Estimating breeding success in Newfoundland Atlantic puffin colonies: A methodological comparison

Raul Zabala Belenguer[1], Katja H. Kochvar[1,2], Antoine Morel[1], Amy C. Wilson[1], Sabina I. Wilhelm[3], Pierre-Paul Bitton[1,*]

[1]Cognitive and Behavioural Ecology, Memorial University of Newfoundland and Labrador, St. John´s, Newfoundland and Labrador, Canada

[2]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey, USA

[3]Canadian Wildlife Service, Environment and Climate Change Canada.

*Corresponding author: Pierre-Paul Bitton, Department of Psychology, Cognitive and Behavioural, Ecology, Memorial University of Newfoundland and Labrador, 232 Elizabeth Avenue, St. John's, NL, A1B 3X9, Canada. Email: pbitton@mun.ca

Competing interests: The authors declare no competing interests

Data availability: Upon acceptance, all data and scripts will be made available on an Open Access site, free of charge

ABSTRACT

Atlantic puffin (*Fratercula arctica*) colonies in the Eastern Atlantic have been experiencing decades of declining numbers, trends revealed by the presence of breeding success monitoring programs. Collecting such data is challenging because puffins have their nest inside burrows, which are usually assessed by hand or burrowscopes. However, it is not clear how comparable the results from these two techniques are. Here, we created a field realistic double-blind experiment to compare the performance of these two assessment methods on Great Island, Newfoundland, Canada. We also surveyed the breeding success of five colonies across the province in 2021 and 2022 and compared them to historical data. Our experiment reveals that handgrubbing and the use of a burrowscope do not generate comparable results and that survey method and observer identity influence breeding success survey outcomes. Burrowscopes generate the most accurate assessments, but they are less consistent between observers than handgrubbing when detecting eggs and chicks. Nevertheless, our results together with the available historical data show successful breeding success across the province, higher than other large Atlantic puffin colonies in the East Atlantic. Our study provides evidence that Newfoundland hosts the largest stable/growing Atlantic puffin colonies in the world for this declining species.

Keywords: Population, Seabird, Update, Double observer, Burrow, Nest

INTRODUCTION

Seabirds are one of the fastest declining groups of birds, with 29 % of species considered at least “Threatened” by the International Union for Conservation of Nature (Spatz *et al.*, 2014). Human-mediated factors are largely contributing to this decline (Croxall, Trathan and Murphy 2002), as they can have severe effects on adult survival, recruitment of new individuals, colony productivity, and immigration/emigration dynamics (Tasker and Furness 2003). To assess seabird population trends, longitudinal demographic studies are employed to regularly estimate the size of breeding populations and their fledging success (Croxall and Rothery 1991). However, most seabird population assessments are difficult due to their breeding biology. Seabirds usually nest in places with difficult access, such as remote islands or steep cliffs, and logistical challenges prevents proper evaluation of all relevant demographic parameters (Tasker and Furness 2003). This is accentuated in burrow-nesting seabirds, which represent about a third of all seabird species (Del Hoyo Elliott and Sargantal 1992).

A seabird burrow is defined as actively breeding when it contains eggs or chicks (Bled, Royle and Cam, 2011). Traditionally, researchers insert their arm in burrows to assess burrow contents, a technique known as handgrubbing (Ambagis 2004). This method has limitations in burrows where incubation chambers are deeper than an arm’s reach (Lavers Hutton and Bond 2019), or when the shape of the tunnel precludes reaching the nest. It can cause significant burrow damage (Ambagis 2004) and affect breeding success (Rodway et al. 1996), which can lead to unreliable breeding estimates (Rexer-Huber et al. 2014). To prevent researcher-induced disturbances, 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 (Ratcliffe et al. 1998, Orben et al. 2019) or camera traps (Bird et al. 2021). Although these techniques are relatively non-invasive, they only provide circumstantial evidence on the presence of an egg or chick. In the 1990s, burrowscopes were introduced as tools for visually assessing burrow contents (Lyver et al. 1998). These devices consist of an infrared camera housed at the tip of a hose connected to a monitor or VR headset. The contents are assessed without coming into direct contact with the animal, potentially minimizing disturbances (Carey 2009). Detection rates for these devices are also not perfect; empirical evidence shows burrowscopes can underestimate breeding success (i.e., when compared to digging burrows; Hamilton 2000, McKechnie et al. 2007). Factors influencing the accuracy of occupation estimates include handler experience, burrow complexity, weather, and hardware limitations (Lyver et al. 1998).

Perhaps the biggest issue is that surveying techniques (i.e., playback calls, handgrubbing, presence of droppings at the burrow entrance) can differ in their detection rates (Ambagis 2004, Lavers Hutton and Bond 2019). If not accounted for, data obtained from one surveying method may not be comparable to data from other methods due to differences in sensitivity, detection accuracy, or detection uncertainty. Moreover, researcher individual expertise is another source of detection estimates discrepancies. To provide accurate estimates, method-specific variation and individual variation should be assessed prior to data collection.

Atlantic puffins (*Fratercula arctica*) are burrow nesting birds considered “Vulnerable” (BirdLife International 2018), but their status is heavily skewed by large declines in the European populations (Kersten et al. 2021). The largest European colonies found in Iceland and Norway have seen their numbers plummet during the past decades mainly due to continuous poor breeding success (Anker-Nilssen et al. 2021, Hansen et al. 2021). This fate is not mirrored in Western-Atlantic populations, where the largest colonies, containing an estimated 900 000 breeding pairs (Lowther et al. 2020), have seen their numbers stable or even increasing (Wilhelm et al. 2015; BirdLife International, 2018). However, with some exceptions (Diamond 2021, Major et al. 2021), most breeding success data from North American colonies come from cross-sectional population surveys where egg or chick fate was not followed through the season. This provides incomplete information on fledging success and productivity (Robertson et al. 2004, Wilhelm et al. 2015), mostly due to the inaccessibility of the colonies and their sensitivity to researcher disturbance. Evidence from Newfoundland colonies demonstrates that fledging success in burrows assessed weekly by handgrubbing can be reduced by as much as 52 % (Rodway et al. 1996, Baillie 2001). This evidence shows that handgrubbing not only can yield imperfect estimates due to technical constraints but can disrupt breeding. A better understanding of how survey methods affect breeding success estimates is necessary to solve trade-offs between data accuracy and impact on the studied burrows.

In this study, we present a small-scale field realistic study to evaluate the consistency and accuracy between two of the most widespread burrow content evaluation methods in Atlantic puffins: burrowscopes and handgrubbing. We also provide updated breeding success from several key Newfoundland Atlantic puffin colonies across a latitudinal range.

METHODS

Method detection reliability

We designed a field relevant experiment to evaluate the consistency between raters and methods on breeding estimates. The experiment was conducted on Great Island (Witless Bay Ecological Reserve, Newfoundland, Canada), during the summer of 2022. Twelve burrows were randomly selected in four different plots within puffin habitat. These 48 burrows were selected by a researcher that did not otherwise participate in the experiment. The content of every identified burrow was then assessed by four different observers (maximum time ~ 1 minute each), two by handgrubbing, and two by burrowscope (EMS2021 Gopher Tortoise Camera System with infrared detection, Environmental Management Services, Canton, Georgia, USA). The same burrows were evaluated twice within the breeding season, once during the incubation period (June 17th 2022) and once during chick rearing (August 2nd 2022). 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 as active if eggs or chicks were found in the burrow, and inactive if not. In case of inactive burrows, the researcher scored them as 'Solved' when they were able to explore the totality of the burrow, or 'Unsolved' if not (e.g., burrow was too deep, too complex for a proper assessment, or obstacles were in the way). Any burrows classified as Unsolved were excluded from the inter-rater agreement analyses. To compare detection consistency between methods, we combined the results of the two assessments per method, deeming a burrow as active if any of the two observers detected an egg or a chick and discarding any burrows marked as Unsolved by both observers. The design is limited in that it did not compare many observers but is representative of field situations where logistics preclude large survey groups.

To compare the two assessment methods, we calculated inter-rater reliability and inter-method reliability values, which represent the extent to which independent evaluations performed by different observers or methods, respectively, are consistent when scoring a qualitative trait (e.g., burrow content). Both estimates are measured using Cohen´s kappa, a coefficient that ranges from 0 to 1, where 0 indicates agreement equal or less than 50 % (i.e., random chance) and 1 when the agreement is complete (McHugh 2012). Cohen´s kappa square root, known as the coefficient of determination (COD, Table S1), corresponds to the percentage of data collected that is validated based on agreement (McHugh 2012). Statistical analyses were conducted using the "CohenKappa" function included in the "DescTools" R package (Signorell 2022), and associated standard error intervals were calculated following the protocol described in McHugh (2012).

To further evaluate the constraints of handgrubbing when assessing burrow content, we measured the incubation chamber depth in 399 active burrows during burrowcope searches for occupancy assessments. Chambers deapth was measured to the nearest 5 cm based on marks placed along the 7.60 m hose. We approached the tip of the hose as close as possible to the egg / chick without direct contact. The breeding depth was calculated as the distance between the tip of the hose to a vertical line between the ceiling and bottom of the burrow entrance. Any assessments with high uncertainty (e.g., chick displacement within the burrow) were discarded.

Breeding success in Newfoundland Atlantic puffin colonies

Five Atlantic puffin colonies along the east coast of Newfoundland were selected to cover a range of colony sizes and latitudes: 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 S2). Colony breeding success was assessed through surveys of breeding plots randomly set in areas with high density of puffin burrows, where series of monitored burrows were tagged with PVC numbered pegs for content assessment (Table S3). Some evidence shows that such high density areas contain and capture the majority of the variance in breeding success (Rodway & Lemon

2011). An area was considered suitable when the density of burrows was higher than the average obtained from previous whole-island surveys (Wilhelm, unpublished data, 2026).

Burrow contents were evaluated two times per breeding season, during late egg incubation and late chick rearing. Timing for these 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). Burrow content was assessed using a burrowscope with infrared detection (EMS2021 Gopher Tortoise Camera System, Environmental Management Services, Canton, Georgia, USA). Every burrow was classified into two different categories: a) Active, if the burrow had evidence of active breeding (egg or chick present), or b) Inactive, if the burrow had no evidence of breeding (only an adult, or empty).

Three breeding parameter estimates were calculated: 1) laying success, the proportion of tagged burrows containing an egg (burrows containing eggs / total number of tagged burrows); 2) fledging success, the proportion of burrows containing an egg where a chick was also found during the second visit (chicks likely to fledge / total number of eggs in tagged burrows); and 3) productivity, the proportion of burrows that fledged a chick (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 last visit would be a successful fledgling, and indeed, there is evidence that fledging success and productivity are usually determined early after hatching (Barrett 2015, Fayet et al. 2017, Diamond 2021, Hansen 2021, Wilhelm et al. 2021). Nevertheless, the last visit was conducted during late chick rearing but before early fledging period to minimize the probability of counting any early fledglings as failed burrows. During the first visit, a series of random burrows located less than 5 m away from the plot were also checked for eggs to calculate occupancy over a larger area (plot-adjacent burrows containing an egg / total plot-adjacent burrows evaluated).

Binomial regression models (link = logit) were used to evaluate the effect of colony and year on each breeding parameter. 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 type I error using a false discovery rate approach (Benjamini and Hochberg 1995).

This study was performed Atlantic puffin colonies, including a protected one within the Witless Bay Ecological Reserve. Animal ethics were covered by an Animal Use Permit (21-03-PB) issued by Memorial University of Newfoundland Animal Care Committee. All research activities were allowed under a Province of Newfoundland and Labrador scientific research permit (wepr2021-23atpucolouration) and a Migratory Bird Research permit (SC4061) issued by Environment and Climate Change Canada.

RESULTS

Method-specific detection consistency

Incubation chambers were 0.55 m deep on average (N = 399, range = 0.05 – 1.15 m, SD = 0.2 m). Considering that the average arm length in North America is 0.65 ± 0.4 m (Claire et al. 1989), as many as 32 % of the burrows contain incubation chambers inaccessible to handgrubbing assessments (Fig. 1), based on chamber depth alone.

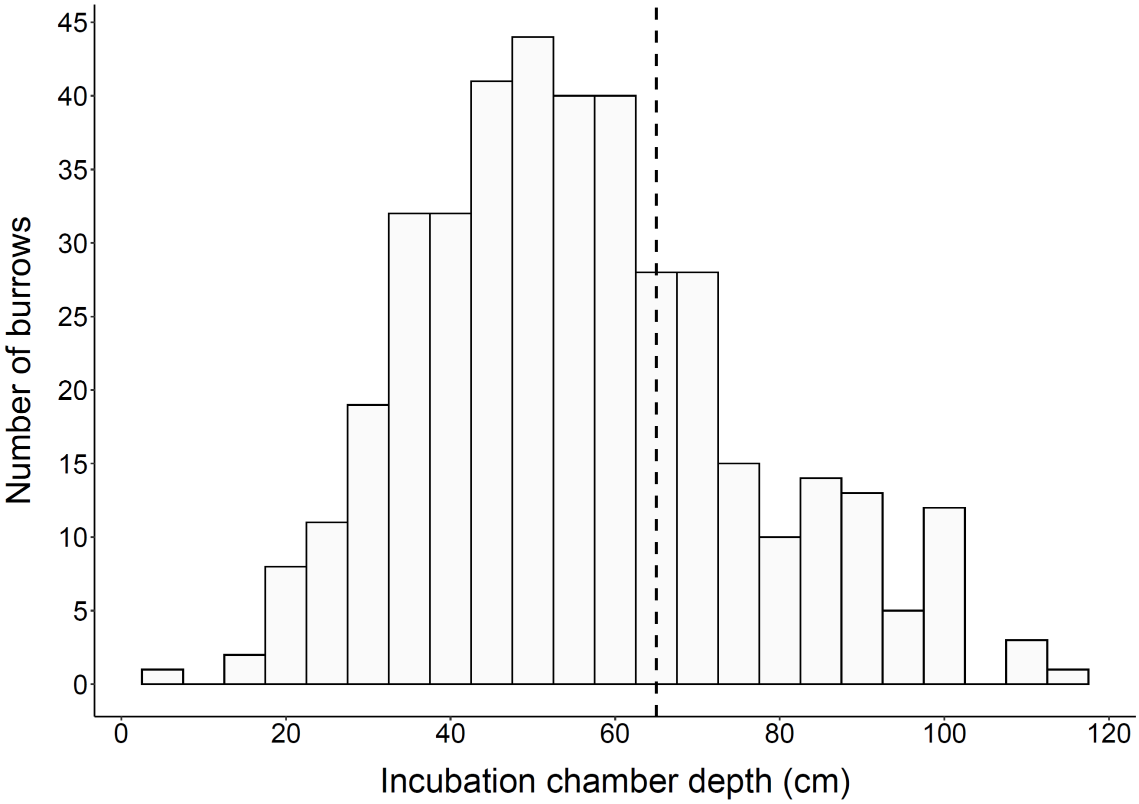

Fig. 1. Incubation chamber depth distribution (where the egg was found) in active Newfoundland puffin burrows (N = 399). Vertical dashed line represents the average arm's reach (see Methods section).

Burrowscope occupancy results were less consistent between observers than handgrubbing (Table 1). No burrows were scored as Unsolved by any of the burrowscope users during egg assessment, but five were described as such during chick assessment. Cohen's kappa for burrowscope detection was 0.51 when detecting eggs and 0.28 for chicks, allowing validation of 25.5 % and 7.9 % of the observations, respectively, due to the score discrepancy between observers. This translated to confidence in the results of roughly 12 / 48 burrow scores during egg assessment, and 3 / 44 for chick assessment. Handgrubbing assessment had 21 Unsolved burrows when detecting eggs, and 24 when detecting chicks. Handgrubbing Cohen's kappa was 0.75 for egg and 0.69 for chick assessment, validating 55.9 % and 48.8 % of the observations. This confirmed the score reliability in 15 / 27 burrows when detecting eggs, and 12 / 24 burrows when detecting chicks.

When comparing between methods, 14 burrows were deemed Unsolved by handgrubbers during egg detection, and 16 during chick detection. Cohen's kappa was 0.66 for egg and 0.35 for chick assessment, making comparable only 43.6 % and 12.6 % of the observations. This made the scores provided by the two different methods comparable in 15 / 34 burrows when detecting eggs, and 4 / 32 when detecting chicks. Nevertheless, burrowscope occupancy rates (N occupied burrows / Total N burrows) were always higher than those obtained by handgrubbing (Table S4).

Table 1. Cohen's kappa and coefficient of determination (COD) values for 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)	-	-
	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)	0.66 (0.12)	43.56 (1.40)
	Chick	0.69 (0.09)	48.80 (0.76)	-	-

Breeding success in Newfoundland puffin colonies

The average occupancy (plot-adjacent burrows with an egg / total plot-adjacent burrows checked) in the assessed colonies for the years 2021 and 2022 was 0.626 (SE = 0.06), and average laying success (plot burrows containing an egg / total burrows in the plot) was 0.937 (SE = 0.040; Table 2). The only extreme value was found on Bakeapple Island in 2022, where the presence of an Arctic fox disrupted breeding in the colony (Belenguer et al. 2024); this datum was therefore excluded from further analyses. Binomial logistic regressions showed no statistical differences in either occupancy or laying success between colonies within the same year (Tables S5 & S6). Average fledging success (chicks likely to fledge / total number of eggs found in the plot) in 2021-2022 was 0.792 (SE = 0.046), and average productivity (plot chicks likely to fledge / total plot burrows) was 0.750 (SE = 0.045; Table 2). Overall, during our study period fledging success was high and consistent except on Bakeapple Island in 2022. Binomial logistic regressions only showed significant lower fledging success on Gull Island compared with the rest of the colonies in 2021, but no differences were found in

2022 (Table S7). Binomial logistic regressions also did not show statistical differences in plot productivity between colonies (Table S8).

We compared our results to available historical data, with the caveat that previous estimates were obtained by different methods such as handgrubbing, presence of feces at the entrance of the burrow, or presence of an adult inside the burrow without the detection of nest content (Robertson, 2004; Wilhelm, personal communications, 2025). We estimate that only 1.8 % (N = 2470) of the burrows checked contained a non-incubating adult, suggesting that our data are comparable to historical values. The average occupancy on Gull Island between 1984 and 2022 was 0.763 (SE = 0.03) and 0.656 (SE = 0.032; Table 3) on Great Island. Average occupancy on Bakeapple and Puffin islands between 2014 to 2022 was 0.370 (SE = 0.072) and 0.651 (SE = 0.07) respectively, although data are very fragmentary (Table 3).

Fledging success were also estimated by handgrubbing. In 2021, 195 / 363 (53.72%) burrows contained chicks without adults. The same was true for 180 / 351 (51.29%) during the year 2022. Historical average fledging success on Gull Island between 1984 and 2019 was 0.671 (SE = 0.021) and the average fledging success for 2021 and 2022 was 0.787 (SE = 0.046). Great Island data was more fragmented; in the years 1992 and 1993, average fledging success was 0.619 (SE = 0.030); for the years included in this study it was 0.732 (SE = 0.057). Other fledging success data from Great Island were obtained using highly invasive methods, and were therefore excluded from this analysis (Nettleship 1972).

Table 2. Breeding parameter estimates obtained during the 2021-2022 breeding seasons through burrowscope assessment: Occupancy (plot-adjacent burrows with an egg / total plot-adjacent burrows checked); 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). The weighted average accounts for the colony sizes relative to the total Newfoundland population. Note that plot burrow laying success and plot productivity could only be calculated for 2022. †Colony affected by the presence of an Arctic fox (see Belenguer et al. 2024), excluded from any posterior analyses.

Colony	**Year**	**Occupancy (SE)**	**Laying Success (SE)**	**Fledging Success (SE)**	**Productivity (SE)**
Great Island	2021	0.623 (0.057)	NA	0.812 (0.037)	NA
	2022	0.647 (0.033)	0.948 (0.025)	0.761 (0.044)	0.721 (0.045)
Gull Island	2021	0.566 (0.060)	NA	0.736 (0.048)	NA
	2022	0.618 (0.057)	0.931 (0.027)	0.864 (0.058)	0.804 (0.043)
Puffin Island	2021	0.681 (0.035)	NA	0.900 (0.042)	NA
	2022	0.715 (0.057)	0.933 (0.042)	0.905 (0.058)	0.844 (0.055)
Bakeapple Island	2021	0.716 (0.050)	NA	0.888 (0.03)	NA
	2022 †	0.309 (0.105)	0.388 (0.069)	0.053 (0.032)	0.02 (0.02)
Elliston Point	2021	0.678 (0.054)	NA	NA	NA
	2022	0.711 (0.098)	0.896 (0.040)	0.814 (0.067)	0.729 (0.064)
Weighted Avg.		**0.626 (0.060)**	**0.937 (0.040)**	**0.792 (0.046)**	**0.750 (0.045)**

Table 3. Historically occupancy (Burrows containing an egg / Total burrows assessed) data per year, number of burrows used to calculate the estimate.

Colony	**Year**	**Burrows**	**Occupancy (SE)**
Great Island	1984[a]	NA	0.843 (0.015)
	1985[a]	319	0.904 (0.011)
	1992[b]	179	0.732 (0.016)
	1993[b]	194	0.871 (0.039)
	2011[c]	423	0.717 (0.036)
	2021[e]	**168**	**0.623 (0.057)**
	2022[e]	**340**	**0.647 (0.033)**
Gull Island	1979[a]	NA	0.441 (NA)
	1984[a]	698	0.798 (0.017)
	1985[a]	716	0.874 (0.011)
	2003[a]	897	0.578 (0.026)
	2012[c]	338	0.722 (0.022)
	2021[e]	**173**	**0.566 (0.060)**
	2022[e]	**102**	**0.616 (0.057)**
Bakeapple Island	2014[d]	62	0.306 (0.051)
	2021[e]	**69**	**0.716 (0.053)**
	2022[e]	**50**	**0.359 (0.114)**
Puffin Island	2014[d]	43	0.558 (0.066)
	2021[e]	**73**	**0.681 (0.041)**
	2022[e]	**74**	**0.715 (0.103)**

a) Robertson, Wilhelm and Taylor 2004; b) Rodway 1996; c) Wilhelm et al. 2015; d) Wilhelm unpublished data, 2026; e) Present study, in bold.

Table 4. Available historical fledging success (chicks ready to fledge / total number of eggs) data per year, number of burrows used to calculate the estimate.

Colony	Year	Burrows	Fledging success (SE)
Great Island	1992[a]	179	0.552 (0.037)
	1993[a]	384	0.685 (0.022)
	2021[f]	**97**	**0.812 (0.037)**
	2022[f]	**92**	**0.761 (0.047)**
Gull Island	1984[b]	43	0.574 (0.073)
	2003[d]	47	0.617 (0.068)
	2004[d]	57	0.632 (0.061)
	2005[d]	94	0.606 (0.048)
	2013[e]	-	0.75 (NA)
	2019[c]	45	0.844 (0.054)
	2021[f]	**95**	**0.736 (0.049)**
	2022[f]	**81**	**0.864 (0.044)**

a) Rodway et al. 1996 b) Rodway 1994, c) Wilhelm et al. 2021, d) Robertson, unpublished data, e) Fayet et al. 2017, f) our study, in bold.

DISCUSSION

Reliability and accuracy of surveying methods

The evaluation of seabird population trends can be affected by imprecise estimates due to methodological constraints, and these imprecisions need to be accounted for when generating population estimates (Johnson 2007; Harris et al. 2020). Our study found that assessment method, observer, and type of nest content (egg or chick) can influence breeding success estimates in Atlantic puffins. Therefore, detection accuracy must be quantified for the surveying methods used, particularly in studies where different surveying methods are employed.

We found evidence that estimates obtained by burrowscope, although imperfect, were consistently higher than those obtained by handgrubbing. Due to the unlikelihood of false positives when visually assessing burrow contents with a burrowscope, these estimates are likely a better representation of true occupancy. Variation in burrow content detection rates between methods has been described in other burrowing species (Ambagis 2004; Sutherland and Dann 2012; Bird et al. 2021). Our research further found that Atlantic puffin breeding estimates varied between observers. These inter-rater discrepancies might be the consequence of several factors, such as different individual capacities, lack of training, or different detection accuracy between devices (Lyver et al. 1998). Burrowscope inter-rater variation was significantly lower during chick assessment compared to egg detection, whereas it remained similar when handgrubbing. This might be explained by the contrast differences between chicks and eggs when visualized through an infra-red camera. Eggs were strikingly white when observed through the infrared lens of the burrowscope, contrasting with the darker background and thus enhancing their likelihood of detection. On the other hand, puffin chicks have dark bodies that blend in with the background, and they can move inside the burrow, further decreasing their likelihood of detection. These detection differences can affect the results of several important population parameters when assessed by burrowscope. A higher detection probability for eggs compared to chicks may result in underestimated fledging success rates, whereas failure to detect eggs would underestimate colony laying success. Also, our results indicate that the presence of an adult in the burrow during the egg incubation period in Witless Bay colonies is a strong indicator of the presence of an egg, at least during high occupancy years. More evidence is needed to extrapolate this observation to low-occupancy years. Our observations highlight the need to consider this masked variation when assessing breeding in Atlantic puffin colonies. *A posteriori* correction is recommended to improve the accuracy and comparability of breeding estimates (Whitehead et al. 2014; Lavers et al. 2019), as it can reduce the uncertainty when correlating them with ecological factors, particularly when they are obtained by different assessment methods.

Even though detection accuracy was higher when assessing Atlantic puffin burrows with burrowscope, trade-offs between accuracy and practicality need to be considered when designing surveys. Burrowscope use is limited by budget restrictions, battery life, access to power during the assessment, damage sensitivity and malfunction, burrow complexity, and/or impossibility to operate them in inclement weather (Lyver et al., 1998). Although handgrubbing lacks these technological and financial challenges and is more consistent between observers, there are still important limitations to consider. We found that the average burrow depth in Newfoundland colonies was 0.55 m; as such, we estimate over 30 % of Newfoundland puffin burrows are out of an average arm's reach. Average burrow depth is even larger in other colonies such as the Isle of May (0.80 – 1.30 m) or the Ainov Islands (1.50 m; Harris and Wanless 2011). This percentage of incubation chambers out of reach is an underestimate since we did not assess other limiting factors such as sharp angles or narrow widths that can also limit the access to the incubation chamber. Moreover, in Newfoundland colonies we observed that Atlantic puffins dug deeper and more complex burrows after handgrubbing assessments. This can ultimately decrease handgrubbing efficacy in burrow content detection when resurveying burrows, particularly for adult and chicks able to move, and increase human hours required for surveys. Furthermore, because handgrubbing involves direct contact with the animal, this method can result in increased stress or injuries to the animals, egg damage, and/or disease transmission (Rodway et al. 1996, Baillie 2001, Carey 2009, Harris and Wanless 2011). All these considerations must be considered *a priori* when deciding on the optimal survey method.

Atlantic puffin breeding success in Newfoundland

The present study provides all the available data (current and historical) on breeding success of Newfoundland puffin colonies, including the largest ones in North America. Breeding success values for the years 2021 and 2022 are higher than any other colony in the species' range, particularly when compared with the declining Eastern Atlantic colonies (Fig. 2, Table S9). With an estimated population size of > 500,000 breeding pairs (Iles et al. 2025, Wilhelm unpublished data, 2026) 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 colonies, Newfoundland fledging success and productivity was 1.68 and 2.77 times higher, respectively, than Icelandic averages between the years 2012-2021 (Hansen 2021). Fledging success in Newfoundland was also 2.07 times higher than the overall average at Norwegian colonies from the years 2008-2021 (Fauchald et al. 2015, Anker-Nilssen et al. 2021). Surveys conducted in 2023 and 2024 in the main Newfoundland colonies (including Gull and Great Islands) show record high breeding pair since systematic surveys started in 1979, suggesting generally good chick production and recruitment (Iles et al. 2025). This makes Newfoundland host to the largest puffin aggregation in the world with successful breeding and stable/increasing numbers.

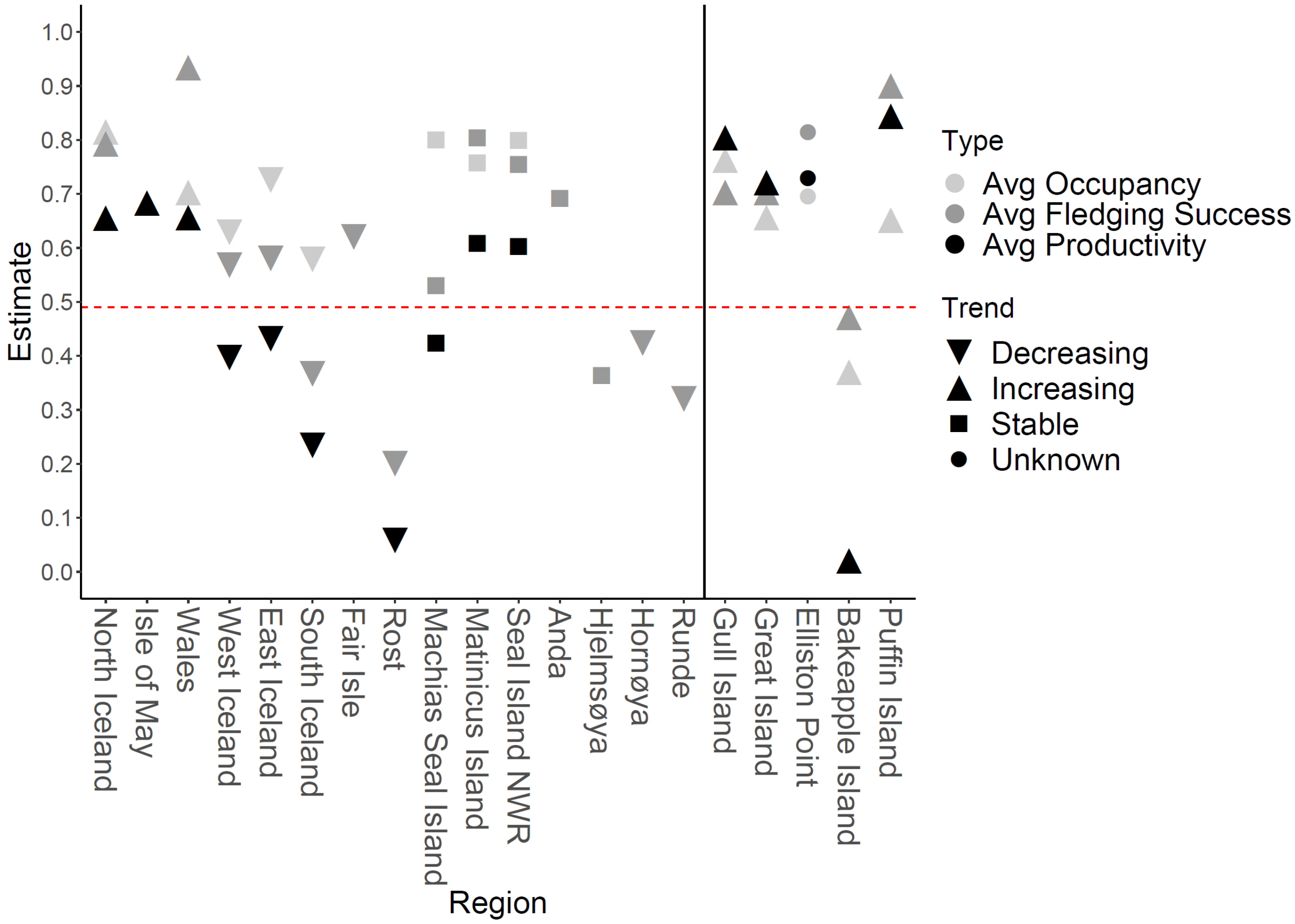

Figure 1. Plot laying success, fledging success, and productivity values from Newfoundland colonies (right of the vertical black line) and other colonies (left of the vertical black line), together with estimated colony population trends (symbol shape). Dotted red line set at 0.49, representing the estimated productivity limit for positive population growth (Hansen 2021). Newfoundland estimates were obtained from the present study, those for the other colonies can be found in Table S9.

The relative success of puffin colonies in Newfoundland seems at odds with their ecological context. Atlantic puffins are central-place foragers during breeding, and their fitness declines as average foraging trip distance from the colony increase (Fayet et al. 2021). Traditionally, capelin (*Mallotus villosus*) has been considered the main prey item consumed 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 to many threats linked with bottom-up processes that crashed led to a population crash (Buren et al. 2014, Murphy et al. 2021). Even so, several studies have 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 diets such as sandlance (Ammodytidae), sculpins (Cottidae), and blenids (Blenidae) (Russell 1999, Baillie and Jones 2003, 2004), although chick rearing and adult body condition is poorer in low-capelin years (Rector et al. 2012; Fitzsimmons et al. 2017) . This plasticity is not found across the whole species' range, as other puffin colonies show stronger dependence on a single prey (Satre et al. 2002; Hansen et al. 2021). This evidence of plasticity during chick food provisioning might make Newfoundland and Labrador puffin colonies more resilient to food availability changes than other colonies, although breeding success variation exists. The years 2011, 2013 and 2016 showed poor chick productivity in Gull and Great Island, both in numbers and body condition, using stranded coastal chicks as a proxy (Wilhelm et al. 2021), although on site breeding success surveys were not conducted for any of these years, pointing to the importance of long-term breeding success surveys to obtain clearer relationships between ecological factors, productivity and eventual recruitment (Visser and Both 2005, Burr et al. 2016, Merkel et al. 2019).

CONCLUSION

In this paper, we have increased the body of evidence suggesting that survey methods and individual researchers add variation in detection of burrowing seabirds, which can lead to imperfect estimates. *A priori* detection assessments might be necessary to quantify this variation. Nevertheless, Newfoundland Atlantic puffin breeding success estimates and population numbers indicate positive colony growth and recruitment during the past decades, despite a suboptimal trophic context, warranting ongoing monitoring of their population status to detect potential changes to these trends.

LITERATURE CITED

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

Anker-Nilssen, T., Aarvak, T., and 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., Moe, B. Reiertsen, T., Bustnes, J., Erikstad, K., Follestad, A., Hanssen, S., Langset, M., Lorentsen, S., Lorentzen, E., and Strøm, H. 2021. Key-site monitoring in Norway 2011-2021, including Svalbard and Jan Mayen. SEAPOP short reports. www.seapop.no/en.

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. and 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. doi: 10.1139/z03-145.

Baillie, S. M. and Jones, I. L. 2004. Response of Atlantic puffins to a decline in Capelin abundance at the Gannet Islands, Labrador. Waterbirds. **27**(1): 102–111. doi: 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. doi: 10.61350/sbj.28.17.

Bates, D., Mächler, M., Bolker, B., and Walker, S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software. **67**(1): 1-48. doi: 10.18637/jss.v067.i01.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B: Statistical Methodology. **57**(1): 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x.

Bird, J., Fuller, R., Pascoe, P., and Shaw, J. 2021. Trialling camera traps to determine occupancy and breeding in burrowing seabirds. Remote Sensing in Ecology and Conservation. 8. 10.1002/rse2.235.

BirdLife International. 2018. Fratercula arctica. The IUCN Red List of Threatened Species 2018: e.T22694927A132581443. http://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T22694927A132581443.en

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

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

Buren, A., Koen-Alonso, M., Pepin, P., Mowbray, F., Nakashima, B., Stenson, G., Ollerhead, N., and Montevecchi, W. 2014. Bottom-Up Regulation of Capelin, a Keystone Forage Species. PloS one. 9. e87589. 10.1371/journal.pone.0087589.

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

Cairns, D. K. Elliot, R. D., Threlfall, W., and Montevecchi, W. A. 1989. A researchers ' guide to Newfoundland seabird colonies. Occasional Papers in Biology No. 14. St. John's, NL: Memorial University of Newfoundland. **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. doi: 10.1080/03014220909510161.

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

Gordon, C., Churchill, T., Clauser, C., Bradtmiller, B., McConville, J. Tebbetts, I., and Walker, R. 1989. Anthropometric Survey of U.S. Personnel: Summary Statistics Interim Report. https://apps.dtic.mil/sti/citations/ADA209600

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

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

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.

Fauchald, P., Anker-Nilssen, T., Barrett, R. T., Bustnes, J.O., Bårdsen, B. J., Christensen-Dalsgaard, S., Descamps, S., Engen, S., Erikstad, K. E., Hanssen, S.A., Lorentsen, S-H., Moe, B., Reiertsen, T. K., Strøm, H., and Systad, G.H. 2015 The status and trends of seabirds breeding in Norway and Svalbard. NINA Report. 1151. 84 pp.

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., Sigurðsson, I. A., Shoji, A., Wanless, S., and 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. doi: 10.1016/j.cub.2017.11.009.

Fayet, A. L., Clucas, G. V., Anker-Nilssen, T., Syposz, M., and 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. doi: 10.1111/1365-2656.13442.

Fitzsimmons, M. G., Rector, M., McKay,. D. W., and Storey, A. E. 2017. High growth and low corticosterone in foodsupplemented Atlantic puffin *Fratercula arctica* chicks under poor foraging conditions. Marine Ecology Progress Series. **565**: 217–226. doi:

10.3354/meps11999.

Grant, P. R. and Nettleship, D. N. 1971. Nesting Habitat Selection by Puffins *Fratercula arctica* L. in Iceland. Ornis Scandinavica, **2**(2): 81-87. doi: 10.2307/3676209.

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. http://doi.org/10.5038/2074-1235.28.1.450

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. and Varpe, Ø. 2021. Centennial relationships between ocean temperature and Atlantic puffin production reveal shifting decennial trends. Global Change Biology. **27**(16): 3753–3764. doi: 10.1111/gcb.15665.

Hansen, E. S. 2021. Puffin population monitoring 2010-2021. Final Report to the Environmental Agency of Iceland. [In Icelandic]

Harris, M. P., Heubeck, M., Bogdanova, M. I., Newell, M. A., Wanless, S., and 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. and Wanless, S. 2011. The Puffin. Edited by T. Poyser and A. D. Poyser. Yale University Press.

Harting, F. 2022. DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.5.0, https://github.com/florianhartig/dharma.

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

JNCC (2021) Seabird Population Trends and Causes of Change: 1986–2019 Report. https://jncc.gov.uk/our-work/smp-report-1986-2019/

Iles, D. T., S. E. Gutowsky, A. M. Calvert, S. I. Wilhelm, J.-F. Rail, A. Hedd, H. L. Major, A. C. Smith, and G. J. Robertson. 2025. Estimating regional trajectories and trends of seabirds from sparse and inconsistent colony counts: case studies from eastern Canada with Leach's Storm-Petrel and Atlantic Puffin. Avian Conservation and Ecology. **20**(2):16. https://doi.org/10.5751/ACE-02931-200216

Johnson, D. H. 2007. Methods of estimating nest success: An historical tour. Studies in Avian Biology. **34**: 1–12.

Kersten, O. Star, B., Leigh, D., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps, S., Erikstad, K., Fitzsimmons, M., Fort, J., Hansen, E., Harris, M, Irestedt, M., Kleven, O., Mallory, M., Jakobsen, K., and Boessenkool, S. 2021. Complex population structure of the Atlantic puffin revealed by whole genome analyses. Communications Biology. **4**(1): 1–12. doi: 10.1038/s42003-021-02415-4.

Lavers, J. L., Hutton, I. and 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**: p. e00579. doi: 10.1016/j.gecco.2019.e00579.

Lowther, P. E., A. W. Diamond, S. W. Kress, G. J. Robertson, K. B. Russell, D. N. Nettleship, G. M. Kirwan, D. A. Christie, C. J. Sharpe, E. Garcia, and P. F. D. Boesman (2020). Atlantic Puffin (*Fratercula arctica*), version 1.0. In Birds of the World (S. M. Billerman, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA.

https://doi.org/10.2173/bow.atlpuf.01

Lyver, P., Hamilton, S., McKenzie, M., Dickson, I., Dooher, M., Broad, T. and Moller, H. 1998. A burrowscope for examining petrel nests in burrows. Conservation Advisory Science Notes No. 209. Department of Conservation, Wellington.

Maine Coastal Islands NWR. 2022. Biological Summary of Islands within Maine Coastal Islands National Wildlife Refuge. Maine Coastal Islands NWR reports. No. 3603. doi: 10.46883/onc.2022.3603.

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

McKechnie, S., Fletcher, D., Moller, H., Scott, D. S., Newman, J., and Bragg, C. 2007. Estimating and Correcting for Bias in Population Assessments of Sooty Shearwaters. Journal of Wildlife Management. **71**(4): 1325–1335. doi: 10.2193/2006-018.

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., Þórarinsson, Þ. L., Wanless, S., and Strøm, H. 2019. Earlier colony arrival but no trend in hatching timing in two congeneric seabirds (*Uria* spp.) across the North Atlantic. Biological Letters. **15**(10): 20190634. doi: 10.1098/rsbl.2019.0634. Epub 2019 Oct 23. PMID: 31640526; PMCID: PMC6832195.

Miles, W. T. S., Mavor, R., Riddiford, N. J., Harvey, P. V., Riddington, R., Shaw, D. N., Parnaby, D., and Reid, J. M. 2015. Decline in an Atlantic Puffin Population: Evaluation of Magnitude and Mechanisms. PLoS ONE 10(7): e0131527. https://doi.org/10.1371/journal.pone.0131527

Murphy, H. M., Adamack, A. T. and 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. doi: 10.1093/icesjms/fsab144.

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

Newell, M.A., Harris, M.P., Wanless, S., Burthe, S., Bogdanova, M., Gunn, C.M., and 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/10.5285/3bf9b65d-727c-4b46-9db6-c0814dc5895f

Newman, L., Blockley, F., Hewitt, J., and 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., and Suryan, R. M. 2019. Comparing imaging, acoustics, and radar to monitor Leach's storm-petrel colonies. PeerJ. **7**:e6721. doi: 10.7717/peerj.6721.

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

Rector, M. E., Kouwenberg, A. L., Wilhelm, S. I., Robertson, G. J., McKay, D. W., Fitzsimmons, M. G., Baker, C. R., Cameron-MacMillan, M. L., Walsh, C. J., and Storey, A. E. 2012. Corticosterone levels of Atlantic puffins vary with breeding stage and sex but are not elevated in poor foraging years. General and Comparative Endocrinology. **178**(2): 408–416. doi: 10.1016/j.ygcen.2012.06.008.

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.

Robertson, G. J., Wilhelm, S. I. and 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. Canadian Wildlife Service Technical Report Series No. 418. Atlantic Region: viii–45. doi: 10.13140/RG.2.2.25229.95200.

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

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

Rodway, M. S., Montevecchi, W. A., and Chardine, J. W. 1996. Effects of investigator disturbance on breeding success of Atlantic puffins Fratercula arctica. Biological Conservation. **76**(3): 311–319. doi: 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., and Anker-Nilssen, T. 2002. Factors affecting the recruitment variability of the Norwegian spring-spawning herring (Clupea harengusL.). ICES Journal of Marine Science. **59**(4): 725–736. doi: 10.1006/jmsc.2002.1180.

Scott, D., Moller, H., Fletcher, D., Newman, J., Jagannath, A., Bragg, C., and 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. doi: 10.1080/03014220909510156.

Signorell, A. 2022. DescTools: Tools for descriptive statistics v0.99.41. Available at: 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., and Croll, D.A. 2014. The biogeography of globally threatened seabirds and island conservation opportunities. Conservation Biology. **28**(5): 1282–1290. doi: 10.1111/cobi.12279.

Spear, L. B., Ainley, D. G., Nur, N., and 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.

Sutherland, D. R., and Dann, P. 2012. Improving the accuracy of population size estimates for burrow-nesting seabirds. Ibis. **154**(3): 488–498. doi: 10.1111/j.1474-919X.2012.01234.x.

Tasker, M., and Furness, R. (2003) Seabirds as Monitors of the Marine Environment, International Council for the Exploration of the Sea. Edited by M. Tasker and R. Furness.

Visser, M. E., and 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. doi: 10.1098/rspb.2005.3356.

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., and 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. doi: 10.1016/j.biocon.2013.11.002.

Wilhelm, S. I., Mailhiot, J., Arany, J., Chardine, J. W., Robertson, G. J., and 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. http://doi.org/10.5038/2074-1235.43.2.1133

Wilhelm, S. I. 2017. CWS: Waterbird colony database (Atlantic region). Version 1', OBIS Canada Digital Collections. doi: 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., and Robertson, G. J. 2021. Effects of land-based light pollution on two species of burrow-nesting seabirds in Newfoundland and Labrador, Canada. Avian Conservation and Ecology. **16**(1):12. https://doi.org/10.5751/ACE-01809-160112

Zabala Belenguer, R., Wilson, A. C., Kochvar, K. H., Wilhelm, S. I., and Bitton, P-P.2024. Effects of a Vagrant Arctic Fox Alopex Lagopus on an Isolated Newfoundland Seabird Community. Marine Ornithology. **52**(1): 73–79.

Table S1. Correspondence between different Cohen's kappa values and the proportion of the reliable data collection due to agreement between evaluations. 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%

Table S2. 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 (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.

Table S3. Permanent plot summary per colony. Indicated number of plots, plot size, number of marked burrows per plot, colony burrow density threshold for an area to be considered as high density (see Methods), and average burrow density on the plot.

Colony name	**Plot**	**Plot size (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.

Table S4. Occupancy (active burrows / total number of burrows) for each observer participating in the study, and for each method using maximum detection rates. Maximum detection rates were obtained from the combination of the detection histories of two observers per method (see Methods).

Assessment	Observer	Occupancy	Method	Occupancy (SE)
Egg	Handgrub 1	0.642	Burrowscope	0.729 (0.064)
	Handgrub 2	0.580		
	Burrowscope 1	0.733	Handgrub	0.618 (0.083)
	Burrowscope 2	0.543		
Chick	Handgrub 1	0.461	Burrowscope	0.510 (0.072)
	Handgrub 2	0.406		
	Burrowscope 1	0.479	Handgrub	0.424 (0.084)
	Burrowscope 2	0.295		

Table s5. Binomial model results for occupancy differences (plot-adjacent burrows with an egg / Total plot-adjacent burrows checked) between colonies in the 2021-2022 breeding seasons (significance level set at 0.05, corrected for Type I error).

	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

	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

Table S6. Binomial model results for plot laying success differences (plot burrows with an egg / Total plot burrows) between colonies during the 2022 breeding season (significance level set at 0.05, corrected for Type I error).

	Gull Island		**Puffin Island**		**Elliston Point**	
	z	**p**	**z**	**p**	**z**	**p**
Great Island	-0.496	0.620	-2.537	0.112	-1.120	0.263
Gull Island	--	--	-2.111	0.352	-0.675	0.500
Puffin Island			--	--	1.267	0.205

Table S7. Binomial model results for fledging success probability (plot chicks likely to fledge / Total eggs in the plot) between colonies during the 2021-2022 breeding seasons (significance level set at 0.05, corrected for Type I error).

Year 2021	**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

Year 2022	**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

Table S8. Productivity (plot chicks likely to fledge / Total plot burrows) binomial model results for the 2022 breeding season (significance level set at 0.05).

	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

Table S9. 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; Machias Seal Island, 1995-2022; Matinicus Island, 2005-2017; Seal Island NWR, 2005-2017).

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[8-9] (Norway)	100 000-250 000[9]	**Stable** Since 2011	NA	0.692 (0.056)	NA
Hjelmsoya[8-9] (Norway)	10,000-50 000[9]	**Stable** Since 2011	NA	0.364 (0.043)	NA
Hornoya[8-9] (Norway)	10,000-50 000[9]	**Decreasing** Since 2011	NA	0.424 (0.067)	NA
Runde[8-9] (Norway)	50 000-100 000[9]	**Decreasing** Since 2011	NA	0.320 (0.051)	NA

Machias Seal I. (Canada)[10-12]	12 555	**Stable** Since 2000	0.8 (0.01)	0.53 (0.04)	0.424 (0.02)
Matinicus Island (USA)[11-12]	563	**Increasing/Stable**	0.757 (0.02)	0.804 (0.04)	0.609 (0.04)
Seal Island NWR (USA)[11-12]	573	**Increasing/Stable**	0.799	0.755 (0.05)	0.603 (0.05)

[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; [10]Major, H. L. , personal communications; [11]Maine Coastal Islands NWR, 2022; [12]Scopel et al., 2019.