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. Butterworth[1]

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

The results of the penguin power analyses based on the data from the island closure experiment were presented to the Panel for the 2016 International Stock Assessment Workshop (IWS 2016). The Panel developed an algorithm for interpreting these results in the context of providing advice as to whether or not the closure experiment should be continued. The four steps of this algorithm (see recommendation C.1.2 of Dunn *et al.* 2016) are reproduced below, with some minor edits to enhance clarity.

1. **Evaluate the impact of fishing by applying the Estimation model (EM) to the actual historical data.**
   Compute the (cumulative) probability that δ is less than the Threshold (currently -0.1), using Equation 2 of MARAM/IWS/DEC16/Peng Clos/P1a (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). (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}[2] 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:

   (a) use the results of "unconditioned" simulations in which the historical data are generated given a set of values specified for δ (with data points generated as for the actual historical data, i.e. no

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

[2]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 proportion of the total simulations for which $P(\delta < T)$ (i.e. the area under the normal curve to the left of the Threshold) is greater than 0.5. For a balanced design, and an unbiased estimator, P_{min} should be close to 0.5.

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. The results of the conditioned simulations (i.e. the data include the historical data and simulated future data; Table 4 of MARAM/IWS/DEC16/Peng_Clos/P1a) should be considered to decide whether or not to continue the experiment.

Appendix A lists a glossary for the terms used in this document.

Results

Table 1 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 data sets do (or alternatively do not) support a biologically meaningful fishing effect, and how long the closure experiment would likely need to be continued before a biologically meaningful fishing effect is likely to be detected if such an effect is present.

Figure 2 shows a graphical representation of some of the information in Table 1.

Figure 3 plots the integrated detection probabilities against years of future simulated data. This Figure is a repeat of Figure 3 of MARAM/IWS/DEC16/Peng Clos/P1a, but showing only the closure EM results. Figures 4a and 4b show the same information as Figure 2, except prior to integrating 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.

References

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.

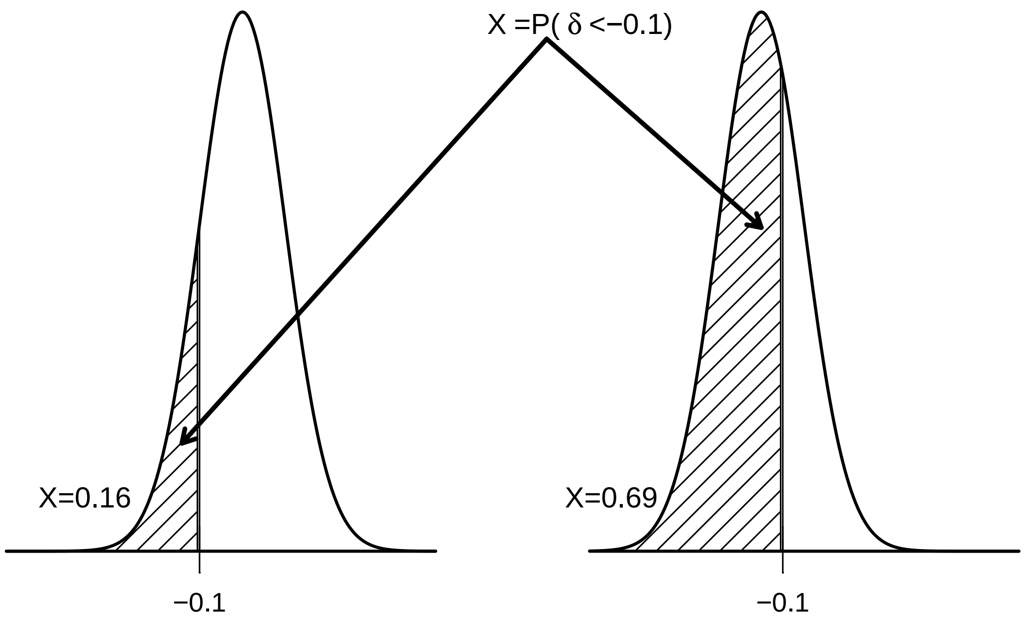

Figure 1a: Examples of two possible δ distributions estimated when applying the EM to the actual historical data and the calculation of X (the area of the shaded portion under the curve - this is the probability of a value less than -0.1 as the area under the whole curve is 1) for each (Step 1 of the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

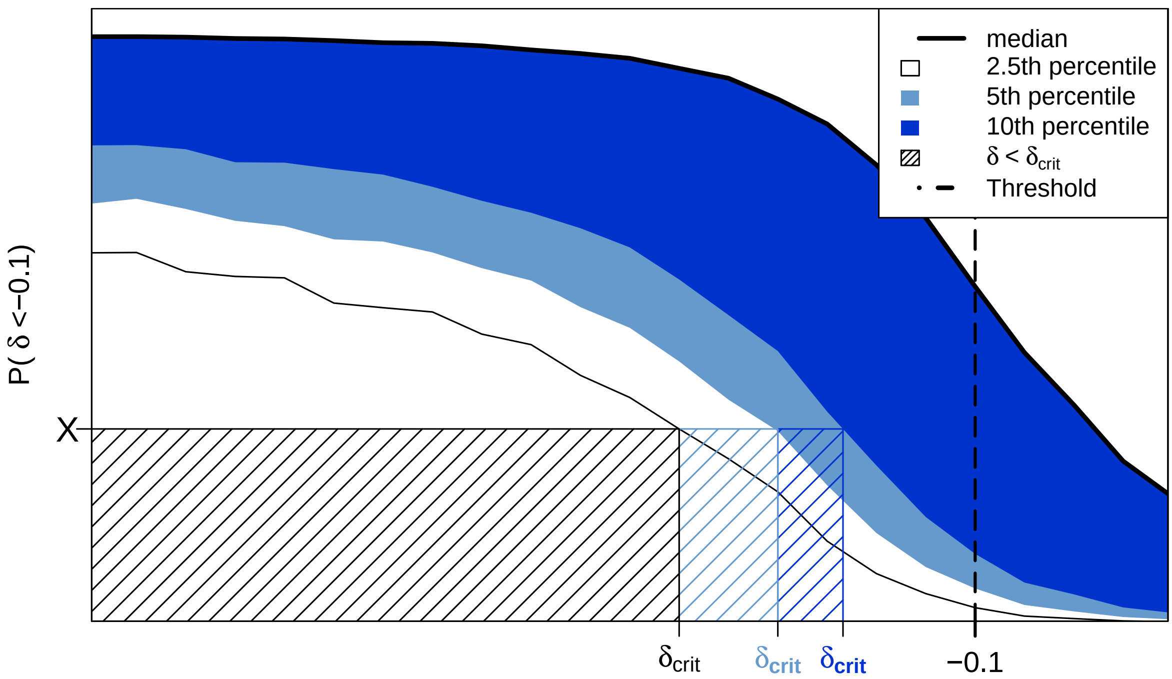

Figure 1b: An example of a detection probility curve to determine which values of δ can be excluded, depending on the risk level adopted (Step 3 of the Algorithm).

Table 1: Summary of the power analysis results for the four data sets and two islands. Results are shown for the closure-only EMs, integrated over the five OMs (closure-only OM, catch-only OM with $\phi = 0.2$ and $\phi = 0.4$ and catch+closure OM with $\phi = 0.2$ and $\phi = 0.4$). 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. Note that these P_{min} values are calculated for when the historical data are replaced with simulated data, i.e. no future data are simulated as for the P_{min} values reported in Table 4 of MARAM/IWS/DEC16/Peng Clos/P1a. If $P(\delta < T) > P_{min}$ 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. Note that the 20% and 50% risk levels shown in the Table are not shown in Figure 2. 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 to conclude from the data (with 80% probability) that δ is less than the Threshold, IF the true δ is indeed below the Threshold. A range shown, such as 2-5, means that at least two but less than five additional years are required.

Island	Data type	EM applied to data				δ_{crit} for a range of risk levels					
		δ_{data}^{EM*}	se	X	P_{min}	2.5%	5%	10%	20%	50%	Years
Dassen	Chick growth	0.38	0.14	0.00	0.55	0.07	> 0.1	> 0.1	> 0.1	> 0.1	> 20
	Chick condition	-0.08	0.22	0.46	0.56	-0.44	-0.37	-0.28	-0.24	-0.07	2-5
	Forage length	0.19	0.29	0.16	0.56	-0.28	-0.22	-0.14	-0.06	> 0.1	11-15
	Fledging success	0.10	0.21	0.18	0.52	-0.37	-0.27	-0.19	-0.07	> 0.1	11-15
Robben	Chick growth	0.03	0.16	0.20	0.53	-0.29	-0.23	-0.16	-0.09	> 0.1	11-15
	Chick condition	-0.13	0.20	0.55	0.57	-0.54	-0.42	-0.32	-0.30	-0.12	2-5
	Forage length	0.16	0.28	0.18	0.53	-0.45	-0.32	-0.19	-0.07	> 0.1	11-15
	Fledging success	-0.25	0.23	0.75	0.52	-0.93	-0.76	-0.61	-0.43	-0.22	0

green There is no evidence in the current data to support a biologically meaningful fishing effect.

blue The experiment needs to continue for 11 to 15 years before a biologically meaningful fishing effect is likely to be detected, if it is present. Note that for this Table, the blue categories would become green if a risk level of some 20% or more were to be adopted.

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.

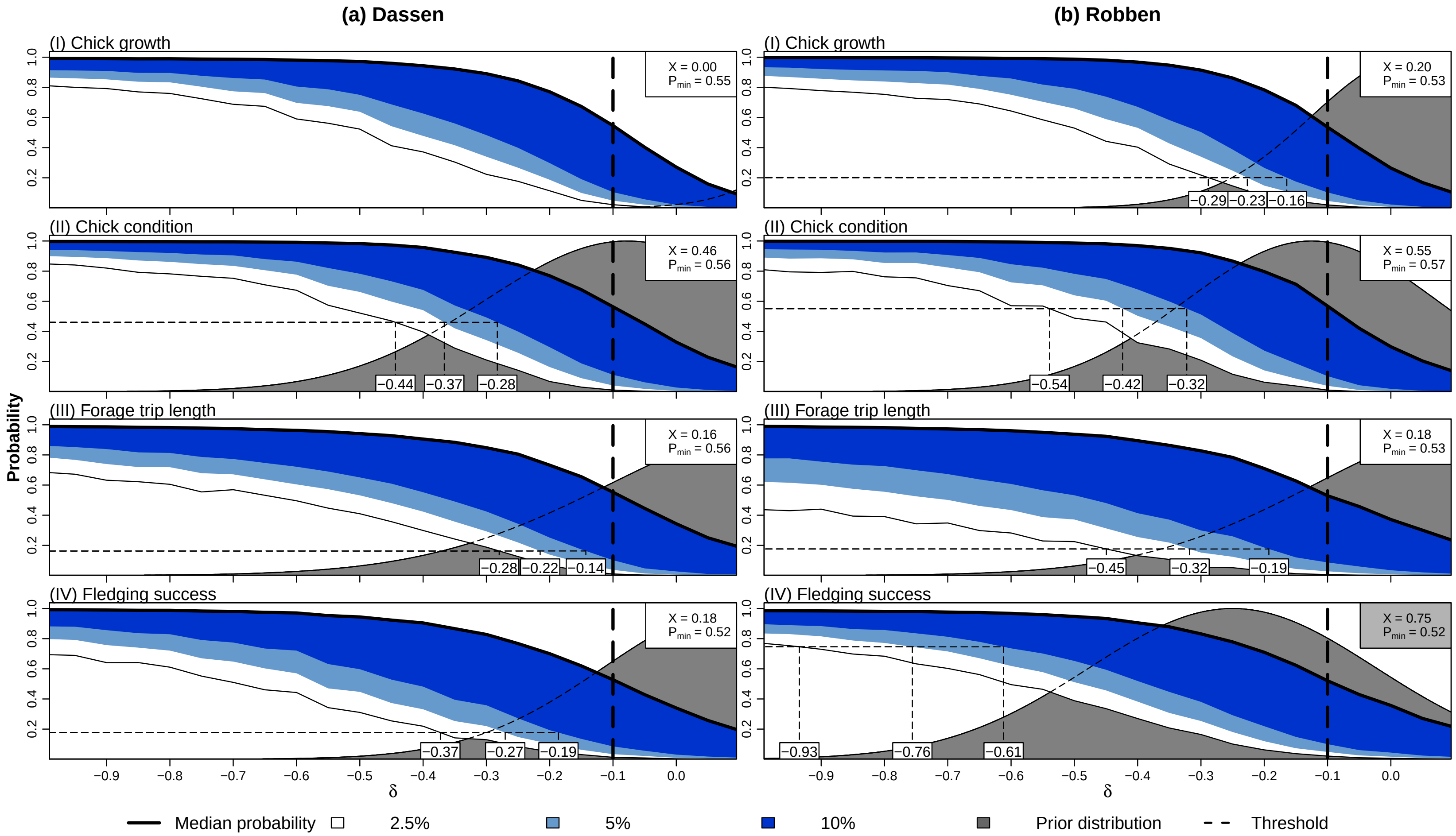

Figure 2: Graphical representation of the information presented in Table 1. The grey shaded area shows the GLM-bias-adjusted normal distribution estimated for δ by the EM. The thick solid black line shows the median detection probability for the 5000 simulations conducted for each value of δ on the range [-1, 0.1]. The shaded regions show the 2.5th percentile (white), 5th percentile (light blue) and 10th percentile (dark blue) of the 5000 simulations. 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 P_{min} from Table 4 of MARAM/IWS/DEC16/Peng Clos/P1a. For each plot, a horizontal line has been drawn at $P(\delta < T)$, 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.

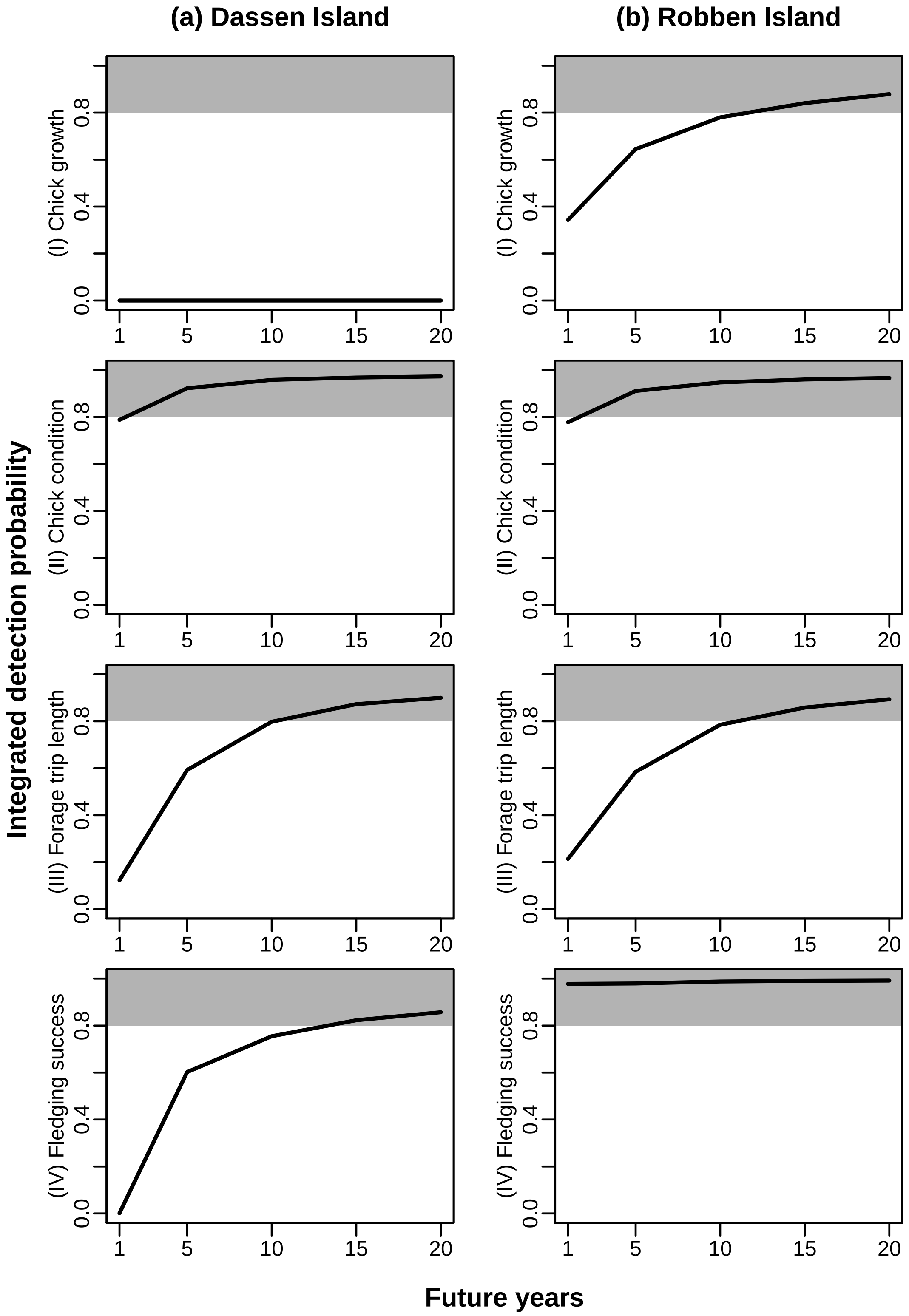

Figure 3: Plot of the integrated detection probabilities from Table 8 of MARAM/IWS/DEC16/Peng Clos/P1a, but for the closure EM only. The probabilities have been weighted across OMs to produce a single detection probability curve for each EM, and show the length of time the closure experiment would need to be continued for before a biologically meaningful fishing effect is likely to be detected from the data, if such an effect is present.

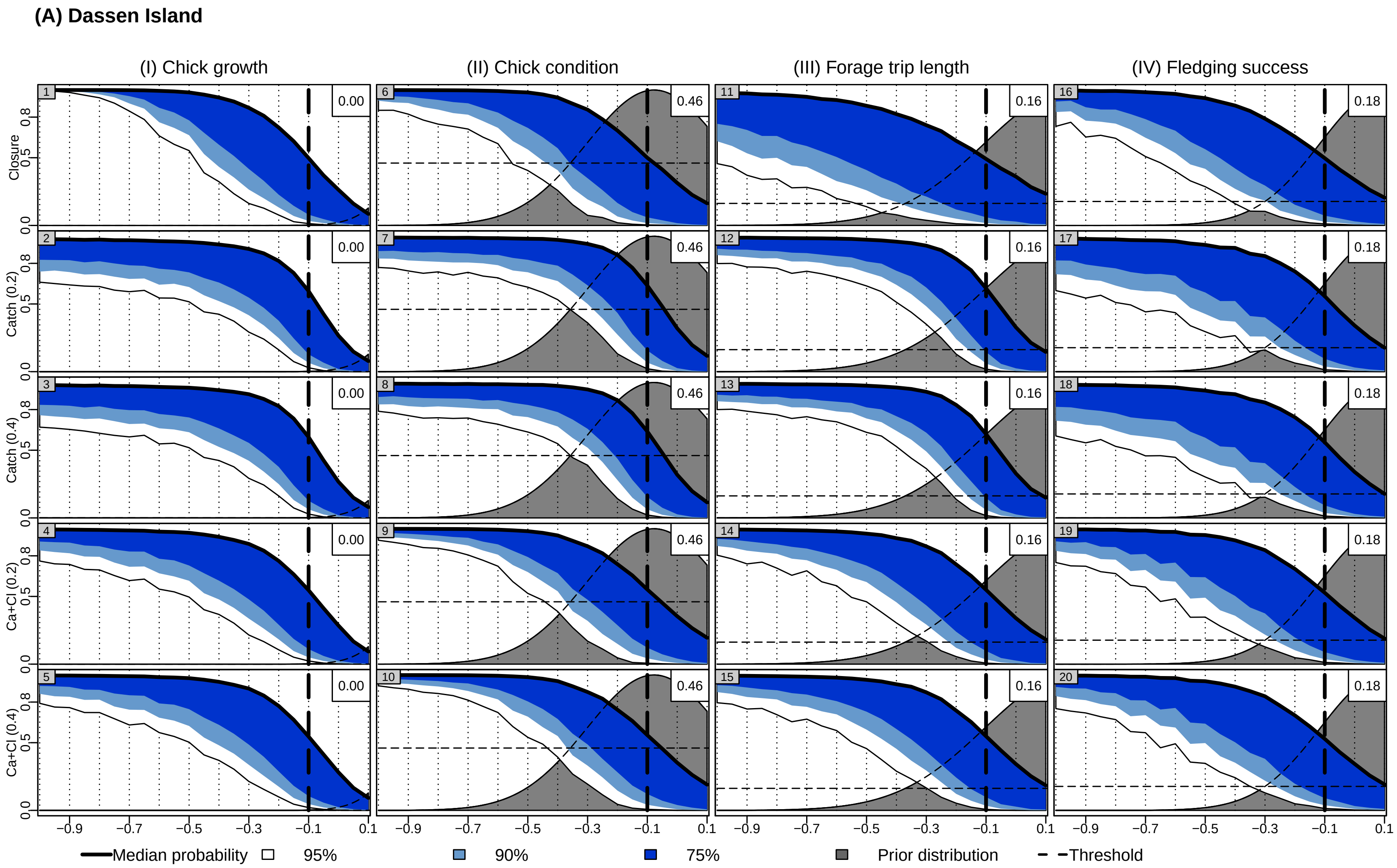

Figure 4a: The detection probabilities and percentile curves for **Dassen Island** prior to integrating across OMs. The Dassen Island plot for each data type in Figure 2 is thus a weighted combination across the five plots for that data type shown here.

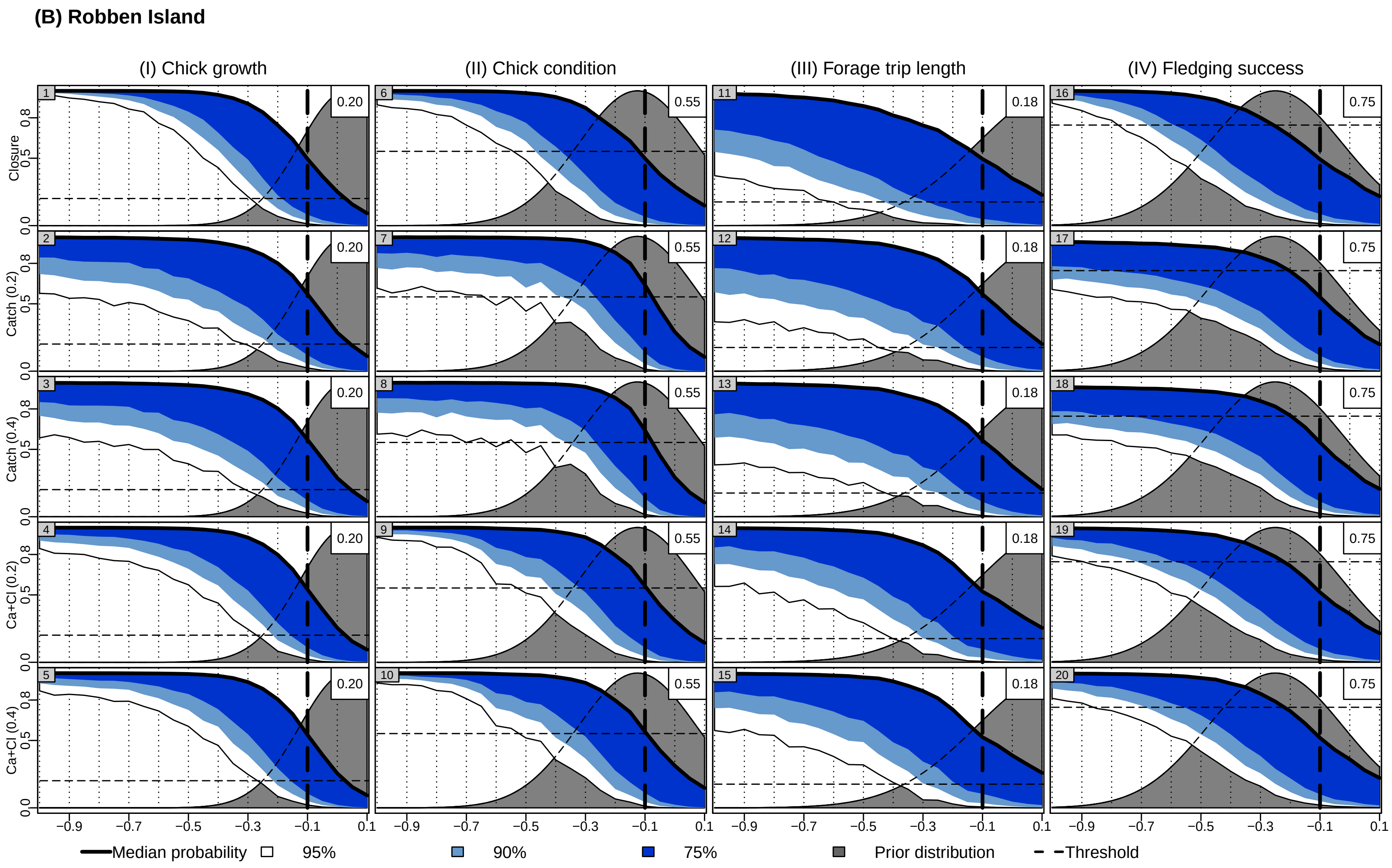

Figure 4b: The detection probabilities and percentile curves for **Robben Island** prior to integrating across OMs.

Appendix A A glossary of terms

OM	The Operating Model. 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 length and fledging success 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.
EM	The Estimation model. 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 final consolidated results of FISHERIES/2016/DEC/SWG-PEL/77 use the closure-only EM alone as per the recommendation of the 2016 IWS Panel.
δ	The closure effect, *viz.* the fishing effect. The δ parameter estimates the extent to which the island being open to fishing negatively impacts the penguin population (see Equation (1) of MARAM/IWS/DEC15/PengD/P1).
T	The Threshold. 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 log-space for the response vairable) 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.

GLM bias	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.
δ_{data}^{EM*}	The GLM-bias-adjusted estimate of δ when the EM is applied to the actual historical data.
se	The estimate of the standard error of δ when the EM is applied to the actual historical data.
P_{min}	P_{min} is the proportion of the total simulations for which $P(\delta < T)$ (i.e. the area under the normal curve to the left of the Threshold) is greater than 0.5. 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. For a balanced design, and an unbiased estimator, P_{min} should be close to 0.5 when $\delta = T$.
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 3 of FISHERIES/2016/DEC/SWG-PEL/77.
X	The (cumulative) probability that δ is less than the Threshold (i.e. $P(\delta < T)$) based on the actual data, using Equation 2 of MARAM/IWS/DEC16/Peng Clos/P1a. 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 -0.1.
detection prob.	The probability of concluding, when applying the EM to past historical and future simulated data under an assumed value of δ, that $\delta < T$.
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.
$P(\delta < T \| \delta^c)$	The detection probability at δ^c, i.e. the probability of concluding that $\delta < T$ if the true δ is equal to δ^c.
δ_{crit}	In the Algorithm of FISHERIES/2016/DEC/SWG-PEL/77, 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 have already been detected. This conclusion is also dependent on a pre-specified risk level.
"unconditioned" simulations	The actual historical data are replaced with simulated data, i.e. the detection probability curve is not conditioned on the data. Remember that for the main power analysis, simulated future data are appended to the actual historical data.