Is pseudo-replication biasing results from analyses from the island closure experiment which model individual penguin responses directly?

D.S. Butterworth and A. Ross-Gillespie[1]

Summary

A simple simulation study is used to investigate the impact of possible pseudo-replication arising from the use of individual penguin observations, in contrast to annually aggregated measures, in analyses of the island closure experiment which attempt to estimate the possible effect on penguins of closure of the neighbourhood of these islands to pelagic fishing. Unlike the case for estimators based on annually aggregated inputs, those based on the use of individual observations are found to lead to possibly substantially negatively biased estimates of the standard errors of the parameter that reflects the effect on penguins of these closures. This means that past results concerning the statistical significance and probabilities that island closures impact penguins from analyses based on individual observations need to be reconsidered. Previous analyses using this approach should ideally be repeated based on year-aggregated inputs, and future analyses need to avoid repeating this earlier approach.

Introduction

A number of analyses of the data from the island closure experiment (e.g. most recently Sherley *et al*., 2018) have directly modelled data from individual penguin observations, in contrast to basing analyses on annual averages of such observations. This approach has, however, been questioned for some time (e.g. see FISHERIES/2016/AUG/SWG-PEL/65), essentially on the grounds that it constitutes a form of pseudo-replication, and as such will lead to over-optimistic estimates of the precision of estimates of the impact of fishing on the penguin reproduction/chick survival process.

This document reports the results of a simple simulation study to investigate this concern further.

Methods

The simulation study used for this investigation is described in detail in the Appendix. Essentially it reflects the design of the island closure experiment, and simplifies models of the type applied in Sherley *et al.* (2018) to their essence by considering only island, year and closure effects (and omitting month, for instance). An additional unknown/hidden covariate is also introduced to allow for the possibility that data collected from different penguins from the same island in a particular year do not constitute fully independent samples. Essentially this reflects some unmeasured/unrealised factor which influences the response variable recorded from a penguin.

Two classes of underlying (operating) models (OMs) are considered: OM1 which includes this hidden covariate, and OM2 which does not. Three estimation models (EMs) are considered: EMA which utilises data from individual penguin observations, EMB which considers only annual average values of such observations, and EMC which duplicates EMB except for estimating year-related parameters as fixed rather than random

[1] Marine Resource Assessment and Management Group, Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701.

effects. Mixed model estimators with both fixed and random effects (such as EMA and EMB) use REML for variance-unbiassed estimation.

Results

Table 1 lists the specifications for the 11 runs for which results are presented in this document. These 11 runs fall into four broad sections. The first three sections are for runs where the number of penguins sampled per island per year is set at *N*=1, 50 and 200 respectively, and the run parameter values have deliberately been chosen to be "large" with the intent to illustrate possible differences in results clearly. In contrast, the fourth section contains runs reflecting a coarse attempt for the simulated data to mimic the values and variability for an actual situation. The *N*=1 runs (only one penguin sampled each year at each island) are not intended to be "realistic"; rather they serve as a check that the code a working correctly and provide information on bounds to outcomes.

Figure 1A shows the results for estimation of the δ parameter for each OM and EM combination, while Figure 1B shows similar results for the estimates of SE_δ(true) and mean(SE_δ). Note that in a few instances, the software used (R software, using the lmer function from the lme4 package) reported a possible convergence query for the model fit, but this occurred in less than 1% of cases for any run, so did not seem of any real concern.

Discussion

Figure 1A shows, as might have been expected given the design of the simulation testing framework, there is no evidence of bias for any of the OM-EM combinations considered: all 95% CIs overlap (or virtually overlap) the true value of the fishing effect parameter δ.

Figure 1B, however, provides more noteworthy results concerning bias in the estimation of the standard error of δ. First, it is very clear that the estimates of this standard error are unbiased for the estimators based on input data which has first been averaged over each year (the fixed effects estimator EMC shows virtually no bias in its estimates of this standard error, while the mixed effects estimator EMB shows only a very small (though consistent) negative bias).

Of more importance though is that the estimator (EMA) based on data from individual penguins is clearly and consistently negatively biased; furthermore, when those data are no longer independent within a year (as in OM1, compared to OM2), the size of this negative bias increases. Behaviour in relation to other specification changes covered in runs 1 to 8 are as might be expected: for example, the standard error of δ (SE(δ)) decreases as the number of years in increased, and the relative bias of the estimate of SE(δ) under EMA (use of data from individual penguins) increases with the number of penguins sampled each year.

The (increase in) bias for this estimate of SE(δ) under EMA is not surprising when moving from OM2 to OM1, given the introduction of non-independence in the individual penguin data generated each year under OM1, which leads to a pseudo-replication effect. However, the reason for bias even in the absence of that effect is perhaps less obvious. It would seem to arise from the "orthogonality" of the effect of interest (the closure effect which is being estimated through disentangling year, island and closure impacts at an **annual** level) from data (the individual penguin observations) whose additional information content relates only to **within-year** variability. (Note that there is no cross-year linkage of these individual penguin data, which arise from random samples each year.) Hence, in the context of the closure effect of interest here, the direct use of individual penguin observations in the estimator reflects pseudo-replication.

The simulation framework applied does not include all the variables which models such as those considered in Sherley *et al.* (2018) have considered, e.g. month. But inclusion of month would not change the fundamental result here that EMA-type estimators lead to negatively biased SE(δ) estimation. Distinguishing

individual data by month, other than adjusting for perhaps unbalanced sampling, contains only within-year information, so would again constitute a form of pseudo-replication in this context.

The simulation parameter value choices made for runs 1-8 were deliberately "large" in their intent to illustrate possible biases. The values used for runs 9-11 were chosen (very coarsely) to be more indicative of what is typical for one of the penguin response variables which is actually being monitored (mean foraging length). The extent of negative bias in SE(δ) here under OM2 is lower than for runs 4-8, and is about 40%. However, if the possibility of non-independence in the data is admitted (under OM1), this bias can increase to about 55%.

More time could be spent on more careful determination of these run parameters so as to better mimic actual situations. However, such an exercise would still not be able to identify the extent of non-independence in such data, and hence would not be able to pin down the extent of the bias in SE(δ). A simpler approach, since EMB- and EMC-type estimators based on annual aggregate data are indicated not to be subject to this bias, would be to determine the bias for EMA-type estimators by re-running previous analyses (such as those in Sherley *et al*., 2018) on annually aggregated rather than individual penguin response data.

Implications

Ignoring the negative bias in estimates of standard errors of the effect of closure to fishing parameter δ when EMA-type estimators based on individual data are used, means both that conclusions of statistical significance may be unfounded, and probabilities that there is such an effect will be over-estimated.

It therefore follows that past results of this nature which have been advised based of such estimation need to be reconsidered. Previous analyses using this approach should ideally be repeated based on year-aggregated inputs, and future analyses need to avoid repeating this earlier individual-based approach.

Many of the additional aspects included in past EMA-type approaches can be included in year-aggregated methods. For example, annual values for response variables can be GLM or GLMM standardised (e.g. to take account of month) before input to the δ estimator to account for sampling that has not followed an exactly balanced design. Concerns about differing sample sizes from year to year are not likely to be serious given that process error typically dominates observation errors in such situations, and if there are years for which sample sizes amounted to very small (<5, say) observations so that representativity becomes questionable, those data points can simply be omitted for the analysis[2].

References

Butterworth, DS 2016. On the use of aggregated vs individual data in assessment models. DAFF Branch Fisheries document: FISHERIES/2016/AUG/SWG-PEL/65: 6pp.

Sherley RB, Barham BJ, Barham PJ, Campbell KJ, Crawford RJM, Grigg J, Horswill C, McInnes A. Morris TL, Pichegru L, Steinfurth A, Weller F, Winker H, Votier SC. 2018. Bayesian inference reveals positive but subtle

[2] Note the recommendation of the 2015 International Stock Assessment Review Panel in this regard: "There is no need to account for sample size when generating data in any simulations given the low observation error relative to process error (MARAM/IWS/DEC15/PengD/P2). However, it is also reasonable to exclude data points based on very small sample sizes (perhaps < 5 points) when conditioning the operating model or to estimate the sample size component of the observation error. "

effects of experimental fishery closures on marine predator demographics. *Proceedings of the Royal Society* B: 285: 20172443.

Table 1: Summary of the specifications for the OM parameters used to generate data for the 11 runs for which results are presented in this document. The table has been divided into four broad sections – the first three sections are for runs where the number of penguins sampled per island per year is set at *N*=1, 50 and 200 respectively, and the fourth section contains results for an coarse attempt for the simulated data to mimic the values and variability for the actual Robben and Dassen island foraging ln(mean length) data. Grey highlighting has been used to indicate where key parameters are changed within each section. In the table below:

M is the number of simulations conducted for each run,
N is the number of penguins sampled each year at each island,
n_b is the number of years considered for each run,
n_c is the number of number of levels considered for the unknown covariate,
a(1, 2) is a vector with the values assumed for the island effect a_i for island *i*,
δ is the value of the closure effect,
σ_b is the standard deviation of the year effect,
σ_c is the standard deviation of the unknown covariate effect,
σ_ϵ is the standard deviation of the error term for OM1, and
$\sigma_{\epsilon 2}$ is the standard error deviation of the error term for OM2.

Run	*M*	*N*	n_b	n_c	*a*(1, 2)	δ	σ_b	σ_c	σ_ϵ	$\sigma_{\epsilon 2}$
1	1000	1	30	1	(0, 0.3)	0.1	0.2	1.00	0.02	1.00
2	10000	1	30	1	(0, 0.3)	0.1	0.2	1.00	0.02	1.00
3	1000	1	60	1	(0, 0.3)	0.1	0.2	1.00	0.02	1.00
4	1000	50	30	5	(0, 0.3)	0.1	0.2	1.00	0.02	1.00
5	1000	200	30	5	(0, 0.3)	0.1	0.2	1.00	0.02	1.00
6	1000	200	30	5	(0, 0.3)	0.1	0.2	0.30	0.02	0.30
7	1000	200	30	10	(0, 0.3)	0.1	0.2	1.00	0.02	1.00
8	1000	200	60	5	(0, 0.3)	0.1	0.2	1.00	0.02	1.00
9	1000	30	30	5	(0, 0.2)	0.1	0.1	0.40[3]	0.00	0.40
10	1000	30	30	5	(0, 0.2)	0.1	0.1	0.35	0.20	0.40
11	1000	30	30	5	(0, 0.2)	0.1	0.1	0.00	0.40	0.40

[3] Values for σ_c follow from the relationship $(\sigma_{\epsilon 2})^2 = (\sigma_\epsilon)^2 + (\sigma_c)^2$.

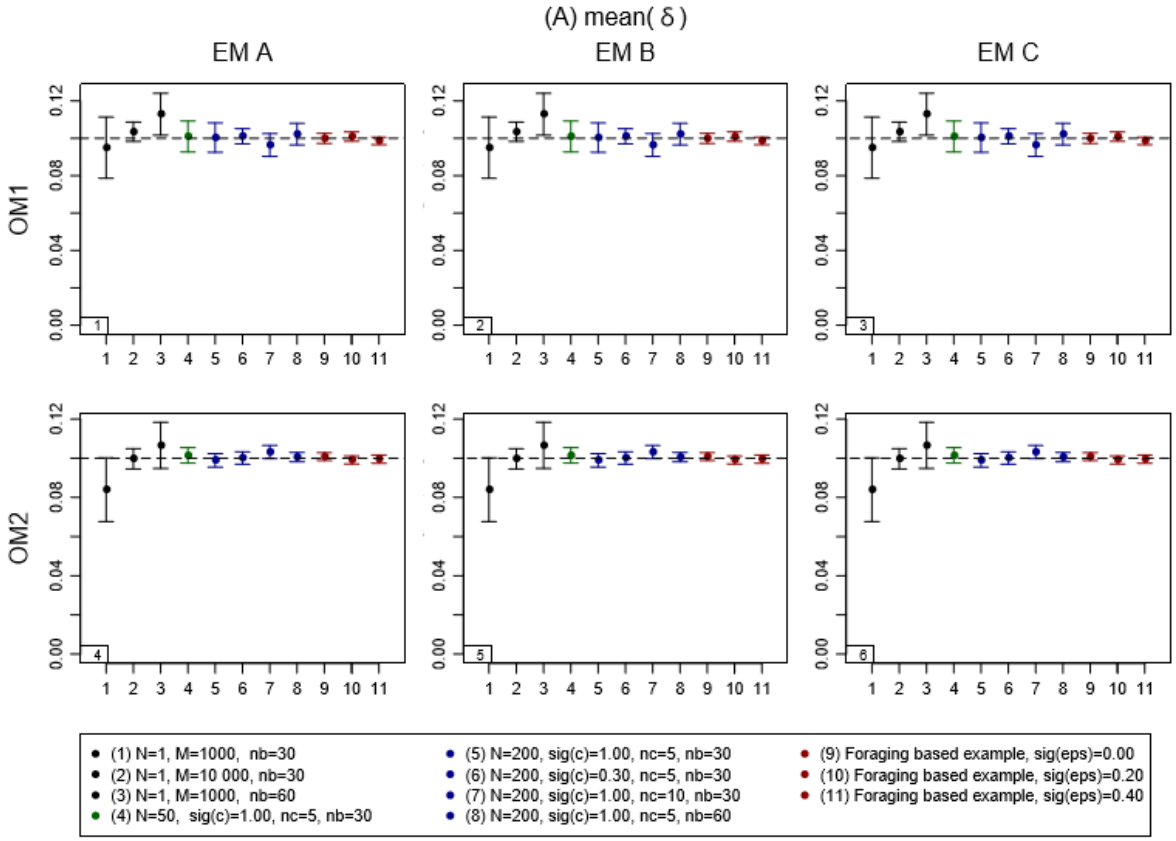

Figure 1A: The mean and 95% confidence interval (taken to be +-1.96SE) for the closure effect $\boldsymbol{\delta}$ estimates across the *M* simulations are shown for each run, OM and EM combination. The horizontal dashed line is at 0.1, the (true) value input for $\boldsymbol{\delta}$ to generate the data – bias is indicated by a difference of the mean value plotted from this line. Note that the numbers in the bottom left corner are for reference purposes. The notation sig(eps) in the legend refers to $\boldsymbol{\sigma_\epsilon}$.

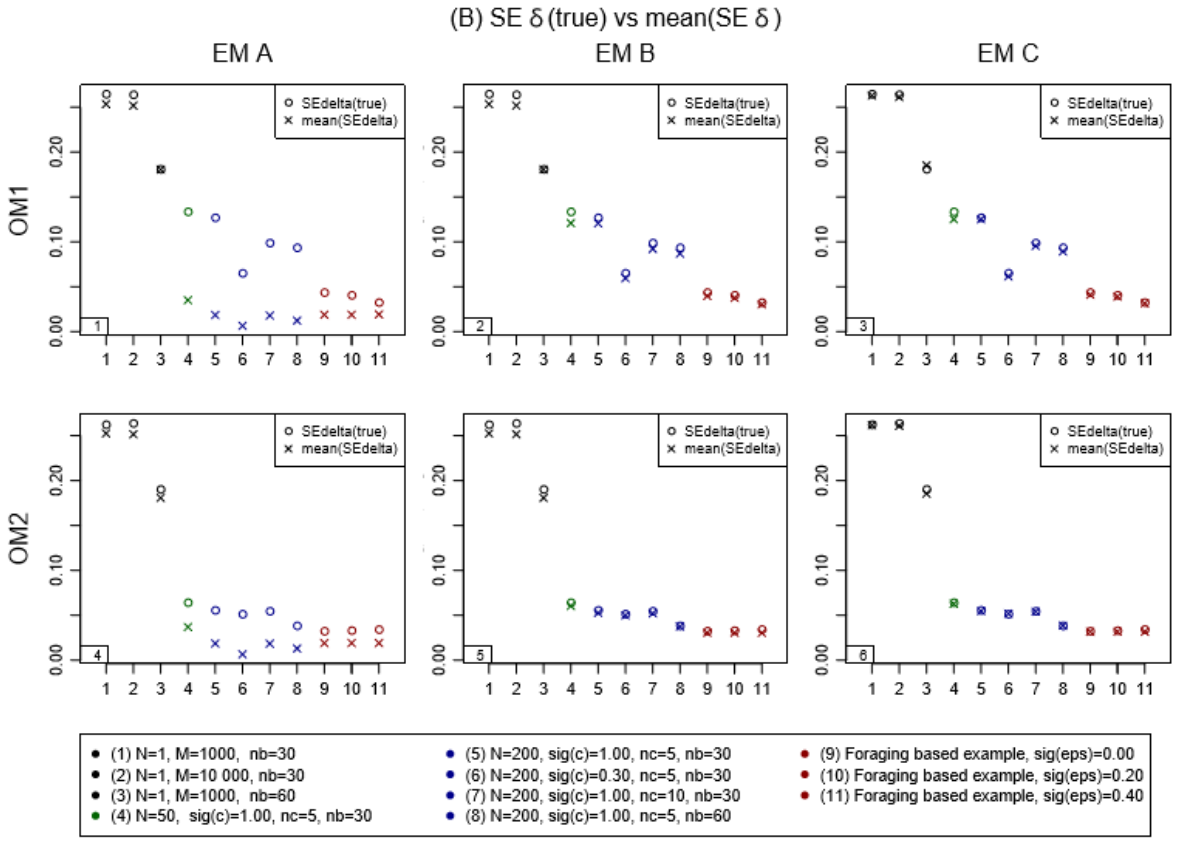

Figure 1B: Values for $\boldsymbol{SE_{\delta}}$(true) are shown by the open circles (o) and the values for mean($\boldsymbol{SE_{\delta}}$) are shown by the crosses (x). $\boldsymbol{SE_{\delta}}$(true) is calculated as the standard deviation of the $\boldsymbol{\delta}$ estimates across the *M* simulations and has been used to calculate the 95% confidence intervals in Figure 1A. The statistic mean($\boldsymbol{SE_{\delta}}$) is calculated as the average across the $\boldsymbol{SE_{\delta}}$ values for the *M* simulations. Note that the numbers in the bottom left corner are for reference purposes. An "x" below an "o" indicates that the estimate of the standard error for that $\boldsymbol{\delta}$ estimate is negatively biased.

Appendix

Simulation testing methodology

The simulation test framework consists of two operating models (OMs) and three estimation models (EMs). The OMs are used to generate the pseudo data, to which the EMs are applied to evaluate their performance.

Operating Models (OM)

1. $F_{i,y,z,j} = a_i + b_y + c_{i,z} + \delta(X_{i,y}) + \epsilon_{i,y,z,j}$
2. $F_{i,y,j} = a_i + b_y + \delta(X_{i,y}) + \epsilon 2_{i,y,j}$

where

$F_{i,y,z,j}$	is the response variable for island *i*, year *y*, unknown covariate *z* and penguin *j,*
a_i	is the island effect for island *i* where *i*=1,2 (fixed effect),
b_y	is the year effect for year *y* where *y*=1*, …, n_b*, and is assumed to be normally distributed with $b_y \sim N(0, (\sigma_b)^2)$,
$c_{i,z}$	is an unknown/hidden covariate effect for island *i* and covariate *z* (e.g. this could reflect different areas within the colony), and is assumed to be normally distributed with $c_{i,z} \sim N(0, (\sigma_c)^2)$,
δ	is the closure effect,
$X_{i,y}$	is a vector of 0's and 1's, with a 0 for years for which island *i* is closed to the fishery, and a 1 where it is open,
$\epsilon_{i,y,z,j}$	is an error term for OM1 for penguin *j*, where $\epsilon_{i,y,z,j} \sim N(0, (\sigma_\epsilon)^2)$
$\epsilon 2_{i,y,j}$	is an error term for OM2 and penguin *j*, where $(\sigma_{\epsilon 2})^2 = (\sigma_\epsilon)^2 + (\sigma_c)^2$ so that the overall variance of the *F* values generated by OM1 and OM2 is the same for the same values of other parameters.

Data generation

Multiple sets (*M* simulations) of data are generated for n_b years for two islands. Island 1 is assumed to be closed to fishing in years 1-3, 7-9, 13-15,… and island 2 to be closed in years 4-6, 10-12, 16-18,… to replicate the design of the island closure experiment. Each year, data are generated for $j = 1,2, \dots, N$ penguins sampled at each of the two islands. For OM1, data are generated in equal numbers for each level for the *z* covariate, i.e. each year N/n_cvalues are generated for each level, where n_c is the number of levels. Note that the role of the z covariate is to introduce non-independence in the individual penguin observations in OM1 (this is not present in OM2). Table 1 in the main text lists the details of the various values assumed to generate data for the different runs.

Estimation models (EM)

A. $Fobs_{i,y,j} = a_i + b_y + \delta(X_{i,y}) + \epsilon_{i,y,j}$

   where a_i and δ are fixed effects and b_y is a random effect, with their values estimated using REML; note the absence of the *z* subscript, as that hidden covariate would not be known to the observation process.

B. As for (A), but generated *F* values are fitted not for each individual penguin observation, but instead are first averaged for each year for each island; hence the *j* subscript no longer appears in the estimator.

C. As for (B) (i.e. the model is fitted to annually aggregated data), but b_y is treated instead as a fixed effect.

Key output statistics

For each simulation $k = 1, 2, \ldots, M$ and for each OM and EM combination, an estimate of δ_k is determined, along with its associated standard error estimate $SE_{\delta,k}$ using the EM under consideration. From these values a mean(δ) and a mean(SE_δ) are calculated. The true SE_δ is given by the standard deviation of the *M* values of δ_k.

Estimation bias is then reflected by the difference of mean(δ) from the actual (true) value of δ input, and for the standard error estimate of δ by: mean(SE_δ) - true SE_δ .