Response to MARAM/IWS/2019/PENG/P6

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

Summary

This document[2] first corrects some misleading results in PENG/5 through specifying the Operating Models OMs more correctly by including process error explicitly. Results from these revised OMs provide resolution of the "self-test" concerns raised in PENG/P6. However, the negative bias in estimates of the precision of the effect of fishing parameter δ remain unless the magnitude of process error is minimal compared to observation error. Since earlier analyses have indicated that process error dominates observation error in the island closure experiment penguin response data, the possibility remains of large negative bias in the estimates of precision from Sherley *et al.* models of the effect of fishing parameter based on the use of individual data. Ultimately only simulation tests will reveal definitively whether or not these random effects approaches do improve estimation precision, and it is pleasing to note that the authors of PENG/P6 are now engaged in pursuing such tests.

Introduction

Before responding to the details of MARAM/IWS/2019/PENG/P6, we must apologise for a glitch in our bias computations for the closure experiment analysis standard error estimators, though this does bring the advantage of resolving one of the queries raised in PENG/P6 (as elaborated below).

This glitch is most easily explained by reference to the update of the original PENG/P5 Appendix, which appears below. Focus for the moment on OM2 and OM4: OM2 is mis-subscripted. The b_y term needs both a deterministic and a process error component, and OM2 showed this as if it were deterministic only. For background, note that the combination $a_i + b_y$ reflects the basic assumption underlying the island closure experiment: that because the pair of islands[3] considered is relatively close, penguins will be responding to a common density of forage fish b (though this will vary with year y). There will though be a constant multiplicative difference (in F - the model reflects the log of the penguin response variable) between the islands, for which account is made by the a_i term. However, this deterministic prescription omits a stochastic process error term (η) to allow for variation about this deterministic description.

OM4 indicates the process error component (η) explicitly, by showing it separately from b_y. The results reported PENG/5 corresponded in fact to OM4, rather than OM2, and for an extreme case of the process error component only, i.e. $\sigma_b = 0$ and σ_η equal to the value specified in the document PENG/P5 for σ_b.

Table 1 lists a set of runs that have now been conducted based on the corrected versions of OM1 and OM2: OM3 and OM4 respectively – see the Appendix below, which also adds EM D to the estimation models, this being a fixed effect equivalent of EM A which is based on input of individual penguin data. The basic specifications are as for Run 5 of PENG/P5, though now including also the option for different choices of the σ_η parameter (though subject to adjusting other variance parameters to maintain a fixed σ_F).

---

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

[2] Note that this document is a revision of MARAM/IWS/2019/PENG/P7, correcting a typo in the EMD description in the Appendix.

[3] There are two pairs of islands in the experiment: Dassen and Robben on the west coast, and Bird and St Croix on the south coast.

Results

Results for the Runs specified in Table 1 are reported in Figure 1 (for OM4), and Figure 2 (comparing OM3 and OM4 outputs), in similar fashion to the corresponding Tables in PENG/P5, so that the basics of this form of presentation, which is explained in PENG/P5, will not be repeated here. Instead we focus on a summary of the key results.

- For what is now the OM4-EMA combination (in place of the OM2-EMA combination whose results were queried in PENG/P6), there is no bias in the estimation of SE_δ for the case of (virtually) no process error ($\sigma_\eta \sim 0$) – in the terms used by PENG/P6, EMA passes a "self-test" (see Figure 1B).

- But the moment process error is introduced, with σ_η greater than zero, the bias returns, increasing as the proportional contribution of process error to the overall variance (of F) increases.

- The true value of SE_δ is effectively independent of the estimator used (EMA to EM D) for the same choices for other run parameters, i.e. independent of whether individual or annually averaged data are used, and of whether the year effect is treated as a fixed or random effect. Hence, as far as the true values of SE_δ are concerned, estimation using the individual data seems to offer no benefit over using the annually averaged data.

- The same is true for the estimated value of SE_δ, except that results differ between the individual data (EMA and EMD) and annually averaged data (EMB and EMC) estimators. The former show negative bias, except in the limiting case of zero process error η; the latter do not show such bias.

- As the annual sample size (N) increases, the true value of SE_δ decreases, but only if the process error η is low. For higher values of the process error, there is virtually no estimation advantage in increasing the size (N) of the annual sample from penguins. Basically, it is the size of the process error σ_η that becomes totally dominant in determining the true value of SE_δ. For estimators based on annually averaged data, there is little bias in the estimate of SE_δ, but for those based on individual data, the extent of the negative bias in the estimate increases with the annual sample size (N).

- From the OM3 vs OM4 comparisons in Figure 2, again there is no bias in the SE_δ estimates for the estimator based on annually averaged data. For individual data, the greater bias for the OM3 model with non-independence in the response variable data is evident, as in PENG/P5, and the extent of the difference is greater as the annual sample size (N) increases, but reduces as the contribution of process error to the overall variance increases (essentially because process error swamps observation error effects, and the latter's effects on differences in estimation precision between OM3 and OM4).

Discussion

Clearly, if individual data are to be considered in analyses (though the results above offer no indications of their leading to improvements in the precision with which δ can be estimated), what is central to whether there needs to be concern about negative bias in estimates of precision (SE_δ) is the relative size of process error to observation error in each's contributions to the overall variance of the response variable.

Earlier evaluations (see sections 2 and 2rev of MARAM/IWS/DEC15/PengD/P2) indicated that for the island closure response variable data, observation error is low compared to process error, as noted by the International Review Panel in 2015. Indeed, the plots and tables in that reference suggest total dominance of process over observation error variance in these data.

Given this, taken together with the results shown in Figures 1B and 2 of this document, there has to be concern that there may be considerable negative bias is the estimates of SE_{δ} reported for the analyses based on individual penguin data which have been reported in various Sherley *et al.* papers. The essence of this concern is that even if the approaches suggested by those authors may be able to extract further information from within-year samples to lessen the impact of observation error on the precision of results, it is the between-year information that relates to the dominant process error effects on precision, and the models advanced by Sherley *et al.* do not appear to address and improve those. In effect then, their use of individual data appears equivalent to pseudo-replication.

Moving on to the specific criticisms of PENG/P6, their concerns about the "self-test" have been resolved. Certainly, our focus on estimation of standard error as the basis for comparison, rather than root-mean-square-error, relies on lack of bias in the estimation of δ, but our testing design intended that in the interests of simplicity (and achieved it, as evidenced by the bias results for δ shown in Figures 1A and 2). Having demonstrated the bias that was the basis for our concern in this simpler situation, there seemed to be no compelling need to consider more complex situations to make our case.

Use of random effects approaches in conjunction with individual data would certainly seem able to lessen the impacts of negative bias in the estimation of precision arising from non-independence of those individual data. But how would it be known by what proportion this negative bias would be reduced – this approach can account for known co-variates, but what about unknown co-variates for which no data are available to include such effects in the models? Indeed PENG/P6 seems at the start of its Discussion section to acknowledge that. However, this seems a secondary concern, given the potentially major impact of process error on individual data-based estimators that is indicated above.

The criticism that PENG/P5 did not test the Sherley *et al.* estimators themselves has some justification. Our conclusions follow more from the inferences that we draw from our results above, which suggest that there is no information content in the data that could allow those models to achieve more precise estimation.

However, we could be wrong in that respect. The ultimate test of an estimator is whether it performs adequately given simulated data (and indeed the results from complex estimators should not be accepted before simulation tests have confirmed such performance). This includes adequate performance – indeed better performance than demonstrated by EM A and EM D above - on data generated from OM3 and OM4. We are pleased to note that the authors of PENG/P6 are engaged in accepting this challenge.

Table 1: Summary of the specifications for the OM parameters used to generate data for the runs for which results are presented in this document. The table has been divided into two sections – the first for the OM4 results presented in Figure 1A and 1B, and the second for the OM3 specifications presented in Figure 2. Note that the OM4 results included in Figure 2 are the same as for Figure 1A and B. Grey highlighting has been used to indicate where key parameters are changed within each section or between sections. A dash indicates the parameter is not included in the OM in question. 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 observation error term for OM3
$\sigma_{\epsilon 2}$ is the standard error deviation of the observation error term for OM4, and
σ_η is the standard error deviation of the process error term for OM3 and OM4.

OM	*M*	*N*	n_b	n_c	*a*(1, 2)	δ	σ_b	σ_c	σ_ϵ	$\sigma_{\epsilon 2}$	σ_η	$\sqrt{\sigma_{\epsilon 2}^2 + \sigma_\eta^2}$
OM4 (Figures 1A and 1B)	1000	10	30	-	(0, 0.3)	0.1	0.2	-	-	1	0.02	1.0002
	1000	10	30	-	(0, 0.3)	0.1	0.2	-	-	0.8663	0.5	1.0002
	1000	10	30	-	(0, 0.3)	0.1	0.2	-	-	0.5	0.8663	1.0002
	1000	10	30	-	(0, 0.3)	0.1	0.2	-	-	0.02	1	1.0002
	1000	30	30	-	(0, 0.3)	0.1	0.2	-	-	1	0.02	1.0002
	1000	30	30	-	(0, 0.3)	0.1	0.2	-	-	0.8663	0.5	1.0002
	1000	30	30	-	(0, 0.3)	0.1	0.2	-	-	0.5	0.8663	1.0002
	1000	30	30	-	(0, 0.3)	0.1	0.2	-	-	0.02	1	1.0002
	1000	200	30	-	(0, 0.3)	0.1	0.2	-	-	1	0.02	1.0002
	1000	200	30	-	(0, 0.3)	0.1	0.2	-	-	0.8663	0.5	1.0002
	1000	200	30	-	(0, 0.3)	0.1	0.2	-	-	0.5	0.8663	1.0002
	1000	200	30	-	(0, 0.3)	0.1	0.2	-	-	0.02	1	1.0002
OM	***M***	***N***	n_b	n_c	***a*(1, 2)**	δ	σ_b	σ_c	σ_ϵ	$\sigma_{\epsilon 2}$	σ_η	$\sqrt{\sigma_\epsilon^2 + \sigma_\eta^2 + \sigma_c^2}$
OM3 (Figure 2)	1000	10	30	5	(0, 0.3)	0.1	0.2	1.00	0.02	-	0.02	1.0002
	1000	10	30	5	(0, 0.3)	0.1	0.2	0.8663	0.02	-	0.5	1.0002
	1000	10	30	5	(0, 0.3)	0.1	0.2	0.50	0.02	-	0.8663	1.0002
	1000	10	30	5	(0, 0.3)	0.1	0.2	0.00	0.02	-	1	1.0002
	1000	200	30	5	(0, 0.3)	0.1	0.2	1.00	0.02	-	0.02	1.0002
	1000	200	30	5	(0, 0.3)	0.1	0.2	0.8663	0.02	-	0.5	1.0002
	1000	200	30	5	(0, 0.3)	0.1	0.2	0.50	0.02	-	0.8663	1.0002
	1000	200	30	5	(0, 0.3)	0.1	0.2	0.00	0.02	-	1	1.0002

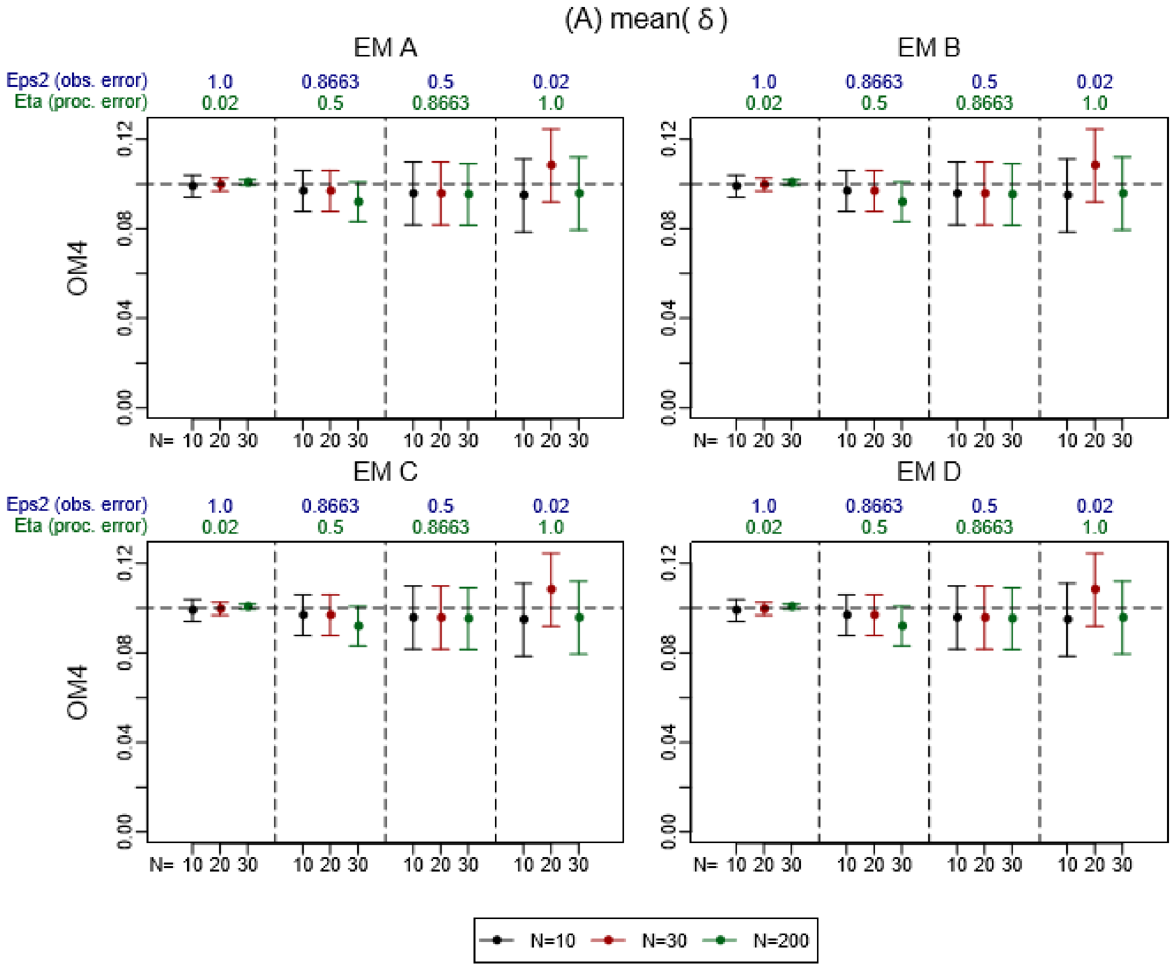

Figure 1A: Estimates of the mean for the closure effect $\boldsymbol{\delta}$ are shown for 1000 simulations of OM4 and the four EMs. In each case, results are shown for a range of four different values for each of the observation error $\boldsymbol{\epsilon 2}$ (values indicated by blue text above the plots) and process error $\boldsymbol{\eta}$ (values indicated by green text above the plots). Results are furthermore shown for a selection of three *N* values, where *N* is the number of penguins sampled each year: $\boldsymbol{N = 10}$ (black leftmost points), $\boldsymbol{N = 30}$ (red centre points) and $\boldsymbol{N = 200}$ (green rightmost points). The plots show the means and 95% confidence intervals[4] for the means of $\boldsymbol{\delta}$. 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.

[4] The 95% CI is taken to be +-1.96 standard error of the mean, which is calculated as the standard deviation divided by the number of simulations.

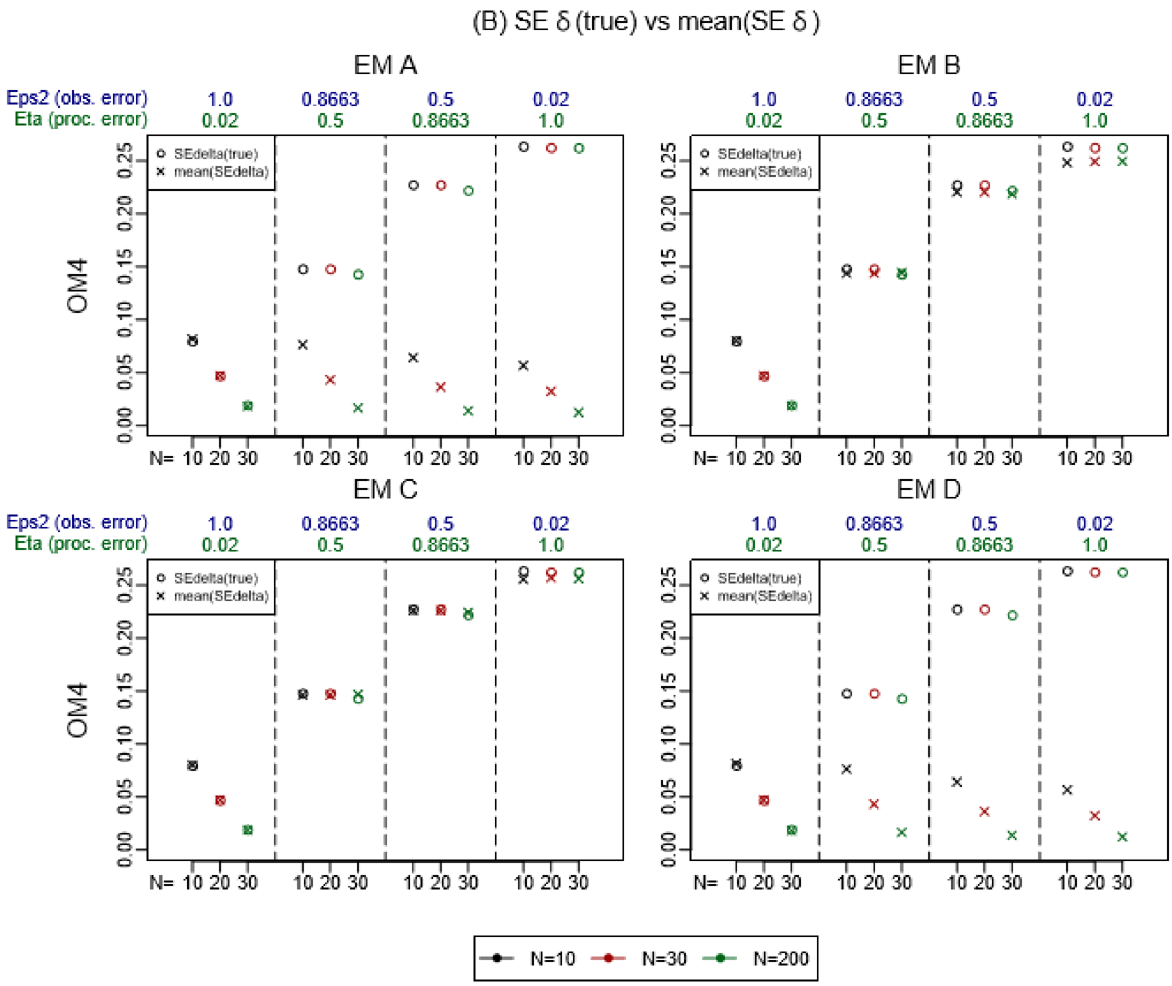

Figure 1B: Values for SE_δ(true) are shown by the open circles (o) and the values for mean(SE_δ) are shown by the crosses (x). SE_δ(true) is calculated as the standard deviation of the δ estimates across the 1000 simulations, and has been used to calculate the 95% confidence intervals for the mean δ estimate in Figure 1A. The statistic mean(SE_δ) is calculated as the average across the SE_δ values for the 1000 simulations. An "x" below an "o" indicates that the estimate of the standard error for that δ estimate is negatively biased.

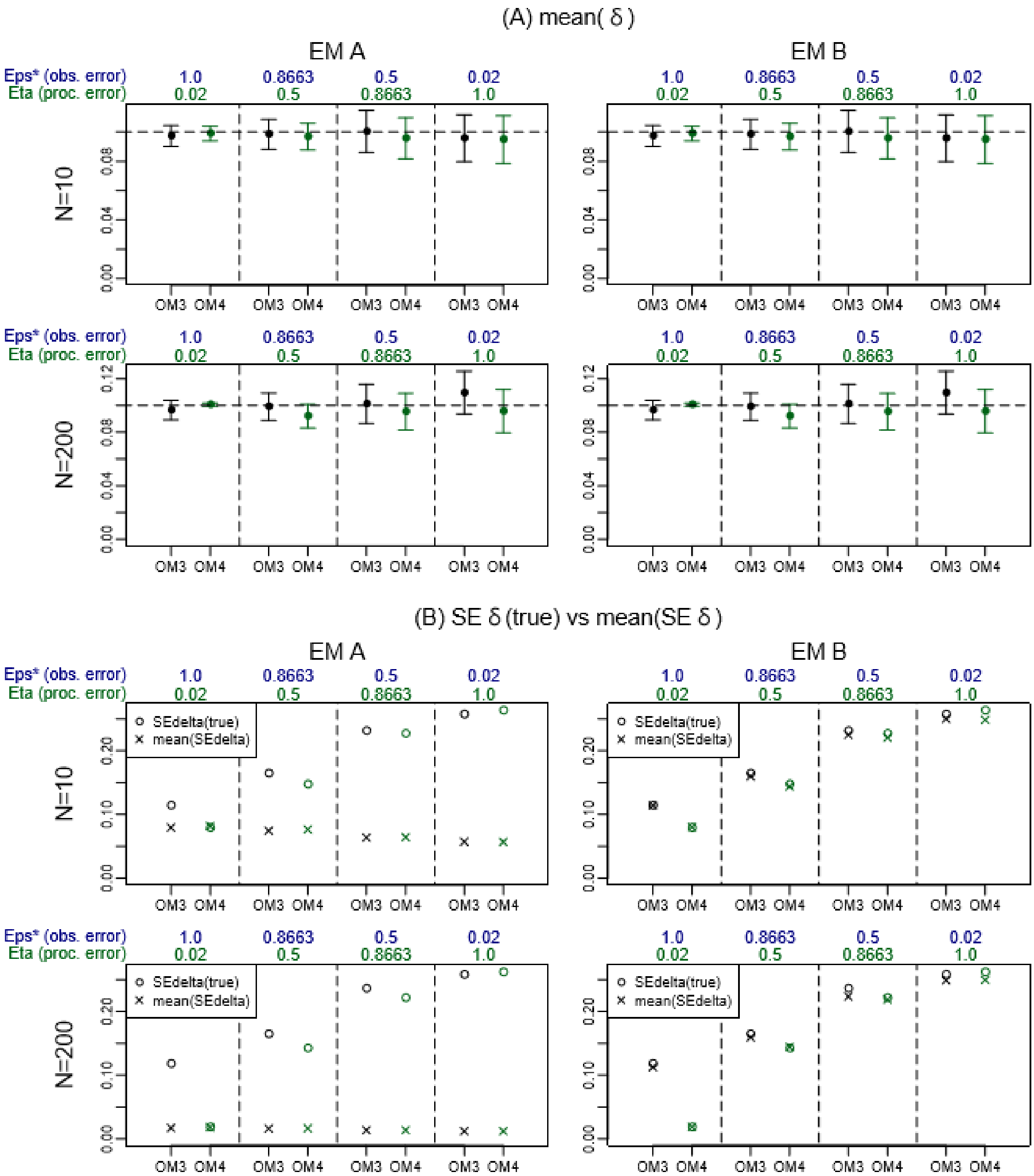

Figure 2: Results are shown for data generated by OM3 and OM4, and Estimation models EMA and EMB. The top section (A) shows the estimates for the mean of $\boldsymbol{\delta}$, while the bottom section (B) shows the estimates for $\boldsymbol{SE_\delta}$(true) and for mean($\boldsymbol{SE_\delta}$). Results are shown for the same range of observation and process errors as in Figures 1A and 1B, but only for $\boldsymbol{N = 10}$ and $\boldsymbol{N = 200}$. Note that the observation error Eps* for OM3 reflects the combination of *c* and $\boldsymbol{\epsilon}$, but for OM4 this is simply $\boldsymbol{\epsilon 2}$. Furthermore, for the last simulation (Eps*=0.02, Eta=1.0), $\boldsymbol{\sigma_c}$ is in fact zero for OM3 to keep the value of $\boldsymbol{\sigma_\epsilon}$ at 0.02 (see Table 1).

Appendix

Simulation testing methodology

The simulation test framework consists of four operating models (OMs) and four 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. $F_{i,y,z,j} = a_i + b_y + c_{i,y,z} + \delta(X_{i,y}) + \epsilon_{i,y,z,j}$
2. $F_{i,y,j} = a_i + b_y + \delta(X_{i,y}) + \epsilon 2_{i,y,j}$
3. $F_{i,y,j} = a_i + b_y + \eta_{i,y} + c_{i,y,z} + \delta(X_{i,y}) + \epsilon_{i,y,z,j}$
4. $F_{i,y,j} = a_i + b_y + \eta_{i,y} + \delta(X_{i,y}) + \epsilon 2_{i,y,j}$

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 $b_y \sim N(0,(\sigma_b)^2)$,
$\eta_{i,y}$	is an error term for island *i* and year *y*, representing process error, normally distributed $\eta_{i,y,j} \sim N\left(0,(\sigma_\eta)^2\right)$,
$c_{i,y,z}$	is an unknown/hidden covariate effect for island *i* , year *y* and covariate *z* (e.g. this could reflect different areas within the colony), and is assumed to be normally distributed with $c_{i,y,z} \sim N(0,(\sigma_c)^2)$,
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 observation error term for OM1 and OM3 for penguin *j*, where $\epsilon_{i,y,z,j} \sim N(0,(\sigma_\epsilon)^2)$
$\epsilon 2_{i,y,j}$	is an observation error term for OM2 and OM4, where $(\sigma_{\epsilon 2})^2 = (\sigma_\epsilon)^2 + (\sigma_c)^2$ so that the overall variance of the *F* values generated by the OM1 and OM2 pair (and alternatively the OM3 and OM4 pair) 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.

D. As for (A) (i.e. the model is fitted to annually dis-aggregated/individual data), but b_y is treated instead as a fixed effect.

Key output statistics

For each simulation $k = 1, 2, \dots, M$ 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 .