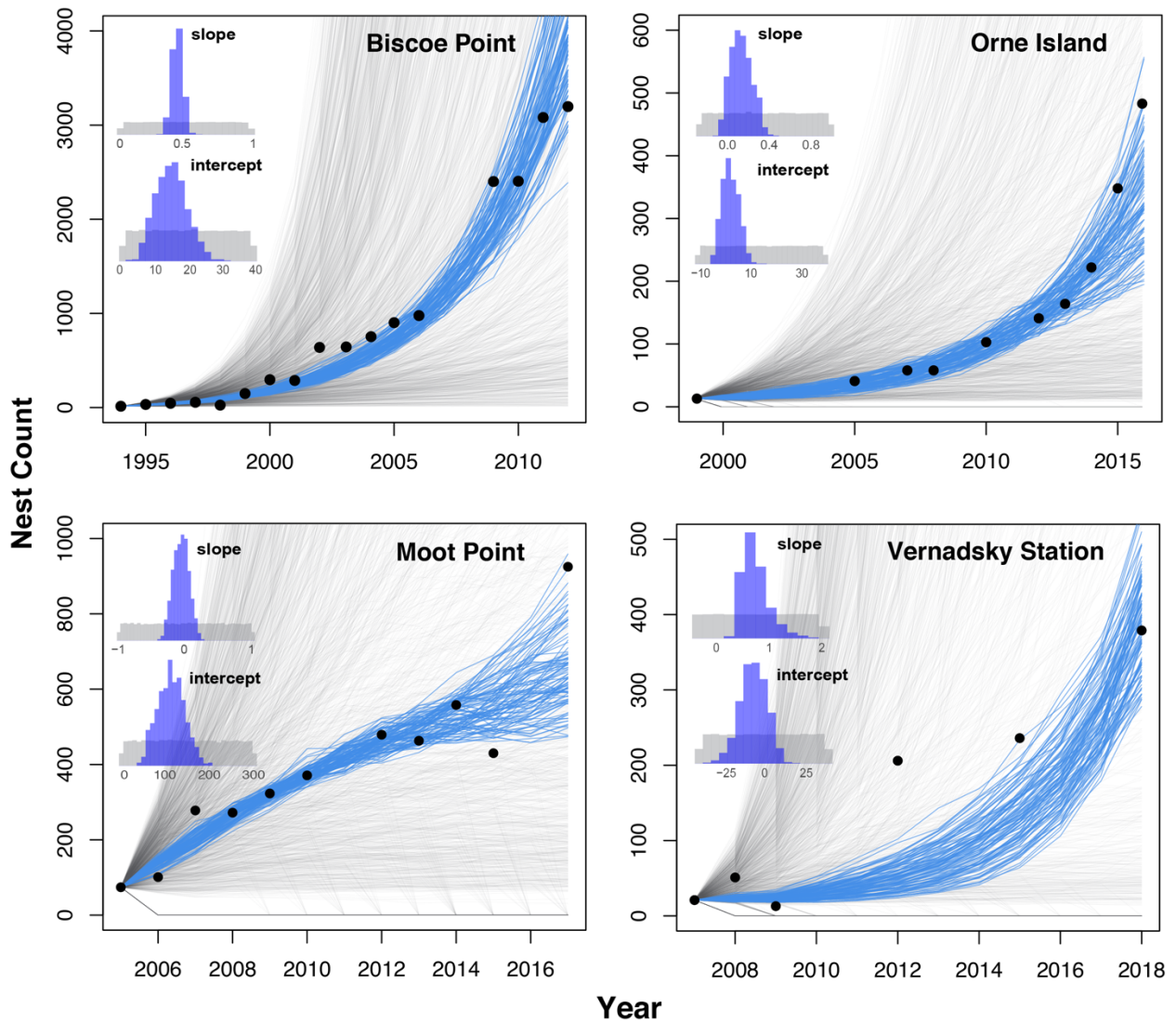
Figure 2.4 Abundance-dependent ABC model results with immigration turned off after the number of years indicated at the upper left of each panel (Vernadsky Station in yellow, Moot Point in teal, Orne Islands in red, Biscoe Point in purple). Results suggest each colony required a certain number of years of immigration in order to sustain the observed population growth. Lines are rejected time series. The observed time series are dark blue lines and squares.
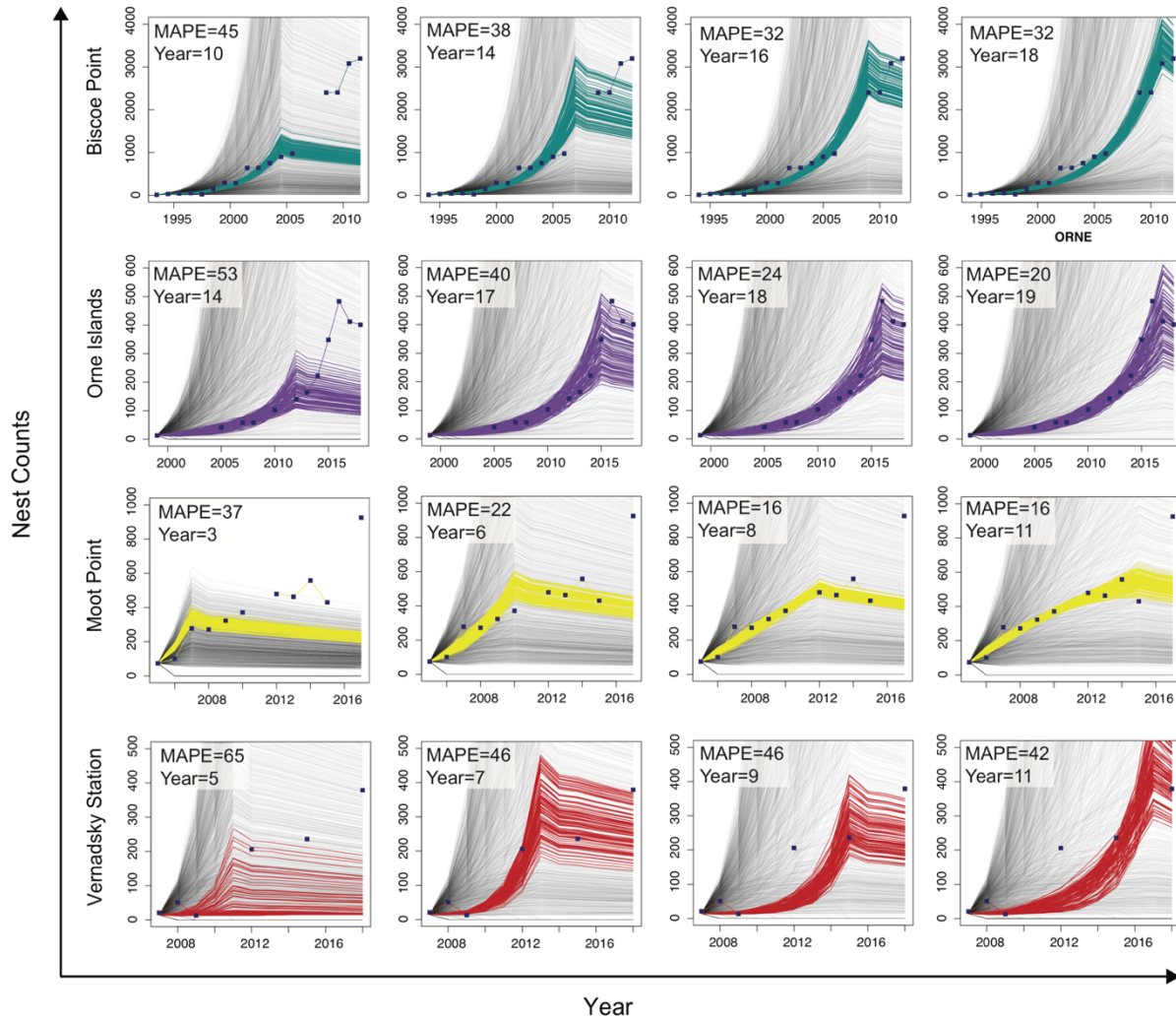
Figure 2.5 Minimum mean absolute percentage error (MAPE) values for acceptance rate of (~0.1%) when immigration is turned off at each year time step. MAPE values level off indicating a threshold of acceptance rate had been reached in the ABC simulations.
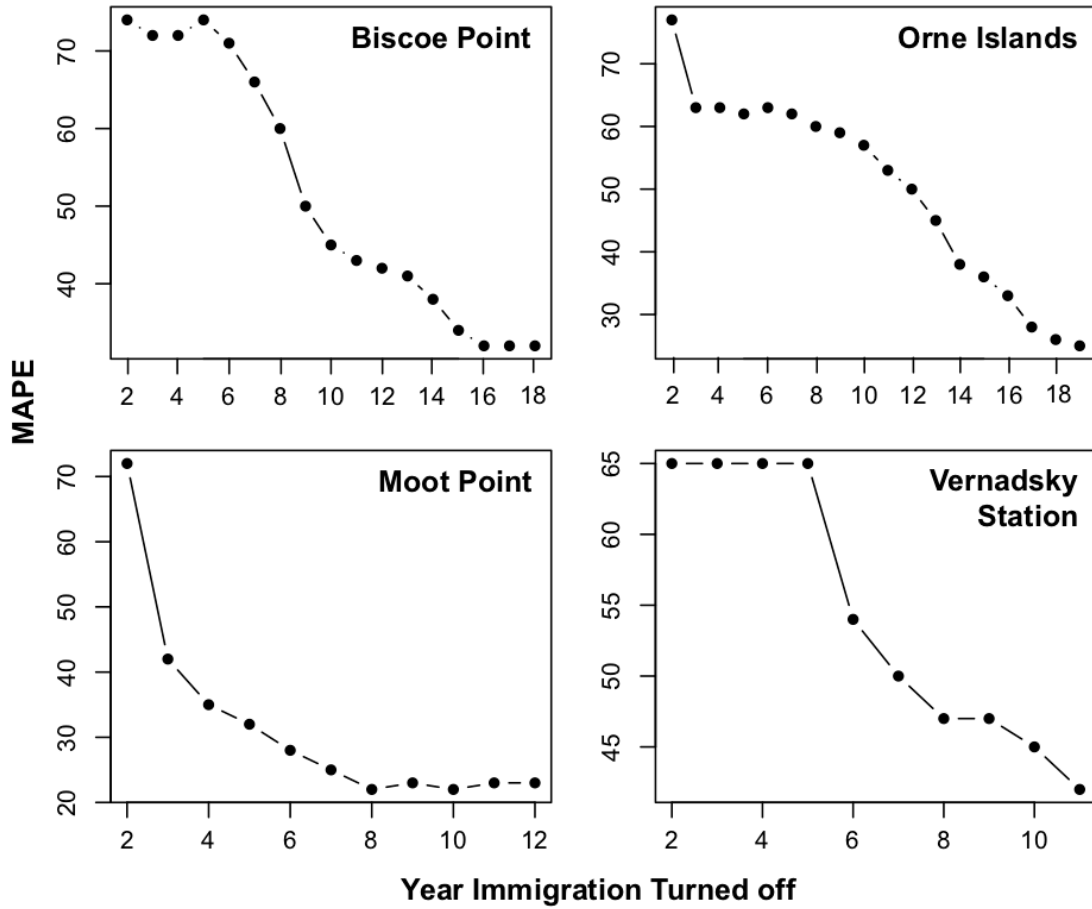
Figure 2.6 Estimated number of immigrants (blue dots) with credible intervals (shaded light purple) for each time series. Number of immigrants appeared to increase over time at Biscoe Point, Orne Islands, and Vernadsky Station, while they decreased over time at Moot Point. Black lines and dots represent observed time series.
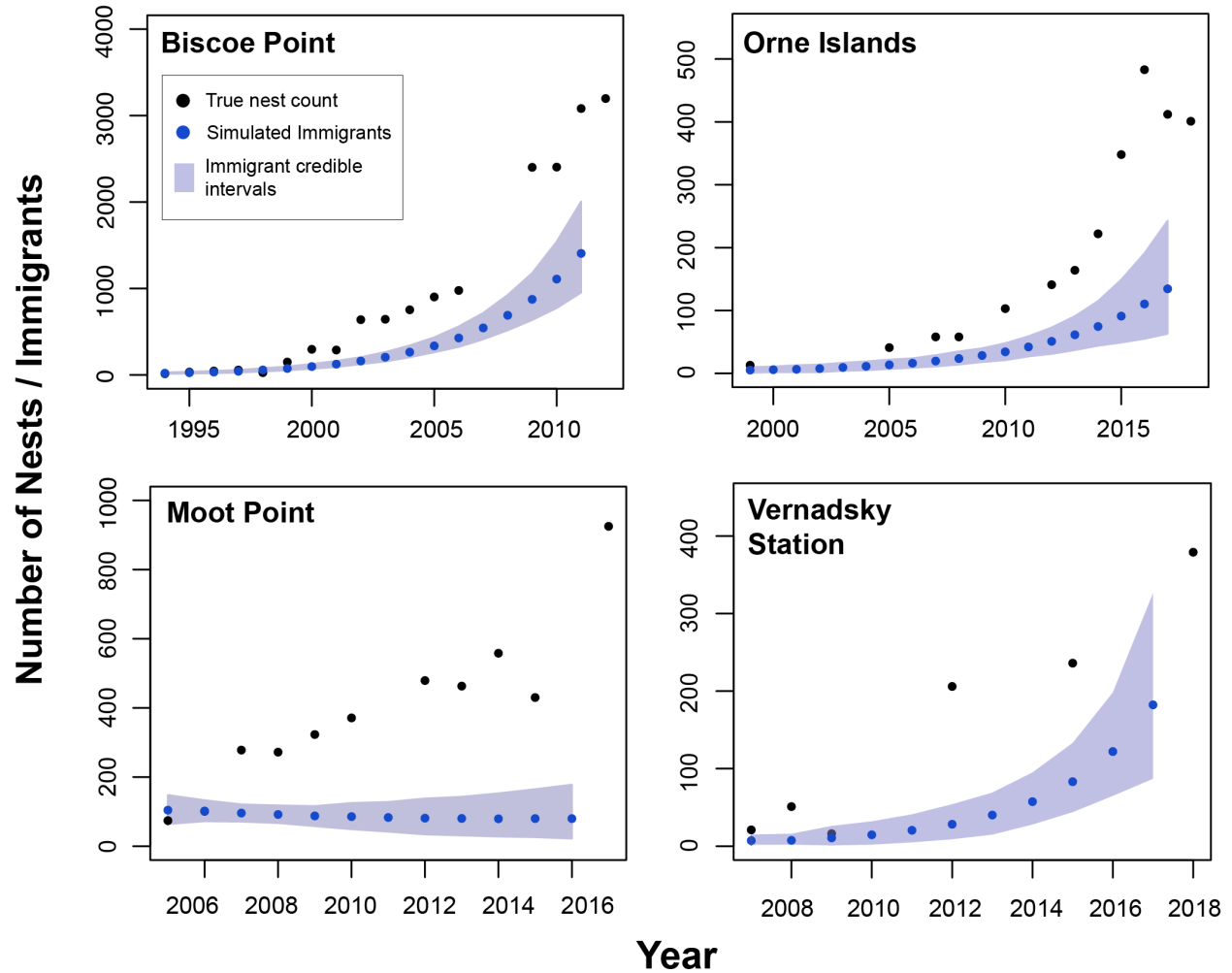
Chapter 3: Whole genome sequencing reveals stepping-stone dispersal buffered against founder effects in a range expanding seabird

3.1 Introduction

Species-level range shifts are ongoing phenomena that often occur during periods of changing climate. While many shifts in species distributions are attributed to the last deglaciation period approximately 10,000 years ago (Van der Putten 2012), ongoing rapid climate perturbations are accelerating such events (Sekercioglu et al. 2008; Germain and Lutz 2020; Gervais et al. 2021). As sea temperatures increase, marine species are shifting their range boundaries poleward as current ranges become unsuitable, and higher latitude waters provide new refugia of suitable habitat (Perry et al. 2005; Poloczanska et al. 2016; Pinsky et al. 2019).

In the Southern Ocean, however, we are also observing poleward range expansions in species that are not necessarily shifting their ranges all together, but rather capitalizing on new suitable habitat. For example, king penguins (*Aptenodytes patagonicus*) have colonized areas on the South Shetland Islands, south of their previous range edge, despite dramatic population increases throughout their historic range (Perón et al. 2012; Petry et al. 2013; Foley et al. 2018). Meanwhile, king crabs (*Paralomis birsteini*) have colonized the Antarctic continental shelf, a region historically too cold for this species, posing substantial threats to the benthic ecosystem (Smith et al. 2012; Aronson et al. 2015). gentoo penguins (*Pygoscelis papua*), another historically sub-Antarctic distributed species, have increased in population size and are expanding their range south along the Western Antarctic Peninsula (WAP) (Lynch et al. 2012, Herman et al. 2020), where declines in sea ice and increases in precipitation have been linked to a suite of oceanographic and ecological changes over the last 40 years (Turner et al. 2005; Turner et al. 2013; Meredith et al. 2017; Lin et al. 2021).

Eight new colonies of gentoo penguins have been established on the WAP in the last 30 years concurrent with local sea ice decline, with continued immigration of individuals sustaining the population growth of (Herman & Lynch, 2022). This movement of individuals was estimated to account for approximately 7.6% of the entire gentoo penguin population on the WAP (Herman & Lynch, 2022), a scale of ongoing range expansion that is to our knowledge unprecedented in

this region. This represented a rare opportunity to study the mechanisms of range expansion facilitated by the climate crisis in real time.

Despite the scale of dispersal involved in this recent southward expansion, the source populations of individuals and dispersal pathways are currently unknown. Identifying source populations can help to determine the causes and mechanisms of range shifts, as well as habitat features that facilitate or restrict dispersal. However, real time observations of dispersal and colonization events can be difficult, particularly in highly mobile, pelagic species such as penguins. While advances in tracking technology have improved our ability to trace the movement of individuals of highly mobile species (Hindell et al. 2020) these methods are not easily applied when source populations are unknown.

Population genomics has been effective at investigating gene flow among populations of seabirds (Clucas et al. 2018; Cristofari et al. 2018; Kersten et al. 2021; Herman et al. 2022), but very few genomics studies address dispersal patterns (Friesen et al. 2007). Colonial seabirds provide a natural model system to explore the balance of gene flow and genetic drift. Seabirds breed along coastlines, island clusters, and island chains, where dispersal models such as the *n*-islands model (random dispersal events independent of distance) and the stepping-stone model (dispersal among adjacent habitat patches) can be tested (Kimura & Weiss, 1964; Slatkin, 1993). Eco-evolutionary theory indicates that higher rates of gene flow among colonies with little population structure would suggest an island model of dispersal, whereas reduced gene flow among colonies with the presence of population divergence would suggest a dispersal pattern that follows the stepping-stone model. The patchy colony distribution and recent range expansion of gentoo penguins on the WAP facilitates the application of robust genomic methods to examine historical and current patterns of dispersal and identify any barriers to gene flow.

Here we present a fine-scale population genomics study of 129 gentoo penguin individuals across 11 colonies along the WAP using whole genome sequencing. To our knowledge this is the first study to use whole-genome wide approaches to investigate fine-scale population structure, demographic history, and dispersal in a top marine predator in the Southern Hemisphere. Our detailed examination of a species undergoing rapid and continuing range expansion reveals a pattern of historical and current dispersal that reflects a modified stepping-stone model with some post-divergence gene flow and buffering against founder effects.

3.2 Methods

Sample Collection

All sampling protocols and procedures employed were ethically reviewed and approved by Stony Brook University's Institutional Animal Care and Use Committee (IACUC). During the 2019/2020 and 2021/2022 breeding seasons, we collected blood samples from 129 breeding gentoo penguins across twelve colonies along the Western Antarctic Peninsula. Nine of these locations represent potential source colonies: Damoy Point, Danco Island, Neko Harbor, Noble Rocks, Jougla Point, Port Charcot, Joubin Islands, Gerlache Island, and Hannah Point, and Petermann Island (Figure 1a). Fifteen samples were collected at each of these sites, except for Damoy Point and Hannah Point where we collected ten samples from each, Noble Rocks where we collected five samples, and Petermann Island where we collected three samples. We also collected samples from two newly established colonies: three samples from Moot Point, and ten samples from Tuxen Rocks. In addition, we collected a blood sample from a non-breeding individual at Armstrong Reef, a location approximately 75 km to the south of their southernmost range limit (Figure 3.1a).

We determined breeding status by observing incubation or chick brooding behavior. We secured adults on the nest using a landing net and carefully picked them up so as not to disturb the nest contents. Eggs or chicks on the nest were covered with a plastic strainer that was secured in place by a hot water bottle to provide warmth and protection from predators. We collected 1-2mL of blood per individual from a small vein between the toes. We sterilized the area with alcohol swabs prior to blood collection and promoted clotting afterwards by placing pressure on the draw site with a clean piece of gauze. We then released adults close to their nests and made sure they returned to their nest to continue incubating or brooding before moving on. Blood samples were stored in ethanol or DNA/RNA Shield at -20°C.

DNA extraction, sequencing and bioinformatic processing

We conducted DNA extractions using Qiagen DNEasy blood and tissue kits. For samples preserved in ethanol, we first dried approximately 80μL of blood and ethanol solution on sterile filter paper for one hour to allow for ethanol evaporation. We then cut around the dried blood spot and placed it in a microcentrifuge tube containing 180μL of Buffer ATL, and then added 20μL of Proteinase K and incubated at 56 degrees C for 6 hours. We then carefully removed the filter paper and followed the standard DNeasy blood and tissue kit protocol. Whole genome re-sequencing

with 150bp paired end was conducted at Beijing Genomics Institute (BGI) using their DNBseq technology with a target average coverage of ~20x. We aligned raw sequence reads to an indexed *Pygoscelis papua* reference genome (Pan et al. 2019) using BWA (Li and Durban, 2009). We then merged read groups using samtools (Li et al. 2009) and marked duplicates using Picard (2.20.4). We called variants using GATK HaplotypeCaller and hard filtered for high quality variants following GATK Best Practices recommendations (McKenna et al. 2010, Van der Auwera and O'Connor, 2020). For SNPs, we used the following GATK VariantFiltration expression filters: QUAL < 50.0, QD < 2.0, FS > 60.0, and SOR > 3.0, and MQ < 40.0. For INDELs, we used the following GATK VariantFiltration expression filters: QD < 2.0, QUAL < 50.0, and FS > 60.0. We then ran a base quality recalibration with GATK BQSR using the hard-filtered high quality variants. We then did our final variant calling step using GATK HaplotypeCaller.

We filtered out likely sex chromosomes by testing for bimodality of the depth of coverage within each contig using Hartigan's Dip Test (Hartigan and Hartigan, 1985). Contigs that had a p-value of less than 0.05 were considered to be located on a sex chromosome and removed. We then took the average depth of coverage across all contigs and removed contigs that were either greater than or less than 5x coverage from the mean genome-wide coverage to filter out repeat elements such as transposable elements. This resulted in a final set containing 5,752,817 SNPs representing ~78% of the full genome across 398 contigs.

SNP-based population genetic analysis

For one set of analyses we randomly thinned the SNP dataset to ~100,000 sites using PLINK (Purcell et al. 2007) to control for linkage disequilibrium. To explore general population structure amongst individuals across colonies, we conducted a principal component analysis (PCA) using smartPCA (Patterson et al. 2006). We also used ADMIXTURE (Alexander et al. 2009) to perform model-based clustering ancestry analyses assuming K=2-12. We ran TreeMix on individuals grouped by colony to infer patterns of past population divergence (Pickrell and Pritchard, 2012) and included an Adelie penguin (*P. adeliae*) genome as the outgroup (Vianna et al. 2020). As there were no outlying residuals with no migration assumed, no migration edges were included in this analysis. Finally, we implemented estimated effective migration surfaces (EEMS), an analysis that visualizes regions of a species' distribution where population divergence deviates from uniform isolation by distance (Petkova et al. 2016). EEMS uses MCMC to estimate migration and diversity

parameters by sampling from their posterior distribution given observed genetic dissimilarities. We omitted samples from Hannah Point in order to highlight fine-scale migration rates around Anvers Island and the mainland WAP colonies. We calculated a distance matrix from our thinned SNP dataset using the program bed2diff_v1 from EEMS. We ran three independent chains from different starting seeds for 1×10^7 MCMC iterations (with a burn-in of 1×10^7 iterations and a thinning rate of 5000 iterations) for each of three different grid sizes (250, 450, and 650 demes). We ran multiple short runs to choose proposal values for migration and diversity parameters that had acceptance rates between 20% and 40%. We plotted the log posteriors using the rEEMSplots package in r to confirm the runs converged, and then combined runs from the three grid sizes to construct the migration surface visualizations (Petkova et al. 2016).

Haplotype-based population genetic analysis

The *P. papua* genome has a low N50, making haplotype-based population genetic analysis difficult. Therefore, we aligned the *P. papua* reference genome (for the sex-chromosome filtered contigs only) to a *Megadyptes antipodes* reference genome (Pan et al. 2019) comprising larger, more complete scaffolds using Ragtag (Alonge et al. 2022) and applied the scaffold merging option. This resulted in 723 *P. papua* contigs placed unambiguously across 77 *Megadyptes* scaffolds. We then calculated pairwise linkage disequilibrium (estimated as the r^2 value using a Pearson's correlation coefficient from diploid genotypes (Purcell et al. 2007)) for each pair of SNPs spanning 10kb on either side of those *P. papua* contigs proposed to be adjacent (including all SNPs with minor allele frequency greater than 10% and with a Hardy-Weinberg p-value greater than 1×10^{-8}). The average r^2 across all pairs was weighted by the inverse of their physical distance assuming no gap between contigs and compared to 1,000 randomly chosen pairs of contigs to generate an empirical p-value for the target pair. Those pairs with p<0.05 (i.e., greater linkage disequilibrium expected than by chance), were placed on the same scaffold, otherwise the proposed *Megadyptes* scaffold was broken into two. We only considered final scaffolds > 5Mb, resulting in a new build (Ppap.V1toMegaLD_5MB) containing 72 scaffolds spanning ~766Mb (~58% of the complete genome). Seven scaffolds in this new build were between 20-40MBb and 26 > 10Mb (compared to 0 and 12 scaffolds in the original build respectively).

We created a chain file and converted genotyped coordinates from *P. papua* to Ppap.V1toMegaLD_5MB using the LiftOverVCF function in GATK and phased the resulting

VCF using Shapeit4 (Delaneau et al. 2019). We then used LDHelmet to construct a *linkage-disequilbrium*-based recombination map (Chan et al. 2012) using the combined samples of Neko Harbor and Danco Island. A grid of ρ values were set to -r 0.0 0.1 10.0 1.0 and 100.0 and θ set to -t 0.0007. We calculated eleven Padé coefficients using pade (Chan et al. 2012) with a population scale mutation rate of 0.0076. We then estimated a recombination map by running rjmcmc (Chan et al. 2012) for 1×10^6 MCMC iterations with burn in rate of 10000, a window size of 50, and block penalty of 50.0. We determined the final ρ-scaled recombination map from the 50% percentile of the sampling distribution, with interpolation used to determine ρ between SNPs. Finally, we converted map to centiMorgans assuming $\theta = 4N_eu$ and $\rho = 4 N_eu$, where $\theta = 0.0007$, $u = 1.5 \times 10^{-8}$ (Bergeron et al. 2023) and thus $N_e = 11667$. We controlled for a minority of regions with spuriously high estimated ρ by capping the maximum value by the 99th percentile of the genome-wide distribution.

We inferred tracts of identity-by-descent (IBD) using Refined IBD and the recombination map generated above (Browning and Browning, 2013). We accepted any IBD segment with a LOD score >3 and constructed a haplotype network of tracts using ForceAtlas2 edge-node algorithm in Gephi (Jacomy et al. 2014). We used PLINK to determine runs of homozygosity (ROH) (--homozyg --homozyg-window-snp 1000 --homozyg-window-missing 0 --homozyg-density 1 --homozyg-gap 1000 --homozyg-window-threshold 0.05 --homozyg-window-het 1) and summed the total number of ROH per population using all segments, and then only segments greater than 5 centimorgans (cM). We excluded Moot Point, Petermann Island, and Armstrong Reef due to low sample sizes. We estimated genome-wide geneaologies using Relate (Speidel et al. 2019). We estimated historical population size and pairwise separation histories between populations assuming a per generation mutation rate of 1.5×10^{-8} based on Adelie penguins (Bergeron et al. 2023) and assumed a generation time of eight years. Finally, we used FineSTRUCTURE to perform chromosome painting on dense haplotype data and perform population clustering analyses (Lawson et al. 2012).

3.3 Results

Genomic Population Structure

We performed whole genome resequencing for 129 individuals from 11 colonies at ~18x mean coverage. A PCA revealed five discrete clusters in genotypic (PC1 and PC2) space that

corresponded closely with geographic location (Figure 3.1b). gentoo penguins at Hannah Point formed their own cluster. The northern cluster on the WAP consisted of Danco Island and Neko Harbor. Interestingly the non-breeding individual sampled at Armstrong Reef more than 75 km away fell within this cluster. The cluster on the eastern side of Anvers Island consisted of Jougla Point, Noble Rocks, and Damoy Point, with the latter being clearly distinguishable in the PC space. Gerlache and Joubin Islands clustered on the western side of Anvers Island, though the latter showed greater spread. The southern cluster consisted of Port Charcot and Petermann Island, two larger and older established colonies, and Moot Point and Tuxen Rocks, two colonies recently established within the last 15 years (Herman et al. 2020). While Port Charcot and Tuxen Rocks overlapped in PC space, the latter were generally more shifted along PC2, corresponding with their more southern latitude.

Model-based clustering analysis via ADMIXTURE with k=5 confirmed the strong geographic clustering identified by the PCA, with little to no evidence of significant admixture. (Figure 3.1c; Appendix 3.1). Beyond k=6 no meaningful structure was observed. Results also indicated that the most recently established colonies sampled at the range edge, Moot Point and Tuxen Rocks, were likely sourced from the closest adjacent colonies sampled, Port Charcot and Petermann Island. This indicated that gentoo penguin southern range expansion is likely sourced by the nearest range-edge colonies. However, the non-breeding indiviual sampled at Armstrong Reef again clustered with the northern colonies, supporting the PCA results.

Reconstruction of population divergences based on patterns of genetic drift using Treemix (Figure 3.2) supported the five discrete clusters found in the PCA and ADMIXTURE results. Introducing migration edges did not improve the model fit based on remaining residuals with no migration and were therefore omitted (Appendix 3.2). The topology of the resulting dendrogram revealed an early split between Hannah Point in the South Shetland Islands and all colonies sampled from the Western Antarctic Peninsula (Figure 3.2). The east Anvers cluster split from the west Anvers cluster, and the southern cluster and northern cluster formed a clade that split from an ancestral population likely closely related to the Damoy Point colony (Figure 3.2). These results suggest gentoo penguins colonized Anvers Island first and moved from west to east in a stepping-stone fashion before colonizing this region of the mainland WAP. An unrooted dendrogram of individuals constructed using FineSTRUCTURE (Appendix 3.2) supported the general topology of the population-based results of Treemix and grouped the individual from Armstrong Reef with

individuals sampled from Danco Island, suggesting the vagrant likely originated from this colony rather than Neko Harbor.

IBD segments were generally short and ranged from 1.5 to 3.8 centimorgans, indicating no major recent episodes of gene flow. A total of 89 shared segments were inferred, with 36 segments shared among individuals from different colonies. Individuals from Hannah Point, Danco Island, and Neko Harbor shared no segments with individuals from other colonies (including those belonging to the same PC and ADMIXTURE clusters), suggesting these colonies had experienced extensive genetic drift since colonization events, with little post divergence gene flow from the other colonies sampled. In contrast, a large network of shared IBD tracts connected individuals from all colonies around Anvers Island and the southern cluster (Figure 3.3a). Notably, the network was made up of three primary components: i) some shared segments amongst the southern colonies and the east Anvers cluster (including Damoy despite no overlap in the PCA), ii) some shared segments amongst the western Anvers cluster, and iii) many shared tracts amongst individuals from all four populations in the southern cluster, likely reflecting the recent common ancestry of these newer colonies. The latter had some shared tracts with both the east and west Anvers cluster possibly reflecting limited geneflow between Anvers Island and the mainland after divergence had occurred.

Explicitly modeling gene flow spatially using EEMS migration suggested reduced migration rates between Anvers Island and mainland WAP (Figure 3.3b), while migration rates between Joubin Islands and Gerlache Island were at least tenfold higher than the region average, with no areas of reduced migration rates between the west Anvers cluster and the southern cluster. However, these results are likely a vast oversimplification of migration surfaces. Because EEMS requires the input of a single closed polygon for spatial analysis, we were not able to include all of the islands and islets present around the Antarctic Peninsula, making it difficult to infer all physical barriers to gene flow.

Individuals from Joubin Islands and Port Charcot had the highest number of ROH greater than both 1cM and 5cM, followed by Tuxen Rocks (Figure 3.4a and b), suggesting relatively small recent effective population sizes (N_e). Gerlache Island had the next highest number of ROH greater than 5cM. All other colonies had much fewer ROH greater than 1cM (Figure 3.4a), and Jougla Point, Noble Rocks, and Neko Harbor had zero ROH greater than 5cM (Figure 3.4b). Allele frequency spectra for each population demonstrate a relative deficiency of singletons, and Tajima's

D estimates were all above zero (Appendix 3.4, Appendix 3.5), indicating that all populations went through a population decline at some point in the past. We used Relate to determine genome-wide geneaologies and used the coalescent rate to model changes in N_e for each individual population over the last ~500,000 years (Figure 4c) as well as cross-coalescent rates between pairs of populations to model periods of reduced gene flow (Appendix 3.6). All populations followed the same declining N_e trajectory until approximately ~500-1000 years ago, presumably reflecting the positive Tajima's D values. Populations appeared to then begin diverging and all populations experienced significant increases in N_e. Danco Island and Hannah Point had the highest recent N_e, while Port Charcot and Tuxen Rocks followed by Joubin Island had the lowest, supporting the observation of increased ROH in these latter populations (Figure 3.4c). Among populations from different clusters, cross-coalescent rates showed clear evidence of reduced population gene flow. While most pairs of populations within clusters were indistinguishable, a notable exception was the west Anvers cluster which showed some evidence of reduced gene flow between Joubin Island and Gerlache Point, the former appeared to have a lower N_e in the present. Interestingly the timing of onset of reduced geneflow between all five clusters appears quite similar, suggesting the expansion across the WAP occurred quite rapidly.

3.4 Discussion

For the first time, the ongoing climate-driven range expansion of gentoo penguins in Antarctica was tracked genetically, revealing a stepping-stone model of dispersal buffered against founder effects. We characterized the population genetic structure and evolutionary history of gentoo penguin colonies along the WAP using a high-coverage whole genome sequencing approach. We identified a region of fine-scale genetic differentiation among five discrete population clusters that correspond closely to geographic location (Figures 3.1a and b). Our results suggest that Anvers Island was likely colonized by gentoo penguins that originated in the South Shetland Islands, and then continued to disperse east quite rapidy in a stepping-stone pattern along the southern coast of Anvers Island before colonizing the mainland of the WAP. gentoo penguins from the northern cluster and southern cluster were likely sourced by individuals from the cluster east of Anvers Island (Figure 3.2, Figure 3.3), identifying the Gerlache Strait as a potential pathway of dispersal for the northern cluster, and the Bismarck/Gerlache Strait as a potential pathway of dispersal to the southern cluster (Figure 3.5). Our results also indicated that the most recently

established colonies sampled at the range edge, Moot Point and Tuxen Rocks, were likely sourced from the closest adjacent colonies, Port Charcot and Petermann Island. Thus, gentoo penguin southern range expansion is likely continuing in a stepping-stone pattern of dispersal, with the leading-edge of expansion acting as the source of colonizing individuals for continued expansion. Surprisingly, a vagrant individual found at Armstrong Reef likely came from the Danco Island in the northern cluster (Figure 3.1b and 3.1c, Appendix 3.3), suggesting that gentoo penguins may sometimes disperse beyond the expected stepping-stone pattern, though this individual was not breeding.

The northern and southern clusters did not share any IBD segments despite being the most genetically similar, whereas the southern cluster shared IBD segments with the east Anvers cluster. One possible model of dispersal could be that colonization of mainland WAP likely stemmed from a single historical colonization event on the WAP. Individuals dispersed in a stepping-stone pattern north along the mainland WAP, eventually colonizing the northern cluster and became genetically isolated through genetic drift. The presence of multiple established gentoo penguin colonies along the mainland WAP south of the northern cluster supports this model concept (Figure 3.5). Meanwhile, gentoo penguins from the same single colonization event on the mainland WAP began dispersing south and colonizing the area at the range edge where the southern cluster is located, resulting in divergence between the north and south (Figure 3.5). Some post-divergence gene flow from the west and east Anvers Island cluster also occurred during this southward expansion, supported by our IBD results (Figure 3.3a). This model is supported by the EEMS visualization showing patches of reduced gene flow between the northern and southern clusters, indicative of the northern cluster becoming isolated over time, and slight extension of increased gene flow from the western Anvers cluster towards the southern cluster with no patches of reduced gene flow between them (Figure 3.3b).

The earliest known presence of gentoo penguins on the Antarctic Peninsula was at Potter Peninsula on King George Island dating back to the mid-Holocene, or 4450-4550 B.P. (del Valle et al. 2002), while the earliest known gentoo penguin breeding colony was located on Byers Peninsula on Livingston Island dating back 1150 B.P. (Emslie et al. 2011), at least 2000 years after the island was deglaciated (Ingólfsson et al. 2003). Emslie et al. (2011) suggested that gentoo penguins may have been slow to colonize recently deglaciated regions at that time, which is supported by our findings of a signal of population expansion started roughly 500-800 years ago

after a decline in effective population size since approximately 20,000 BP. While Relate results are highly dependent on input parameters and should therefore be interpreted with caution, the decline in historical population size may have coincided with the last glacial maximum (LGM) during which the ice shelf extended to the continental shelf (Davies et al. 2012), making the region unsuitable for gentoo penguins that had to take refuge on the closest sub-Antarctic islands such as South Georgia and the South Sandwich Islands. In contrast, two studies that used mitochondrial marker HVR-1 (Peña et al. 2014; Clucas et al. 2014) detected signals of population expansion estimated around 13,000 and 15-20,000 years ago, respectively, suggesting that the populations were expanding soon after the last LGM out of glacial refugia. This order of magnitude difference in results may be due to known issues with mitochondrial markers as unreliable for inferring population history without the inclusion of fossil calibrations (Galtier et al. 2009; Allio et al. 2017). Whole-genome sequencing, however, provides high resolution with greater accuracy for inferring demographic histories, and our results likely provide an update on these historical population sizes and demographic history of gentoo penguins to the literature. Furthermore, the lack of paleofossil evidence prior to 4550 BP also supported our inferred timing of historical population expansion (del Valle et al. 2002; Emslie et al. 2011), though it is possible that earlier fossil evidence has yet to be discovered.

Species that have shifted or expanded their ranges from low latitudes to higher latitudes tend to exhibit a loss of genetic diversity due to founder effects (Hewitt 2000). The gentoo penguin colonies sampled did not exhibit any major bottleneck signatures indicative of founder effects, which is also associated with the classic stepping-stone model. Instead, the colonies maintained relatively high and similar genetic diversity throughout the stepping-stone colonization process, suggesting they colonize new locations in large numbers or with continued migration for a period of time. This is true even of the southern cluster colonies which are known to be recently established yet have only a moderate increase in ROH and short between-colony IBD segments consistent with only a minor decrease in effective population size. Herman & Lynch (2022) found that continued - and in some cases increasing – immigration is essential for observed population growth in recently colonized gentoo penguin colonies, which is supported by our findings. This may help to buffer against Allee effects, as colonial seabird fledglings are vulnerable to predation and larger cohorts reduce the probability of being depredated (Schippers et al. 2011). In the closely related Adélie penguin it has been shown that fledgling survival is inversely related to cohort size

due to a combination of increased predation risk by aerial predators at the colony, by marine predators at sea, and by reduced cooperative behavior for foraging efficiency (Emmerson & Southwell, 2022). The ongoing dispersal behavior of gentoo penguins may therefore be an evolutionary adaptation that has allowed the species to increase in number and expand their range in recent decades, demonstrating robustness to environmental changes and predator presence.

A meta-analysis by Friesen et al. (2007) found that isolation by distance was present in many colonial seabird species, but evidence of long-range colonization in other species suggested that distance is not a universal barrier to dispersal. The gentoo penguin sampled at Armstrong Reef that likely came from Danco Island, approximately 215 km to the north, demonstrated this paradox. Seabirds are known to carry out prospecting behaviors to evaluate future potential breeding locations, with this behavior particularly common in pre- or failed breeders (Campioni et al. 2017, Ponchon et al. 2017). Individual variation in the spatial scale of prospecting and dispersal behaviors is found throughout seabirds (Phillips et al. 2017, Ponchon et al. 2017) and it is possible that such variation in gentoo penguins may be inherent plasticity in response to environmental changes. While there is evidence of selection against immigration in wild seabird populations (Barbraud & Delord 2021), vagrancy or dispersal in seabirds can occur in response to changes in local environmental conditions (Garnier et al. 2023) or increasing population sizes of the source population (Acosto Alamo et al. 2022). This suggests there is likely an adaptive advantage in individual variation in dispersal behaviors that override distance barriers, and gentoo penguins may sustain this variation in behavior.

In conclusion, our study showed a fine-scale and complex network of historical and recent dispersal events indicative of the stepping-stone model of dispersal, illustrating how gentoo penguins have spread and colonized along the Western Antarctic Peninsula. However, the detection of post-divergence gene flow and lack of population bottlenecks suggests a less dramatic form of the classic stepping-stone model. This method of dispersal likely serves as a buffer against founder effects, exemplified by the robustness of gentoo penguin populations and their continued southward expansion. To our knowledge, this study is the first to use whole-genome resequencing to investigate such fine-scale networks of dispersal in a species undergoing range expansion. Harnessing high-resolution genomic data to explore such systems provides an effective tool for determining dispersal networks in species that are otherwise difficult or impossible to track using alternative methods. Such results provide a foundation for the forecasting of distributions of

species undergoing range expansion and shifts in response to climate driven environmental changes. Our findings of sustained genetic diversity and population growth at the range edge, coupled with evidence of more southerly prospecting, highlight the need for ongoing monitoring of gentoo penguins on the West Antarctic Peninsula as it is clear their range expansion may be ongoing.

3.5 Tables and Figures

Figure 3.1 A) Map of colony locations on the Western Antarctic Peninsula. Each circle or square represents a colony, each circle color represents a colony established prior to 20 years ago, and each square color represents a recently established colony within the past 20 years. Black dots on inset represent all known gentoo penguin colonies. Color patterns are used throughout the manuscript to indicate colonies. **B**) Principal component analysis results from smartPCA of 129 individuals projected onto PC axes 1 and 2. **C**) ADMIXTURE results with k=5. Bars indicate individual samples and vertical white lines separate individual sample columns by colony. Colors show ancestry clusters with very little admixture. gentoo photo by R. Herman.
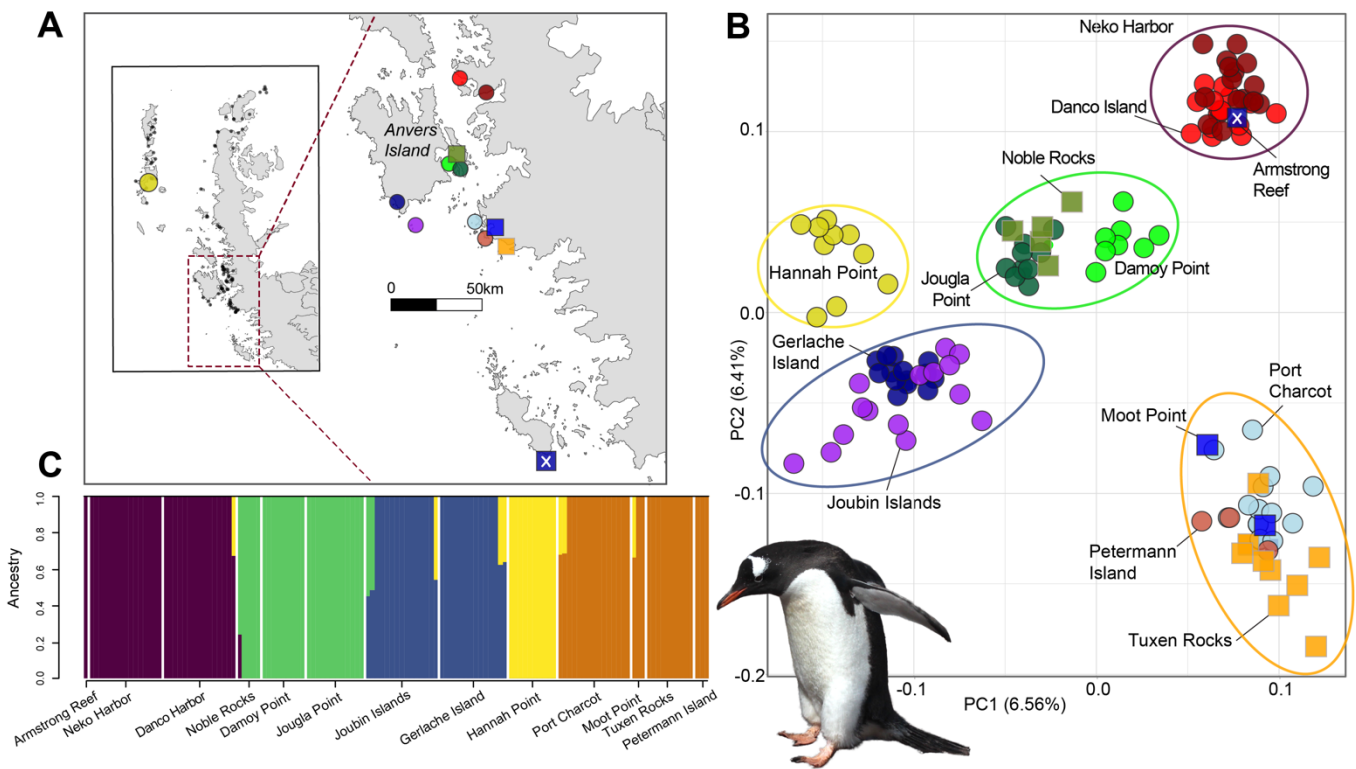
Figure 3.2 Treemix analysis that measures historical relationships among populations. Lengths of branches equal level of genetic drift.
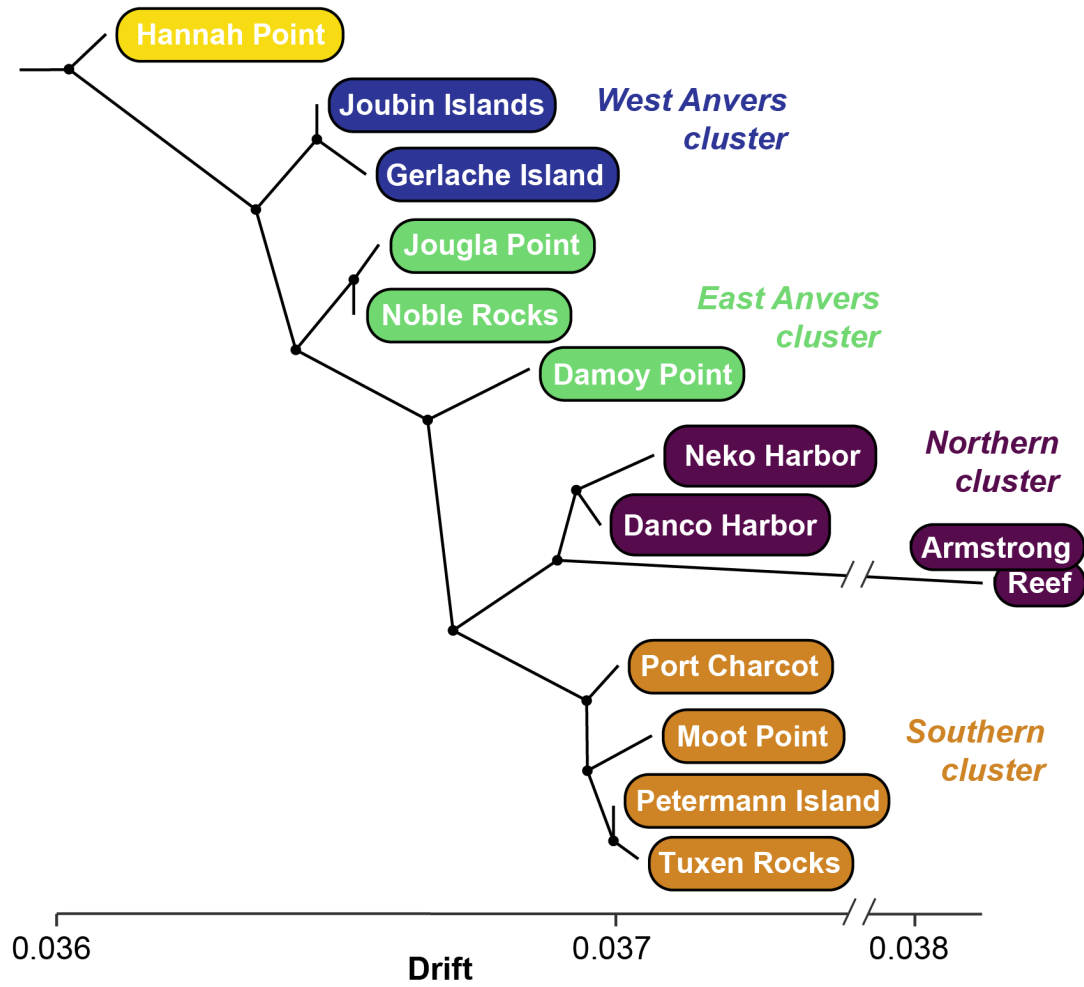
Figure 3.3 A) Gephi network analysis using RefinedIBD shared IBD segments. **B)** EEMS results visualization of gene flow. Colors indicate respective colony locations consistent with Figure 1.
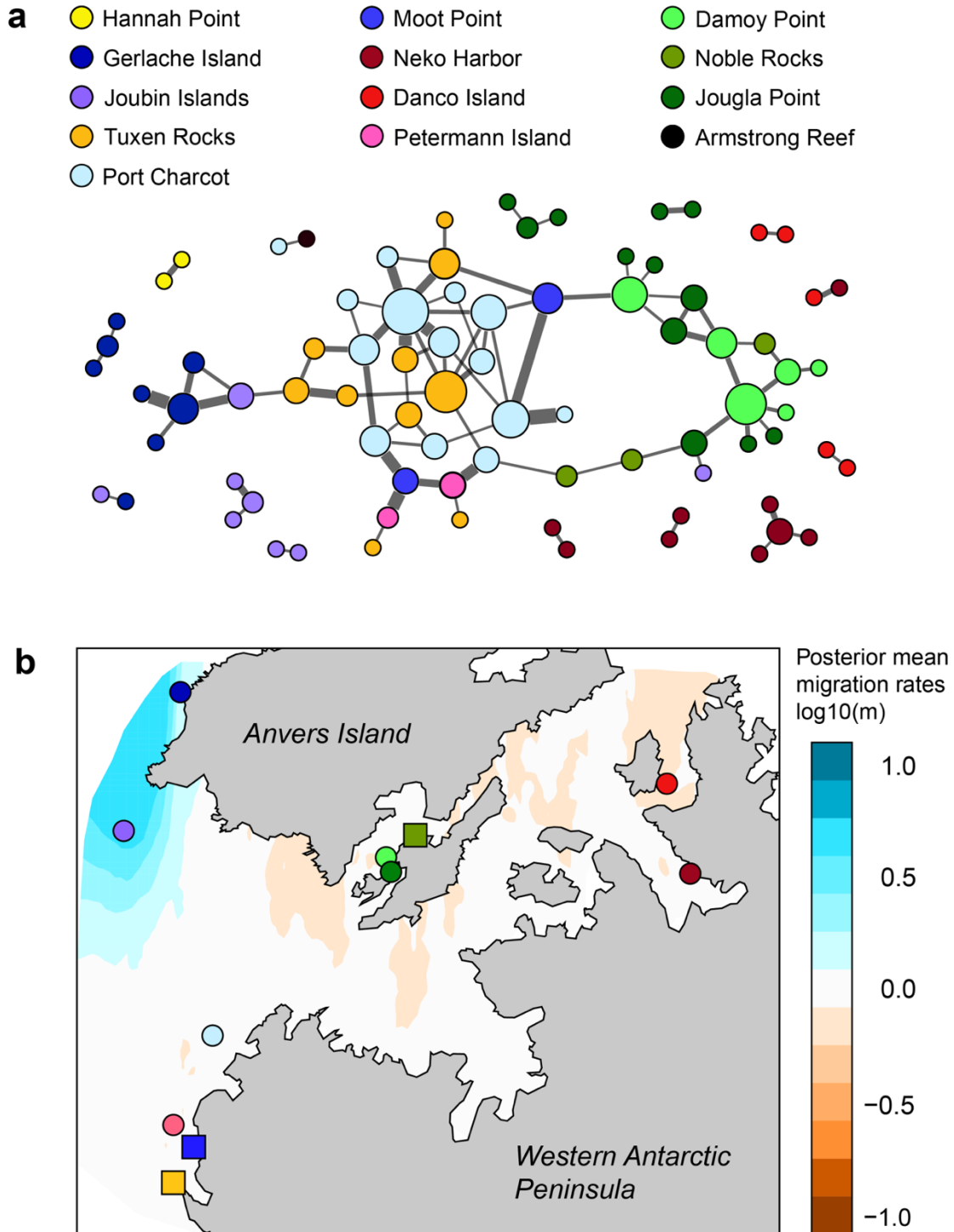
Figure 3.4: A) Violin plots showing sums of all runs of homozygosity for each population. **B)** Violin plots showing sums of runs of homozygosity greater than 5cM for each population. **C)** Skyline plots of effective population size (Ne) through time for each population in the last 500,000 years.
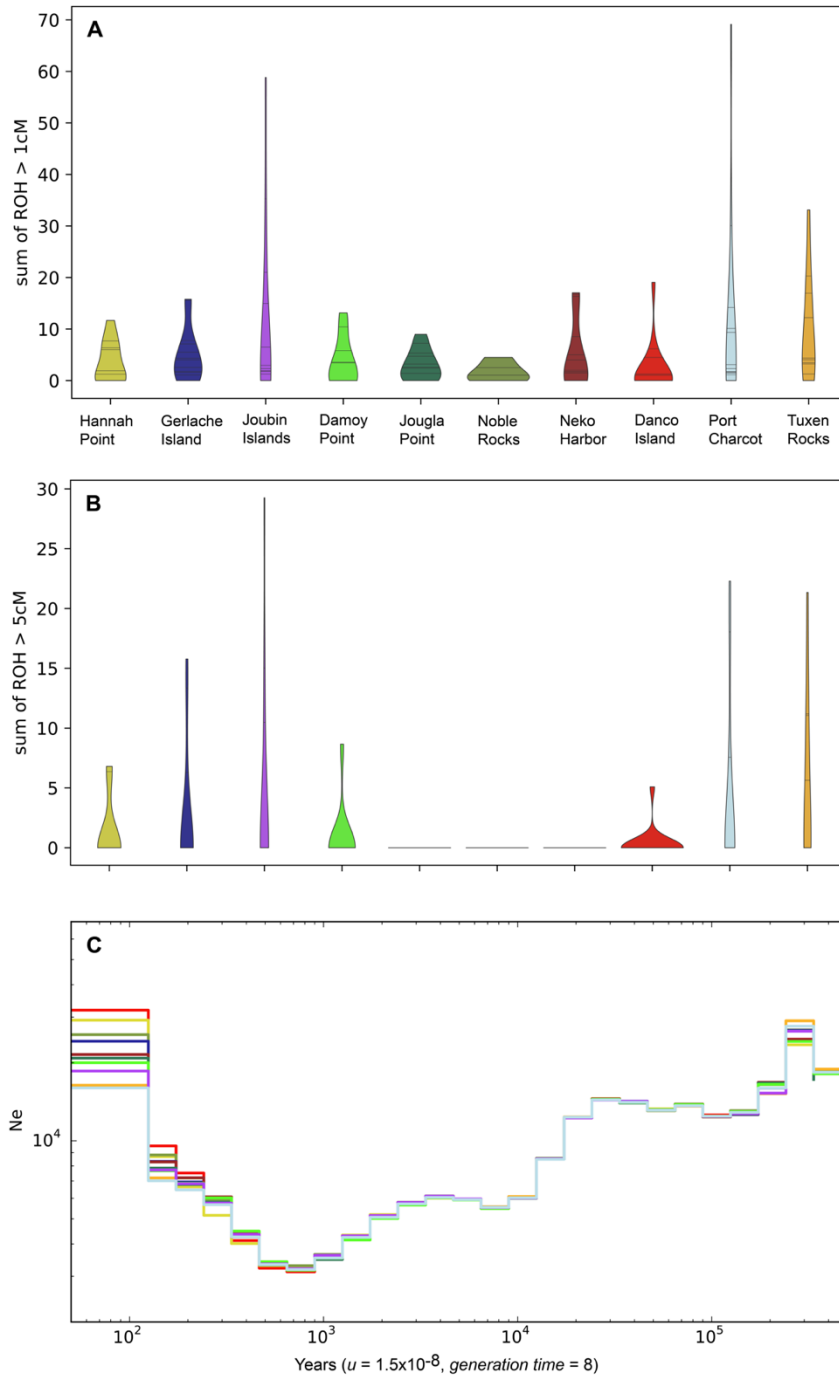
Figure 3.5 Map of inferred dispersal pathways (red arrows) among major gentoo penguin population clusters. Black circles indicate all other gentoo penguin colonies. Pink shaded region indicates post-divergence gene flow.
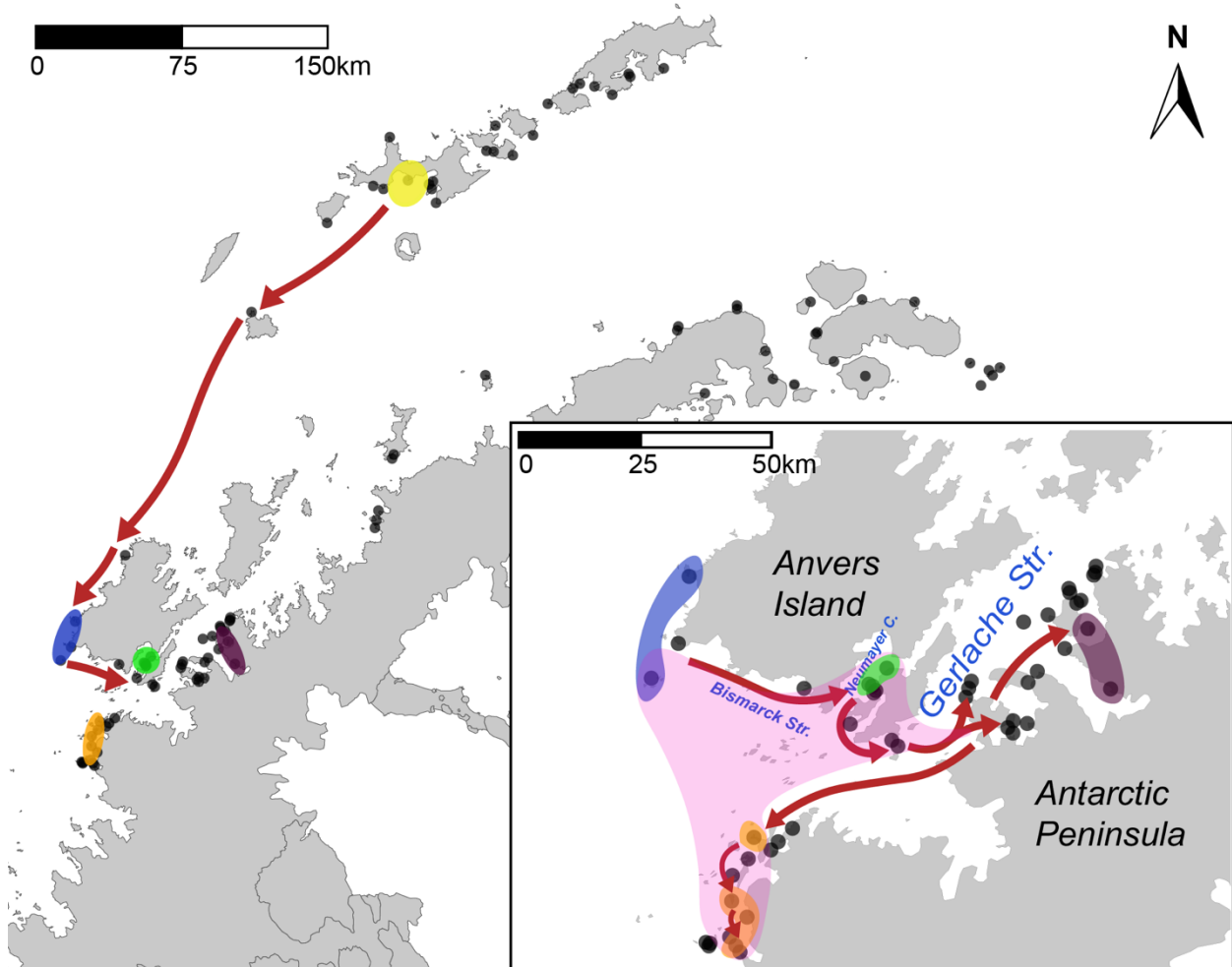
Chapter 4: Forecasting colonization events and range expansion of gentoo penguins along the Western Antarctic Peninsula using agent-based modeling

4.1 Introduction

Understanding the ecological consequences of climate change in rapidly transforming regions has become a major focus in the fields of ecology and conservation. The Western Antarctic Peninsula (WAP) has emerged as a hotspot for investigating the ecological consequences of climate change due to its unparalleled warming rates compared to the rest of the Antarctic continent (Vaughan et al. 2003). gentoo penguins (*Pygoscelis papua*) have become a focal species of interest, as this species has been both rapidly increasing in breeding population size along the WAP and expanding its range further south since the early 1990s (Lynch et al. 2012; Herman et al. 2020). In contrast to several Antarctic species facing population declines (Forcada et al. 2006), gentoo penguins appear to be climate change "winners", prompting further investigation into the underlying mechanisms driving this species' ability to adapt to shifting environmental conditions (Clucas et al. 2014).

Previous research has shown that the rapid population growth observed among gentoo penguins at recently colonized territories is likely driven by sustained and even accelerating immigration over many years (Herman and Lynch 2022). This has also been supported by genomic research that found genetic diversity to be similar across the majority of colonies along the southern range edge. This suggests gentoo penguins are buffered against the typical founder effects expected during colonization events (Herman et al. *in Review*). This capacity for synchronous movement among sites may allow them to continue colonizing novel breeding locations within and beyond their current range and exploit newly suitable habitats created by the WAP's highly variable conditions. Whereas other colonial seabirds with high site fidelity tend to only occupy portions of available suitable breeding habitat available to them (Matthiopoulos et al. 2005), gentoo penguins may have more plasticity in their level of site fidelity, enabling them to exploit newly suitable habitats created by the WAP's highly variable conditions.

Forecasting animal movement and range expansion relies on comprehensive and up-to-date data, which can be particularly difficult to obtain in inaccessible regions that are undergoing rapid environmental change. This is particularly true for studying species like gentoo penguins in

the Antarctic, where traditional fieldwork is often logistically challenging and expensive. Agent-based models (ABMs) are a powerful tool for predicting movement, dispersal, and range shifts. By simulating the behavioral decisions and movements of individuals, ABMs naturally accommodate individual-level interactions with others as well as with the environment. (Grimm and Railsback 2006; Railsback and Grimm, 2012; DeAngelis and Grimm, 2014).

Several previous studies have investigated oceanographic and climatic variables that drive gentoo penguin breeding distribution patterns and southward range expansion (Cimino et al, 2013; Lynch et al. 2012; Pertierra et al. 2020; Korczak-Abshire et al. 2021). While a few studies have investigated the role of the terrestrial environment on nesting in Adélie penguins (McDowall and Lynch 2019; Schmidt et al. 2021) we are aware of no studies on the gentoo penguins. Applying ABMs to investigate this species may facilitate a more mechanistic approach that can incorporate demography, behavior, the patchiness of suitable habitat, and the stochasticity that is known to dominate the Antarctic environment. For example, McDowall and Lynch (2019) used ABMs to demonstrate that nest aggregations observed in Adélie penguin (*P. adeliae*) colonies were the result of complex interactions among self-organizing dynamics and landscape terrain characteristics, sometimes leading to sub-optimal patchiness.

Here we present the first attempt to systematically predict future occupancy of new gentoo penguin breeding habitat south of their current range edge along the WAP. We apply a custom agent-based modeling approach using a combination of geological, oceanographic, and ecological variables to forecast dispersal and colonization events. As climate change continues to shape the dynamics of the Antarctic ecosystem, understanding the likely trajectory of gentoo penguin range expansion into previously unoccupied territories will be essential for informed decision-making and sustainable resource management.

4.2 **Methods**

Our model simulates the dispersal and colonization of gentoo penguins in a realistic terrestrial and oceanographic space that reflects the features and characteristics of the southern breeding range (Figure 4.1). It encompasses all known established colonies and extends from the lower half of Anvers Island to just south of Renaud Island along the WAP. We determined the northern boundary based on genomic evidence that southern range edge colonies were not sourced by individuals from colonies further north and the southern boundary based on the presence of a single

gentoo penguin vagrant at Armstrong Reef (Herman et al. *in review*). Population counts of all colonies (specifically, counts of active nests or number of chicks during the breeding season) are available at the Mapping Application for penguin Populations and Projected Dynamics (MAPPPD; http://www.penguinmap.com; Che-Castaldo et al. 2023).

Within the study area, we restricted all terrestrial habitat to areas with exposed bare rock using a high-resolution Landsat 8-derived rock outcrop shapefile map (Burton-Johnson et al. 2016). We included only bare rock features that were within 1km of the coastline, as gentoo penguins do not typically nest further inland. We applied a 100 x 100m grid to the study area and extracted centroid values for each grid pixel that contained bare rock. All pixels were assigned an ID number, and pixels that were located within known colony locations were marked as "present". These areas were confirmed by visually inspecting Landsat imagery for gentoo penguin guano stains which reflect the presence of breeding colonies. All other bare rock pixels were marked as absent. We used current breeding census counts for every known colony within the study area to populate all bare rock pixels marked as "present" (Che-Castaldo et al. 2023). For example, if a colony's breeding count was 200, and there were ten pixels designated within that colony, we assigned 20 individuals to each of the ten pixels. We also applied this to bare rock pixels not currently inhabited by assigning all to pixels to new potential colony clusters using the k-means function in R and setting a Euclidean distance of 1km for the cluster assignment step.

We extracted mean elevation, slope, and aspect values within each pixel derived from 2-meter digital elevation models provided by the Polar Geospatial Center (Polar Geospatial Center 2023). We also incorporated areas of krill density hotspot persistence. Krill hotspots are defined as the number of times during the chick-rearing period (December – March) simulated krill accumulation within a grid cell was greater or equal to the mean regional accumulation (Santora and Veit 2013), and hotspot persistence is defined as the proportion of observations the grid cell was tagged as a hotspot across the years 2006-2012, and 2018-2020 derived from krill density simulations using the WAP version Regional Ocean Modeling System (ROMS) (Haidvogel et al. 2008; Graham et al. 2016; Hudson et al. 2021; Hudson et al. 2022). We used 0.25 as the threshold for a high hotspot persistence and calculated the number of years that hotspot persistence within a 10km grid cell was equal to or above this value. This final value was presented as the fraction of years a grid cell was above this threshold.

We ran a logistic regression analysis to evaluate the above covariates for their predictability

of gentoo penguin pixel occupancy. We confined this analysis to pixels within the known current southern range distribution. We modelled occupancy using the following:

$$Y_i \sim Bernoulli(P_i)$$

$$\log\left(\frac{P_i}{1 - P_i}\right) = \beta_0 + \beta_1 X_{slope} + \beta_2 X_{elevation} + \beta_3 X_{aspect.dir} + \beta_4 X_{meanHP}$$

where we include slope, elevation, aspect direction (*aspect.dir*), and mean hotspot persistence (*meanHP*) as covariates. We then ran the predict base function in R using the above model output to fit the covariate values for every pixel within the bounded study area. The resulting probability values of occupancy per pixel were then used as the suitability indices.

Figure 4.2 depicts the ABM procedures (survival, reproduction, dispersal, and choice of dispersal location) executed by each gentoo penguin individual in each time step, where one time step is equivalent to one year. To realistically simulate the phenology and movement of individuals over time, we parameterized our model using published apparent survival rates for gentoo penguins such that adult survival was the result of a Bernoulli trial with probability of 0.82 and survival to the first and second years were a result of a Bernoulli trial with probability of 0.43 (Williams 1995). We allowed for individual to start reproducing at year two based on published observational data (Lynch et al. 2012). To account for Allee effects, we parameterized individual's probability of reproduction such that individual fitness was a function of the size of the population. For each individual reproductive procedure, if the size of the population within an individual's current colony location was less than five, we conducted a Bernoulli trial with a probability equal to the total population size divided by five. If the size of the population was greater than or equal to five, we assigned a probability of 0.9 for the Bernoulli trial. We based this on observations of early colonization events where 2-3 pairs of gentoo penguins were successful at reproducing. Annual productivity (chicks créched per nest) ranges from 1.23 to 1.52 (Lynch et al. 2010), but we chose to allow for one chick to be produced per successful reproduction to reduce model complexity. We applied a dispersal probability of 0.4 based on an estimate of 60% site-fidelity observed at a gentoo colony at King George Island (Williams and Rodwell 1992).

If an individual dispersed, all bare rock pixels within a 50km radius of the individual's current location were binned per 10kms and assigned a distance penalty to facilitate a stepping-

stone pattern during dispersal decision making. These distance penalties per pixel were then multiplied by each respective pixel's suitability index and combined as multinomial probabilities from which one multinomial trial was conducted to select the final pixel location of dispersal (Figure 4.2).

To validate the model, we populated pixels using colony breeding census counts prior to 1994, when only 15 of the 27 current colonies existed in the study area. We ran the ABM simulation for twenty time-steps (twenty years) for 1000 replicates to compare to the current colony distributions. We calculated the mean and variance of the number of individuals that populated a given pixel after twenty years across all 1000 simulations. After model validation, we ran the ABM simulation for twenty years using the current colony breeding census counts as of 2023 to forecast future gentoo penguin range expansion and colonization events. All model development was conducted in R version 4.2.2 (R Development Core Team 2017) and the complete code is provided in Appendix 4.1.

4.3 Results

Results from the logistic regression indicated that gentoo penguin occupancy significantly increased with proximity to hotspot persistence ($p < 0.001$), while significantly decreased with slope ($p < 0.001$) and aspect ($p < 0.001$). Elevation was not significant ($p = 0.92$). We therefore included krill hotspot persistence, slope, and aspect in the final suitability index calculations for both validation and forecasting model runs.

The ABM was able to reproduce the colonization of eight colonies between 1994 and 2023 with low variance across replicates. (Figures 4.3, 4.4). However, the simulated population growth of many of these colonies was much slower than the true time series counts. The model did not simulate any colonization events in areas inside the coastal inlet east of Anvers Island, reflecting major areas where gentoo penguins are not observed to breed, despite the presence of bare rock. The model did simulate small colonization events throughout Waewermans Islands – an area with no known gentoo penguin breeding colonies. Despite this, the model performed well and reproduced the general spatial patterns of colonization in this region.

The twenty-year ABM forecast resulted in further expansion west and southwest of the gentoo penguin southern range edge, comprising 16 new colony locations (Figure 4.5). Again, the model did not forecast any colonization events in areas inside the inlets east along the WAP

coastline. The forecast also resulted in the colonization of Waewermans Islands, similar to the validation model. Many areas along the southern coastline of Anvers Island were also colonized in the forecasting simulations as well.

4.4 Discussion

Modeling what determines species distributions and colonization events is important for improving our understanding of ecological systems that are undergoing rapid change due to climate shifts and extreme events. Identifying important predictor variables for colonization events can provide a foundational model to test the future of an ecosystem under various climatological scenarios. We used an ABM to determine what and how individual decisions and movements of gentoo penguins facilitate their patterns of colonization and range expansion. We demonstrated that our model could reproduce individual and population level patterns of dispersal and colonization under realistic terrestrial and oceanographic conditions (Figures 4.3, 4.4). We also forecasted colonization twenty years into the future and identified areas that gentoo penguins are predicted to colonize as they continue to expand their range (Figure 4.5).

The proximity to persistence of krill hotspots seems to the strongest predictor of gentoo penguin presence (Figure 4.5). This is not surprising considering Antarctic krill is a major prey source for this species. Lack of hotspots in the eastern inlets along the main coastline may be the reason why gentoo penguins have not colonized these areas despite the presence of accessible bare rock features. gentoo penguins are known to be central-place foragers, and while they have a more flexible diet relative to their sister species, they tend to forage primarily on krill during the chick rearing phase (Wawrzynek-Borejko et al. 2022). It would therefore make sense to choose breeding locations that are in close proximity to temporally reliable food sources. It is unclear how gentoo penguins determine that a breeding location is suitable due to proximity to a persistent prey source, particularly from year to year. It is possible that successful chick rearing may be an indicator to this species, and only over time can breeding pairs decide whether a breeding location is suitable enough.

It remains unclear why gentoo penguins do not breed in the Waewermans Islands. The Waewermans may be too exposed and steep for gentoo penguins, despite including slope as a covariate in our model. It is also possible that these islands have been colonized, and the colonies have escaped detection over the last several years in which surveys have been limited by the Covid-

19 pandemic, or that they will be colonized in the near future. Certainly, our results identify a clear set of testable hypotheses about where colonization is likely to occur, and only future monitoring of these areas will tell us for sure whether the model is making accurate predictions or whether additional covariates, such as fetch (Schrimpf et al. 2021), are required. It is also possible that our model populated these islands as well as many additional locations along southern coastline of Anvers Island in between much larger established colonies due an absence of stronger colonial aggregation incentive. Adelie penguins are known to have strong self-organizing aggregations in their breeding colonies, sometime resulting in suboptimal configurations, that are largely driven by nest site fidelity and predator avoidance (McDowall and Lynch 2019). While our model did incorporate breeding success based on Allee effects, we did not include any individual decision-making based on spatial aggregations or nest density. This has been demonstrated as important in other seabirds (Matthiopoulus et al. 2005; Robert et al. 2014) though the structure of an agent-based model naturally accommodates such behaviors. By identifying important terrestrial and ecological features that drive the presence and absence of gentoo penguins, our ABM provides concrete testable predictions about future colonization events for gentoo penguins moving southward along the Western Antarctic Peninsula and, in doing so, may help to inform the management of krill fishing management and Antarctic tourism.

4.5 Tables and Figures

Figure 4.1 Map of study boundary area (red border) and locations of all known gentoo penguin colonies (yellow circles). The purple circle marks location of vagrant individual observed at Armstrong Reef.
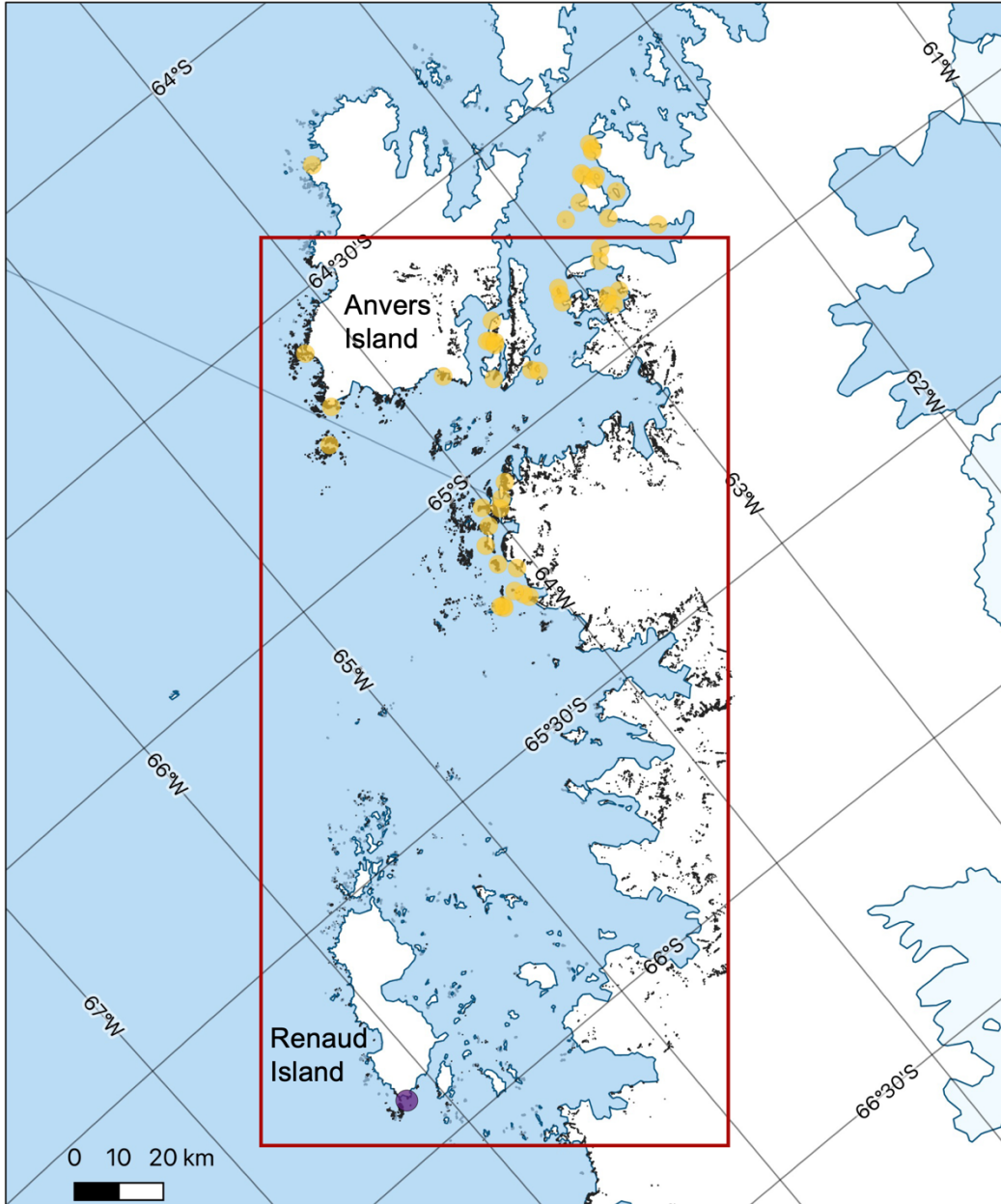
Figure 4.2 Agent-based model procedures schematic.
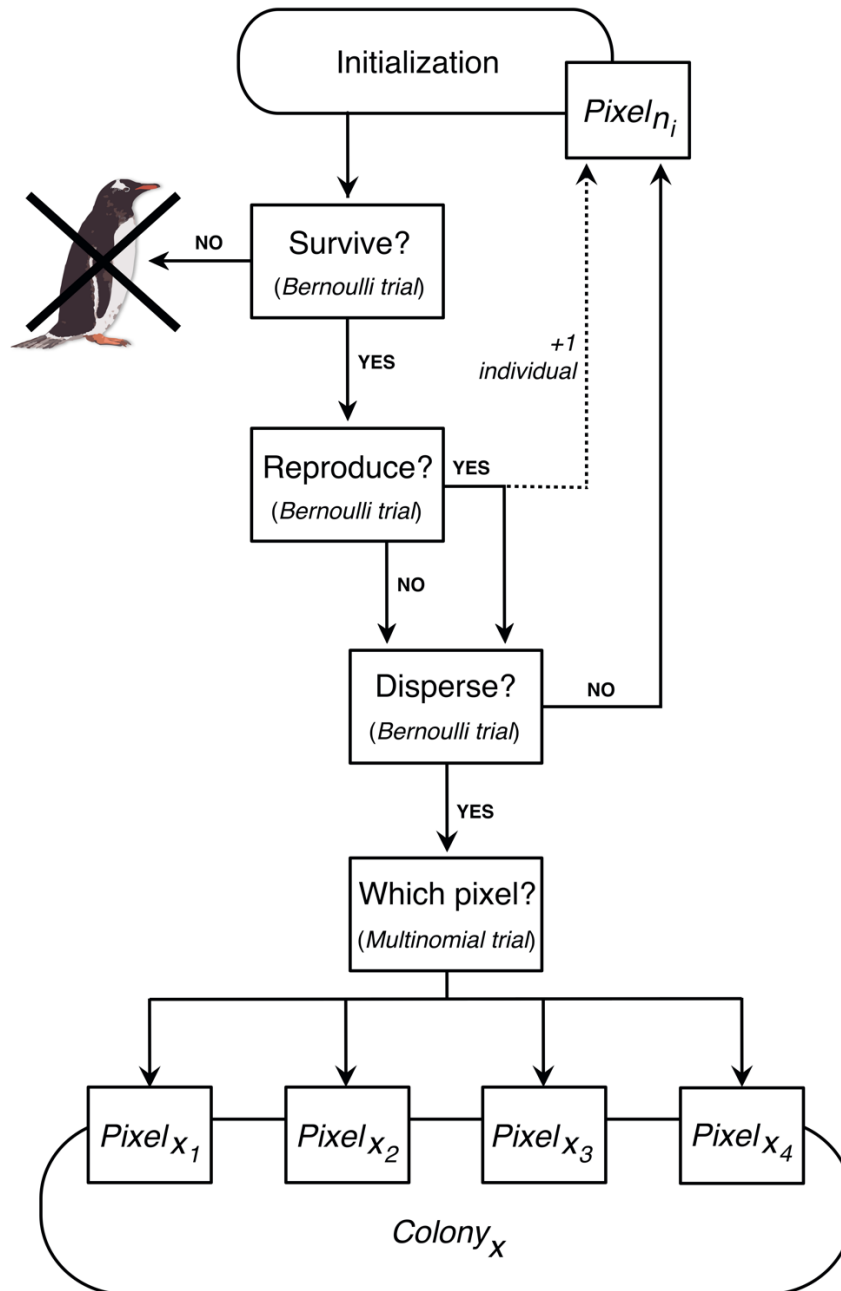
Figure 4.3 ABM model validation results of individual pixel colonization and abundance. The top panel displays colony locations as of 1994 and respective population sizes. The bottom panel shows model results of 20-year simulation since 1994. Colors represent variance across simulations per pixel, and size of circles reflects mean number of individuals per pixel. Pixels with three individuals and lower were omitted from the figure for clarity.
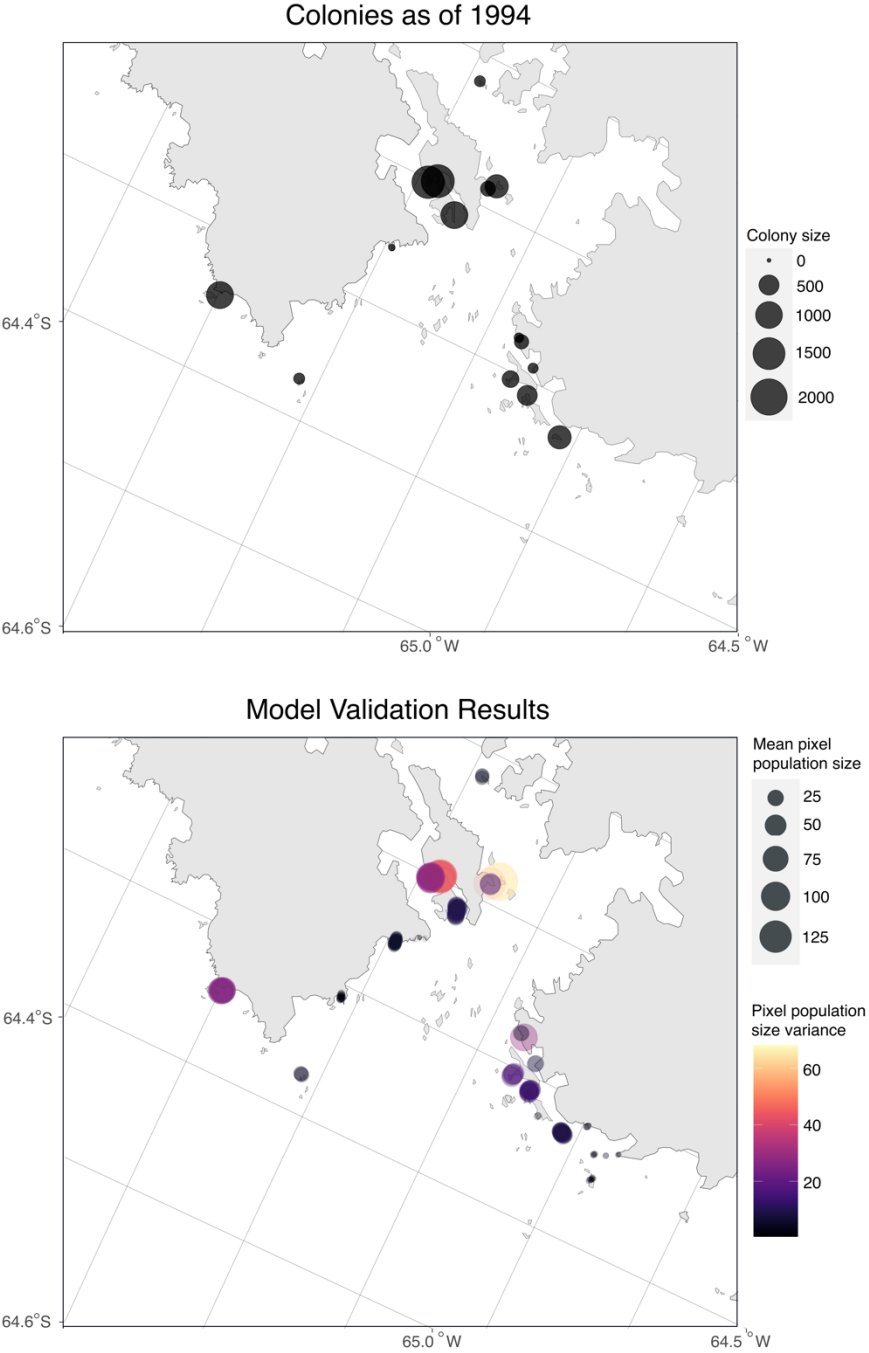
Figure 4.4 ABM model validation results of clustered colony aggregates and abundance. The top panel displays colony locations and population sizes as of 2023. The bottom panel shows model results of 20-year simulation since 1994 to 2023. The bottom panel displays the mean population sizes per colony cluster across simulations.
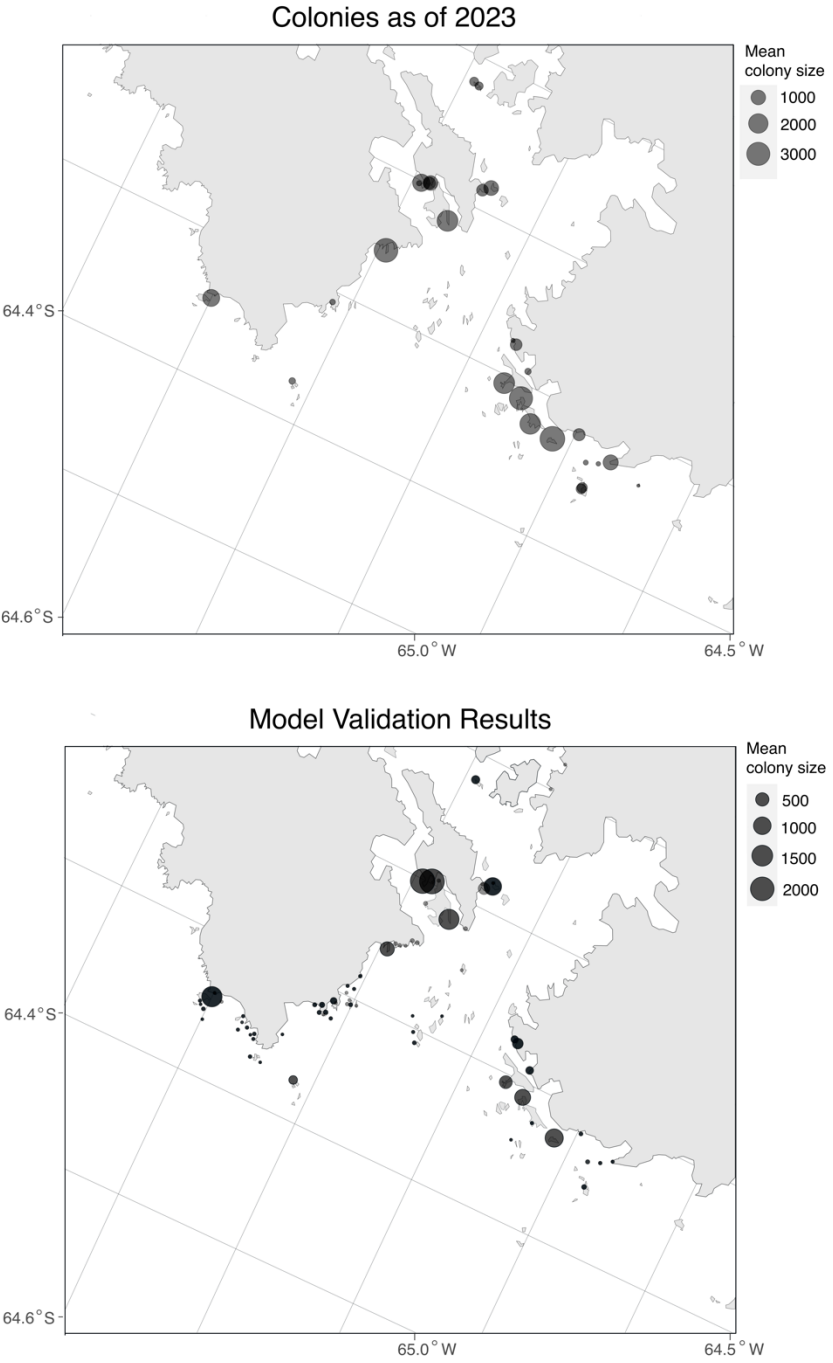
Figure 4.5 20-year forecast results and krill hotspot persistence distributions.
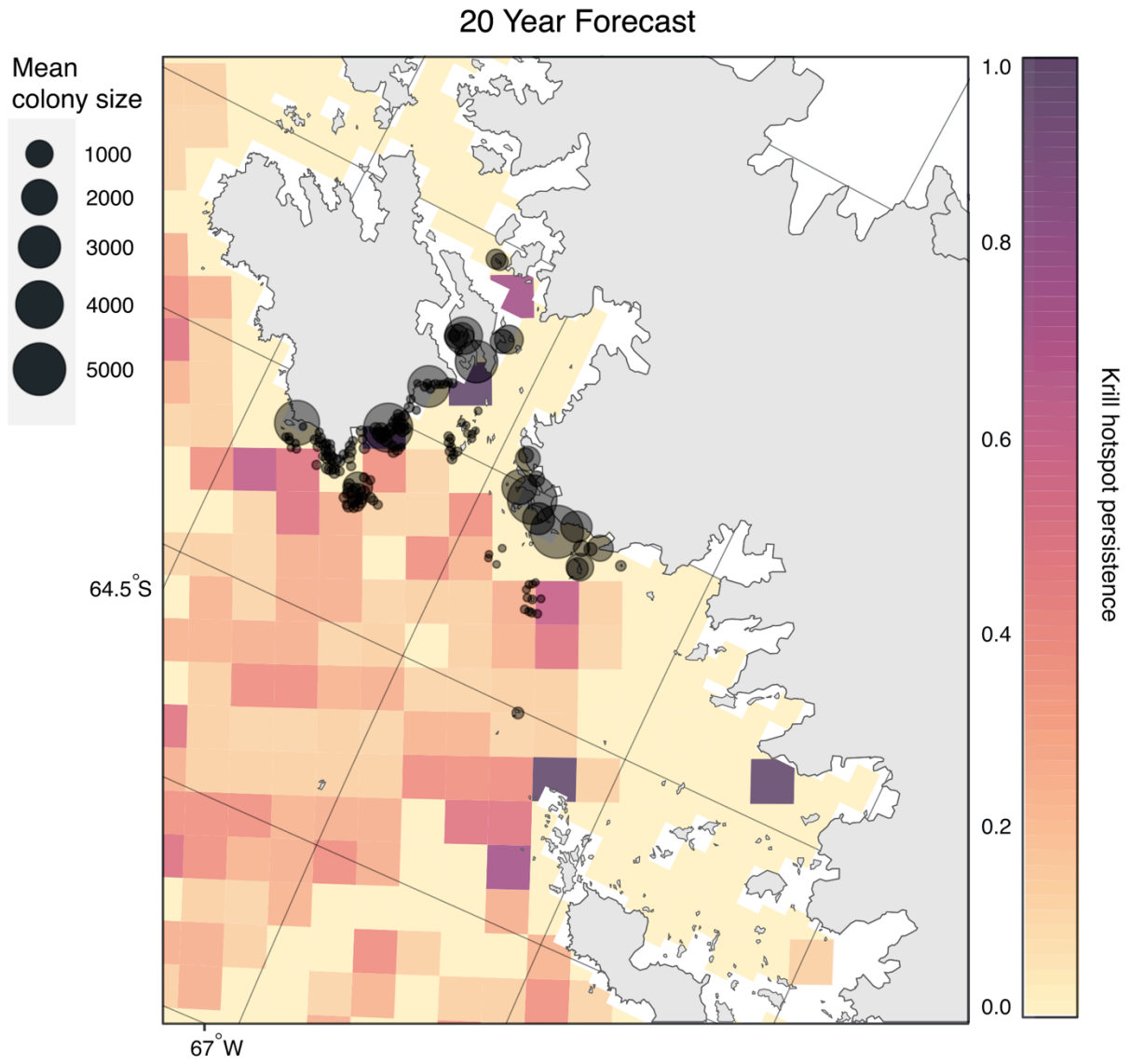
Conclusions

The main objectives of this dissertation were to investigate the mechanisms of gentoo penguin southern range expansion on the WAP, and what characteristics allow them to be successful colonizers during a time of rapid climate change and environmental variability. In the first chapter, we produced a comprehensive review and update on the global distributions and population sizes of gentoo penguins, accounting for many new colonies and recent population growth trends. For the second chapter, we utilized the updated population growth trends to model the contribution of immigration to rapid population growth observed in recently colonized breeding locations. Using a combination of population time series and age-structured matrix model within an approximate Bayesian computation framework, we determined that sustained immigration over several years was required to generate the rapid population growth observed, with some sites even showing evidence of an accelerating immigration rate following initial colonization. We demonstrated that our method is capable of estimating the contribution of immigration to population growth in a species where mark-recapture datasets are unavailable.

In the thirds chapter, we conducted a fine-scale population genomics study of gentoo penguins near their southern range edge using whole genome sequencing. Through a detailed examination of fine-scale population structure, admixture, and population divergence, we inferred that gentoo penguins historically dispersed rapidly in a stepping-stone pattern from the South Shetland Islands leading to the colonization of the Palmer Archipelago, and then the adjacent mainland Western Antarctica Peninsula. Recent southward expansion along the Western Antarctic Peninsula also followed a stepping-stone dispersal pattern coupled with some post-divergence gene flow from colonies on the Palmer Archipelago. Genetic diversity appeared to be maintained across colonies during the historical dispersal process, and range edge populations are still growing, suggesting continued gene flow and high numbers of migrants provided a buffer against founder effects typically expected in the classic stepping-stone model. These results coupled with a continued increase in effective population size since approximately 500-800 years ago distinguishes gentoo penguins as a robust species that is highly adaptable and resilient to changing climate.

In the fourth and final chapter, we combined the findings of the first three chapters to inform and parameterize an agent-base model to forecast continued gentoo penguin colonization events and range expansion along the WAP. We identified important terrestrial features such as bare rock, slope, and aspect direction as significant predictors of gentoo penguin breeding occupancy. We also identified that the locations of year-to-year persistence of krill abundance hotspots were also significant predictors of gentoo penguin occupancy. We demonstrated that our model could reproduce individual and population level patterns of dispersal and colonization using these predictor variables. Finally, we forecasted colonization twenty years into the future and identified areas that gentoo penguins are predicted to colonize as they continue to expand their range.

The combination of these four chapters provides a detailed study of the population dynamics, population genetics, demographic history, and future distributions of gentoo penguin - a robust and resilient species that continues to excel in the face of changing climate conditions. These findings will hopefully contribute to future research of interactions and ecosystem dynamics of the WAP, as well as provide decision support for policy makers with regards to implementing and expanding marine protected areas, managing the commercial krill fishing industry, and Antarctic tourism.

Bibliography

Abadi, F., Barbraud, C., & Gimenez, O. (2017). Integrated population modeling reveals the impact of climate on the survival of juvenile emperor penguins. Global Change Biology, 23(3), 1353–1359.

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

Acosta Alamo, M., Manne, L. L., & Veit, R. R. (2022). Does population size drive changes in transatlantic vagrancy for gulls? A study of seven North Atlantic species. Frontiers in Ecology and Evolution, 10, 275.

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19(9), 1655–1664.

Allio, R., Donega, S., Galtier, N., & Nabholz, B. (2017). Large variation in the ratio of mitochondrial to nuclear mutation rate across animals: Implications for genetic diversity and the use of mitochondrial DNA as a molecular marker. Molecular Biology and Evolution, 34(11), 2762–2772.

Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biology, 23(1), 258.

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. International Journal of Forecasting, 8(1), 69–80.

Aronson, R. B., Smith, K. E., Vos, S. C., McClintock, J. B., Amsler, M. O., Moksnes, P. O., Ellis, D. S., Kaeli, J., Singh, H., Bailey, J. W., Schiferl, J. C., van Woesik, R., Martin, M. A., Steffel, B. V., Deal, M. E., Lazarus, S. M., Havenhand, J. N., Swalethorp, R., Kjellerup, S., & Thatje, S. (2015). No barrier to emergence of bathyal king crabs on the Antarctic shelf. Proceedings of the National Academy of Sciences of the United States of America, 112(42), 12997–13002.

Ashbrook, K., Wanless, S., Harris, M. P., & Hamer, K. C. (2010). Impacts of poor food availability on positive density dependence in a highly colonial seabird. Proceedings. Biological Sciences, 277(1692), 2355–2360.

Barbraud, C., & Delord, K. (2021). Selection against immigrants in wild seabird populations.

Ecology Letters, 24(1), 84–93.

Barbraud, C., Rolland, V., Jenouvrier, S., Nevoux, M., Delord, K., & Weimerskirch, H. (2012). Effects of climate change and fisheries bycatch on Southern Ocean seabirds: A review. Marine Ecology Progress Series, 454, 285–307.

Baylis, A. M., Crofts, S., & Wolfaardt, A. C. (2013). Population trends of gentoo penguins *Pygoscelis papua* breeding at the Falkland Islands. Marine Ornithology, 41, 1–5.

Baylis, A. M., Orben, R. A., Arnould, J. P., Christiansen, F., Hays, G. C., & Staniland, I. J. (2015). Disentangling the cause of a catastrophic population decline in a large marine mammal. Ecology, 96(10), 2834–2847.

Beaumont, M. A. (2010). Approximate Bayesian computation in evolution and ecology. Annual Review of Ecology, Evolution, and Systematics, 41(1), 379–406.

Bergeron, L. A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M. F., Quintard, B., Hoffman, J. I., Li, Z., St Leger, J., Shao, C., Stiller, J., Gilbert, M. T. P., Schierup, M. H., & Zhang, G. (2023). Evolution of the germline mutation rate across vertebrates. Nature, 615(7951), 285–291.

Borowicz, A., McDowall, P., Youngflesh, C., Sayre-McCord, T., Clucas, G., Herman, R., Forrest, S., Rider, M., Schwaller, M., Hart, T., Jenouvrier, S., Polito, M. J., Singh, H., & Lynch, H. J. (2018). Multi-modal survey of Adélie penguin mega-colonies reveals the Danger Islands as a seabird hotspot. Scientific Reports, 8(1), 3926.

Bost, C. A., & Clobert, J. (1992). gentoo penguin *Pygoscelis papua*: Factors affecting the process of laying a replacement clutch. Acta Oecologica, 13(5), 593–605.

Browning, B. L., & Browning, S. R. (2013). Improving the accuracy and efficiency of identity-by-descent detection in population data. Genetics, 194(2), 459–471.

Campioni, L., Granadeiro, J. P., & Catry, P. (2017). Albatrosses prospect before choosing a home: Intrinsic and extrinsic sources of variability in visit rates. Animal Behaviour, 128, 85–93.

Chan, A. H., Jenkins, P. A., & Song, Y. S. (2012). Genome-wide fine-scale recombination rate variation in Drosophila melanogaster. PLOS Genetics, 8(12), e1003090.

Chapron, G., Wikenros, C., Liberg, O., Wabakken, P., Flagstad, Ø., Milleret, C., Månsson, J., Svensson, L., Zimmermann, B., Åkesson, M., & Sand, H. (2016). Estimating wolf (Canis lupus) population size from number of packs and an individual based model. Ecological Modelling, 339, 33–44.

Che-Castaldo, C., Humphries, G., & Lynch, H. (2023). Antarctic penguin Biogeography Project: Database of abundance and distribution for the Adélie, chinstrap, gentoo, emperor, macaroni and king penguin south of 60 S. Biodiversity Data Journal, 11.

Chen, I. C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. Science, 333(6045), 1024–1026.

Cherel, Y., Hobson, K. A., Guinet, C., & Vanpe, C. (2007). Stable isotopes document seasonal changes in trophic niches and winter foraging individual specialization in diving predators from the Southern Ocean. Journal of Animal Ecology, 76(4), 826–836.

Cimino, M. A., Fraser, W. R., Irwin, A. J., & Oliver, M. J. (2013). Satellite data identify decadal trends in the quality of Pygoscelis penguin chick-rearing habitat. Global Change Biology, 19(1), 136–148.

Clements, J. F., Schulenberg, T. S., Iliff, M. J., Roberson, D., Fredericks, T. A., Sullivan, B. L., & Wood, C. L. (2017). The eBird/Clements checklist of birds of the world: v2016. edu/clementschecklist/download. http://www.birds.cornell

Clucas, G. V., Dunn, M. J., Dyke, G., Emslie, S. D., Levy, H., Naveen, R., Polito, M. J., Pybus, O. G., Rogers, A. D., & Hart, T. (2014). A reversal of fortunes: Climate change "winners" and "losers" in Antarctic Peninsula penguins. Scientific Reports, 4(1), 1–7.

Clucas, G. V., Younger, J. L., Kao, D., Emmerson, L., Southwell, C., Wienecke, B., Rogers, A. D., Bost, C. A., Miller, G. D., Polito, M. J., Lelliott, P., Handley, J., Crofts, S., Phillips, R. A., Dunn, M. J., Miller, K. J., & Hart, T. (2018). Comparative population genomics reveals key barriers to dispersal in Southern Ocean penguins. Molecular Ecology, 27(23), 4680–4697.

Cole, T. L., Dutoit, L., Dussex, N., Hart, T., Alexander, A., Younger, J. L., Clucas, G. V., Frugone, M. J., Cherel, Y., Cuthbert, R., Ellenberg, U., Fiddaman, S. R., Hiscock, J., Houston, D., Jouventin, P., Mattern, T., Miller, G., Miskelly, C., Nolan, P., . . . Waters, J. M. (2019). Receding ice drove parallel expansions in Southern Ocean penguins. Proceedings of the National Academy of Sciences of the United States of America, 116(52), 26690–26696.

Coulson, J. C. (2001). Colonial breeding in seabirds. In E. A. Schreiber & J. Burger (Eds.), Biology of marine birds (pp. 87–113). CRC Press.

Crawford, R. J., Dyer, B. M., Upfold, L., & Makhado, A. B. (2014). Congruent, decreasing trends of gentoo penguins and Crozet shags at sub-Antarctic Marion Island suggest food limitation through common environmental forcing. African Journal of Marine Science, 36(2), 225–231.

Crawford, R. J., Whittington, P. A., Upfold, L., Ryan, P. G., Petersen, S. L., Dyer, B. M., & Cooper, J. (2009). Recent trends in numbers of four species of penguins at the Prince Edward Islands. African Journal of Marine Science, 31(3), 419–426.

Cristofari, R., Liu, X., Bonadonna, F., Cherel, Y., Pistorius, P., Le Maho, Y., Raybaud, V., Stenseth, N. C., Le Bohec, C., & Trucchi, E. (2018). Climate-driven range shifts of the king penguin in a fragmented ecosystem. Nature Climate Change, 8(3), 245–251.

Crofts, S., & Stanworth, A. (2019). Falkland Islands Seabird Monitoring Programme – Annual report 2018/2019 (SMP26). Falklands conservation, Stanley.

Croxall, J. P., & Kirkwood, E. D. (1979). The distribution of penguins on the Antarctic Peninsula and islands of the Scotia Sea. British Antarctic Survey.

Csilléry, K., Blum, M. G., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian computation (ABC) in practice. Trends in Ecology and Evolution, 25(7), 410–418.

Davies, B. J., Hambrey, M. J., Smellie, J. L., Carrivick, J. L., & Glasser, N. F. (2012). Antarctic Peninsula ice sheet evolution during the Cenozoic era. Quaternary Science Reviews, 31, 30–66.

de Dinechin, M., Dobson, F. S., Zehtindjiev, P., Metcheva, R., Couchoux, C., Martin, A., Quillfeldt, P., & Jouventin, P. (2012). The biogeography of gentoo penguins (*Pygoscelis papua*). Canadian Journal of Zoology, 90(3), 352–360.

DeAngelis, D. L., & Grimm, V. (2014). Individual-based models in ecology after four decades. F1000prime reports, 6.

Del Valle, R. A., Montalti, D., & Inbar, M. (2002). Mid-Holocene macrofossil-bearing raised marine beaches at Potter Peninsula, King George Island, South Shetland Islands. Antarctic Science, 14(3), 263–269.

Delaneau, O., Zagury, J. F., Robinson, M. R., Marchini, J. L., & Dermitzakis, E. T. (2019). Accurate, scalable and integrative haplotype estimation. Nature Communications, 10(1), 5436.

Dugger, K. M., Ballard, G., Ainley, D. G., & Barton, K. J. (2006). Effects of flipper bands on foraging behavior and survival of Adélie penguins (*Pygoscelis adeliae*). Auk, 123(3), 858–869.

Dunlop, J. N. (2009). The population dynamics of tropical seabirds establishing frontier colonies on islands off south-Western Australia. Marine Ornithology, 37, 99–105.

Dunn, M. J., Forcada, J., Jackson, J. A., Waluda, C. M., Nichol, C., & Trathan, P. N. (2019). A long-term study of gentoo penguin (*Pygoscelis papua*) population trends at a major Antarctic tourist site, Goudier Island, Port Lockroy. Biodiversity and Conservation, 28(1), 37–53.

Emmerson, L., & Southwell, C. (2022). Environment-triggered demographic changes cascade and compound to propel a dramatic decline of an Antarctic seabird metapopulation. Global Change Biology, 28(24), 7234–7249.

Emslie, S. D., Baumann, K., & van Tuinen, M. (2011). Late Holocene occupation of gentoo penguins (*Pygoscelis papua*) on Byers Peninsula, Livingston Island, Antarctica. Polar Biology, 34(2), 283–290.

Fay, R., Weimerskirch, H., Delord, K., & Barbraud, C. (2015). Population density and climate shape early-life survival and recruitment in a long-lived pelagic seabird. Journal of Animal

Ecology, 84(5), 1423–1433.

Foley, C. M., & Lynch, H. J. (2020). A method to estimate pre-exploitation population size. Conservation Biology, 34(1), 256–265.

Foley, C. M., Hart, T., & Lynch, H. J. (2018). King penguin populations increase on South Georgia but explanations remain elusive. Polar Biology, 41(6), 1111–1122.
z

Forbes, L. S., & Kaiser, G. W. (1994). Habitat choice in breeding seabirds: When to cross the information barrier. Oikos, 70(3), 377–384.

Forcada, J., & Trathan, P. N. (2009). Penguin responses to climate change in the Southern Ocean. Global Change Biology, 15(7), 1618–1630.

Forcada, J., Trathan, P. N., Reid, K., Murphy, E. J., & Croxall, J. P. (2006). Contrasting population changes in sympatric penguin species in association with climate warming. Global Change Biology, 12(3), 411–423.

Friesen, V. L., Burg, T. M., & McCoy, K. D. (2007). Mechanisms of population differentiation in seabirds. Molecular Ecology, 16(9), 1765–1785.

Fulton, E. A. (2011). Interesting times: Winners, losers, and system shifts under climate change around Australia. ICES Journal of Marine Science, 68(6), 1329–1342.

Galtier, N., Nabholz, B., Glémin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: A reappraisal. Molecular Ecology, 18(22), 4541–4550.

Germain, S. J., & Lutz, J. A. (2020). Climate extremes may be more important than climate means when predicting species range shifts. Climatic Change, 163(1), 579–598.

Gervais, C. R., Champion, C., & Pecl, G. T. (2021). Species on the move around the Australian coastline: A continental-scale review of climate-driven species redistribution in marine systems. Global Change Biology, 27(14), 3200–3217.

Graham, J. A., Dinniman, M. S., & Klinck, J. M. (2016). Impact of model resolution for on-shelf heat transport along the West Antarctic Peninsula. Journal of Geophysical Research: Oceans, 121(10), 7880–7897.

Grémillet, D., & Boulinier, T. (2009). Spatial ecology and conservation of seabirds facing global climate change: A review. Marine Ecology Progress Series, 391, 121–137.

Grimm, V., & Railsback, S. F. (2006). Agent-based models in ecology: Patterns and alternative theories of adaptive behaviour. Agent-based computational modelling: Applications in demography, social, economic and environmental sciences (pp. 139–152).

Haidvogel, D. B., Arango, H., Budgell, W. P., Cornuelle, B. D., Curchitser, E., Di Lorenzo, E., Fennel, K., Geyer, W. R., Hermann, A. J., Lanerolle, L., Levin, J., McWilliams, J. C., Miller, A. J., Moore, A. M., Powell, T. M., Shchepetkin, A. F., Sherwood, C. R., Signell, R. P., Warner, J. C., & Wilkin, J. (2008). Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the Regional Ocean Modeling System. Journal of Computational Physics, 227(7), 3595–3624.

Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. Annals of Statistics, 13(1), 70–84.

Cushing, J. M., & Henson, S. M. (2018). Periodic matrix models for seasonal dynamics of structured populations with application to a seabird population. Journal of Mathematical Biology, 77(6–7), 1689–1720.

Herman, R. W., Valls, F. C. L., Hart, T., Petry, M. V., Trivelpiece, W. Z., & Polito, M. J. (2017). Seasonal consistency and individual variation in foraging strategies differ between and within Pygoscelis penguin populations in the Antarctic Peninsula region. Marine Biology, 164(5), 115.

Herman, R., Borowicz, A., Lynch, M., Trathan, P., Hart, T., & Lynch, H. (2020). Update on the global abundance and distribution of breeding gentoo penguins (*Pygoscelis papua*). Polar Biology, 43(12), 1947–1956.

Herman, R. W., & Lynch, H. J. (2022). Age-structured model reveals prolonged immigration is key for colony establishment in gentoo penguins. Ornithological Applications, 124(3), duac014.

Herman, R. W., Winger, B. M., Dittmann, D. L., & Harvey, M. G. (2022). Fine-scale population genetic structure and barriers to gene flow in a widespread seabird (*Ardenna pacifica*). Biological Journal of the Linnean Society, 137(1), 125–136.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. Nature, 405(6789), 907–913.

Hickling, R., Roy, D. B., Hill, J. K., Fox, R., & Thomas, C. D. (2006). The distributions of a wide range of taxonomic groups are expanding polewards. Global Change Biology, 12(3), 450–455.

Hindell, M. A., Reisinger, R. R., Ropert-Coudert, Y., Hückstädt, L. A., Trathan, P. N., Bornemann, H., Charrassin, J. B., Chown, S. L., Costa, D. P., Danis, B., Lea, M. A., Thompson, D., Torres, L. G., Van de Putte, A. P., Alderman, R., Andrews-Goff, V., Arthur, B., Ballard, G., Bengtson, J., . . . & Raymond, B. (2020). Tracking of marine predators to protect Southern Ocean ecosystems. Nature, 580(7801), 87–92.

Hinke, J. T., Salwicka, K., Trivelpiece, S. G., Watters, G. M., & Trivelpiece, W. Z. (2007). Divergent responses of Pygoscelis penguins reveal a common environmental driver. Oecologia, 153(4), 845–855.

Hudson Gallagher, K., Lynch, H., Dinniman, M. S., Youngflesh, C., & Che-Castaldo, C. (2022,

December). Pygoscelis penguin colony locations and diet compositions along the west Antarctic Peninsula could be driven by high retention and accumulation of simulated krill. In AGU Fall Meeting Abstracts, 2022.

Hudson, K. L., Oliver, M. J., Kohut, J., Cohen, J. H., Dinniman, M. S., Klinck, J., Statscewich, H., Bernard, K. S., & Fraser, W. (2022). Subsurface eddy facilitates retention of diel vertical migrators in a biological hotspot. Authorea Preprints.

Humphries, G. R. W., Naveen, R., Schwaller, M., Che-Castaldo, C., McDowall, P., Schrimpf, M., & Lynch, H. J. (2017). Mapping application for penguin populations and projected dynamics (MAPPPD): Data and tools for dynamic management and decision support. Polar Record, 53(2), 160–166.

Ingólfsson, Ó., Hjort, C., & Humlum, O. (2003). Glacial and climate history of the Antarctic Peninsula since the Last Glacial Maximum. Arctic, Antarctic, and Alpine Research, 35(2), 175–186.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLOS ONE, 9(6), e98679.

Jouventin, P. (1994). Les populations d'oiseaux marins des T.A.A.F.: Résumé de 20 années de recherché. Aluada, 62, 44–47.

Kersten, O., Star, B., Leigh, D. M., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps, S., Erikstad, K. E., Fitzsimmons, M. G., Fort, J., Hansen, E. S., Harris, M. P., Irestedt, M., Kleven, K., S. D., Irons, D. B., Nysewander, D. R., & Buck, C. L. (2005). Formation and growth of new seabird colonies: The significance of habitat quality. Marine Ornithology, 33, 49–58.

Kimura, M., & Weiss, G. H. (1964). The stepping stone model of population structure and the decrease of genetic correlation with distance. Genetics, 49(4), 561–576. https://doi.org/10.1093/genetics/49.4.561

Korczak-Abshire, M., Hinke, J. T., Milinevsky, G., Juáres, M. A., & Watters, G. M. (2021). Coastal regions of the northern Antarctic Peninsula are key for gentoo populations. Biology Letters, 17(1), 20200708.

LaRue, M. A., Ainley, D. G., Swanson, M., Dugger, K. M., Lyver, P. O'B., Barton, K., & Ballard, G. (2013). Climate change winners: Receding ice fields facilitate colony expansion and altered dynamics in an Adélie penguin metapopulation. PLOS ONE, 8(4), e60568.

Lawson, D. J., Hellenthal, G., Myers, S., & Falush, D. (2012). Inference of population structure using dense haplotype data. PLOS Genetics, 8(1), e1002453.

Lesage, L., Crête, M., Huot, J., Dumont, A., & Ouellet, J. P. (2000). Seasonal home range size and philopatry in two northern white-tailed deer populations. Canadian Journal of Zoology,

78(11), 1930–1940.

Lescroël, A., & Bost, C. A. (2005). Foraging under contrasting oceanographic conditions: The gentoo penguin at Kerguelen Archipelago. Marine Ecology Progress Series, 302, 245–261.

Lescroël, A., & Bost, C. A. (2006). Recent decrease in gentoo penguin populations at Iles Kerguelen. Antarctic Science, 18(2), 171–174.

Levy, H., Clucas, G. V., Rogers, A. D., Leaché, A. D., Ciborowski, K. L., Polito, M. J., Lynch, H. J., Dunn, M. J., & Hart, T. (2016). Population structure and phylogeography of the gentoo penguin (Pygoscelis papua) across the Scotia Arc. Ecology and Evolution, 6(6), 1834–1853.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics, 25(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. Bioinformatics, 25(16), 2078–2079.

Lin, Y., Moreno, C., Marchetti, A., Ducklow, H., Schofield, O., Delage, E., Meredith, M., Li, Z., Eveillard, D., Chaffron, S., & Cassar, N. (2021). Decline in plankton diversity and carbon flux with reduced sea ice extent along the western Antarctic Peninsula. Nature Communications, 12(1), 4948.

Lynch, H. J. (2013). gentoo penguins (*Pygoscelis papua*). In P. G. Borboroglu & P. D. Boersma (Eds.), Penguins: Natural history and conservation (1st ed) (pp. 75–88). University of Washington Press.

Lynch, H. J., Fagan, W. F., Naveen, R., Trivelpiece, S. G., & Trivelpiece, W. Z. (2012). Differential advancement of breeding phenology in response to climate may alter staggered breeding among sympatric pygoscelid penguins. Marine Ecology Progress Series, 454, 135–145.

Lynch, H. J., Naveen, R., & Casanovas, P. V. (2013). Antarctic Site Inventory breeding bird survey data. Ecology 94:2653, 2012/13, 1994/95.

Lynch, H. J., Naveen, R., Trathan, P. N., & Fagan, W. F. (2012). Spatially integrated assessment reveals widespread changes in penguin populations on the Antarctic Peninsula. Ecology, 93(6), 1367–1377.

Lynch, H. J., White, R., Naveen, R., Black, A., Meixler, M. S., & Fagan, W. F. (2016). In stark contrast to widespread declines along the Scotia Arc, a survey of the South Sandwich Islands finds a robust seabird community. Polar Biology, 39(9), 1615–1625.

Lynch, H. J., Fagan, W. F., & Naveen, R. (2010). Population trends and reproductive success at a frequently visited penguin colony on the western Antarctic Peninsula. Polar Biology, 33(4), 493–503.

Marschner, I. C. (2010). Stable computation of maximum likelihood estimates in identity link Poisson regression. Journal of Computational and Graphical Statistics, 19(3), 666–683.

Masello, J. F., Kato, A., Sommerfeld, J., Mattern, T., & Quillfeldt, P. (2017). How animals distribute themselves in space: Variable energy landscapes. Frontiers in Zoology, 14(1), 33.

Matthiopoulos, J., Harwood, J., & Thomas, L. E. N. (2005). Metapopulation consequences of site fidelity for colonially breeding mammals and birds. Journal of Animal Ecology, 74(4), 716–727.

McDowall, P. S., & Lynch, H. J. (2019). When the "selfish herd" becomes the "frozen herd": Spatial dynamics and population persistence in a colonial seabird. Ecology, 100(10), e02823.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–1303.
McMahon, K. W., Michelson, C. I., Hart, T., McCarthy, M. D., Patterson, W. P., & Polito, M. J. (2019). Divergent trophic responses of sympatric penguin species to historic anthropogenic exploitation and recent climate change. Proceedings of the National Academy of Sciences of the United States of America, 116(51), 25721–25727.

Meredith, M. P., Stammerjohn, S. E., Venables, H. J., Ducklow, H. W., Martinson, D. G., Iannuzzi, R. A., Leng, M. J., van Wessem, J. M., Reijmer, C. H., & Barrand, N. E. (2017). Changing distributions of sea ice melt and meteoric water west of the Antarctic Peninsula. Deep Sea Research Part II, 139, 40–57.

Millsap, B. A. (2018). Demography and metapopulation dynamics of an urban Cooper's Hawk subpopulation. Condor, 120(1), 63–80.

Minter, A., & Retkute, R. (2019). Approximate Bayesian Computation for infectious disease modelling. Epidemics, 29, 100368.

Mott, P. G., & Wiggins, W. D. C. (1965). Falkland Islands and dependencies aerial survey expedition 1955–57. Geographical Journal, 131(3), 430–432.

Naveen, R., Lynch, H. J., Forrest, S., Mueller, T., & Polito, M. (2012). First direct, site-wide penguin survey at Deception Island, Antarctica, suggests significant declines in breeding chinstrap penguins. Polar Biology, 35, 1879–1888.

Kersten, O., Star, B., Leigh, D. M., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps, S., Erikstad, K. E., Fitzsimmons, M. G., Fort, J., Hansen, E. S., Harris, M. P., Irestedt, M., Kleven, O., Mallory, M. L., Jakobsen, K. S., & Boessenkool, S. (2021). Complex population structure of the Atlantic puffin revealed by whole genome analyses. Communications Biology, 4(1), 922.

Oro, D., & Ruxton, G. D. (2001). The formation and growth of seabird colonies: Audouin's gull as a case study. Journal of Animal Ecology, 70(3), 527–535.

Pan, H., Cole, T. L., Bi, X., Fang, M., Zhou, C., Yang, Z., Ksepka, D. T., Hart, T., Bouzat, J. L., Argilla, L. S., Bertelsen, M. F., Boersma, P. D., Bost, C. A., Cherel, Y., Dann, P., Fiddaman, S. R., Howard, P., Labuschagne, K., Mattern, T., . . . & Zhang, G. (2019). High-coverage genomes to elucidate the evolution of penguins. GigaScience, 8(9), giz117.

Paolo, F. S., Fricker, H. A., & Padman, L. (2015). Ice sheets. Volume loss from Antarctic ice shelves is accelerating. Science, 348(6232), 327–331.

Parks, & Wildlife Service. (2006). Macquarie Island nature reserve and world heritage area management plan. Parks and Wildlife Service, Department of tourism, arts, and the environment. Hobart.

Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. PLOS Genetics, 2(12), e190.

Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I. C., Clark, T. D., Colwell, R. K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R. A., Griffis, R. B., Hobday, A. J., Janion-Scheepers, C., Jarzyna, M. A., Jennings, S., . . . Williams, S. E. (2017). Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. Science, 355(6332), eaai9214.

Peña M, F., Poulin, E., Dantas, G. P., González-Acuña, D., Petry, M. V., & Vianna, J. A. (2014). Have historical climate changes affected gentoo penguin (*Pygoscelis papua*) populations in Antarctica? PLOS ONE, 9(4), e95375.

Péron, C., Weimerskirch, H., & Bost, C. A. (2012). Projected poleward shift of king penguins' (*Aptenodytes patagonicus*) foraging range at the Crozet Islands, southern Indian Ocean. Proceedings. Biological Sciences, 279(1738), 2515–2523.

Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. Science, 308(5730), 1912–1915.

Pertierra, L. R., Segovia, N. I., Noll, D., Martinez, P. A., Pliscoff, P., Barbosa, A., Aragón, P., Raya Rey, A., Pistorius, P., Trathan, P., Polanowski, A., Bonadonna, F., Le Bohec, C., Bi, K., Wang-Claypool, C. Y., González-Acuña, D., Dantas, G. P. M., Bowie, R. C. K., Poulin, E., & Vianna, J. A. (2020). Cryptic speciation in gentoo penguins is driven by geographic isolation and regional marine conditions: Unforeseen vulnerabilities to global change. Diversity and Distributions, 26(8), 958–975.

Petkova, D., Novembre, J., & Stephens, M. (2016). Visualizing spatial population structure with estimated effective migration surfaces. Nature Genetics, 48(1), 94–100.

Petry, M. V., Valls, F. C. L., Petersen, E. S., Finger, J. V. G., & Krüger, L. (2018). Population trends of seabirds at Stinker Point, Elephant Island, Maritime Antarctica. Antarctic Science, 30(4), 220–226.

Petry, M. V., Basler, A. B., Valls, F. C. L., & Krüger, L. (2013). New southerly breeding location of king penguins (*Aptenodytes patagonicus*) on Elephant Island (Maritime Antarctic). Polar Biology, 36(4), 603–606.

Phillips, R. A., Lewis, S., González-Solís, J., & Daunt, F. (2017). Causes and consequences of individual variability and specialization in foraging and migration strategies of seabirds. Marine Ecology Progress Series, 578, 117–150.

Picard Tools, & Broad Institute. http://broadinstitute.github.io/picard/

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLOS Genetics, 8(11), e1002967.

Pinsky, M. L., Eikeset, A. M., McCauley, D. J., Payne, J. L., & Sunday, J. M. (2019). Greater vulnerability to warming of marine versus terrestrial ectotherms. Nature, 569(7754), 108–111.

Pistorius, P. A., Baylis, A., Crofts, S., & Pütz, K. (2012). Population development and historical occurrence of king penguins at the Falkland Islands. Antarctic Science, 24(5), 435–440.

Polar Geospatial Center. (2023). High-Resolution Imagery Service. Retrieved from https://www.pgc.umn.edu/data/.

Polito, M. J., Trivelpiece, W. Z., Patterson, W. P., Karnovsky, N. J., Reiss, C. S., & Emslie, S. D. (2015). Contrasting specialist and generalist patterns facilitate foraging niche partitioning in sympatric populations of Pygoscelis penguins. Marine Ecology Progress Series, 519, 221–237.

Poloczanska, E. S., Burrows, M. T., Brown, C. J., García Molinos, J., Halpern, B. S., Hoegh-Guldberg, O., Kappel, C. V., Moore, P. J., Richardson, A. J., Schoeman, D. S., & Sydeman, W. J. (2016). Responses of marine organisms to climate change across oceans. Frontiers in Marine Science, 3, 62.

Ponchon, A., Iliszko, L., Grémillet, D., Tveraa, T., & Boulinier, T. (2017). Intense prospecting movements of failed breeders nesting in an unsuccessful breeding subcolony. Animal Behaviour, 124, 183–191.

Pozzi, L. M., García Borboroglu, P. G., Boersma, P. D., & Pascual, M. A. (2015). Population regulation in Magellanic penguins: What determines changes in colony size? PLOS ONE, 10(3), e0119002.

Pritchard, H. D., Ligtenberg, S. R. M., Fricker, H. A., Vaughan, D. G., Van den Broeke, M. R., & Padman, L. (2012). Antarctic ice-sheet loss driven by basal melting of ice shelves. Nature, 484(7395), 502–505.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575.

Pyk, T. M., Weston, M. A., Bunce, A., & Norman, F. I. (2013). Establishment and development of a seabird colony: Long-term trends in phenology, breeding success, recruitment, breeding density and demography. Journal of Ornithology, 154(1), 299–310.

QGIS Development Team. (2018). QGIS Geographic Information System. Open-Source Geospatial Foundation Project. http://qgis.osgeo.org

R Development Core Team. (2017). R: A Language and Environment for Statistical Computing. http://www.r-project.org/. R Foundation for Statistical Computing.

Railsback, S. F., & Grimm, V. (2019). Agent-based and individual-based modeling: A practical introduction. Princeton University Press.

Raya Rey, A., Rosciano, N., Liljesthröm, M., Sáenz Samaniego, R., & Schiavini, A. (2014). Species-specific population trends detected for penguins, gulls and cormorants over 20 years in sub-Antarctic Fuegian Archipelago. Polar Biology, 37(9), 1343–1360.

Robert, A., Paiva, V. H., Bolton, M., Jiguet, F., & Bried, J. (2014). Nest fidelity is driven by multi-scale information in a long-lived seabird. Proceedings. Biological Sciences, 281(1793), 20141692.

Sandvik, H., Erikstad, K. E., Barrett, R. T., & Yoccoz, N. G. (2005). The effect of climate on adult survival in five species of North Atlantic seabirds. Journal of Animal Ecology, 74(5), 817–831.

Santora, J. A., & Veit, R. R. (2013). Spatio-temporal persistence of top predator hotspots near the Antarctic Peninsula. Marine Ecology Progress Series, 487, 287–304.

Santoro, S., Green, A. J., & Figuerola, J. (2016). Immigration enhances fast growth of a newly established source population. Ecology, 97(4), 1048–1057.

Schaub, M., & Abadi, F. (2011). Integrated population models: A novel analysis framework for deeper insights into population dynamics. Journal of Ornithology, 152(S1), 227–237.

Schippers, P., Stienen, E. W. M., Schotman, A. G. M., Snep, R. P. H., & Slim, P. A. (2011). The consequences of being colonial: Allee effects in metapopulations of seabirds. Ecological Modelling, 222(17), 3061–3070.

Schmidt, A. E., Ballard, G., Lescroël, A., Dugger, K. M., Jongsomjit, D., Elrod, M. L., & Ainley, D. G. (2021). The influence of subcolony-scale nesting habitat on the reproductive success of Adélie penguins. Scientific Reports, 11(1), 15380.

Schrimpf, M., & Lynch, H. (2021). The role of wind fetch in structuring Antarctic seabird breeding occupancy. Ibis, 163(2), 695–705. https://doi.org/10.1111/ibi.12910

Scranton, K., Knape, J., & de Valpine, P. (2014). An approximate Bayesian computation approach to parameter estimation in a stochastic stage-structured population model. Ecology, 95(5), 1418–1428.

Secor, D. H., Kerr, L. A., & Cadrin, S. X. (2009). Connectivity effects on productivity, stability, and persistence in a herring metapopulation model. ICES Journal of Marine Science, 66(8), 1726–1732.

Sekercioglu, C. H., Schneider, S. H., Fay, J. P., & Loarie, S. R. (2008). Climate change, elevational range shifts, and bird extinctions. Conservation Biology, 22(1), 140–150.

Shepherd, A., Fricker, H. A., & Farrell, S. L. (2018). Trends and connections across the Antarctic cryosphere. Nature, 558(7709), 223–232.

Sherley, R. B., Abadi, F., Ludynia, K., Barham, B. J., Clark, A. E., & Altwegg, R. (2014). Age-specific survival and movement among major African penguin *Spheniscus demersus* colonies. Ibis, 156(4), 716–728.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. Evolution; International Journal of Organic Evolution, 47(1), 264–279.

Smagol, V. M., Dzhulay, A. O., Dykyy, I. V., Milinevsky, G. P., & Dykyy, E. O. (2018). Pygoscelis penguin census in the Vernadsky Antarctic station area. Working, paper SC-CAMLR-XXXVII/BG/06 for the Commission for the Conservation of Antarctic Marine Living Resources.

Smith, C. R., Grange, L. J., Honig, D. L., Naudts, L., Huber, B., Guidi, L., & Domack, E. (2012). A large population of king crabs in Palmer Deep on the west Antarctic Peninsula shelf and potential invasive impacts. Proceedings. Biological Sciences, 279(1730), 1017–1026.

Somero, G. N. (2010). The physiology of climate change: How potentials for acclimatization and genetic adaptation will determine "winners" and "losers". Journal of Experimental Biology, 213(6), 912–920.

Speidel, L., Forest, M., Shi, S., & Myers, S. R. (2019). A method for genome-wide genealogy estimation for thousands of samples. Nature Genetics, 51(9), 1321–1329.

Stammerjohn, S. E., Martinson, D. G., Smith, R. C., & Iannuzzi, R. A. (2008). Sea ice in the western Antarctic Peninsula region: Spatio-temporal variability from ecological and climate change perspectives. Deep Sea Research Part II, 55(18–19), 2041–2058.

Sweanor, L. L., Logan, K. A., & Hornocker, M. G. (2000). Cougar dispersal patterns, metapopulation dynamics, and conservation. Conservation Biology, 14(3), 798–808.

Szostek, K. L., Schaub, M., & Becker, P. H. (2014). Immigrants are attracted by local pre-breeders and recruits in a seabird colony. Journal of Animal Ecology, 83(5), 1015–1024.

Thomas, C. D., Franco, A. M., & Hill, J. K. (2006). Range retractions and extinction in the face of climate warming. Trends in Ecology and Evolution, 21(8), 415–416.

Trathan, P. N., Daunt, F. J. H., & Murphy, E. J. (1996). South Georgia: An ecological atlas. Br Antarct Surv.

Trathan, P. N., Grant, S. M., Siegel, V., & Kock, K. H. (2013). Precautionary spatial protection to facilitate the scientific study of habitats and communities under ice shelves in the context of recent, rapid, regional climate change. CCAMLR Science, 20, 139–151.

Trivelpiece, W. Z., Hinke, J. T., Miller, A. K., Reiss, C. S., Trivelpiece, S. G., & Watters, G. M. (2011). Variability in krill biomass links harvesting and climate warming to penguin population changes in Antarctica. Proceedings of the National Academy of Sciences of the United States of America, 108(18), 7625–7628.

Turner, J., Phillips, T., Marshall, G. J., Hosking, J. S., Pope, J. O., Bracegirdle, T. J., & Deb, P. (2017). Unprecedented springtime retreat of Antarctic sea ice in 2016. Geophysical Research Letters, 44(13), 6868–6875.

Turner, J., Lachlan-Cope, T., Colwell, S., & Marshall, G. J. (2005). A positive trend in western Antarctic Peninsula precipitation over the last 50 years reflecting regional and Antarctic-wide atmospheric circulation changes. Annals of Glaciology, 41, 85–91.

Turner, J., Maksym, T., Phillips, T., Marshall, G. J., & Meredith, M. P. (2013). The impact of changes in sea ice advance on the large winter warming on the western Antarctic Peninsula. International Journal of Climatology, 33(4), 852–861.

Van der Auwera, G. A., & O'Connor, B. D. (2020). Genomics in the cloud: Using Docker, GATK, and WDL in Terra. O'Reilly Media.

Van der Putten, W. H. (2012). Climate change, aboveground-belowground interactions, and species' range shifts. Annual Review of Ecology, Evolution, and Systematics, 43(1), 365–383.

van der Vaart, E., Beaumont, M. A., Johnston, A. S. A., & Sibly, R. M. (2015). Calibration and evaluation of individual-based models using Approximate Bayesian Computation. Ecological Modelling, 312, 182–190.

Vaughan, D. G., Marshall, G. J., Connolley, W. M., Parkinson, C., Mulvaney, R., Hodgson, D. A., King, J. C., Pudsey, C. J., & Turner, J. (2003). Recent rapid regional climate warming on the Antarctic Peninsula. Climatic Change, 60(3), 243–274.

Veit, R. R. (1990). Do vagrant birds in Massachusetts reflect population growth and dispersal rather than weather patterns? Bird Observations, 18, 86–91.

Veit, R. R. (1997). Long-distance dispersal and population growth of the Yellow-headed Blackbird *Xanthocephalus xanthocephalus*. Ardea, 85, 135–144.

Veit, R. R. (2000). Vagrants as the expanding fringe of a growing population. Auk, 117(1), 242–246.

Veit, R. R., Zawadzki, L. C., Manne, L. L., Cales, P., Fibikar, D., Curley, S., Dluhos, E., & Norton, R. L. (2016). Vagrancy and colonization of St. Thomas and St. John, US Virgin Islands, by Adelaide's Warblers (*Setophaga adelaidae*). Journal of Caribbean Ornithology, 29, 47–50.

Vianna, J. A., Fernandes, F. A. N., Frugone, M. J., Figueiró, H. V., Pertierra, L. R., Noll, D., Bi, K., Wang-Claypool, C. Y., Lowther, A., Parker, P., Le Bohec, C., Bonadonna, F., Wienecke, B., Pistorius, P., Steinfurth, A., Burridge, C. P., Dantas, G. P. M., Poulin, E., . . . Bowie, R. C. K. (2020). Genome-wide analyses reveal drivers of penguin diversification. Proceedings of the National Academy of Sciences of the United States of America, 117(36), 22303–22310.

Vianna, J. A., Noll, D., Dantas, G. P. M., Petry, M. V., Barbosa, A., González-Acuña, D., Le Bohec, C., Bonadonna, F., & Poulin, E. (2017). Marked phylogeographic structure of gentoo penguin reveals an ongoing diversification process along the Southern Ocean. Molecular Phylogenetics and Evolution, 107, 486–498.

Wawrzynek-Borejko, J., Panasiuk, A., Hinke, J. T., & Korczak-Abshire, M. (2022). Are the diets of sympatric Pygoscelid penguins more similar than previously thought? Polar Biology, 45(10), 1559–1569.

Weimerskirch, H., Inchausti, P., Guinet, C., & Barbraud, C. (2003). Trends in bird and seal populations as indicators of a system shift in the Southern Ocean. Antarctic Science, 15(2), 249–256.

Weimerskirch, H., Zotier, R., & Jouventin, P. (1989). The avifauna of the Kerguelen Islands. Emu – Austral Ornithology, 89(1), 15–29.

Williams, T. D., & Rodwell, S. (1992). Annual variation in return rate, mate and nest-site fidelity in breeding gentoo and macaroni penguins. Condor, 94(3), 636–645.

Williams, T. D., & Busby, J. (1995). Bird families of the world. The penguins: Spheniscidae (2nd ed). Oxford University Press.

Woehler, E. J. (1993). The distribution and abundance of Antarctic and sub Antarctic penguins. Scientific Publishing Commission on Antarctic Research. 2017.

Younger, J., Emmerson, L., Southwell, C., Lelliott, P., & Miller, K. (2015). Proliferation of East Antarctic Adélie penguins in response to historical deglaciation. BMC Evolutionary Biology, 15, 236.

APPENDIX

Appendix 1.1 See published electronic supplementary material at:
https://link.springer.com/article/10.1007/s00300-020-02759-3

Appendices 2.1-2.5 See published electronic supplementary material at:
https://doi.org/10.1093/ornithapp/duac014

Appendix 3.1 – ADMIXTURE results for k=2 through k=6.

K=2

K=3

K=4

K=5

K=6

Colony
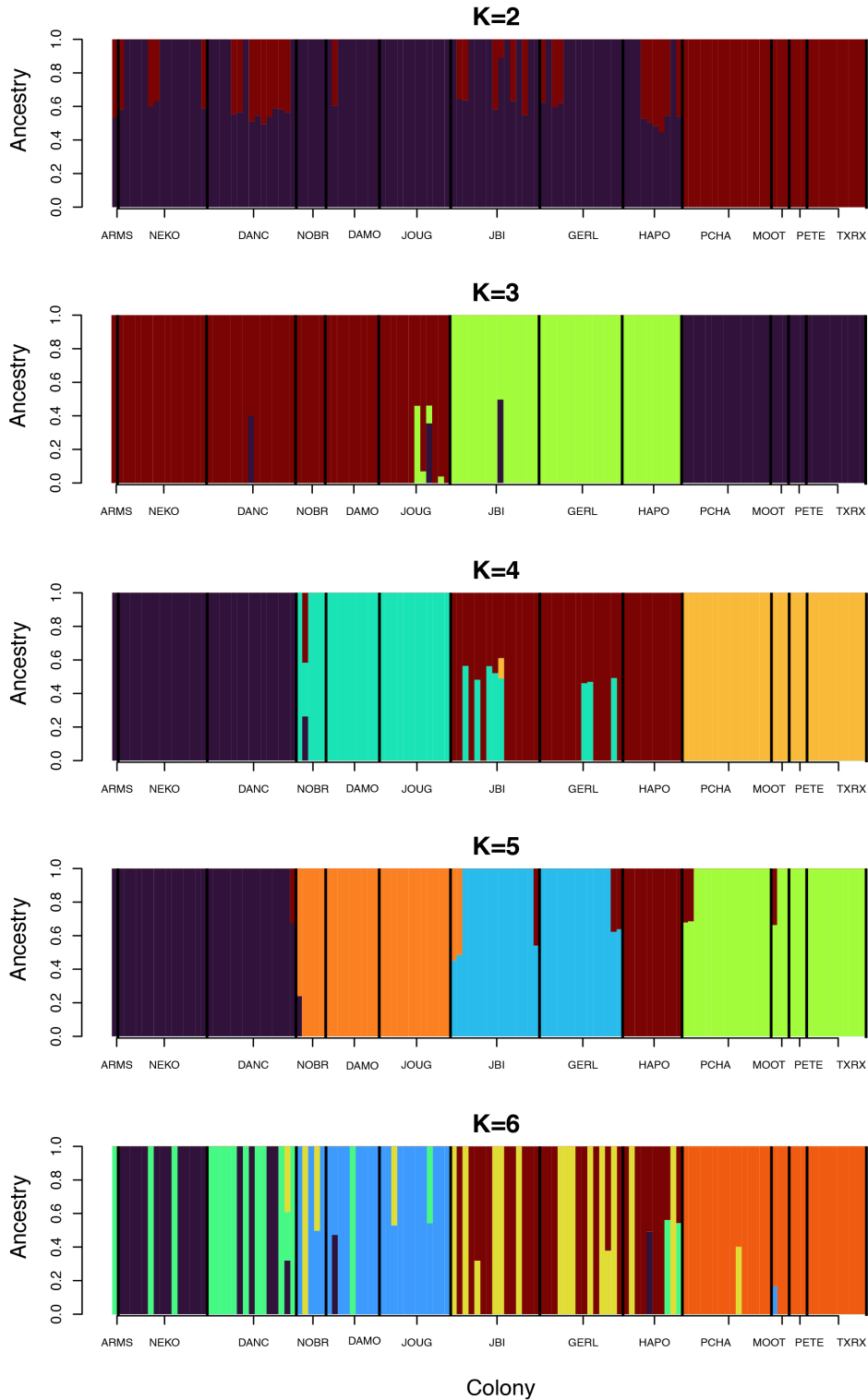
Appendix 3.2 – Pairwise comparisons of residual fits from Treemix output with no migration edges.
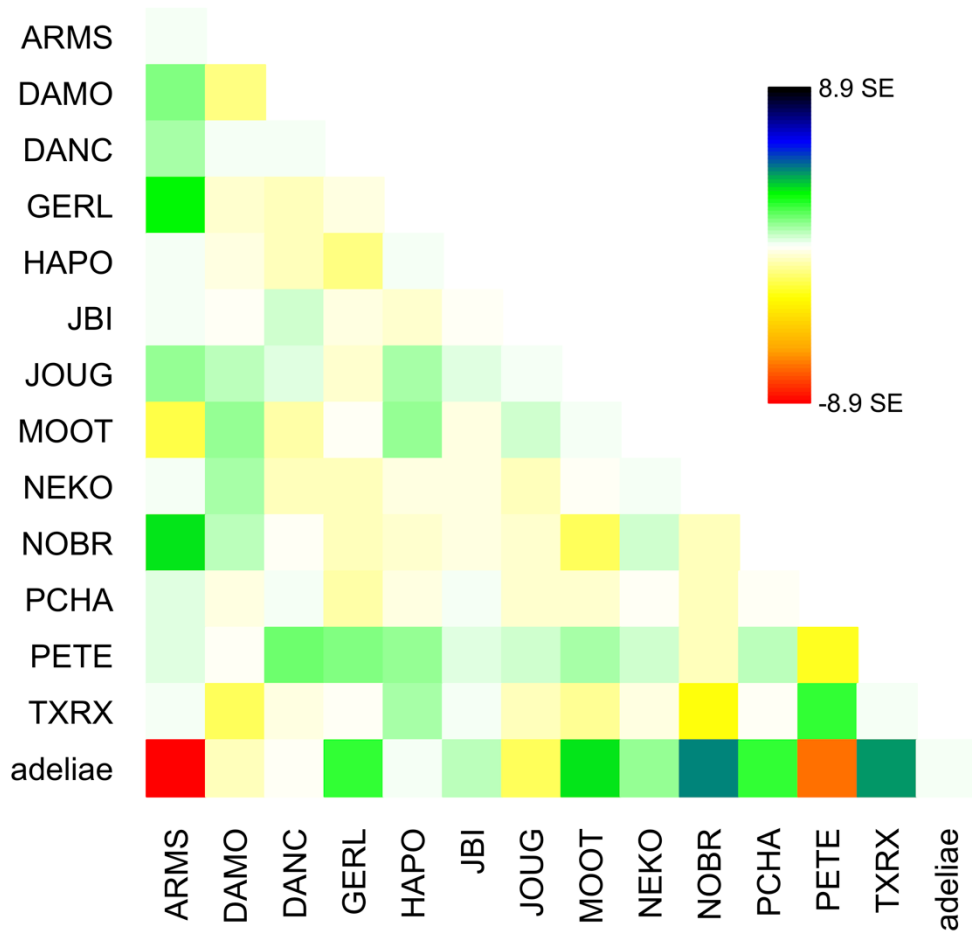
Appendix 3.3 – FineSTRUCTURE unrooted dendrogram. DAMO = Damoy Point, DANC = Danco Island, NEKO = Neko Harbor, JOUG = Jougla Point, HAPO = Hannah Point, GERL = Gerlache Island, JBI = Joubin Islands, TXRX = Tuxen Rocks, MOOT = Moot Point, PETE = Petermann Island, PCHA = Port Charcot. ARMS = Armstrong Reef.
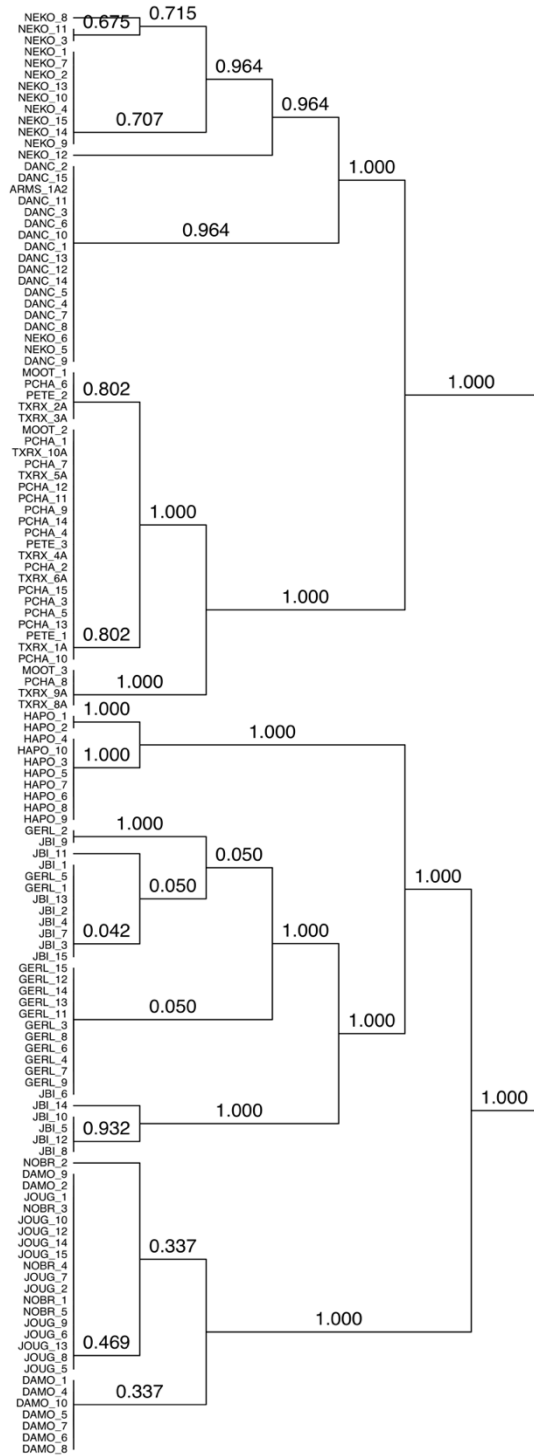
Appendix 3.4 – Allele frequency spectra for each population.
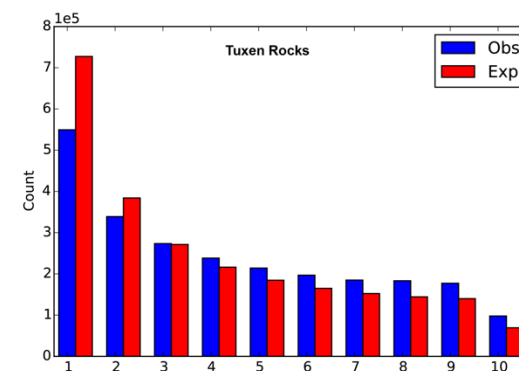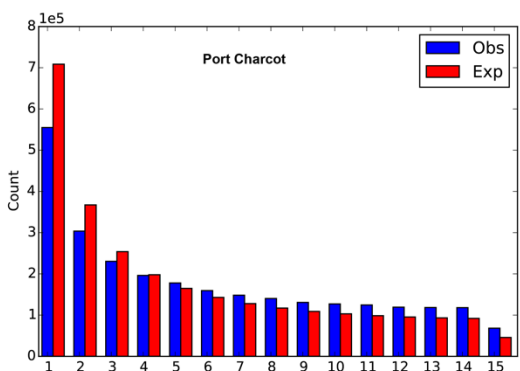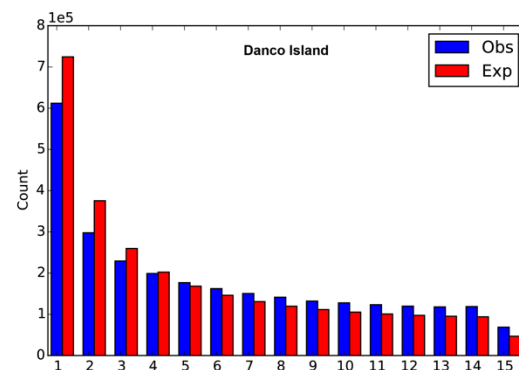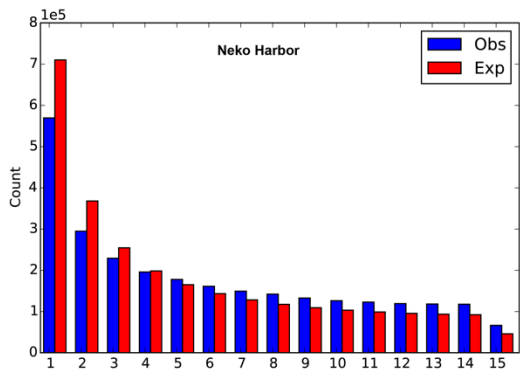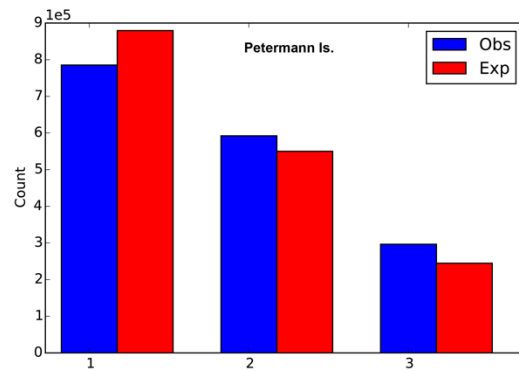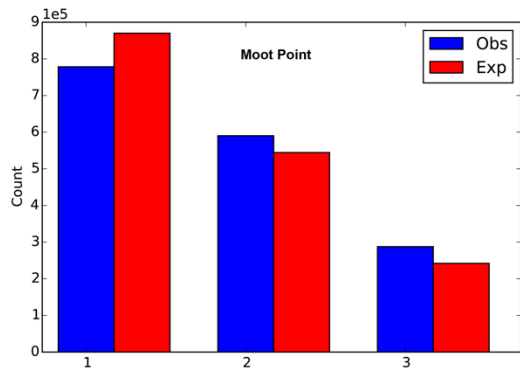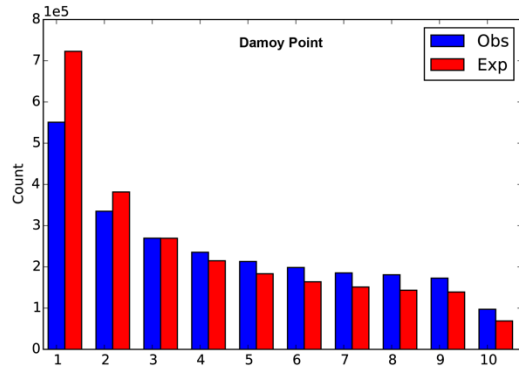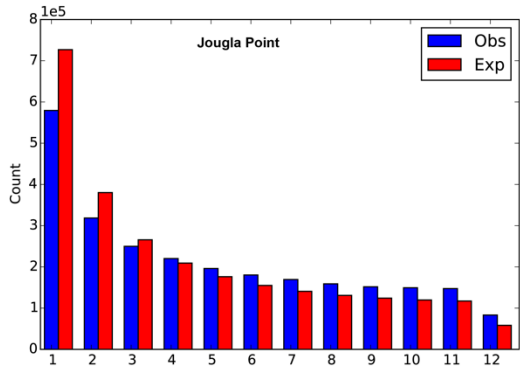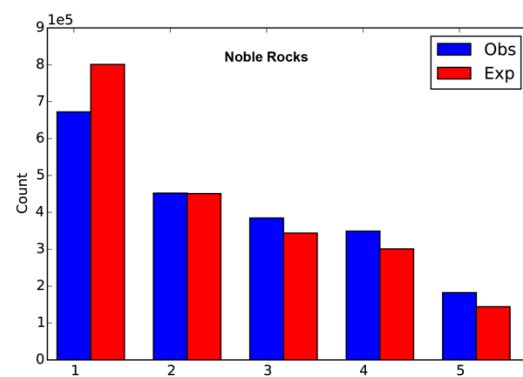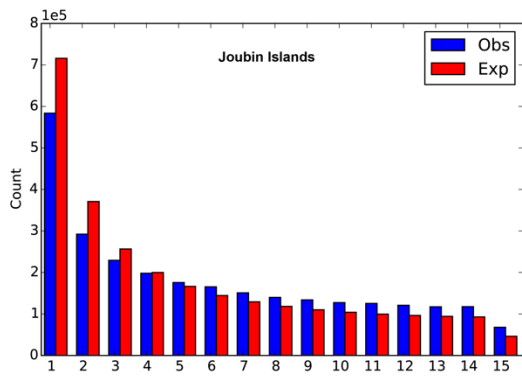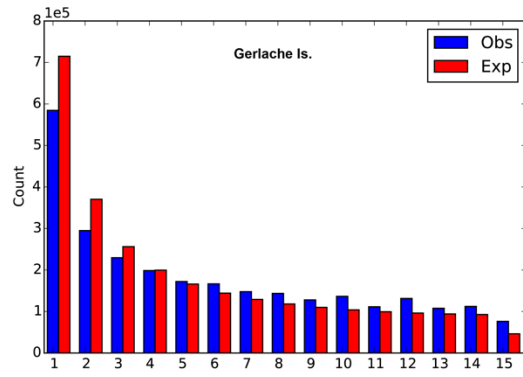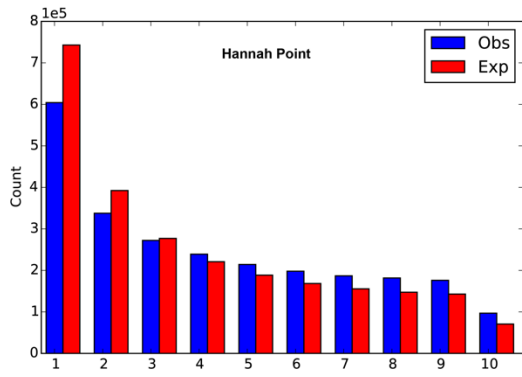
Appendix 3.5 – Table of colonies sampled with respective Tajima's D estimates, years of first observation, and current breeding population size. Year of first observation for Noble Rocks, Moot Point, and Tuxen Rocks are years of colony establishment.
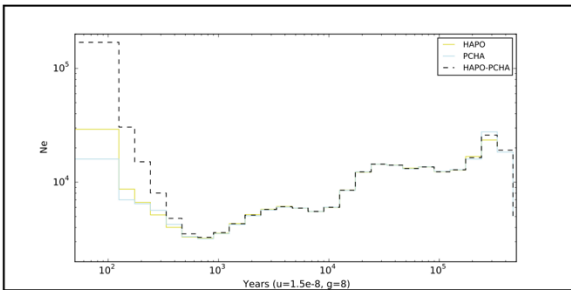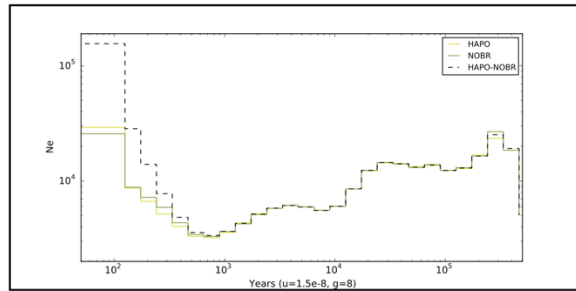
Region	Colony	Tajima's D	Year of first observation	Current breeding population size
South Shetland Islands	Hannah Point	0.401	1958	2820 (2019)
Northern cluster WAP	Danco Island	0.433	1964	732 (2021)
Northern cluster WAP	Neko Harbor	0.480	1962	603 (2021)
East Anvers WAP	Gerlache Island	0.463	< 1904	1005 (2019)
East Anvers WAP	Joubin Islands	0.473	< 1904	615 (2021)
East Anvers WAP	Jougla Point	0.449	< 1914	973 (2021)
West Anvers WAP	Damoy Point	0.465	< 1954	2675 (2021)
West Anvers WAP	Noble Rocks	0.285	2010	34 (2019)
Southern cluster WAP	Moot Point	0.193	2005	629 (2023)
Southern cluster WAP	Port Charcot	0.492	1907	1556 (2022)
Southern cluster WAP	Petermann Is.	0.199	1914	3498 (2019)
Southern cluster WAP	Tuxen Rocks	0.470	2013	30 (2019)

Appendix 3.6 – Generalized skyline plots for pairwise separation histories between populations assuming a per generation mutation rate of 1.5x10^{-8} and a generation time of eight years. DAMO = Damoy Point, DANC = Danco Island, NEKO = Neko Harbor, JOUG = Jougla Point, HAPO = Hannah Point, GERL = Gerlache Island, JBI = Joubin Islands, TXRX = Tuxen Rocks, MOOT = Moot Point, PETE = Petermann Island, PCHA = Port Charcot.

Ne

Years (u=1.5e-8, g=8)

JBI
NOBR
JBI-NOBR

JBI
PCHA
JBI-PCHA

JBI
TXRX
JBI-TXRX

JOUG
NEKO
JOUG-NEKO

JOUG
NOBR
JOUG-NOBR

JOUG
PCHA
JOUG-PCHA

JOUG
TXRX
JOUG-TXRX

NEKO
NOBR
NEKO-NOBR
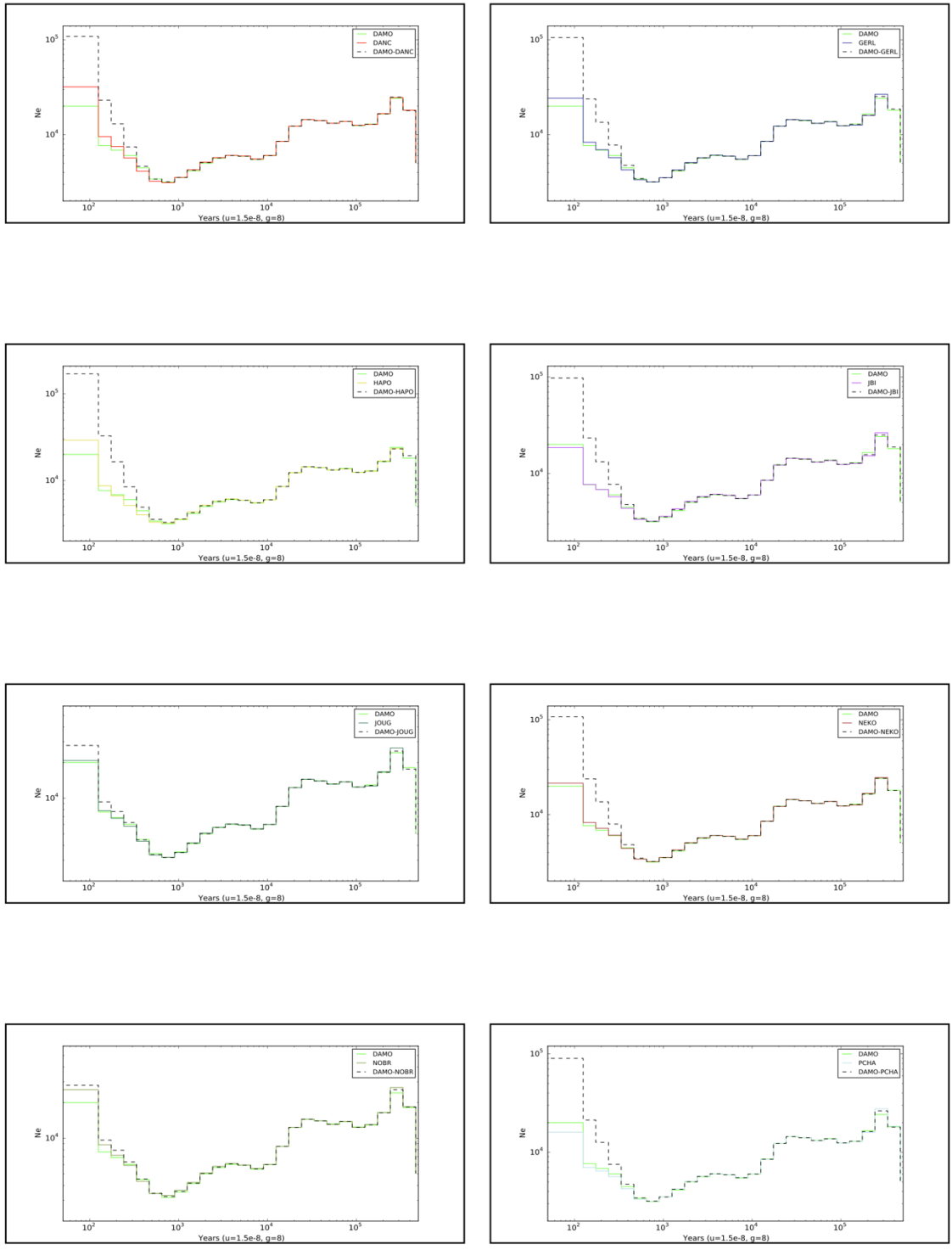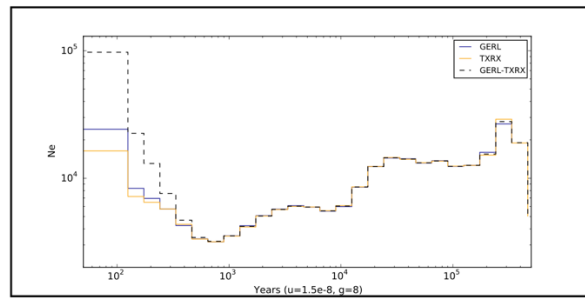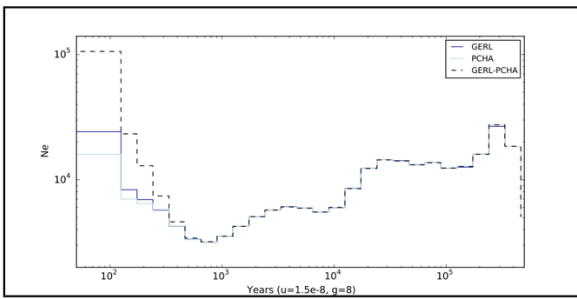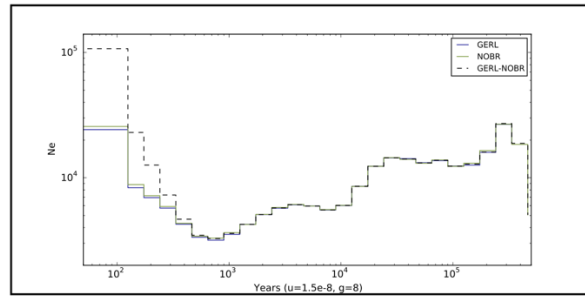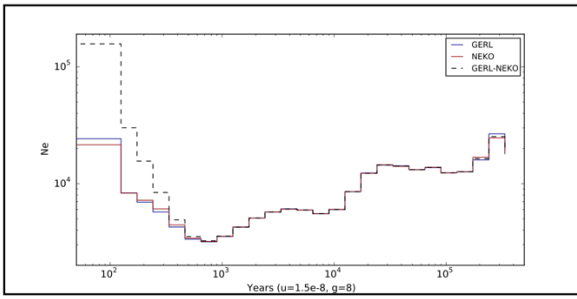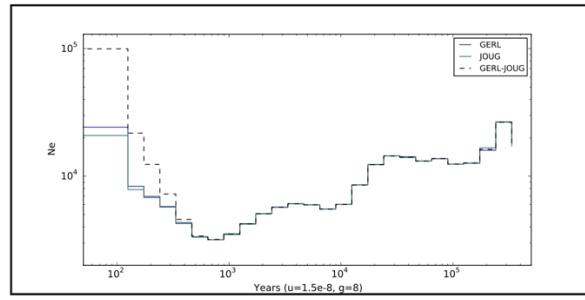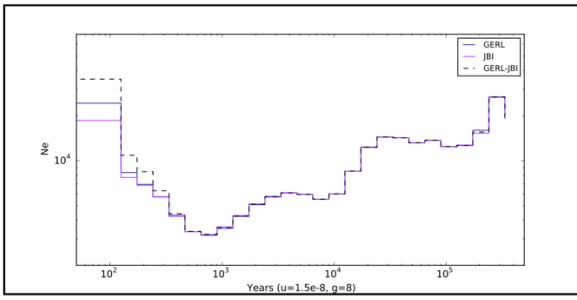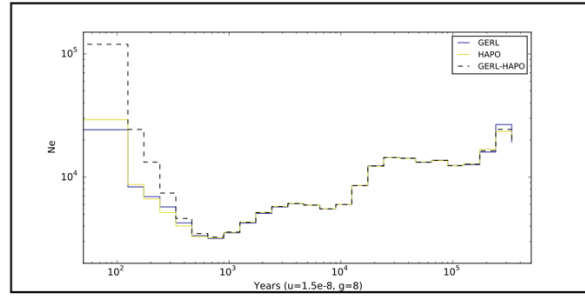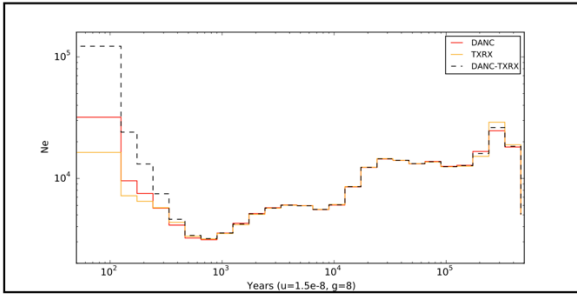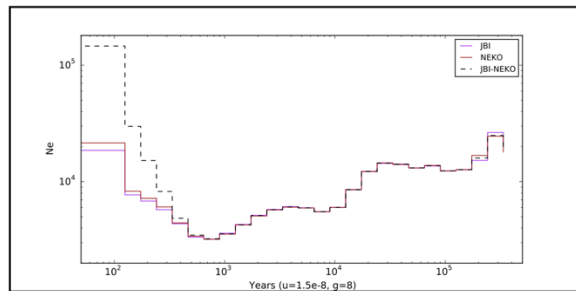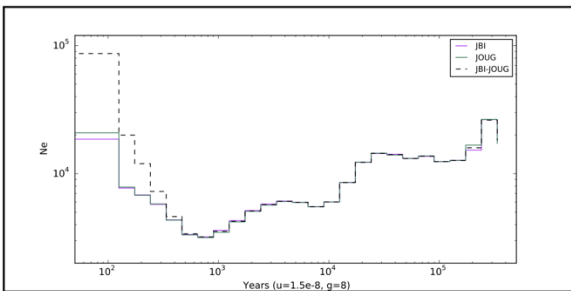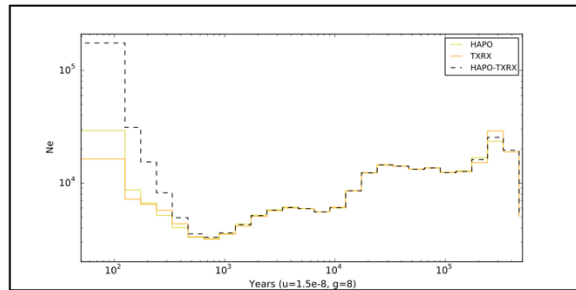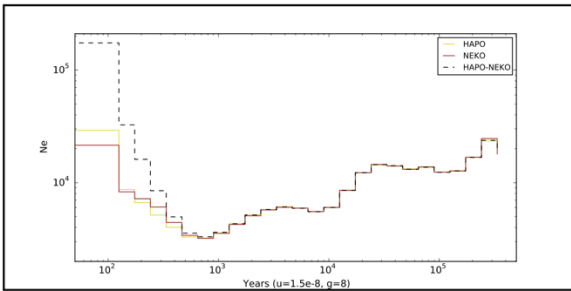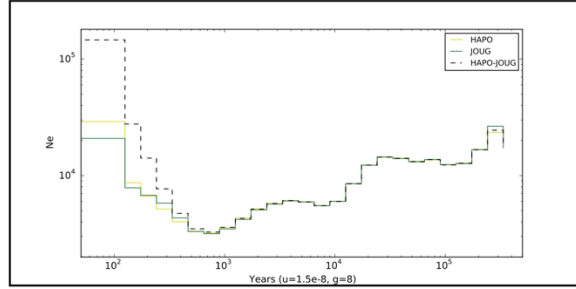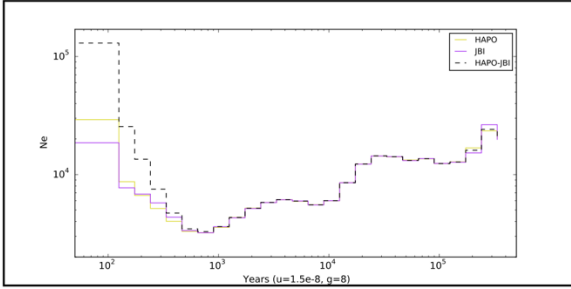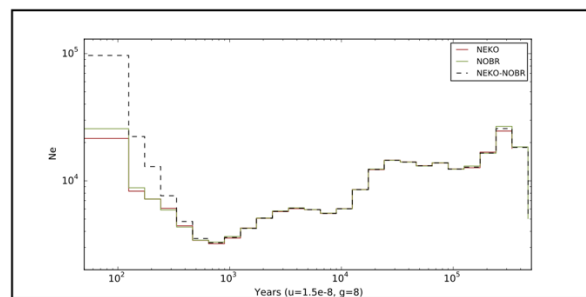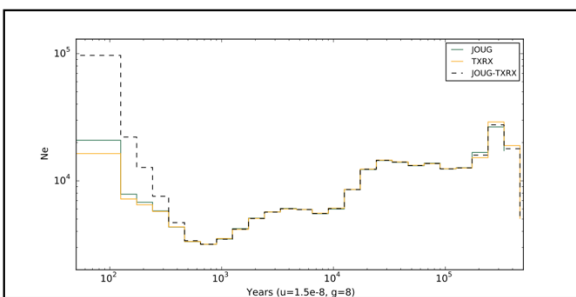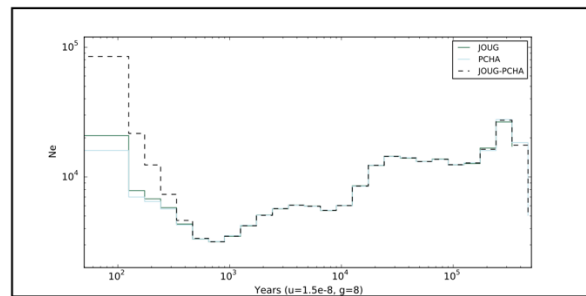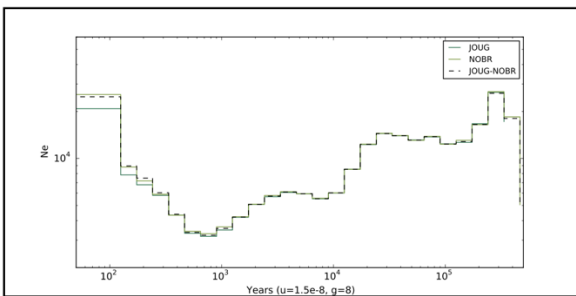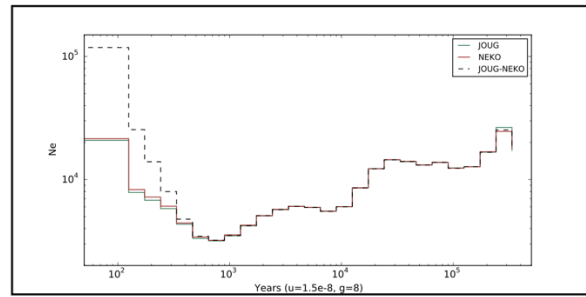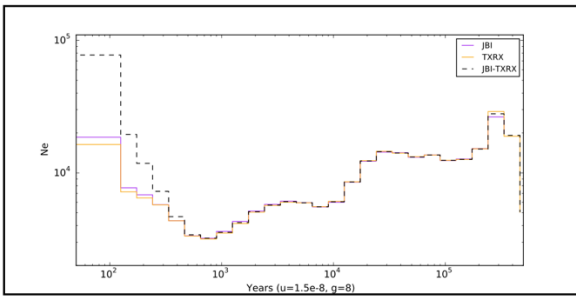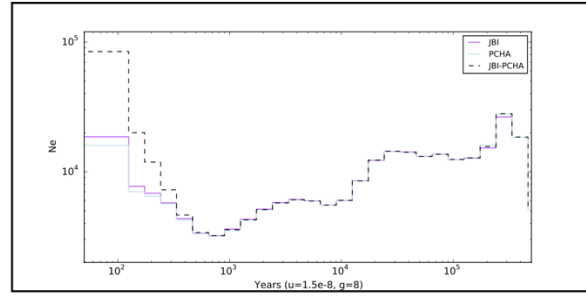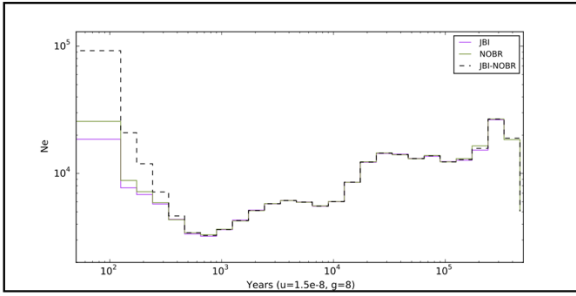
Appendix 4.1: ABM model code

```
### Load files
library(sf)

individual_track_start <- read.csv("/gpfs/projects/LynchGroup/Gentoo_ABM/indies_all_e
xpand_final.csv", header = TRUE)
dist_pixels <- read.csv("/gpfs/projects/LynchGroup/Gentoo_ABM/combined_clustered_suit
abilities.csv", header = TRUE)
dist_pixels_sf <- st_read("/gpfs/projects/LynchGroup/Gentoo_ABM/dist_pixels_sf.shp")
dist_pixels_matrix <- st_distance(dist_pixels_sf)

rownames(dist_pixels_matrix) <- dist_pixels[,1]
colnames(dist_pixels_matrix) <- dist_pixels[,1]

# Get unique colony values
unique_colonies <- unique(dist_pixels$colony)

# Initialize a vector to store total pixel counts
total_pixels <- numeric(length(unique_colonies))

# Loop through unique colonies and calculate total pixel count
for (i in 1:length(unique_colonies)) {
  colony <- unique_colonies[i]
  colony_pixels <- individual_track_start$pixel[individual_track_start$colony == colo
ny]
  total_pixels[i] <- length(colony_pixels)
}

# Create a summary dataframe
colony_summary <- data.frame(colony = unique_colonies, total_pixels = total_pixels)


### Start here
# Define the number of iterations
# num_iterations <- c(1:250)

# Run the loop
for (iter in 651:700){
  individual_track <- individual_track_start
  prob_surv <- 0
  penalty <- 0.9
  N = 20
  t <- 0 # Initiate counter of years
  while(t < N)
  {
    # Set up allee effects data
    # Loop through unique colonies and calculate total pixel count
   for (i in 1:length(unique_colonies)) {
      colony <- unique_colonies[i]
      individuals_alive <- individual_track[!is.na(individual_track$age), ]
# Missings in x1
      colony_pixels <- individuals_alive$pixel[individuals_alive$colony == colony]
      total_pixels[i] <- length(colony_pixels)
```

```r
    }

    # Create a summary dataframe
    colony_summary <- data.frame(colony = unique_colonies, total_pixels = total_pixel
s)
    for (i in 1:nrow(individual_track)){
      if(individual_track$age[i] == 0 & !is.na(individual_track$disperse[i])){ # if i
ndividual is age 0
        prob_surv <- rbinom(1, 1, 0.43) # draw probability of survival for juvenile
      }
      if(individual_track$age[i] == 1 & !is.na(individual_track$disperse[i])){ # if i
ndividual is age 1
        prob_surv <- rbinom(1, 1, 0.82) # draw probability of survival for juvenile
      }
      if(individual_track$age[i] == 2 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 3 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 4 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 5 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 6 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 7 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 8 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 9 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 10 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 11 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 12 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 13 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 14 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 15 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 16 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 17 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 18 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 19 & !is.na(individual_track$age[i])){
        prob_surv <- rbinom(1, 1, 0.82)}
      if(individual_track$age[i] == 20 & !is.na(individual_track$disperse[i])){ # if
individual is 15
        individual_track$age[i] <- NA # kill!
        individual_track$reproduce[i] <- NA
        individual_track$disperse[i] <- NA
```

```r
      }
      if(prob_surv == 1) {individual_track$age[i] <- as.numeric(individual_track$age[
i]) + 1 # if survive, add + 1 to age
      }
      if(prob_surv == 0){
        individual_track$age[i] <- NA # if not, kill!
        individual_track$reproduce[i] <- NA
        individual_track$disperse[i] <- NA
      }

      if(!is.na(individual_track$age[i])){

        allee <- colony_summary[colony_summary$colony %in% individual_track$colony[i]
,]
        ifelse(allee$total_pixels < 5, penalty <- allee$total_pixels / 5, penalty <-
0.9)

        reproduce <- rbinom(1,1,(penalty))  # draw probability of reproducing
        if(reproduce == 1){
          individual_track$reproduce[i] <- "yes" # change repro status
        }
        if(reproduce == 0){
          individual_track$reproduce[i] <- "no" # change repro status
        }
        if(reproduce == 1){
          chick <- c(individual_track$colony[i],individual_track$pixel[i], 0, "no", "
no") # create new chick
          individual_track[nrow(individual_track) + 1,] <- chick # add new chick row
to dataframe
        }
      }
      if(!is.na(individual_track$age[i])){ # if not dead
        dispersal <- rbinom(1,1,0.4) # draw probability of dispersal
        ifelse(dispersal == 1, individual_track$disperse[i] <- "yes", individual_trac
k$disperse[i] <- "no" ) # update dispersal status

      }
      if (individual_track$disperse[i] == "yes" & !is.na(individual_track$age[i])){
        distance <- as.numeric(dist_pixels_matrix[rownames(dist_pixels_matrix) == ind
ividual_track$pixel[i],])

        prospect_pixels_40 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 400
00 & distance < 50000])
        prospect_pixels_30 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 300
00 & distance < 40000])
        prospect_pixels_20 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 200
00 & distance < 30000])
        prospect_pixels_10 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 100
00 & distance < 20000])
        prospect_pixels_5 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 5000
& distance < 10000])
        prospect_pixels_1 <- as.numeric(colnames(dist_pixels_matrix)[distance > 0 & d
istance < 5000])

        #prospect_pixels_20 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 20
000 & distance < 30000])
```

```
        #prospect_pixels_10 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 10
000 & distance < 20000])
        #prospect_pixels_5 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 500
0 & distance < 10000])
        #prospect_pixels_1 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 100
0 & distance < 5000])
        #prospect_pixels_0.5 <- as.numeric(colnames(dist_pixels_matrix)[distance >= 5
00 & distance < 1000])
        #prospect_pixels_0.1 <- as.numeric(colnames(dist_pixels_matrix)[distance < 50
0])

        suitabilities_40 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_40,]
        suitabilities_40$suitability <- suitabilities_40$suitability * 0.2

        suitabilities_30 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_30,]
        suitabilities_30$suitability <- suitabilities_30$suitability * 0.4

        suitabilities_20 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_20,]
        suitabilities_20$suitability <- suitabilities_20$suitability * 0.6

        suitabilities_10 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_10,]
        suitabilities_10$suitability <- suitabilities_10$suitability * 0.8

        suitabilities_5 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_5,]
        suitabilities_5$suitability <- suitabilities_5$suitability * 0.9

        suitabilities_1 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_1,]
        suitabilities_1$suitability <- suitabilities_1$suitability

    #suitabilities_20 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_20,]
        #suitabilities_20$suitability <- suitabilities_20$suitability * 0.5

        #suitabilities_10 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_10,]
        #suitabilities_10$suitability <- suitabilities_10$suitability * 0.6

        #suitabilities_5 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_5,]
        #suitabilities_5$suitability <- suitabilities_5$suitability * 0.7

        #suitabilities_1 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_1,]
        #suitabilities_1$suitability <- suitabilities_1$suitability * 0.8

        #suitabilities_0.5 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_0.5,
]
        #suitabilities_0.5$suitability <- suitabilities_0.5$suitability * 0.9

        #suitabilities_0.1 <- dist_pixels[dist_pixels$pixel %in% prospect_pixels_0.1,
]
        #suitabilities_0.1$suitability <- suitabilities_0.1$suitability

        suitabilities <- rbind(suitabilities_40, suitabilities_30,suitabilities_20, s
uitabilities_10, suitabilities_5, suitabilities_1)

    #suitabilities <- rbind(suitabilities_20, suitabilities_10, suitabilities_5,
        #                    suitabilities_1, suitabilities_0.5, suitabilities_0.1
)
```

```r
        multinom <- rmultinom(1,1, prob = (suitabilities$suitability))# draw suitabil
ity from multinomial
        colonies <- suitabilities$colony
        dispersal_choices <- cbind(colonies, suitabilities, multinom)
        final_choice <- subset(dispersal_choices, multinom==1)
        individual_track$pixel[i] <- final_choice$pixel
        individual_track$colony[i] <- final_choice$colony
    }
    t <- t + 1 #update counter
  }
  # Your code for each iteration goes here

  # Assuming you have a dataframe named "result_df" with the results
  result_df <- data.frame(individual_track)  # Your data

  # Generate a unique file name based on the iteration number
  file_name <- paste("result_iteration_", iter, ".csv", sep = "")

  # Save the dataframe to a CSV file
  write.csv(result_df, file = file_name, row.names = FALSE)

  # Clear variables if needed to free up memory
  rm(result_df)
  }
}
```