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

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

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

Breeding success data in burrowing birds can be obtained using several techniques, but hand search (or handgrubbing) and burrowscopes are two of the most widespread. Breeding success performance can be important in detecting positive or negative population trends; however, it is not clear how accurate, consistent, and 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 in Atlantic puffin (*Fratercula arctica* Linaeus, 1758) burrows on Great Island, Newfoundland, Canada and produced correction factors to ensure the comparability between the data obtained from the two different methods. We complement these results by surveying the breeding success status of five colonies across the province in 2021 and 2022 and comparing them to historical data. Our results suggest that burrowscopes generate the most accurate assessments, but they might be less consistent between observers than handgrubbing when detecting eggs and chicks. Furthermore, our combination of surveys and historical data, corrected for comparability, show high breeding success across the province, better 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: Atlantic puffin, Breeding success, survey method, burrowscope, handgrubbing

INTRODUCTION

Roughly 29 % of seabird species are 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 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 its contents, a technique known as handgrubbing (Ambagis 2004). This method has limitations: incubation chambers are often deeper than an arm’s reach (Lavers Hutton and Bond 2019), it may cause significant burrow damage (Ambagis 2004), affect breeding success (Rodway et al. 1996), and can stress the attending adult, all of 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, both in seabirds and other bird groups: detection of fresh diggings, feces, or pellets at burrow entrances (Conway et al. 2008; Rexer-Huber et al. 2014), CO2 detectors within the burrow (Mondain-Monval & Sharp 2018), use of flashlights in

narrow tunnels (Silva et al. 2022), 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 in seabirds and other bird groups (Lyver et al. 1998). These devices consist of an infrared camera housed at the tip of a hose connected to a monitor or virtual reality headset. The contents are assessed without coming into direct contact with the animal, potentially minimizing disturbances (Carey 2009). This, together with the greater reach of the camera hose and smaller diameter compared to an arm, makes them an interesting alternative to handgrubbing, and they are being used in multiple burrowing species such as burrowing owls *Athene cunicularia* Molina, 1782 (Baladrón 2021); sand martins *Riparia riparia* Linnaeus, 1758 (Mondain-Moval & Sharp 2018); Amazonian motmots *Momotus momota* Linnaeus, 1758; rufous-tailed jacamars *Galbula ruficauda* Cuvier, 1816; and swallow tanagers *Tersina viridis* Illiger, 1811 (Silva et al. 2023), as well as many burrowing seabirds (Hamilton 2000, McKenchie et al. 2007).

Perhaps the biggest issue is that surveying techniques (i.e., playback calls, handgrubbing, burrowscopes, 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. Factors influencing the accuracy of the assesments include handler experience, burrow complexity, weather, and hardware limitations (Lyver et al. 1998). Some studies have tried to compare method-specific detection rates to true occupancy data, but this involved the complete excavation of the assessed burrows (Hamilton 2000, McKechnie et al. 2007). Because of the binomial probability of detection, recent studies have

applied capture-mark-recapture analysis (CMR) to produce reliable estimates, reducing the need for burrow excavation (Newman et al. 2009, Whitehead et al. 2014). Since assessments are usually conducted in teams, individuals can also act as another source of variation in detection estimates. 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 mostly due to the inaccessibility of the colonies and their sensitivity to researcher disturbance (Robertson et al. 2004, Wilhelm et al. 2015). Evidence from Newfoundland colonies demonstrates that fledging success in burrows frequently (i.e., weekly) assessed by handgrubbing can be reduced by as much as 52 % (Rodway et al. 1996, Baillie 2001). This shows that handgrubbing not only can yield imperfect estimates due to technical constraints but can disrupt breeding. This can further make comparisons between different surveying techniques more challenging.

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 and generate a correction factor between the two assessment methods. We then use these correction factors to evaluate the 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, two by handgrubbing, and two by burrowscope (EMS2021 Gopher Tortoise Camera System with infrared detection, Environmental Management Services, Canton, Georgia, USA). The number of observers was limited by availability of trained personnel and burrowscopes at the time of the experiment. 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 and training 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

141 burrow, or ‘Unsolved’ if not (e.g., burrow was too deep, too complex for a proper
142 assessment, or obstacles were in the way). Any burrows classified as Unsolved were
143 excluded from the inter-rater agreement analyses. To compare detection consistency
144 between methods, we combined the results of the two assessments per method, deeming
145 a burrow as active if any of the two observers detected an egg or a chick and discarding
146 any burrows marked as Unsolved by both observers. The design is limited in that it did
147 not compare many observers but is representative of field situations where logistics
148 preclude large survey groups.

149 To compare the two assessment methods, we calculated inter-rater reliability and
150 inter-method reliability values, which represent the extent to which independent
151 evaluations performed by different observers or methods, respectively, are consistent
152 when scoring a qualitative trait (e.g., burrow content). Both estimates are measured
153 using Cohen´s kappa, a coefficient that ranges from 0 to 1, where 0 indicates agreement
154 equal or less than 50 % (i.e., random chance) and 1 when the agreement is complete
155 (McHugh 2012). Cohen´s kappa square root, known as the coefficient of determination
156 (COD, Table A1), corresponds to the percentage of data collected that is validated (total
157 agreement in the scores produced by the two observers) based on agreement (McHugh
158 2012). Statistical analyses were conducted using the “CohenKappa” function included
159 in the “DescTools” R package (Signorell 2022), and associated standard error intervals
160 were calculated following the protocol described in McHugh (2012).

161 To further evaluate the constraints of handgrubbing when assessing burrow
162 content, we measured the incubation chamber depth in 399 active burrows during the
163 burrowcope evaluation to the nearest 5 cm based on marks placed along the 7.6 m hose.
164 We approached the tip of the hose as close as possible to the egg / chick without direct
165 contact. The breeding depth was calculated as the distance between the tip of the hose to

a vertical line between the ceiling and floor of the burrow entrance. Any assessment with high uncertainty (e.g., chick displacement within the burrow) were discarded.

Method-specific detection probability

We estimated the probability of detecting an egg or chick in an occupied burrow with a given surveying method (p) using the presence/absence data to create a set of capture/recapture incubation and chick rearing histories that could be analyzed using mark–recapture modelling (Jones et al. 2003, Whitehead et al. 2014). An initial release occasion that represented egg laying was added at the beginning of every history, resulting in mark-recapture histories composed of five occasions, always starting with 1, the egg laying instance (Whitehead et al. 2014). These encounter histories were fitted to open population Cormack-Jolly-Seber (CJS) models, included in MARK 6.0 (White and Burnham 1999). Goodness-of-fit of the 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. AIC values were then corrected for overdispersion (Q), and small sample size (QAICc). Observer surveys were conducted successively (less than an hour apart). In these instances, survival (θ) was fixed at 1 in all models containing a time-dependent effect (t). This assumes a 100 % chance of survival in surveys conducted less than 1 hour apart. Any inactive burrows, as well as histories with “Unsolved” checks, were omitted from the analysis.

Due to sample size constraints, a different approach was used to estimate the detection accuracy for handgrubbing. Following the suggested methods of Lavers et al. (2019), the burrow occupancy histories yielded by the two observers of each method were merged into one, considering a burrow as active if one of the observers detected

190 breeding contents. The difference of occupancy values between the two methods was
191 then used to obtain the handgrubbing detection accuracy.

192 **Breeding success in Newfoundland Atlantic puffin colonies**

193 Five Atlantic puffin colonies along the east coast of Newfoundland were selected to
194 cover a range of colony sizes and latitudes: Bakeapple and Puffin Island, part of the
195 Little Fogo archipelago; Elliston Point, Bonavista Peninsula; and Great and Gull Island
196 in Witless Bay Ecological Reserve (Table A2). Colony breeding performance was
197 assessed through surveys of breeding plots randomly set in areas with high density of
198 puffin burrows, where series of monitored burrows were tagged with PVC numbered
199 pegs for content assessment (Table A3). Some evidence shows that such high density
200 areas contain and capture the majority of the variance in breeding performance (Rodway
201 & Lemon 2011). An area was considered suitable when the density of burrows was
202 higher than the average obtained from previous whole-island surveys (Wilhelm,
203 unpublished data, 2026).

204 Burrow contents were evaluated two times per breeding season, during late egg
205 incubation and late chick rearing. Timing for these visits was based on bibliographical
206 references (Rodway et al. 1998), personal observations, and expected duration of
207 incubation (42 days; Harris & Wanless 2011) and rearing time (42 days; Grant &
208 Nettleship 1971, Rodway et al. 1998). Burrow content was assessed using a
209 burrowscope with infrared detection (EMS2021 Gopher Tortoise Camera System,
210 Environmental Management Services, Canton, Georgia, USA). Every burrow was
211 classified into two different categories: a) Active, if the burrow had evidence of active
212 breeding (egg or chick present), or b) Inactive, if the burrow had no evidence of
213 breeding (only an adult, or empty).

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

229 Binomial regression models (link = logit) were used to evaluate the effect of
230 colony and year on each breeding parameter. Binomial regressions were conducted in R
231 (R Core Team 2021), using the package "lme4" (Bates *et al.* 2015). Model
232 overdispersion was assessed with the "testDispersion" function, included in the package
233 "DHARMA" (Harting 2022). All pairwise comparisons were corrected for type I error
234 using a false discovery rate approach (Benjamini and Hochberg 1995).

235

236 RESULTS

237 **Method-specific detection consistency**

238 Our occupancy results comparing survey methods showed less consistency
239 between observers when produced by burrowscope, compared to handgrubbing (Table
240 1). Nevertheless, occupancy rates (N occupied burrows / Total N burrows) were always
241 higher when obtained by burrowscope (egg = 0.729, SE = 0.064; chick = 0.51, SE =
242 0.072) than those obtained by handgrubbing (egg = 0.618, SE = 0.083; chick = 0.424,
243 SE = 0.084; Table A4).

244 No burrows were scored as Unsolved by any of the burrowscope users during
245 egg assessment, and 5 during chick assessment, while handgrubbing assessment had 21
246 Unsolved burrows when surveying for eggs, and 24 when surveying for chicks.
247 Burrows classified as Unsolved by one of the two observers could not be used to
248 calculate method-specific Cohen´s kappa. Cohen's kappa for burrowscope detection was
249 0.51 when detecting eggs and 0.28 for chicks, so that the two observers agreed on the
250 status of 25.5% of nests with eggs and only 7.9% of nests with chicks. These
251 percentages translate to certainty in the agreement of the burrowscope scores produced
252 by the two observers in 12 / 48 burrows during egg assessment, and 3 / 43 for the chick
253 assessment. Handgrubbing Cohen's kappa was 0.75 for egg and 0.69 for chick
254 assessment, an agreement between observers for 55.9 % of nests with eggs and 48.8 %
255 of nests with chicks. These results confirm the agreement between handgrubbing
256 observers in 15 / 27 burrows when detecting eggs, and 12 / 24 burrows when detecting
257 chicks.

258 When comparing between methods, 14 burrows were deemed Unsolved by both
259 handgrubbers during egg detection, and 16 during chick detection. These burrows were

discarded for the calculation of Cohen´s kappa between methods. Cohen's kappa results were 0.66 for egg and 0.35 for chick assessment, confirming the agreement of 43.6 % and 12.6 % of the observations between methods. This translates into a confident agreement of the scores provided by the two different methods in 15 / 34 burrows when detecting eggs, but only 4 / 32 when detecting chicks.

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)	-	-

Our incubation chamber depth 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 before considering other types of barriers (Fig. 1).

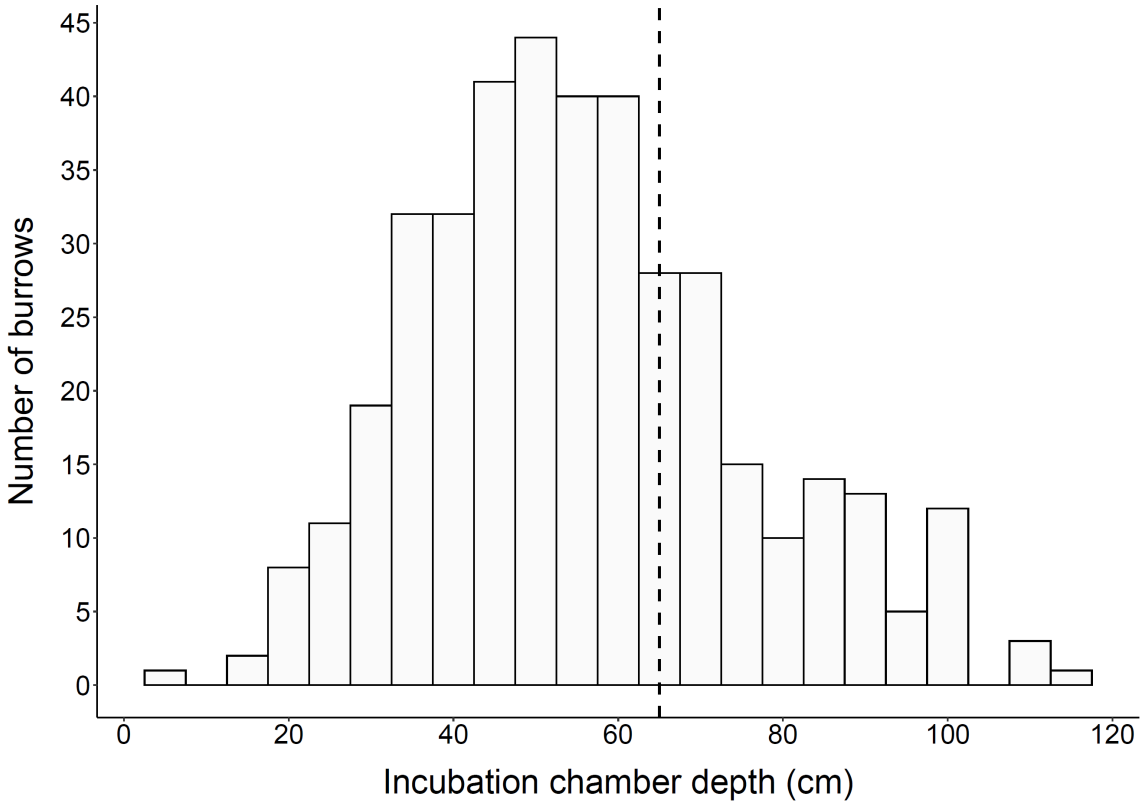


Figure 1. Incubation chamber depth distribution (where the egg was found) in active Newfoundland puffin burrows (N = 399). Vertical dashed line represents the average arm's reach (see Methods section).

Method-specific detection probability

CMR analyses could only be performed on complete capture-recapture histories with at least two positive detection instances in the sequence of five (one representing the egg laying plus a posterior detection). For burrowscope detection, we could use 36 / 48 burrows to estimate probabilities of burrowscope detection. 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), three models could be considered as likely candidates as they fell within 2 AIC of one another (Table A5). The best supported model included survival probability as time dependent, and probability

of detection as time independent. The next likely candidate included both probability of detection and survival as time dependent. The third candidate model included 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. We used the best fitting model as our reference to calculate detection probabilities. The output of the best model gave a detection probability of 0.865 for both eggs and chick assessments (95% CI = 0.679-0.955).

Handgrubbing CMR analyses were not possible because of the presence of unsolved burrows in almost every encounter history (45 unsolved burrows / 192 assessments), leaving only 12 / 48 complete burrow histories usable for analyses. The parametric bootstrap showed an extreme value of overdispersion ($\hat{c}$=3.798), probably due to sample size. To obtain detection estimates, we compared the differences in occupancy estimates by the two methods. When detecting eggs, burrowscope occupancy was 0.729, but 0.618 for handgrubbing. For chick detection, the burrowscope estimate was 0.510, and 0.424 for handgrubbing (Table A4). Handgrubbing detection rate was then calculated as the difference between the burrowscope detection rate (0.865) and the occupancy estimate difference between the two methods. Therefore, handgrubbing detection rate was 0.754 for egg assessment and 0.779 for chicks. This method did not allow for calculation of 95 % confidence intervals. A correction factor to make handgrubbing occupancy estimates comparable to burrowscope was obtained dividing both detection rates; a correction factor of 1.15 for egg detection, and 1.11 for chick detection.

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 *Vulpes lagopus*, Linnaeus, 1758 disrupted breeding in the colony (Zabala 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 A6 & A7). 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 A8). Binomial logistic regressions also did not show statistical differences in plot productivity between colonies (Table A9).

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 Zabala 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	0.616 (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 †	0.359 (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)**

We compared our results to available historical data, with the caveat that previous estimates were obtained by different methods, mainly handgrubbing confirmation of the presence of either a solo adult or an adult plus an egg or chick inside the burrow (Robertson, 2004; Wilhelm, personal communications, 2025). Therefore, as a previous step to validate the comparison between data sources, we 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. In 10 / 122 burrows (8.2%) of the burrows checked the same years, adults were present without chicks. This also allows for the use of correction factor in the data, as the presence of an adult puffin inside the burrow almost guarantees the presence of a chick. Uncorrected average occupancy in Great Island between 1984 and 2022 was 0.763 (SE = 0.03), and corrected average occupancy was 0.848 (Table 3). On Gull Island, occupancy was 0.656 (uncorrected, SE = 0.032), corrected to 0.728. Uncorrected average occupancy in Bakeapple and Puffin islands between 2014 to 2022 was 0.460 (SE = 0.072) and 0.651 (SE = 0.07), while corrected values were 0.475 and 0.679 respectively.

Table 3. Historically occupancy (Burrows containing an egg / Total burrows assessed) data per year, number of burrows used to calculate the estimate. Corrected occupancy obtained by applying correction factors to the data (see Results). Values above 100% occupancy were capped at 1 and are indicated by *. Occupancy estimates obtained in the present study are represented in bold.

Colony	Year	Burrows	Occupancy (SE)	Corrected Occupancy
Great Island	1984[a]	NA	0.843 (0.015)	0.967
	1985[a]	319	0.904 (0.011)	1*
	1992[b]	179	0.732 (0.016)	0.840
	1993[b]	194	0.871 (0.039)	0.999
	2011[c]	423	0.717 (0.036)	0.822
	2021[e]	**168**	**0.623 (0.057)**	
	2022[e]	**340**	**0.647 (0.033)**	
Gull Island	1979[a]	NA	0.441 (NA)	0.506
	1984[a]	698	0.798 (0.017)	0.915
	1985[a]	716	0.874 (0.011)	1*
	2003[a]	897	0.578 (0.026)	0.663
	2012[c]	338	0.722 (0.022)	0.828
	2021[e]	**173**	**0.566 (0.060)**	
	2022[e]	**102**	**0.616 (0.057)**	
Bakeapple Island	2014[d]	62	0.306 (0.051)	0.351
	2021[e]	**69**	**0.716 (0.053)**	
	2022[e]	**50**	**0.359 (0.114)**	
Puffin Island	2014[d]	43	0.558 (0.066)	0.640
	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.

Uncorrected average fledging success between 1992 and 2022 in Great Island was 0.702 (SE = 0.036), and the corrected estimate was 0.737 (Table 4). Other fledging success data from Great Island exist but the values were obtained using highly invasive

366 methods, and therefore we excluded it from this paper (Nettleship 1972). The
367 uncorrected average fledging success in Gull Island between the years 1984 and 2022
368 was 0.702 (SE = 0.057), while the corrected estimate was 0.758 (Table 4).

369

370 **Table 4.** Available historical fledging success (chicks ready to fledge / total number of
371 eggs) data per year, number of burrows used to calculate the estimate. Corrected
372 occupancy obtained by applying correction factors to the data (see Results). Fledging
373 success estimates obtained in the present study are represented in bold.

Colony	Year	Burrows	Fledging success (SE)	Corrected Fledging success
Great Island	1992[a]	179	0.552 (0.037)	0.613
	1993[a]	384	0.685 (0.022)	0.760
	2021[f]	**97**	**0.812 (0.037)**	
	2022[f]	**92**	**0.761 (0.047)**	
Gull Island	1984[b]	43	0.574 (0.073)	0.637
	2003[d]	47	0.617 (0.068)	0.685
	2004[d]	57	0.632 (0.061)	0.702
	2005[d]	94	0.606 (0.048)	0.673
	2013[e]	-	0.750 (NA)	0.833
	2019[c]	45	0.844 (0.054)	0.937
	2021[f]	**95**	**0.736 (0.049)**	
	2022[f]	**81**	**0.864 (0.044)**	

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

376

377

378

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). We suggest that even highly qualified researchers with ample experience in surveying techniques require evaluation of their data collection accuracy. We suggest the test of both equipment and researcher´s accuracy in controlled set ups using artificial burrows with known contents. 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 in Witless Bay colonies is a strong indicator of the presence of an egg or a chick, 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.

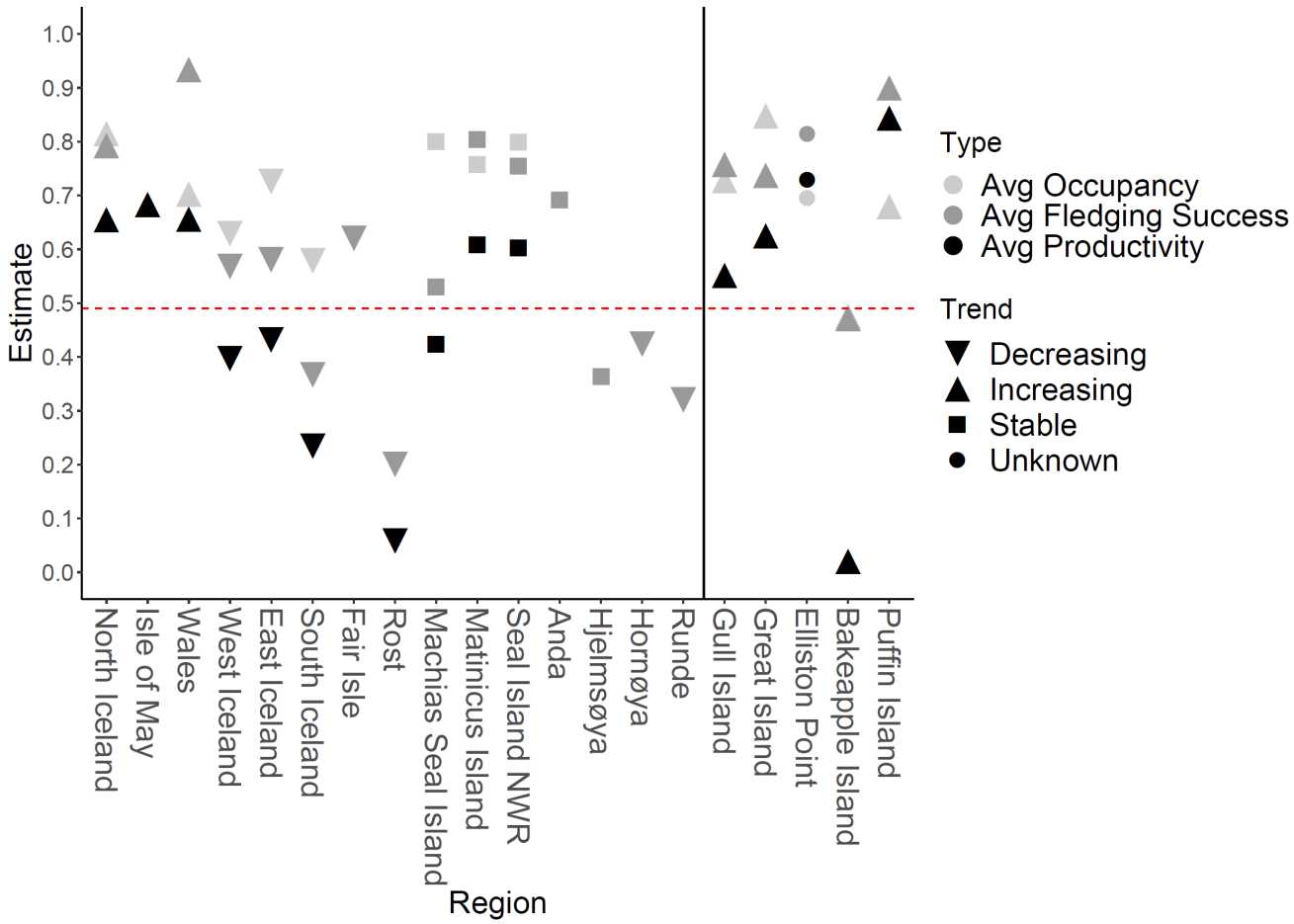
Even though burrowscopes yield estimates closer to true occupancy when assessing Atlantic puffin burrows, trade-offs between accuracy and practicality need to be considered when designing surveys. Burrowscope use is limited by budget restrictions, battery life, access to power during the assessment, damage sensitivity and malfunction, burrow complexity, and/or impossibility to operate them under inclement weather (Lyver et al., 1998). Some of these technical limitations can be more impactful in the estimates over time, as the device's battery and other elements are exposed to the elements and usage damage. Our results also highlight some of the handgrubbing limitations. A total of 45 / 96 (~ 47 %) of the handgrubbing assessments were not

satisfactory (unresolved burrows) to the researcher performing them. This uncertainty is expected, as 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 (Zabala, personal observations). 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. This uncertainty in the assessment of the burrow impedes the utilization of the collected data in CMR analyses, which are used by multitude of long-term datasets to assess Atlantic puffin colony survival (i.e. Breton et al., 2005, Lahoz-Monfort et al., 2011, St. John Glew et al., 2019). 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 further 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 A10). 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 was more than 50 % higher than Icelandic averages, and productivity was more than double than Icelandic averages between the years 2012-2021 (Hansen 2021). This is also true to 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.



473

474 **Figure 2.** Plot laying success, fledging success, and productivity values from
475 Newfoundland colonies (right of the vertical black line) and other colonies (left of the
476 vertical black line), together with estimated colony population trends (symbol shape).
477 Dotted red line set at 0.49, representing the estimated productivity limit for positive
478 population growth (Hansen 2021). Newfoundland estimates were obtained from the
479 present study, those for the other colonies can be found in Table A9.

480 The relative success of puffin colonies in NL seems at odds with their ecological
481 context. Atlantic puffins are central-place foragers during breeding, and their fitness
482 declines as average foraging trip distance from the colony increase (Fayet et al. 2021).
483 Traditionally, capelin (*Mallotus villosus* Muller, 1776) has been considered the main
484 prey item consumed by Atlantic puffins during chick rearing in Western Atlantic
485 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 researchers add variation in detection of burrowing seabirds, which can lead to imperfect estimates. *A priori* detection assessments might be necessary to quantify variation in the data collection by different observers. Even so, our results suggest that even with potential individual discrepancies between observers, burrowscopes are the method that delivers the estimates that are closer to reality, while avoiding undesirable,

harmful effects on the surveyed burrowing seabirds. 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.

517 LITERATURE CITED

518 Ambagis, J. 2004. A comparison of census and monitoring techniques for Leach's
519 Storm Petrel. Waterbirds. **27**(2): 211–215. doi: 10.1675/1524-
520 4695(2004)027[0211:ACOCAM]2.0.CO;2.

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

524 Anker-Nilssen, T., Barrett, R., Christensen-Dalsgaard, S., Dehnhard, N., Descamps, S.,
525 Systad, G., Moe, B. Reiertsen, T., Bustnes, J., Erikstad, K., Follestad, A., Hanssen, S.,
526 Langset, M., Lorentsen, S., Lorentzen, E., and Strøm, H. 2021. Key-site monitoring in
527 Norway 2011-2021, including Svalbard and Jan Mayen. SEAPOP short reports.
528 www.seapop.no/en.

529 Baillie, S. M. 2001. Atlantic Puffin response to changes in capelin abundance in
530 Newfoundland and Labrador: An intercolony and interdecade comparison, MSc thesis.
531 Memorial University of Newfoundland.

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

535 Baillie, S. M. and Jones, I. L. 2004. Response of Atlantic puffins to a decline in Capelin
536 abundance at the Gannet Islands, Labrador. Waterbirds. **27**(1): 102–111. doi:
537 10.1675/1524-4695(2004)027[0102:ROAPTA]2.0.CO;2.

538 Baladrón, A. V., Cavalli, M., Bó, M. S., & Isacch, J. P. 2021. Nest dimensions, burrow-
539 lining, and decoration behavior of burrowing owls in the Pampas. Journal of Raptor

Research. **55(2):** 255-266.

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.

Breton, A. R., Diamond, A.W., and Kress, S. W. 2005. Adult Survival Estimates from Two Atlantic Puffin (Fratercula arctica) Colonies in the Gulf of Maine. The Auk. **122 (3)**: 773-782. https://www.jstor.org/stable/4090584

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

Conway, C. J., Garcia, V., Smith, M. D., & Hughes, K. 2008. Factors affecting detection of Burrowing Owl nests during standardized surveys. The Journal of Wildlife Management. **72(3)**: 688-696.

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

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

590 Croxall, J. P., Trathan, P. N. and Murphy, E. J. 2002. Environmental change and
591 Antarctic seabird populations. Science. **297**(5586): 1510–1514. doi:
592 10.1126/science.1071987.

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

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

600 Freeman, R., Anker-Nilssen, T., Diamond, A., Erikstad, K. E., Fifield, D., Fitzsimmons,
601 M. G., Hansen, E. S., Harris, M. P., Jessopp, M., Kouwenberg, A.-L., Kress, S., Mowat,
602 S., Perrins, C. M., Petersen, A., Petersen, I. K., Reiertsen, T. K., Robertson, G. J.,
603 Shannon, P., Sigurðsson, I. A., Shoji, A., Wanless, S., and Guilford, T. 2017. Ocean-
604 wide drivers of migration strategies and their influence on population breeding
605 performance in a declining seabird. Current Biology. **27**(24): 3871-3878.e3. doi:
606 10.1016/j.cub.2017.11.009.

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

610 2656.13442.

611 Fitzsimmons, M. G., Rector, M., McKay,. D. W., and Storey, A. E. 2017. High growth
612 and low corticosterone in foodsupplemented Atlantic puffin *Fratercula arctica* chicks
613 under poor foraging conditions. Marine Ecology Progress Series. **565**: 217–226. doi:
614 10.3354/meps11999.

615 Grant, P. R. and Nettleship, D. N. 1971. Nesting Habitat Selection by Puffins
616 *Fratercula arctica* L. in Iceland. Ornis Scandinavica, **2**(2): 81-87. doi:
617 10.2307/3676209.

618 Hamilton, S. 2000. How precise and accurate are data obtained using an infra-red scope
619 on burrow-nesting sooty shearwaters *Puffinus griseus*?. Marine Ornithology. **28**(1): 1–6.
620 http://doi.org/10.5038/2074-1235.28.1.450

621 Hansen, E. S., Sandvik, H., Erikstad, K. E., Yoccoz, N. G., Anker-Nilssen, T., Bader, J.,
622 Descamps, S., Hodges, K., Mesquita, M. d. S., Reiertsen, T. K. and Varpe, Ø. 2021.
623 Centennial relationships between ocean temperature and Atlantic puffin production
624 reveal shifting decennial trends. Global Change Biology. **27**(16): 3753–3764. doi:
625 10.1111/gcb.15665.

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

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

632 Harris, M. P. and Wanless, S. 2011. The Puffin. Edited by T. Poyser and A. D. Poyser.

633 Yale University Press.

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

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

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

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

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

648 Kersten, O. Star, B., Leigh, D., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps,
649 S., Erikstad, K., Fitzsimmons, M., Fort, J., Hansen, E., Harris, M, Irestedt, M., Kleven,
650 O., Mallory, M., Jakobsen, K., and Boessenkool, S. 2021. Complex population structure
651 of the Atlantic puffin revealed by whole genome analyses. Communications Biology.
652 **4**(1): 1–12. doi: 10.1038/s42003-021-02415-4.

653 Lahoz-Monfort, J.J., Morgan, B.J.T., Harris, M.P., Wanless, S., Freeman, S.N. 2011. A
654 capture–recapture model for exploring multi-species synchrony in survival. Methods in
655 Ecology and Evolution. **2(1)**: 116-124. https://doi.org/10.1111/j.2041-

210X.2010.00050.x

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.,

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

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

687 Mondain-Monval, T. O. and Sharp, S. P. 2018. Burrow depth, carbon dioxide and
688 reproductive success in Sand Martins *Riparia riparia*. Bird Study. **65(1)**: 123-131.

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

693 Nettleship, D. N. 1972. Breeding success of the common puffin (Fratercula arctica L.)
694 on different habitats at Great Island, Newfoundland. Ecological Monographs. **42**(2):
695 239–268. doi: 10.2307/1942264.

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

700 Newman, J., Fletcher, D., Moller, H., Bragg, C., Scott, and D., McKechnie. 2009.
701 Estimates of productivity and detection probabilities of breeding attempts in the sooty
702 shearwater (*Puffinus griseus*), a burrow-nesting petrel. Wildlife Research. **36**: 159-168.

https://doi.org/10.1071/WR06074

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

726 breeding performance of Atlantic Puffins. Waterbirds. **21**(2): 171–184. doi:
727 10.2307/1521904.

728 Rodway, M. S. and Lemon, M. J. F. 2011. Use of permanent plots to monitor trends in
729 burrow-nesting seabird populations in British Columbia. Marine Ornithology. **39**(2):
730 243–253.

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

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

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

740 Satre, R., Toresen, R., and Anker-Nilssen, T. 2002. Factors affecting the recruitment
741 variability of the Norwegian spring-spawning herring (Clupea harengusL.). ICES
742 Journal of Marine Science. **59**(4): 725–736. doi: 10.1006/jmsc.2002.1180.

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

748 Signorell, A. 2022. DescTools: Tools for descriptive statistics v0.99.41. Available at:

749 https://cran.r-project.org/web/packages/DescTools/index.html.

750 Spatz, D. R., Newton, K. M., Heinz, R., Tershy, B., Holmes, N. D., Butchart, S. H. M.,
751 and Croll, D.A. 2014. The biogeography of globally threatened seabirds and island
752 conservation opportunities. Conservation Biology. **28**(5): 1282–1290. doi:
753 10.1111/cobi.12279.

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

757 St. John Glew, K., Wanless, S., Harris, M.P., Daunt, F., Erikstad, K.E., Strøm, H.,
758 Speakman, J.R., Kürten, B., and Trueman, C.N. 2019. Sympatric Atlantic puffins and
759 razorbills show contrasting responses to adverse marine conditions during winter
760 foraging within the North Sea. Movement Ecology. **7**: 33.
761 https://doi.org/10.1186/s40462-019-0174-4

762 Sutherland, D. R., and Dann, P. 2012. Improving the accuracy of population size
763 estimates for burrow-nesting seabirds. Ibis. **154**(3): 488–498. doi: 10.1111/j.1474-
764 919X.2012.01234.x.

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

768 Visser, M. E., and Both, C. 2005. Shifts in phenology due to global climate change: The
769 need for a yardstick. Proceedings of the Royal Society B: Biological Sciences.
770 **272**(1581): 2561–2569. doi: 10.1098/rspb.2005.3356.

771 Whitehead, A.L., Lyver, P. O. B., Jones, C. J., Bellingham, P. J., MacLeod, C. J.,

772 Coleman, M., Karl, B. J., Drew, K., Pairman, D., Gormley, A. M., and Duncan, R.P.
773 2014. Establishing accurate baseline estimates of breeding populations of a burrowing
774 seabird, the grey-faced petrel (*Pterodroma macroptera gouldi*) in New Zealand.
775 Biological Conservation. **169**: 109–116. doi: 10.1016/j.biocon.2013.11.002.

776 Wilhelm, S. I., Mailhiot, J., Arany, J., Chardine, J. W., Robertson, G. J., and Ryan, P. C.
777 2015. Update and trends of three important seabird populations in the western North
778 Atlantic using a geographic information system approach. Marine Ornithology. **43**(2):
779 211–222. http://doi.org/10.5038/2074-1235.43.2.1133

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

783 Wilhelm, S. I., Dooley, S. M., Corbett, E. P., Fitzsimmons, M. G., Ryan, P. C., and
784 Robertson, G. J. 2021. Effects of land-based light pollution on two species of burrow-
785 nesting seabirds in Newfoundland and Labrador, Canada. Avian Conservation and
786 Ecology. **16**(1):12. https://doi.org/10.5751/ACE-01809-160112

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

790

791

792