

**Response to comments on Winker and Sherley's brief reply to FISHERIES/NOV/2019/SWG-PEL/34**

D S Butterworth

MARAM, UCT

**Abstract**

This document provides a point by point response to comments made by Winker and Sherley on paper FISHERIES/NOV/2019/SWG-PEL/34.

For readers' ease, comments have been inserted in *red italics* in the original text below and the response to the response is provided in *blue italics*.

**Brief reply to Butterworth and Ross-Gillespie: "Is pseudo-replication biasing results from analyses from the island closure experiment which model individual penguin responses directly?"**

Henning Winker & Richard Sherley

Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34) implemented a simple simulation experiment to explore potential bias in precision estimates resulting from pseudo-replication when fitting Generalized Mixed Effect Models (GLMMs) to individual observations. To attempt this, they introduced a 'hidden covariate' into the operating (simulation) model (OM), which they then ignored in their Estimation Model (EM) by only assuming a random year-effect for their simulation-estimation evaluations of the precision estimates.

*The 'hidden covariate' is only one mechanism potentially leading to (standard error) estimation bias that is explored in SWG-PEL/34; it is incorporated in operating model OM1. Importantly though (see below), operating model OM2 includes no such effect.*

This 'hidden covariate' is commonly known as unobservable 'latent effect' or 'latent state variable'. Indeed, it is widely accepted that, if ignored, such "latent states will generally cause model residuals to be correlated, violating the assumption of statistical independence" (Thorson and Minto, 2014), which can then lead to over-estimated precision and type II errors.

*Agreed.*

However, modelling individual observations with an appropriate hierarchal random-effects structure typically provides superior statistical power over an approach that uses aggregated means. This is exactly the reason why hierarchical mixed-effects models have been strongly advocated in both fisheries and ecological sciences over the past three decades as an important tool for **estimating the relative contribution of different hierarchical sources of variation** (e.g. Hilborn and Liermann, 1998; Gelman and Hill, 2007; Pinheiro and Bates, 2009; Zuur et al., 2009; Thorson and Minto, 2014).

*Indeed, they may for the purposes **grey highlighted**, though that may not be relevant in all cases to the quantity whose value it is of primary interest to estimate.*

*It is relevant to prevent loss of statistical power by omitting the information content (N given variance) associated with individual observations. The aggregated mean model is left about 8 degrees of freedom in the real world applications (even after ignoring the month effect), which is unsurprisingly too little for isolating an island effect in one of the most variable marine ecosystems in the world. The best example for poor predictive skill under a stochastic environment is probably sardine, which recently fell outside the range of projection envelope of plausible outcomes based on an arguably substantially more complex and comprehensive set OMs and conditioning models under the previous OMP. It is perhaps a different philosophy, in that FISHERIES/2019/NOV/SWG-PEL/34 aims optimally use the available degrees freedom, while accounting for non-independence through nested random effect structures and FISHERIES/2019/NOV/SWG-PEL/32 actively decreases the degree freedom to a minimum to avoid over-precision by all means, which come, however, with the trade-off of substantially reduced statistical power to detect effects in a variable system.*

Unlike the purposefully miss-specified EM in Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34),

*The estimator EMA using individual data in SWG-PEL/34 may indeed be mis-specified for OM1, which includes a “hidden” co-variate (though since what this co-variate is is unknown, it could not be otherwise). But be that as it may, EMA is NOT mis-specified for OM2, which does not include such an effect. And **importantly the bottom left panel in Fig 1B of SWG-PEL/34 indicates that there remains a substantial bias in the EMA estimates of the standard error of the closure effect parameter in this situation, once the number of individual penguin observations per island per year (N) exceeds 1.** Nothing in Winker and Sherley’s “Brief reply” negates this fundamental and important result, which even in isolation is sufficient to confirm the conclusions drawn by SWG-PEL/34.*

*The evaluation EMA for OM2 is essentially a ‘self test’. If this indeed holds true, SWG-PEL/34 just rejected the validity of fitting linear regressions to observations in principle, which would even make the principles of a length-weight relationship invalid. At least we can agree that N = 1 observation is an unbiased estimator of the population mean – albeit with zero variance I guess.*

models in Sherley et al. (2018) and Sherley et al. (FISHERIES/2019/NOV/SWG-PEL/32) do in fact account for hierarchical sources of variation that are implicit to the nested sampling design. For example, for the response ‘chick survival’ a random effect for ‘nest’ is introduced (Sherley et al. 2018), which is nested within the year effect, to accommodate latent effects (‘hidden covariates’) that cause variation in chick survival (e.g. due to different fitness of parents or area effects).

*The inclusion of the ‘nest’ effect in the SWG-PEL/32 model in this way is perfectly appropriate, and will indeed account for what would otherwise be one source of non-independence in the data; consequently, it will lessen the negative bias in variance estimation in individual based estimates. But the specific issue here is that the nest effect is not necessarily the only source of such non-independence. There are other sources which are not known (“hidden”), and for that reason could not be measured in conjunction with a particular penguin observation; hence also, they cannot be explicitly incorporated in a model to lessen this negative bias further. [There may be some semantic confusion here: “hidden” is used SWG-PEL/34 in the sense of the covariate responsible not being identified.]*

*There is no confusion here: a “hidden [covariate]” is used SWG-PEL/34 in the sense of the covariate responsible not being identified is widely modelled as latent state variable. More importantly, ‘nest’,*

*as well all other consider higher hierarchical random effects, are nested with in year. Therefore, SWG-PEL/34 explicitly considers that “those data [observations] are no longer independent within a year” to minimize the risk of pseudo-reflections.*

To conclude, we agree with Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34) that ignoring latent effects at a finer scale than accommodated by the random year effect increases the risk of negatively biased precision estimates. Given this is not the case in the analyses put forward by Sherley et al. (2018) and Sherley et al. (FISHERIES/2019/NOV/SWG-PEL/32), we refute the conclusions by Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34), and in particular that “past results concerning the statistical significance and probabilities that island closures impact penguins from analyses based on individual observations need to be reconsidered”.

*These conclusions have **not** been refuted by this “Brief reply”. Certainly, approaches such as including a “nest effect” will lessen the negative bias in precision estimates. But one has no a priori knowledge of the proportion by which this negative bias will be reduced – what about all the other “hidden” contributors to this bias through their introduction of further non-independence in the individual observations? HOWEVER, as emphasised above, this is a secondary concern. Even if this consideration is put to one side, as is the case when EMA is applied to OM2, the corresponding results show clear evidence of negative bias in standard error estimates when individual observations are used as input to the estimator which uses individual data.*

*FISHERIES/2019/NOV/SWG-PEL/34 essentially provides (i) ‘self tests’ with surprisingly varying results between aggregated mean and individual (regressions OM2) and (ii) a simplified misspecified EMs (OM2). It argues that the lack of independence of observations within the year effect increases the risk of negatively biased precision estimates whereas all model performs fine with regards to adequate Confidence Interval coverage. Given that the models in FISHERIES/2019/NOV/SWG-PEL/34 attempt to account for the non-independence – how can a simplified simulation experiment that is ignoring this can propose to refute the results published in Sherley et al. (2018) and then not be refuted?*

Instead we argue that the EM GLMM by Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34) for individual observations should have been correctly specified by introducing an additional (nested) random effect at a lower hierarchical structure to prevent similarly moot and even misleading conclusions in future.

*Again, this fails to explain the results obtained for EMA applied to OM2. Winker and Sherley’s hypothesis/theorem is that their estimator provides (near) unbiased estimates of variance for the effect of closure parameter  $\delta$ . Only one counter-example is required to falsify a theorem – the EMA-applied-to-OM2 results provide that counter-example.*

*None other than a very simple estimation procedure should be accepted (particularly when the results inform management decisions) without simulation testing. This is the current norm in fisheries science, and the testing approach followed in SWG-PEL/34 is absolutely standard in fisheries assessment work.*

*Agreed – I am indeed looking forward bringing, for example, Replacement Yield models to a simulation test. However, I am not sure if this is common practice providing evidence for a pre-cautionary approach to provide advice for steeply declining, high conservation priority species though – at least I am not aware of examples – other than IWC of course.*

*If Winker and Sherley wish to defend the individual-observation-based estimator which they promote as providing unbiased (or at least not negatively biased) estimates of variance, the onus is on them to provide results from simulation studies that demonstrate that this is the case. Such studies would need to include OM2 as one of the operating models considered. Data sets generated using OM2 will readily be provided to them for this purpose, should they so request.*

**No comment**

## **References**

- Gelman A, Hill J. 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge, UK.
- Hilborn R, Liermann M. 1998. Standing on the shoulders of giants: Learning from experience in fisheries. *Reviews in Fish Biology and Fisheries*, 8: 273–283.
- Pinheiro J, Bates D. 2009. Mixed-Effects Models in S and S-PLUS. Springer, New York, NY.
- 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 effects of experimental fishery closures on marine predator demographics. *Proceedings of the Royal Society B: Biological Sciences*, 285: 20172443.
- Thorson JT, Minto C. 2014. Mixed effects: a unifying framework for statistical modelling in fisheries biology. *ICES Journal of Marine Science*, 72: 1245–1256.
- Zuur AF, Ieno EN, Walker NJ, Saveliev AA, Smith GM. 2009. Mixed effects models and extensions in ecology with R. Springer, New York, NY.