

1Farallon Institute, Petaluma, CA, USA
2School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA, USA
3School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY, USA
4U.S. Geological Survey, Alaska Science Center, Anchorage, AK, USA
5Center for Ecosystem Sentinels and Department of Biology, University of Washington, Seattle, WA, USA
6Department of Biology, Boston University, Boston, MA, USA
7Department of Wildlife, Fish, and Conservation Biology, University of California Davis, Davis, CA, USA
8Centre for Ecology and Conservation, University of Exeter, Penryn, Cornwall, UK

*corresponding author: wsydeman@faralloninstitute.org; phone: 707-981-8033
SM1: Modeling the effects of harvest control rules on seabirds with varying foraging strategies

Koehn et al. (2021) use a structured seabird population model to test the impact of various harvest control rules for fisheries targeting sardine and anchovy on two hypothetical seabird species with contrasting life history types: 1) a spatially-restricted, low fecundity specialist (the “restricted” scenario), and 2) a wide-ranging, high fecundity generalist (the “flexible” scenario). The “restricted” seabird more closely matches the life history of the African penguin *Spheniscus demersus*. Koehn et al. found that the “flexible” seabird was moderately robust to constant fishing pressure, while the “restricted” seabird population declined or went extinct when anchovy/sardine were fished at a constant fishing mortality (M). Dietary specialisation (as seen in the African penguin) was one of the most important determinants in population declines for the “restricted” scenario. The impacts of fishing on seabirds in the models of Koehn et al. were substantially reduced when prey were fished with a moderate hockey-stick control rule (note: this is consistent with the default control rules proposed by Pikitch et al. 2012; \( B_{\text{lim}} = 0.4B_0, B_{\text{target}} = 0.8B_0 \) and \( F_{\text{max}} = 0.5M \)), which would see the fisheries closed when biomass falls below 40% of the unfished biomass. “Restricted” populations would maintain a median population size of 80% and 85% (relative to no fishing) when anchovy/sardine, respectively, were fished with the moderate hockey-stick control rule, compared to 0–50% and 4–51% for fishing at a constant rate. While hockey-stick control rules would result in the best outcome with fishing for all seabird scenarios, it also resulted in 13–17% loss of landings compared to a maximum catch control rule, higher variation in catch from year to year, and more years with zero catch than most of the other control rules tested (Koehn et al. 2021).

Currently, the management of the South African sardine *Sardinops sagax* fishery is under “exceptional circumstances” and contains a “Critical Biomass” threshold of 300,000 metric tonnes (de Moor 2018), or ~0.35 of the long-term mean spawner biomass (852,030 tonnes, 1984–2020; Coetzee et al. 2020). When the annual spawner biomass falls below this “Critical Biomass” threshold, Total Allowable Catch (TAC) is reduced to 10,000–65,000 tonnes (de Moor 2018). The minimum directed sardine TAC of 10,000 tonnes was implemented in the 2018 Operational Management Procedure (OMP) to reflect “the expectation that the directed fishery would never be closed completely in practice” (de Moor 2018). Above this “Critical Biomass” threshold, the South African sardine OMP has explicit mechanisms designed to minimise extreme variation in TAC; the maximum proportion by which the TAC could be reduced from one year to the next was 0.2 prior to the implementation of the 2018 OMP and is 0.5 in the 2018 OMP (de Moor 2018). The 2018 OMP also contains a ‘preventative red flag’, which triggers a spatial management component (to offer greater protection to the more productive west coast component of the stock) by which at most 40% of the directed sardine TAC may be taken west of Cape Agulhas, when the biomass of sardine surveyed west of Cape Agulhas is below 100,000 tonnes (de Moor 2018). While this is positive, there is no pre-defined level at which the catch west of Cape Agulhas would be set to zero (de Moor 2018). In other words, two of the key sacrifices that needed to be borne by the fishing industry in Koehn et al.’s (2021) models to maintain the restricted seabird population at 80% of the no fishing level – namely high annual variation in catches and some years of zero catch – are directly guarded against in the South African OMP for sardine to safeguard against socioeconomic concerns.

Additionally, although the sardine spawner biomass in South Africa has only been below 0.4 of the long-term mean six times in the last 20 years, five of those were the last five years (2016–2020, inclusive). In four of the last five years, the spawner biomass was also below the 300,000 tonnes “Critical Biomass” threshold and in three of the last five years the biomass west of Cape Agulhas was below 100,000 tonnes (Coetzee et al. 2020). Even when the November 2018 biomass survey returned an estimated sardine biomass of 90,745 tonnes, the lowest since 1985 (Coetzee 2019) and less than 10% of the long-term (1984–2017) mean of 910,512 tonnes at the time, a TAC of 12,250 tonnes (plus a total allowable
bycatch of 9,400 tonnes with directed anchovy *Engraulis encrasicolus* fishing) was approved by the Department of Forestry, Fisheries and the Environment (it should be noted, however, not using the normal OMP process). Had all of those fish been caught, this would have represented exploitation rates of 14–24%, higher than the long-term average prior to 2007 of ~11% and similar to some of the exploitation rates seen west of Cape Agulhas between 2001 and 2004, just before the eastward shift of spawners occurred (Coetzee et al. 2008). However, the industry only managed to catch 2,145 tonnes during the 2019 fishing season.

In summary, it is not clear that it will be possible or practical to implement harvest control rules that are likely to provide better, timelier or less costly (to industry) solutions to the dire situation of the African penguin population in the short term.

SM2: Breeding success as a key variable in assessing the effects of fisheries closures on African penguins

Butterworth and Ross-Gillespie (2022; hereafter B&R-G) did not agree with our focus on breeding success as the key metric of the experiment. The aim of the original feasibility study referenced by B&R-G was to develop monitoring protocols that would better:

> “inform the design of a possible subsequent experiment to detect the efficacy of using such closed areas to improve the reproductive success and survival rates of African Penguins at the associated colonies” (Seabird Task Group 2007).

B&R-G argue that all variables measured should be considered equally, even when some of these variables are not easily interpretable as directly affecting reproductive success or penguin population growth rates (e.g., some foraging behaviours). When Sherley et al. (2019) and Sherley (2020a) proposed an approach to approximate integrated effects of the fishing closures (while taking estimation precision into account and easily allowing for different statistical weighting), these studies were strongly criticized:

> “How best to combine the results across both different response variables and islands is not an easy question. First though, for reasons given above, combination across islands is inappropriate.... Further, a formal “statistical equal weighting” is questionable. For St Croix, the distributions for condition index and maximum distance are entirely of opposite sign, i.e. completely non-overlapping, so that no probabilistic interpretation can be placed on the combined result presented above – rather if some real impact of closure on penguins is indeed occurring, at least one of the two models assumed for these two analyses must be wrong. There is no simple way to address this matter; discussions are first needed about the relative reliability of the response variables as measures of forage fish availability on penguin reproductive success, together with some quantification of the likely impact of each on penguin dynamics” (Butterworth 2020).

Later, Butterworth and Ross-Gillespie (2021a) argued:

> “…not every variable considered should be equally weighted a priori (i.e. before taking estimation precision into account). In essence, there are only three or fewer “independent” sources of information for each colony, related to chick growth, adult foraging and chick survival”.

The quotes above appear to be at odds with using “an equally weighted average across the available resource variables [sic] types” to “combine these results across islands” (Butterworth and Ross-Gillespie 2021b). In other words, Butterworth (2020) and Butterworth and Ross-Gillespie (2021a) have previously argued against using the kind of approach that lead them (in B&R-G) to conclude that the “impact of future island closure for the two
Western Cape colonies are roughly estimated to be 0% with a 95% PI of [-0.5%; +0.5%], and for the two Eastern Cape colonies to be -0.25% with a 95% PI of [-0.75%; +0.25%].

Moreover, the contradiction between the approach advocated for by B&R-G and their statements elsewhere are problematic for two other reasons. First, B&R-G’s approach explicitly assumes that all variables contribute equal statistical evidence for the impact of future island closures. A glance at their Figure 1 shows that this assumption is invalid. Five of the 24 estimates of change in the penguin population growth rate in their Figure 1 provide at least moderate statistical evidence one way or the other (i.e., the 95% confidence intervals do not overlap zero; Muff et al. 2022a). Of those, four (80%) suggest that the closures will benefit penguin population growth rates. The remaining 19 are so poorly estimated (or too variable) that they provide little or no evidence one way or the other. ‘No evidence’ does not imply that an effect is absent; it only means that the data do not have the power to reveal a potential effect, typically because of large uncertainty (Muff et al. 2022b).

Second, we agree with B&R-G that the relative reliability of all of the foraging metrics as measures of the impact of forage fish availability on penguin reproductive success is unclear. We outline the rationale for this in SM3 below and have also discussed this elsewhere (e.g., Sydeman et al. 2017).

In summary, B&R-G’s suggestion that all variables are considered could be reasonable, but only if a framework to link some (or all) of the foraging metrics to demography can be developed, or a weighting scheme for measured variables can be agreed by all stakeholders, given differences in demographic significance. Unfortunately, a flaw in the experimental design was that it was not practical to measure breeding success at the colonies in the Eastern Cape (particularly at St. Croix), which precludes integration of data from different regions, as suggested by B&R-G.

………………………………………………………………………………………………………………

SM3: The relative importance of foraging variables in analyses
The relative reliability of all of the foraging metrics as measures of the impact of forage fish availability on penguin reproductive success is unclear. The maximum distance that seabirds travel from the colony during a foraging trip has a straight-forward empirical link to resource competition (e.g., Wakefield et al. 2013, Jovani et al. 2016, Weber et al. 2021), which has been linearly linked to reproductive success in Magellanic penguins (Boersma and Rebstock 2009), a closely related species. African penguins have recently been found to linearly respond to local prey abundance (measured by hydroacoustic surveys) around Robben Island (Campbell et al. 2019). In contrast, penguin foraging path length and trip duration appear to show threshold-like or bell-shaped responses to local prey availability (also measured by hydroacoustic surveys) around St Croix and Bird Island (McInnes et al. 2019). Moreover, penguins have delayed gastric emptying, and use a 2-stage strategy where they spend some time foraging for themselves, rebuilding their body resources, and some time foraging for food to bring back to chicks. This split strategy can result in a bimodal distribution in trip duration with short trips being associated with provisioning of the chicks (high feeding frequency and larger meals), while long trips are made when parent body mass is low to enable them to rebuild their reserves (Saraux et al. 2011a). This makes foraging metrics difficult to interpret.

Critically, the distance seabirds may travel from their colony to forage has been linked to localised prey depletion by a purse-seine fishery. By tracking seabirds progressively through their breeding season from the day on which the Peruvian anchovy Engraulis ringens fishery opened for the year, Bertrand et al. (2012) demonstrated that once catches reached a threshold, there was a significant positive relationship between the cumulative catches of the Peruvian anchovy fishery and the maximum distance that Peruvian boobies Sula variegata travelled from their breeding colony, but not an effect on trip duration. The same study also
found a significant positive relationship between the cumulative catches by the fishery and the maximum distance from the colony at which foraging dives occurred (Peruvian boobies are plunge divers). In short, the more the fishery reduced the quantity of prey fish available in the local area, the farther from the colony these breeding seabirds went to obtain food. In other words, while it is clear what travelling further away from a breeding colony means for a seabird, it is difficult to say conclusively what a long path length or a long trip duration mean in terms of access to prey.

In summary, while we agree with B&R-G that the foraging data should not be ignored, with the exception of the maximum distance metric, it is not obvious that these variables provide any information on the direction (sign) of the closure effect.

SM4: On the threshold of population growth selected for management action
B&R-G claim that the management threshold value of 1% (improvement in population growth) has:

“yet to be agreed by the PWG” [Small Pelagic Working Group], was just the result of “loose choices of words in a number of documents submitted to the PWG in recent years” and “was never recommended by the panel”.

However, this statement is misleading.

Ross-Gillespie and Butterworth (2020) define the 1% threshold in their Glossary as:

“a biological value determined externally to the power analysis [that] relates a decision on what is considered to be a biologically meaningful change in the penguin population growth rate” and report in their Table 2 “evidence in the data of a fishing effect that is biologically meaningful with respect to penguin demographics” in respect of this 1% threshold for four penguin metrics.

Ross-Gillespie and Butterworth (2016) state (pg. 3):

“a change of 1% in the penguin data… was used for the base assumption for these analyses… based on recommendation A.2.1 of the [2015] panel report, which states that the threshold “should be computed using a population dynamics model such as the simple model in MARA/WS/DEC15/PengD/BG4, or the penguin population dynamics developed by Robinson et al. (2015) given a management objective of a pre-specified change in population growth rate following elimination of fishing near islands (and assuming that fishing impacts only one population dynamics parameter) (Dunn et al. 2015)”.

Cochrane (2016), in introducing documents from the “Technical Team”§ (TT) formed by the PWG in 2016 to oversee analysis of the data from the Island Closures Experiment, stated that (pg. 2):

“the TT concluded that the previously agreed value of 1% as the pre-specified change in population growth rate should be retained.”

§ The TT consisted of M.O. Bergh, D.S. Butterworth, K.L. Cochrane (chair), T.L. Morris, R.B. Sherley and H. Winker. A. Ross-Gillespie undertook, on behalf of the Team, analyses and tests, under the supervision of D.S. Butterworth (Cochrane 2016).
Additionally, Appendix B of Ross-Gillespie and Butterworth (2021) begins:

“...In analyses of penguin data to date, estimates of the fishing effect parameter $\delta$ have been taken (on earlier advice by the Panel) to be biologically meaningful if $\delta < -0.1$ in cases where the response variable is analysed in log-space. This threshold was based on changes in survival estimates, and corresponds to a value below which (i.e., a $\delta$ value more negative than $-0.1$) analysis indicates that the population growth rate will decrease by 1% or more if fishing occurs in the neighbourhood of the island.”

SM5: Importance of chick and juvenile survival

We agree with B&R-G that it is appropriate to transform the estimates of changes in the penguin response variables to predicted changes in components of demography (for foraging variables), or changes in population growth rates. We also agree that there is a relationship available for chick condition in Macaroni penguins *Eudyptes chrysolophus* (Horswill et al. 2014) that can be used (and was used by Sherley et al. 2018*) to estimate a key demographic attribute. However, B&R-G appear to have misunderstood that the relationship described by Horswill et al. (2014) corresponds to an impact on juvenile (i.e., post-fledging) survival, not chick survival (survival over the first three months of life). Perhaps this stems from their argument that the observed measures of chick survival and chick body condition:

“...are, in fact, independent (sets of) estimates of the same quantity, viz. the differential impact of closure on chick survival over the first 2-3 months of life”.

This statement is incorrect.

Individual penguin chicks (even those that survive to fledging) show high plasticity in their growth and condition. For example, the 125 chicks at Robben Island in 2004 on which the chick condition index was based, all fledged successfully (i.e., chick survival = 1 for this cohort); however, their body condition indices ranged from -0.85 to 1.65 (Lubbe et al. 2014). In other words, while chick condition may have a weak effect on chick survival in the nest, it can have a more important impact on chick survival at sea after fledging, hence recruitment (i.e., “juvenile survival”). This is because young seabirds are not as proficient at foraging as mature adults, and must develop both their musculature (Wilson 1985) and their foraging skills (Daunt et al. 2007, Votier et al. 2017) if they are to survive, especially in the period immediately after fledging. Fledging with a better body condition correlates with better first-year survival, as shown for a variety of seabirds (e.g., Gaston 1997, Saraux et al. 2011b, Horswill et al. 2014) because it provides a physiological buffer to the processes of learning to forage (note: this is also an example of the “complex adaptive ecosystem” in which the experiment occurred).

The 2016 international review panel also noted:

“...that the analyses have not attempted to integrate information across the response variables. However, while chick condition and chick growth are likely correlated, chick condition/growth and fledging success [chick survival] affect processes that are sequential in the life history of penguins, which means that a fishery effect on each of chick

While neither B&R-G nor Butterworth and Ross-Gillespie (2021a,b) cite either Horswill et al. (2014) or Sherley et al. (2018), Ross-Gillespie and Butterworth (2021) make clear in their Appendix B that the basis of their approach is the relationship first proposed by Sherley et al. (2018).
condition/growth and fledgling success in combination could lead to a biologically meaningful population effect" (Dunn et al. 2016).

To say this a different way, at least two distinct biological process contribute to survival of penguin offspring from hatching to the end of their first year of life (Weller et al. 2016), so the effects of the island closures on chick survival and chick condition should be applied sequentially to i) pre-fledging survival, then ii) post-fledging survival (e.g., Sherley et al. 2018, 2021). The population model used by B&R-G (and by extension in Figure 1 in their main text) does not apply these effects in combination (see Robinson et al. 2014, Ross-Gillespie and Butterworth 2021) despite the 2016 panel recommendation (above). In a situation like that observed at Robben Island, where the estimated closure effects would appear to increase both chick survival and chick condition (thus juvenile survival) a model that does not apply these effects sequentially to pre-fledging survival and post-fledging survival underestimates recruitment and the potential population value of fisheries closures. For example, if 75% of chicks survive to fledging at three months old and 20% of those survive to their first year as juveniles (Sherley et al. 2014a, 2021), then 15% of chicks that hatch survive to age 1. If the closures were to simultaneously improve both chick survival and juvenile survival (through an improvement in chick condition) by (a hypothetical) 5 percentage points (to 80% and 25%) then 20% of chicks that hatch survive to age 1. But, if only chick survival or juvenile survival is changed by 5 percentage points, then either 16% or 19% (respectively) of chicks that hatch survive to age 1. These differences correspond to ~8–23 more chicks each year surviving to age 3 to recruit into a breeding population of 1,000 pairs, assuming an adult survival rate of 0.77 (see SM4), or a ~0.8–2.3 percentage point underestimate of recruitment. To give this context, recruitment rates for African penguins have only been ~10% in recent decade (e.g., Sherley et al. 2014b).

In summary, for the reasons outlined above, and because – as we note in the main text – reducing the costs of reproduction could have positive effects on adult survival, the effect sizes in Figure 1 (in the main text) should be viewed as the minimum improvement in population growth rate from the 20-km closure of fishing around the penguin breeding colonies.

SM6: International review panels support various modeling approaches

B&R-G criticize our perspective over long-standing arguments about statistical approaches used to analyse the fisheries closure experiment. However, the 2020 international review panel were asked:

"Are individual data-based analyses unreliable and consequently unacceptable for consideration in developing management recommendations regarding possible future island closures?" and "Should the aggregated approach not be used to provide results on which management advice for island closures is to be based?". They directly answered "No." in both cases (Haddon et al. 2020). In other words, the results of both approaches appear valid and provide useful insight.

Pseudo-replication is, and remains, one of many issues discussed by review panels, and indeed we present a detailed assessment of this concern in the Supplemental Materials of Sydeman et al. (2021). Consequently, we feel there is no need to redraft that discussion here, except to reiterate that there remains debate in the peer-reviewed literature about how the random effect structure should best be selected in mixed-models.

For example, Arnqvist (2020) says, in an opinion article:
“Trust the full model – the random effects structure should be determined by your design rather than by your data”.

While Matuschek et al. (2017) conclude from a simulation study that:

“Although it is true that fitting a model with only random intercepts may lead to higher Type I error, fitting a maximal model also has a cost: it can lead to a significant loss of power. We demonstrate [that] higher power is achieved without inflating Type I error rate if a model selection criterion is used to select a random effect structure that is supported by the data”.

Importantly, the 2019 and 2020 review panels were very clear that model selection methods should be used (Die et al. 2019, Haddon et al. 2020). As such, in the Supplemental Materials of Sydeman et al. (2021) we use the disaggregated data to fit the model with the random effect structure (Year/Island or Year + Year:Island) recommended by the 2020 panel (Haddon et al. 2020) and demonstrate that an objective, information-criterion-based approach favours the model structures originally used in Sherley et al. (2018, 2021) over that recommended by the 2020 panel.

Furthermore, B&R-G (2022) confirm our results in their Table SM-3 in their Supplemental Materials, but they also add candidate models with the random effect structure Year/Month/Island (for chick condition) and Year/NestID/Island (for chick survival). For chick survival, their best-fitting model indicates a statistically significant (differs from zero at the 5% level) and biologically meaningful (see SM4) positive closure effects on penguin population growth rates at both Robben and Dassen islands (see model D3 in Figure 1A in the main text). This result matches the inference from the best supported model in the Supplemental Materials of Sydeman et al. (2021; labelled D5 in Figure 1A in the main text).

B&R-G then fit an additional model in which they introduce additional uncertainty by fixing the standard deviation of the year random effect to 0.33. The use of this additional variance term is neither supported by the data, nor an information theoretic approach to select the best-supported random effect structure. Nevertheless, although the closure effect on the penguin population growth rate is no longer statistically significant (at the 5% level) at Robben Island when this additional variance is added to model, it remains both statistically significant and biologically meaningful at Dassen Island (model (D3b, Figure 1A in the main text).

For chick condition, their best-fitting model only indicates a marginally positive closure effect on penguin population growth rate at Robben Island (the 95% CIs just overlaps zero), but this result is at odds with both B&R-G’s preferred approach using the aggregated data (Figure 1 in the main text), and all of the models in the Supplemental Materials of Sydeman et al. (2021). This is also the first time that a model has failed to detect an effect on chick condition at Robben Island using the data up to 2018 (Sherley et al. 2018, 2021, Ross-Gillespie and Butterworth 2020) and to what extent this Year/Month/Island nesting structure is appropriate is unclear. Even the 2020 international review panel was unsure:

“It is unclear if month should be nested within Year or Year:Island or an alternative model formulation would be most appropriate”(Haddon et al. 2020).

Clearly, exactly how the model is specified affects uncertainty in the estimated effect size – this is not at all surprising. We demonstrated in Sydeman et al. (2021) that the fixed effect structure (whether the Island × Closure interaction is retained or not) can have as big an impact on the precision of the closure effects as changing the random effect structure (see Figure S2 in the Supplemental Materials of Sydeman et al. 2021).
In summary, (1) B&R-G’s best fitting chick survival model (D3) produces effect sizes that exceed the 1% management threshold for an improvement in the penguin population growth rate at both Robben Island and Dassen Island. (2) B&R-G’s preferred model D3b (where the standard deviation of the year random effect is fixed to 0.33) also produces an effect on the penguin population growth rate >1% at Dassen Island, as does their favoured approach using aggregated data (Figure 1 in the main text). This 1% threshold has been accepted as indicative of a biologically meaningful impact of fisheries closures on the African penguin population since 2016 at least (Cochrane 2016, Ross-Gillespie and Butterworth 2016, 2020; see also SM4).

SM7: Minor differences in inference

Perhaps the most critical thing of all is that the differences from the two modelling approaches generally lead to minor differences in inference (see Figure 1 in the main text). Although we show this again in Figure 1 in the main text for two of the penguin metrics (chick survival and condition), this is not new information. Comparison of results from the disaggregated and aggregated data have been implemented before – following the spirit of advice in Arnqvist (2020): “evaluate your data also using much simpler models of group means… If you find much larger or more significant fixed effects in your complex LMM than in the simpler tests, you have likely not modeled the random effects structure in an appropriate manner…”.

Sherley (2016a,b) first provided the 2016 international review panel with comparisons of precisions (standard errors) of closure estimates derived from aggregated and disaggregated data; the panel then explored some of the consequences of the impact of these lower standard errors on the power analyses carried out by Ross-Gillespie and Butterworth (2016) and “found them to be small” (Dunn et al. 2016). Later, Sherley and Winker (2019) used the lmer package in R to present a comparison of the two modelling approaches to the 2019 international review panel, using the maximum distance dataset from the Eastern Cape island pair and the chick condition dataset for the Western Cape island pair. They found statistically significant closure effects (in the direction benefiting penguins) using both the aggregated and disaggregated data approaches (Sherley and Winker 2019). The 2019 international review panel later recommended “shifting the debate from estimation models to the consequences of the estimates” (Die et al. 2019).

Subsequently, Sherley (2020b) compared results modelling aggregated and disaggregated data using a Bayesian approach and found >95% probability that the closures led to a positive effect for the penguins using the aggregated data for the chick condition data (96% at Robben Island) and the chick survival data (99.7% averaged over both islands) from the Western Cape island pair, and the maximum distance foraging metric data (99.9% at St. Croix) from the Eastern Cape island pair.

Finally, when, in the spirit of ‘failing to disagree’ (Norberg et al. 2022), another member of Butterworth’s Marine Resource Assessment and Management (MARAM) Group at the University of Cape Town was asked in 2020 to summarize the main take-home messages from the annually aggregated and individual-level data approaches, they concluded:

“It therefore appears that:
- Both sets of analyses currently show that the closure of Robben Island to fishing will benefit penguins.
- Both sets of analyses currently show that the closure of St. Croix Island to fishing will benefit penguins.
- Both sets of analyses currently show that the closure of Dassen Island to fishing will benefit penguin chick survival, but the overall benefit is not clear given some results from Butterworth and Ross-Gillespie which suggest the
closure of Dassen Island to fishing may be detrimental to chick growth, path length and trip duration?

- Sherley’s results suggests this experiment could not inform on the impact of fishing to penguins on Bird Island, while a power analysis from Butterworth and Ross-Gillespie may help ascertain whether extending the experiment at this Island may produce more informative results” (de Moor 2020).

SM8: Future research: Importance of adult survival

Mean (± SE) adult African penguin survival was 0.711 ± 0.02 based on data from 11 colonies and over 30,000 adults marked and resighted on the Western Cape between 1994/95 and 2010/11 (Sherley et al. 2014a). It was slightly higher (0.743) at Robben Island and slightly lower (0.680) at Dassen Island (the two colonies that contributed the bulk of the data) over the same period. After a rapid decline in sardine abundance west of Cape Agulhas between 2001–2003, penguin survival was worse. After 2002, penguin survival averaged 0.657 at Robben Island 0.590 at Dassen Island (Sherley et al. 2014a). Notably, the sardine decline in the early 2000s led to spatially disproportionate fishing pressure and sardine exploitation rates between 2004 and 2007 that were more than double the long-term average (Coetzee et al. 2008).

All of the survival rates quoted above are poor in the general context of seabird life-history metrics. They are all within the bottom 3.5% of survival rates of 117 seabird species (species in the Charadriiformes, Pelecaniformes, Procellariiformes, Sphenisciforms and Suliformes) collated in a recent analysis of generation lengths of birds; moreover, only two seabird species in that database had survival rates lower than 0.711 (Bird et al. 2020). Although mean (± SE) survival rates at Robben Island have recovered slightly to 0.77 ± 0.02 in recent years (2013–2020; Leith et al. 2022), they are still within the bottom 5% of the 117 survival rates for seabird species, and no other penguin species had mean survival rates <0.80 (Bird et al. 2020). Even in a population of the closely related Magellanic penguin that was declining at ~1.3% per annum, adult survival was >0.8 in 23 of 25 study years (92%; Gownaris and Boersma 2019).

SM9: Experimental results need perspective

The demographic impacts of long-term closures have not been assessed by the experiment to date, and cannot truly be assessed using the data available. B&R-G state that the patterns of alternating closures around one of each island in a pair were implemented to maximise statistical contrast so as to improve estimation precision. We agree that this approach adds statistical contrast, but it is worth noting that the 3-year cycle that was implemented (a 2-year cycle was initially used for the feasibility study) was selected partly opportunistically (de Moor 2020), and partly as a trade-off designed to:

“optimize the outcome of the study taking account of the sometimes conflicting objectives of alternation to maximise contrast for more precise estimation, a slower alternation to take account of possible autocorrelation in the penguin indices being monitored, and the desirability to integrate the feasibility study and experiment” (Coetzee 2010).

In other words, the experiment was not designed to maximise the effect size (or detectability) of demographic responses of penguins to the possible benefits of greater food availability over the long-term. Indeed, it was argued early on by seabird biologists that rapidly alternating closures would almost certainly ensure that the signal (of not fishing around the islands) would be swamped by the noise of the alternations, and that population
responses of penguins were only likely to be detectable after five years of closures (e.g., Crawford 2006, 2010, Wanless and Moseley 2010) because of deferred breeding (birds take at least three years to recruit).

As de Moor (2020) stated:

“it is worth stepping back and noting that these analyses have evaluated the ability to detect impacts of fishing around the penguin islands based from this particular experiment… [which] has shown that closure of some islands to fishing benefits some penguin performance data. One needs to remember that for the cases where impacts have not been detected, this does not imply these impacts could not potentially be detected with a different experiment (e.g. 2 or 4 year opening/closing or 5 year closed-2 year open rotation etc.). While I’m not suggesting we start a completely new experiment, I am simply noting the ‘limitations’ of these results.”

SM10: Other factors affecting the penguin

It is worth highlighting here that spatial restrictions on fishing within the core foraging range of breeding birds is only one of many management actions currently implemented and/or under consideration to conserve the African penguin. A holistic approach to management of the species’ conservation is guided by the African Penguin Biodiversity Management Plan (AP-BMP, Department of Environmental Affairs 2013), a second version of which is currently in draft form. While many of the actions listed in the first AP-BMP, especially those that could be implemented in breeding colonies (like improving nesting habitat, e.g., Sherley et al. 2012), were or are being successfully implemented, the species continued to decrease in South Africa after the implementation of the first AP-BMP, and it retains a listing of Endangered. Although it has yet to be officially published in the government gazette, the latest draft version of the second AP-BMP states:

“While fishing closures around penguin colonies alone will not prevent this species from going extinct, and fishing closures will have an economic impact on the purse-seine industry, the critical state of this species requires all contributing conservation measures to be implemented urgently… The scarcity of food for African penguins makes it likely that the attainment of several of the APBMP’s objectives will necessitate the effective management of local competition with the purse-seine fishing industry for sardine and anchovy, through exclusion of fishing in areas that surround South Africa’s important penguin colonies” (pg. 25; Department of Forestry, Fisheries and the Environment 2022).

B&R-G do not suggest any alternative management actions that are not already being implemented that might result in a 1% increase in the penguin population growth rate. Rather, they dismiss the value of closing areas around penguin breeding colonies to purse-seine fishing and suggest more research in the form of “Models of Intermediate Complexity for Ecosystem” (MICE) to rule out competition with seals. As the long-term population viability of the species is at stake, we cannot “fiddle while Rome burns” to untangle all of the effects, find perfect analytical solutions to complicated data (if they even exist), nor simply continue to research the problem and watch a dire situation become a conservation crisis.

SM11: Data used by Butterworth and Ross-Gillespie are subject to rebuttal

B&R-G argue that the main justification in Free et al. (2021) in support of closures around seabird breeding colonies was based on the results in Pichegru et al. (2010), which they say:
“were controversial and the subject of a rebuttal” and “have been overtaken by” the results for foraging related variables in their Figure 1.

Free et al. (2021) also cite Pichegru et al. (2012), which was not the subject of a rebuttal, and B&R-G omit to mention that the key results reported in Pichegru et al. (2010, 2012) are consistent with those in their Figure 1 for the Eastern Cape colonies. Moreover, the data used by B&R-G for their Figure 2 came from Free et al. (2021). Initially these data were used by Hilborn et al. (2017), which was subject to a rebuttal by Pikitch et al. (2018). Thus, following B&R-G’s logic, since the data that they used for their Figure 2 comes from a source that is “controversial and the subject of a rebuttal”, it should also be excluded from consideration.

……………………………………………………………………………………………………………

SM12: Model averaging

To generate the overall means (and 95% credible intervals), we assume that each modelled mean effect size ($\mu_m$) was independently drawn from a latent normal distribution:

$$\mu_m \sim N(\varphi, \hat{\tau}_m),$$

where $\varphi$ is the unknown true mean effect size and $\hat{\tau}_m$ is the standard error for the modelled effect size (e.g., Sherley et al. 2015). This approach is based on hierarchical Bayesian meta-analyses (e.g., McCarthy and Masters 2005) and meant we could estimate an overall mean effect accounting for the uncertainty associated with each modelled estimate. The code and data to reproduce Figure 1 in the main text and to calculate the mean effect sizes are available on GitHub: https://github.com/rbsherley/IJMS_AP_IC (Mod_Comp_GitHub.R and MetaAnalysis.rdata).

……………………………………………………………………………………………………………

Supplementary references:


Sydeman et al. 2022

PEL_35%20Re-
analysis%20of%20the%20ICE%20results.pdf?sequence=1&isAllowed=y


