Estimating breeding status in Atlantic puffin colonies across Newfoundland: A methodological comparison

By

Raul Zabala Belenguer

Submitted in partial fulfilment of the requirements for the degree of

Master of Science

Cognitive and Behavioural Ecology

Memorial University of Newfoundland

April 2023

St. John’s Newfoundland and Labrador

Açó es pa tu iaio. On estigues, pero sempre amb mi.

Abstract

The largest colonies of Atlantic puffins (*Fratercula arctica*) have been experiencing decades of declining population growth linked to poor breeding performance, particularly in the Eastern Atlantic. These trends have been revealed by the presence of colony-specific monitoring programs. Such data are fragmented and not updated for Newfoundland (Canada) colonies, the largest in the Western Atlantic. Here, I have assessed the burrow laying success, fledging success, and productivity of five colonies at different latitudes in the 2021-2022 breeding season through the establishment of permanent plots. Direct comparisons between current and historical estimates were not possible due to differences in burrow assessment methods. As a remedy, I compared detection probabilities obtained by two different methods, burrowscoping and handgrubbing, and estimated a correction factor to allow for comparisons. Inter-rater reliability of the estimates was also evaluated. My findings show that estimates can be influenced by both data collection method and double-observer, even with experienced individuals. Nevertheless, every breeding parameter remained high in all colonies included in this study, suggesting an overall healthy breeding status in Newfoundland puffin populations, even in those where no historical data are available. This makes Newfoundland colonies the largest puffin aggregation worldwide with no signs of breeding failure in this declining species.

General Summary

Atlantic puffin (*Fratercula arctica*) colonies are declining worldwide. The origin of this decline has several components, but a large part of it is due to long-term breeding failures. Thanks to a multi-colony breeding performance evaluation, I have confirmed that Newfoundland Atlantic puffins breeding status is healthy, producing enough chicks to sustain the population numbers. I have also designed an experiment to evaluate the effects of the assessment method in the breeding performance estimates, visual observation of the nest contents with a burrowscope versus hand inspection without visual confirmation, resulting in substantial differences in the results between methods.

Acknowledgements

Som la gent que no es rendeix

Fràgils, humils, supervivents

Tota una vida a la deriva

Amb l'esperança entre les dents

Here I am. Close to the surface, almost catching my breath again. Mild winds bringing me to port after the storm. The victory lap, and the warrior's rest. In this empty page, soon to be written, I see all of you. Writing this document was not an easy job, but if it is here, is largely thanks to you. Thanks to you I'm stronger, wiser, and more confident. All of you, thanks for telling me that I can, because turned out you were right. I would like to start with my supervisor, Pierre-Paul, because you parked all this. Thanks for trusting me, for giving me independence and advice when I needed one, the other, or both. Thanks for creating all the controlled (most of the time) chaos my project and our lab is, and for putting together an incredible human team without whom this document, and my past two years, wouldn't exist. Thanks to my committee, Bill Motevecchi and Sabina Wilhelm, for all the kind words, ideas, and experience. You have helped in making an idea become reality. None of your input was in vain, and it helped tremendously to increase the quality of this project. I would now like to thank the humans behind the act, my lab mates, and my everlasting friends: Amy (and Forrest), Antoine, Fiona, Jill, Bobby. This would go forever but thanks for all the meals, the field season craziness, the fires, the random adventures, the nights out, the nights in, the baked goodies, and the support. I hope I meant to you as much as you meant to me, and I hope you remember me with the same appreciation. You are simply the best. Thanks to the Zoo, and all the living creatures inside. Now it is emptier, in many ways, but that room will always be filled with your presence. Of these living organisms, I would like to give a special mention to Sydney and Kyle. Jokes aside, you have brought me light and warmth with your friendship. Thanks for the socks, the pizza and movie nights, and the weird fruits that-you-shouldn't-be-able-to-eat-in-Newfoundland-but-somehow-I'm-eating-them. I still don't like sour beers, sorry. Thanks to everyone that has helped me with my project. I must bring forward Lily and Chandler, your positive and willing attitude has made impossible situations become more manageable. Thanks to Chris and Josh, you gave me a raft in the middle of the storm, quite literally. Thanks to everyone that has supported my insane fieldwork. Fogo Islands will be always in my heart, and besides for their breathtaking nature, it will be for the quality of their people. Gordon and Randy from Shorefast, Austin and his brother, Joelle and her family… The list goes on and on. Thanks to the people at Willbur Hobbs in Elliston Point, for giving me a hand when I most needed it. Thanks to Chavela and George, down at Bauline East. Those coffees and pastries after five days without sleeping were like balsam in my soul.

Now, like a tree that has roots at the base, I will go to my family. I have confronted opinions in which language this was going to be written, because none of you speak much English, but this is an English text. But since I like Bad Bunny's attitude, I'll do it in Valenciano, and whoever is reading can use Google translate if interested. Moltes gràcies a tots. Se que he estat lluny, i que probablement seguiré estant-lo. Se que en aquestes pàgines hi ha alegria i orgull, però també tristesa i falta. Papà i mamà, no se com de molt enteneu el que faig, però gràcies per no qüestionar el que m'agrada i em fa feliç. Gràcies per mostrar orgull i estar content per mi. I per cridar-me trenta voltes per setmana. Sou molt pesats, però jo també, i vos vull. Lucieta, no se en que muntanya estàs ni a que cabra estàs mirant, però dona igual, tu eres ma germana i no hi ha altra persona en el món millor que tu. Gràcies al tio, a les ties, i cosines, me voldríeu veure més, i jo també a vosaltres. Però seguiu sent la meua casa quan estic lluny. Ara ve la part més que més m'uneix amb un barri que es el que m'ompli. Iaios. Els dos, una que viu encara al barri, l'altre que viu ja sols al meu cap, i els dos al meu cor. Si la meua vida fos una casa, vosaltres serieu les parets. Em protegiu de tot mal, i no importa on i quant. Vos pense a cada dia que passa, sou la memòria de la terra que m`ha parit i la llengua que m'entén i m'explica. And now, because we love a good loop, back to English. I know you are getting itchy already Katja, so this is for you. Thanks for being my anchor in the storm, and my compass in the mist. Thanks for letting me be a part of you, and for accepting being a part of me. I love you beyond reason. I loved you from the moment I saw you, and I will love you until the moment I cease to see. Thanks for everything. I would need a hundred more pages to say it enough.

And puffins, I tried to be the best I could to you. I swear, this was for you. I have no interest in messing with your breeding or anything, I just wanted to know you are alright in Canada at least. Seems like it, and I'm happy. You are the best seabird in the Northern hemisphere, and I hope you forgive me if I've bothered you.

Last but not least, thanks to Memorial University and SGS, for funding and supporting this project. It was quite the ride. I will always have a dear place in my heart for you.

Thanks again,

Raul

Table of Contents

List of Figures

List of Tables

Appendix list

CO-AUTHORSIP STATEMENT

Raul Zabala Belenguer - Conceptualization, Methodology, Data Collection, Data analysis, Writing

Pierre Paul Bitton - Conceptualization, Methodology, Writing (Review and edits), Supervision

William A. Montevecchi – Methodology, Writing (Review and Edits)

Sabina I. Wilhelm – Data collection, Methodology, Writing (Review and edits)

1. Introduction

1.1 Population dynamics

Populations are in close relationship to the resources of the environment they live in, and their numbers depend strongly on environmental factors. In situations where resources are unlimited, species with continuous breeding will grow exponentially (e.g., bacteria, or humans), whereas pulse breeding species will grow geometrically (e.g., deer) (Edwards & Edwards, 2011). However, populations cannot increase perpetually, and therefore their size will be in continuous fluctuation in response to environmental limiting factors (Edwards & Edwards, 2011). These include any type of ecological conditions that limit population growth, like food or space availability, through their effects on breeding, mortality, or migration (Edwards & Edwards, 2011; Messier, 1991).

The influence that limiting factors exert on population growth can be dependent or independent of the density of the population. Density dependent factors are usually biotic in origin (including disease, competition, and predation) and can have a negative or positive relationship with the population size. On the other hand, density independent factors are usually abiotic and include climate, pollutants, and seasonal cycles (such as monsoons or hurricanes). Understanding the relative contribution of each limiting factor on population dynamics is challenging, as they usually interact with each other, modifying their net effects (Bendik & Dries, 2018; Fayet et al., 2021; Wauters & Lens, 1995).

The population size at which a species sees its intrinsic growth rate limited or impeded by the environment is known as carrying capacity (K), above which the population numbers will decline, and below which will remain constant or potentially grow. More specifically, K describes the interaction between organisms and their environment, the capacity of the environment to provide resources for the population to sustain their numbers, or the capacity of a species to exploit environment resources (Fuentes & Ferrada, 2017). In ecology, r/K selection (or r/K continuum theory) describes the combination of life traits that link an organism's breeding capacity to the resources stability and availability in the environment they live in (MacArthur & Wilson, 1967). Stable environments favour K-selected species, in which the strategy that maximizes fitness is the production of few, high-quality offspring. The population growth rate of these species is usually low, maintaining stable numbers once K is reached. These species are long lived, develop slowly, delay sexual maturity, have low reproduction rates, and invest large amounts of resources in each offspring (MacArthur & Wilson, 1967). On the other hand, unstable ecosystems promote the dominance of r-selected species, with breeders that produce a large quantity of offspring, resulting in rapid population growth. These species tend to be short lived, develop rapidly, mature early, have low parental investment, and high breeding rates. Their population dynamics usually follow bursts of exponential population growth followed by rapid population collapses (MacArthur & Wilson, 1967).

This ecological theory implies that r and K are antagonistic strategies: one is maximized at the expense of the other. Therefore, both K and r are linked to the specific relationship between a species and the stability of the environment they live in. As an environment becomes less predictable, its carrying capacity will fluctuate so K-strategists will see their numbers declining as the environment will no longer sustain previous population size, up to a level where the population is no longer viable. There are many reasons environmental conditions can fluctuate unpredictably: accelerated rates of change in climate, habitat loss and fragmentation, chemical contamination, anthropogenic-related changes in nutrient cycles, catastrophic events such as earthquakes or volcanic eruptions, or regime shifts (Bernhardt et al., 2020; Buren et al., 2014, and references therein). The effect of these fluctuations on the population dynamics is usually hard to assess, especially when these effects are non-linear show a non-linear relationship with the population's dynamics (Vilar & Rubi, 2018).

1.2 Seabird demography

Seabirds are a polyphyletic taxon that includes any species of birds that depend on the marine environment for at least part of their lives (BirdLife International, 2012). Traditionally, this includes the orders Sphenisciformes, Procellariformes, Pelicaniformes, Suliformes and Charadriiformes (Keogan et al., 2018). Together, these orders contain over 340 species, accounting for around 3.5 % of the total bird diversity (Croxall et al., 2012). Over 95 % of seabirds are colony nesters, a proportion higher than any other bird group (Lack, 1968). Seabirds are considered a paradigm for K-strategy. Social monogamy and bi-parental care is a widespread breeding strategy in this group (Lack, 1968). Adult survival is generally high, and pairs produce small clutches, ranging from 1-3 eggs, with a long chick-rearing period (Pianka, 2008). Furthermore, in species like common murres (*Uria aalge*) or murrelets (*Brachyramphus* sp.), chicks fledge before having functional flight feathers, and one parent, usually the male, continues feeding the chick until it can fly (Winkler, 2016). This extensive parental care leads to increased chick survival when conditions are optimal and parental quality is robust (Pianka, 2008). Once successfully fledged, seabird chicks take several years to reach sexual maturity (Tasker & Furness, 2003). These strategies and delayed sexual maturity (Table 1) translate into low reproductive rates (Croxall & Rothery, 1991; Pianka, 2008).

Table 1. Mean demographic and biological characteristics of the main seabird families. Adapted and corrected from Croxall & Rothery, 1991.

Group	Age (years) at first breeding	Adult annual survival rate (%)	Clutch size	Chick-rearing period (days)
Sphenisciformes				
Spheniscidae (penguins)	4–8	75–85	1–2	50-80
Procellariiformes				
Diomedeidae (albatrosses)	7–13	92–97	1	116–150
Procellariidae (petrels)	4–10	90–96	1	42-120
Hydrobatidae (storm petrels)	4–5	90+	1	55-70
Pelacanoididae (diving petrels)	2–3	75–80	1	45–55
Pelacaniformes				
Pelacanidae (pelicans)	3–4	85	2–3	55–60
Sulidae (gannets, boobies)	3–5	90–95	1–2	90–120
Phaethontidae (tropic birds)	NA	NA	1	60–90
Fregatidae (frigate birds)	9–10	?	1	140–170+
Phalacrocoracidae (cormorants)	4–5	85–90	2–3	60–90
Charadriiformes				
Alcidae (auks)	2–5	80–93	1–2	15–42
Lariiformes				
Laridae (gulls)	2–5	80–85	2–3	25–50
Sternidae (terns)	2–5	77–90	1–3	25–40
Stercorariidae (skuas)	3–8	93	2	25–40

Population dynamics in seabirds are ruled by four main parameters: adult survival, recruitment, productivity, and immigration-emigration dynamics (Croxall & Rothery, 1991; Tasker & Furness, 2003). In practical terms, demographic studies usually try to estimate the size of the breeding population and its fledging success (Croxall & Rothery, 1991). A general equation for seabird population dynamics is:

$$R_{t+k} = N_t * b_t * s_j * (s_a)^{k-1} * s_a$$

Where R_{t+k} is the size of a cohort of potential recruits available in a population; N_t is the number of breeding pairs k years previously; b_t is the number of chicks fledged per pair k years previously; s_j is the first-year survival rate; $(s_a)^{k-1}$ is the annual survival rate from first-year until recruitment; k = age at recruitment; s_a = annual adult survival rate (Tasker & Furness, 2003). Migration dynamics are considered to have a neutral net effect on the population (immigration equals emigration) due to assessment challenges. A general scheme of these parameters, and how they relate to each other, can be found in Figure 1.

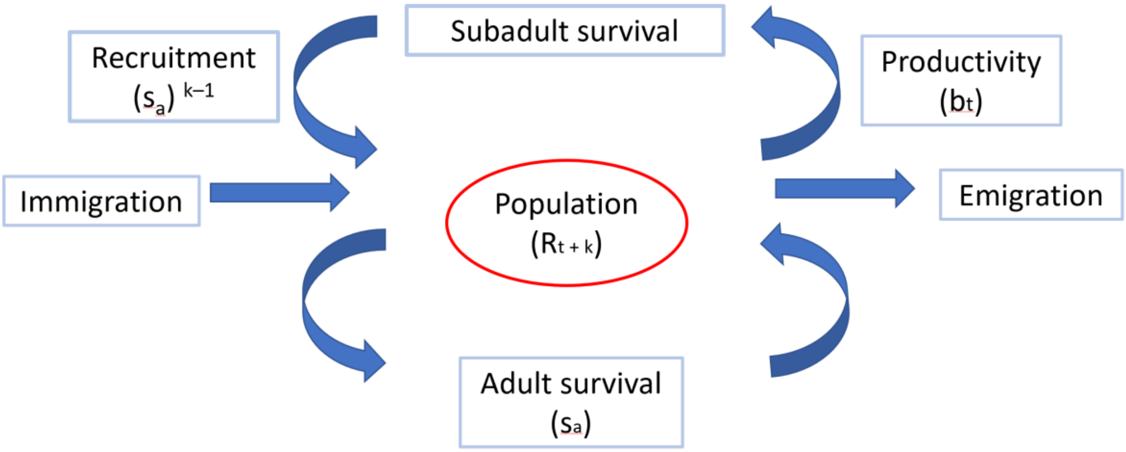


Figure 1. Scheme of the most important parameters regulating seabird populations. Adapted from Tasker & Furness, 2003.

Some of these parameters are difficult to obtain due to the specific biology of seabirds. For instance, juveniles in many species spend most of their immature years at sea, making estimates of subadult survival imprecise (Croxall & Rothery, 1991). Extrapolations from adult survival cannot be made, as there is a body of evidence that, subadults and adults can have different survival rates (Dobson, 1990; Hudson, 1985; Ricklefs, 1973). Other population parameters are easier to obtain. Tasker & Furness (2003) compiled a series of demographic parameter estimates in seabirds and classified them subjectively according to how relevant the data are for understanding population dynamics (desirability), and how feasible data collection is due to logistic or economic reasons (Table 2).

Theoretical models show that in extreme K-selected species like seabirds (extreme high adult survival, minimal reproductive output), small decreases in adult survival will affect population numbers most, while in less extreme K-selected species (high adult survival, moderate reproductive output), the effect of small decreases in juvenile survival and fledging success will have a relative stronger effect in population numbers (Croxall & Rothery, 1991; Tasker & Furness, 2003; Table 3).

Table 2. List of relevant seabird life traits classified based on their utility to understand population demographics (Desirability), and feasibility of data collection to obtain the estimates (+++ highest; ++ high; + moderate; ± more or less; - low). Adapted from Tasker & Furness (2003)

Parameter	Desirability	Feasibility
Population		
Population size (breeders)	+++	+++
Incidence of non-breeding	+	+
Adult survival	+++	++
Subadult survival to breeding	+++	±
Recruitment		
% of fledglings	++	±
% recruits in population	+	±
Recruitment age	+	±
Immigration rate	+	-
Emigration rate	+	-
Sex ratio of the breeding population	±	-
Mean age of breeders	++	±
Reproduction and food provisioning		
Reproductive success	++	++
Laying date	+	++
Clutch size	+	+++
Egg size	+	+++
Chick growth rate	+	++
Mass of fledglings	+	++
Adult nest and brood attendance	++	++
Degree of kleptoparasitism	+	++
Body condition of breeders	+	+

Table 3. Population declines (%) associated with an absolute reduction of 0.05 % in any of the three demographic parameters in two hypothetical species: Species P, with an extremely high adult survival (95 %), low reproductive output (1 egg clutch); Species S, with moderate adult survival (80 %), higher reproductive output (2-3 egg clutch). Adapted from Croxall & Rothery, 1991.

Parameter	Species P	Species S
Adult survival	4.2 %	3.6 %
Juvenile survival	1.5 %	2.4 %
Productivity	0.8 %	1.3 %

Perhaps surprisingly, productivity contributes the least to seabird demography. This is because low productivity can be buffered by their increased individual longevity (Croxall & Rothery, 1991). These theoretical models consider small decreases in the population parameters, but if the average fledging success or juvenile survival decrease is large and constant, the effects will be noticeable in any population (Halpern et al., 2005). For example, in the extreme K-selected species Tristan Albatross (*Diomedea dabbenena*) and Atlantic Petrel (*Pterodroma incerta*), extremely low productivity and juvenile survival due to mice predation, was associated with steady population decline (Cuthbert et al., 2013; Davies et al., 2015). Moreover, colony productivity is the life trait parameter in seabirds that fluctuates the most due to environmental changes, such as variations in sea surface temperature wind patterns, in a bottom-up, indirect fashion (Catry et al., 2013; Ventura et al., 2021). These effects seem to be geographically dependent (Catry et al., 2013; Devney et al., 2009); therefore, long-term monitoring protocols are needed to elucidate the specific environmental factors affecting a particular region. When in place, such studies can reveal changes in environmental conditions, making seabird productivity studies key to assessing ecosystem changes (Catry et al., 2013; Ventura et al., 2021).

Due to this sensitivity to environmental changes, seabirds are considered monitors of the ecosystem (Piatt et al., 2007). Changes in the environment they live in are linked to changes in demographic, behavioural, and reproductive parameters (Piatt, et al., 2007). This sensitivity to environmental fluctuations contributed greatly to making seabirds the most threatened marine animal group. At present, 29 % of species are considered at least “Threatened” by the International Union for Conservation of Nature’s (Spatz et al., 2014). There is a myriad of environmental causes behind this phenomena, both human and non-human mediated: predation by local or introduced species, fishing bycatch, climate change, direct and indirect disturbances, pollution, exploitation, deterioration of breeding space, or food shortages (Croxall et al., 2002). Seabirds show a range of non-linear responses to environmental factors, particularly food availability, obscuring this relationship (Piatt, Harding, et al., 2007). While all these factors can lead to both adult and chick mortality, some can also affect breeding, both in timing (phenology) or success. Therefore, demographic studies should be as complete as possible to have a broader understanding of the population trends and the factors at play. Specifically, most monitoring efforts should include fledging success in some form (Newman et al., 2009; Scott et al., 2009).

1.3 Seabird breeding biology

Seabirds display a diversity of breeding cycles. Some species are year-round breeders; eggs can be produced throughout the year, others are seasonal breeders; the breeding season only occurs at a specific time of the year (Harris, 1969). These strategies are selected depending on habitat-specific conditions. For instance, species that inhabit higher latitudes have shorter breeding windows due to extreme resource seasonality (Descamps et al., 2019). Outside this temporal window conditions can be suboptimal (Durant et al., 2007). Therefore, the relationship between a species' timing of breeding, known as phenology, and resource availability is constrained by adaptation (Walther et al., 2002). This parameter is defined by the mean laying and hatching dates in seabird species with seasonal breeding (Keogan et al., 2018). Different populations of the same species can present different phenologies depending on resource availability, a concept known as allochrony, which can lead to sympatric speciation (Medrano et al., 2022). The match-mismatch hypothesis states that if the most energy expensive part of the breeding phenology of a species occurs at the same time as the peak resource availability, recruitment is maximized (Cushing, 1990; Durant et al., 2007). Inability to adjust phenology to new environmental conditions can have negative effects on an individual's fitness (Miller-Rushing et al., 2010). The environmental factor that is best known to influence in a bottom-up fashion the distribution, abundance, and phenology of seabird prey species is sea surface temperature, or SST (Cheung et al., 2013; Hipfner, 2008; Buren et al., 2014). Unpredictable SST has been associated with increased variation in food availability prior to egg laying, which can lead to decreased fitness (Shultz et al., 2009). Even though matching new environmental conditions is a powerful evolutionary driver, the degree of response varies among species, with species at higher trophic levels being slower to adapt to new conditions due to a longer generation span (Thackeray et al., 2016). Oceanic warming due to climate change has influenced the phenology of Arctic seabirds with some, but not all, species breeding earlier (Descamps et al., 2019). Moreover, comprehensive multi-species analysis show that seabirds have not adapted their phenology to environmental change (Keogan et al., 2018).

Fledging success and productivity are demographically relevant reproductive parameters. The number of chicks produced by a colony depends on the proportion of breeding pairs that incubate an egg, the proportion of these that successfully hatch, and the proportion of chicks that fledge successfully. Therefore, colony productivity (the proportion of total nests in a colony that fledged a chick) can be broken down into two different population parameters: 1) Laying success: the proportion of burrows containing an egg(s) in a given year, 2) Fledging success: the proportion of nest with eggs that successfully fledged a chick (Walsh et al., 1995). These parameters can be highly variable between years and locations, even within the same species (e.g., Rodway et al., 1998). These differences can reflect variances in both biotic and abiotic factors linked to the breeding biology of the species, such as SST (Hansen et al., 2021; Montevecchi et al., 2021), climatic oscillations (Lewis et al., 2009; Surman & Nicholson, 2009), changes of preferred prey items availability (Guillemette et al., 2018), increased predatory pressure (Rodway & Lemon, 2011; Stenhouse et al., 1999; Wanless et al., 2007), or colony size (Lewis et al., 2009). Other factors affecting breeding performance are intrinsic to the bird: age-related breeder experience (Froy et al., 2017), senescence (Reed et al., 2008), previous fledging success (Lewis et al., 2009), or stress levels (Doody et al., 2008). Since different species respond differently to environmental factors, multi-species analysis at different latitudes can be helpful in disentangling intrinsic-extrinsic factors affecting productivity.

1.4 Breeding assessment challenges

The study of each of the different demographic parameters presents intrinsic challenges. Seabird colonies are often located in remote areas, where access is limited, and constant nest monitoring is often not possible making nest assessment difficult. This is particularly true for burrowing species for which direct visual evaluation is not possible.

Burrowing seabirds comprise about a third of the total seabird species (Del Hoyo et al., 1992), and burrow content (presence of eggs or chicks) must be checked to evaluate breeding. In reproductive assessments, a seabird nest is defined as active when there is evidence of breeding within the nest (Bled et al., 2011; Cuthbert et al., 2013; d'Entremont et al., 2020; Diamond, 2021; Priddel et al., 2006). The collection of these data is particularly difficult in burrowing species due to the impossibility of direct visual assessment, the narrow and winding nature of their burrows, and the fragility of the soil they inhabit (Savard & Smith, 1985; Schumann et al., 2013). For these reasons, count units for laying success calculations can be nests with confirmed active breeding, but apparently occupied nests (nests with fresh diggings at the entrance, presence of an adult in the nest, etc.) can be used when nest assessment is difficult (Walsh et al., 1995). Nevertheless, precision in the estimates is expected to be higher with direct assessments. To obtain direct evidence of breeding, researchers have traditionally used a method known as "handgrubbing", during which the assessment is conducted by hand, introducing the arm inside the burrow to check its content (Ambagis, 2004). This practice has limitations in burrows where incubation chambers are deeper than an arm's reach (Lavers et al., 2019), and can also lead to significant burrow damage (Ambagis, 2004) that can affect breeding (Rodway et al., 1996). To reduce these negative effects, researchers might not check the burrow thoroughly to reduce stress to the incubating bird. This can lead to unreliable breeding estimates (Rexer-Huber et al., 2014). To prevent researcher disturbance, a variety of non-invasive methods have been developed, including the detection of fresh diggings or feces at burrow entrances (Rexer-Huber et al., 2014), response to audio playbacks (Orben et al., 2019; Ratcliffe et al., 1998) or camera traps (Bird et al., 2021). Due to the non-invasive nature of these techniques, they only provide indirect information on the presence of an egg or chick.

1.4.1 Burrowscopes as burrow assessment tools

In the 1990s, burrowscopes were introduced as tools for assessing nest contents in burrowing species (Lyver et al., 1998). Most of these devices consist of an infrared camera housed at the tip of a hose connected to a visualizing device (e.g., monitor). The cable with the infrared light source is introduced inside the burrow, and its contents are assessed without direct contact with the animal, potentially minimizing disturbance (Carey, 2009). Unfortunately, detection rates are not perfect. Some studies have compared burrowscope detection rates to the true data obtained after researchers dug the burrow sites. These studies showed that burrowscopes can underestimate between 17% (Mckenchie et al., 2007) to 34% (Hamilton, 2000) the laying success in Sooty Shearwaters (*Ardena grisea*). Factors influencing the accuracy of occupation estimates include handler experience, burrow complexity, and weather (Lyver et al., 1998). Because of the binomial nature of detection, some studies have applied mark-recapture analysis to establish method-specific detection probabilities without the need for plot excavation (Whitehead et al., 2014). Burrowscope prototypes were very rudimentary, but as their application as detection tools expanded, the technology improved yielding better detection rates (Lavers et al., 2019). Despite these challenges, burrowscopes have been increasingly used for determining breeding success, being used as comparison baseline to assess the detection accuracy of other methods (Albores-Barajas et al., 2018). Nevertheless, animals dig and inhabit burrows with different traits, and assessment of possible detection biases need to be conducted for each study species.

1.5 Study species

1.5.1 Breeding biology

The Atlantic puffin (*Fratercula arctica*, hereafter puffins) is an iconic migratory and pelagic seabird that inhabits the Northern Atlantic. Its distribution extends from the Arctic Western Russian islands to Atlantic Canada, with southern limits reaching the northern United States and France. Puffins are pursuit-divers, and individuals can dive up to 80 meters and actively chase prey underwater at high speeds, an energetically demanding activity (Harris & Wanless, 2011). Individuals are found in land colonies only during the breeding season, where females lay a single egg (Harris & Wanless, 2011). The species is monogamous and mated individuals rarely change burrows (Harris & Wanless, 2011). The species exhibits natal philopatry, though it is highly variable among colonies (Kersten et al., 2021). Atlantic puffins reach sexual maturity approximately at 4-5 years of age, and can live up to 32 years (Harris & Wanless, 2011). Adult (reproductive) survival is considered to be high, with survival and breeding rates decreasing with age (Breton & Diamond, 2014; Harris et al., 1997). Both parents are involved in incubation (39 to 43 days on average) and chick rearing (38 to 44 days on average) (Harris & Wanless, 2011). These require heavy energetic investment from both parents, through incubation and direct food provisioning during chick rearing (Harris & Wanless, 2011). After rearing, chicks fledge on their own, mostly at night, when they reach about 60-80 % of adult body mass (Lowther et al., 2020). Fledging success is negatively correlated with prey scarcity (Diamond, 2021; Durant et al., 2003), but high adult survival and longevity can buffer population numbers during low productivity years (Grosbois et al., 2009).

1.5.2 Long term monitoring in Atlantic puffins

Atlantic puffins are no exception to the general seabird decline. This auk species is considered “Vulnerable” (BirdLife International, 2018), an assessment heavily influenced by large population declines in the European populations (Kersten et al., 2021). Several factors have been proposed to explain this phenomenon: fluctuations in the oceanic conditions (Durant et al., 2003; Fayet et al., 2021; Mesquita et al., 2015), introductions of predators (Evans & Nettleship, 1985), extreme winter conditions (Anker-Nilssen et al., 2017; Harris et al., 2005), and/or decline of traditional prey species (Diamond, 2021; Fayet et al., 2021; Harris & Wanless, 2011). All these factors can affect populations at different levels, from long-term breeding failures (Fayet et al., 2021; Hansen, 2021), low juvenile recruitment (Miles et al., 2015), or increased adult mortality (Anker-Nilssen et al., 2017; Evans & Nettleship, 1985).

The effects of such declines are noticeable. Two of the species’ largest colonies have experienced steep declines in the past decades: once the largest Atlantic puffin colony in the world, Røst (Norway) has suffered an 81 % decline since 1980 (Anker-Nilssen et al., 2020; Fayet et al., 2021) and a 45 % decline has been observed in Westmann Islands, south Iceland (Hansen, 2021). Long term monitoring studies have informed these population trends (Table 4), demonstrating a consistent, long-term breeding failure, as one of the factors behind this decline (Durant et al., 2004; Frederiksen et al., 2004; Hansen, 2021). Some of these studies use different burrow inspection methods. There is no specific information on content detection rates in Atlantic puffin burrows when using different assessment methods, so comparisons can be challenging.

Table 4. Summary of available data on proportion laying success (burrows with eggs / total burrows), fledging success (fledged chicks / total number of eggs), and productivity (fledged chicks / total burrows) across the Atlantic puffin range, with colony population size and population trend estimated for each colony. Values averaged over the total study period (Icelandic colonies, 2010-2021; Isle of May, 1977-2021; Wales, 2013-2021; Fair Isle, 1987-2013).

Region/Colony	**Population size**	**Trend**	**Avg. Laying success (SE)**	**Avg. Fledging success (SE)**	**Avg. Productivity (SE)**
North Iceland[1]	509 000	**Increasing** +59.6% since 2010	0.815 (0.0144)	0.802 (0.0155)	0.655 (0.0247)
West Iceland[1]	562 000	**Decreasing** -21.8% since 2003	0.629 (0.0544)	0.568 (0.092)	0.397 (0.0741)
East Iceland[1]	483 000	**Decreasing** -13.3% since 2003	0.726 (0.0311)	0.581 (0.0936)	0.432 (0.0708)
South Iceland[1]	1 125 000	**Decreasing** -46% since 2007	0.579 (0.0473)	0.367 (0.0936)	0.234 (0.0630)
Isle of May (Scotland)[2, 3]	78 400	**Increasing** +48% since 1989	NA	NA	0.683 (0.015)
Fair Isle (Scotland)[4]	10 700	**Decreasing** -47% since 1987	NA	0.621(0.02)	NA
Wales[5, 6]	48 216	**Increasing** +125% since 2004	0.703	0.934	0.656
Røst[7] (Norway)	274 000	**Decreasing** -81% since 1979	NA	0.2	NA
Anda[8] (Norway)	100 000-250 000[9]	**Stable** Since 2011	NA	0.692 (0.056)	NA
Hjelmsoya[8] (Norway)	10,000-50 000[9]	**Stable** Since 2011	NA	0.364 (0.043)	NA
Hornoya[8] (Norway)	10,000-50 000[9]	**Decreasing** Since 2011	NA	0.424 (0.067)	NA
Runde[8] (Norway)	50 000-100 000[9]	**Decreasing** Since 2011	NA	0.320 (0.051)	NA

1: Hansen, 2021; 2: Newell et al., 2022; 3: Joint Nature Conservation Committee, 2021; 4: Miles et al., 2015; 5: Newman et al., 2021; 6: Brown & Eagle, 2021; 7: Fayet et al., 2021; 8: Anker-Nilssen et al., 2021; 9: Anker-Nilssen et al., 2003.

1.5.3 Breeding success in Canadian Atlantic puffin populations

There is limited information on breeding success data on Western Atlantic puffin colonies, most of it coming from the Machias Seal Island population (Diamond, 2021; Major et al., 2021). There are no updated, systematically collected data on breeding status of other North American colonies, including Newfoundland colonies (Table 5), despite being the largest in North America with an estimated population size of 897,566 breeding pairs (BirdLife International, 2018; Lowther et al., 2020; Wilhelm, unpubl. data). Most data available from Newfoundland colonies comes from population surveys, where egg or chick fate was not followed through the season, providing incomplete information on fledging success and productivity (Robertson et al., 2004; Wilhelm et al., 2015; Wilhelm, pers. comm.). This is mostly due to the inaccessibility of the colonies and their sensitivity to researcher disturbance. Evidence shows that fledging success in frequently assessed burrows by handgrubbing can be reduced as much as 52 % in Newfoundland colonies (Baillie, 2001; Rodway et al., 1996; Scott et al., 2009; Spear et al., 1995). This makes it challenging to assess the breeding status of Newfoundland Atlantic puffin colonies and highlights a need for more consistent assessments with a less invasive method.

Table 5. Collectionof relevant historical data available for this study on population size and trend, laying success (number of eggs / burrows checked), fledging success (fledged chicks / eggs monitored), productivity (fledged chicks / burrows monitored), and mean hatching date in every colony considered for this research across the province of Newfoundland and Labrador.

Colony (Trend, Pop. Size)	Year	Laying success	Fledging success	Productivity	Hatching day
Great Island, Witless Bay, Newfoundland (Increasing, 174 491)[4]	1968[1]	-	0.42	-	184 (mean)
	1969[1]	-	0.58	-	181 (mean)
	1972[4]	0.54	-	-	-
	1984[4]	0.84	-	-	-
	1985[4]	0.9	-	-	-
	1992[2]	0.76	0.55	0.42	190 (mean)
	1993[2]	0.87	0.68	0.58	189 (mean)
	1994[b]	-	-	0.6	-
	2011[4]	0.72	-	-	-
Gull Island, Witless Bay, Newfoundland (Increasing, 118 401)[5]	1998[3]	0.63	0.57	0.36	187 (mean)
	2012[5]	0.71	-	-	-
	2013[6]	-	0.75	-	-
	2019[3]	-	0.84	-	-
Bakeapple Island, Little Fogo Islands, Newfoundland (Increasing, 2 072)[5]	2014[5]	0.31	-	-	-
Puffin Island, Little Fogo Islands, Newfoundland (Increasing, 4 125)[5]	2014[5]	0.56	-	-	-

1: Nettleship, 1972; 2: Rodway, 1994; 3: Wilhelm, 2021; 4: Wilhelm et al., 2015; 5: Wilhelm, unpublished data; 6: Fayet et al., 2017.

1.6 Study goals

There is a need for updated laying success, fledging success, productivity, and mean hatching date estimates in Newfoundland puffin colonies. Burrowscopes have been showing an improvement in detection rates of burrow contents, but there is a lack of data on method-related biases when estimating these parameters. This potential bias needs to be addressed through a correction factor to compare estimates obtained by different methods.

The aims of this study are to:

a) Evaluate biases associated with two widespread surveying methods, handgrubbing and burrowscopes.
b) Survey burrow depth in different colonies across Newfoundland.
c) Establish method-specific detection probabilities and reliability estimates when assessing Atlantic puffin burrows.
d) With this data, elaborate correction factors to make comparisons between these methods.
e) Compare past and present breeding parameters among different colonies in Newfoundland, using the forementioned correction factors when needed.
f) Compare body mass-wing length ratios between puffins from colonies situated at different latitudes.

2. Methods

2.1 Methodological comparison

2.1.1 Study site

The experiment took place on Great Island, Witless Bay (Figure 2) in 2022. The site selected for the experiment was based on three factors assessed during the survey: a high-density of burrows, easy access for researchers, and stable soil conditions to prevent burrow destruction.

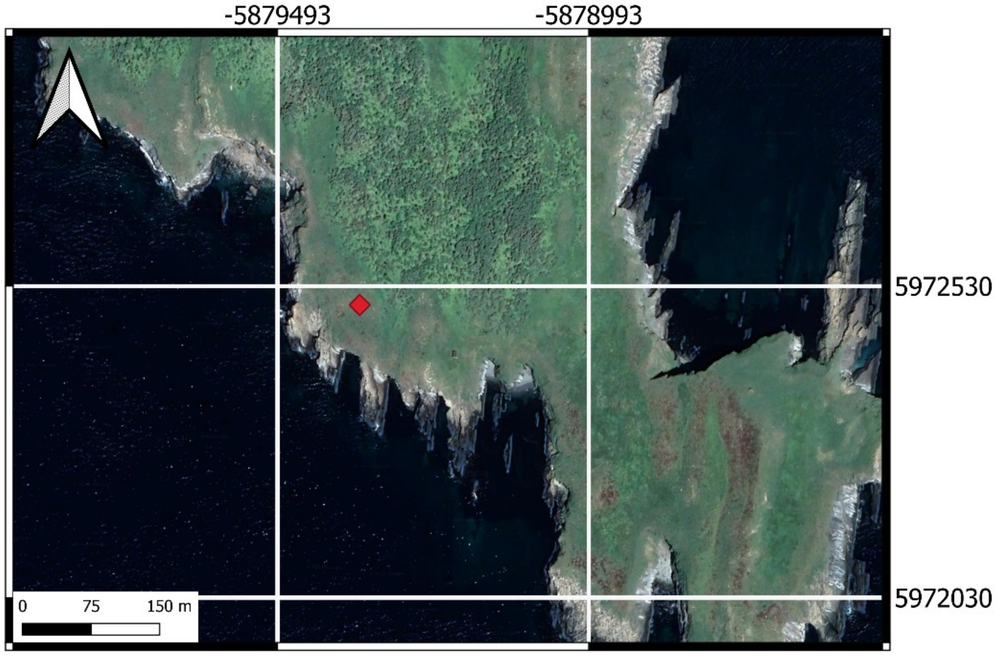


Figure 2. General location (47.1833, -52.8119) where the methodological comparison experiment was conducted on Great Island, Witless Bay, Newfoundland.

2.1.2 Experimental design

The experiment followed a standard capture-mark-recapture approach, based on Whitehead et al., 2014. Twelve burrows were selected randomly within the selected site, in four different areas deemed as representative of puffin habitat (48 burrows in all). A researcher independent of the posterior experiment selected and flagged burrows that were suitable to host a breeding pair. The content of every flagged burrow was then assessed by a team of four observers: two by handgrubbing, and two by burrowscope (EMS2021 Gopher Tortoise Camera System with infrared detection, Environmental Management Services, Canton, Georgia, USA). Burrow contents were evaluated twice, once during the incubation period (egg assessment), and once during brood rearing (chick assessment), for a total of eight checks per burrow. To avoid assessment biases, every plot was out of the visual range from the others, and only one observer was on any given plot during content evaluation. All observers had extensive experience in the method they used for the assessment (> 200 burrows checked using that method before the start of the experiment).

Each burrow was classified by each observer into one of these two categories: a) Active, defined as burrows with breeding activity content (egg or chick); b) Inactive, defined as burrows where no breeding activity was detected, either fully empty burrows or just an adult. In case of inactive burrows, the researcher scored them as a) Solved, when the researcher was able to explore the totality of the burrow; or b) Unsolved, if areas were out of reach (e.g., burrow was too deep, too complex for a proper assessment, or obstacles were in the way).

To obtain a proportion of unsuitable nests for handgrubbing assessment, the incubation chamber depth was measured in a subset of the burrowscope assessed burrows during the 2022 breeding season. For that, burrowscope hoses were marked every 5 cm. This measurement was obtained by approximating the tip of the hose to the egg / chick, without direct contact. Then, the depth of the chamber was estimated by approximating the number of 5 cm bands between the tip of the hose to the vertical line between the ceiling and bottom of the burrow entrance. Any measurements with high uncertainty associated (e.g., chick displacement within the burrow) were discarded.

2.1.3 Statistical analyses

2.1.3.1 Inter-reliability tests

Inter-rater reliability (IRR) is defined as the extent to which data collectors (raters) assign the same score to a trait to validate the data collected (McHugh, 2012). Cohen's kappa was the index used to evaluate the IRR between two observers when estimating plot laying success (Active burrows / Total burrows in the plot) within assessment method. This coefficient ranges from 0, when the inter-rater agreement is entirely due to chance, and 1, when the agreement between observers is complete (McHugh, 2012). The value of this coefficient is correlated with the percentage of data collection that is a true representation of the burrow contents after the assessment (Table 6). The experimental design fits the three pre-requisites for the use of this coefficient: 1) independent observers; 2) clear evaluation scales; and 3) assessment of the same sample by different observers. Statistical analyses were conducted using the "CohenKappa" function included in the "DescTools" R package (Signorell, 2022). Standard error intervals associated to the coefficient we calculated following the protocol described in McHugh, 2012. Statistical differences between the laying success estimates obtained by the two observers were assessed using Fisher's exact tests.

Table 6. Correspondence between different Cohen's kappa values and the percentage of the data collection that is valid. Table adapted from McHugh, 2012.

Kappa value	Level of Agreement	Reliable data (%)
0-0.20	None	0 - 4 %
0.21-0.39	Minimal	4 - 15 %
0.41-0.59	Weak	16 - 35 %
0.60-0.79	Moderate	36 - 63 %
0.80-0.90	Strong	64 - 81 %
Above 0.90	Perfect	82 - 100%

Cohen’s kappa was also used to compare the agreement between total handgrubbing and burrowscope assessments. For this, the results of the two individual assessments for each method were combined into one, scoring a burrow as active if any of the observers detected breeding a chick or an egg, and inactive if not. False positives (e.g., detection of an egg or chick when there is none) are very rare. The laying success values (eggs or chicks/total number of burrows) were then calculated using the data obtained from the two different methods eliminating all the “Unsolved” burrows. Fisher’s exact tests were used to analyze statistical differences in the laying success estimates yielded from the two methods.

2.1.3.2 Correction factors

Mark-recapture models are used to calculate the probability of survival (θ) of a marked individual between recapture events, correcting for its probability of detection (p). Both probabilities are binomial (0 or 1). Due to the nature of the double blind surveys conducted, mark-recapture analysis is appropriate to calculate probability of detection between surveys (Jones et al., 2003; Whitehead et al., 2014). This detection probability was used as a correction factor for posterior analyses.

To estimate the probability of detecting an egg or chick present in an occupied burrow (p), presence/absence data from the standard and double-blind surveys was used to create a set of encounter histories (Figure 3) that could be analyzed using mark–recapture modelling (Jones et al., 2003; Whitehead et al., 2014). Any encounter histories without any presence (1), as well as histories with "Unsolved" checks, were eliminated from the analysis. To increase the number of estimable parameters, an 'initial release' occasion to represent 'laying' to the recapture record (1) was added at the beginning of every history, giving mark recapture histories composed of five occasions, always starting with 1 (Whitehead et al., 2014). This assumes that eggs and near-fledging chicks do not move between burrows, which is biologically reasonable due to the independent nature of puffin burrows (Whitehead et al., 2014). Open population Cormack-Jolly-Seber (CJS) models, included in MARK 6.0 (White & Burnham, 1999) were used to estimate detection probabilities for eggs and chicks. Goodness-of-fit of this model to the observed data was assessed with the variance inflation factor ($\hat{c}$) and the bootstrapped deviation to compare it with the observed deviation using the parametric bootstrap procedure in MARK, set at 1000 iterations (White, 2002; White and Burnham, 1999). AIC values were then corrected for overdispersion (Q), and small sample size (QAICc). Double-blind surveys were conducted successively (less than an hour apart), therefore survival parameters (θ) for both eggs and chicks were fixed at 1 between each double-blind survey in all models containing a time-dependent effect on survival (t). This assumes a 100% chance of survival in surveys conducted less than 1 hour apart, which is likely.

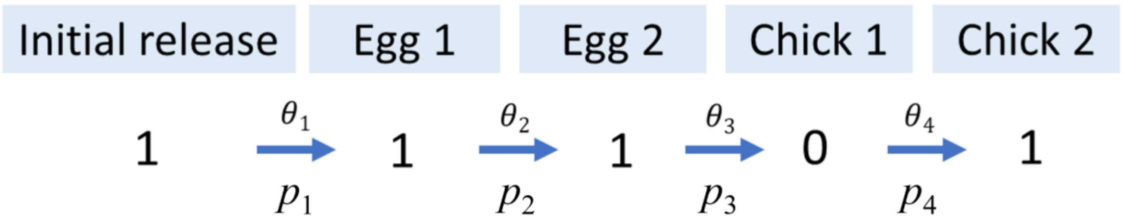


Figure 3. Mark-recapture model sample scheme. Data of each burrow were converted to a 5-occasion history (0 - 1), 1 being occupied, 0 being empty. θ represents survival rate between occasions, and p probability of detection. Probability assessments involved the estimate of p_2 (probability of detection in egg stage), and p_4 (probability of detection in chick stage).

Due to sample size constraints described in the results section, a different approach was used to estimate the correction factor for handgrubbing. The burrow occupancy rates yielded by the two observers per method were merged into one assessment. These merged occupancy rates considered that a burrow was occupied if one of the observers detected breeding contents. This resulted in histories that reflected the maximum occupancy rate estimated by each method. These maximum detection histories were then compared for differences in occupancy rates produced between methods. This method is useful when the true status of each burrow cannot be assessed (Lavers et al., 2019).

2.2 Breeding success estimates across Newfoundland

2.2.1 Study sites

Permanent plots were placed on five islands around the island of Newfoundland. The different colonies selected were aggregated in three different regions: Bakeapple and Puffin Island, part of the Little Fogo archipelago; Elliston Point, Bonavista Peninsula; and Great and Gull Island in Witless Bay Ecological Reserve (Table 7, Figure 4). Study sites were selected to cover different colony sizes: large colonies containing hundreds of thousands of breeding pairs in Witless Bay, medium sized colonies containing thousands in Little Fogo, and a small colony in Bonavista containing hundreds of breeders (Table 7). Study sites were also selected to cover a latitudinal gradient along Newfoundland eastern coast, to capture possible latitudinal variations in breeding (Figure 4).

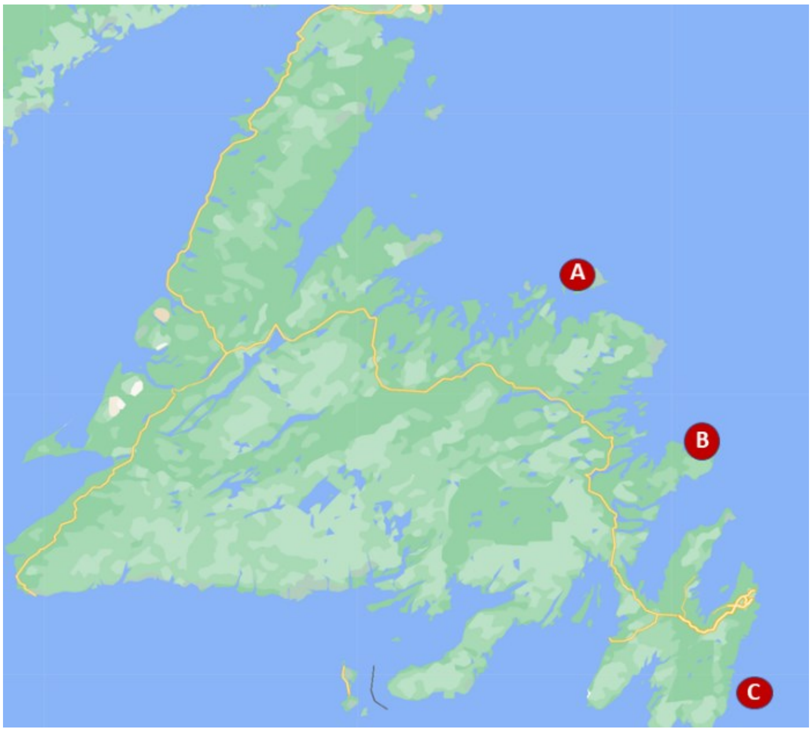


Figure 4. Location of colonies considered for breeding success monitoring. A: Little Fogo Islands, B: Bonavista colonies, C: Witless Bay colonies. Map taken from Google maps and modified.

Table 7. Colonies included in this study with population numbers (breeding pairs), year of their last assessment, and percentage of the total Newfoundland population that each colony represents.

Region	Colony name (Coordinates WGS84)	Estimated breeding pairs (Year assessed)	Proportion of the Total Population (%)
Little Fogo Islands	Bakeapple Island (49.8155, -54.1127)	2 072 ± 776[1] (2014)	0.23 ± 0.09
	Puffin Island (49.8155, -54.1127)	4 125 ± 1 832[1] (2014)	0.46 ± 0.21
Bonavista Peninsula	Elliston Point (48.6282, -53.0195)	400[2] (1985)	0.01
Witless Bay	Great Island (47.1871, -52.8136)	174 491 ± 53 864[3] (2011)	19.44 ± 3.01
	Gull Island (47.2625, -52.7733)	118 401 ± 8 987[1] (2012)	13.19 ± 2.02

1: Wilhelm, 2017; 2: Cairns et al., 1986; 3: Wilhelm et al., 2015.

2.2.2 Plot setting

Permanent plots were set in the year 2021 to assess fledging success in Atlantic puffin colonies. Evidence in other auk species with similar biology show that between 6-8 plots are needed to reveal accurate trends in the population in large colonies like the one in Witless Bay (Rodway & Lemon, 2011), but time and personnel constraints, as well as habitat disturbance concerns, were limiting factors during plot setting. These constraints were addressed by establishing permanent plots in high density areas, since evidence shows that such areas often contain and capture the majority of the variance in fledging success and burrow laying success (Rodway & Lemon, 2011). Any part of the colony was considered as a high-density area when the density of burrows was higher than the island average, based on previously assessed colony specific parameters (Wilhelm, unpublished data; summary in Table 8). Newfoundland Atlantic puffins are highly sensitive to disturbance (Rodway et al., 1996), so potential high density areas were searched in colony sections where access and research transit would cause as little disturbance as possible, while maximizing personal safety. Such areas were selected after full colony surveys in small colonies (Bakeapple and Puffin islands, and Elliston Point). In large colonies (Great and Gull islands), plots were distributed across the colony, to cover as much variation as constraints allowed (Figure 5).

Table 8. Permanent plot summary per colony. Indicated number of plots, area that these plots cover, number of marked burrows per plot, colony burrow density threshold for an area to be considered as high density, and average burrow density on the plot. No previous burrow density data available for Elliston Point.

Colony name	Plot	Plot area (m^2)	Tagged burrows	Density threshold (burrow/m^2)
Great Island	A	12	25	0.92[1]
	B	12	25	0.92[1]
	C	12	25	0.92[1]
	D	12	25	0.92[1]
Gull Island	A	27	50	1.36[2]
	B	15	50	1.36[2]
Elliston Point	A	18	50	NA
Bakeapple Island	A	30	25	0.65 [2]
	B	18	25	0.65 [2]
Puffin Island	A	21	50	0.88 [2]

1: Wilhelm et al., 2015; 2: Wilhelm, unpublished data.

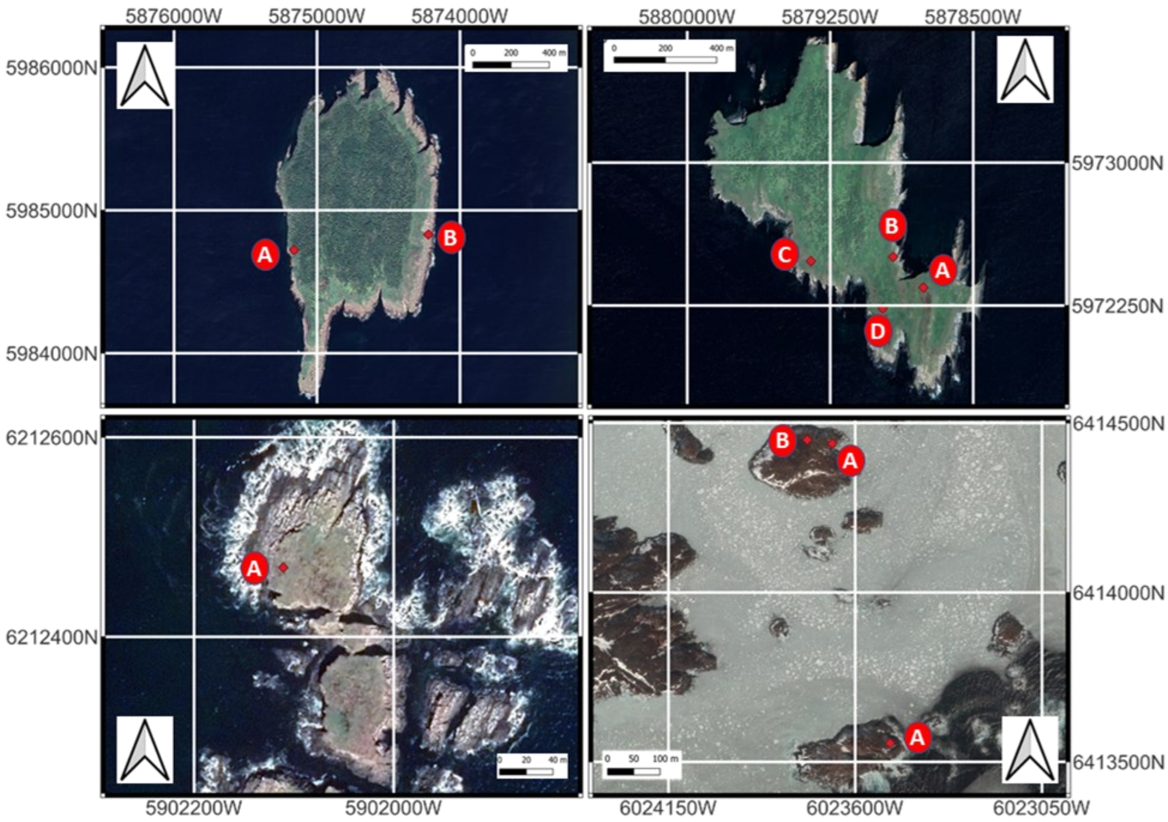


Figure 5. Locations of permanent plots selected for this study (red circles A to D). Gull Island, top left; Great Island, top right; Elliston Point, bottom left. Bakeapple Island, bottom right, (top island); Puffin Island, bottom right (bottom island), with one plot. Images obtained from Google Satellite imagery.

When a suitable area was found, plots were established by deploying a series of 3 by 3-meter grids fixed to the ground with spikes. A burrow was defined as any excavation within puffin habitat longer than half an arm's reach (Rodway et al., 1998). Any excavation within puffin habitat that was shorter, or was deemed as non-suitable to host a pair (e.g., entrance leading to an excavation with large roof sections missing), was classified as a hole and excluded. Since Atlantic puffin burrows are known to have two chambers (Harris & Wanless, 2011), single entrances that diverged in two internal sections were considered as one burrow. If an entrance was found at the limit of the grid, the contents were checked only if the burrow was excavated towards the grid.

Every excavation within the limits of the grid was checked using a burrowscope EMS2021 Gopher Tortoise Camera System with infrared detection (Environmental Management Services, Canton, Georgia, USA). Only nests with the presence of an egg or a chick were included in this study. Once a burrow was confirmed to contain an egg, it was marked with a 30 cm PVC peg secured with an engraved aluminum tag and a unique number identifier. After all the excavations within the plot were checked, the grid was set in an adjacent area (Figure 6), and the process was repeated until the desired number of active burrows was achieved (Table 8). This systematic process prevented double counting and allowed for a calculation of an estimated density and total covered area during the process (See summary in Table S1).

Figure 6. Plot setting scheme using 3x3 m grids. Areas in pale blue are checked areas, navy blue areas are newly placed grids. Note that newly placed grids are adjacent to previously checked ones, preventing double counting.

2.2.3 Visit regime

Due to the extensive geographic range considered in this monitoring project, a visit regime involving frequent monitoring within the season was not possible. In addition, frequent burrow assessment can lead to biased estimates due to researcher's disturbance (Baillie, 2001; Rodway et al., 1998). The visit regime for this study was designed to obtain reliable information on breeding parameters and phenology of the different colonies with these considerations in mind. When possible, each colony was visited three times during the seasons of 2021 and 2022, mirroring the visit scheme followed in the Atlantic puffin monitoring program in Iceland (Hansen, 2021), which also covers a large geographic area. Every burrow included in the study was burrowscope assessed, with its contents classified into two different categories: a) Active, if the burrow had evidence of active breeding (egg or chick present); b) Inactive, if the burrow had no evidence of breeding (only an adult, or empty).

Timing for the visits was based on bibliographical references (Rodway et al., 1998), personal observations, and expected duration of incubation (42 days; Harris & Wanless, 2011) and rearing time (42 days; Grant & Nettleship, 1971; Rodway et al., 1998). The first visit (laying success visit hereafter) was conducted midway into incubation, when only eggs were present in the colony. This maximized the probability that burrows that are found empty are non-active for the season, instead of late breeders. After this visit, a proportion of experimental burrows that contained an egg was obtained. The second visit (phenology visit hereafter) was conducted mid-way into hatching of the eggs, when a mix of hatched and unhatched eggs was expected to be found in the colony. This allowed for the estimation of the mean hatching date in the plots and colonies (see "Breeding parameter calculation"). The third visit (productivity visit hereafter) was planned midway into the brood rearing period (63-70 days after expected laying, 21-28 days after expected hatching), when only grown chicks were predominantly found in the colony, but before first fledglings occurred. This prevented the classification of a burrow as "Inactive" due to the fledging of a chick, rather than a failed breeding attempt. A proportion of burrows in the plot that contained chicks was obtained with this visit.

The timing of visits was different for each colony. Previous information suggests that the breeding cycle in northern colonies is delayed when compared to southern colonies, which was later confirmed. For this reason, the visit regime had to be adapted to the specific timing of each colony (see summary in Table S2).

In 2022, burrow density in the plot area was reassessed. Between two to five 3 by 3-meter grids were deployed (Table S1) in an area less than 10 m away from the plot center, and of similar terrain conditions. Every burrow within the grid was assessed and classified as previously described.

2.2.4 Present breeding parameter estimates

2.2.4.1 Parameter calculation

Burrow density (burrows / m^2) in the plot area was calculated by dividing the number of burrows found within grids by the area (m^2) covered. General laying success was obtained by dividing the number of active burrows within the grids by the number of burrows in all the area covered (Active burrows / Total number of burrows in the grid).

The information obtained from tagged burrows during the different visits was used to calculate the following breeding parameters: Plot laying success (LS), the proportion of tagged burrows containing an egg (Burrows containing eggs / Total number of tagged burrows); Plot fledging success (FS), the proportion of plot burrows containing an egg during the laying success visit where a chick was also found during the productivity visit (Chicks likely to fledge / Total number of eggs in tagged burrows); and plot productivity (P), the proportion of tagged burrows that fledged a chick (LS x FS, or Chicks likely to fledge / Total number of tagged burrows). The calculation of these parameters relied on the assumption that any chick found alive in the productivity visit would be a successful fledgling. There is evidence that fledging success and productivity are usually determined early after hatching (Barrett, 2015; Diamond, 2021; Fayet et al., 2017; Hansen, 2021; Wilhelm et al., 2021), so there is support for the reliability of such estimates. Any burrows where were lost pegs (within or between years) were excluded from the analyses.

Mean hatching date of each location was also estimated. Traditionally, this parameter is calculated by constant direct observation (Baillie & Jones, 2003; Nettleship, 1972; Rodway et al., 1998), but time constraints related to the geographic extent of this study precluded a constant monitoring. Mean hatching dates were obtained using the cumulative distribution function of the hatching date frequency. For that, a proportion of hatched and unhatched eggs was obtained during the phenology visit (second visit). Then, the cumulative normal distribution function was used to estimate the mean hatching date, or the specific date where the proportion of eggs and chicks found in the colony is 0.5. In this case, the cumulative normal distribution function calculated consisted of three parameters: day of the year of the visit, the standard deviation of the hatching distribution (in days), and hatching proportion on the assessment date. The proportion of hatched eggs is obtained from the phenology visit. This value was used for every other colony due to lack of colony specific data, and any eggs that did not hatch in posterior assessments were excluded from the analysis (i.e., burrows with eggs found during the third visit). To test the accuracy of this method in estimating a correct mean hatching date, a control plot was set on Gull Island with 60 burrows that were checked every four days to minimize researcher disturbance. Hatch date was determined based on the contents of the burrow at each visit, following a standardized protocol (Kochvar, in prep.). If a burrow contained a chick during a given visit, and contained an egg during the previous visit, the hatch date was assumed to be the midpoint between the visits (Baillie & Jones, 2004). If the chick still appeared to be wet, the hatch date was designated as the date of the current visit. If a burrow contained an egg, the burrow was re-checked after 3-5 days. However, if the chick appeared to be hatching (cracks or holes in the egg), the hatch date was designated as the date after the current visit, and the burrow was re-checked during the next visit to confirm there was a chick present. Any egg with a hatching date precision larger than five days was excluded from the analysis. To assess the precision of this method, the resulting observed plot mean hatching date was compared to the estimated mean hatching date for every proportion of eggs and chicks found in each visit to the control plot.

2.2.4.2 Historical data

The breeding parameters obtained during this study were compared to available historical data found in the bibliography (Table S3). Historical data were collected by handgrubbing, therefore was corrected to make it comparable with current burrowscope data (see “Correction factors” for more information).

2.2.4.3 Statistical analysis

Binomial regression models (link = logit) were used to evaluate the effect of colony and year on the different breeding parameters. Binomial regressions were conducted in R (R Core Team, 2021), using the package “lme4” (Bates et al., 2015). Model overdispersion was assessed with the “testDispersion” function, included in the package “DHARMA” (Harting, 2022). All pairwise comparisons were corrected for False Discovery Rate (type I error) using the function “p.adjust” included in R base package.

2.2.5 Chick wing length-mass ratio comparison between colonies

To investigate potential chick differences at the extremes of the breeding range in Newfoundland, I compared chick wing length-mass ratios between Gull and Puffin islands. During the 2022 breeding season we collected the wing chord and mass of 19 chicks from Puffin Island, and 22 from Gull Island. Mass was obtained with a 600 g Pesola to the nearest 5 g, wing chord length was measured with a flat ruler to the nearest 1 mm. Normality of the ratio data was tested using both a visual approach (Q-Q plot) and Shapiro-Wilks tests (significance at 0.05). Linear models were used to find statistical differences in the ratio between colonies.

3. Results

A total of 3,261 burrows were checked for the different parts of the present study. This took more than 1,000 human hours, in teams ranging from 1 to 5 people, spread over the course of the summers of 2021 and 2022.

3.1 Method comparison

3.1.1 Burrow depth analysis

Incubation chambers were 55 cm deep on average (N = 399, range = 5 - 115 cm, SD = 19.92 cm). With consideration that the average arm length in North America is 65 ± 4 cm (Claire et al., 1989), 32 % of the incubation chambers would be out of reach (Fig. 7). This excludes other limiting factors for this method, like sharp angles or narrow burrows, which can prevent the researcher's from reaching the incubation chamber.

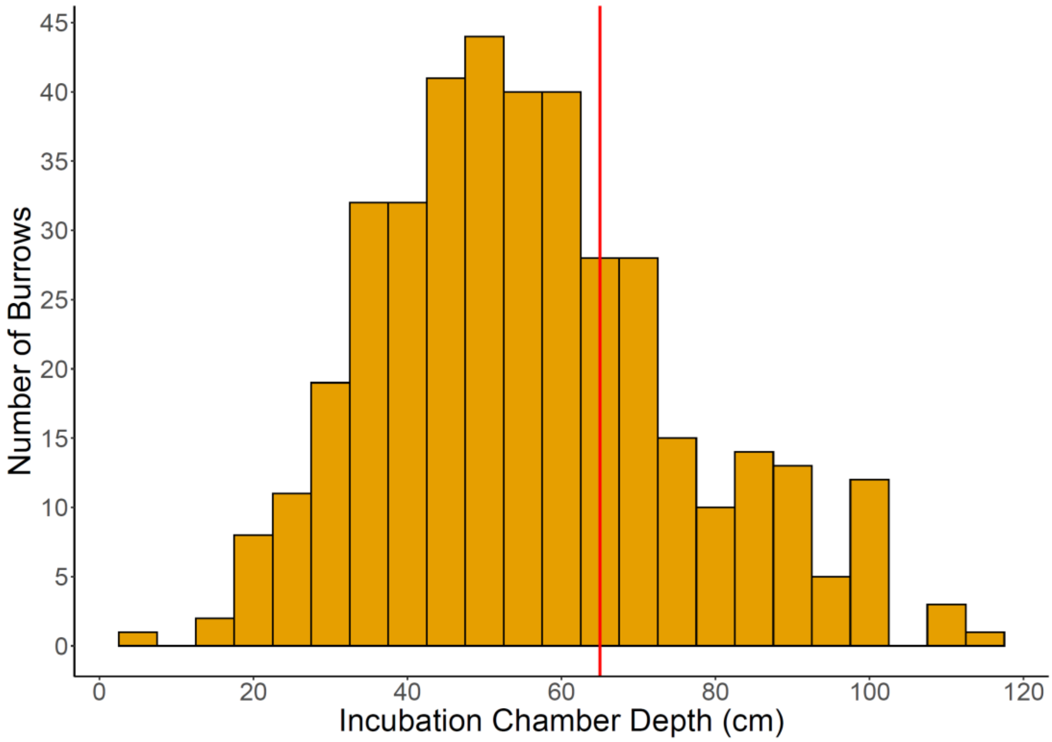


Figure 7. Incubation chamber depth distribution in Newfoundland puffin burrows (N = 399). Vertical red line represents the average arm's reach (65 cm).

3.1.2 Inter-rater and method reliability tests

Inter-rater values showed a large discrepancy between methods. I found significant differences in laying success (Active burrows / Total number of burrows) estimates between burrowscope observers when assessing the presence of eggs and chicks (Fisher's exact test, $p < 0.001$). Cohen's kappa values were extremely low for burrowscope evaluation. The estimate for the egg assessment was 0.51, validating 25.5 % of the data collected, while the estimate for chick assessment was 0.28, validating only a 7.9 % of the data (Table 9). These values show an extreme discrepancy between the two observers. Hand-grubbing showed an improvement in inter-rater agreement. Cohen's kappa was 0.75 for the egg and 0.69 for the chick assessment, validating 55.9 % and 48.8 % of the data respectively. Still, detection results were significantly different between observers (Fisher's exact test, $p < 0.001$).

Table 9. Cohen’s kappa and Coefficient of Determination (COD) values for both inter-rater and inter-method comparisons.

		Inter-rater (SE)		Inter-method (SE)	
Method	**Check**	**Cohen's Kappa**	**COD (%)**	**Cohen's Kappa**	**COD (%)**
Burrowscope	Egg	0.51 (0.12)	25.52 (1.50)	0.66 (0.12)	43.56 (1.4)
	Chick	0.28 (0.13)	7.93 (1.17)	0.35 (0.11)	12.60 (1.58)
Handgrubbing	Egg	0.75 (0.08)	55.93 (0.67)	-	-
	Chick	0.69 (0.09)	48.80 (0.76)	-	-

When comparing between methods (Table 10. Handgrubbing and burrowscope maximum occupancy rates), Laying success estimates were significantly higher when obtained by burrowscope (Fisher’s exact test, p < 0.001). Cohen’s kappa was 0.66 for egg, and 0.35 for chick assessments, making comparable 43.56 % and 12.6 % of the data between the two methods respectively (Table 9).

Table 10. Handgrubbing and burrowscope maximum occupancy rates obtained from the combination of the detection histories of two observers per method. Any completely unsolved burrow was excluded from the calculations.

Assessment	Method	Laying success (SE)
Egg	Burrowscope	0.729 (0.064)
	Handgrubbing	0.618 (0.083)
Chick	Burrowscope	0.510 (0.072)
	Handgrubbing	0.424 (0.084)

3.2 Mark recapture analyses

3.2.1 Burrowscope detection

The histories of 36 burrows, out of 48 initially selected, were used to estimate probabilities of burrowscope detection. The rest were discarded for being inactive. Bootstrapped iteration on the global model showed a strong degree of overdispersion ($\hat{c}$=2.506), so models were adjusted. Following QAICc selection criteria (Akaike, 1985), the best supported model included survival probability as time dependent, and probability of detection as time independent (Table 11). The next candidate showed strong statistical support, with both probability of detection and survival as time dependent. The third candidate model was also within 2 QAICc from the best supported model, including survival probability as time dependent, and probability of detection as time independent with fixed survival between consecutive observations. The rest of the candidate models received close to no support. The output of the best model gave a detection probability of 0.865 (95% CI = 0.679-0.955)

3.2.2 Handgrubbing detection

The mark-recapture histories of 36 burrows were unusable due to the presence of unsolved detection occasions, leaving only 12 histories available for the analyses. The parametric bootstrap showed an extreme value of overdispersion ($\hat{c}$=3.798), probably due to the lack of sample size. This prevented any posterior analysis.

Table 11. Summary of model selection results for probability of survival, Phi, between sampling periods and detection, p, for Atlantic puffin eggs and chicks. Models are ranked by Delta QAICc, the difference between each model and the best model, based on Akaike's information criterion (Akaike, 1985), corrected for over-dispersion (Q) and small sample size (QAICc). Models are defined as follows: (.) parameter constant; (t) parameter varies with time of sampling; # denotes models for which the survival rates between surveys carried out on the same day were fixed at the constant value 1.0.

Model	QAICc	Delta QAICc	QAICc Weight	Model Likelihood	Parameters	Qdeviance
phi(t)p(.)	62.3426	0	0.29470	1	5	9.2536
phi(t)p(t)#	62.3959	0.0533	0.28695	0.9737	6	7.0477
phi(t)p(.)#	62.8513	0.5087	0.22852	0.7754	3	14.1423
phi(.)p(.)	64.5952	2.2526	0.09555	0.3242	2	18.0100
phi(t)p(t)	64.7022	2.3596	0.09057	0.3073	7	7.0467
phi(.)p(t)	71.0990	8.7564	0.00370	0.0126	5	18.0100

3.2.3 Maximum detection rates

Due to the impossibility of obtaining specific detection rates for hand-grubbing, I compared the differences in laying success estimates by the two methods. When detecting eggs, burrowscope laying success estimate was 0.729, and 0.618 for handgrubbing, resulting in an 11.1 % difference. (Table 10. Handgrubbing and burrowscope maximum occupancy rates). For chick detection, the burrowscope laying success estimate was 0.51, and 0.424 for handgrubbing, an 8.6 % difference. The average difference in laying success between the two assessments was 9.85 %.

The handgrubbing specific detection rate was then calculated as the difference between the burrowscope detection and the discrepancy between the two methods. Therefore, the detection rate for this method was 0.754 for eggs and 0.779 for chicks. This calculation method did not allow for 95% confidence intervals. The correction factor for handgrubbing-burrowscope estimates was obtained dividing these values by the burrowscope detection rate. This resulted in a correction factor of 1.147 for egg detection and 1.11 for chick detection.

3.3 Present breeding success estimates

In only three cases the visit regime was compromised: a) in 2021, due to logistic complications, Elliston Point could not be assessed during the first two visits; b) during the 2021 phenology visit to Bakeapple and Puffin islands, due to lack of previous breeding data of the region, only grown chicks were found, instead of a proportion of hatched and unhatched eggs; c) in 2022 at Bakeapple Island, breeding at this colony was disrupted due to the presence of an Arctic fox (*Vulpes lagopus*) that led to an almost total breeding failure (Zabala et al., in prep.).

3.3.1 Burrow density and general laying success

Burrow density at plot level was variable between colonies. The average burrow density in the plots (Total burrows / m^2) was 3.369 (SE = 0.33, range 1.51 to 4.33), although it was highly variable between years and colonies (Table 12). Holes (entrances that did not lead to burrows) were not extensively present with 4.39 % of the entrances checked being too short to host a puffin nest (N = 637, range 8 % at Puffin Island to 2.94 % at Great Island). The average general burrow laying success (Active burrows / Total burrows checked) was 0.626 (SE = 0.06, range = 0.309 – 0.716). This value was high in every colony checked (Table 12). Larger colonies (Gull and Great islands) tended to have lower general laying success than smaller ones, but low sample size precluded a quantitative assessment. The only extreme value was found on Bakeapple Island in 2022. This low burrow laying success was due to the presence of an Arctic fox in the colony, which disturbed the normal breeding in the colony. Dozens of excavated Leach's Storm-Petrel burrows and puffin chick carcasses were found around the island, but no direct impact was observed on puffin burrows (e.g., signs of excavation) or direct predation on adult puffins. This evidence suggests that the presence of the fox prevented puffin breeding through indirect mechanisms (e.g., increased individual stress) rather than through active adult predation (Zabala Belenguer and Bitton, in preparation). Therefore, this data point was excluded from any posterior analyses.

Inter-annual comparisons in general laying success were only possible between Great, Gull, and Puffin islands. Binomial logistic results showed a significantly higher general laying success on Gull Island in 2022 ($Z(1) = 2.838$, $p = 0.005$; Appendix S5). Models did not show over-dispersion (Great Island, $\hat{c} = 1.0019$, $p = 0.976$; Gull Island, $\hat{c} = 1.0079$, $p = 0.856$; Puffin Island, $\hat{c} = 1.0074$, $p = 0.952$). Binomial logistic regressions results showed no statistical differences in general laying success probabilities in both years (Appendix tables S 6 & S 7). Neither model suffered from over-dispersion (Year 2021, $\hat{c} = 1.0015$ $p = 0.952$; Year 2022, $\hat{c} = 1.005$ $p = 0.936$).

Table 12. Burrow density general laying success (Active burrows / Total burrows checked) values obtained for 2021-2022. * Bakeapple Island data for 2022 is non-comparable due to the presence of an Arctic fox in the colony

Colony	Year	Surveyed area (m^2)	Burrows checked	Burrow density (Burrows/m^2) (SE)	Laying success (Active /Total) (SE)
Great Island	2021	144	168	1.17 (0.11)	0.623 (0.057)
	2022	216	340	1.57 (0.06)	0.647 (0.033)
Gull Island	2021	126	173	1.37 (0.12)	0.566 (0.06)
	2022	90	102	1.13 (0.13)	0.618 (0.057)
Puffin Island	2021	63	74	1.17 (0.16)	0.681 (0.035)
	2022	45	41	0.91 (0.12)	0.715 (0.057)
Bakeapple Island	2021	81	69	0.85 (0.14)	0.716 (0.05)
	2022*	72	50	0.69 (0.11)	0.309 (0.105)
Elliston Point	2021	54	78	1.44 (0.17)	0.678 (0.054)
	2022	81	104	1.29 (0.1)	0.711 (0.098)
Average	**-**	**93.6**	**120**	**1.16 (0.11)**	**0.626 (0.06)**

3.3.2 Plot laying success, fledging success, and productivity

Plot specific laying success could only be calculated for 2022. Average laying success (Active burrows / Total burrows in the plot) was 0.819 (SE = 0.04, range = 0.388 - 0.948). Overall, values were very high in every colony assessed, giving evidence that a large proportion of pairs that bred in 2021 also bred the following year (Table 13). Binomial logistic regressions were performed to estimate the effect of colony on the probability that a burrow would be occupied in the plot (Table 14). Bakeapple Island showed significantly lower plot laying success to every other colony (Great Island, Z (1) = -6.184, p = <0.001; Gull Island, Z (1) = -5.943, 0 = <0.001; Puffin Island, Z (1) = 3.951, p = <0.001). There were no differences between the rest of the colonies. This is explained again by the disruptive presence of the Arctic fox in the colony. The proportion of egg laying at Bakeapple showed a 41.59 % decrease when compared to the neighbour Puffin Island in that same year. Interestingly, none of the study burrows showed any sign of active predation (i.e., digging evidence). This suggests that the sole presence of a predator on the island can have a very large impact on puffin egg laying.

Table 13. Breeding parameter estimates obtained during the 2021-2022 breeding seasons through burrowscope assessment: Plot laying success (Plot active burrows / Total plot burrows), Fledging success (Plot chicks likely to fledge / Total number of eggs found in the plot), and Productivity (Plot chicks likely to fledge / Total plot burrows) Average weighted by colony size relative to the total Newfoundland population. Note that plot burrow laying success and plot productivity could only be calculated for 2022.

Colony name	Year	Plot Laying success (SE)	Fledging Success (SE)	Plot Productivity (SE)
Great Island	2021	NA	0.812 (0.037)	NA
	2022	0.948 (0.025)	0.761 (0.044)	0.721 (0.045)
Gull Island	2021	NA	0.736 (0.048)	NA
	2022	0.931 (0.027)	0.864 (0.058)	0.804 (0.043)
Puffin Island	2021	NA	0.90 (0.042)	NA
	2022	0.933 (0.042)	0.905 (0.058)	0.844 (0.055)
Bakeapple Island	2021	NA	0.888 (0.03)	NA
	2022	0.388 (0.069)	0.053 (0.032)	0.02 (0.02)
Elliston Point	2021	NA	NA	NA
	2022	0.896 (0.04)	0.814 (0.067)	0.729 (0.064)
Average		**0.937 (0.04)**	**0.792 (0.046)**	**0.750 (0.045)**

Table 14. Binomial model results for plot burrow laying success differences (Plot active burrows / Total plot burrows) between colonies during the 2022 breeding season (significance level set at 0.05, corrected for Type I error). Colony included as fixed factor.

	Gull Island		**Bakeapple Island***		**Puffin Island**		**Elliston Point**	
	z	**p**	**z**	**p**	**z**	**p**	**z**	**p**
Great Island	-0.496	0.620	-6.184	**<0.001**	-2.537	0.112	-1.120	0.263
Gull Island	--	--	-5.943	**<0.001**	-2.111	0.352	-0.675	0.500
Bakeapple Island*			--	--	3.951	**<0.001**	4.736	**<0.001**
Puffin Island					--	--	1.267	0.205

Average fledging success (Chicks likely to fledge / Total number of eggs found in the plot) in 2021-2022 was 0.748 (SE = 0.046, range = 0.053-0.905). Overall, fledging success was extremely high, always above 70 %, and it did not vary much between colonies except on Bakeapple Island in 2022 (Table 13). Inter-annual comparisons in fledging success were only possible between Great, Gull, and Puffin islands. Binomial logistic regression results showed that fledging success between 2021 and 2022 was only different on Gull Island (Z(1) = 3.258, p = 0.001) (Table S 5). Models did not show over-dispersion (Great Island, $\hat{c}$ = 1.0011, p = 0.947; Gull Island, $\hat{c}$ = 1.0089, p = 0.849 Puffin Island, $\hat{c}$ = 1.007, p = 0.961).

Binomial logistic regressions only showed significant lower fledging success on Gull Island in comparison with the rest of the colonies in 2021 (Great Island, Z (1) = -3.156, p = 0.002; Bakeapple Island, Z (1) = -2.801, p = 0.005; Puffin Island, Z (1) = -3.115, p = 0.002; Table 15). No differences were found in 2022 (Table 16). Models did not show over-dispersion (Year 2021, $\hat{c}$ = 1.0001, p = 0.985; Year 2022, $\hat{c}$ = 1.009, p = 0.995).

Table 15. Binomial model results for fledging success probability (Plot chicks likely to fledge / Total eggs in the plot) between colonies during the 2021 breeding season (significance level set at 0.05, corrected for Type I error). Elliston Point data not available.

	Gull Island		**Bakeapple Island**		**Puffin Island**	
	z	**p**	**z**	**p**	**z**	**p**
Great Island	-3.156	**0.002**	-0.476	0.634	-0.909	0.364
Gull Island	--	--	-2.801	**0.005**	-3.115	**0.002**
Bakeapple Island			--	--	0.391	0.696

Table 16. Binomial model results for fledging success probability (Plot chicks likely to fledge / Total eggs in the plot) between colonies during the 2022 breeding season (significance level set at 0.05). Bakeapple data not included.

	Gull island		**Puffin Island**		**Elliston Point**	
	z	**p**	**z**	**p**	**z**	**p**
Great Island	1.707	0.088	1.576	0.115	-0.449	0.653
Gull Island	--	--	0.262	0.794	0.995	0.320
Puffin Island			--	--	1.062	0.288

Average plot productivity (Plot chicks likely to fledge / Total plot burrows) was 0.587 (SE = 0.045, range = 0.02-0.844). Plot productivity could only be calculated in 2022, as plot laying success data was only available for that year (Table 13). Binomial logistic regressions were performed to estimate the effect of colony on the probability that a burrow egg would produce a successful fledgling at the end of the breeding season (Table 17). Binomial regression only included colony as a fixed factor. Models did not show over-dispersion ($\hat{c}$ = 1.0009, p = 0.992). No statistically significant differences were found in productivity between colonies. Data from Bakeapple Island was excluded due to the influence of the fox presence (Table 17).

Table 17. Productivity (Plot chicks likely to fledge / Total plot burrows) binomial model results for the 2022 breeding season (significance level set at 0.05). Colony included as fixed factor. *Data from Bakeapple 2022 excluded.

	Gull Island		**Puffin Island**		**Elliston Point**	
	z	**p**	**z**	**p**	**z**	**p**
Great Island	1.422	0.155	1.431	0.153	0.188	0.851
Gull Island	--	--	0.302	0.763	-1.005	0.315
Puffin Island			--	--	-1.12	0.263

3.3.3 Mean hatching date

3.3.3.1 Control plot

Accurate hatching dates and histories were obtained for 48 eggs in the control plot. Fledging success (Eggs hatched / Total burrows) in the plot was 0.80, which suggests a limited effect of researcher disturbance. Normality of the distribution could not be assessed because burrow content evaluation was not conducted every day. Hatching occurred over a 31-day period (Figure 8) from June 17 to July 18 (day of year 168 to 199). Peak hatching occurred between June 26 to July 3 (day of year 177 to 184). The mean hatching date was June 30 (mean = 181.33, rounded to day of the year 181, with a standard deviation of 7.61 days), and the median hatching date was June 29.

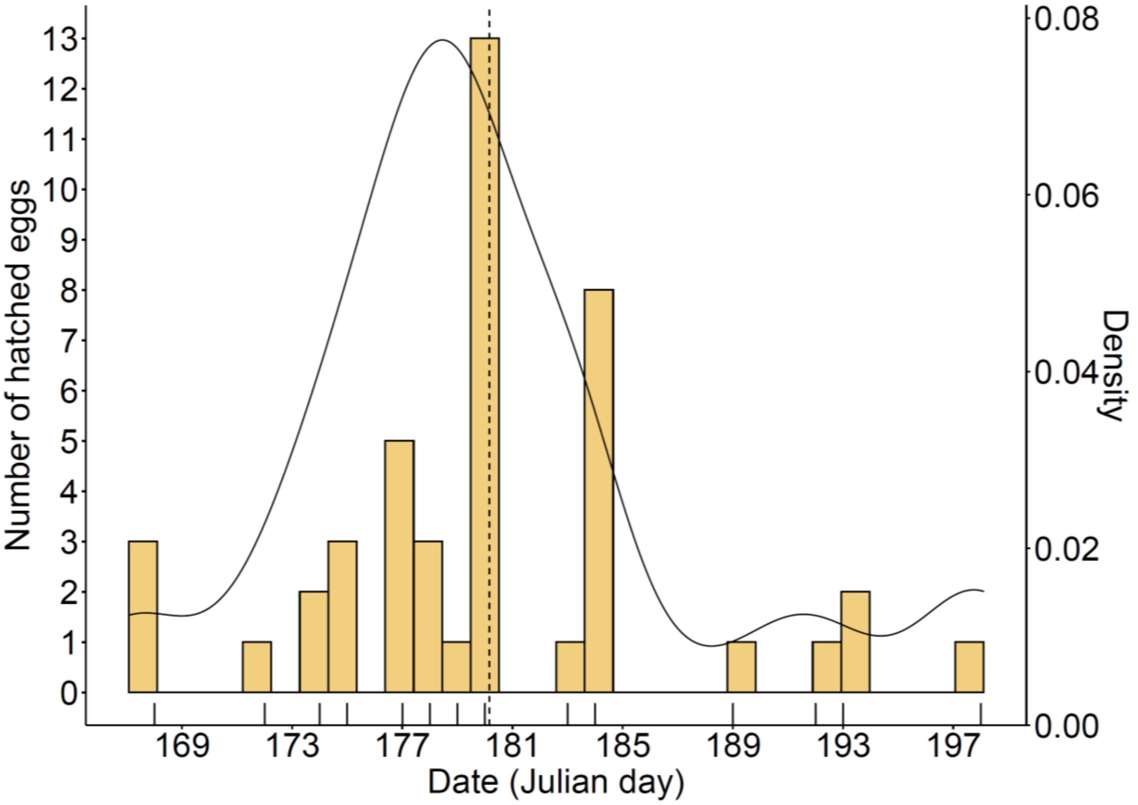


Figure 8. Gull Island control plot hatching distribution histogram (number of hatched eggs in a given day) and density plot (N = 45). Mean hatching date = 181 day of year (dotted line).

I compared the mean hatching dates obtained using the cumulative distribution function method to the real mean hatching date using the hatching proportions obtained during the 14 visits to the control plot. The standard deviation value used in the calculations was 7.61 days, obtained from the control plot. This value is very similar to the standard deviation of 7.51 days found in the 1993 hatching distribution on Great Island (Rodway et al., 1996). Overall, precision was moderately high. Mean hatching dates obtained with the cumulative distribution function corresponded to the real estimate in 3/14 cases (21.43 %), were accurate ± 1 day in 7/14 (50 %), and accurate ± 3 days in 12/14 cases (85.71 %). The average mean hatching date from all values obtained by this method was 182 (SE = 0.732 days). This method showed relatively higher accuracy at very low hatching proportions, and less accuracy when the proportion of chicks was higher than eggs (Table 18).

Table 18. Visit dates, number of chicks and eggs found during the assessment, hatching proportion (Burrows with chicks / Burrows with eggs), and estimated mean hatching dates calculated by the cumulative distribution function to the hatching proportion, compared to the real mean hatching date of 181 day of year date. Tested day of year dates and hatching proportions were obtained from the observed values during plot visits.

Visit date (Day of year)	Chicks	Egg	Hatching proportion	Estimated mean hatching date (Day of year)
168	3	46	0.063	**180**
172	4	45	0.083	**182**
174	6	43	0.125	**183**
175	9	40	0.188	**182**
177	14	35	0.292	**181**
178	17	32	0.354	**181**
179	18	31	0.375	**182**
180	31	18	0.646	**178**
183	32	17	0.667	**180**
184	40	9	0.833	**177**
189	41	8	0.854	**181**
192	42	7	0.875	**184**
193	44	5	0.917	**184**
198	45	4	0.938	**188**

3.3.3.2 Colony mean hatching dates

Mean hatching dates varied between colonies. Great and Gull islands mean hatching dates were similar, around late June-early July (Table 19). Data from both Elliston Point and Puffin Island could only be collected for 2022. Mean hatching dates in these colonies was considerably later, around the third week of July. This suggests a latitudinal phenology delay in northern colonies. No data from Bakeapple could be used for this analysis.

Table 19. Estimated mean hatching dates (MHD) in the different study colonies in 2021-2022. Mean hatching date obtained using the cumulative distribution method, together with the number of eggs and chicks used for the estimate, and the date of the visit.

Colony	Year	Assessment date (Day of the ear)	Eggs	Chicks	Estimated MHD
Great Island	2021	181	34	31	181
	2022	185	33	55	183
Gull Island	2021	190	17	73	183
	2022	182	45	57	181
Puffin Island	2021	-	-	-	-
	2022	208	8	29	202
Elliston Point	2021	-	-	-	-
	2022	207	5	37	198

3.4 Chick wing length-mass ratio comparison between colonies

Chick wing length-weight ratio data followed a normal distribution in both colonies (Gull Island, Shapiro Wilks test, W (25) = 0.967, p = 0.572; Puffin Island, Shapiro Wilks test, W (20) = 0.967, p = 0.572). Wing length-weight ratio showed differences between colonies (Figure 9). Puffin Island chicks had a statistically heavier body weight for a given wing length when compared to those from Gull Island (F (1,43) = 17.32, p < 0.001). The largest chick found on Puffin Island was 370 g and with 126 mm wing length, while the largest from Gull Island was 350 g, and a 106 mm wing length.

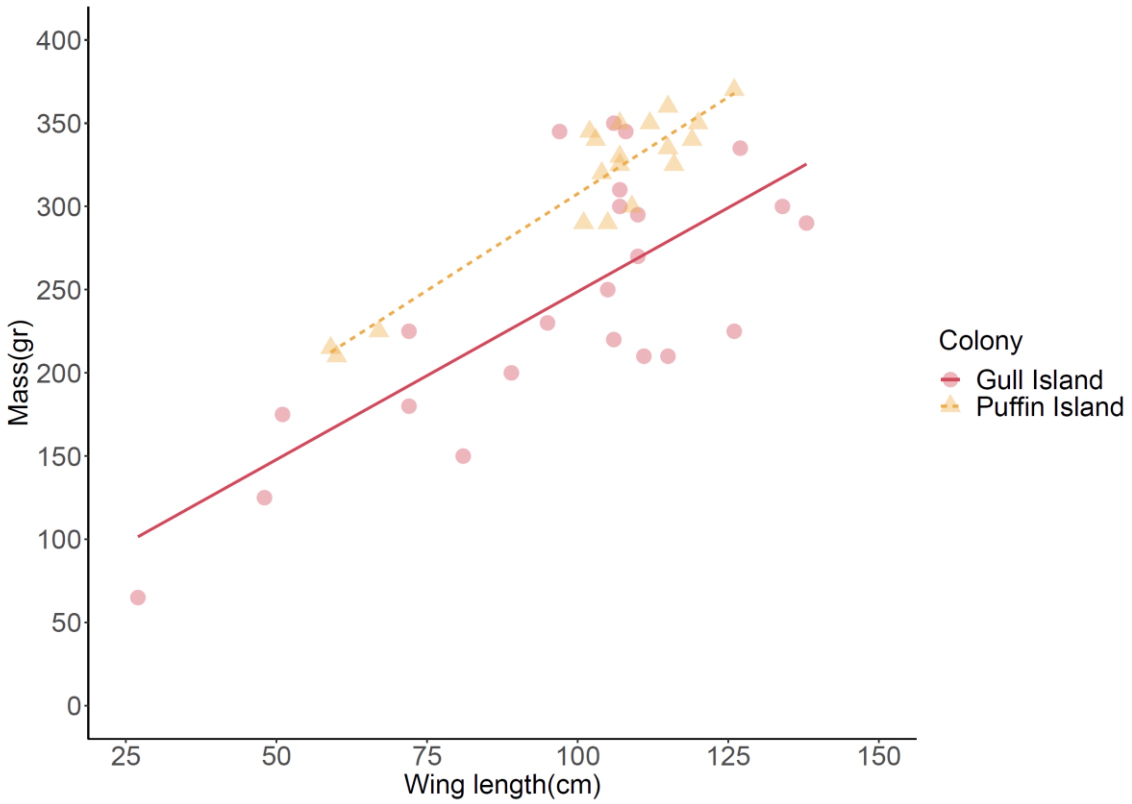


Figure 9. Relationship between mass (g) and wing length (cm) in puffin chicks found in two colonies, Gull Island (Witless Bay), and Puffin Island (Little Fogo Islands).

3.5 Historical breeding success in Newfoundland

The available Newfoundland historical data differed in its laying success definition from the one used in this study: all studies had classified burrows as active even if only an adult was present. As a previous step to validate the comparison between the two data sources, I assessed the percentage of burrows checked with burrowscope that only contained an adult and no egg or chick. Only 1.8 % (N = 2470) of the burrows checked during the years 2021-2022 contained a non-incubating adult, allowing for comparison between the data.

Historical laying success estimates available were obtained by handgrubbing. These estimates were multiplied by the egg correction factor to make them comparable with our estimates (Table 20). After applying correction factors, average historical laying success was 0.786 (SD = 0.138) in Great Island, 0.732 (SD = 0.123) in Gull Island, 0.475 (SD = 0.17) in Bakeapple Island, and 0.678 (0.03) in Puffin Island.

Binomial logistic regressions were performed to assess the effects of year on laying success, both with corrected and uncorrected estimates. Laying success binomial models included year as a fixed factor. None of the laying success models showed evidence of over-dispersion, both for uncorrected (Great Island, $\hat{c}$ = 1.0001, p = 0.928; Gull Island, $\hat{c}$ = 1.001, p = 0.976; Bakeapple Island, $\hat{c}$ = 1.0147, p = 0.744; Puffin Island, $\hat{c}$ = 1.0122, p = 0.84) and corrected data (Great Island, $\hat{c}$ = 0.99891, p = 0.944; Gull Island, $\hat{c}$ = 0.99932, p = 0.968; Bakeapple Island, $\hat{c}$ = 1.0084, p = 0.740; Puffin Island, $\hat{c}$ = 1.0139, p = 0.84).

The results of the binomial regression comparisons showed discrepancies between corrected and uncorrected data. On Great Island, 53.34 % of the total Great Island pairwise comparisons (8 / 15 cases) changed significance when applying the correction factor, 75 % (6 / 8 cases) of which involved data collected by different methods (Table 21). A 20 % (3 /15 cases) of Gull Island comparisons changed significance, 66.7 % (2 / 3 cases) of which also involved data collected by different methods (Table 22). This shows a larger effect of the correction factor when data collection methods differ. Values for both Bakeapple and Puffin islands remained the same before and after applying the correction factor (Tables S 8 & S 9). The only instance where the corrected estimate was encompassed in the uncorrected data standard error interval was on Bakeapple 2014 (Figure 10).

After applying correction factors, laying success in Great Island during the years 2021 and 2022 was significantly lower than every other year on record (Table 21). Similarly, Gull Island laying success was significantly lower in 2021 and 2022 than any other year except 2012 (Table 22). Laying success data from Bakeapple and Puffin islands was too scarce to evaluate trends.

Table 20. Historically available laying success (Eggs present in burrows / Total burrows assessed) data. Number of burrows used to calculate the different parameters and the source are included. Corrected estimate obtained by multiplying the original estimate (handgrubbing) by 1.147, the correction factor for egg detection based on burrowscope estimates.

Colony	Year	Study burrows	Original estimate	Corrected estimate
Great Island	1985[a]	319	0.718 (0.02)	0.824
	1992[b]	179	0.732 (0.01)	0.840
	1993[b]	194	0.871 (0.03)	0.999
	2011[c]	423	0.717 (0.03)	0.822
	2021[d]	168	0.623 (0.06)	-
	2022[d]	340	0.647 (0.03)	-
Gull Island	1984[a]	698	0.715 (0.02)	0.820
	1985[a]	716	0.788 (0.01)	0.904
	2003[a]	897	0.575 (0.02)	0.660
	2012[c]	338	0.722 (0.02)	0.828
	2021[d]	173	0.566 (0.06)	-
	2022[d]	102	0.616 (0.06)	-
Bakeapple Island	2014[c]	62	0.306 (0.05)	0.351
	2021[d]	69	0.716 (0.05)	-
	2022[d]	50	0.359 (0.11)	-
Puffin Island	2014[c]	43	0.558 (0.06)	0.640
	2021[d]	74	0.681 (0.04)	-
	2022[d]	41	0.715 (0.1)	-

a) Robertson, unpub. data; b) Rodway, 1994; c) Wilhelm, unpub. data; d) Present study.

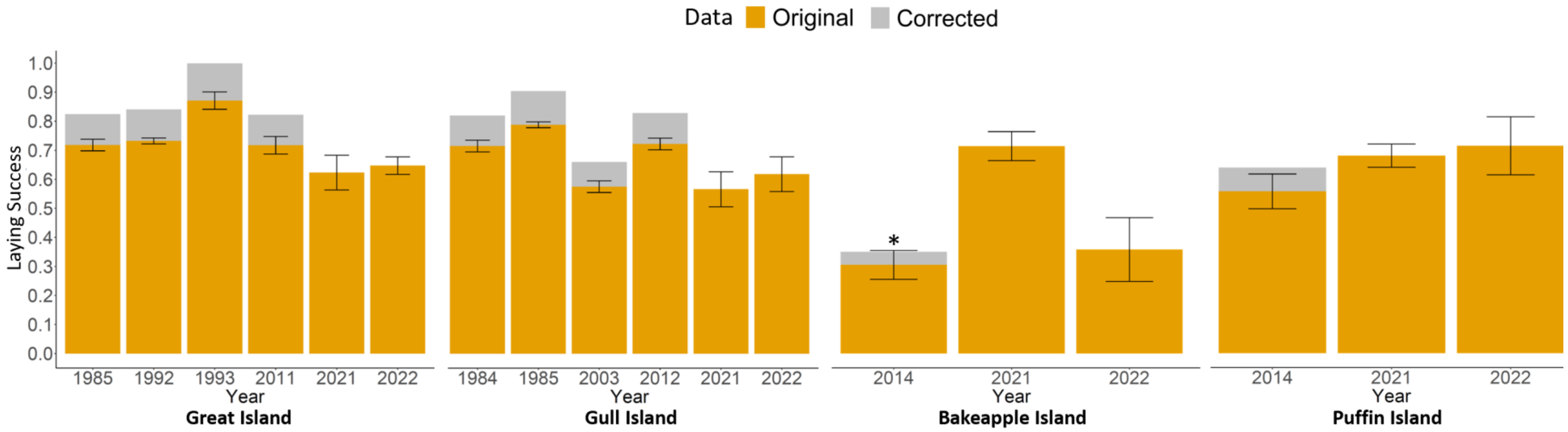


Figure 10. Yearly uncorrected (orange) and corrected (grey) laying success estimates in Great Island, Gull Island, Bakeapple Island, and Puffin Island. Uncorrected estimates present 95 % confidence intervals, which could not be calculated for corrected estimates (see text). Asterix indicates a case where the corrected estimated is encompassed in the 95% confidence interval or the uncorrected estimate.

Table 21. Binomial model results on Great Island corrected and uncorrected laying success estimates. Changes in significance are values are presented in bold (significance set at 0.05, corrected for type I error).

Uncorrected Laying success

	1992		**1993**		**2011**		**2021**		**2022**	
	z	**p**	**z**	**p**	**z**	**p**	**z**	**p**	**z**	**p**
1985	0.334	0.795	3.943	**<0.001**	0.024	0.981	-0.903	**0.500**	-1.790	**0.139**
1992	--	--	3.326	**0.026**	-0.330	0.795	-1.087	**0.462**	-1.827	**0.139**
1993			--	--	-4.054	0.598	-4.301	<0.001	-5.270	<0.001
2011					--	--	-0.966	**0.052**	-1.948	**0.130**
2021							--	--	-0.574	0.708
Corrected Laying success										
1985	0.385	0.876	0.036	**0.972**	-0.062	0.972	-3.614	**<0.001**	-4.915	**<0.001**
1992	--	--	0.036	**0.972**	-0.454	0.876	-3.425	**0.001**	-4.347	**<0.001**
1993			--	--	-4.054	0.641	-4.301	<0.001	-5.270	<0.001
2011					--	--	-0.966	**0.032**	-1.948	**0.037**
2021							--	--	-0.574	0.849

Table 22. Binomial model results on Gull Island for corrected and uncorrected laying success estimates. Changes in significance are values are presented in bold (significance set at 0.05, corrected for type I error).

	Uncorrected Laying success									
	1985		**2003**		**2012**		**2021**		**2022**	
	z	**p**	**z**	**p**	**z**	**p**	**z**	**p**	**z**	**p**
1984	3.160	0.004	-5.721	<0.001	0.234	0.831	-3.728	0.001	-1.996	**0.063**
1995	--	--	-8.869	<0.001	-2.351	**0.032**	-5.845	<0.001	-3.724	0.002
2003			--	--	4.683	<0.001	-0.214	**0.831**	0.821	0.475
2012					--	--	-3.508	<0.001	-2.000	0.0463
2021							--	--	0.832	0.475
	Corrected Laying success									
1984	3.24	0.002	-7.019	<0.001	0.826	0.409	-6.833	<0.001	-4.567	**<0.001**
1995	--	--	-9.912	<0.001	-1.832	**0.084**	-9.045	<0.001	-6.511	<0.001
2003			--	--	6.065	<0.001	-2.344	**0.026**	-0.852	0.409
2012					--	--	-6.523	<0.001	-4.682	<0.001
2021							--	--	0.832	0.409

Historical fledging success values were also estimated by handgrubbing. These estimates were multiplied by the chick correction factor to make them comparable with our estimates (Table 23). Historical average fledging success was 0.732 (SD = 0.493) on Great Island, and 0.746 (SD = 0.092) on Gull Island.

Fledging success binomial models from corrected and uncorrected data were compared. None of the laying success models showed evidence of over-dispersion, both for uncorrected (Great Island, $\hat{c}$=1.0001, p = 0.952; Gull Island, $\hat{c}$= 1.013, p = 0.88; and corrected data (Great Island, $\hat{c}$=1.0016, p = 0.96; Gull Island, $\hat{c}$= 1.0128, p = 0.912).

Table 23. Available historical fledging success (Chicks ready to fledge / Total number of eggs) data. Number of burrows used to calculate the different parameters and the source are included. Corrected estimate obtained by multiplying the original estimate by 1.11, the correction factor for chick detection established in this study.

Colony	Year	Burrows	Original estimate	Corrected estimate
Great Island	1992[a]	179	0.552 (0.037)	0.613
	1993[a]	384	0.685 (0.022)	0.760
	2003[a]	47	0.617 (0.068)	0.685
	2004[a]	57	0.632 (0.061)	0.702
	2005[a]	94	0.606 (0.048)	0.673
	2021	97	0.812 (0.037)	-
	2022	92	0.761 (0.047)	-
Gull Island	1984[b]	43	0.574 (0.073)	0.637
	2019[c]	45	0.844 (0.054)	0.937
	2021	95	0.736 (0.049)	-
	2022	81	0.864 (0.044)	-

a) Robertson, unpub. data; b) Rodway, 1994, c) Wilhelm, 2021.

The results of the binomial regression comparisons showed discrepancies between corrected and uncorrected data. A 23.81 % of the total Great Island pairwise comparisons (5 / 21 cases) changed significance when applying the correction factor, 80 % (4 / 5 cases) of which involved data collected by different methods (Table 24). Only in 1 / 6 cases changed significance when applying correction factors on Gull Island fledging success estimates, also involving different methods (Table 25). The only instances where uncorrected fledging success standard errors encompassed the corrected values were Great Island 2003 and Gull Island 1984 (Fig. 11).

Table 24. Binomial model results on Great Island corrected and uncorrected fledging success data. Changes in significance values are presented in bold (significance set at 0.05, corrected for type I error).

	Not corrected											
	1993		**2003**		**2004**		**2005**		**2021**		**2022**	
	z	**p**	**z**	**p**	**z**	**p**	**z**	**p**	**z**	**p**	**z**	**p**
1992	3.025	**0.011**	0.786	0.504	1.041	0.481	0.845	0.504	4.201	0.001	3.291	**0.010**
1993	-	-	0.937	0.504	0.803	0.504	1.446	0.271	2.480	**0.039**	1.422	0.271
2003			-	-	0.153	0.903	0.122	0.903	2.519	**0.039**	1.758	0.184
2004					-	-	0.309	0.838	2.481	**0.039**	1.682	0.195
2005							-	-	3.117	0.011	2.246	0.065
2021									-	-	0.899	0.369
	Corrected											
1992	3.791	**0.003**	0.971	0.475	1.333	0.426	1.084	0.475	3.502	0.005	2.560	**0.073**
1993	-	-	1.186	0.475	0.956	0.475	1.784	0.268	1.128	**0.475**	0.009	0.993
2003			-	-	0.230	0.904	0.127	0.944	1.770	**0.268**	1.007	0.475
2004					-	-	0.403	0.801	1.599	**0.330**	0.796	0.526
2005							-	-	2.257	0.126	1.365	0.426
2021									-	-	0.899	0.484

Table 25. Binomial model results on Gull Island corrected and uncorrected fledging success data. Changes in significance values are presented in bold (significance set at 0.05, corrected for type I error).

	Not corrected					
	2019		**2021**		**2022**	
	z	**p**	**z**	**p**	**z**	**p**
1984	3.093	0.006	1.811	0.101	3.397	0.006
2019			1.987	**0.094**	0.398	0.690
2021					2.056	0.477
	Corrected					
1984	3.131	0.009	1.291	0.236	2.934	0.009
2019			2.509	**0.024**	1.160	0.246
2021					2.056	0.060

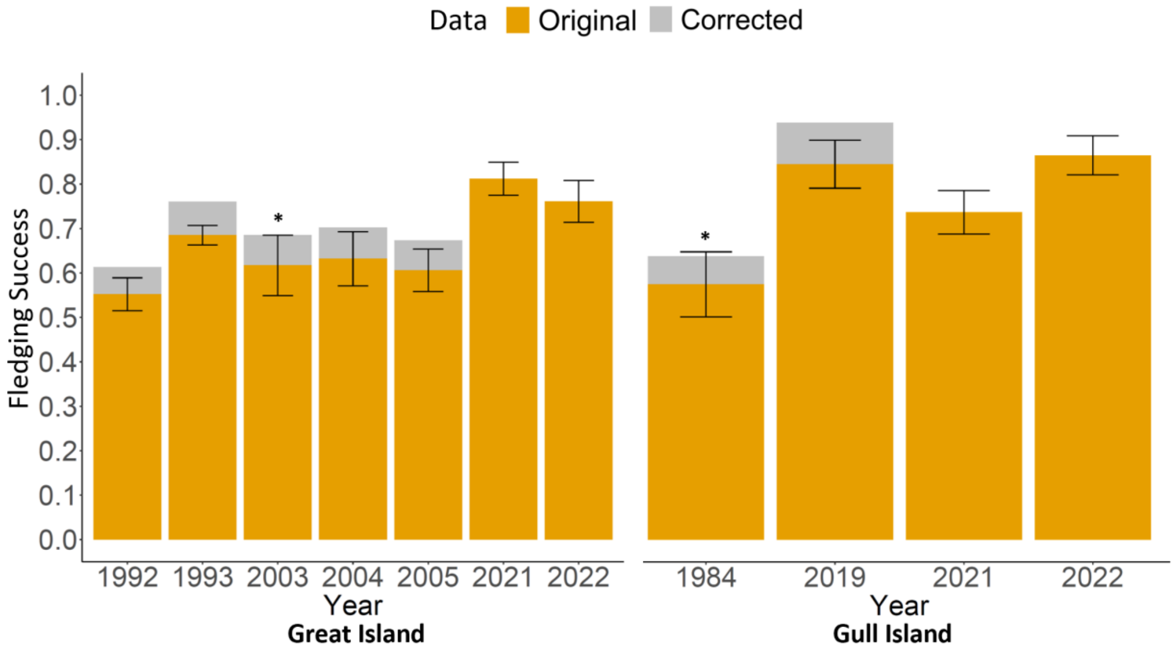


Figure 11. Bar plot with yearly uncorrected (orange) and corrected (grey) fledging success estimates in Great Island, and Gull Island. Uncorrected estimates present 95% confidence intervals, which could not be calculated for corrected estimates (see text). Asterixis indicate cases where the corrected estimated is encompassed in the 95% confidence interval or the uncorrected estimate.

4. Discussion

4.1 Methodological comparison

The interpretation of seabird population trends can be flawed by imprecise estimates due to methodological constraints. Methods that can correct or account for these discrepancies need to be developed and validated (Harris et al., 2020). The present study reveals that assessment method, observer, and type of nest content (egg or chick) can influence burrow laying success and other assessments of Atlantic puffins. Therefore, correction factors are needed when data collection involves different methods.

Overall, I found that estimates obtained by burrowscope, even though not perfect, were consistently higher than those yielded by handgrubbing. Due to the unlikeliness of false positives, this indicates that burrowscope estimates are a better representation of true laying success values. Individual variation was significant between observers in both methods, and between methods, even though it was higher in burrowscope assessment. Variation in burrow content detection rates between methods has been described in other burrowing species (Ambagis, 2004; Bird et al., 2021). This variation might be the consequence of several factors, such as different individual capacities, lack of training, or detection differences between devices. Burrowscope inter-rater variation was significantly lower during the chick assessment, while it remained similar in handgrubbing. This might be explained by the contrast differences between chicks and eggs when visualized through an infra-red camera. Eggs are strikingly white when observed through the infrared lens of the burrowscope, contrasting with the darker background and making their detection easier. On the other hand, puffin chicks have dark bodies that blend with the background, and they can move inside the burrow, making their localization more challenging. These detection differences can affect the results of several important population parameters when assessed by burrowscope. A larger detection probability for eggs when compared to chicks may result in an underestimated productivity, where failure to detect eggs will underestimate colony laying success. All this masked variation needs to be accounted for. *A posteriori* correction is necessary to improve the accuracy and comparability of the estimates, reducing the uncertainty when correlating these with ecological factors, particularly when estimates are obtained by different assessment methods.

Even though detection accuracy differed between methods, there is a tradeoff between accuracy and practicality. Here we have only tested method accuracy and reliability, but there is a myriad of limitations associated to each method. Even though burrowscope is potentially more accurate, it showed extreme discrepancies between observers. On top of that, their use is limited by other factors such as budget restrictions, battery life, access to power during the assessment, damage sensitivity and malfunction, or strong weather dependency. Handgrubbing, which showed a higher inter-rater consistency, faces other challenges. Over 30 % of Newfoundland puffin burrows were out of an arm's reach, making them unsuitable for proper hand assessment. This percentage is an underestimate, since other factors such as burrow morphology can prevent a thorough check. This can decrease time efficiency in data collection, increasing human hours required for colony assessment. Together with the need of direct contact with the animal for manual assessment, this can result in increased stress on the animals, egg damage, disease transmission, or injuries to the animal or the observer.

Nevertheless, the experimental design conducted has limitations, mainly the lack of data of the true contents of each burrow. Experiments involving assessment of artificial burrows with known contents, placed by a researcher, can be useful in establishing even more accurate detection rates and inter-rater reliability.

4.2 Atlantic puffin breeding status in Newfoundland

The present study assessed the breeding status of several Newfoundland puffin colonies. The assessed colonies showed, during the 2021-2022 breeding seasons, higher values in laying success, fledging success, and productivity than any other colony from which data are available, particularly when compared with declining populations (Fig. 12). The only exception was Bakeapple Island in 2022, due to the disruptive presence of an Arctic fox in the colony. With an estimated population size of 897,566 breeding pairs (Wilhelm, unpublish. data), Newfoundland's Atlantic puffin population is the third largest in the world, only after Iceland, with ~ 2 000 000 breeding pairs (Lowther et al., 2020), and Norway, with ~ 1 884 000 breeding pairs (Fauchald et al., 2015). Despite these high population numbers, both Norwegian and Icelandic colonies have experienced extremely low productivity and fledging success, paired with steady population decline, in the past decades (Birdlife International, 2018). When compared to these, Newfoundland average laying success, fledging success, and productivity was 2.44, 1.68, and 2.77 times higher than Icelandic averages between the years 2012-2021, respectively (Hansen, 2021). Fledging success in Newfoundland was 2.07 times higher than the overall average at Norwegian colonies from the years 2008-2021 (Anker-Nilssen et al., 2021; Fauchald et al., 2015). This makes Newfoundland host of the largest puffin aggregation in the world with successful breeding and stable/increasing numbers.

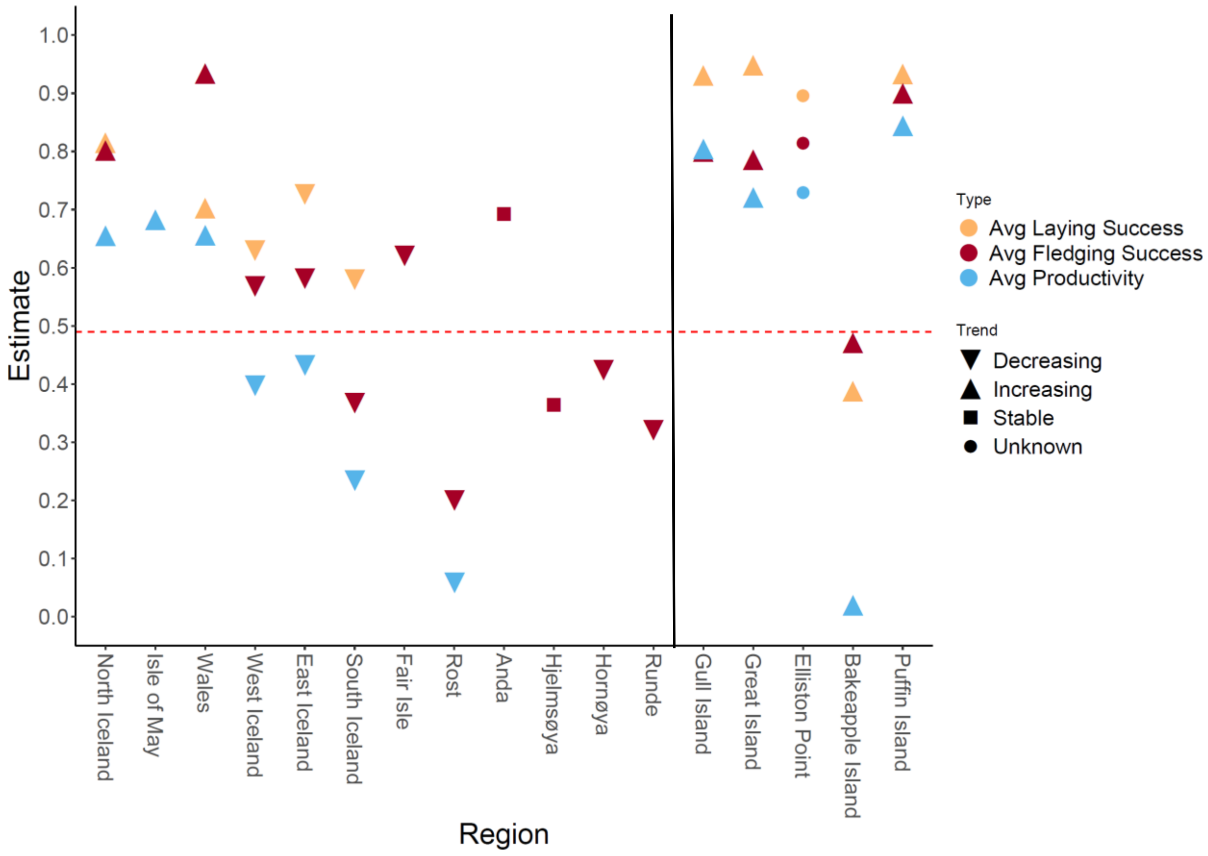


Figure 12. Plot laying success, fledging success, and productivity values from Newfoundland colonies (right of the vertical black line) and other colonies (left of the vertical black line), together with estimated colony population trends (point shape). Dotted red line set at 0.49, representing the estimated productivity limit for positive population growth (Hansen, 2021). Newfoundland estimates obtained from the present study. Study years and references for the other colonies indicated in Tables 4 and 5.

Successful breeding has been consistent throughout the years, as the historical data suggests. Both historical laying success and fledging success remained high in every colony where data are available. General laying success on Gull and Great islands was significantly lower during the 2021-2022 breeding seasons. This could reflect a slight decline in the number of breeding adults at the colony (Tables 21 & 22). Since 2021 and 2022, general laying success estimates are obtained from high density areas, whole island assessments are needed to evaluate this potential decline. These two islands not only host the largest Atlantic puffin colonies in the Western Atlantic, but also a diverse and abundant seabird community. This can lead to both intra- and inter-specific competition due to saturation of the nesting habitat's carrying capacity (Durant et al., 2012; Petalas et al., 2021). Carrying capacity fluctuates with environmental conditions, and the effects of such fluctuations have a greater impact on nearly saturated habitats (Goyert et al., 2018). Decreases in the carrying capacity of the habitat linked to persistent environmental changes could be the cause for this decline in laying success.

The only exception to this high laying success rates is Bakeapple Island, showing the largest fluctuations. As seen in this study, predation events might be common on this island. Its proximity to Labrador, and the high volume of sea ice that reaches the adjacent waters (Petrie & Anderson, 1983), can make these colonies particularly vulnerable to the introduction of land predators (see for example, Burke et al., 2011). Due to the lack of physical barriers, it is likely that this vulnerability is extended to every island in the archipelago. The extremely low laying success value found in Bakeapple Island during the 2014 could be related to predation. The low laying success on the island, with few adults actively breeding as a response to the presence of predators, has been described in other auk species (Major et al., 2006). Another explanation for this low value could reside in collapse-recolonization dynamics. Burrow density on Bakeapple Island was approximately half compared to the neighboring Puffin Island. Puffin colonies can collapse as consequence of digging processes that lead to complete erosion, and several years are needed for the soil to be suitable for recolonization (Furness, 1991). These lower density values might reflect a recent colony collapse-recolonization process, which could indicate that in the recent past there were a high proportion of unusable and degraded burrows. A third hypothesis suggests gradual reintroduction of puffins in the archipelago. Little Fogo Islands was used traditionally by humans as fishing hub during the summer until 1992, when the Canadian Cod Moratorium was initiated (Hutchings & Myers, 1994). Cod fisheries relied heavily on gill-net, which has a large impact on diving species such as auks (Regular et al., 2013). Between the years 1992-2009, the waters surrounding Little Fogo Islands suffered from the largest incidence of gill-net fishing in the whole of Newfoundland, resulting in population declines in several auk species, including puffins (Regular et al., 2013). Therefore, the combination of human presence and by-catch in Little Fogo Islands region could have displaced puffins to other areas due to disturbance, or decreased their numbers due to by-catch, process that was reversed after the implementation of the Moratorium. The puffin population in this archipelago has seen their numbers slowly increasing from ~150 breeding pairs in the 1970's (Cairns et al., 1989, Wilhelm, pers. comm.) to several thousands. Besides increased burrow laying success, I found anecdotical evidence supporting this steady population growth. A full island count on one of the previously human inhabited islands, St. Anne's Island (coordinates 49.811598, -54.119403) is hosting a active puffin breeding population for the first time in record. On this island, 48 % of the Atlantic puffin burrows had fledgling chicks (N = 58) (Zabala, pers. obs.). Further comparative research between the different colonies on the archipelago is needed to understand the specific dynamics and factors that drive this puffin community. Even with the presence of circumstantial trends in the data, its fragmentary nature impedes more thorough analyses. Moreover, inter-annual comparison might not be entirely possible in large colonies, since historical data comes from randomly distributed plots in colony-wide surveys, while 2021-2022 data was obtained from targeted areas. Longer temporal series with more consistent data collection are needed to understand puffin breeding estimates in Newfoundland.

All data available indicates that Newfoundland puffin colonies have been successfully breeding during the past decades. These results seem in contradiction with the ecological context that Atlantic puffins face in Newfoundland. Atlantic puffins are central place foragers during breeding, seeing their fitness decreased as average foraging trip from the colony increases, with a maximum recorded foraging trip length of 74 km (Fayet et al., 2021). Traditionally, capelin (*Mallotus villosus*) has been considered the focal prey item used by Atlantic puffins during chick rearing in Western Atlantic colonies (Carscadden et al., 2002). Capelin stocks and recruitment have been declining in the province since the early 1990s, due multifactorial reasons but linked with bottom-up climate-trophic processes that crashed the population (Buren et al., 2014; Murphy et al., 2021). Even so, several studies failed to link low capelin chick diets with increased breeding failure in Newfoundland Atlantic puffins, mainly due to the presence of alternative prey in chick's diet (Baillie & Jones, 2003, 2004; Russell, 1999). This plasticity is not found in the whole species' range, with other puffin colonies showing stronger dependence on a single prey that leads to massive colony failures when these fish species abundance are low, or out of reach (Hansen, 2021; Satre et al., 2002).

Furthermore, adult mortality reduction associated to the removal of cod fisheries might produce an overall population increase that obscures the effect of preferred prey limitation. No chick diet data studies were available for the years assessed here, but my results suggest that food availability was not a limiting factor for fledging success. This evidence of plasticity during chick food provisioning (Baillie, 2001; Baillie & Jones, 2004) might make Newfoundland and Labrador puffin colonies more resilient to climate change mediated changes in the oceanic food chain. Nevertheless, to obtain clear relationships between ecological factors and fledging success, long-term regional data series are required (Burr et al., 2016; Merkel et al., 2019; Visser & Both, 2005).

Nevertheless, the intent of these comparisons was to show potential changes in multi-year surveys when data collection protocols are modified. My results stress the importance of data collection methods and origin on the calculation of estimates. This study has shown the importance of correction factors when comparing historical breeding estimates obtained by different methods. Data collection protocols should be as detailed as possible, and changes in data collection methods should be included in scientific reports.

4.3 Newfoundland Atlantic puffin phenology

This study has explored the use of the cumulative distribution function as a proxy to calculate mean hatching dates when consistent monitoring of the colonies is not possible. Even so, constant monitoring is strongly preferred since the relatively large error associated to estimates by the cumulative function method makes them not suitable for fine environmental match-mismatch analysis, although it can lead to disturbance. I have evidence of phenological differences between Newfoundland colonies. Estimated mean hatching date in northern colonies was approximately 20 days later than those of the more southern colonies. Timing of breeding can be a heritable trait, and therefore can respond to natural selection, producing local phenological adaptations (Miller-Rushing et al., 2010). This variation in phenology can respond to differences in resource availability, particularly food. As previously stated, capelin is the preferred prey item for Atlantic puffins in Newfoundland colonies, but its availability is limited by distance from the colony during the breeding season. Capelin overwinters offshore, travelling inshore every spring to spawn on beaches (Carscadden et al., 1997). This beach arrival is highly variable, with colder water temperatures during capelin gonadal development (February-June) delaying capelin spawning in coastal communities (Carscadden et al., 1997; Murphy et al., 2021; Regular et al., 2009). The stronger influence of the Labrador Current on the northern Newfoundland coast makes the adjacent oceanic areas to be 1.58 ºC (±0.18 ºC) colder than those on southern locations during capelin gonadal development months (Fig. S10), which can help explain delayed puffin phenologies to match capelin spawning in northern colonies. In fact, food availability in the previous year has been shown to be a strong predictor of hatching date in some puffin populations (Durant et al., 2004). This gradient hypothesis is in concordance with other Western Atlantic puffin colonies.

Gannet Island Ecological Reserve (Labrador) puffin colonies in the Canadian Arctic, have mean hatching date on early August (Baillie & Jones, 2004). Machias Seal Island (Gulf of Maine) colonies, situated further south of Newfoundland, have mean hatching dates on June 22. This latitudinal phenology gradient within the same population has also been described in other regions (Burr et al., 2016; Hansen, pers. obs.). This local adaptation to specific environmental conditions can make this species particularly sensitive to climate change, with different parts of the population affected differently. It is essential to understand the particularities of these local relationships.

Traditionally, Newfoundland puffins have been described as a single population, but genetic analysis included individuals from only Witless Bay colonies (Kersten et al., 2021). Mismatched breeding phenologies can lead to temporal isolation between populations, which is a driver of genetic differentiation (Coyne & Orr, 2004). Chicks in Witless Bay were significantly smaller than those found in Little Fogo islands, even though no hard physical barriers are present between these colonies. Differences in chick size and weight has been previously described between Labrador and Witless Bay colonies (Baillie & Jones, 2003). Our data concur with this hypothesis. The presence of genetically separated populations resulting from temporal isolation could explain the evidence of different chick growth ratios between northern and southern colonies. Multi-colony genetic and morphometric studies are needed to define the population structure in Western Atlantic puffins.

Phenology and fledging success are interconnected, as breeding cycles are adapted to match food availability. A timing mismatch can lead to breeding failure in some seabird species (Piatt et al., 2007), or increased energetic expenditure by rearing parents to buffer for lack of food, often at parent’s energetic reservoir expense (Regular et al., 2014; Wilhelm et al., 2008). Due to this buffering capacity, relationships between ecological factors and breeding usually follow a nonlinear fashion (Hansen et al., 2021; Piatt et al., 2007), and affects different seabird species in different ways depending on their specific biology (Piatt et al., 2007). Capelin spawning date has been used as a proxy for its availability to colony seabirds in Newfoundland (Regehr & Rodway, 1999; Regular et al., 2014; Wilhelm et al., 2008), although capelin coastal congregations might be present before beach spawning (Nakashima & Wheeler, 2002). First evidence of capelin beach spawning in Witless Bay was the on 7 July in 2021, and 28 June in 2022 (ECapelin, 2017). Under the assumption that capelin needs to be available for puffin chick rearing after the first day of spawning, this resulted in a 6-day mismatch between capelin spawning and the colony mean hatching date in 2021, and no mismatch in 2022. Elliston Point adjacent beaches had capelin spawning as early as July 16 (ECapelin, 2017), corresponding to the estimated mean hatching date of the puffin colony in 2022, suggesting no mismatch either. These differences are much lower than previously recorded (Fitzsimmons et al., 2017), although the behavioural plasticity of pursuit-divers such as puffins makes hard to establish a mismatch threshold (Baillie & Jones, 2004; Regular et al., 2014). Because both fledging success and productivity values were similar between 2021 and 2022, and with the historical data available, this supports the hypothesis that puffins can buffer changes in food availability to certain extent. Nevertheless, species with different phenologies, existing at different latitudes, are affected differently by climate change (Visser & Both, 2005). This is being seen in Western Atlantic puffin populations, where colonies at the southernmost limit are showing signs of poor breeding success (Major et al., 2021). Therefore, full range monitoring is needed to assess potential changes in the fate of local populations.

5. General conclusions

My research provides evidence that breeding estimates are influenced by observers and survey method and that these biases need to be accounted for. The number of researchers involved in this study, together with their high level of expertise, can be interpreted as a limitation. This biases are not such, since colony assessments usually involve small teams of highly experienced individuals, due to the inaccessibility of the colonies and the potential researcher impact on the colonies. Some surveys involve large teams working together over an entire breeding season. Nevertheless, a replication of this experiment with more raters in a myriad of conditions, or assessments of artificial nests with known contents, might be necessary to further assess sample size effects on estimates. Individual variation in detection consistency could be improved with supervised training, both in controlled and natural set ups, whereas continuation of individual detection rates over several seasons could help improve estimates obtained by different team members. Technological improvements, such as better infrared sensors, thermal cameras, better batteries, or automatized devices, will translate in new assessment methods with improved detection rates. Protocols and results should be updated to include methodological changes. To analyze long temporal series, studies need to corroborate estimates obtained by different methods to make analyses more precise.

The total Newfoundland and Labrador Atlantic puffin population accounts for almost 900,000 breeding pairs. Unlike their European counterparts, these colonies seem to be stable, even growing (Birdlife International, 2018). Their good condition is reflected in a successful breeding cycle, that is potentially providing generational replacement. This makes Newfoundland and Labrador puffin colonies important reservoirs for a declining species. In the face of a changing climate that can affect these positive trends, this establishes the need for more constant monitoring. This study has implemented a breeding monitoring protocol in some of the main puffin colonies across Newfoundland, but some important colonies were left out. Colonies such as the one found on Baccalieu Island (48.131347, -52.801967) were last assessed in 2005, where the total population estimate was ~75,000 breeding pairs. This colony has given signs of a large population increase over the past decades, probably due to high fledging success and the population rebound effect of the gillnet removals associated with the fishery moratorium (Regular et al., 2013). The inclusion of colonies like this one as part of a consistent monitoring protocol would help create a more complete picture of the breeding status of Newfoundland colonies. Moreover, constant breeding failure has been the first sign of decline in other puffin colonies, so the elaboration of long-term data series can result in early detection and proactive implementation of corrective measures if these successful breeding trends are reversed.

6. References

Akaike, H. (1985). Prediction and entropy. In A. C. Atkinson & S. E. Fienberg (Eds.), *A Celebration of Statistics* (pp. 1–24). Springer, New York.

Albores-Barajas, Y. V., Soldatini, C., Ramos-Rodríguez, A., Alcala-Santoyo, J. E., Carmona, R., & Dell'Omo, G. (2018). A new use of technology to solve an old problem: Estimating the population size of a burrow nesting seabird. *PLoS ONE*, *13*(9), 1–15. https://doi.org/10.1371/journal.pone.0202094

Ambagis, J. (2004). A comparison of census and monitoring techniques for Leach's Storm Petrel. *Waterbirds*, *27*(2), 211–215. https://doi.org/10.1675/1524-4695(2004)027[0211:ACOCAM]2.0.CO;2

Anker-Nilssen, T., Harris, M. P., Kleven, O., & Langset, M. (2017). Status , origin , and population level impacts of Atlantic Puffins killed in a mass mortality event in southwest Norway early 2016. *Seabird*, *30*(March), 1–14.

Anker-Nilssen, T, Aarvak, T., & Bangjord, G. (2003). Mass mortality of Atlantic puffins *Fratercula arctica* off Central Norway, spring 2002: causes and consequences. *Atlantic Seabirds*, *5*(2), 57–72.

Anker-Nilssen, T, Barrett, R., Christensen-Dalsgaard, S., Dehnhard, N., Descamps, S., Systad, G. H., Moe, B., Reiertsen, T. K., Bustnes, J. O., Erikstad, K.-E., Follestad, A., Hanssen, S. A., Langset, M., Lorentsen, S.-H., Lorentzen, E., & Strøm, H. (2020). Key-site monitoring in Norway 2018, including Svalbard and Jan Mayen. *Seapop*, 1–15.

Anker-Nilssen, Tycho, Barrett, R. T., Børge Moe, Reiertsen, T. K., Systad, G., Jan, B., Christensen-Dalsgaard, S., Descamps, S., Erikstad, K. E., Follestad, A., Hansen, S., Langset, M., Loretsen, S.-H., Loretzen, E., & Strøm, H. (2021). Key-site monitoring in Norway 2011-2021, including Svalbard and Jan Mayen. *SEAPOP Short Report*, 1–13.

Baillie, S. M. (2001). Atlantic Puffin response to changes in capelin abundance in Newfoundland and Labrador: An intercolony and interdecade comparison. *MSc Thesis. Memorial University of Newfoundland*.

Baillie, S. M., & Jones, I. L. (2003). Atlantic Puffin (*Fratercula arctica*) chick diet and reproductive performance at colonies with high and low capelin (*Mallotus villosus*) abundance. *Canadian Journal of Zoology*, *81*(9), 1598–1607. https://doi.org/10.1139/z03-145

Baillie, S. M., & Jones, I. L. (2004). Response of Atlantic puffins to a decline in Capelin abundance at the Gannet Islands, Labrador. *Waterbirds*, *27*(1), 102–111. https://doi.org/10.1675/1524-4695(2004)027[0102:ROAPTA]2.0.CO;2

Barrett, R. T. (2015). Atlantic Puffin *Fratercula arctica* chick growth in relation to food load composition. *Seabird*, *28*, 17–29.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bendik, N. F., & Dries, L. A. (2018). Density-dependent and density-independent drivers of population change in Barton Springs salamanders. *Ecology and Evolution*, *8*(11), 5912–5923. https://doi.org/10.1002/ece3.4130

Bernhardt, J. R., Connor, M. I. O., Sunday, J. M., Gonzalez, A., & Bernhardt, J. R. (2020). Life in fluctuating environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*.

Bird, J. P., Fuller, R. A., Pascoe, P. P., & Shaw, J. D. S. (2021). Trialling camera traps to determine occupancy and breeding in burrowing seabirds. *Remote Sensing in Ecology and Conservation*, 1–11. https://doi.org/10.1002/rse2.235

BirdLife International. (2012). *Spotlight on seabirds.Presented as part of the BirdLife State of the world's website.*

BirdLife International. (2018). *Fratercula arctica*, Atlantic Puffin. *The IUCN Red List of Threatened Species*, *8235*.

Bled, F., Royle, J. A., & Cam, E. (2011). Assessing hypotheses about nesting site occupancy dynamics. *Ecology*, *92*(4), 938–951. https://doi.org/10.1890/10-0392.1

Breton, A. R. & Diamond, A. W. (2014). Annual survival of adult Atlantic Puffins Fratercula arctica is positively correlated with Herring Clupea harengus availability. *Ibis,* 156, 35–47.

Brown, R., & Eagle, G. (2021). *Skokholm Bird Observatory Seabird Report series*. https://www.welshwildlife.org/sites/default/files/2022-03/Seabird-Report-2021.pdf

Buren, A. D., Koen-Alonso, M., Pepin, P., Mowbray, F., Nakashima, B., Stenson, G., Ollerhead, N., & Montevecchi, W. A. (2014). Bottom-up regulation of capelin, a keystone forage species. *PLoS ONE*, *9*(2), 1–11. https://doi.org/10.1371/journal.pone.0087589

Burr, Z. M., Varpe, Ø., Anker-Nilssen, T., Erikstad, K. E., Descamps, S., Barrett, R. T., Bech, C., Christensen-Dalsgaard, S., Lorentsen, S., Moe, B., Reiertsen, T. K., & Strøm, H. (2016). Later at higher latitudes: large-scale variability in seabird breeding timing and synchronicity. *Ecosphere*, *7*(5), 1–12. https://doi.org/10.1002/ecs2.1283

Cairns, D. K., Elliot, R. D., Threlfall, W., & Montevecchi, W. A. (1989). Researchers ' guide to Newfoundland seabird colonies. *Memorial University of Newfoundland Ocassional Papers*, *10*, 1–50.

Carey, M. J. (2009). The effects of investigator disturbance on procellariiform seabirds: A review. *New Zealand Journal of Zoology*, *36*(3), 367–377. https://doi.org/10.1080/03014220909510161

Carscadden, J. E., Montevecchi, W. A., Davoren, G. K., & Nakashima, B. S. (2002). Trophic relationships among capelin (*Mallotus villosus*) and seabirds in a changing ecosystem. *ICES Journal of Marine Science*, *59*(5), 1027–1033. https://doi.org/10.1006/jmsc.2002.1235

Carscadden, J. E., Nakashima, B. S., & Frank, K. T. (1997). Effects of fish length and temperature on the timing of peak spawning in capelin (*Mallotus villosus*). *Canadian Journal of Fisheries and Aquatic Sciences*, *54*(4), 781–787. https://doi.org/10.1139/cjfas-54-4-781

Catry, T., Ramos, J. A., Catry, I., Monticelli, D., & Granadeiro, J. P. (2013). Inter-annual variability in the breeding performance of six tropical seabird species: Influence of life-history traits and relationship with oceanographic parameters. *Marine Biology*, *160*(5), 1189–1201. https://doi.org/10.1007/s00227-013-2171-2

Cheung, W. W. L., Watson, R., & Pauly, D. (2013). Signature of ocean warming in global fisheries catch. *Nature*, *497*(7449), 365–368. https://doi.org/10.1038/nature12156

Claire, C. G., Thomas, C., Clauser, C. E., Bradtmiller, B., McConville, J. T., & Walker, R. A. (1989). *Anthropometric Survey of U.S. Personnel: Summary Statistics Interim Report*.

Coyne, J. A., & Orr, H. A. (2004). *Speciation (Vol. 37)*. Sunderland, MA: Sinauer associates.

Croxall, J. P., Butchart, S. H. M., Lascelles, B. E. N., Stattersfield, A. J., Sullivan, B., Symes, A., & Taylor, P. (2012). Seabird conservation status and threats: A global assessment of priorities. *Bird Conservation International*, *22*, 1–34.

Croxall, J. P., & Rothery, P. (1991). Population regulation of seabirds: implication of their demography for conservation. In C. M. Perrins, J. D. Leberton, & G. J. Hirons (Eds.), *Bird population studies: relevance to conservation and management.* (pp. 272–296). Oxford University Press.

Croxall, J. P., Trathan, P. N., & Murphy, E. J. (2002). Environmental change and Antarctic seabird populations. *Science*, *297*(5586), 1510–1514. https://doi.org/10.1126/science.1071987

Cushing, D. H. (1990). Plankton production and year-class strength in fish populations: An update of the match/mismatch hypothesis. In *Advances in Marine Biology* (Vol. 26, Issue C). https://doi.org/10.1016/S0065-2881(08)60202-3

Cuthbert, R. J., Louw, H., Lurling, J., Parker, G., Rexer-Huber, K., Sommer, E., Visser, P., & Ryan, P. G. (2013). Low burrow occupancy and breeding success of burrowing petrels at Gough Island: A consequence of mouse predation. *Bird Conservation International*, *23*(2), 113–124. https://doi.org/10.1017/S0959270912000494

D'Entremont, K. J. N., Zitske, L. M., Gladwell, A. J., Elliott, N. K., Mauck, R. A., & Ronconi, R. A. (2020). Breeding population decline and associations with nest site use of Leach's storm-petrels on Kent Island, New brunswick from 2001 to 2018. *Avian Conservation and Ecology*, *15*(1), 1–11. https://doi.org/10.5751/ACE-01526-150111

Davies, D., Dilley, B. J., Bond, A. L., Cuthbert, R. J., & Ryan, P. G. (2015). Tendances et tactiques de prédation par des souris sur les poussins d'Albatros de Tristan *Diomedea dabbenena* à l'île Gough, océan Atlantique Sud. *Avian Conservation and Ecology*, *10*(1).

Del Hoyo, J., Elliott, A., & Sargantal, J. (1992). *Handbook of the birds of the world (Vo.1, No. 8)* (Barcelona: Lynx edicions (ed.)).

Descamps, S., Ramírez, F., Benjaminsen, S., Anker-Nilssen, T., Barrett, R. T., Burr, Z., Christensen-Dalsgaard, S., Erikstad, K. E., Irons, D. B., Lorentsen, S. H., Mallory, M. L., Robertson, G. J., Reiertsen, T. K., Strøm, H., Varpe, Ø., & Lavergne, S. (2019). Diverging phenological responses of Arctic seabirds to an earlier spring. *Global Change Biology*, *25*(12), 4081–4091. https://doi.org/10.1111/gcb.14780

Devney, C. A., Short, M., & Congdon, B. C. (2009). Sensitivity of tropical seabirds to El Niño precursors. *Ecology*, *90*(5), 1175–1183. https://doi.org/10.1890/08-0634.1

Diamond, A. W. (2021). Seabirds in a changing ocean: an Overview of 20 Years of research and monitoring on Machias Seal Island , Bay of Fundy , Canada. *Proceedings of the Nova Scotian Institute of Science*, *51*(2), 365–409.

Dobson, A. (1990). Survival rates and their relationship to life-history traits in some common British bird. In D. M. Power (Ed.), *Current Ornithology Vol. 7* (pp. 115–146). Plenum Press, New York, NY, USA.

Doody, L. M., Wilhelm, S. I., McKay, D. W., Walsh, C. J., & Storey, A. E. (2008). The effects of variable foraging conditions on common murre (*Uria aalge*) corticosterone concentrations and parental provisioning. *Hormones and Behavior*, *53*(1), 140–148. https://doi.org/10.1016/j.yhbeh.2007.09.009

Durant, J., Anker-Nilssen, T., & Stenseth, N. C. (2003). Trophic interactions under climate fluctuations: The Atlantic puffin as an example. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1523), 1461–1466. https://doi.org/10.1098/rspb.2003.2397

Durant, J. M., Anker-Nilssen, T., Hjermann, D., & Stenseth, N. C. (2004). Regime shifts in the breeding of an Atlantic puffin population. *Ecology Letters*, *7*(5), 388–394. https://doi.org/10.1111/j.1461-0248.2004.00588.x

Durant, J. M., Hjermann, D., Ottersen, G., & Stenseth, N. C. (2007). Climate and the match or mismatch between predator requirements and resource availability. *Climate Research*, *33*(3), 271–283. https://doi.org/10.3354/cr033271

Durant, J. M., Krasnov, Y. V., Nikolaeva, N. G., & Stenseth, N. C. (2012). Within and between species competition in a seabird community: Statistical exploration and modeling of time-series data. *Oecologia*, *169*(3), 685–694. https://doi.org/10.1007/s00442-011-2226-3

ECapelin. (2017). *eCapelin: An online observation tool of capelin and capelin spawn distribution along East Canadian coasts.* ECapelin, Saint-Lawrence Global Observatory, Rimouski, Québec, Canada. https://ecapelan.ca/?lg=en

Edwards, W. J., & Edwards, C. T. (2011). Population limiting factors. *Nature Education Knowledge*, *3*(10), 1.

Evans, P., & Nettleship, D. N. (1985). Conservation of the Atlantic Alcidae. In D. N. Nettleship & T. R. Birkhead (Eds.), *The Atlantic Alcidae: Evolution, Distribution and Biology of the Auks Inhabiting the Atlantic Ocean and Adjacent Water Areas* (pp. 428–488). Academic Press, London, United Kingdom.

Fauchald, P., Anker-Nilssen, T., Barrett, R. T., Ove Bust-nes, J., Bårdsen, B.-J., Christensen-Dalsgaard, S., Descamps, S., Engen, S., Einar Erikstad, K., Are Hanssen, S., Lorentsen, S.-H., Moe, B., Reiertsen, T. K., Strøm, H., Helge Systad, G., Bustnes, J. O., Bårdsen, B.-J., Christensen-Dalsgaard, S., Engen, S., … Systad, G. H. (2015). The status and trends of seabirds breeding in Norway and Svalbard. In *NINA Report* (Vol. 1151).

Fayet, A. L., Clucas, G. V., Anker-Nilssen, T., Syposz, M., & Hansen, E. S. (2021). Local prey shortages drive foraging costs and breeding success in a declining seabird, the Atlantic puffin. *Journal of Animal Ecology*, *90*(5), 1–13. https://doi.org/10.1111/1365-2656.13442

Fayet, A. L., Freeman, R., Anker-Nilssen, T., Diamond, A., Erikstad, K. E., Fifield, D., Fitzsimmons, M. G., Hansen, E. S., Harris, M. P., Jessopp, M., Kouwenberg, A. L., Kress, S., Mowat, S., Perrins, C. M., Petersen, A., Petersen, I. K., Reiertsen, T. K., Robertson, G. J., Shannon, P., … Guilford, T. (2017). Ocean-wide drivers of migration strategies and their influence on population breeding performance in a declining seabird. *Current Biology*, *27*(24), 3871-3878.e3. https://doi.org/10.1016/j.cub.2017.11.009

Fitzsimmons, M. G., Rector, M. E., McKay, D. W., & Storey, A. E. (2017). High growth and low corticosterone in food supplemented Atlantic puffin *Fratercula arctica* chicks under poor foraging conditions. *Marine Ecology Progress Series*, *565*, 217–226. https://doi.org/10.3354/meps11999

Frederiksen, M., Harris, M. P., Daunt, F., Rothery, P., & Wanless, S. (2004). Scale-dependent climate signals drive breeding phenology of three seabird species. *Global Change Biology*, *10*(7), 1214–1221. https://doi.org/10.1111/j.1529-8817.2003.00794.x

Froy, H., Lewis, S., Nussey, D. H., Wood, A. G., & Philips, R. A. (2017). Contrasting drivers of reproductive ageing in albatrossess. *Evolutionary Ecology*, *86*(5), 1022–1032. https://doi.org/10.1111/ijlh.12426

Fuentes, M. A., & Ferrada, E. (2017). Environmental fluctuations and their consequences for the evolution of phenotypic diversity. *Frontiers in Physics*, *5*(MAY), 1–7. https://doi.org/10.3389/fphy.2017.00016

Furness, R. (1991). The occurrence of burrow-nesting among birds and its influence on soil fertility and stability. *Symposia of the Zoological Society of London- the Environmental Impact of Burrowing Animals and Animal Burrows*, 53–57.

Goyert, H. F., Garton, E. O., & Poe, A. J. (2018). Effects of climate change and environmental variability on the carrying capacity of Alaskan seabird populations. *Auk*, *135*(4), 975–991. https://doi.org/10.1642/AUK-18-37.1

Grant, P. R., & Nettleship, D. N. (1971). Nesting Habitat Selection by Puffins *Fratercula arctica* L. in Iceland. *Ornis Scandinavica*, *2*(2), 81. https://doi.org/10.2307/3676209

Grosbois, V., Harris, M.P., Anker-Nilssen T., et al. (2009). Modeling survival at multi-population scales using mark-recapture data. *Ecology*. 90(10), 2922-2932. doi:10.1890/08-1657.1

Guillemette, M., Grégoire, F., Bouillet, D., Rail, J. F., Bolduc, F., Caron, A., & Pelletier, D. (2018). Breeding failure of seabirds in relation to fish depletion: Is there one universal threshold of food abundance? *Marine Ecology Progress Series*, *587*(January), 235–245. https://doi.org/10.3354/meps12442

Halpern, B. S., Gaines, S. D., & Warner, R. R. (2005). Habitat size, recruitment, and longevity as factors limiting population size in stage-structured species. *American Naturalist*, *165*(1), 82–94. https://doi.org/10.1086/426672

Hamilton, S. (2000). How precise and accurate are data obtained using an infra-red scope on burrow-nesting sooty shearwaters *Puffinus griseus*? *Marine Ornithology*, *28*(1), 1–6.

Hansen, E. S. (2021). Puffin population monitoring 2010-2021. *Final Report to the Environmental Agency of Iceland. [In Icelandic].*, 1–41.

Hansen, E. S., Sandvik, H., Erikstad, K. E., Yoccoz, N. G., Anker-Nilssen, T., Bader, J., Descamps, S., Hodges, K., Mesquita, M. d. S., Reiertsen, T. K., & Varpe, Ø. (2021). Centennial relationships between ocean temperature and Atlantic puffin production reveal shifting decennial trends. *Global Change Biology*, *27*(16), 3753–3764. https://doi.org/10.1111/gcb.15665

Harris, M. . (1969). Breeding seasons of sea-birds in the Galapagos Islands. *Journal of Zoology*, *159*, 145–165.

Harris, M. P., Anker-Nilssen, T., McCleery, R. H., Erikstad, K. E., Shaw, D. S., & Grosbois, V. (2005). Effect of wintering area and climate on the survival of adult Atlantic puffins Fratercula arctica in the eastern Atlantic. *Marine Ecology Progress Series*, *297*(August 2014), 283–296. https://doi.org/10.3354/meps297283

Harris, M. P., Freeman, S. N., Wanless, S., Morgan, B. J. T., & Wernham, C. V. (1997). Factors Influencing the Survival of Puffins *Fratercula arctica* at a North Sea Colony over a 20-Year Period. *Journal of Avian Biology*, *28*(4), 287. https://doi.org/10.2307/3676941

Harris, M. P., Heubeck, M., Bogdanova, M. I., Newell, M. A., Wanless, S., & Daunt, F. (2020). The importance of observer effort on the accuracy of breeding success estimates in the Common Guillemot *Uria aalge*. *Bird Study*, *67*(1), 93–103. https://doi.org/10.1080/00063657.2020.1779654

Harris, M. P., & Wanless, S. (2011). *The Puffin* (T. Poyser & A. D. Poyser (eds.)). Yale University Press.

Harting, F. (2022). *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models* (R package version 0.4.6).

Hipfner, J. M. (2008). Matches and mismatches: Ocean climate, prey phenology and breeding success in a zooplanktivorous seabird. *Marine Ecology Progress Series*, *368*, 295–304. https://doi.org/10.3354/meps07603

Hudson, P. J. (1985). Population parameters for the Atlantic Alcidae. In D. N. Nettleship & T. R. Birkhead (Eds.), *The Atlantic Alcidae: The Evolution, Distribution and Biology of the Auks Inhabiting the Atlantic Ocean and Adjacent Water Areas* (pp. 233–261). Academic Press, New York.

Hutchings, J. A., & Myers, R. A. (1994). What can be learned from the collapse of a renewable resource? Atlantic Cod, Gadus morhua, of Newfoundland and Labrador. *Can. J. Fish. Aquat. Sci.*, *51*, 2126–2146.

JNCC. (2021). *Seabird Population Trends and Causes of Change: 1986–2019 Report*.

Jones, C., Bettany, S., Moller, H., Fletcher, D., Lyver, P., & De Cruz, J. (2003). Burrow occupancy and productivity at coastal sooty shearwater (*Puffinus griseus*) breeding colonies, South Island, New Zealand: Can mark-recapture be used to estimate burrowscope accuracy? *Wildlife Research*, *30*(4), 377–388. https://doi.org/10.1071/WR01050

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., & Iredell, M. (1996). The NCEP/NCAR reanalysis 40-year project. *Bulletin of the American Meteorological Society*, *77*, 437–471.

Keogan, K., Daunt, F., Wanless, S., Phillips, R. A., Walling, C. A., Agnew, P., Ainley, D. G., Anker-Nilssen, T., Ballard, G., Barrett, R. T., Barton, K. J., Bech, C., Becker, P., Berglund, P. A., Bollache, L., Bond, A. L., Bouwhuis, S., Bradley, R. W., Burr, Z. M., … Lewis, S. (2018). Global phenological insensitivity to shifting ocean temperatures among seabirds. *Nature Climate Change*, *8*(4), 313–317. https://doi.org/10.1038/s41558-018-0115-z

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), 1–12. https://doi.org/10.1038/s42003-021-02415-4

Lack, D. (1968). *Ecological adaptations for breeding in birds*. Methuen, London.

Lavers, J. L., Hutton, I., & Bond, A. L. (2019). Changes in technology and imperfect detection of nest contents impedes reliable estimates of population trends in burrowing seabirds. *Global Ecology and Conservation*, *17*, e00579. https://doi.org/10.1016/j.gecco.2019.e00579

Lewis, S., Elston, D. A., Daunt, F., Cheney, B., & Thompson, P. M. (2009). Effects of extrinsic and intrinsic factors on breeding success in a long lived seabird. *Oikos*, *118*(4), 521–528. https://doi.org/10.1111/j.1600-0706.2009.17308.x

Lowther, P. E., Diamond, A. W., Kress, S. W., Robertson, G. J., Russell, K., Nettleship, D. N., Kirwan, G. M., Christie, D., Sharpe, C., Garcia, E., & Boesman, P. F. D. (2020). Atlantic Puffin (*Fratercula arctica*). *Birds of the World*. https://doi.org/10.2173/bow.atlpuf.01

Lyver, P. O., Hamilton, S., Mckenzie, M., Dickson, I., Broad, M. D. T., & Moller, H. (1998). A burrowscope for examining petrel nests in burrows. *Conservation Advisory Science Notes No.*, *209*(October).

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

Major, H.L., Durham, S.E., Fana, N., Rivers, J.E., Diamond, A.W. (2021). Contrasting phenological and demographic responses of Atlantic Puffin (*Fratercula arctica*) and Razorbill (*Alca torda*) to climate change in the Gulf of Maine. *Elementa*, *9(1)*:1-19. doi:10.1525/elementa.2021.00033

Major, Heather L., Ian L. Jones, G. Vernon Byrd, and Jeffrey C. Williams. (2006.) Assessing the Effects of Introduced Norway Rats (*Rattus Norvegicus*) on Survival and Productivity of Least Auklets (Aethia Pusilla). *Auk,* 123(3), 681–94.

McHugh, M. L. (2012). Lessons in biostatistics interrater reliability : the kappa statistic. *Biochemica Medica*, *22*(3), 276–282. https://hrcak.srce.hr/89395

Mckenchie, S., Fletcher, D., Moller, H., Scott, D. S., Newman, J., & Bgragg, C. (2007). Estimating and correcting for bias in population assessments of Sooty shearwaters. *Journal of Wildlife Management*, *71*(4), 1325–1335. https://doi.org/10.2193/2006-018

Medrano, F., Militao, T., Gomes, I., Sardá-Serra, M., de la Fuente, M., Dinis, H., & Gonzalez-Solis, J. (2022). Phenological divergence, population connectivity and ecological segregation between two allochronic seabird populations. *Frontiers in Marine Science*, *9*, 975716. https://doi.org/10.3389/fmars.2022.975716

Merkel, B., Descamps, S., Yoccoz, N. G., Danielsen, J., Daunt, F., Erikstad, K. E., Ezhov, A. V., Grémillet, D., Gavrilo, M., Lorentsen, S. H., Reiertsen, T. K., Steen, H., Systad, G. H., Pórarinsson, P. L., Wanless, S., & Strøm, H. (2019). Earlier colony arrival but no trend in hatching timing in two congeneric seabirds (*Uria* spp.) across the North Atlantic. *Biology Letters*, *15*(10), 1–6.

Mesquita, M. D. S., Erikstad, K. E., Sandvik, H., Barrett, R. T., Reiertsen, T. K., Anker-Nilssen, T., Hodges, K. I., & Bader, J. (2015). There is more to climate than the North Atlantic Oscillation: A new perspective from climate dynamics to explain the variability in population growth rates of a long-lived seabird. *Frontiers in Ecology and Evolution*, *3*(APR). https://doi.org/10.3389/fevo.2015.00043

Messier, F. (1991). The significance of limiting and regulating factors on the demography of moose and white-tailed deer. *The Journal of Animal Ecology*, *60*(2), 377. https://doi.org/10.2307/5285

Miles, W. T. S., Mavor, R., Riddiford, N. J., Harvey, P. V., Riddington, R., Shaw, D. N., Parnaby, D., & Reid, J. M. (2015). Decline in an Atlantic Puffin population: Evaluation of magnitude and mechanisms. *PLoS ONE*, *10*(7), 1–25. https://doi.org/10.1371/journal.pone.0131527

Miller-Rushing, A. J., Høye, T. T., Inouye, D. W., & Post, E. (2010). The effects of phenological mismatches on demography. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1555), 3177–3186. https://doi.org/10.1098/rstb.2010.0148

Murphy, H. M., Adamack, A. T., & Cyr, F. (2021). Identifying possible drivers of the abrupt and persistent delay in capelin spawning timing following the 1991 stock collapse in Newfoundland, Canada. *ICES Journal of Marine Science*, *78*(8), 2709–2723. https://doi.org/10.1093/icesjms/fsab144

Nakashima, B. S., & Wheeler, J. P. (2002). Capelin (*Mallotus villosus*) spawning behaviour in Newfoundland waters - The interaction between beach and demersal spawning. *ICES Journal of Marine Science*, *59*(5), 909–916. https://doi.org/10.1006/jmsc.2002.1261

Nettleship, David N. (1972). Breeding success of the common puffin (*Fratercula arctica* L.) on different habitats at Great Island, Newfoundland. *Ecological Monographs*, *42*(2), 239–268. https://doi.org/10.2307/1942264

Newell, M. A., Harris, M. P., Wanless, S., Burthe, S., Bogdanova, M., Gunn, C. M., & Daunt, F. (2022). *The Isle of May long-term study (IMLOTS) seabird annual breeding success 1982-2021*. NERC EDS Environmental Information Data Centre. https://doi.org/https://doi.org/10.5285/3bf9b65d-727c-4b46-9db6-c0814dc5895f

Newman, J., Fletcher, D., Moller, H., Bragg, C., Scott, D., & McKechnie, S. (2009). Estimates of productivity and detection probabilities of breeding attempts in the sooty shearwater (*Puffinus griseus*), a burrow-nesting petrel. *Wildlife Research*, *36*(2), 159–168. https://doi.org/10.1071/WR06074

Newman, L., Blockley, F., Hewitt, J., & Wood, M. J. (2021). *Seabird monitoring on Skomer Island in 2021*.

Orben, R. A., Fleishman, A. B., Borker, A. L., Bridgeland, W., Gladics, A. J., Porquez, J., Sanzenbacher, P., Stephensen, S. W., Swift, R., McKown, M. W., & Suryan, R. M. (2019). Comparing imaging, acoustics, and radar to monitor Leach's storm-petrel colonies. *PeerJ*, *7*. https://doi.org/10.7717/peerj.6721

Petalas, C., Lazarus, T., Lavoie, R. A., Elliott, K. H., & Guigueno, M. F. (2021). Foraging niche partitioning in sympatric seabird populations. *Scientific Reports*, *11*(1), 1–12. https://doi.org/10.1038/s41598-021-81583-z

Petrie, B., & Anderson, C. (1983). Circulation on the Newfoundland continental shelf. *Atmosphere - Ocean*, *21*(2), 207–226. https://doi.org/10.1080/07055900.1983.9649165

Pianka, E. R. (2008). Optimal Reproductive Tactics. In S. Jorgensen & B. Fath (Eds.), *Encyclopedia of Ecology* (pp. 2567–2572). Elsevier. https://doi.org/10.1016/B978-008045405-4.00841-7

Piatt, J. F., Harding, A. M. A., Shultz, M., Speckman, S. G., Van Pelt, T. I., Drew, G. S., & Kettle, A. B. (2007). Seabirds as indicators of marine food supplies: Cairns revisited. *Marine Ecology Progress Series*, *352*(1987), 221–234. https://doi.org/10.3354/meps07078

Piatt, J. F., Sydeman, W. J., & Wiese, F. (2007). Introduction: A modern role for seabirds as indicators. *Marine Ecology Progress Series*, *352*, 199–204. https://doi.org/10.3354/meps07070

Priddel, D., Carlile, N., Fullagar, P., Hutton, I., & O'Neill, L. (2006). Decline in the distribution and abundance of flesh-footed shearwaters (*Puffinus carneipes*) on Lord Howe Island, Australia. *Biological Conservation*, *128*(3), 412–424. https://doi.org/10.1016/j.biocon.2005.10.007

Ratcliffe, N., Vaughan, D., Whyte, C., & Shepherd, M. (1998). Development of playback census methods for storm petrels *Hydrobates pelagicus*. *Bird Study*, *45*(3), 302–312. https://doi.org/10.1080/00063659809461101

RCoreTeam. (2021). A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. In *Nature* (Vol. 333, Issue 6176, pp. 852–855).

Reed, T. E., Kruuk, L. E. B., Wanless, S., Frederiksen, M., Cunningham, E. J. A., & Harris, M. P. (2008). Reproductive senescence in a long-lived seabird: Rates of decline in late-life performance are associated with varying costs of early reproduction. *American Naturalist*, *171*(2). https://doi.org/10.1086/524957

Regehr, H. M., & Rodway, M. S. (1999). Seabird breeding performance during two years of delayed capelin arrival in the Northwest Atlantic : A multi-species comparison. *Waterbirds*, *22*(1), 60–67.

Regular, P. M., Hedd, A., Montevecchi, W. A., Robertson, G. J., Storey, A. E., & Walsh, C. J. (2014). Why timing is everything: Energetic costs and reproductive consequences of resource mismatch for a chick-rearing seabird. *Ecosphere*, *5*(12), 1–13. https://doi.org/10.1890/ES14-00182.1

Regular, P. M., Shuhood, F., Power, T., Montevecchi, W. A., Robertson, G. J., Ballam, D., Piatt, J. F., & Nakashima, B. (2009). Murres, capelin and ocean climate: Inter-annual associations across a decadal shift. *Environmental Monitoring and Assessment*, *156*(1–4), 293–302. https://doi.org/10.1007/s10661-008-0484-2

Regular, P., Montevecchi, W., Hedd, A., Robertson, G., & Wilhelm, S. (2013). Canadian fishery closures provide a largescale test of the impact of gillnet bycatch on seabird populations. *Biology Letters*, *9*(4). https://doi.org/10.1098/rsbl.2013.0088

Rexer-Huber, K., Parker, G. C., Ryan, P. G., & Cuthbert, R. J. (2014). Burrow occupancy and population size in the Atlantic Petrel *Pterodroma incerta*: A comparison of methods. *Marine Ornithology*, *42*(2), 137–141.

Ricklefs, R. F. (1973). Fecundity, Mortality, and Avian Demography. In S. Farmer (Ed.), *Breeding Biology of Birds* (pp. 366–435). National Academy of Sciences, Washington D.C., USA.

Robertson, G. J., Wilhelm, S. I., & Taylor, P. A. (2004). *Population size and trends of seabirds breeding on Gull and Great Islands, Witless Bay Islands Ecological Reserve, Newfoundland up to 2003*. *418*, viii–45. https://doi.org/10.13140/RG.2.2.25229.95200

Rodway, M. S. (1994). Intracolony Variation in Breeding Success of Atlantic Puffins: An Application of Habitat Selection Theory. In *MSc thesis. Memorial University of Newfoundland*. Memorial Unviersity of Newfoundland.

Rodway, M. S., Chardine, J. W., & Montevecchi, W. A. (1998). Intra-colony variation in breeding performance of Atlantic Puffins. *Waterbirds*, *21*(2), 171–184. https://doi.org/10.2307/1521904

Rodway, M. S., & Lemon, M. J. F. (2011). Use of permanent plots to monitor trends in burrow-nesting seabird populations in British Columbia. *Marine Ornithology*, *39*(2), 243–253.

Rodway, M. S., Montevecchi, W. A., & Chardine, J. W. (1996a). Effects of investigator disturbance on breeding success of Atlantic puffins *Fratercula arctica*. *Biological Conservation*, *76*(3), 311–319.

Rodway, M. S., Montevecchi, W. A., & Chardine, J. W. (1996b). Effects of investigator disturbance on breeding success of Atlantic puffins Fratercula arctica. *Biological Conservation*, *76*(3), 311–319. https://doi.org/10.1016/0006-3207(94)00118-9

Russell, J. (1999). *Chick diet and nestling condition among Atlantic Puffins at three Northwest Atlatic colonies. M.Sc. thesis.* Memorial University of Newfoundland, St. John's, Newfoundland.

Satre, R., Toresen, R., & Anker-Nilssen, T. (2002). Factors affecting the recruitment variability of the Norwegian spring-spawning herring (*Clupea harengus*L.). *ICES Journal of Marine Science*, *59*(4), 725–736. https://doi.org/10.1006/jmsc.2002.1180

Savard, J. P. L., & Smith, G. E. J. (1985). Comparison of survey techniques for burrow-nesting seabirds. *In: Canadian Wildlife Service Progress Notes No. 151. Environment Canada.*, *151*, 1–7.

Schumann, N., Dann, P., Hoskins, A. J., & Arnould, J. P. Y. (2013). Optimizing survey effort for burrow-nesting seabirds. *Journal of Field Ornithology*, *84*(1), 69–85. https://doi.org/10.1111/jofo.12007

Scott, D., Moller, H., Fletcher, D., Newman, J., Aryal, J., Bragg, C., & Charleton, K. (2009). Predictive habitat modelling to estimate petrel breeding colony sizes: Sooty shearwaters (*Puffinus griseus*) and mottled petrels (*Pterodroma inexpectata*) on Whenua Hou Island. *New Zealand Journal of Zoology*, *36*(3), 291–306. https://doi.org/10.1080/03014220909510156

Shultz, M. T., Piatt, J. F., Harding, A. M. A., Kettle, A. B., & Van Pelt, T. I. (2009). Timing of breeding and reproductive performance in murres and kittiwakes reflect mismatched seasonal prey dynamics. *Marine Ecology Progress Series*, *393*, 247–258. https://doi.org/10.3354/meps08136

Signorell, A. (2022). *DescTools: Tools for descriptive statistics v0.99.41* (R package version 0.99.45). https://cran.r-project.org/web/packages/DescTools/index.html

Spatz, D. R., Newton, K. M., Heinz, R., Tershy, B., Holmes, N. D., Butchart, S. H. M., & Croll, D. A. (2014). The biogeography of globally threatened seabirds and island conservation opportunities. *Conservation Biology*, *28*(5), 1282–1290. https://doi.org/10.1111/cobi.12279

Spear, L. B., Ainley, D. G., Nur, N., & Howell, S. N. G. (1995). Population Size and Factors Affecting At-Sea Distributions of Four Endangered Procellariids in the Tropical Pacific. *The Condor*, *97*(3), 613–638.

Surman, C. A., & Nicholson, L. W. (2009). The good, the bad and the ugly: ENSO-driven oceanographic variability and its influence on seabird diet and reproductive performance at the Houtman Abrolhos, Eastern Indian Ocean. *Marine Ornithology*, *37*(2), 129–138.

Tasker, M., & Furness, R. (2003). Seabirds as Monitors of the Marine Environment. In M. Tasker & R. Furness (Eds.), *International Council for the Exploration of the Sea* (Issue 258).

Thackeray, S. J., Henrys, P. A., Hemming, D., Bell, J. R., Botham, M. S., Burthe, S., Helaouet, P., Johns, D. G., Jones, I. D., Leech, D. I., MacKay, E. B., Massimino, D., Atkinson, S., Bacon, P. J., Brereton, T. M., Carvalho, L., Clutton-Brock, T. H., Duck, C., Edwards, M., … Wanless, S. (2016). Phenological sensitivity to climate across taxa and trophic levels. *Nature*, *535*(7611), 241–245. https://doi.org/10.1038/nature18608

Ventura, F., Lukacs, P. M., Granadeiro, J. P., Matano, R., & Catry, P. (2021). Demographic responses to environmental change of the black-browed albatross, sentinel of the Patagonian Shelf Large Marine Ecosystem. *Marine Ecology Progress Series*, *668*, 107–120. https://doi.org/10.3354/meps13743

Vilar, J. M. G., & Rubi, J. M. (2018). Determinants of population responses to environmental fluctuations. *Scientific Reports*, *8*(1), 1–12. https://doi.org/10.1038/s41598-017-18976-6

Visser, M. E., & Both, C. (2005). Shifts in phenology due to global climate change: The need for a yardstick. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1581), 2561–2569. https://doi.org/10.1098/rspb.2005.3356

Walsh, P., Halley, D., Harris, M., Del Nevo, A., Sim, I., & Tasker, M. (1995). Seabird monitoring handbook for Britain and Ireland. *JNCC/RSPB/ITE/Seabird Group, Peterborough.*

Walther, G., Post, E., Convey, P., Menzel, A., Parmesank, C., Beebee, T. J. C., Fromentin, J., I, O. H., & Bairlein, F. (2002). Ecological response to recent climate change. *Nature*, *416*, 389–395.

Wanless, R. M., Angel, A., Cuthbert, R. J., Hilton, G. M., & Ryan, P. G. (2007). Can predation by invasive mice drive seabird extinctions? *Biology Letters*, *3*(3), 241–244. https://doi.org/10.1098/rsbl.2007.0120

Wauters, L. A., & Lens, L. (1995). Effects of food availability and density on red squirrel (*Sciurus vulgaris*) reproduction. *Ecology*, *76*(8), 2460–2469. https://doi.org/10.2307/2265820

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

Whitehead, A. L., Lyver, P. O. B., Jones, C. J., Bellingham, P. J., MacLeod, C. J., Coleman, M., Karl, B. J., Drew, K., Pairman, D., Gormley, A. M., & Duncan, R. P. (2014). Establishing accurate baseline estimates of breeding populations of a burrowing seabird, the grey-faced petrel (*Pterodroma macroptera gouldi*) in New Zealand. *Biological Conservation*, *169*, 109–116. https://doi.org/10.1016/j.biocon.2013.11.002

Wilhelm, S. I. (2017). CWS: Waterbird colony database (Atlantic region). Version 1. *OBIS Canada Digital Collections*. https://doi.org/https://doi.org/10.15468/vlrjfx accessed via GBIF.org on 2022-01-12.

Wilhelm, S. I., Dooley, S. M., Corbett, E. P., Fitzsimmons, M. G., Ryan, P. C., & Robertson, G. J. (2021). Effects of land-based light pollution on two species of burrow-nesting seabirds in Newfoundland and Labrador , Canada Effets de la pollution lumineuse terrestre sur deux espèces d ’ oiseaux marins nichant dans des terriers. *Avian Conservation and Ecology*, *16*(1), 12.

Wilhelm, S. I., Mailhiot, J., Arany, J., Chardine, J. W., Robertson, G. J., & Ryan, P. C. (2015). Update and trends of three important seabird populations in the western North Atlantic using a geographic information system approach. *Marine Ornithology*, *43*(2), 211–222.

Wilhelm, S. I., Walsh, C. J., & Storey, A. E. (2008). Time budgets of common murres vary in relation to changes in inshore capelin availability. *Condor*, *110*(2), 316–324. https://doi.org/10.1525/cond.2008.8440

Winkler, D. (2016). Breeding biology of birds. In I Lovette & J. Fitzpatrick (Eds.), *Handbook of bird biology* (Third Edit, pp. 407–452).

Appendix

S 1. Number of 3 by 3-meter grids deployed in the plot area. The content of every burrow within the grid was assessed and classified (see text for details).

Colony	**Plot**	**Year**	**Grids deployed**
Great Island	A	2021	4
		2022	2
	B	2021	4
		2022	2
	C	2021	4
		2022	6
	D	2021	4
		2022	3
Gull Island	A	2021	9
		2022	3
	B	2021	5
		2022	7
Puffin Island	A	2021	7
		2022	5
Bakeapple Island	A	2021	10
		2022	4
	B	2021	6
		2022	4
Elliston Point	A	2021	6
		2022	9

S 2. Summary of the visits conducted for 2021-2022 in the different colonies included in this study.

Colony name	Year	Visit		
		Laying success	**Phenology**	**Productivity**
Gull Island	2021	June 9	July 9	July 29
	2022	June 13	July 1	August 3
Great Island	2021	June 13	June 30	July 28
	2022	June 16	July 4	August 4
Elliston Point	2021	NA	NA	July 24
	2022	June 11	July 26	August 14
Bakeapple Island	2021	July 3	NA	August 7
	2022	June 29	July 27	August 14
Puffin Island	2021	July 3	NA	August 7
	2022	June 29	July	August 14

S 3. Summary of available historical data used. Number of burrows used to calculate the different parameters, and the source, are included.

Colony	Year	Breeding parameter	Number of burrows	Reference
Gull Island	1984	Laying success	676	Robertson, unpublish.
	1998	Fledging success	84	Baillie & Jones, 2003
	2012	Laying success	338	Wilhelm, unpub. data
Great Island	1992	Laying success	179	(Rodway, 1994)
		Fledging success	179	(Rodway, 1994)
	1993	Laying success	194	(Rodway, 1994)
		Fledging success	384	(Rodway, 1994)
	2011	Laying success	423	Wilhelm, unpub. data
Bakeapple Island	2014	Laying success	32	Wilhelm, unpub. data
Puffin Island	2014	Laying success	42	Wilhelm, unpub. data

S 4. Laying success detection contingency tables, for inter-observer and inter-method comparison.

Hand-grubbing Contingency table

Egg Check		Observer 1			
		Inactive	Active	Unsolved	Total
Observer 2	Inactive	11	0	2	13
	Active	1	15	1	17
	Unsolved	1	3	14	18
	Total	13	18	17	48

Chick check		Observer 1			
		Unoccupied	Occupied	Unsolved	Total
Observer 2	Inactive	12	2	0	14
	Active	0	11	1	12
	Unsolved	7	0	15	22
	Total	19	13	16	48

Burrowscope Contingency table

Egg Check		Observer 1			
		Inactive	Active	Unsolved	Total
Observer 2	Inactive	12	1	2	15
	Active	9	24	0	33
	Unsolved	0	0	0	0
	Total	21	25	2	48

Chick check		Observer 1			
		Inactive	Active	Unsolved	Total
Observer 2	Inactive	19	2	4	25
	Active	12	11	0	23
	Unsolved	0	0	0	0
	Total	31	13	4	48

Inter-method contingency table

Egg check		Handgrubbing		
		Inactive	Active	Total
Burrowscope	Inactive	8	0	8
	Active	5	21	26
	Total	13	21	34

Chick check		Handgrubbing		
		Inactive	Active	Total
Burrowscope	Inactive	15	1	16
	Active	4	13	17
	Total	19	14	33

S 5. Binomial model results for inter-annual general laying success and fledging success differences for the years 2021 and 2022.

	General laying success		**Fledging success**	
	z	**p**	**z**	**p**
Great Island	1.938	0.053	-1.454	0.146
Gull Island	2.838	**0.005**	3.258	**0.001**
Puffin Island	0.625	0.532	-0.292	0.77

S 6. Binomial model results for general burrow laying success differences (Active burrows / Total burrows checked) between colonies in the 2021 (significance level set at 0.05). Elliston Point data not available.

	Gull Island		**Bakeapple Island**		**Puffin Island**	
	z	**p**	**z**	**p**	**z**	**p**
Great Island	0.373	0.709	-1.893	0.058	1.705	0.088
Gull Island	--	--	-1.394	0.163	1.353	0.176
Bakeapple island			--	--	0.235	0.814

S 7. Binomial model results for general burrow laying success differences (Active burrows / Total burrows checked) between colonies in the 2022 (significance level set at 0.05). Bakeapple data excluded from the analyses.

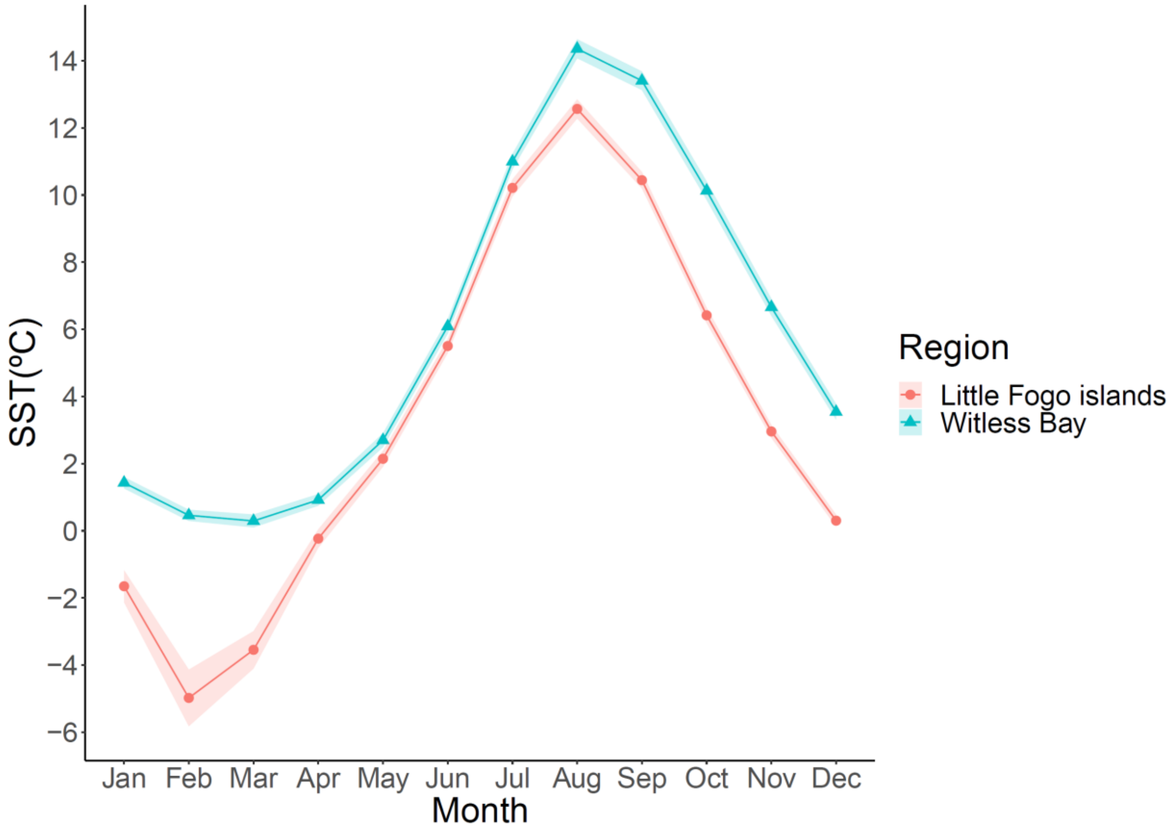
	Gull Island		**Puffin Island**		**Elliston Point**	
	z	**p**	**z**	**p**	**z**	**p**
Great Island	2.123	0.084	1.142	0.253	-1.684	0.092
Gull Island	--	--	-0.289	0.773	0.396	0.692
Puffin Island			--	--	0.011	0.991

S 8 Binomial model results on Bakeapple Island corrected and uncorrected laying success data. Changes in significance are presented in bold (significance set at 0.05).

	Uncorrected				**Corrected**			
	2021		**2022**		**2021**		**2022**	
	z	**p**	**z**	**p**	**z**	**p**	**z**	**p**
2014	4.478	**<0.001**	0.598	0.550	3.98	**<0.001**	0.057	0.9548
2021	--	--	-3.711	**<0.001**	--	--	-3.711	**0.002**

S 9. Binomial model results on Puffin Island corrected and uncorrected laying success data. Changes in significance are presented in bold (significance set at 0.05).

	Uncorrected				**Corrected**			
	2021		**2022**		**2021**		**2022**	
	z	**p**	**z**	**p**	**z**	**p**	**z**	**p**
2014	1.267	0.205	1.409	0.159	0.271	0.7863	0.55	0.582
2021	--	--	0.350	0.726	--	--	0.35	0.726



S 10. Average sea surface temperatures (1948-2022) with 95 % confidence interval (shadowed area) in the adjacent 2 degrees of oceanic waters from both Little Fogo islands and Witless Bay. Elliston Point data not available. Data retrieved from the National Center for Environmental Prediction (NCEP) and the National Center for Atmospheric Research. (NCAR, https://psl.noaa.gov/cgiin/data/timeseries/timeseries1.pl, Kalnay et al., 1996).