

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 Winker¹ & Richard Sherley^{2,3}

¹*Department of Environment, Forestry and Fisheries (DEFF), South Africa*

²*Environment and Sustainability Institute, University of Exeter, UK*

³*FitzPatrick Institute of African Ornithology, University of Cape Town, South Africa*

Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34) implemented a simple simulation experiment to explore bias in precision estimates of the island closure effect that can arise from pseudo-replication when fitting Generalized Mixed Effect Model (GLMMs) to individual observations. Based on their findings they made wide ranging conclusions, for example, that “Previous analyses [by Sherley et al. 2018 and Sherley PEL32] should ideally be repeated based on year-aggregated inputs, and future analyses need to avoid repeating this earlier approach.” In this paper, we use a ‘common sense approach’ in a first instance to refute this claim by arguing that their experiment (1) fails a critical ‘self-test’, which potentially renders any conclusion by FISHERIES/2019/NOV/SWG-PEL/34 as invalid based on first principle and (2) that the authors have mis-specified the relevant estimation model to support their claim, which prevents any meaningful inference about the performance of more adequately specified Bayesian hierarchical mixed-effect models by Sherley et al. 2018 and Sherley PEL32.

The Butterworth and Ross-Gillespie simulation experiment

To explore potential bias on precision estimates of the island closure effect δ as a result of pseudo-replication (violation of independence in the data), the authors introduced a ‘hidden covariate’ into their first operating (simulation) model (OM1) and subsequently omitted the ‘hidden covariate’ again from OM2, presumably to provide a ‘control’. Both OMs were intended to broadly resemble the design of the island closure experiment on penguin response metrics in a simplified manner, such that:

$$F_{i,y,z,j} = a_i + b_y + \delta(X_{i,y}) + c_{i,z,y} + \varepsilon_{i,y,z,j} \quad (\text{OM1})$$

$$F_{i,y,j} = a_i + b_y + \delta(X_{i,y}) + \varepsilon_{i,y,j} \quad (\text{OM2})$$

where $F_{i,y,z,j}$ is the response variable on log-scale, a_i is the island effect for $i = 1, 2$ (fixed effect), b_y is the normally distributed year effect $b_y \sim N(0, \sigma_b^2)$ (random effect) and δ is the binary closure effect

for a vector with a sequence of 0's (closed years) and 1's (open years), such that a negative δ implies a positive closure effect (opposite to Sherley et al. 2018). The key difference between the two OMs is the inclusion of the 'hidden covariate' $c_{i,z,y}$ term into OM1, which is realized by the fixed effect term $c_{i,z,y}$ with 1, 5 or 10 factorial levels z within each island i in year y and normally distributed effect sizes $c_{i,z,y} \sim N(0, \sigma_c^2)$. The error term in OM1 is assumed to be normally distributed for the measurement error of penguin j , given covariate z on island i in year y , such that $\varepsilon_{i,y,z,j} \sim N(0, \sigma_\varepsilon^2)$. The error term $\varepsilon_{i,y,z,j}$ is then adjusted for OM2 by incorporating the variance σ_c^2 from the now omitted 'hidden covariate' into the OM2 error term, such $\varepsilon_{i,y,j} \sim \sigma_c^2 + \sigma_\varepsilon^2$.

The authors consider three estimation models (EMs) in their simulation experiment, which specifically focusses on potential bias in the estimated precision for δ in the form of the standard errors (SE). They define bias as the difference between the $mean(SE_\delta)$, computed from the mean of the estimated $SE_{\delta,k}$ across the simulation replicates k , and the 'true' $SE_\delta(true)$, computed from the standard deviation of the estimated δ_k across 1000 simulation replicates k . We note that this bias calculation seems to have the undesirable property of relying on unbiased estimates of δ_k in the first instance (here confirmed to be approximately unbiased for runs 4-11 in Fig. 1A), which may not necessarily hold in a less idealized OM setup that considers for example an unbalanced sampling design. The three estimation models are specified as:

$$F_{i,y,j} = a_i + b_y + \delta(X_{i,y}) + \varepsilon_{i,y,j} \quad (EMA)$$

$$\bar{F}_{i,y} = a_i + b_y + \delta(X_{i,y}) + \varepsilon_{i,y} \quad (EMB)$$

$$\bar{F}_{i,y} = a_i + \beta_y + \delta(X_{i,y}) + \varepsilon_{i,y} \quad (EMC)$$

where $\bar{F}_{i,y}$ in EMB and EMC denotes the mean value of the measured penguin response variable for island i and year y and β_y denotes the change to a fixed effect term for the year effect in EMC.

At this point, we point to two important properties of the simulation experiment design, which are critical for our argument below:

- (1) The combination of OM2 and EMA represent a "self-test" given the structure of both are identical, so that EMA is correctly specified with regards to OM2.
- (2) The combination of OM1 and EMA represents a clear misspecification of EMA, which, unlike Sherley et al. (FISHERIES/2019/NOV/SWG-PEL32), omits the appropriate nested random effects structure to account for non-independence introduced by the 'hidden covariate' in OM1.

1. Self-Test Failure

If the model specification of a data generating OM is identical to the model specification of the EM that is fitted to the generated data, all estimators and their variances should be in principle unbiased. OM2 and EMA represent such identical model specification, yet Fig. 1B in FISHERIES/2019/NOV/SWG-PEL/34 shows a strong negative bias, which we, in contrast to the authors, judge simply as implausible. We suspect that the authors recorded the routine standard error outputs for the estimated coefficient δ (e.g. using the function `se.coef()` for the fitted `lmer()` object in the R package `lme4`). However, from the R package documentation we understand that the returned standard errors are only approximations for the fixed effects and not the joint (sums of) random and fixed effects coefficients, which would typically require use of the delta method, bootstrap or MCMC to be estimated reliably. If our suspicion should indeed hold true, this would render the conclusions in FISHERIES/2019/NOV/SWG-PEL/34 obviously invalid in its current form.

2. Misleading EMA misspecification for OM1

Firstly, it is important to reiterate that all models presented in Sherley et al. 2018 and Sherley et al. FISHERIES/2019/NOV/SWG-PEL32 consider a hierarchical random effect structures with a random effect for year and a second random effect that is nested within year. Each of the choices are justified carefully but the common motivation was to minimize the risk of overestimating precision due to non-independence of observations. For example, for chick survival we specified a hierarchical random effects term for nest identity within year to account for non-independence of chick survival information within the same nest. This random effect for nest may also account for other unobservable (latent) effects such as the fitness of parents or some hidden covariate associated with the spatial location of the nest. The latter also provides an easy to interpret link to the hidden covariate $c_{i,z,y}$ as introduced in OM1, which essentially blocks each island in each year into n_c groups of penguin observations that are similar (non-independent) within each group and can vary notably across groups with respect to the response variable. With this in mind it is then straight forward to generalize a correctly specified EM for OM1 with the desirable property that it is also consistent with the model structure used in Sherley et al. (FISHERIES/2019/NOV/SWG-PEL32), such that:

$$F_{i,y,j} = a_i + b_y + d_{i,y} + \delta(X_{i,y}) + \varepsilon_{i,y,j} \quad (\text{EMD})$$

where $b_y \sim N(0, \sigma_b^2)$ is the random effect for year and $d_{i,y} \sim N(0, \sigma_d^2)$ is a random effect for island i , nested within year y . Note that the 'hidden covariate' still remains 'hidden' to EMD, but it is now

correctly specified by providing a means to account for the variation resulting from an unknown number groupings n_c on the island within a year that are associated with an unobservable effect on the response $F_{i,y,j}$.

We argue that not considering a hierarchical random effects term $d_{i,y}$ results in a severe misspecification of EMA with respect OM1. Therefore, it is evident that the interpretation of the results in Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34) is highly misleading and that simulation experiment in its current form fails to provide the basis for a 'fair' and meaningful comparison to the more adequately specified Bayesian hierarchical mixed-effect models used in Sherley et al. 2018 and Sherley et al. (FISHERIES/2019/NOV/SWG-PEL32).

Discussion

The term 'hidden covariate' is also 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. However, modelling individual observations with an appropriate hierarchical random-effects structure typically provides superior statistical power over an approach that uses aggregated means, which is basically left with severely decimated degrees of freedom. This is probably one of the main reasons 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). Unlike the miss-specified EMA in Butterworth and Ross-Gillespie (FISHERIES/2019/NOV/SWG-PEL/34), 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. We have highlighted this for the response 'chick survival', where we had specified a random effect for 'nest' (Sherley et al. 2018), which was nested within the year to accommodate latent effects ('hidden covariates') that can cause variation in chick survival (e.g. due to different fitness of parents or area effects).

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. We argue, however, that their EMA for individual observations should have been correctly specified by introducing an additional (nested) random

effect at a lower hierarchical structure (i.e. EMD) to prevent similarly moot and even misleading conclusions about the hierarchical mixed-effects models used in Sherley et al. (2018) and Sherley et al. (FISHERIES/2019/NOV/SWG-PEL32) in future. Finally, we suggest that it should have been on the onus of the authors to identify and first further explore the apparent inconsistency related to the 'self-test failure' (EMA-OM1) before choosing to make such strong recommendations that "past results concerning the statistical significance and probabilities that island closures impact penguins from analyses based on individual observations need to be reconsidered", which we herewith refute.

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.