Introducing an imbalance in the sampling from the unknown covariate from OM3

A. Ross-Gillespie and D.S. Butterworth

This working paper presents results for an "OM5", which is a variant of OM3 (the Operating Model which includes the unknown covariate effect as well as process error) where each of the 200 penguins sampled at each island is allocated to one of the five covariates at random, rather than equal number of samples allocated to each covariate as is the case for OM3 (a test requested by Andre Punt). In practice, this was done by choosing the five $c_{i,y,z}$ values for each year from $N(0, \sigma_c^2)$, and using the sample function in R to randomly allocate a covariate number between 1 and 5 to each of the 200 annual samples from each island.

Results are shown for OM3, OM4 and OM5, and Table 1 provides the run specifications assumed for these, which correspond to the second $N = 200$ run for OM3 and OM4 in Table 1 of Peng/P7. Figure 1 plots the estimated values for mean(δ) and compares SEδ(true) with mean(SEδ). Figure 2 plots the generated response variables for the three OMs – these plots have been included for reference purposes and with the aim of clarifying the difference in data generated by OM3 (with unknown covariate) and OM4 (no unknown covariate).

There is very little difference between the results from OM3 and OM5 for the estimation properties for δ.

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. 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
$\sigma_{\epsilon2}$ 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	$\sigma_{\epsilon2}$	σ_η	$\sqrt{\sigma_{\epsilon2}^2 + \sigma_\eta^2 + \sigma_c^2}$
OM3	1000	200	30	5	(0, 0.3)	0.1	0.2	0.8660	0.02	-	0.5	1.0002
OM4	1000	200	30	-	(0, 0.3)	0.1	0.2	-	-	0.8663	0.5	1.0002
OM5	1000	200	30	5	(0, 0.3)	0.1	0.2	0.8660	0.02	-	0.5	1.0002

(A) mean(δ)

(B) SE δ (true) vs mean(SE δ)
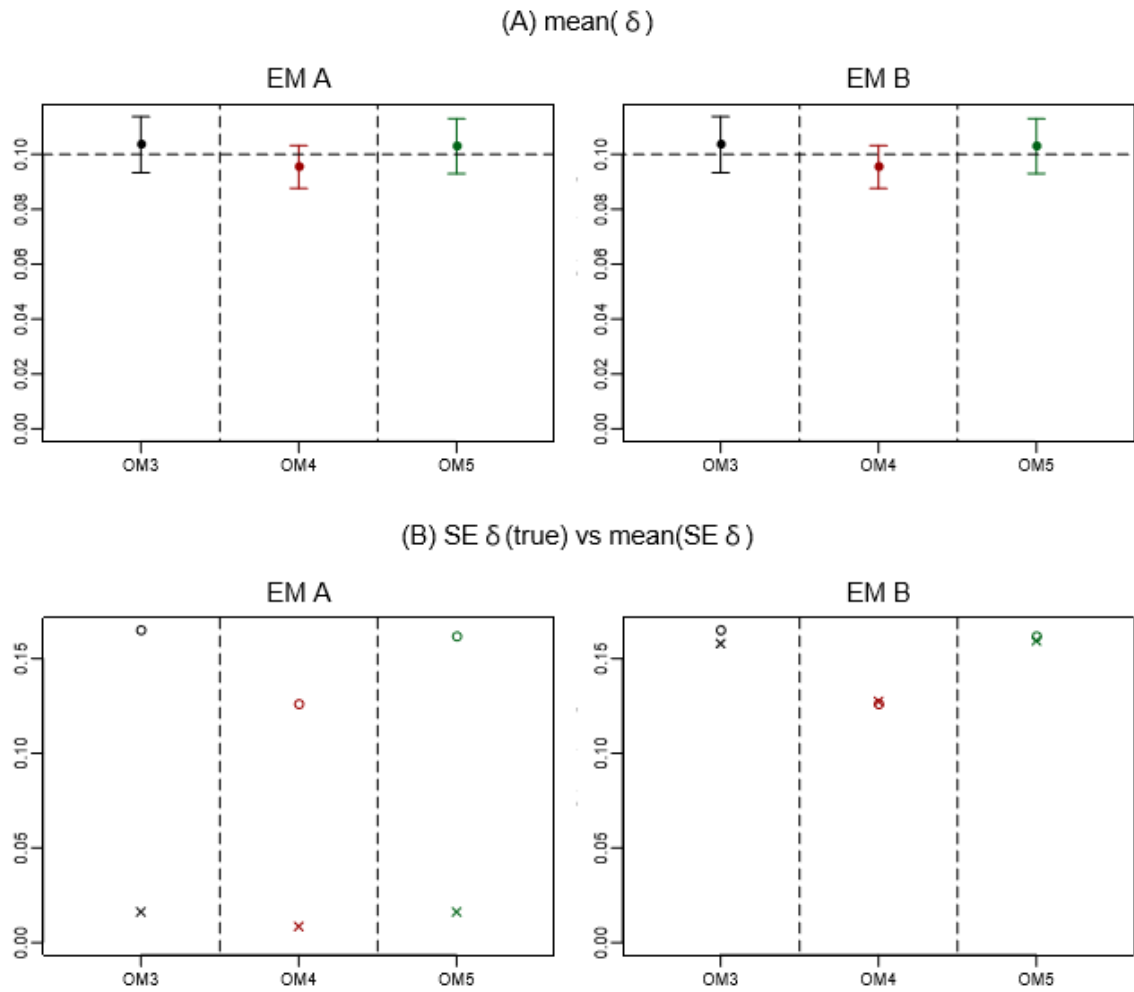

Figure 1: The top two plots show the estimates of mean(δ) and the 95% confidence intervals[1] are shown for OM3, OM4 and OM5 subject to estimators EMA and EMB. The bottom two plots show SEδ(true) (open circles) and mean(SEδ) (crosses).

---

[1] 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.
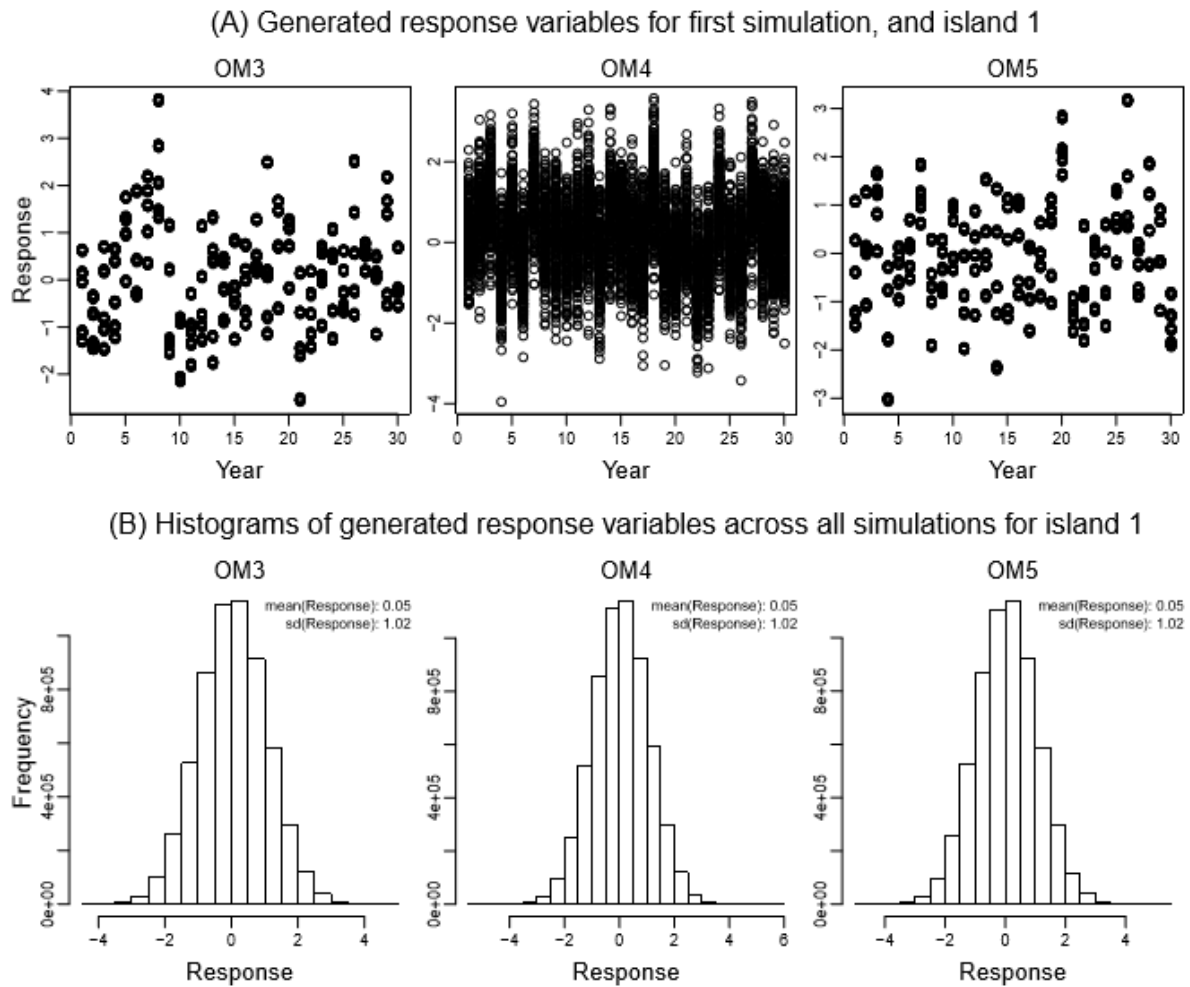
Figure 2: The top row shows the generated response variables for OM3, OM4 and OM5, for the first of the 1000 simulations, and for island 1. The bottom row shows histograms of the distribution of the response variable for island 1, for all the years and simulations combined.