Additional analysis suggested in response to differences in variance estimates between Sherley (2016) and Ross-Gillespie & Butterworth (2016).

Richard B. Sherley

Environment and Sustainability Institute, University of Exeter, Penryn Campus, Cornwall, TR10 9FE, United Kingdom; Department of Biological Sciences, University of Cape Town, Private Bag X3, Rondebosch 7701, South Africa. E-mail: r.sherley@exeter.ac.uk

Sherley (2016) outlined a Bayesian approach to understand the effect sizes, uncertainty and demographic impact associated with purse-seine fishing closures around African penguin colonies. *Inter alia*, the approach in Sherley (2016) used linear mixed effects models to analyse 9,436 individual chick condition observations. Linear mixed effect models were used because of "the flexibility they offer in modeling the within-group correlation often present in grouped data" (Pinhero & Bates 2000) because they can "account for dependencies within hierarchical groups through the introduction of random-effects" (Zuur et al. 2009). Their use is now commonplace in ecological analyses and they have been advocated for and used in fisheries management for some time (Venables & Dichmont 2004; Miller & Anderson 2004; Punt et al. 2006; Thorson & Minto 2015; Thorson et al. 2016), including by members of MARAM (e.g. Brandão et al. 2004). To quote Venables & Dichmont (2004), writing over a decade ago: "One of the most important benefits of using mixed models is their capacity to 'borrow strength' from one part of the data to another, thus often providing a more realistic analysis of large fragmentary data sets, which are the norm in fisheries research". Or as Punt et al. (2006) put it: "there is value in using a mixed-effects approach to allow the years for which the dataset is large to 'provide support' for the years for which the data are sparse".

In contrast, Butterworth (2016) has argued that there is "nothing to be gained in terms of improved estimation performance by fitting to the individual data for each year rather than to their means", that "one cannot assume that a random effects estimator will fully correct for non-independence of data; rather it seems likely to yield estimates of standard errors for parameters which are negatively biased to some extent" and based on a simulated dataset with "fairly strong non-independence" ignoring the non-independence in the data would yield standard error estimates "an order of magnitude too small". In other words, there is a concern that the smaller standard error estimates for the effect of closure on chick condition in Sherley (2016) compared to those in Ross-Gillespie & Butterworth (2016) results not from the additional power gained by using mixed models, but rather because the analysis in Sherley (2016) fail to appropriately correct variance estimates for the effects of non-independence in the underlying data.

Unfortunately, a number of technical issues make it difficult to compare the standard error estimates associated with effect sizes (μ_{data}^{EM} in Ross-Gillespie & Butterworth 2016) between the two analyses directly. Notably, Ross-Gillespie & Butterworth (2016) worked in log-space on the aggregated means and Sherley (2016) worked in normal space on the raw data. The datasets used also differed: 2004, 2008–2013 in Ross-Gillespie & Butterworth (2016) and 2008–2015 in Sherley (2016). Thus as Butterworth (2016) notes "actual truly comparable results have yet to be calculated", but a number of suggestions have been made by members of the Penguin Task Team as to how we might determine whether any differences will greatly impact the key results from these analyses:

(1) Mike Burgh suggested comparing results from Sherley (2016), with those using a shorter (than one month) time period for the lowest level of the hierarchical random effect, specifically a two-week period. In addition, Doug Butterworth suggested (2) omitting biomass estimates as explanatory variables; refitting the model to the time range of data used in in Ross-Gillespie & Butterworth (2016), taking values from the joint posterior distribution for the differences between open and closed years at each island, logging each value in each pair, taking the differences for each pair, and then computing the standard deviation of those differences and (3) comparing directly between models fit to the same disaggregated data and the corresponding aggregated means. To address (3), I used the 'nlme' library in R.

Here, I have dealt with all four suggestions, but have used the entire joint posterior (rather than sample from it) and have also (4) repeated Doug's first suggestion (2) with the data used in the original Sherley (2016) model (2008–2015 for both islands). Details of the Bayesian hierarchical models used can be found in Sherley (2016) and are not repeated here.

Results

Note I have used closed years as the baseline here so that negative effects sizes from all models correspond to negative estimates in Ross-Gillespie and Butterworth (2016) and represent high chick condition in closed years.

Using the same aggregated data as Ross-Gillespie and Butterworth (2016) and working with the natural log of the condition data, I obtain near identical estimates (compare A and 3 below), with the remaining difference likely due to rounding.

Using the same data range (2004, 2008–2013) as Ross-Gillespie and Butterworth (2016), but refitting the model in Sherley (2016) yielded essentially unchanged results, though the effect size at Dassen is closer to zero (compare B and C below), and the precision estimates still appear an order of magnitude smaller than in Ross-Gillespie and Butterworth (2016; compare A and C); but note, the estimates in A are in log space and those from B and C in normal space.

Reducing the time-step used as the lowest level of the hierarchical random effect to fortnight (still using 2004, 2008–2013 data) produced essentially unchanged precision estimates in normal space (compare C and 1 in the table below) and log space (compare 1 and 2), though both point estimates are now negative as in Ross-Gillespie and Butterworth (2016).

Finally, Doug's suggestion of calculating the standard deviation of the differences of the logged posterior distribution for open and closed years yielded precision estimates that were no longer an order of magnitude smaller than those of Ross-Gillespie and Butterworth (2016), but about 50% smaller (compare A and 2 in the table below). This was not greatly influenced by which data time period was used (compare 2 and 4, and both with A in the Table below).

Source/Data range	Model type	Data type	Island	Effect size	SE/SD
(A) Ross-Gillespie and Butterworth 2016 from Table 6[1]	Log LMM(?)	Agg.	Dassen	−0.08	0.22
			Robben	−0.13	0.20
(B) Sherley 2016 (results from fit with biomass omitted)[2]	LMM (JAGS)	Disagg.	Dassen	0.02	0.02
			Robben	−0.11	0.03
(C) As Sherley 2016, but fit to 2004, 2008–2013 data[2]	LMM (JAGS)	Disagg.	Dassen	0.003	0.03
			Robben	−0.12	0.03
(1) Adding fortnight to the hierarchical random effect (Year/Month/Fortnight)[1 and 2 shown]	LMM (JAGS)	Disagg.	Dassen	−0.05	0.03[2] (0.10[1])
			Robben	−0.06	0.03[2] (0.09[1])
(2) As Sherley 2016, but fit to 2004, 2008–2013 data, SD of difference of logged joint posterior[1]	LMM (JAGS)	Disagg.	Dassen	0.003	0.10
			Robben	−0.12	0.09
(3) 2004, 2008–2013 aggregated data, no biomass, year random effect[1]	Log LMM (nlme)	Agg.	Dassen	−0.08	0.23
			Robben	−0.12	0.20
(4) As (2) but fit to 2008–2015 disaggregated data[1]	LMM (JAGS)	Disagg.	Dassen	0.02	0.09
			Robben	−0.11	0.08

Notes: 1. Results are in log space; 2. Results are in normal space; Agg. = aggregated data, meaning that the annual means are used; Disagg. = disaggregated data, meaning each of the original observations made in the field is used; nlme = model fit using the nlme library in R; JAGS = model fit using Bayesian inference and Just Another Gibbs Sampler (JAGS); LMM = linear mixed model; Log LMM linear mixed model on log transformed data.

References

Brandão A., Butterworth D.S., Johnston S.J., Glazer J.P. 2004. Using a GLMM to estimate the somatic growth rate trend for male South African west coast rock lobster. Fisheries Research 70: 399–349.

Butterworth D.S. 2016. On the use of aggregated vs individual data in assessment models. Department of Agriculture, Forestry and Fisheries Report No. FISHERIES/2016/NOV/SWG-PEL/65. Pp. 1–6.

Pinhero J.C. & Bates D.M. 2000. *Mixed-effects models in S and S-Plus*. Springer Verlag, New York.

Punt A.E., Hobday D.K. & Flint R. 2006. Bayesian hierarchical modelling of maturity-at-length for rock lobsters, *Jasus edwardsii*, off Victoria, Australia. Marine and Freshwater Research 57: 503–511.

Ross-Gillespie A. & Butterworth D. S. 2016. Penguin power analyses using the approach recommended by the international panel: methods and results. FISHERIES/2016/NOV/SWG-PEL/Peng/01. Pp. 1–31.

Sherley R.B. 2016. A Bayesian approach to understand the effect sizes, uncertainty and demographic impact associated with purse-seine fishing closures around African penguin colonies. Department of Agriculture, Forestry and Fisheries Report No. FISHERIES/2016/NOV/SWG-PEL/Peng/02. Pp. 1–28.

Thorson J.T. & Minto C. 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. ICES Journal of Marine Science 72: 1245–1256.

Thorson J.T., Rindorf A., Gao J., Hanselman D.H. & Winker H. 2016. Density-dependent changes in effective area occupied for sea-bottom-associated marine fishes. Proceedings of the Royal Society B: Biological Sciences 283: 20161853.

Venables W.N. & Dichmont C.M. 2004. GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research. Fisheries Research 70: 319–337.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A. & Smith G.M. 2009. *Mixed effects models and extensions in Ecology with R*. Springer, New York.