

Updated implementation of the Algorithm recommended by the Panel for the 2016 International Stock Assessment Workshop for assessing whether or not to continue with the penguin island closure experiment

A. Ross-Gillespie and D. S. Butterworth¹

Contact email: *mlland028@myuct.ac.za*

Summary

The analysis conducted late in 2016 to apply the algorithm recommended by the international Panel to evaluate the results of the island closure experiment is extended to include the further data now available, as well as to incorporate results further response variables for which data have been provided, making for a total of seven variables in all. For Dassen Island, all the response variables except for chick survival give little to no indication of a biologically meaningful negative impact of fishing on the penguin population. For Robben island, two of the data sets (chick condition and fledging success) provide evidence for a meaningful negative impact of fishing; the other response variables give little indication of such an impact. For all the variables not currently reflecting such an impact (except for chick survival at Robben island), at least an additional 20 years' data would be required before there is an 80% probability of detecting such an impact, if it is indeed present.

Introduction

In late 2016, a power analysis was conducted using data arising from the penguin island closure experiment in order to try to evaluate the power which this experiment has to detect a biologically meaningful impact of fishing on the penguin population at each island (see Ross-Gillespie and Butterworth 2016a). The Panel for the 2016 International Stock Assessment Workshop (IWS 2016) had developed an algorithm for interpreting the results of this power analysis, in the context of providing advice as to whether or not the closure experiment should be continued (Dunn *et al.* 2016). The algorithm was implemented in Ross-Gillespie and Butterworth (2016b), and those analyses are now updated given further data now available. The data used in 2016 and now in 2020 are summarised in Table 1.

The broad steps of the power analysis approach, and the Panel's algorithm, are to (a) set a threshold (which has a negative value) for the magnitude of the fishing effect on the penguin population at an island colony, below which the impact of fishing can be considered biologically meaningful², (b) simulate past and future data using an

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

² More details regarding the threshold are provided in the glossary in Appendix A. As for the 2016 analyses, the value of the threshold has been taken to be -0.1. This threshold has a negative value because the fishing effect parameter value is the more negative, the greater the adverse effect of fishing; a fishing effect parameter value *below* the threshold is considered biologically meaningful.

Operating Model (OM)³ assuming a range of values for the true fishing effect, and (c) apply the Estimation Model (EM) advocated to the simulated data to calculate the probability of concluding that the fishing effect is less than the threshold (and hence biologically meaningful), given the knowledge of the value of this effect assumed to generate the data. In other words, the analyses aim to answer the question that if the true effect of fishing on the penguin population is less than the negative threshold, what is the probability of coming to this conclusion by analysing the data arising from the experiment. A further aspect of the analyses is that if there is insufficient information in the data to conclude whether there is a biologically meaningful effect of fishing, then (a) are there values for the fishing effect that can be excluded (as these should have been detected by now), and (b) how long does the experiment need to continue before a meaningful effect is likely to be detected, if it is present.

Methodology

The four steps of the Panel's algorithm (see recommendation C.1.2 of Dunn *et al.* 2016) are reproduced below, with some minor edits to enhance clarity. Note that there are five Operating Models (OMs) used to generate data (a catch-only OM with a catch-anchovy biomass correlation of $\phi = 0.2$, a catch-only OM with $\phi = 0.4$, a closure-only OM, and two catch+closure OM with correlations of 0.2 and 0.4) and one Estimation Model (EM) – the closure only EM – applied to these generated data to estimate the fishing effect. The fishing effect estimated by the catch models is denoted by λ , and the fishing effect for the closure models is denoted by δ . Since the Panel recommend the use of the closure EM only for the Dassen and Robben analyses, the letter δ will be used to denote fishing effect in the descriptions that follow.

1. Evaluate the impact of fishing by applying the Estimation model (EM) to the actual historical data.

Compute the (cumulative) probability that the fishing effect δ is less than the Threshold (currently -0.1), using Equation 2 of Ross-Gillespie and Butterworth (2016a) (this cumulative probability is denoted " X "). See Figure 1a for some examples of such cumulative probabilities. The Panel recommended using the closure EM only (i.e. to exclude results from the catch-only EM), as the closure-only EMs resulted in P_{min} bias correction factors that were closer to 0.5 and were more robust across the different Operating Models (OMs) for the Dassen and Robben analyses (See recommendation C.1.5.1 of Dunn *et al.* 2016).

2. If there is evidence in the data of a biologically meaningful fishing effect, no further work is needed.

If X is greater than P_{min} ⁴ for the response variable concerned, it can be concluded that there is a fishery effect because δ is less than the threshold corresponding to a biologically meaningful impact on the penguin growth rate. No further observations are required for this response variable.

3. If there is not sufficient evidence that the fishing effect is biologically meaningful, are there values of the closure effect δ that can be excluded?

If X is less than P_{min} then there is not sufficient information to conclude that $\delta < T$. In this case:

³ Details of the mixed effects model used for the OMs and EM are given in Appendix B.

⁴ In order to evaluate P_{min} , the historical data are replaced with simulated data, which are generated assuming that the true δ is equal to the Threshold. P_{min} is the value for which the proportion of the total simulations for which $P(\delta < T) > P_{min}$ is 0.5 (i.e. half the total simulations have an area under the normal curve to the left of the Threshold that is greater than P_{min}). For a balanced design, and an unbiased estimator, P_{min} should be close to 0.5.

- (a) generate historical data given a set of values specified for δ (with data points generated as for the actual historical data, i.e. no future data are simulated) to approximate the distribution of $P(\delta < T)$ for each value of δ where the distribution is integrated across the OMs;
- (b) for each candidate value of δ , say δ^c , evaluate whether X is less than a chosen percentile of the simulated distribution of $P(\delta < T|\delta^c)$ and reject δ^c if it is. Identify the maximum value of δ (denoted δ_{crit} ; Figure 1b) that can be rejected by this rule. It can be concluded that the current data allow values for δ less than δ_{crit} to be excluded because if $\delta < \delta_{crit}$ that should have already been detected.

4. How long does the experiment need to be continued before a biologically meaningful fishing effect will be detected, if it is present?

Step 3 implies that the current data may not be able exclude values for δ between δ_{crit} and the Threshold. Generate future data and append these to the actual historical data to evaluate the integrated future detection probability, which in essence tries to estimate the likelihood of concluding that $\delta < T$ for a range of assumed values for δ . The integrated detection probability has been calculated by taking 11 equally spaced points for δ over the range $[\delta_{data}^{EM*} - 3(se); T]$ and evaluating the detection probability (i.e. the proportion of simulations for which $P(\delta > T) > P_{min}$) at each of these values δ_i . This detection probability curve is then integrated, with adjustment by the prior distribution for δ_{data}^{EM*} , by taking a weighted average, $P_{int} = \sum_i P_{det}^i p(\delta_i) / \sum_i p(\delta_i)$ where $p(\delta_i)$ is the value at δ_i of the normal distribution with mean δ_{data}^{EM*} and standard deviation se .

Variance of the random year effect

For the 2016 analyses, the estimates for the variance of the random year effect (σ_α) when some of the OMs were applied to the data were at times zero (see Table B.1 of Ross-Gillespie and Butterworth 2016a), and an approach (use of the median from a posterior distribution evaluated using MCMC) was developed to deal with this issue as a non-zero value was needed to generate simulated data. With the further data now available, the estimates of variance for the random effects are now all non-zero. However, in certain instances when data are generated assuming a range of δ values, zero estimates for the year effect variance do arise when the EM is applied to the generated data. To deal with this, a lower limit of 0.1 was applied to all σ_α , as this was considered a reasonable lower limit, and a run with and without the lower limit imposed yielded very similar results.

Results

Table 2 is a summary of key results of the four steps of the algorithm and includes the values of X , P_{min} and δ_{crit} . This Table aims to summarize which response variables do (or alternatively do not) support a biologically meaningful fishing effect at each island at this time, and furthermore how long the closure experiment would need to be continued before a biologically meaningful fishing effect is likely to be detected if such an effect is present. Note that as in 2016, the analyses have been conducted including all response variable data available, i.e. before as well as from the time the closure experiment started.

Table 3 compares estimates of the fishing effect parameter with and without the GLM bias adjustment.

Figure 2 shows comparison plots for the GLM-bias-adjusted estimates along with their 95% confidence intervals for the fishing effect δ for the 2016 and 2020 analyses, corresponding to the information in Table 2. Figure 3 provides a graphical representation of some of the information in Table 1.

Figure 4 plots the integrated detection probabilities against years of future simulated data. Figure 5a and Figure 5b show the same information as Figure 3, but prior to averaging across the OMs. These final sets of Figures are shown purely for background information to illustrate the differences in results amongst the various OMs before a weighted average is taken.

Discussion

Some key discussion points are listed in bullet-point form below.

- Figure 2 shows that the 2020 estimates of fishing effects lie well within the confidence bounds of the 2016 analyses, and in most cases the point estimates for 2016 and 2020 are fairly similar. This is as would be expected, with the biggest impact of adding further data being to reduce the size of the confidence intervals. This does not apply to chick growth there have been no further data for that response variable after 2014, so results there are unchanged. While results are given integrating over the five OMs considered, there is actually relatively little difference amongst the results for each when considered separately (Figure 5a and Figure 5b).
- The results listed in Table 3 show that the GLM bias adjustment makes only minimal difference to estimates of the impact of fishing parameter δ .
- Plots of the standardised residuals (Figure 6) indicate no clear systematic trends (which might indicate model mis-specification), except possibly for the last few years for chick growth at Dassen). However, this is of lesser concern as that response variable has not been collected after 2014, implying that it is to be considered of relatively less importance relative to the other response variables.
- Care must be taken when considering these results in combination, as the different response variables are not entirely independent of each other. For example, results are now shown for all three measures arising from the foraging data in the interests of completeness, though the Panel previously advised to choose and consider only that which was considered the best indicator of food availability (previously considered to be forage length). Furthermore, data on chick survival and fledging success are not independent. A further Panel comment related to the importance of establishing relationships between changes in these response variables and resultant changes in penguin population growth rates.

In summary

- For Dassen Island, all the different data types (apart from chick survival) give little to no indication of a biologically meaningful impact of fishing on the penguin population.
- For Robben island, two of the data sets (chick condition and fledging success) provide evidence for a meaningful impact of fishing on the penguin population. Given an additional 2-5 years' data, a meaningful effect for chick survival should become evident, if it is indeed present. The remainder of the data sets, give little indication of a biologically meaningful impact of fishing on the penguin population (all these currently reflect a point estimate suggesting the reverse of a negative impact caused by fishing).

- For all the variables not currently reflecting a biologically meaningful negative impact of fishing on the penguin population, with the exception of the one mentioned above, at least an additional 20 years' data would be required to demonstrate such an impact, if it is indeed present.

Acknowledgements

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: <http://hpc.uct.ac.za>. The data providers indicated in Table 1 are gratefully acknowledged.

References

- 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.
- Dunn, A., Haddon, M., Parma, A. M. and Punt, A. E. 2016. International Review Panel Report for the 2016 International Fisheries Stock Assessment Workshop, 18 Nov - 2 December 2016, UCT. 17pp.
- OLSPS. 2020. A simple restructuring of penguin chick survival data to estimate survivorship. Fisheries branch document FISHERIES/2020/JAN/SWG-PEL/06.
- Penguin Island Closure Task Team. 2015. Specifications for Operating Models to evaluate bias in estimation methods in accordance with recommendation A.1 of the 2014 International Review Panel. Document MARAM/IWS/DEC15/PengD/P1 submitted to the 2015 International Stock Assessment Review Workshop.
- Ross-Gillespie, A. and Butterworth, D.S. 2016a. 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 Review Workshop.
- Ross-Gillespie, A. and Butterworth, D.S. 2016b. Implementation of the Algorithm recommended by the Panel for the 2016 International Stock Assessment Workshop for assessing whether or not to continue with the penguin island closure experiment. Fisheries Branch document FISHERIES/2016/DEC/SWG-PEL/77rev.

Table 1: Summary of the data used in the 2016 analyses and the 2020 analyses. The 2020 analyses incorporate up to five years additional data in some instances and correspond to the data provided in Coetzee and Merkle (2019). This is apart from active nest proportion data which were not considered in 2016 either, and the chick survival estimates arising from the analyses of OLSPS (2020) (namely the estimates for survivorship to 74 days). A dash indicates that the data type was not included in the 2016 analyses. Note that no new chick growth data have become available since 2014.

Data type	2016 analyses	2020 analyses	Data providers
Chick growth	1989-2014	1989-2014	DEA/Richard Sherley
Chick condition	2008-2013	2008-2018	DEA/CapeNature, Richard Sherley
Fledging success	1989-2013	1989-2015	DEA/Richard Sherley
Foraging trip length	2003-2013	2003-2018	DEA/UCT/Cape Nature, Richard Sherley/Lorien Pichegru
Foraging trip duration	-	2003-2018	DEA/UCT/Cape Nature, Richard Sherley/Lorien Pichegru
Max. foraging trip length	-	2008-2018	Lorien Pichegru
Chick survival ⁵	-	2001-2015	Richard Sherley (raw data), OLSPS (processed data)

⁵ Similar to the treatment of the fledging success, chick growth and chick condition datasets, chick survival is transformed prior to analyses by taking natural logarithms, so that the estimates of fishing impacts from different data types are roughly comparable in relative terms. The foraging data are transformed taking the negative of the logarithm, so that similar to the other data types a decrease in value of the transformed data corresponds to a negative impact on the penguin population.

Table 2: Summary of the power analysis results. Section (a) repeats the 2016 results from Ross-Gillespie and Butterworth (2016b) for reference purposes, while section (b) lists the results for the 2020 update and furthermore includes results for additional response variables. Results are shown for the closure-only EMs, integrated over the five OMs. The closure-only OMs receive twice the weighting of the others to take into account that the catch-only and catch+closure OMs are implemented for two catch-biomass correlation values. The first column of numbers shows the GLM-bias-corrected estimates of δ and the second column the EM estimates of the standard error. The third column shows the area (corresponding to probability) under the normal curve with mean δ_{data}^{EM*} and standard deviation se that lies to the left of the Threshold (i.e. “X” from the main text). The fourth column lists the P_{min} values. If $X > P_{min}$ then there is evidence in the data of a fishing effect that is biologically meaningful with respect to penguin demographics. In such cases the table entry has been highlighted in grey. The next five columns show the values below which the true δ is not likely to be, for a range of (one-tailed) risk levels from which one might be adopted for decision purposes. Cells that have been highlighted in yellow indicate that the value below which δ is not likely to lie, is above the Threshold. The last column shows the number of years that the experiment would need to be continued for before it is likely to be possible (with 80% probability) to conclude from the data that δ is less than the Threshold, IF the true δ is indeed below the Threshold. Note that this is the number of years of additional data required, which given the different years the data span doesn't always correspond to the same end-point (see Figure 4). Note further that a computation error was found for the code used in 2016 to calculate the number of years the experiment needs to continue (resulting in an overestimation of the detection probabilities). Consequently, those erroneous values have not been included in section (a), and the blue and orange highlighting (which depends on these “Years” estimates) have been removed.

(a)		EM applied to data				δ_{crit} for a range of risk levels					Years
Island	Data type	δ_{data}^{EM*}	se	X	P_{min}	2.50%	5%	10%	20%	50%	
Dassen	Chick growth	0.38	0.14	0.00	0.55	0.07	>0.1	>0.1	>0.1	>0.1	
	Chick condition	-0.08	0.22	0.46	0.56	-0.44	-0.37	-0.28	-0.24	-0.07	
	Fledging success	0.10	0.21	0.18	0.52	-0.37	-0.27	-0.19	-0.07	>0.1	
	Forage length	0.19	0.29	0.16	0.56	-0.28	-0.22	-0.14	-0.06	>0.1	
Robben	Chick growth	0.03	0.16	0.20	0.53	-0.29	-0.23	-0.16	-0.09	>0.1	
	Chick condition	-0.13	0.20	0.55	0.57	-0.54	-0.42	-0.32	-0.30	-0.12	
	Fledging success	-0.25	0.23	0.75	0.52	-0.93	-0.76	-0.61	-0.43	-0.22	
	Forage length	0.16	0.28	0.18	0.53	-0.45	-0.32	-0.19	-0.07	>0.1	
(b)		EM applied to data				δ_{crit} for a range of risk levels					Years
Island	Data type	δ_{data}^{EM*}	se	X	P_{min}	2.50%	5%	10%	20%	50%	
Dassen	Chick growth	0.39	0.14	0.00	0.52	0.09	>0.10	>0.10	>0.10	>0.10	>20
	Chick condition	-0.03	0.14	0.32	0.58	-0.24	-0.20	-0.16	-0.14	0.01	>20
	Fledging success	0.10	0.16	0.10	0.55	-0.30	-0.23	-0.18	-0.10	0.05	>20
	Forage length	0.35	0.18	0.01	0.56	-0.01	0.04	>0.10	>0.10	>0.10	>20
	Forage duration	0.54	0.15	0.00	0.57	>0.10	>0.10	>0.10	>0.10	>0.10	>20
	Max length	0.04	0.18	0.21	0.58	-0.22	-0.18	-0.13	-0.09	0.07	>20
	Survival	-0.13	0.10	0.61	0.56	-0.63	-0.53	-0.39	-0.26	-0.12	>20 ⁶
Robben	Chick growth	0.03	0.16	0.20	0.55	-0.29	-0.23	-0.16	-0.08	0.08	>20
	Chick condition	-0.14	0.13	0.62	0.61	-0.37	-0.31	-0.26	-0.27	-0.13	10
	Fledging success	-0.20	0.17	0.71	0.53	<-1.00	-0.97	-0.70	-0.37	-0.24	0
	Forage length	0.03	0.18	0.24	0.56	-0.34	-0.24	-0.17	-0.10	0.05	>20
	Forage duration	0.10	0.16	0.10	0.57	-0.16	-0.10	-0.06	0.00	>0.10	>20
	Max length	0.03	0.19	0.24	0.59	-0.28	-0.22	-0.16	-0.10	0.05	>20
	Survival	-0.04	0.11	0.29	0.57	-0.32	-0.25	-0.20	-0.13	0.02	2-5

- green There is no evidence in the current data to support a biologically meaningful fishing effect, as the lower bound of the normal distribution, $\delta_{data}^{EM*} - 2(se)$, lies above the threshold.
- blue The experiment needs to continue for more than 10 years before a biologically meaningful fishing effect is likely to be detected, if it is present.
- orange The experiment needs to continue for 2 to 5 years before a biologically meaningful fishing effect is likely to be detected, if it is present.
- red There is evidence in the current data of a biologically meaningful fishing effect because $X > P_{min}$.

⁶ It might seem counter-intuitive that here, and for Robben Island chick condition, the estimates of how long the experiment needs to continue are non-zero, even though the conclusion is drawn that there is evidence to support a biologically meaningful impact of fishing on the penguin population. The values reported here (in the “Years” column) estimate how long the experiment would need to continue for before it becomes likely (with 80% probability) that a fishing effect would be detected if it is there. However, a detection probability of 40-50%, as is the case here, does not preclude the possibility of already having detected a meaningful effect now, rather a higher likelihood of having missed detecting it. It should also be noted that for both of these cases, the X variable is only marginally bigger than P_{min} , suggesting that the evidence for a biologically meaningful impact is less strong than for Robben fledging success.

Table 3: Table listing the estimates of the fishing effect estimated when the closure-only EM is applied to the actual historical data. Both the GLM-bias-adjusted estimates (δ_{data}^{EM*}) and the original unadjusted values (δ_{data}^{EM}) are listed. The se value applies to the original estimate δ_{data}^{EM} .

Island	Data type	EM applied to data		
		δ_{data}^{EM*}	δ_{data}^{EM}	se
Dassen	Chick growth	0.39	0.38	0.14
	Chick condition	-0.03	-0.04	0.14
	Fledging success	0.10	0.10	0.16
	Forage length	0.35	0.36	0.18
	Forage duration	0.54	0.53	0.15
	Max length	0.04	0.05	0.18
	Survival	-0.13	-0.13	0.10
Robben	Chick growth	0.03	0.03	0.16
	Chick condition	-0.14	-0.14	0.13
	Fledging success	-0.20	-0.20	0.17
	Forage length	0.03	0.03	0.18
	Forage duration	0.10	0.11	0.16
	Max length	0.03	0.04	0.19
	Survival	-0.04	-0.03	0.11

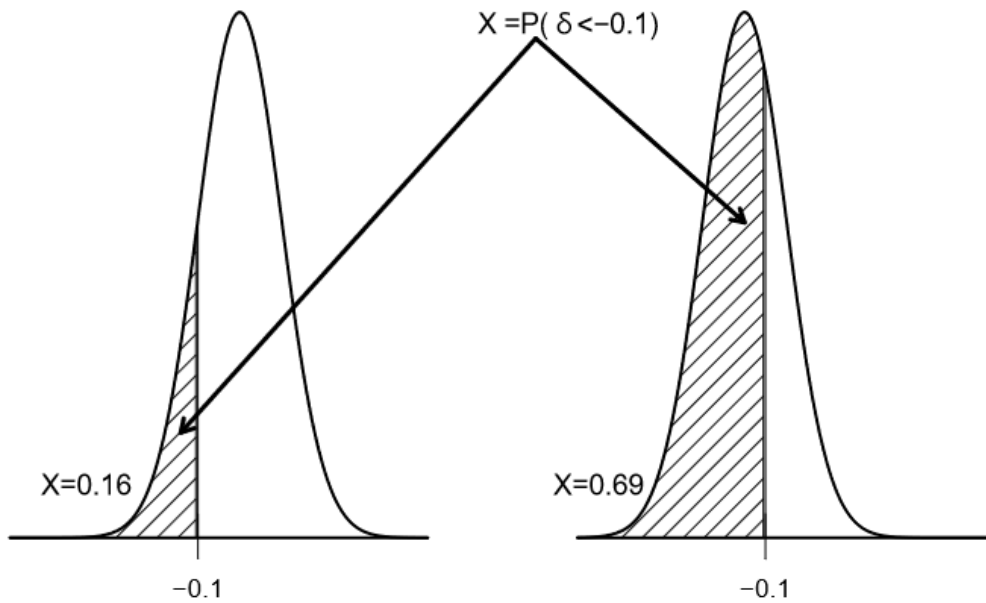


Figure 1a: Examples of two possible δ distributions estimated when applying the EM to the actual historical data and the calculation of X (the area of the shaded portion under the curve - this is the probability of a value less than -0.1 as the area under the whole curve is 1) for each (Step 1 of the Algorithm).

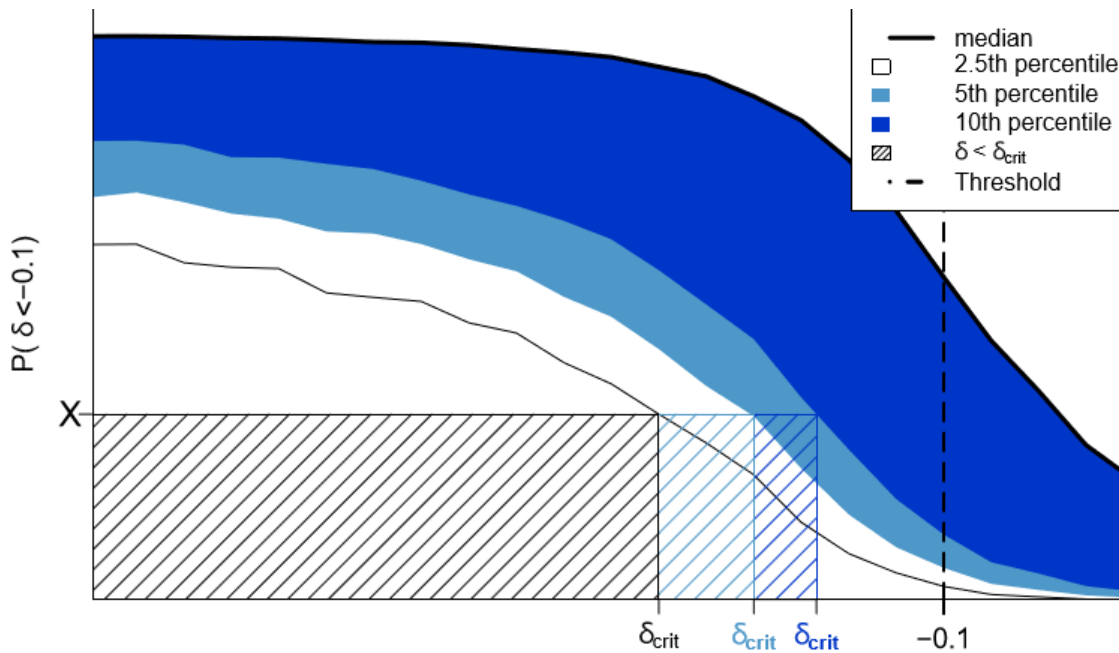


Figure 1b: An example of a detection probability curve to determine which values of δ can be excluded, depending on the risk level adopted (Step 3 of the Algorithm). The detection probability shows, for any given δ_i on the horizontal axis, what the probability is of concluding from the data that $\delta < T$ if the true fishing effect $\delta = \delta_i$. In other words, the asymptote towards 1 on the left of the plot shows that the more negative the true δ is, the more likely that a conclusion will be reached that $\delta < T$. For the value of X shown, the three values of δ_{crit} shown correspond (from left to right) to risk levels of 2.5%, 5% and 10%.

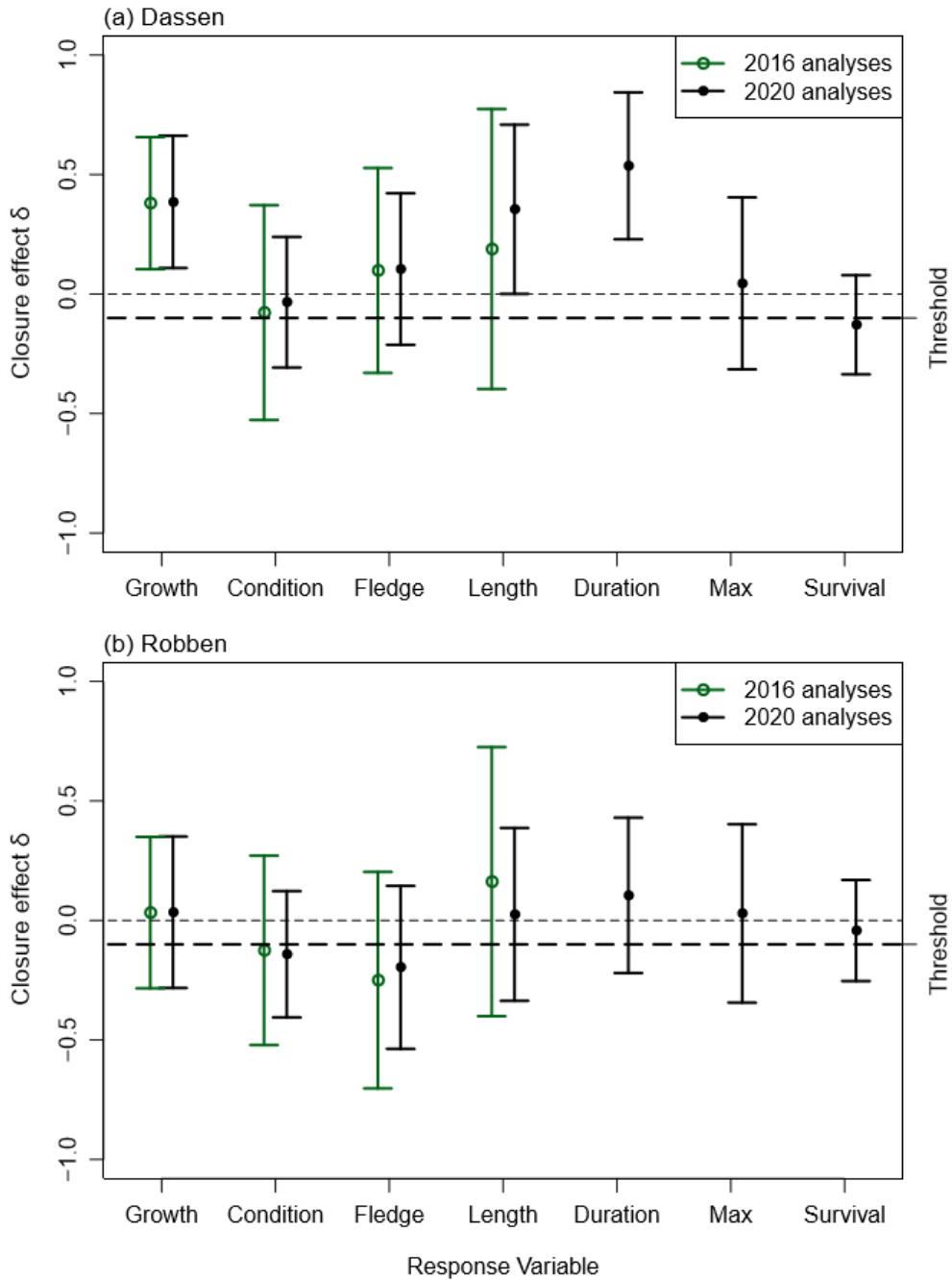


Figure 2: GLM-bias-adjusted estimates of the closure effect δ when the closure-only EM is applied to the various response variable datasets are shown along with the 95% confidence intervals (corresponding to twice the standard error estimates). Results are shown in green for the 2016 analyses and in black for the 2020 analyses, where an additional three data sets have been included in the analyses. The thin horizontal dashed line marks zero, and the thick horizontal dashed line marks the threshold of -0.1. Note that no new chick growth data have become available since 2014, so that the 2016 and 2020 results are based on the same data and therefore identical.

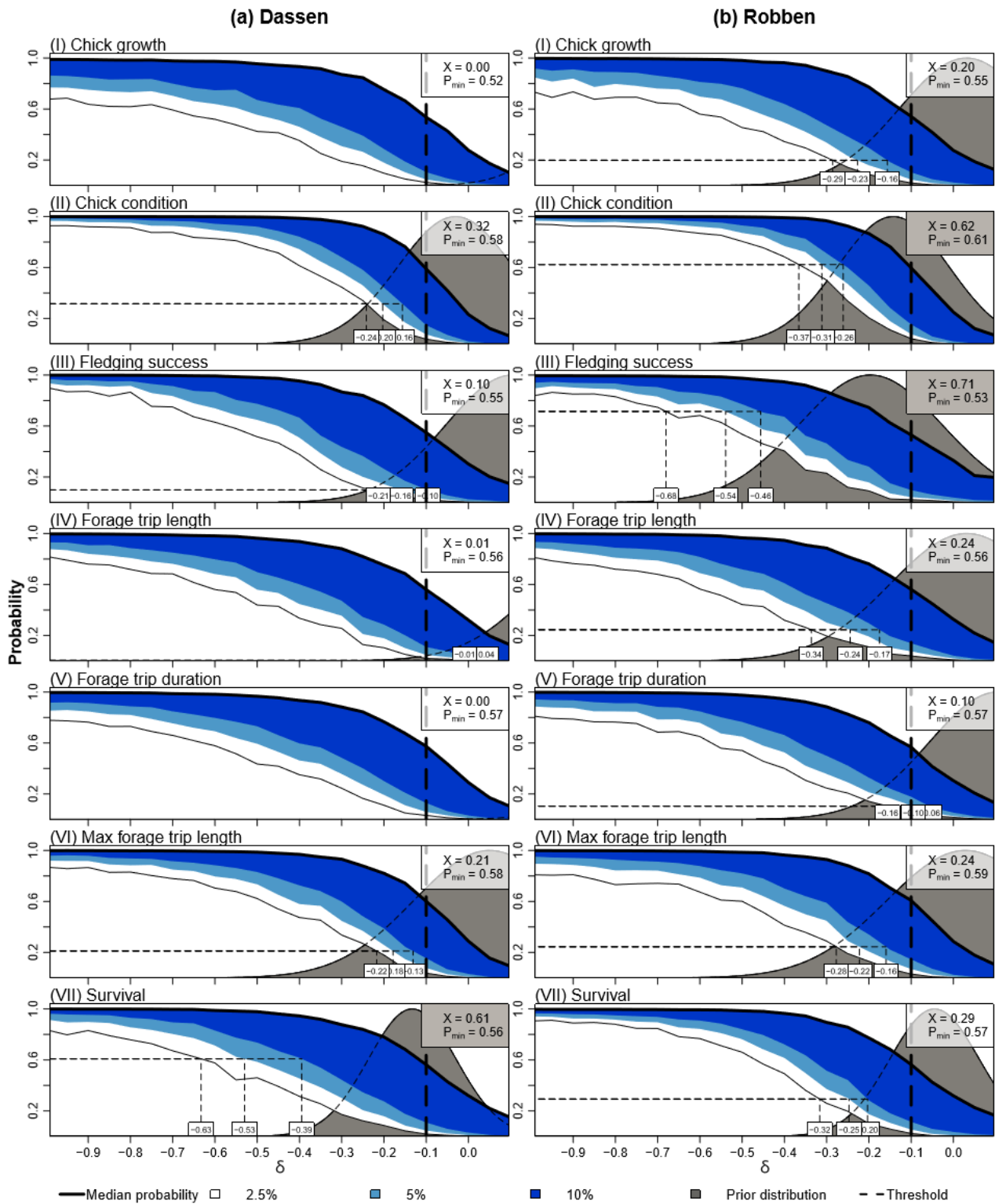


Figure 3: Graphical representation of the information presented in Table 1. The grey shaded area shows the GLM-bias-adjusted normal distribution estimated for δ by the EM. The thick solid black line shows the median detection probability for the 1000 simulations conducted for each value of δ over the range $[-1, 0.1]$. The shaded regions show the 2.5th percentiles (white), 5th percentiles (light blue) and 10th percentiles (dark blue) of the 1000 simulations over this δ range. The legend in the top right corner shows X , the area (corresponding to a probability) under the grey-shaded normal curve to the left of the Threshold, as well as the value of the OM-weighted P_{min} . For each plot, a horizontal line has been drawn at $X = P(\delta < T)$, and the values at which this line intersects the three percentile curves are displayed. These are the values below which the true δ is not likely to be, depending on the risk level adopted (i.e. which of these three percentiles, or conceivably some other, might be chosen for decision purposes).

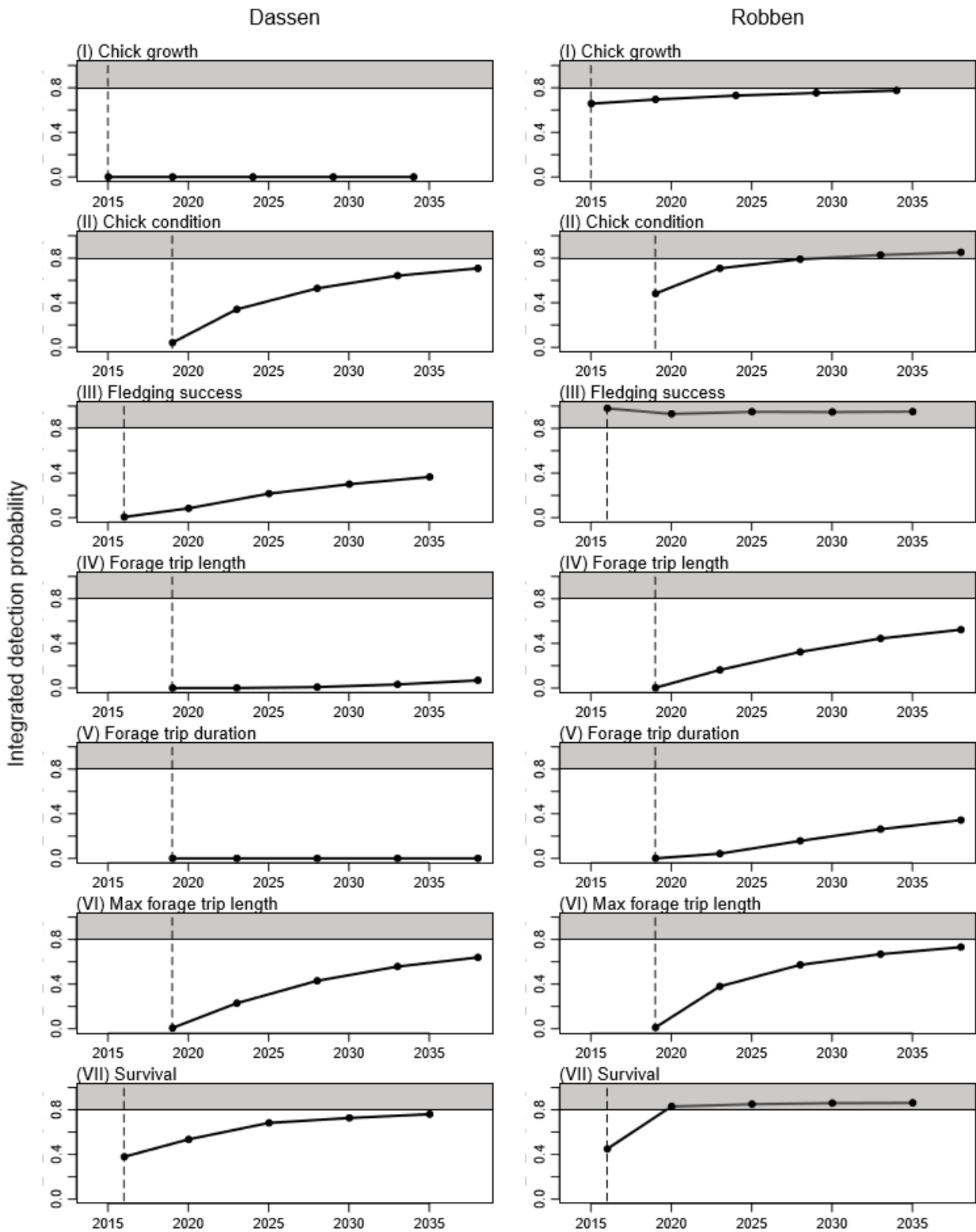


Figure 4: Plot of the integrated detection probabilities. The probabilities have been weighted across OMs to produce a single detection probability curve for each EM, and show the length of time the closure experiment would need to be continued for before a biologically meaningful fishing effect is likely to be detected from the data, if such an effect is present.

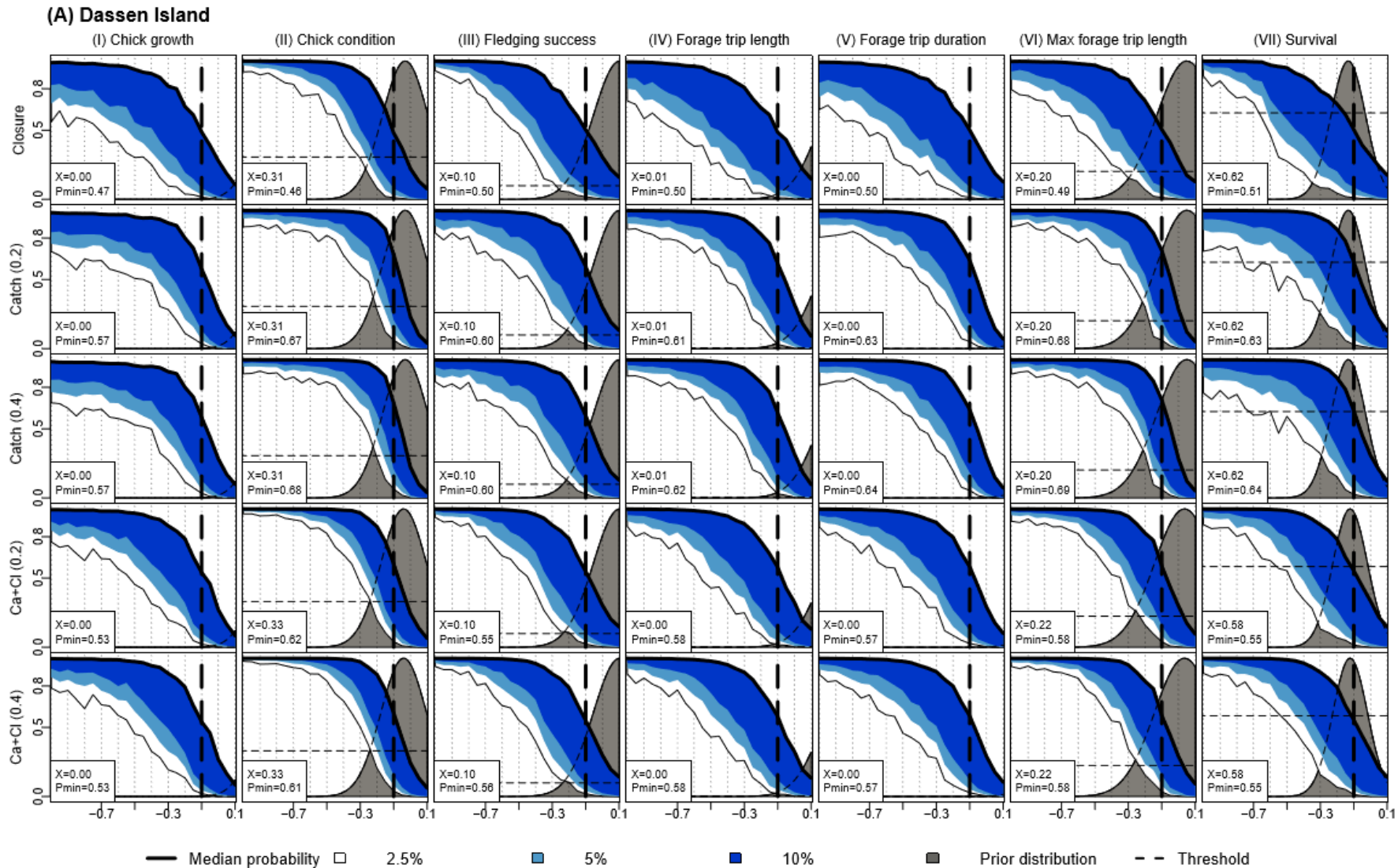


Figure 5a: The detection probabilities and percentile curves for **Dassen Island** prior to integrating across OMs. The Dassen Island plot for each data type in Figure 3 is thus a weighted combination across the five plots for that data type shown here. Note that the curves are not as smooth as those shown in Ross-Gillespie and Butterworth (2016b) because those results corresponded to 5000 simulations compared to the 1000 conducted here. Note also that variation in the X parameter across the range of OMs is as a result of simulation variation, as the GLM bias adjustment factor (and therefore the centre of the normal distributions) is based on simulations.

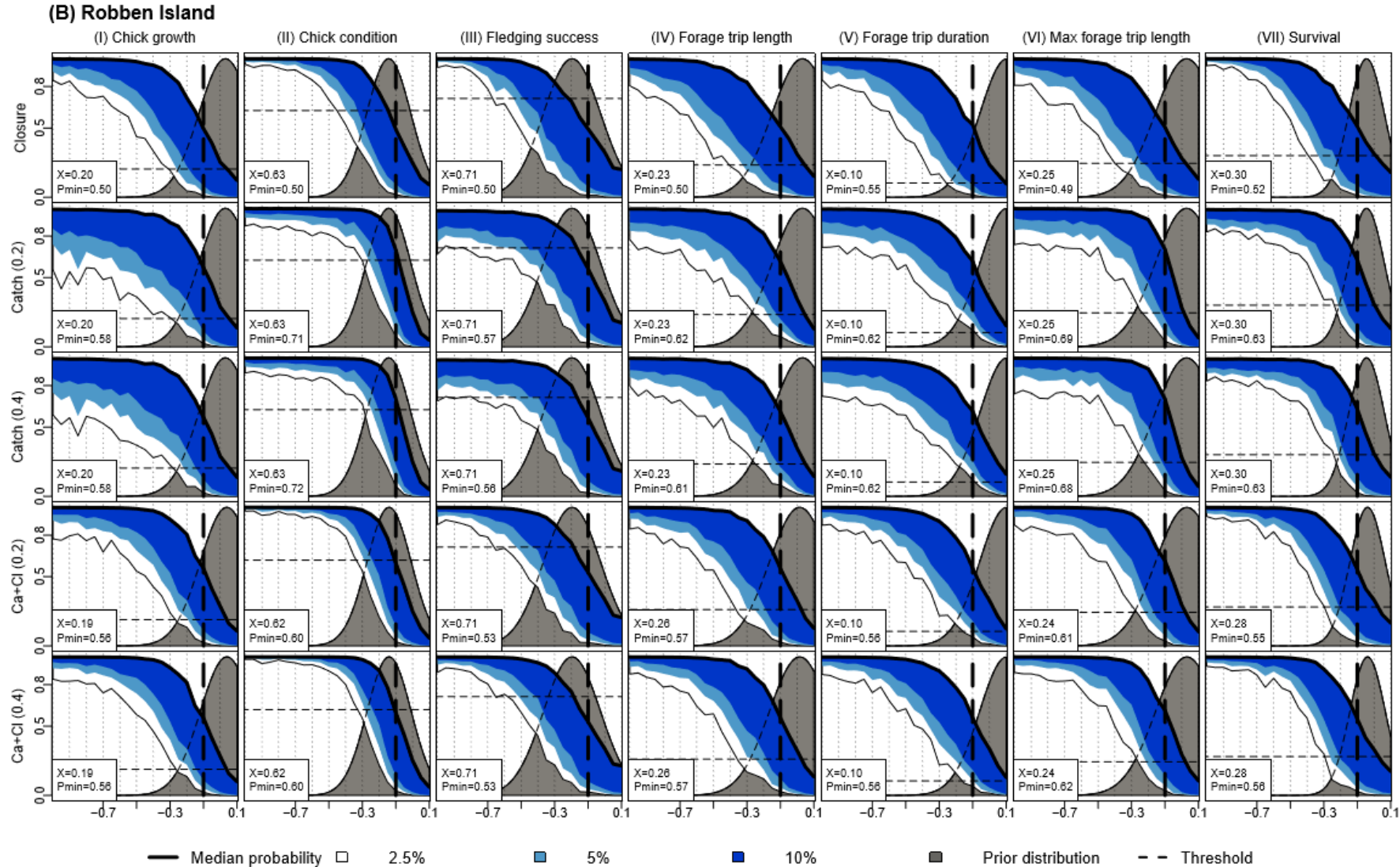


Figure 5b: The detection probabilities and percentile curves for Robben Island prior to integrating across OMs.

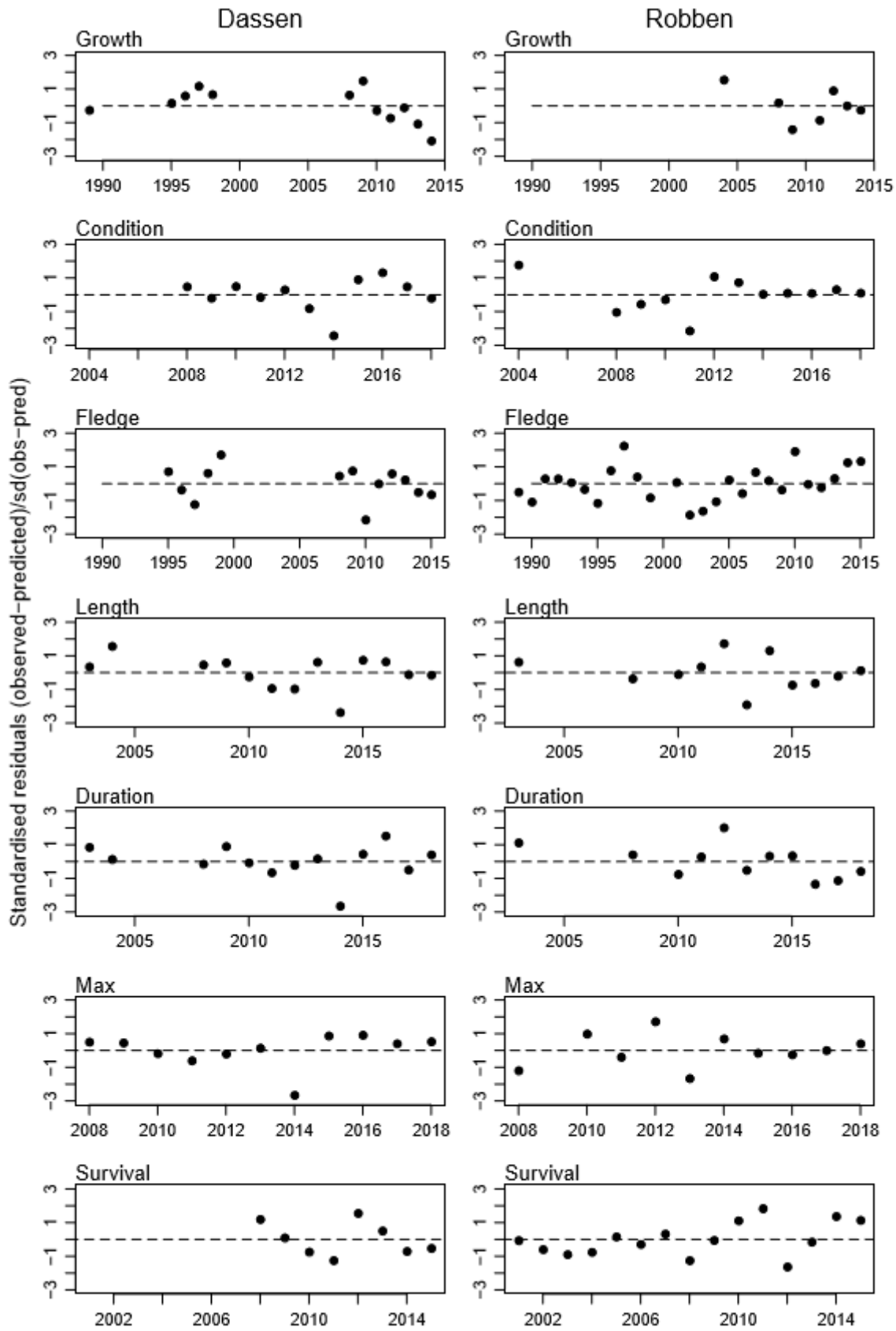


Figure 6: Plots of standardised residuals when the closure only EM is applied to the actual historical data. Note that the vertical axes are the same throughout, but the horizontal (year) axes vary for the different data types.

Appendix A A glossary of terms

OM	<p>The Operating Model.</p> <p>A model of the actual underlying (but unknown) situation which is used to generate simulated penguin response data (i.e. to generate possible values for chick growth, chick condition, foraging trip data, fledging success and chick survival observations). Three different OMs have been implemented, each of which assumes that anchovy abundance is a major driver of how the penguin response variable behaves, and in addition that: (1) anchovy catch is a major driver in the catch-only OM, (2) island closure is a major driver in the closure-only OM, and (3) both catch and closure are major drivers in a catch+closure OM. The power analysis results are integrated across the OMs since it is not known which better reflects the true underlying situation that is giving rise to the data observed. The details of the data generation process can be found in MARAM/IWS/DEC15/PengD/P1.</p>
EM	<p>The Estimation Model.</p> <p>A model that is applied to the historical or simulated data to estimate the impact of fishery on a penguin response variable and hence the penguin population. The results reported in this document use the closure-only EM alone as per the recommendation of the 2016 IWS Panel.</p>
δ	<p>The closure effect, <i>viz.</i> the fishing effect.</p> <p>The δ parameter estimates the extent to which the island being open to fishing negatively impacts the penguin population, as indicated for the response variable under analysis (see Equation (1) of MARAM/IWS/DEC15/PengD/P1). If fishing is having a negative effect, the value of δ will be negative.</p>
T	<p>The Threshold.</p> <p>The value of the Threshold is a biological value determined externally to the power analysis. It relates a decision on what is considered to be a biologically meaningful change in the penguin population growth rate, with this change then being linked to a multiplicative change in an observable penguin population response variable (such as fledging success, chick growth etc.). Since the equations for the power analysis have been formulated in a manner that a decrease in the response variable observed equates to a negative impact on the penguin population, the Threshold also has a negative value. A change of 1% in the penguin population growth rate, corresponding closely to a Threshold (evaluated in logspace for the response variable) of -0.1 in δ space for the fledging success response variable, has been used for the base assumption for these analyses. This same relationship between the penguin population growth rate change and the Threshold was assumed for the other response variables.</p>
GLM bias	<p>In the analyses undertaken for IWS 2015, it was established that when data are simulated using the GLM point estimates estimated from the actual data, and when the same GLM was applied to those simulated data, the estimated δ values differed from the true original point estimate for δ used to simulate the data. This bias is termed the “GLM bias”. It arises primarily from a lack of balanced design in the historical data available. See immediately below for how this bias is calculated.</p>

δ_{data}^{EM*}	The GLM-bias-adjusted estimate of δ when the EM is applied to the actual historical data. The bias is evaluated by applying the EM to the actual historical data to estimate the fishing effect δ , as well as the other parameters of the model. These parameters are then used by an EM-like-OM to generate multiple sets of simulated historical data to which the EM is again applied. The difference between the δ used to generate the data, and the mean of the new estimates obtained from the simulated data determines the GLM bias.
se	The estimate of the standard error of δ when the EM is applied to the actual historical data.
P_{min}	In order to evaluate P_{min} , the historical data are replaced with simulated data, which are generated assuming that the true δ is equal to the Threshold. P_{min} is the value for which the proportion of the total simulations for which $P(\delta < T) > P_{min}$ is 0.5 (i.e. half the total simulations have an area under the normal curve to the left of the Threshold that is greater than P_{min}). For a balanced design, and an unbiased estimator, P_{min} should be close to 0.5.
P_{min} bias	This bias is defined as the extent to which P_{min} (calculated as described above) differs from 0.5 when δ is equal to the Threshold. This P_{min} bias is taken into account when calculating the integrated detection probabilities in Figure 4.
X	The (cumulative) probability that δ is less than the Threshold (i.e. $P(\delta < T)$) is based on the actual data, using Equation 2 of Ross-Gillespie and Butterworth (2016a), which is reproduced below. $P_i(\delta_i < T) = \int_{-\infty}^T \frac{1}{\sqrt{2\pi\sigma_{\delta_i}}} e^{-(x-\hat{\delta}_i)^2 / (2\sigma_{\delta_i})^2} dx$ <p>In other words, X is the area of the normal curve for δ estimated from the actual data (i.e. a normal curve with mean δ_{data}^{EM*} and standard deviation se) to the left of the Threshold (taken here to be -0.1).</p>
detection prob.	The probability of concluding, when applying the EM to past historical and future simulated data under an assumed value of δ , that $\delta < T$. $P(\delta < T \delta^c)$ is the detection probability at δ^c , i.e. the probability of concluding that $\delta < T$ if the true δ is equal to δ^c .
integrated detection prob.	The probability of concluding, when applying the EM to past historical and future simulated data under a range of assumptions for the true value of δ , that $\delta < T$ when the true δ is in fact below the Threshold.
δ_c	Any given value of δ on the interval over which the detection probability curve is evaluated.
δ_{crit}	In the Panel's algorithm, if one cannot conclude that $\delta < T$, then the next step is to identify what values of δ^c can be excluded. δ_{crit} is the maximum value of δ that can be rejected, i.e. it can be concluded that the current data allow values for δ less than δ_{crit} to be excluded because if $\delta < \delta_{crit}$ that should already have been detected. This conclusion is also dependent on a pre-specified risk level. The risk level is the probability that δ^c could be less than the maximum value (δ_{crit}) calculated.

Appendix B The mixed effects model

The mixed effects model assumed for the OMs and EM is given by the equation below (Ross-Gillespie and Butterworth 2016a):

$$\ln(F_{y,i,s}) = K + \alpha_y + \gamma_s + \lambda_i \frac{C_{y,i}}{\bar{C}_i} + \delta_i X_{y,i} + \epsilon_{y,i,s} \quad (1)$$

where

- $F_{y,i,s}$ is the penguin response variable for series s for island i in year y ,
- 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),
- γ_s is an island effect (simply equal to island in most cases, but for fledging success there are two series for Robben Island,
- 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 anchovy catch taken in year y in the 18.148km radius neighbourhood of island i ,
- \bar{C}_i is the average catch taken over the years for which island i was open to the fishery between 2008-2013⁷,
- δ_i is a fishing effect occasioned by whether or not the neighbourhood around the island was closed to fishing,
- $X_{y,i}$ is a closure variable that is set to 1 in years where island i is open to the fishery, and 0 where the 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. Thus, for the closure-only EM application in this document $\lambda = 0$.

⁷ 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.