2019 Updated GLMM results for the South Coast penguin colony foraging data

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

Summary

The GLMM model to estimate the impact of fishing on penguins at the St Croix and Bird Island colonies is updated given further data. For trip length and duration data, there is no evidence to suggest a statistically significant impact of fishery on the penguin populations on these two islands, unless the influential 2016 data points are excluded from the analyses, in which case a significant impact would be indicated for the St Croix forage trip length variable. However, no *a priori* independent reason to treat 2016 data differently has been put forward to justify such an exclusion. If the maximum distance from the colony is considered, a significant (at the 5% level) negative impact of the fishery is indicated for St Croix.

Introduction

New foraging data have become available for the major penguin colonies on the South Coast, at St Croix and Bird islands, since the analyses of fishing impact were last conducted in 2016 (FISHERIES/2017/DEC/SWG-PEL/38). The catch and closure mixed effects models are re-applied to include these new data, as well as to the old data (which have been modified to a limited extent since 2016) for comparison purposes. This document has been revised from the original version to include results for maximum distance foraged as a response variable.

Methods

The mixed effects model is given by the equation:

$$-ln(F_{y,i}) = \alpha_y + \beta_i + \lambda_i \frac{C_{y,i}}{\bar{C}_i} + \delta_i X_i + \epsilon_{y,i} \quad (1)$$

where

$F_{y,i}$ is the foraging trip response variable (foraging trip length, duration or maximum length), which has been transformed by taking the negative of the logarithm so that a larger negative number implies a more negative impact on the penguin population,

α_y is a random year effect reflecting prevailing environmental conditions (assumed to be the same each year, random variation excepted, for both islands in a pair),

β_i is an island effect,

λ_i is a fishing effect occasioned by the size of the catch taken in the neighbourhood of the island,

$C_{y,i}$ is the sardine catch taken in year *y* in the 18.148km radius neighbourhood (the area effectively closed on occasion – see further discussion below) of island *i*,

$\bar{C}_i$ is the average catch taken over the years for which island *i* was open to the fishery between 2008-2013[2],

δ_i is a fishing effect occasioned by whether or not the neighbourhood around the island was closed to fishing,

X_i is a vector with an entry of 1 in years where island *i* is open to the fishery, and 0 where the

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

[2] The purpose of the $\bar{C}_i$ factor is to scale the catch values in Equation (1) and consequently lead to more readily interpreted λ estimate values, which become more comparable to estimated δ values. The year range 2008-2013 is as was used previously (e.g. MARAM/IWS/DEC14/Peng/B12rev, FISHERIES/2017/DEC/SWG-PEL/38), and has therefore also been used here to maintain comparability of the new with the earlier λ estimates.

island is closed, and
$\epsilon_{y,i}$ is an error term.

For the Catch model δ is set to zero and for the Closure model λ is set to zero.

The catch series used here correspond to the 18.148km radius "closure area" catches, as these are considered to define catches in the neighbourhood of an island which are representative of the area to which the fishery closures correspond. While the foraging data series have simply been extended from 2016 to 2018, the entire 18km catch series has been revised from what was used in FISHERIES/2017/DEC/SWG-PEL/38, as well as extended to 2018. This is because accurate catch positions have only been reported since mid-2011, and sardine catch data from later years are used to allocate catches for earlier years (see FISHERIES/2019/NOV/SWG-PEL/23**)** – thus addition of new data points results on a revision of earlier catch values. Consequently, the 18km catch series used in FISHERIES/2017/DEC/SWG-PEL/38 is not identical to the series available now and are reported in Table 2.

In light of this, results for five data and model combinations have been reported on in this document:

1. Closure-only model with the 2016 foraging data (i.e. duplication of results reported in FISHERIES/2017/DEC/SWG-PEL/38)
2. Closure-only model with the updated 2018 foraging data
3. Catch-only model with the 2016 foraging data and 2016 catch series (i.e. duplication of results reported in FISHERIES/2017/DEC/SWG-PEL/38)
4. Catch-only model with the 2016 foraging data and 2018 catch series, truncated at 2016
5. Catch-only model with the 2019 foraging data and 2018 catch series

Model (4) has been included to determine the impact (if any) that the updated catch series on their own have on the estimates of fishing impact. Models (2) and (5) provide the most up to date results. R software was used for the analyses, using the lmer function from the lme4 package; this implements REML estimation which provides unbiased estimates of variance.

Table 1 lists the foraging data as they are currently available, excluding the 2017 and 2018 values which are under a confidentiality agreement between the data collectors and DEA for a period of six months. Table 2 lists the 18.148km catch series that were available in 2016 and the updated series available now.

A key requirement for lack of (especially negative) bias in the results from these analyses is that there is no strong indication of a positive correlation between sardine catch and biomass, as taken to apply for the previous analyses of FISHERIES/2017/DEC/SWG-PEL/38; this appears to remain a justifiable assertion, for reasons provided in more detail in the Appendix.

Results and Discussion

Table 3 presents the estimates of fishing effect (λ for the catch model and δ for the closure model) for these five model fits. Figure 1 shows Zeh plots of the fishing effect along with the 95% confidence intervals. In addition to the five main models, Figure 1 also shows results for variants where the 2016 data points have been excluded, as these data are very influential, reflecting poor foraging results for St Croix despite the island being closed to fishing that year.

Figure 2a plots the data points along with the model predictions for the five models. The data points corresponding to 2017 and 2018 have been excluded in these plots in terms of a six-month confidentiality agreement with DEA. Figure 2b plots the residuals for the five model fits.

For Bird island, there is no evidence to suggest a significant (at the 5% level) impact of the fishery on the penguin foraging behaviour. For St Croix, the impact is estimated to be significant (relative to the -0.1 threshold value, which has been used in the West Coast power analyses as the level below which the impact of fishery on the penguin population would be biologically meaningful) for only the closure model applied to

the 2018 maximum distance data (Model 2). For all other models and data combinations, there is no evidence to suggest a significant impact of the fishery, unless the 2016 data points are excluded, in which case the fishing impact for St Croix for the closure model does become significantly less than zero, and also significantly less than the threshold value of -0.1. However, no *a priori* independent reason to treat 2016 data differently has been put forward to justify such an exclusion.

In general, for St Croix and forage trip length and maximum distance, the catch models estimate a smaller impact of fishing on the penguin population than the closure models. For Bird island and forage trip length, the effect is reversed, as the closure models estimate a slightly less negative impact. For forage trip duration, the patterns are less clear.

In summary, the addition of two further years of data has resulted in weaker indications than previously of any deleterious effect of fishing around these colonies. The Catch model indicates no effect at either, though the Closure model provides a weak indication of such at St Croix.

References

Bergh, M. O. 2014. Further clarification of the biases in and interpretation of regressisons where catch is a predictor of penguin response. Document MARAM/IWS/DEC14/Peng/A10 submitted to the 2014 International Stock Assessment Workshop, Cape Town, December 2014.

Butterworth, D.S., Moosa, N., Furman, L.B., Robinson, W.M.L. and Johnston, S.J. 2014. Updated Analyses of the Results from the Island Closure Feasibility Study for the Dassen/Robben and St Croix/Bird Island Pairs given Revised Data and Responses to Matters Raised in Documents MARAM/IWS/DEC14/Peng/A3. Document MARAM/IWS/DEC14/Peng/B12rev submitted to the 2014 International Stock Assessment Workshop, Cape Town, December 2014.

Coetzee, J. and Merkle, D. 2019. The current set of available data for evaluation of the Island Closure Experiment. Fisheries Branch document FISHERIES/2019/NOV/SWG-PEL/23.

Robinson WML. 2013. Modelling the impact of the South African small pelagic fishery on African penguin dynamics. PhD thesis, University of Cape Town. xiv + 207 pp.

Ross-Gillespie, A. and Butterworth, D.S. 2016. Penguin power analyses using the approach recommended by the international panel: methods and the complete set of results. Document MARAM/IWS/DEC16/Peng Clos/P1a_rev submitted to the 2016 International Stock Assessment Workshop, Cape Town, December 2016.

Ross-Gillespie, A. and Butterworth, D.S. 2017. Updated GLM results for the South Coast penguin colony foraging data. Fisheries Branch document FISHERIES/2017/DEC/SWG-PEL/38.

Table 1: The updated foraging data set (J. Coetzee *pers. comm.*). Note: Under Closure, 1 means open to the fishery and 0 means closed. The newest data, for 2017-2018, have been included in the analyses presented in this paper, are under a six-month confidentiality agreement with DEA and have thus not been reported in this table.

			Trip length (km)			Duration (hours)			Max dist from island (km)		
Island	Year	Closure	Mean	se	n	Mean	se	n	Mean	se	n
St Croix	2008	1	79.39	6.74	20	22.88	1.62	20	33.18	2.02	20
	2009	0	47.88	4.54	15	16.36	1.08	14	19.67	1.87	15
	2010	0	66.79	5.81	18	26.11	2.19	17	24.90	1.66	22
	2011	0	64.19	3.60	20	20.05	1.06	23	21.86	1.59	21
	2012	1	92.30	4.54	20	27.10	2.27	20	36.72	1.50	20
	2013	1	75.38	2.66	66	21.88	0.83	68	27.74	1.09	70
	2014	1	79.06	6.41	26	27.29	2.13	28	30.53	1.91	28
	2015	0	57.38	5.83	14	21.00	1.77	14	21.00	2.23	13
	2016	0	110.45	13.45	11	42.55	5.03	12	35.96	6.32	12
	2017	0									
	2018	1									
Bird	2008	1	37.73	2.04	30	14.91	0.80	20	14.33	1.23	30
	2009	1	41.02	2.20	15	18.19	1.19	15	12.71	0.97	22
	2010	1	51.99	5.14	26	20.20	1.42	18	15.37	1.37	35
	2011	1	41.06	3.51	26	20.70	1.60	20	15.42	1.53	48
	2012	0	56.03	1.56	45	18.88	0.63	20	16.02	0.87	55
	2013	0	38.89	1.70	77	12.59	0.61	66	12.30	0.48	82
	2014	0	54.83	2.88	25	18.43	1.10	26	16.70	0.65	30
	2015	1	43.93	2.35	29	15.95	1.09	14	14.32	0.83	31
	2016	1	44.13	2.49	27	17.52	1.15	11	12.85	0.75	27
	2017	1									
	2018	0									

Table 2: Catch (tons) and closure data are given for the two islands. A closure value of 1 means the island was open to the fishery and a value of 0 that it was closed. The catches correspond to the sardine catches taken within an 18.148km radius around each island, corresponding to the area from which the fishery is excluded in closed years (J. Coetzee, *pers. comm.*). The series under the column "Catch (2016)" lists the catch values as they were available in 2016 (and which were used in the analyses of FISHERIES/2017/DEC/SWG-PEL/38), while the column "Catch (2018)" lists the updated series as currently advised by J. Coetzee.

St Croix				Bird			
Year	Closure	Catch (2016)	Catch (2018)	Year	Closure	Catch (2016)	Catch (2018)
1987	1	0	0	1987	1	0	0
1988	1	0	0	1988	1	0	0
1989	1	7	6	1989	1	0	0
1990	1	108	90	1990	1	5	26
1991	1	302	256	1991	1	3	3
1992	1	144	118	1992	1	8	34
1993	1	92	73	1993	1	0	0
1994	1	234	183	1994	1	0	31
1995	1	316	253	1995	1	0	0
1996	1	421	323	1996	1	0	0
1997	1	25	19	1997	1	0	0
1998	1	1757	1387	1998	1	0	0
1999	1	1065	879	1999	1	0	0
2000	1	435	341	2000	1	0	0
2001	1	749	579	2001	1	0	0
2002	1	2813	2157	2002	1	2	6
2003	1	10097	7770	2003	1	0	0
2004	1	2679	2070	2004	1	11	56
2005	1	708	552	2005	1	15	23
2006	1	6188	5027	2006	1	209	510
2007	1	1638	1448	2007	1	627	2079
2008	1	312	264	2008	1	601	1136
2009	0	197	159	2009	1	218	1135
2010	0	1230	1017	2010	1	568	598
2011	0	617	276	2011	1	45	19
2012	1	3149	2732	2012	0	40	40
2013	1	2149	2065	2013	0	0	0
2014	1	407	401	2014	0	0	0
2015	0	0	0	2015	1	0	0
2016	0	126	0	2016	1	106	90
2017	0	-	479	2017	1	-	123
2018	1	-	5182	2018	0	-	0

Note: J. Coetzee (*pers. commn*) advises that catches taken within the 18.148 km closed area during years in which islands are closed to fishing are located just inside the boundary circle in the majority of cases, and may arise from vessels passively drifting into the closed area during pursing of the catch. Incorrect reporting of the catch position as well as deliberate disregard for the permit conditions cannot, however, always be ruled out.

Table 3: GLM estimates of the fishing effect parameter (δ for the Closure and λ for the Catch model) are given for these models applied to different data sets and foraging trip response variables. Standard errors are given in brackets. (Note that a negative value for either of these parameters implies fishing reduces penguin reproductive success, and *vice versa*.) The Closure model (1 and 2) and the Catch model (3 to 5) were applied to the foraging trip data set as it was available in 2016, and to the extended series available now in 2019, which includes values up to 2018. Additionally, there are two catch series available: the 18.148km closure catches as used in the 2016 analyses, and the 18.148km series available currently, which has been revised as well as extended to 2018 (see the main text for more details). The foraging trip response variables were foraging trip length, duration and maximum distance from island, which enter the regressions as the negative of the logarithm of the variable. The significance levels indicated are approximate because there are different views of how to account for degrees of freedom in random effects models. Grey highlighting has been used to indicate the model runs using the most recent data (2 and 5).

	St Croix	Bird
(1) **Closure** model, **2016** foraging data		
Length	-0.21 (0.13)*	0.13 (0.14)
Duration	-0.05 (0.16)	-0.09 (0.16)
Max distance	-0.28 (0.11)**	0.05 (0.11)
(2) **Closure** model, **2018** foraging data		
Length	-0.19 (0.12)*	0.06 (0.12)
Duration	-0.03 (0.13)	-0.14 (0.13)
Max distance	-0.30 (0.10)**	0.08 (0.11)
(3) **Catch** model, **2016** foraging data, **2016** catch data		
Length	-0.12 (0.13)	0.03 (0.11)
Duration	-0.03 (0.14)	-0.02 (0.12)
Max distance	-0.16 (0.11)	-0.00 (0.10)
(4) **Catch** model, **2016** foraging data, **2018** catch data		
Length	-0.12 (0.13)	0.07 (0.11)
Duration	-0.03 (0.14)	-0.01 (0.12)
Max distance	-0.16 (0.11)*	0.04 (0.09)
(5) **Catch** model, **2018** foraging data, **2018** catch data		
Length	-0.02 (0.07)	0.05 (0.10)
Duration	0.02 (0.07)	-0.03 (0.11)
Max distance	-0.05 (0.07)	0.04 (0.10)

*Estimate is more than 1.44se's away from zero (~15% significance level)

**Estimate is more than 1.96se's away from zero (~5% significance level)

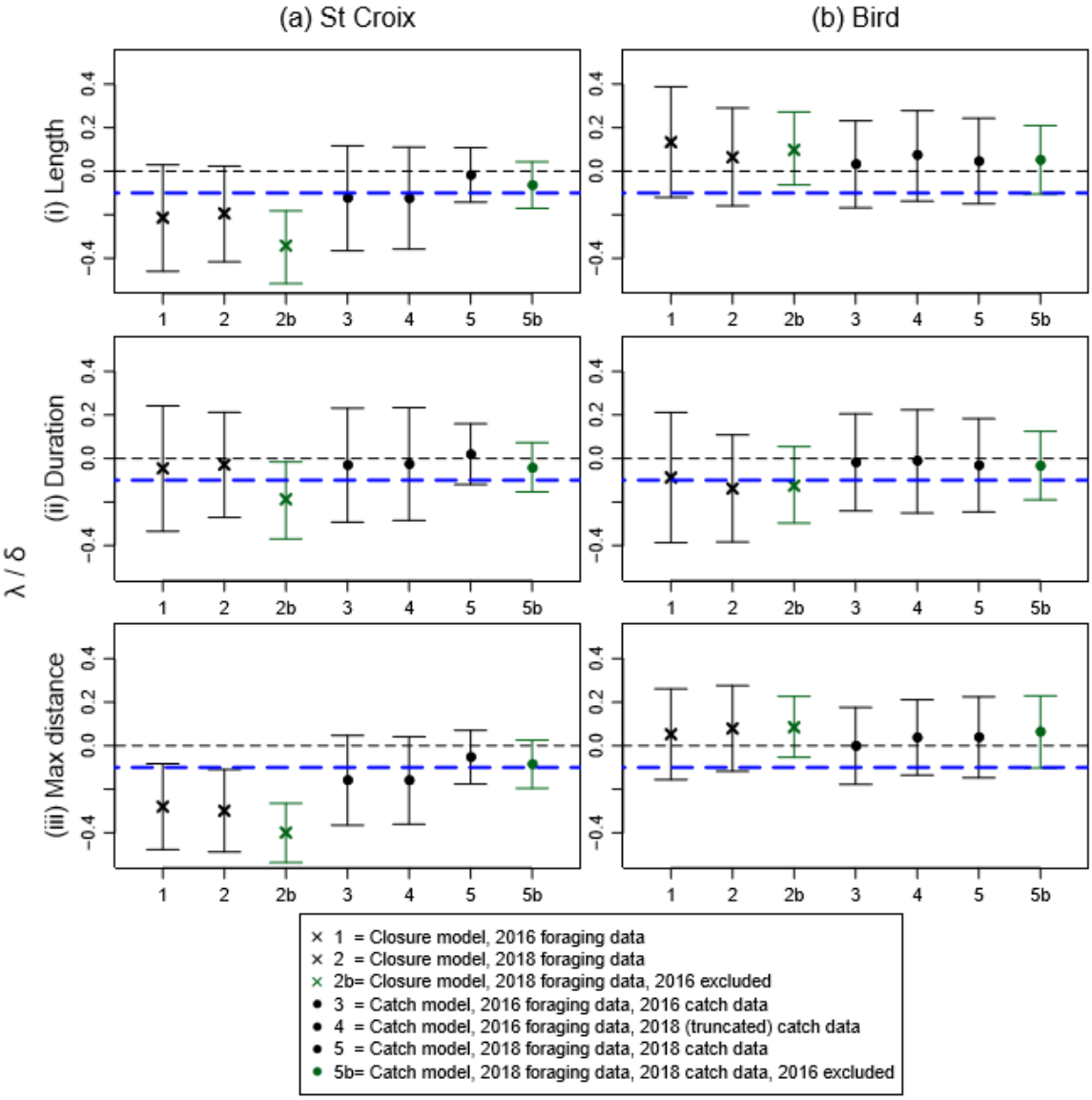

Figure 1: Zeh plots showing estimates of fishing effects in terms of $\boldsymbol{\lambda}$ (catch model) and $\boldsymbol{\delta}$ (closure model) with their 95% confidence intervals. The black horizontal dashed line marks zero, while the blue dashed line marks the threshold value of -0.1, which was considered in the West Coast power analyses as the value below which the impact of fishing on the penguin population would be considered biologically meaningful. The values for models 1-5 match those listed in Table 3. Additionally, 2b is a variant of 2 (closure model using the 2018 foraging data) where the 2016 data points are excluded (for both islands) – see the main text for the associated rationale. Similarly, 5b is a variant of 5 where those 2016 data points are excluded.

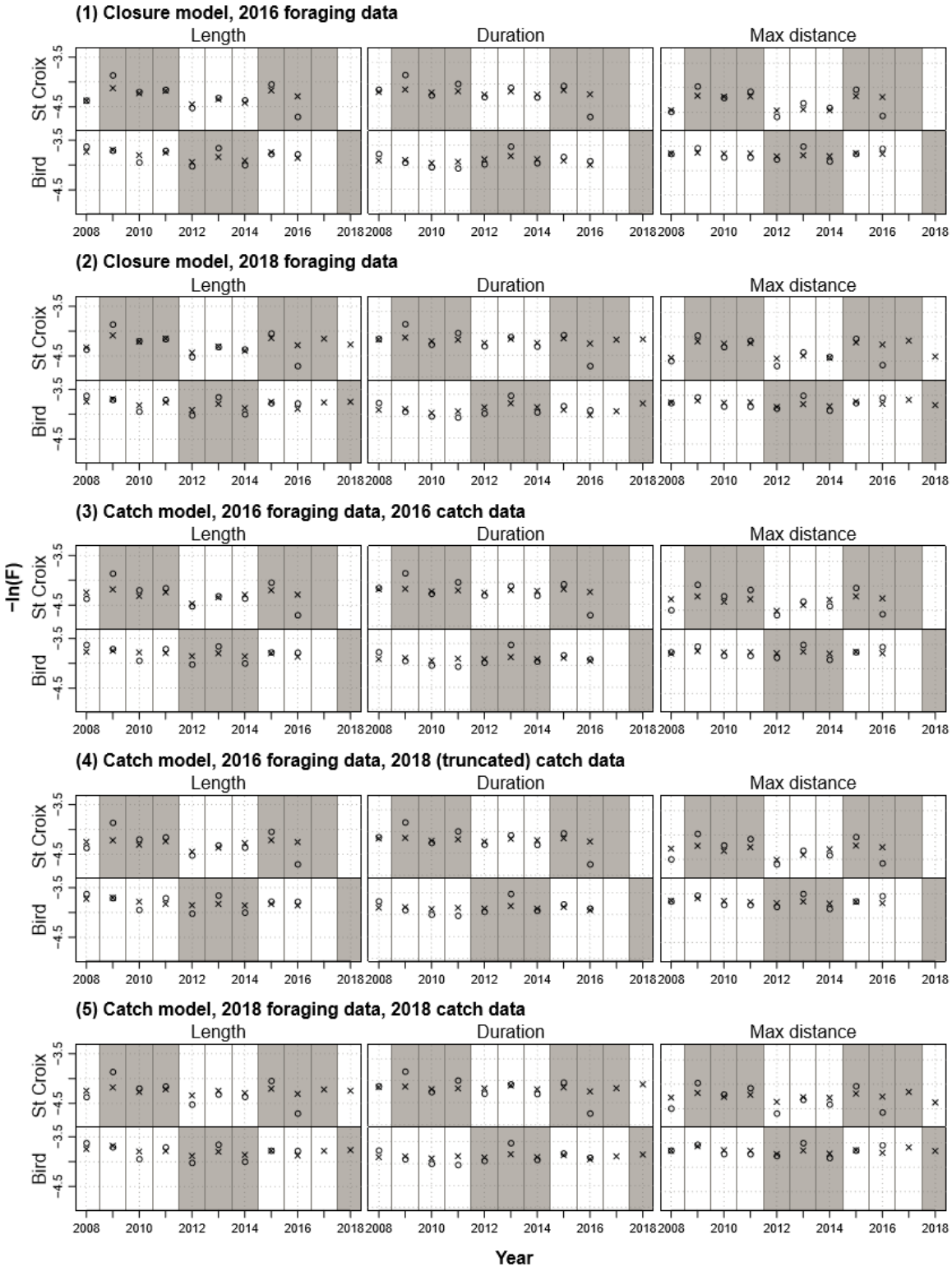

Figure 2a: Plots of the data points (open circles) and the model fits (crosses) for when the mixed model is applied to the full new data set. The grey shaded blocks mark the years for which the island in question was closed. Note that the 2017 and 2018 data points have been excluded from the plots in terms of a six-month confidentiality agreement with DEA.

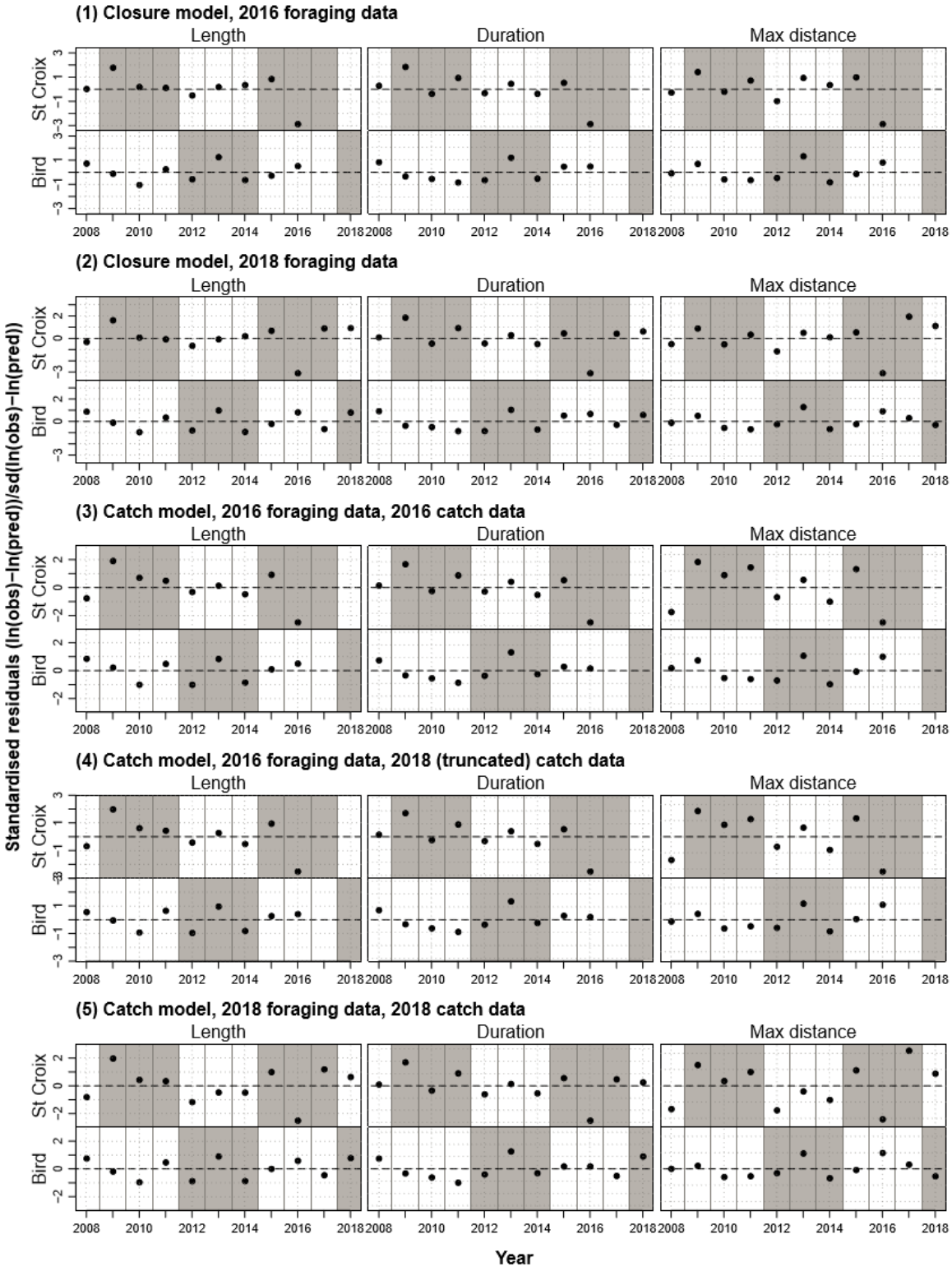

Figure 2b: Plots of the standardised residuals, calculated as ($\epsilon_{i,y}$)/sd(ϵ), where $\epsilon_{i,y} = \ln(obs_{i,y}) - \ln(pred_{i,y})$. The grey shaded blocks mark the years for which the island in question was closed.

Appendix: Evaluating the catch-biomass correlation for various 2016 and 2018 closure catch series

One might "expect" the catch model to be negatively biased as regards identifying a positive fishing impact on penguins because of the "Bergh effect" (which results when there is a positive correlation between catch and resource biomass – see Bergh (2014) - as was found to be the case for anchovy off Robben and Dassen Islands, see Robinson, 2013). In the previous assessment of the South Coast St Croix and Bird islands (FISHERIES/2017/DEC/SWG-PEL/38) it was found that there was no strong evidence for a positive correlation between catch and biomass, so that the Bergh affect could be ignored, and a simple mixed model could be applied rather than the complicated power analysis procedure (see MARAM/IWS/DEC16/Peng Clos/P1a rev) that was followed for the West Coast islands[3]. This Appendix explores whether this conclusion remains justified with two years' additional data available.

Table A. 1 lists the estimates of catch-biomass correlation, along with the 95% confidence intervals, for the 18km closure catch series as it was available in 2016 and for the current 18km closure series. These results suggest that there is no evidence for either island of a positive correlation between catch and biomass. Since the biomass estimates for Mossel Bay to Port Alfred are not available prior to 1996 (see Table 12 of FISHERIES/2019/NOV/SWG-PEL/23), the correlation analyses were also conducted for (c) the 20nm catch series and the sardine biomass estimates east of Agulhas, where data are available from 1987. For St Croix, when the entire series (1987-2016/18) is considered, the catch-biomass correlation is significantly greater than zero; however when the series corresponding to the years for which the experimental closures were implemented is considered (2008-2016/18), the correlation values are no longer significantly different from zero.

In order to ascertain whether the positive correlation between catch and biomass found for St Croix for the 20nm series when the whole data series is considered is a cause for concern, an analysis was performed whereby the starting year considered in the correlation calculation was increased incrementally from 1987 to 2008. These results are shown in Figure A. 1, and indicate that the correlation values are significantly greater than zero only when the early (pre-1997) values are included, a period for which where the catch values were all close to zero. It thus seems that these early years are highly influential in concluding "significant" correlations for the 1987-2018 period for the 20nm catch series in Table A. 1(c); when a year-range from 1998 onwards (when catches had increased to more substantial levels) is considered, the catch-biomass correlation is no longer significantly greater than zero. In light of this result – significance being found only when data from years where the fishery was clearly operating quite differently - it seems defensible to continue with the view that there is no substantial evidence for a positive catch-biomass correlation, and hence that the Bergh effect can continue to be ignored.

[3] Note that a glitch was found in the 2016 calculations for the correlation between C_y and B_{y-1}, and $(B_y + B_{y-1})/2$, which has been corrected for the values reported here.

Table A. 1a: Estimates of the correlation coefficient for the catch-biomass regression are given along with the 95% confidence intervals for the two islands for a regression between the sardine catches within 18.148km and sardine biomass from Mossel Bay to Port Alfred. The catch series used are (a) the 18km closure catches as used in 2016 and (b) the updated series taking data up to 2018 into account. Since the biomass estimates for Mossel Bay to Port Alfred are not available prior to 1996 (see Table 12 of FISHERIES/2019/NOV/SWG-PEL/23), the correlation analyses were also conducted for (c) the 20nm catch series and the sardine biomass estimates east of Agulhas, where data are available from 1987. The regressions were conducted using (i) catch in year *y* regressed against biomass in year *y*, (ii) catch in year *y* regressed against biomass in year *y*-1, and (iii) catch in year *y* regressed against the average biomass from year *y*-1 to year *y*. The correlation coefficients are shown for all the available data from 1987-2016/2018 and for the data subset from 2008-2016/2018. Values marked with an asterisk indicate a correlation coefficient significantly different from zero at the 15% significance level.

	St Croix					
(a) 18km original	(i) C_y to B_y		(ii) C_y to B_{y-1}		(iii) C_y to $(B_{y-1}+B_y)/2$	
1997-2016	0.36	(-0.18, 0.74)	0.41	(-0.13, 0.76)	0.44	(-0.10, 0.78)
2008-2016	-0.42	(-0.98, 0.91)	0.73	(-0.78, 0.99)	0.63	(-0.84, 0.99)
1997-2018	-	-	-	-	-	-
2008-2018	-	-	-	-	-	-
(b) 18km updated	(i) C_y to B_y		(ii) C_y to B_{y-1}		(iii) C_y to $(B_{y-1}+B_y)/2$	
1997-2016	0.33	(-0.22, 0.72)	0.37	(-0.17, 0.74)	0.40	(-0.14, 0.76)
2008-2016	-0.47	(-0.99, 0.90)	0.67	(-0.82, 0.99)	0.57	(-0.87, 0.99)
1997-2018	0.21	(-0.31, 0.64)	0.27	(-0.26, 0.68)	0.27	(-0.26, 0.68)
2008-2018	-0.62	(-0.97, 0.58)	0.32	(-0.78, 0.94)	0.19	(-0.83, 0.92)
(c) 20nm updated	(i) C_y to B_y		(ii) C_y to B_{y-1}		(iii) C_y to $(B_{y-1}+B_y)/2$	
1987-2016	0.49*	(0.12, 0.74)	0.67*	(0.38, 0.84)	0.61*	(0.29, 0.81)
2008-2016	-0.66	(-0.99, 0.82)	0.75	(-0.75, 0.99)	0.66	(-0.82, 0.99)
1987-2018	0.43*	(0.06, 0.70)	0.62*	(0.31, 0.81)	0.55*	(0.21, 0.77)
2008-2018	-0.85	(-0.99, 0.12)	0.37	(-0.76, 0.94)	0.17	(-0.84, 0.91)
	Bird					
(a) 18km original	(i) C_y to B_y		(ii) C_y to B_{y-1}		(iii) C_y to $(B_{y-1}+B_y)/2$	
1997-2016	-0.43	(-0.76, 0.06)	-0.37	(-0.72, 0.13)	-0.45	(-0.77, 0.03)
2008-2016	-0.34	(-0.90, 0.65)	-0.24	(-0.88, 0.71)	-0.42	(-0.92, 0.59)
1997-2018	-	-	-	-	-	-
2008-2018	-	-	-	-	-	-
(b) 18km updated	(i) C_y to B_y		(ii) C_y to B_{y-1}		(iii) C_y to $(B_{y-1}+B_y)/2$	
1997-2016	-0.40	(-0.74, 0.10)	-0.31	(-0.69, 0.21)	-0.40	(-0.74, 0.10)
2008-2016	-0.55	(-0.94, 0.47)	0.08	(-0.78, 0.84)	-0.50	(-0.93, 0.52)
1997-2018	-0.38	(-0.72, 0.11)	-0.28	(-0.66, 0.21)	-0.37	(-0.71, 0.12)
2008-2018	-0.54	(-0.92, 0.36)	0.16	(-0.68, 0.81)	-0.45	(-0.90, 0.46)
(c) 20nm updated	(i) C_y to B_y		(ii) C_y to B_{y-1}		(iii) C_y to $(B_{y-1}+B_y)/2$	
1987-2016	-0.20	(-0.54, 0.19)	-0.06	(-0.43, 0.33)	-0.14	(-0.49, 0.26)
2008-2016	-0.42	(-0.92, 0.59)	-0.60	(-0.95, 0.41)	-0.52	(-0.94, 0.50)
1987-2018	-0.19	(-0.53, 0.19)	-0.06	(-0.42, 0.32)	-0.13	(-0.48, 0.25)
2008-2018	-0.37	(-0.88, 0.53)	-0.06	(-0.78, 0.73)	-0.38	(-0.88, 0.52)

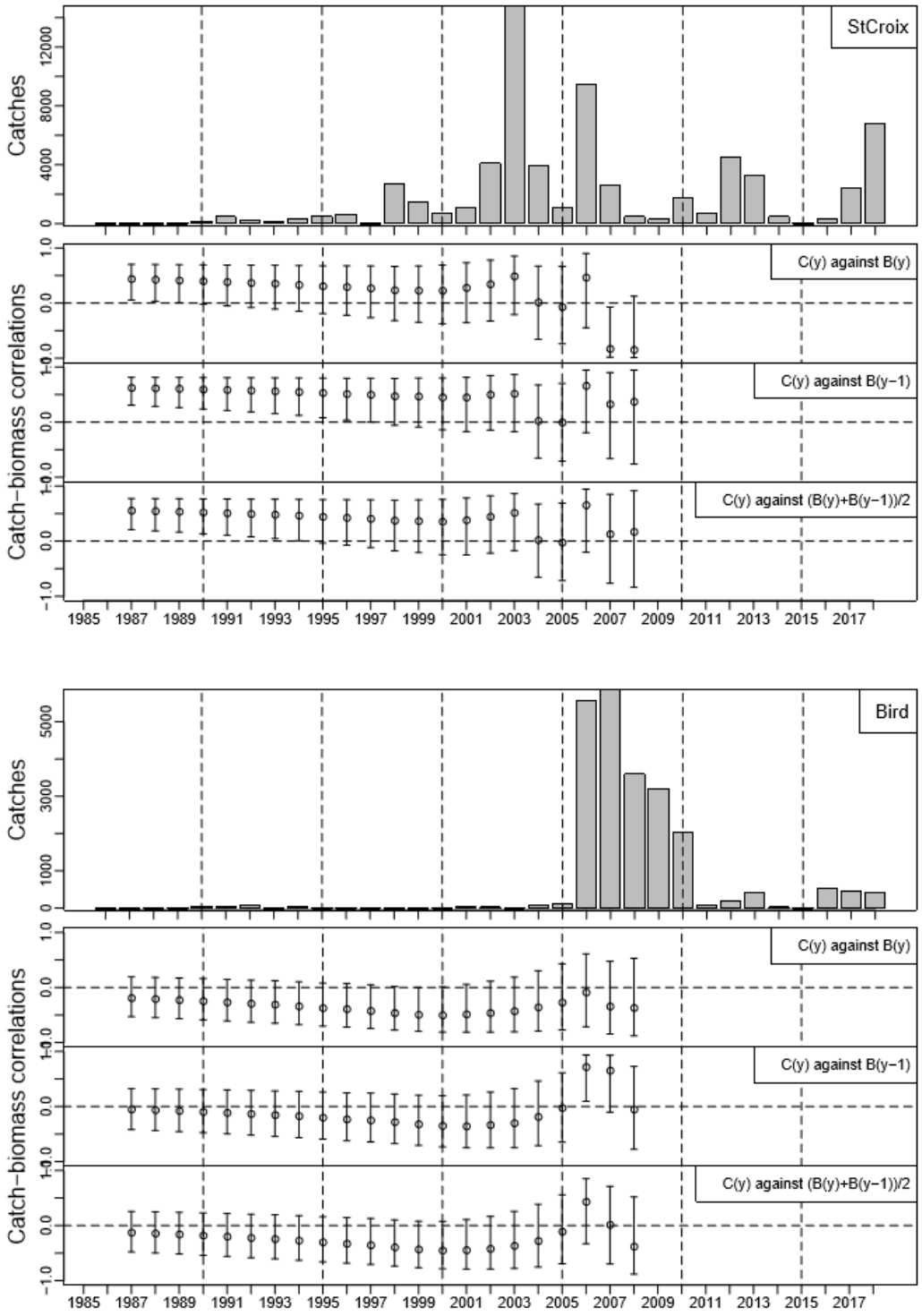

Figure A. 1: Catch-biomass correlation are calculated for the 20nm catch series and the sardine biomass estimates east of Agulhas while incrementally increasing the first year considered in the correlation calculation. For each island, the top panel shows a bar plot of the 20nm catch series as it is currently available. The three panels below show Zeh plots for the catch-biomass correlation values and 95% Cis for the three biomass series considered. The point at 1987 means data from 1987-2018 were used to calculate the correlation, the point at 1988 means data from 1988-2018 were used, and so on.