REPORT OF THE INTERNATIONAL REVIEW OF SOME ASPECTS OF THE ISLAND CLOSURE EXPERIMENT
3-9 December 2020
[Malcolm Haddon[1], Ana Parma[2], André E Punt[3]*, Michael J. Wilberg[4]]

Background

The Panel, which met virtually, was established to provide responses to three key questions related to two different approaches used to estimate the effects of fishery closures around penguin colonies on the dynamics of penguin populations. The expected outcomes of the review were YES/NO answers, with a short rationale (see Appendix for the Terms of Reference for the review, including the expected outcomes). Unlike previous Panel reports, there are no detailed lists of research recommendations, although there are several avenues for additional research (see "additional details and comments"). The Panel appreciated the provision of well-written and structured questions and briefing documents, as well as the analyses conducted during the review. In particular, the two sets of analyses differ in several ways, and the Panel made some requests for additional analyses from Dr. Richard Sherley using his Bayesian estimation approach to better understand the effects of using individual versus aggregated data, i.e. the effects of weighting the data for each year/island (and month in some cases) by sample size rather than assuming that the residual error is homogeneous.

Some general principles and overarching comments
- The work of the Panel was made more difficult because of the lack of like-vs-like comparisons. It is desirable to use common data sets and common model structures for some of the comparisons.
- Previous Panels have suggested a preference for use of individual-based data in a hierarchical model structure. In the context of this report, this should primarily be understood (within the context of the types of models under consideration) to intend that strata (e.g. year-island-month combinations) with more observations and/or lower among-individual variation are given more weight during model fitting and that the correlation structure in the data be accounted for to avoid overestimating precision, particularly, of the estimated closure effect size.
- There are many differences between the implementations of the two approaches. Some are likely to be relatively unimportant (e.g. Bayesian vs maximum likelihood estimation), but others (how data are weighted) more critical. It is also important how variation that is not due to closures (e.g. year-effects, month-effects, other covariates) is modelled.
- The primary difference between using individual versus aggregated data <u>within the current model structures</u> is whether the variance within each stratum (e.g., month-year-island) is inversely proportional to, or independent of, sample size. This could be an important consideration not only for the estimated precision of the closure effect size, but also for the estimated effect size itself given the data sets are unbalanced in terms of sample size.

---

[1] CSIRO (retired)
[2] Center for the Study of Aquatic Systems, CENPAT-CONICET, Puerto Madryn, Argentina.
[3] School of Aquatic and Fishery Sciences, University of Washington, USA
[4] Chesapeake Biological Laboratory, University of Maryland, Center for Environmental Science, Solomons, MD, USA

- The use of hierarchical models is supported by the Panel given the nature of the experiment (the primary sampling unit is island but with year, and perhaps month within year, impacting on the response) and the data collection protocol, which does not lead to the same sample size for all strata. The use of random effects for some covariates (e.g. year and month) is justified given the limited number of degrees of freedom.
- Given the complexity of the situation no model will be "correct". However, appropriately structured and tested models will provide an adequate basis to develop management recommendations regarding possible island closures.
- Random-effects models are appropriate to mitigate the effects of unmeasured covariates. However, given that the experiment could not eliminate all possible factors, even the best random-effects model may not account for an important (but unknown) covariate. Nevertheless, this is not an *a priori* reason not to (a) use the data from the experiment, and (b) reject the use of model-based inference (i.e., this is a problem common to fixed and random effects models).
- Individual-based models are vulnerable to underestimating the standard errors of important model outputs if the model is incorrectly specified; aggregated models are vulnerable to assigning inappropriate weights by stratum. Both types of error can lead to biased estimates of closure effects and the standard errors of the estimates of these effects.

Question 1. Are individual data-based analyses (for reasons given in FISHERIES/2020/AUG/SWG-PEL/82) unreliable and consequently unacceptable for consideration in developing management recommendations regarding possible future island closures

Panel conclusion: **No**. Individual data-based analyses can be appropriate for developing management recommendations regarding possible island closures.

Rationale: The individual-based approach has the advantage of analysing the data at the level they are collected, but the model needs to appropriately capture the factors and sources of variability (observed or unobserved) impacting the observations, other than closure alone.

Additional work:
- The model selection process recommended by the 2019 Panel should continue to be applied but a final test should also be conducted to determine if the variance of the residuals is similar (though not necessarily identical) to the variance of the observations at the level of island/year/month (or the strata kept in the model selected). Exact matching is not a requirement, but a model would likely be mis-specified (and inappropriate) if the variance of the residuals substantially exceeds that of the observations. The aggregated approach aims to capture unmodelled covariates using a process error term, but the disaggregated approach assumes that the model is structured so that the residual error is due only to observation error within a stratum.
- There may be some instances where the same penguin is measured in multiple years or where sources of unexplained variation affect all observations within a stratum. Assumptions of the method are that the modelled random structure accounts for all such variability and there is no additional correlation among individual observations within a stratum such that the standard error of the mean is unrepresentative of the true level of variation at the stratum level.
- The current estimation framework does not implement a hierarchical random effects structure (i.e. Year + Year:Island) (See Fig. 1). This may be

2

inconsequential in terms of the estimate of the closure effect, but Island nested within Year seems more plausible than independent Year*Island effects. Additional models with that structure should be included in the model selection process.

- The issue raised in the previous bullet is complicated by the inclusion of month as a random component. It is unclear if month should be nested within Year or Year:Island or an alternative model formulation would be most appropriate.

Question 2. Does the difference in chick survival rates after 50 days between the Kaplan Meier estimator and a log-normal hazard model at Robben (but not Dassen) island mean that conclusions regarding the effects of fishing on penguin population not use results for chick survival until the reasons for the differences are better understood.

Panel conclusion: **No**. The analyses based on chick survival rate should be used when drawing conclusions regarding the effects of fishing on penguin populations <u>if decisions are to be made now</u>, but additional work is warranted if a decision is not required immediately.

Rationale: Three reasons support the Panel view to support use of this metric if a decision is needed now:
  a) The patterns in survival with day evident from the Kaplan-Meier estimator can be explained by biological phenomena differing between Robben and Dassen Island – such differences are not unexpected *a priori*.
  b) Chick survival is perhaps the parameter that can be best linked to penguin population dynamics of the data types analysed from the island closure experiment, and should only be excluded from consideration now with more evidence than provided by the residual patterns.
  c) Looking only at the raw data (rather than model results, which account for covariates), the effect of using the "raw" (Kaplan-Meier) estimates of survival at 74 days is to strengthen (not weaken) the closure effect compared to using the results of the log-normal hazard model (i.e., there would be a larger closure effect with the Kaplan-Meier survival estimates on Robben Island).

Additional details and comments:
- The reasons for the differences between Dassen and Robben Island should continue to be investigated.

Question 3. Should the aggregated approach not be used to provide results on which management advice for island closures is to be based

Panel conclusion: **No**. The aggregated approach can be appropriate for developing management recommendations regarding possible island closures.

Rationale: The advantages of the aggregated approach include that (a) it avoids needing to include model structure to represent variation more than that due to island, year and month by allowing for "process error" (i.e. estimating a residual standard deviation that might substantially exceed the standard error of the mean at the stratum level), and (b) it does not need to make assumptions about within-stratum correlation. The disadvantage of the approach as applied at the moment is that no attempt is made to model the sources of process error and each data point is given equal weight (although the equal weighting is appropriate if process error is substantially larger than observation error). Different alternatives for weighting the (averaged) data points are suggested below in order to account for the heteroscedasticity resulting from the unbalanced design and differences in observation error variances among strata.

Additional details and comments:
- Include "month" in the model (where such data are available), as month clearly has an effect on some of the variables being measured.
- Consider models for which either: (a) there are more random effects (but still fitting to the aggregated data) but with the error variance replaced by $\hat{\sigma}^2_{i,j,k} / n_{i,j,k}$ where $\hat{\sigma}_{i,j,k}$ is the standard deviation of the observations in stratum $(i.j,k)$, where i is island, j is year and k is month, and $n_{i,j,k}$ is the number of observations for stratum $(i.j,k)$ [the check outlined for the aggregated approach - first point under "additional details and comments" for Q1 - should be applied], or (b) the error variance is replaced by $\hat{\sigma}^2_{i,j,k} / n_{i,j,k} + \tau^2$ where τ^2 is the process error variance.
- Models (a) and (b) could be combined, for which case it would be expected that the estimate of τ^2 would be close to zero.
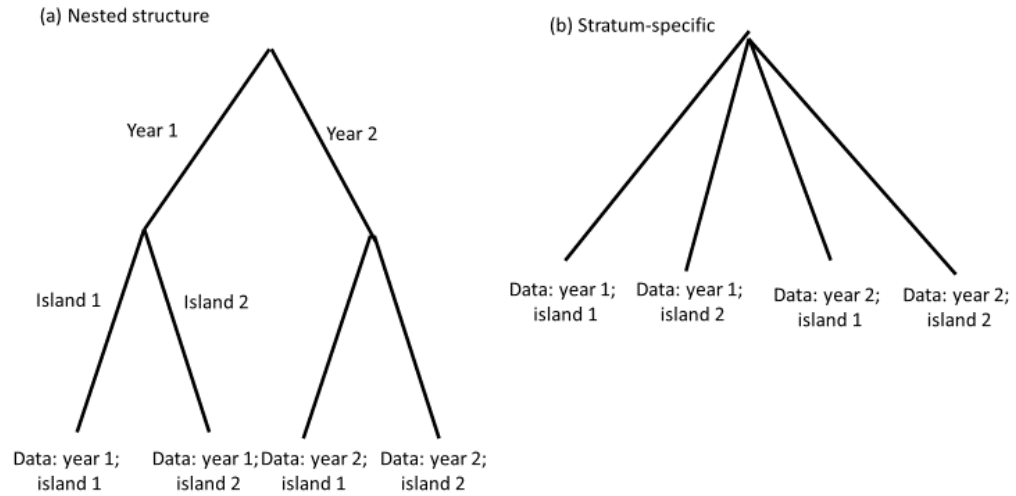
Figure 1. Illustration of a model in which island is nested within year (a) and in which no nested structure is imposed on year and island (b). The models developed by Dr. Richard Sherley are based on (b). The differences between (a) and (b) are that they estimate the parameters that describe the distributions of the data in different ways. In the nested structure (a) year effects are drawn from a normal distribution, the island effects are drawn from a year-specific normal distribution and the data are drawn from a normal distribution based on the estimated island effect (nested within year) (response ~ fixed_effects + (1|Year)+(1|Year:Island)) in *lmer*). For the stratum-specific structure (b), each year-island combination is drawn from a common normal distribution, and the data are drawn from normal distributions with year-island specific parameters (response ~ fixed_effects + (1|stratum)). Thus, the two structures will result in different estimates of the means and variances at the data level because of the distribution imposed by the differences in model structure.

Appendix A: International review of some aspects of the Island Closure Experiment Terms of Reference and Organisation Details

Format:
Virtual review meeting via Skype 3-4 December & 7-9 December 2020.
International review panel members: Ana Parma, Andre Punt, Malcolm Haddon, Michael Wilberg. Chair: Janet Coetzee

Process:
Two, alternating "closed sessions" with the review panel members for Richard Sherley and Doug Butterworth/Mike Bergh. The first round of closed sessions will allow for presentation of the main methods/results/comments to the review panel. Members and observers of the SWG-PEL will be invited to join these sessions as a "silent audience". This will be followed by a second closed session with focussed discussion of those methods/results/comments that will enable the review panel to address the key questions. These are "private sessions". A 5th plenary session will be held with the entire SWG-PEL (including observers) where the panel will deliver its report containing answers to the key questions (see Table 1 for proposed meeting schedule and Table 2 for session detail).

Expected outcome:
Written report with clear, unambiguous YES/NO answers to the key questions. This may include optional remarks suggesting further work or rationale for answers, preferably restricted to a few sentences. Given the limited time available for this review and the need to focus on KEY QUESTIONS only, no other issues are to be considered. Note the list of questions below is in **PRIORITY ORDER.**

Key Questions:
Q1: It has been asserted (see FISHERIES/2020/SEP/SWG-PEL/96rev) that the estimates of island closure effects provided in FISHERIES/2020/JUL/SWG-PEL/53REV, which are based on individual data-based analyses, are (for reasons given in FISHERIES/2020/AUG/SWG-PEL/82) unreliable and consequently unacceptable for consideration in developing management recommendations regarding possible future island closures. Do the reasons given justify this assertion?

Q2: It has been asserted (see FISHERIES/2020/SEP/SWG-PEL/96rev) that the marked (and apparently relatively precisely estimated) change in the estimated survival rate at Robben (but not Dassen) island from the Kaplan-Meier estimates of chick survival rates after some 50 days exposure needs to be better understood before the associated results could be used with confidence as inputs to estimators of island closure effects – see the reasons given in FISHERIES/2020/AUG/SWG-PEL/82 (bottom of page 32) and FISHERIES/2020/AUG/SWG-PEL/84 (last paragraph on page 2). Is this assertion justified?

Q3: It is acknowledged that the limited degrees of freedom available with the aggregated data approach hampers efforts to obtain precise closure (fishing effect) estimates from the island closure experiment. In 2016, a power analysis approach was finalised to advise on the period needed for the closure experiment to have to continue before being able to provide biologically meaningful results; this was based on an aggregated data approach. In 2019, the Panel recommended that "given the nature of the experiment, use of individual data is to be preferred" (first bullet, page 10, MARAM/IWS/DEC19/General/5). Does it therefore follow that the aggregated approach should not be used to provide results on which management advice for island closures is to be based?

Table 1. Dates and times.

	Thur 3 Dec	Fri 4 Dec	Mon 7 Dec	Tue 8 Dec	Wed 9 Dec
SAST (UTC+2)	15:15 – 16:45	15:15 – 16:45	15:15 – 16:45	15:15 – 16:45	15:15 – 16:45
Ana Parma (ART) UTC-3 hours 10:15	Not available	(grey)	(grey)	(grey)	(grey)
Andre Punt (PST) UTC-8 hours 05:15	(grey)	(grey)	(grey)	(grey)	(grey)
Malcolm Haddon (AEDT) UTC+11 00:15 **AM** (Fri 4th, Sat 5th)	(grey)	(grey)	Not available	Not available	Not available
Michael Willberg (EST) UTC-5 hours 08:15	(grey)	(grey)	Available up to 16:00	(grey)	(grey)
Richard Sherley UTC 13:15	(grey)		(grey)		(grey)
Mike Bergh SAST 15:15		(grey)	(grey)	(grey)	(grey)
Doug Butterworth SAST 15:15	(grey)	(grey)		(grey)	(grey)
Kim Prochazka/Janet Coetzee SAST 15:15	(grey)	(grey)	(grey)	(grey)	(grey)

Grey = attendance required

Table 2. Session detail

Date	Time (SAST)	
Thursday 3 December	15:15 – 16:15	Richard Sherley presentation to the Panel ; Members and observers of the SWG-PEL attend as "silent audience"
	16:15 – 16:45	Panel zoom meeting
Friday 4 December	15:15 – 16:15	Doug Butterworth/Mike Bergh presentation to the Panel Members and observers of the SWG-PEL attend as "silent audience"
	16:15 – 16:45	Panel zoom meeting
Monday 7 December	15:15 – 16:15	Panel discussion with Richard Sherley
	16:15 – 16:45	Panel zoom meeting
Tuesday 8 December	15:15 – 16:15	Panel discussion with Doug Butterworth/Mike Bergh
	16:15 – 16:45	Panel zoom meeting
Wednesday 9 December	15:15 – 15:45	Panel zoom meeting
	15:45 – 16:45	Panel report to the SWG-PEL

Panel Zoom meeting to be set up independently by Andre Punt for panel members only.