Revisiting the key results in MARAM/IWS/2019/PENG/P4 in light of the 2019 Panel recommendations

Richard B. Sherley[1,2]

1. Centre for Ecology and Conservation, College of Life and Environmental Sciences, University of Exeter, Penryn Campus, Cornwall, TR10 9FE, United Kingdom.
2. FitzPatrick Institute of African Ornithology, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch 7701, South Africa.

Email: r.sherley@exeter.ac.uk

Summary: Here, I have refit the three key models in MARAM/IWS/2019/PENG/P4 (Sherley et al. 2019) using model selection, model averaging and best-practise guidelines (Zuur et al. 2009) to choose the best fitting random effect structure (as recommended by Die et al. 2019). I have also updated the chick survival analysis for the Western Cape island pair (Robben Island and Dassen Island) to include the most recent data, so all analyses now span 11 years from 2008 to 2018. Based on the best fitting models, Chick Condition at Robben Island improved by 23% (Highest Posterior Density Interval, HPDI: −5–51%) when purse-seine fishing was banned around the island. Although the 95% HPDI overlapped zero for this effect, 96% of the posterior samples were positive. There was also weak evidence for a Closure effect at Dassen Island, with 85% of all the posterior estimates being positive. Four models were well supported for the Eastern Cape Maximum Distance dataset; based on the model averaged results, the penguins foraged 28% (HPDI: 14–44%) closer to St. Croix Island during 'Closed' years. Finally, based on the best fitting model, Chick Survival increased by 10% (HPDI: 6–15%) at Robben Island and 11% (HPDI: 5–16%) at Dassen Island during 'Closed' years, relative to 'Open' years. By updating the Overall Closure Effect, I found there is 3 times more evidence for a positive Closure effect than for no effect at present. Further consideration of whether it is sensible to include 'Island' in the random component of these models is warranted. Nevertheless, the inference about the effect of the fishing closures is now stronger than reported in Sherley et al. (2019). Moreover, the results presented here, in Sherley et al. (2019), and in the recent update of the MARAM power analysis (Ross-Gillespie and Butterworth 2020) broadly agree that biologically meaningful effects of fishing around African penguin breeding colonies are apparent on variables (e.g. chick survival) that impact the demographic process.

1. Introduction

The Panel at the 2019 International Stock Assessment Workshop made the following key recommendations regarding the work outlined in MARAM/IWS/2019/PENG/P4 (hereafter Sherley et al. 2019).

- Model selection methods should be applied to select an appropriate random effects structure;
- Use of a Bayesian fitting process combined with the Widely Applicable Information Criterion (WAIC) comparisons should be used to check model selection;
- The set of covariates to consider in the analysis should be identified by the relevant DEA working group. This requires an understanding of how the individual data are collected;
- Best practise for fitting mixed effects models (e.g., Zuur et al., 2009) should be followed (if this is not the case already). This should include standard residual analysis as well as residual analyses that are tailored to the problem at hand (e.g., temporal, spatial or within-season plots of residuals).

Here, I have refit the three key models in Sherley et al. (2019), namely Eastern Cape Maximum Distance, Western Cape Chick Condition and Western Cape Chick Survival, using model

selection and followed the guidelines in Zuur et al. (2009) to choose the best fitting random effect structure. I then present a series of residual analysis for each best fitting model, using standardized residuals for models with Gaussian errors and Pearson residuals for models with Gamma errors (Zuur et al. 2009, 2013). I have retained the fixed effects used in Sherley et al. (2019) because these were selected and/or approved by seabird biologists with understanding of how the individual data were collected and that sit on the most relevant DEA working group (the Seabird Task Team of the Top Predator Working Group).

The Panel also suggested developing a simple guide to mixed effects models to broaden the group of scientists capable of discussing these models meaningfully (Die et al. 2019). Although many good simple guides exist (e.g. Harrison et al. 2018), I will attempt to go some way to meeting this Panel recommendation here. Mixed effects model contains two components: a fixed effect component (the explanatory variables) and the random effects. In most cases, the variables we are interested in the effects of are the fixed effects (e.g. the closure effect in our case). But there may also be sources of variation that we wish to control for when we estimate the fixed effects, but that would require a large number of parameters to estimate as fixed effects themselves. For example, the year effect in our case would require 11 parameter estimates if included as a fixed effect. This can lead to models that are over-parameterised (e.g. Robinson and Butterworth 2014). Zuur et al., (2009) provide the example of such a model that has 45 data points and estimates 17 model parameters, saying "the number of parameters used by this model is excessively high". Over-parameterisation can lead to inflated standard errors for parameter estimates (compromising statistical power). Instead, to save on model parameters, we can use random effects.

To select the appropriate random effect structure, Zuur et al. (2009) recommend a top-down strategy (pp. 121–122) that begins with a model that contains all the fixed components of interest and as many interactions as possible, or this is impractical e.g. due to a large number of explanatory variables or interaction, a selection of covariates that you think are most likely to contribute to the optimal model. For the analysis here, I retained the fixed effect structures from Sherley et al. (2019). This means I retain an 'Island' × 'Closure' interaction in the fixed effect component as this leaves open the possibility of island-specific closure (or fishing) effects, an approach that has been almost unanimously applied for some time (e.g. Robinson 2013, Robinson & Butterworth 2014, Hagen et al. 2014, Sherley et al. 2018). Using these fixed effect model structures, I then find the optimal structure of the random component. As Zuur et al. (2009) put it, "Because we have as many explanatory variables as possible in the fixed component, the random component (hopefully) does not contain any information that we would like to have in the fixed component". In other words, effects we are interested in directly should be in the fixed component, those we only wish to account for should be random effects. In the context of the Island Closures experiment, this means that Island would be best placed in the fixed (not random) component of the model (plus, it only used 1 additional degree of freedom in the fixed component. Nevertheless, since models containing Island in the random component were specifically asked for by one participant at the 2019 International Stock Assessment Workshop, I have included some here (but see methods section).

Below I present the results of models based on a Bayesian fitting process and provide the Widely Applicable Information Criterion (WAIC; Watanabe 2010) for each model, as recommended by Die et al. (2019). However, since leave-one-out (LOO) cross-validation is now generally recommended over WAIC for Bayesian model selection (Gelman et al. 2014, Piironen & Vehtari 2017, Vehtari et al. 2017), I base my model selection on Pareto smoothed importance sampling (PSIS) LOO cross-validation (PSIS–LOO; Vehtari et al. 2019a) and provide model averaged results based on stacking of predictive distributions (Yao et al. 2018).

2. Methods

Datasets and the basic analytical approach remain the same as in Sherley et al. (2019), with the exception that the chick survival models (section 3.3) which now use data from 2008–2018 and are now implemented using a log-normal hazard function rather than an exponential hazard function (initial analysis indicated this provided a slightly better model fit, see Appendix 1 for details). All models were implemented using Markov Chain Monte Carlo (MCMC) estimation in JAGS (v. 4.3.0; Plummer 2003), with all priors as specified in Sherley et al. (2019). The fixed component of each model is specified in the results section along with the various random-effects structures tested. To implement the WAIC and PSIS–LOO model selection I fit each candidate model using the *jags* function in the 'jagsUI' library (version 1.5.0; Kellner 2018) for R (version 3.5.2; R Core Team 2019) and traced the log-likelihood for each observation using the log density distributions implemented in JAGS (Plummer 2017). I then pass the posterior samples for the log-likelihood for each model to the *waic* and *loo* functions in the 'loo' library (version 2.2.0; Vehtari et al. 2019b) for R to calculate the WAIC and PSIS–LOO value respectively for each model. Finally, stacking weights (Yao et al. 2018) for all the candidate models were calculated by passing the expected log pointwise predictive density (elpd) values for each model observation (the pointwise output from the *loo* function) to the *stacking_weights* function in the 'loo' library (see Vehtari & Gabry 2019). All JAGS models were fit using 3 MCMC chains of 120 000 iterations each, with the first 20 000 iterations discarded as burn-in and a thinning rate of 10, leaving 30 000 samples for inference. Unless otherwise specified, we present means and 95% highest posterior density intervals (HPDI) as the credible intervals. Convergence of all models was checked visually and using Gelman–Rubin diagnostics. All models unambiguously converged (all $\hat{R}$ values ≤ 1.001).

Lastly, we used the best fitting models for each dataset to convert all effect sizes to a percentage effect following the approach in Sherley et al. (2019). We then we combine these six percentage effect posteriors with the four from the datasets that have not been updated here (two islands for Eastern Cape Chick Condition, two islands for Western Cape Maximum Foraging Distance), but that featured in Sherley et al. (2019). We then recalculate the Overall Closure Effect (%) presented in Sherley et al. (2019) based on the updated results presented here. This approach follows that developed recently to combine regional % declines into a weighted global change % for IUCN Red List assessments (Sherley et al. 2020). Here all posterior samples are of length 30 000, so each data set is automatically given equal weighting. We plot this combined distribution, and calculate the mean, median, 95% credible intervals and the percentages of the Overall Closure Effect posterior that are above and below zero and the pre-identified 10% threshold for management action (Cochrane 2016).

Note that models with 'Island' alone as the only random effect were omitted because such a model structure effectively accounts for the exact same source of variation twice (in both the fixed and random component) and, because there are only two islands (two levels of the effect), the models cannot estimate the standard deviation for the Island random effect (because you cannot reliably estimate a variance from two observations). The result is that the models cannot partition the variance between the fixed and random effects, and either would not converge (after 500 000 iterations) or produced biologically impossible estimates for the Island fixed effect (e.g. chick condition estimates spanning the positive to negative 100s, Appendix 2). I noted this potential issue with having 'Island' in the random effect during the discussions at the 2019 International Stock Assessment Workshop.

3. Results and Discussion

3.1. Chick Condition, Western Cape

The fixed effect structure used for these models had the following components: 'Island' main effect, 'Closure' main effect, 'Island' × 'Closure' interaction, a sardine biomass main effect and an anchovy biomass main effect. The random effects structures tested are shown in Table 1, along with their corresponding WAIC, PSIS–LOO scores and stacking weights. All models are random intercept models.

The best fitting model contained the random 'Island/Year/Month' intercept and yielded positive point estimates for the closure effect at Robben Island of 22.9% (HPDI: −4.5–51.1%; mean effect size = 0.07, 95% HPDI: −0.010–0.14) and at Dassen Island of 13.1% (HPDI: −12.4–39.0%; mean effect size = 0.03, 95% HPDI: −0.03–0.10). Although neither effect was credibly different from zero based on the 95% HPDI range, 95.6% of all the posterior estimates for the closure effect were > 0 at Robben Island (down from 99.99% in Sherley et al. 2019). Although not directly comparable to a frequentist p-value, this could be thought of as corresponding to a p-value of 0.044 on a one-tailed test (i.e. assuming *a priori* that fishing cannot improve chick condition). In addition, 81.7% of the posterior estimates exceeded a 10% effect size. Although this is down from 99.8% in Sherley et al. (2019), there is still strong evidence (> 80% probability) from this model for closure effect at Robben Island that exceeds the 10% threshold (Cochrane 2016). Moreover, the evidence for a closure effect at Dassen Island is greater based on M1 (Table 1) than it was in Sherley et al. (2019), with 84.6% of all the posterior estimates being > 0 (up from 45.5% in Sherley et al. 2019) and 57.2% exceeding a 10% effect size (up from 9.4% in Sherley et al. 2019).

Table 1. Model selection results for the candidate models with different random effect structures, tested to assess the impact of the fishing closures on African penguin chick condition at Robben and Dassen Islands. M3 (Year/Month) corresponds to the original model presented in Sherley et al. (2019). Effect sizes marked in bold text are credibly different from zero (≥ 97.5% of the posterior > 0). Models are ranked by PSIS–LOO value (the *smaller* the PSIS–LOO, the better the relative model fit).

Model Number	Random effects structure	WAIC	PSIS–LOO	Stacking weight	Robben Closure effect mean (95% HPDI)	Dassen Closure effect mean (95% HPDI)
M1	Island/Year/Month	10365.9	10366.2	0.946	0.07 (−0.01–0.14)	0.03 (−0.03–0.10)
M3	Year/Month	10680.5	10680.7	0.022	**0.10 (0.05–0.14)**	−0.002 (−0.05–0.04)
M4	Island/Month	11348.0	11348.0	0.002	**0.10 (0.08–0.12)**	0.01 (−0.02–0.03)
M6	Month	11449.9	11449.9	0.019	**0.10 (0.08–0.12)**	−0.002 (−0.03–0.02)
M2	Island/Year	11499.6	11499.6	0.000	0.08 (−0.01–0.16)	0.02 (−0.07–0.10)
M5	Year	11582.6	11582.6	0.012	**0.11 (0.06–0.15)**	0.01 (−0.03–0.06)
Model-averaged results					0.07 (−0.01–0.14)	0.03 (−0.03–0.10)

Notes: / denotes nesting of the random effects, thus Island/Year/BirdID = Month nested in Year, nested in Bird Identity. WAIC = Widely Applicable Information Criterion (Watanabe 2010). PSIS–LOO = Pareto smoothed importance sampling, leave-one-out cross-validation (PSIS–LOO; Vehtari et al. 2019a). HPDI = highest posterior density interval.

The best-fitting model in this case was given ~95% of the stacking weights (Table 1), so averaging across all the models based on the stacking weights gives similar results to the best fitting model (Table 1). The model averaged closure effect at Robben Island represented an improvement during closed years of 23.6% (−4.9–51.9%) with 96% of all the posterior estimates > 0 and 82% > 10% (Figure 1). To put this in perspective, to be considered credibly different from zero (and thus bold in Table 1), 97.5% would need to be > 0. For Dassen Island, the corresponding model averaged estimates represented an increase of 13% (−12–39%) with 83% of all the posterior estimates > 0 and 55% > 10% (Figure 1).

A notable difference between the best fitting model (M1, Table 1) and the model presented in

Sherley et al. (2019; M3, Table 1) is also that the estimate for the 'Island' × 'Closure' interaction was not credible different from zero in the former, but is in the latter (Figure 2). Proceeding to simplify the fixed effects in M1 by dropping the 'Island' × 'Closure' interaction first yields a 'Closure' main effect of 0.047 (−0.0018–0.095) with 97.1% of the posterior estimates > 0, which corresponds to an 18% increase at Dassen Island and a 15% increase at Robben Island. The PSIS–LOO of this simplified model = 10366.0, thus there is no evidence from either classical Bayesian inference or model selection to retain the interaction in the model. Simplifying this model further, to drop the 'Island' main effect (which also overlapped zero) as well yields a 'Closure' main effect of 0.046 (−0.0005–0.094) with 97.33% of the posterior estimates > 0, which corresponds to a 17% increase in chick condition during 'Closed' years at both islands. The PSIS–LOO of this simplified model = 10367.0, so again this model would be considered the more parsimonious. See Appendix 3 for the candidate set of models tested and model selection results for the fixed component of the model.
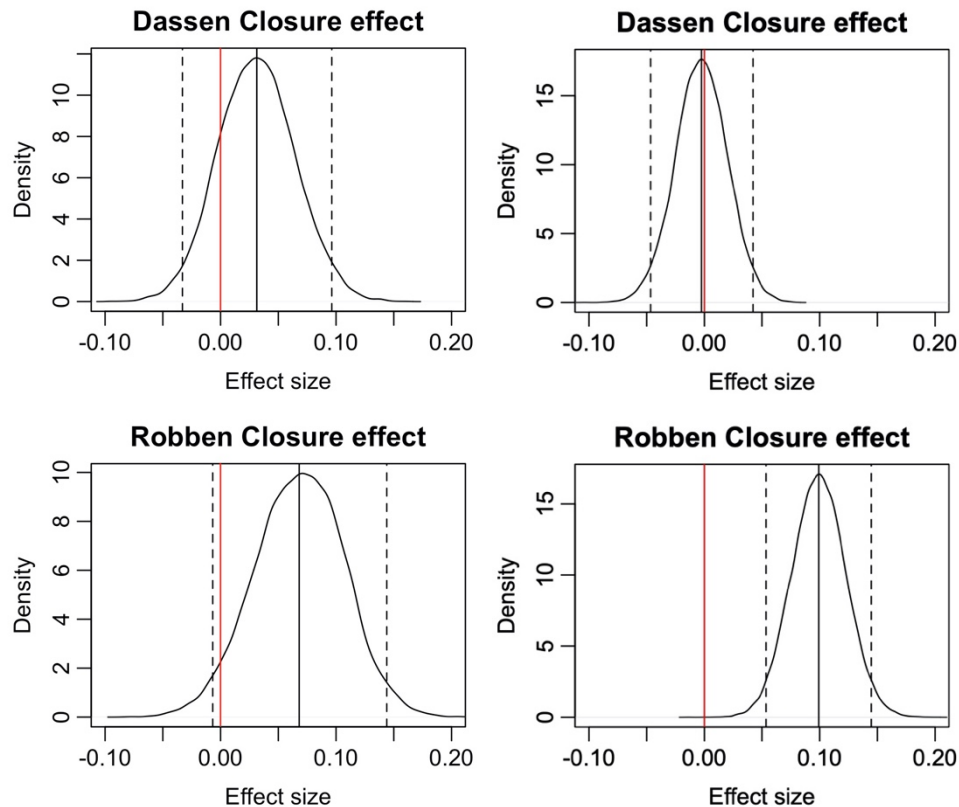

Figure 1. Posterior probability density plots along with the mean (solid black line) and 95% highest posterior density interval (HPDI, dashed lines) for the effect of Closure on chick condition at Dassen Island (top) and Robben Island (bottom) from (left) the model averaged results based on stacking (Table 1) and (right) the original model fit from Sherley et al. (2019, right). In both cases, the 95% highest posterior density intervals for the Robben Island Closure effect do not overlap zero (shown by the solid red line).

Finally, something to note is that the inclusion of 'Island' in the random component of these models consistently decreased the precision of the 'Island' fixed effect in all models but not the 'Closure' fixed effect or the 'Island' × 'Closure' interaction (Figure 2). This hints that, even with the 'Island' random effect nested, the models are struggling to partition the variance due to the different island between the fixed and random effects. Further advice should be sought from the 2019 Panel on whether 'Island' should be included in both the fixed and random effects based on this observation (also see Appendix 2) and the statement in the introduction from Zuur et al. (2009).

To conclude here, although the effect is not unambiguous (as it was in Sherley et al. 2019), the results from M1 (Table 1) provide strong evidence for a meaningful 'Closure' effect at Robben Island (>95% of all the posterior > 0) and simplifying this model strongly suggests an overall 'Closure' effect (i.e. improvements at both Dassen and Robben Island) on chick condition on the Western Cape (>97% of all the posterior > 0). Just as it has been argued that because of the random effect structure used, Sherley et al. (2019) made a type I error in concluding that a 'Closure' effect was apparent on chick condition at Robben Island (Butterworth and Ross-Gillespie 2019), I suggest that ignoring the evidence presented here strongly risks making a Type II error. Updating the chick condition analysis to include 2019 data should be seen as a priority.
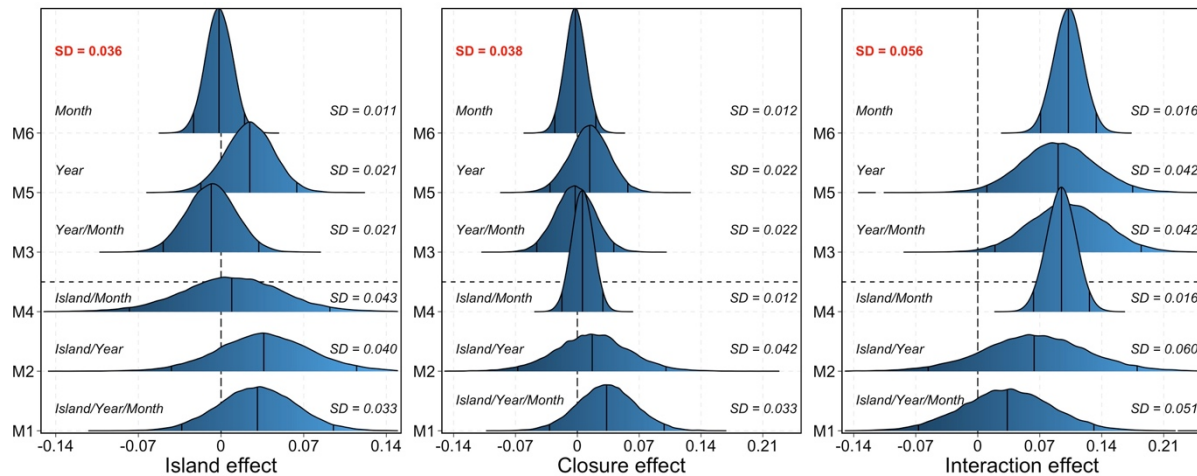

Figure 2. Posterior distributions, means and 95% credible intervals for the 'Island' main effect (left), 'Closure' main effect (middle, which represents the difference at Dassen Island in these models) and 'Island' × 'Closure' interaction (right, which represents the effect at Robben Island) for all 6 models compared using model selection to find the best fitting random effect (Table 1). The models are split above and below the horizontal dashed line based on whether 'Island' was included in their random effect structure (M1, M2 and M4), which is specified on the left-hand side, or not (M3, M5, M6). Each model's standard deviation (SD) is specified to the right-hand side. The SD shown in red in the top left-hand corner is the SD for the corresponding parameter from a frequentist model using the aggregated annual means and 'Year' as a random effect.

3.2. Maximum distance travelled, Eastern Cape

The fixed effect structure used for these models had the following components: 'Island' main effect, 'Closure' main effect, 'Island' × 'Closure' interaction, a sardine biomass main effect, an anchovy biomass main effect and a brood mass main effect (implemented as an imputed z-score standardised total brood mass (see Sherley et al. 2019). The random effects structures tested are shown in Table 2, along with their corresponding WAIC, PSIS–LOO scores and stacking weights. All models are random intercept models.

For this dataset, there were four well supported models (M6, M3, M2 and M1, Table 2), but the estimated closure effect sizes and 95% HPDI were essentially identical in all four models (Table 2). The results of all four were unchanged from the result reported in Sherley et al. (2019). They were also unambiguous, with 100% of the posterior indicating a positive effect of the Closure (max. distance decreasing in closed years) in all four models and 99.99% of the posteriors for the effect size at St. Croix exceeding a 10% effect size. Averaging across the models, according to the stacking weights assigned to each model, the effect size at St Croix Island was −0.33 (−0.49–−0.19; Figure 3), corresponding to breeding penguins foraging 28% (14–44%) closer to St. Croix Island during 'Closed' years, or a mean maximum distance travelled of 20.9 (17.1–24.4) km at St. Croix during 'Closed' years and 29.4 (26.5–32.8) km during 'Open' years. In all models, the closure effect at Bird Island was not credibly different from zero, as reported in Sherley et al. (2019).

In conclusion <u>the inference here remains unchanged</u> from that presented in Sherley et al. (2019) <u>for Maximum distance travelled at the Eastern Cape islands</u>.

Table 2. Model selection results for the candidate models with different random effect structures, tested to assess the impact of the fishing closures on the maximum distance travelled to forage by breeding African penguins at Bird and St. Croix Islands. M3 (Year/BirdID) corresponds to the original model presented in Sherley et al. (2019). Effect sizes marked in bold text are credibly different from zero (≥ 97.5% of the posterior < 0). Models are ranked by PSIS–LOO value (the *smaller* the PSIS–LOO, the better the relative model fit).

Model Number	Random effects structure	WAIC	PSIS–LOO	Stacking weight	St. Croix Closure effect mean (95% HPDI)	Bird Closure effect mean (95% HPDI)
M6	BirdID	5719.9	5754.4	0.552	**−0.32 (−0.43–−0.21)**	−0.01 (−0.08–0.07)
M3	Year/BirdID	5719.4	5754.8	0.081	**−0.32 (−0.43–−0.21)**	−0.01 (−0.09–0.07)
M2	Island/BirdID	5720.5	5756.1	0.000	**−0.32 (−0.43–−0.21)**	−0.01 (−0.08–0.07)
M1	Island/Year/BirdID	5720.0	5757.6	0.000	**−0.32 (−0.43–−0.21)**	−0.01 (−0.08–0.07)
M4	Island/Year	5787.1	5787.4	0.328	**−0.34 (−0.53–−0.15)**	0.04 (−0.14–0.22)
M5	Year	5793.0	5793.0	0.040	**−0.41 (−0.58–−0.24)**	0.07 (−0.09–0.22)
Model-averaged results					**−0.33 (−0.49–−0.19**	0.01 (−0.11–0.16)

Notes: / denotes nesting of the random effects, thus Island/Year/BirdID = Month nested in Year, nested in Bird Identity. WAIC = Widely Applicable Information Criterion (Watanabe 2010). PSIS–LOO = Pareto smoothed importance sampling, leave-one-out cross-validation (PSIS–LOO; Vehtari et al. 2019a). HPDI = highest posterior density interval.
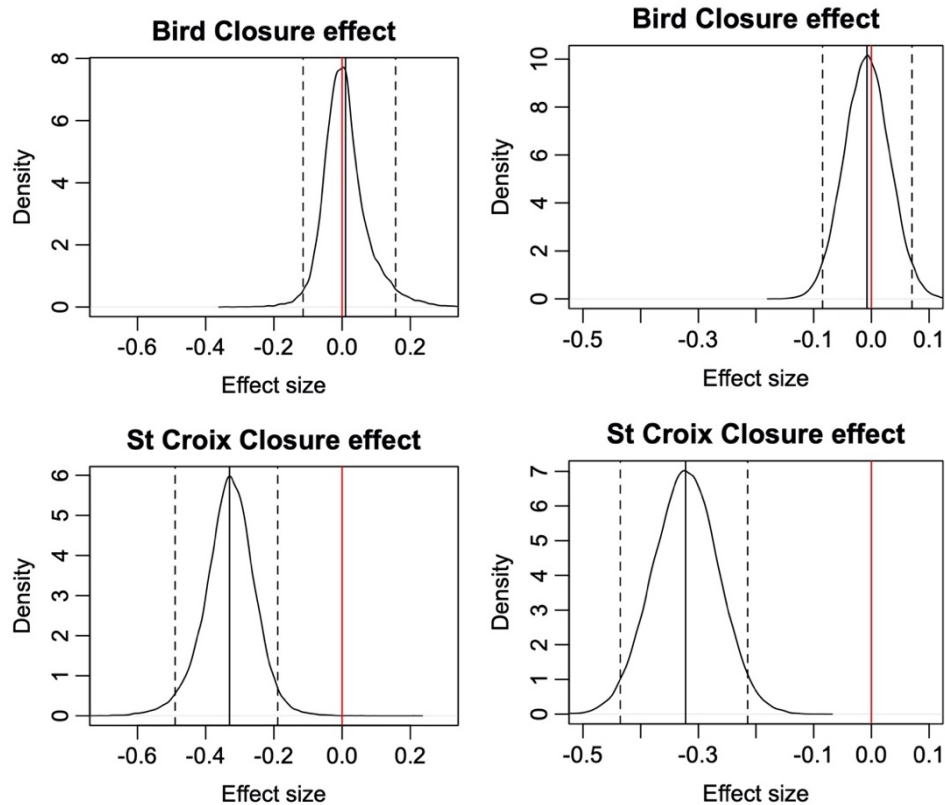

Figure 3. Posterior probability density plots along with the mean (solid black line) and 95% highest posterior density interval (HPDI, dashed lines) for the effect of Closure on maximum

foraging distance at Bird Island (top) and St. Croix Island (bottom) from (left) the model averaged results based on stacking (Table 2) and (right) the original model fit from Sherley et al. (2019, right). In both cases, the 95% highest posterior density intervals for the St. Croix Island Closure effect do not overlap zero (shown by the solid red line).

3.3. Chick Survival, Western Cape

The fixed effect structure used for these models had the following components: 'Island' main effect, 'Closure' main effect, a sardine biomass main effect, and an anchovy biomass main effect (see Sherley et al. 2019). The shared frailty term structures (akin to random intercepts) tested are shown in Table 3, along with their corresponding WAIC, PSIS–LOO scores and stacking weights.

The best fitting model contained the random 'Island/Year/Month' intercept and yielded an estimated Closure effect size of 0.38 (HPDI: 0.21–0.58). This corresponds to an improvement in survival of 10.3% (5.4–15.2%) at Robben Island and 10.6% (5.2–16.2%) at Dassen Island when the closure was in place. And, as with Max. foraging distance at the Easern Cape islands, this effect was unambiguous, with 100% of the posterior indicating a positive effect of the Closure. Respectively, 53% and 57% of the posterior distribution exceeded the 10% threshold for management action at Robben Island and Dassen Island.

Averaging across all the models based on the stacking weights (Table 3) gives similar results to the best fitting model. The model averaged 'Closure' effect estimate was 0.36 (HPDI: 0.18–0.54), with 99.9% of the posterior indicating a positive effect (Figure 4). At Dassen Island, chick survival improved by 10.1% (0.2–19.7%) with 97.7% of the posterior estimates > 0 and 51% > 10%. At Robben Island, chick survival also improved by 9.9% (1.1–18.2%) with 97.9% of all the posterior estimates > 0 and 49% > 10%.

In conclusion, the closure effect is credible, even when taking the uncertainty of all model structures into account. And, at present the inference here remains unchanged from that presented in Sherley et al. (2018, 2019) for chick survival at the Western Cape islands, even with 3 more years of data added to the analysis.

Table 3. Model selection results for the candidate models, tested to assess the impact of the fishing closures on African penguin chick survival at Robben and Dassen Islands. M3 (Year/NestID) corresponds to the original random effect structure presented in Sherley et al. (2019). Effect sizes marked in bold text are credibly different from zero (≥ 97.5% of the posterior > 0). Models are ranked by PSIS–LOO value (the *smaller* the PSIS–LOO, the better the relative model fit).

Model Number	Random effects structure	WAIC	PSIS–LOO	Stacking weight	Closure effect mean (95% HPDI)
M1	Island/Year/NestID	15140.0	15675.4	0.768	**0.38 (0.21–0.55)**
M3	Year/NestID	16474.0	16595.6	0.003	**0.39 (0.27–0.52)**
M5	Year	17144.9	17145.5	0.201	**0.30 (0.16–0.44)**
M4	Island/Year	17151.5	17151.6	0.024	0.36 (−0.02–0.74)
M2	Island/NestID	17170.4	17171.0	0.004	**0.37 (0.24–0.50)**
M6	NestID	17201.0	17201.1	0.000	**0.36 (0.23–0.49)**
	Model-averaged results				**0.36 (0.18–0.54)**

Notes: / denotes nesting of the random effects, thus Island/Year/BirdID = Month nested in Year, nested in Bird Identity. WAIC = Widely Applicable Information Criterion (Watanabe 2010). PSIS–LOO = Pareto smoothed importance sampling, leave-one-out cross-validation (PSIS–LOO; Vehtari et al. 2019a). HPDI = highest posterior density interval.
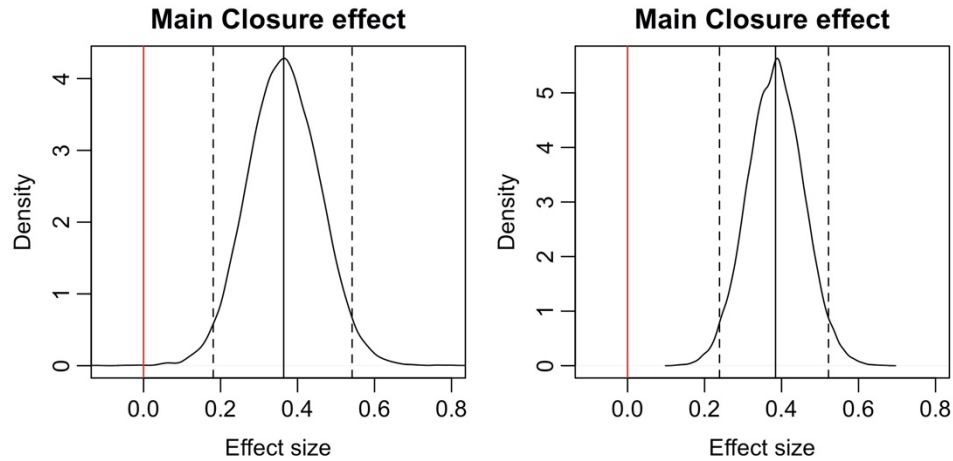
Figure 4. Posterior probability density plots along with the mean (solid black line) and 95% highest posterior density interval (HPDI, dashed lines) for the main effect of Closure on chick survival at Dassen Island and Robben Islands from (left) the model averaged results based on stacking (Table 3) and (right) the original random effect structure from Sherley et al. (2019; M3, Table 3). In both cases, the 95% highest posterior density intervals for the Robben Island Closure effect do not overlap zero (shown by the solid red line).

3.4. Overall Closure Effect

The Overall Closure Effect (%) is calculated by combining the posterior distributions for the percentage effect sizes for all ten penguin responses (Figure 5) into a single vector in R using the *c* (combine) function. The posterior distributions are combined with equal weighting (30 000 posterior samples each) given to each penguin response. This is akin to the approach developed recently to combine regional % declines into a weighted global change % for IUCN Red List assessments (Sherley et al. 2020). We plot this combined distribution, with all negative posterior samples (≤ 0) shown in red, all positive posterior samples (>0) shown in green (Figure 5), and calculate the mean, median, 95% HPDI and the percentages of the Overall Closure Effect posterior that are above both zero and the 10% threshold for management action (Cochrane 2016).

Table 4. Comparison of the Overall Closure Effect presented in Sherley et al. (2019) and the one calculated here.

Parameter	Sherley et al. 2019	This report
Mean (%)	8.5	8.3
Median (%)	8.5	8.7
95% HPDI (%)	(−36–45)	(−36–40)
% positive (>0)	71.7	75.2
Ratio of evidence for a closure effect	2.53	3.03
% exceeding management threshold (>10%)	44.3	44.3
Ratio >10% to <−10%	3.69	3.96

The mean (95% HPDI) of the combined posterior distribution for the Overall Closure Effect mean was 8.3 (−36–40)% (Figure 5E; the median was 8.7%). Overall, 75.2% of the posterior was positive and 24.8% was negative (Figure 5E). Taking a 50:50 split in the posterior as the null, then the 50.4% point difference between the two halves of the posterior suggests that the evidence for a closure effect is 3 times greater than the evidence for no effect at present (75.2/24.8 = 3.03). Moreover, if only the posterior samples exceeding a 10% effect are considered (in line with the pre-identified threshold for management action; Cochrane 2016), the evidence for a closure effect is stronger: 44.3% of the posterior samples in Figure 5E were > 10% and 11.2% were < −10%. This suggests that the evidence for a closure effect exceeding the threshold for management action is almost 4 times greater than the evidence for no effect at present (44.3/11.2 = 3.96). In general, these results provide slightly stronger evidence for an Overall Closure Effect than the results presented in Sherley et al. (2019) (Table 4).
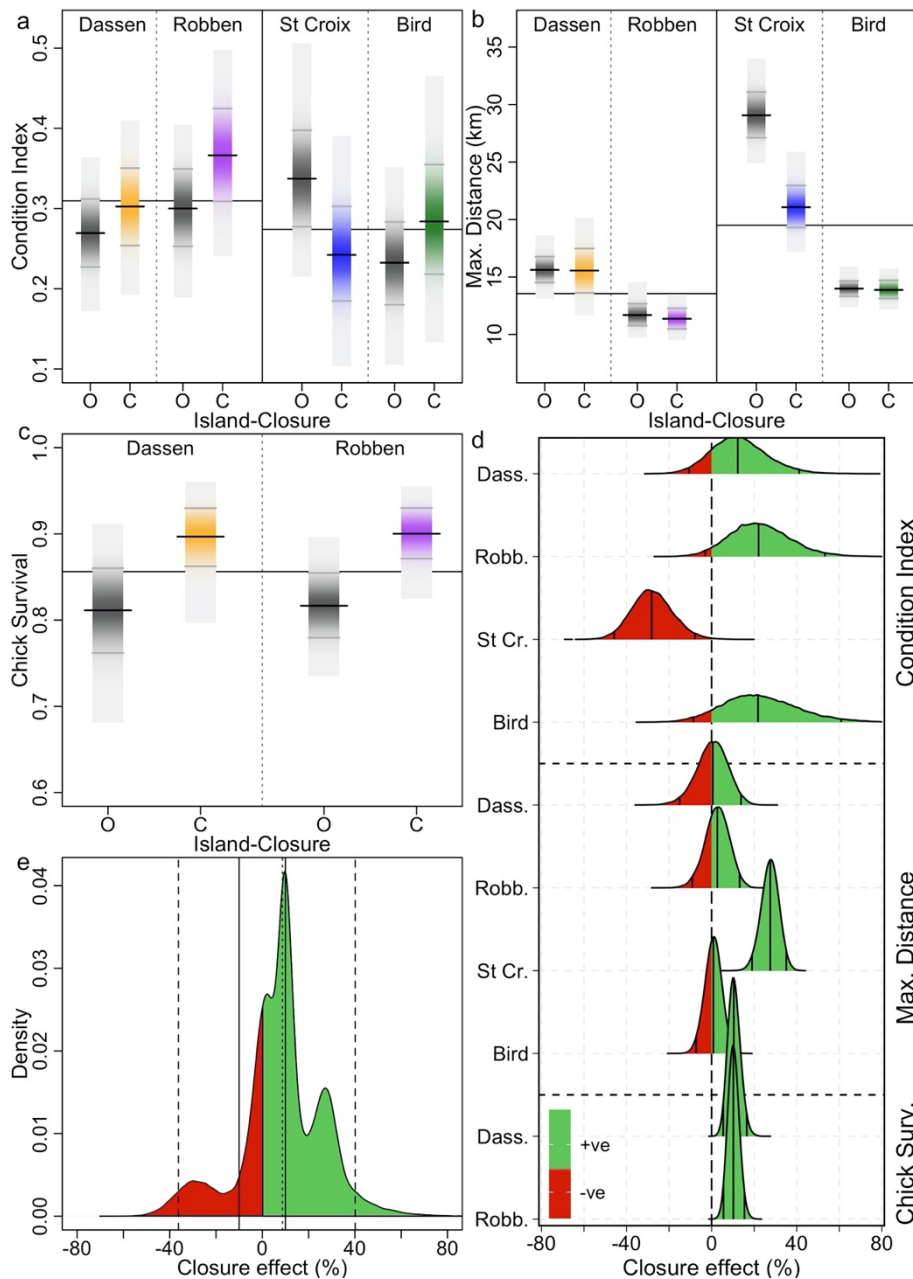
Figure 5. Posterior distributions, means and 95% highest posterior density interval (HPDI) for (A) chick body condition index and (B) the maximum distance travelled by foraging penguins at Dassen Island, Robben Island, St Croix Island and Bird Island, and for (C) chick survival at Dassen Island and Robben Island for years where fishing was permitted ['O'] or not permitted ['C']. Open ['O'] results are shown in black, Closed ['C'] are in orange for Dassen, purple for Robben, blue for St Croix, dark green for Bird. Black tick marks denote the posterior mean (calculated at mean anchovy and sardine biomass in both A and B and at mean brood mass in B), grey ticks the 95% HPDI and grey polygons the range of the posterior distribution. The horizontal solid black lines show the overall mean at each island pair for chick condition (in A) or maximum distance (in B) and chick survival (in C) across 11 years (2008–2018). (D) Posterior distributions for the percentage difference between 'Closed' years and 'Open' years for chick body condition [Condition Index], the maximum distance travelled from the island by foraging penguins [Max. Distance] and chick survival [Chick Surv.] at Dassen Island [Dass.], Robben Island [Robb.], Bird Island [Bird] and St Croix Island [St Cr.]. The mean and 95% credible intervals are shown on each posterior distribution as solid black lines. The zero axis (no effect of closure) is shown as a dashed black line. (E) Posterior distribution (polygon), median (dotted black line) and 95% HPDI (dashed black lines) for the Overall Closure Effect (%), the overall difference between 'Closed' years and 'Open' years based on combining the 10 individual posteriors in D. The solid black lines show a pre-agreed 10% threshold for management action. In (D) and (E),

all 10 individual posteriors distribution have equal weighting (30 000 posterior samples) and all samples yielding a positive % effect for penguins shown in green [+ve] and those yielding a negative % effect of the closure on the penguins are shown in red [−ve]. Chick body condition index and chick survival results for Dassen Island and Robben Island, and maximum distance travelled by foraging penguin results for St Croix Island and Bird Island are based on the best fitting models presented here (see Tables 1–3). Chick body condition index results for St Croix Island and Bird Island and maximum distance travelled by foraging penguin results for Dassen Island and Robben Island are unchanged from Sherley et al. 2019.

3.5. Conclusions and next steps

It has been argued that the effects presented in Sherley et al. (2019) were not robust because the allegedly poorly chosen random effect structure resulted in fixed effect estimates that were overly precise (Butterworth and Ross-Gillespie 2019). However, the results presented here suggest little meaningful change in inference whether or not Island was included as a higher-level random effect. This result has also just been supported by a new simulation study on the robustness of linear mixed-effects models which concluded that "missing random effect predictors had little effect on the fixed effect estimates but had systematic effects on the estimates of random effects. The variance due to unmodelled higher level predictors [like Island in our case] was almost completely absorbed by the nested random effect variance of interest" (Schielzeth et al. 2020).

As a consequence of our updated results, it has now been shown that meaningful, unbiased, positive effects of the fishing closures are apparent in all three of the following cases:

a) When the observation level (disaggregated) data were modelled using random effects that were selected *a priori* by the analyst to adequately reflect the sampling structure of the data (Sherley et al. 2018, 2019);
b) When the observation level (disaggregated) data were modelled using random effects that were selected using model selection (this document), as advised by the 2019 Panel (Die et al. 2019);
c) When the aggregated data (annual means) are modelled appropriately (e.g. Sherley and Winker 2019) to avoid over-parameterization (as in Robinson 2013, Robinson and Butterworth 2014), even when those models account for the differing precision of each annual estimate (Sherley et al. 2015).

Moreover, the most recent update of the MARAM power analysis (Ross-Gillespie and Butterworth 2020) also finds "evidence in the current data of a biologically meaningful fishing effect" on chick condition at Robben Island, fledging success at Robben Island (chick survival measures a component of fledging success), and chick survival at Dassen Island. Those results from Ross-Gillespie and Butterworth (2020) concur with results presented here in section 3.1 and 3.3. Plus, Ross-Gillespie and Butterworth (2020) also find that "a biologically meaningful fishing effect is likely to be detected" if the experiment continues for 2–5 years (using a dataset that ended in 2015) for chick survival at Robben Island. This broadly concurs with the results presented here in section 3.3 as well. In other words, **we have now iterated to a place where two independent sets of analyses agree that biologically meaningful effects of fishing around African penguin breeding colonies are apparent** and, importantly, that some of those effects are on variables (chick survival, fledging success) that contribute to the demographic process.

I propose the next steps should be:
i) Seek guidance from the 2019 Panel on the sense of including 'Island' in both the random and fixed components of these models (given the results herein and recommendations in Zuur et al. 2009).
ii) Based on the advice pertaining to i), if necessary, undertake model selection as presented here for the other variables examined in Sherley et al. (2019),

namely chick condition in the Eastern Cape and maximum foraging distance in the Western Cape.

iii) Based on the advice pertaining to i), if necessary, recalculate the Overall Closure Effect presented in Sherley et al. (2019) using these updated model structures.

iv) Seek written confirmation from the 2019 International Stock Assessment Workshop Panel that the results herein [and if necessary the outcomes of steps i) to iii)] satisfy the recommendations in Die et al. (2019) and, if so, make a final management recommendation regarding the Island Closures Experiment to the Department of Environment, Forestry and Fisheries by December 2020.

References

Butterworth DS and Ross-Gillespie A. 2019. Is pseudo-replication biasing results from analyses from the island closure experiment which model individual penguin responses directly? Department of Environment, Forestry and Fisheries Report: FISHERIES/2019/NOV/SWG-PEL/34. Pp. 1–10.

Christensen R, Johnson W, Branscum A and Hanson TE. 2010. *Bayesian Ideas and Data Analysis. An Introduction for Scientists and Statisticians.* CRC Press, Boca Ranton.

Cochrane K. 2016. Chair's Introduction to documents from the Technical Team on the Penguin Island Closure Experiment. Department of Environment, Forestry and Fisheries Report: MARAM/IWS/DEC16/Peng Clos/P6. Pp. 1–4.

Die DJ, Punt AE, Tiedemann R, Waples R and Wilberg MJ. 2019. International Review Panel Report for the 2019 International Fisheries Stock Assessment Workshop, 2–5 December 2019, UCT. Department of Environment, Forestry and Fisheries Report: MARAM/IWS/2019/General/5. Pp. 1–18.

Gelman A, Hwang J, and Vehtari A. 2014. Understanding predictive information criteria for Bayesian models. *Statistics and Computing* 24: 997–1016.

Hagen C, Jarre A, Shannon L, Sherley RB, Steinfurth A, Crawford RJM, van der Merwe L, Wanless RM, Oosthuizen H, Pichegru L, Robinson K, Weller F, McInnes A, Winker H, Altwegg R, Ludynia K, Waller L and Makhado AB. 2014. Evaluating the state of knowledge on fishing exclusions around major African Penguin colonies. Department of Environment, Forestry and Fisheries Report: MARAM/IWS/DEC14/PENG/A1. Pp. 1–12.

Harrison XA, Donaldson L, Correa-Cano ME, Evan J, Fisher DN, Goodwin CED, Robinson BS, Hodgson DJ and Inger R. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6: e4794.

Jackson C. 2016. flexsurv: A Platform for Parametric Survival Modeling in R. *Journal of Statistical Software* 70: 1–33.

Piironen J, and Vehtari A. 2017. Comparison of Bayesian predictive methods for model selection. *Statistics and Computing* 27: 711–735.

Plummer M. 2003. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In: Hornik K., Leisch F., Zeileis A. (eds.), Proceedings of the Third International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria. ISSN 1609-395X, http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/Plummer.pdf

Plummer M. 2017. JAGS Version 4.3.0 user manual. https://sourceforge.net/projects/mcmc-jags/files/Manuals/4.x/jags_user_manual.pdf/download

Robinson WML. 2013. *Modelling the impact of the South African small pelagic fishery on African penguin dynamics.* PhD Thesis, University of Cape Town.

Robinson WML and Butterworth DS. 2014. Island closure feasibility study power analysis results for Dassen and Robben islands. Department of Environment, Forestry and Fisheries Report: FISHERIES/2014/MAR/SWG-PEL/ICTT/5. Pp. 1–11.

Ross-Gillespie A and Butterworth DS. 2020. 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.

Department of Environment, Forestry and Fisheries Report: FISHERIES/2020/JAN/SWG-PEI/09. Pp. 1–17.

Schielzeth H, Dingemanse N, Nakagawa S, Westneat DF, Allegue H, Teplitsky C, Réale D, Dochtermann NA, Garamszegi L and Araya-Ajoy Y. 2020. Robustness of linear mixed-effects models to violations of distributional assumptions. *Methods in Ecology and Evolution*. DOI: 10.1111/2041-210X.13434.

Sherley RB, Barham BJ, Barham PJ, Campbell KJ, Crawford RJM, de Blocq A, Grigg J, Le Guen C, Hagen C, Ludynia K, Makhado AB, McInnes A, Meyer A, Morris T, Pichegru L, Steinfurth A, Upfold L, van Onselen M, Visagie J, Weller F and Winker H. 2019. A Bayesian approach to understand the overall effect of purse-seine fishing closures around African penguin colonies. Department of Environment, Forestry and Fisheries Report: MARAM/IWS/2019/PENG/P4. pp. 1–25.

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 and 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.

Sherley RB, Winker H, Altwegg R, van der Lingen CD, Votier SC and Crawford RJM 2015. Bottom-up effects of a no-take zone on endangered penguin demographics. *Biology Letters* 11: 20150237.

Sherley RB, Underhill LG, Barham BJ, Barham PJ, Coetzee JC, Crawford RJM, Dyer BM, Leshoro TM and Upfold L. 2013. Influence of local and regional prey availability on breeding performance of African penguins *Spheniscus demersus. Marine Ecology Progress Series* 473: 291–301.

Sherley RB, Winker H, Rigby CL, Kyne PM, Rollom R, Pacoureau N, Herman K, Carlson JK, Yin JS, Kindsvater HK and Dulvy NK. 2020. Estimating IUCN Red List population reduction: JARA – a decision-support tool applied to pelagic sharks. *Conservation Letters* 13: e12688.

Therneau TM, Grambsch PM and Pankratz VS. 2003. Penalized Survival Models and Frailty. *Journal of Computational and Graphical Statistics* 12: 156–175.

Vehtari A, Gelman A, and Gabry J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27: 1413–1432.

Vehtari A, and Gabry J. 2019. Bayesian Stacking and Pseudo-BMA weights using the loo package. https://cran.r-project.org/web/packages/loo/vignettes/loo2-weights.html

Vehtari A, Simpson D, Gelman A, Yao Y, and Gabry J. 2019. Pareto Smoothed Importance Sampling. *arXiv* 1507.02646v6.

Watanabe S. 2010. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* 11: 3571–3594.

Yao Y, Vehtari A, Simpson D, and Gelman A. 2018. Using Stacking to Average Bayesian Predictive Distributions (with Discussion). *Bayesian Analysis* 13: 917–1007.

Zuur AF, Ieno EN, Walker NJ, Saveliev AA, and Smith GM. 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer, New York.

Zuur AF, Hilbe JM, and Ieno EN. 2013. *A Begginer's Guide to GLM and GLMM with R: A frequentist and Bayesian perspective for ecologists*. Highland Statistics Ltd., Newburgh, UK.

Appendix 1: JAGS output for the chick condition model with 'Island' as the only random effect

Chick Condition, Western Cape: Having 'Island' as the only random effect results in poorly estimated SDs for the main Island fixed effect "beta.island" and for the 'Island' random effect "sigma.int" (because the model cannot partition the variance), with the downstream effect of biologically implausible values in the range of estimates for chick condition at each island in the Open and Closed years (i.e. mu.oD, which is the Open years at Dassen Island). The relevant values in the Table are highlighted in bold text.

Parameter	mean	sd	2.50%	50%	97.50%	overlap0	f	Rhat	n.eff
mu.oD	−0.25	**57.69**	**−127.47**	0.08	**124.78**	TRUE	0.50	1	30000
sigma	0.41	0.00	0.41	0.41	0.42	FALSE	1.00	1	13649
mu.cD	−0.24	**57.69**	**−127.46**	0.09	**124.81**	TRUE	0.50	1	30000
beta.close	0.01	0.01	-0.01	0.01	0.03	TRUE	0.84	1	30000
beta.island	0.71	**81.03**	**−175.13**	0.14	**178.40**	TRUE	0.50	1	30000
sigma.int	**49.93**	**28.77**	2.56	49.84	**97.39**	FALSE	1.00	1	7589
beta.sard	−0.19	0.04	−0.27	−0.19	−0.10	FALSE	1.00	1	30000
beta.anch	−0.01	0.01	−0.03	−0.01	0.00	TRUE	0.90	1	13888
beta.inter	0.10	0.02	0.07	0.10	0.13	FALSE	1.00	1	20448
bC.bl	0.11	0.01	0.09	0.11	0.13	FALSE	1.00	1	11459
mu.oR	0.46	57.41	**−124.16**	0.24	**126.65**	TRUE	0.51	1	30000
mu.cR	0.57	57.41	**−124.06**	0.37	**126.76**	TRUE	0.51	1	30000

Notes: mean = posterior mean; sd = posterior standard deviation; 2.50% = the lower 95% credible limit; 50% = the posterior median; 97.50% = the upper 95% credible limit; overlap0 = whether the 95% credible interval overlaps zero; f = the proportion of the posterior with the same sign as the mean (i.e., our confidence that the parameter is positive or negative), Rhat is the Gelman-Rubin diagnostic, indicating convergence at Rhat = 1) and n.eff is a crude measure of the effective sample size (iterations used to estimate that parameter, higher is better).

Appendix 2: Details of the new Chick Survival models.

Previously I have estimated chick failure rates (deaths/unit time of exposure or hazard functions) using an exponential hazard function and used an exponential distribution to transform these failure rates into survival estimates (Sherley et al. 2013, 2015, 2018) as this is the simplest model structure (it assumes hazards [mortality rates] are constant over time). However, preliminary analysis using the *flexsurvreg* function from the 'flexsurv' library in R (Jackson 2016) that used a frequentist approach with null models and no frailty terms to compare exponential, Weibull, Gompertz and log-normal hazard functions supported the use of log-normal hazards (ΔAICc = 49.1 over the next best model). We confirmed this by comparing the pareto smoothed importance sampling leave-one-out cross-validation (PSIS–LOO; Vehtari et al. 2019a) scores between a model specified in JAGS using an exponential hazard function (PSIS–LOO = 21620.7) and one using a log-normal hazard function (PSIS–LOO = 16594.0). Both models used the shared frailty structure presented in previous analyses (Year/Nest ID; Sherley et al. 2018, 2019) and included the fixed effects outlined in eqn. A2.1. The log-normal hazard function gave a substantially improved model fit (ΔPSIS–LOO = 5027.7) and estimates of chick survival were not credibly different from those presented in the past (Figure A2.1).

Thus, here we estimated failure rates (deaths/unit time of exposure or hazard functions) for 'Open' or 'Closed' years using a log-normal error distribution (Christensen et al. 2010). We used Year, Island and Nest Identity to specify hierarchical shared frailty terms (see Table 3 in the main text), which are analogous to random effects (Therneau et al. 2003); i.e. the survival rates of chicks within the same nest, year and island could be considered non-independent. The mean hazard function (Λ) was estimated as:

$$t_{i,j,k,l} \sim \text{Lognormal}(\Lambda_{i,j,k,l}, \tau),$$
$$\Lambda_{i,j,k,l} = \alpha + \beta_1 x_j + \beta_2 z_i + \beta_3(S_j - \bar{S}) + \beta_4(A_j - \bar{A}) + \omega_{j,k},$$
$$i = 1,2, \quad j = 1,\dots,8, \quad k = 1,\dots,m_{i,j} = 239, \quad l = 1,\dots n_{i,j,k} = 3\,699, \quad \omega_{j,k} \sim \text{Normal}(0, \sigma^2)$$

(eqn. A2.1)

$t_{i,j,k,l}$ denotes the observed time of exposure for each chick (l), in nest k, in year j, at island i; $\omega_{j,k}$ denotes the shared frailty term (here, nest nested in year); α is the intercept; the β's are the coefficients to be estimated for the fixed-effects; x_j is a binary covariate for the island closure effect ('Open' = 0, 'Closed' = 1); z_i is a binary covariate denoting to which island a chick belongs (Dassen = 0, Robben = 1); S_j and A_j are sardine and anchovy (respectively) biomass estimates associated to year j and $\bar{S}$ and $\bar{A}$ the mean biomass of each species over the years considered; and the variance (σ) shared frailty term and the standard deviation (τ) for the log-normal distribution were estimated from the data.

The Λs were then transformed into estimates of chick survival (S) at $d = 74$ days (following Sherley et al. 2013, 2015, 2018) as:

$$S(d) = 1 - \Phi((\log(d) - \Lambda) \times \sqrt{\tau})$$

(eqn. A2.2)

where Φ is the cumulative distribution function of the normal distribution (Christensen et al. 2010).
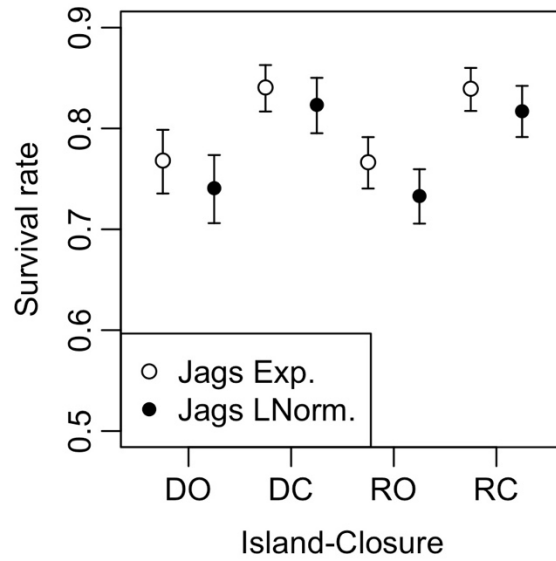
Figure A2.1. Comparisons of the mean (points) and 95% highest posterior density intervals (whiskers) for Chick Survival estimates for Dassen Island (D) and Robben Island (R) during years that were 'Open' (O) or 'Closed' (C) to fishing from a parametric survival model specified in JAGS using an exponential hazard function (white points) (as in Sherley et al. 2018, 2019) and a log-normal hazard function (black points). Both models had the same random effect (shared frailty terms; nest ID nested in year) and the same fixed effect structure (eqn A2.1).

Appendix 3: Fixed-effect model selection for the Chick Condition, Western Cape dataset.

Table A3.1. Model selection results for the candidate models with different fixed effect structures, tested to assess the impact of the fishing closures on African penguin chick condition at Robben and Dassen Islands. The random effect structure used in all models was the best supported model from Table 1, M1 (Island/Year/Month). Models are ranked by PSIS–LOO value (the *smaller* the PSIS–LOO, the better the relative model fit) and the best supported model from Table 1 (M1) is included for reference.

Model Number	Fixed effects	WAIC	PSIS–LOO	Stacking weight	Effect size* (95% HPDI)	% >0
MF5	C + I + C×I + S	10364.8	10365.4	0.291	0.07 (−0.01–0.14)	95.5
MF3	C + S	10364.9	10365.5	0.312	0.05 (0.00–0.10)	96.1
MF4	C + I + S	10364.9	10365.6	0.253	0.05 (0.00–0.09)	97.0
MF1	C + I + S + A	10365.8	10366.0	0.145	0.05 (0.00–0.09)	97.1
M1	C + I + C×I + S + A	10365.9	10366.2	0.000	0.07 (−0.01–0.14)	95.6
MF2	C + S + A	10366.4	10367.0	0.000	0.05 (0.00–0.09)	97.3

Notes: WAIC = Widely Applicable Information Criterion (Watanabe 2010). PSIS–LOO = Pareto smoothed importance sampling, leave-one-out cross-validation (PSIS–LOO; Vehtari et al. 2019a). HPDI = highest posterior density interval. C = Closure main effect, I = Island main effect, C×I = Island-Closure interaction, S = sardine biomass main effect, A = anchovy biomass main effect. * For models containing the Island-Closure interaction, the effect size given is the interaction (effect of closure at Robben Island) for models without the Island-Closure interaction, the effect given is the main Closure effect.

Averaging across all the models in Table A3.1 according to the stacking weights gives an effect size of 0.05 (−0.01–0.12) with 96.2% of the combined posterior > 0 (Figure A3.1).
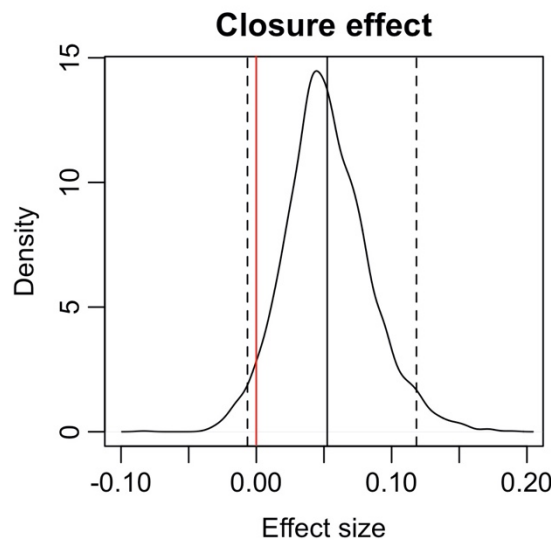
Figure 4. Posterior probability density distribution, mean (solid black line) and 95% highest posterior density interval (HPDI, dashed lines) for the effect of Closure on chick condition at Dassen Island and Robben Island from the model averaged results based on stacking (Table A3.1). A zero effect is shown by the solid red line.

Appendix 4: Diagnostic plots for the three best fitting models.

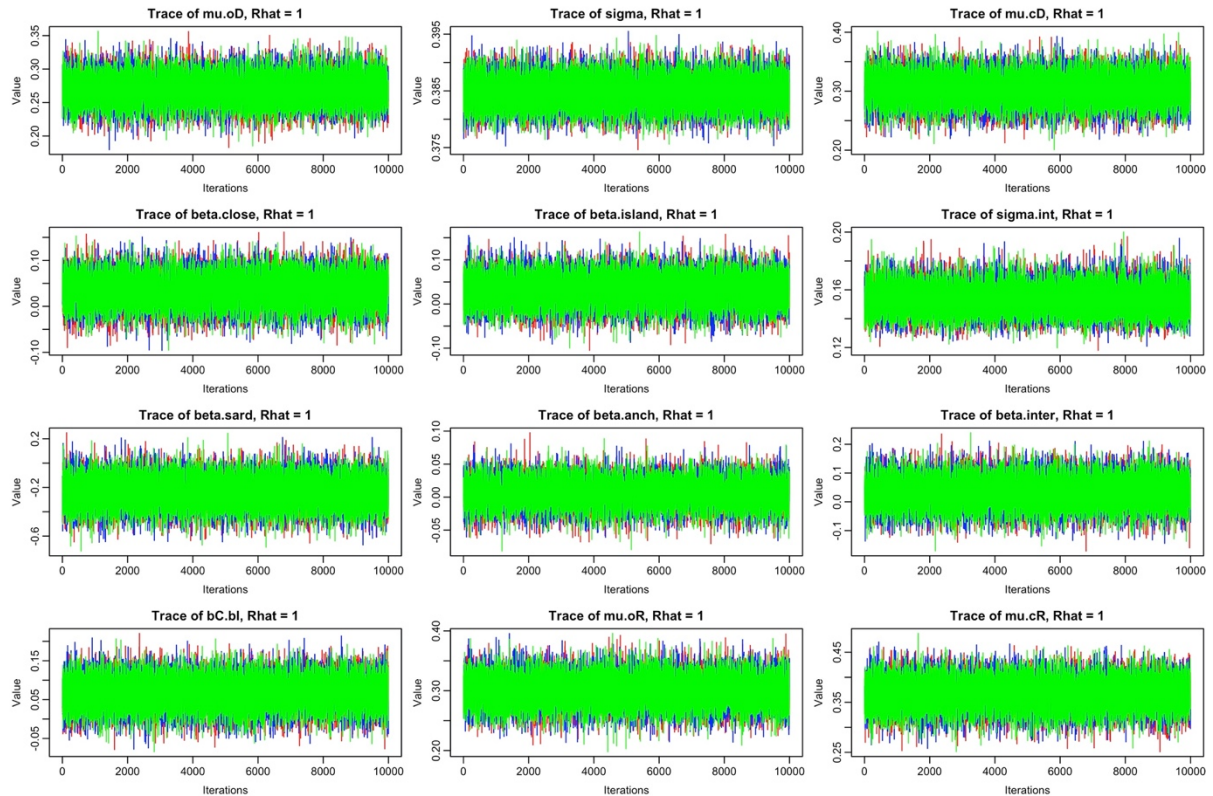## Chick Condition, Western Cape

Figure A4.1. Trace plots used to assess convergence of the three MCMC chains used to estimate parameters in the best fitting Western Cape Chick Condition model (M1, Table 1).
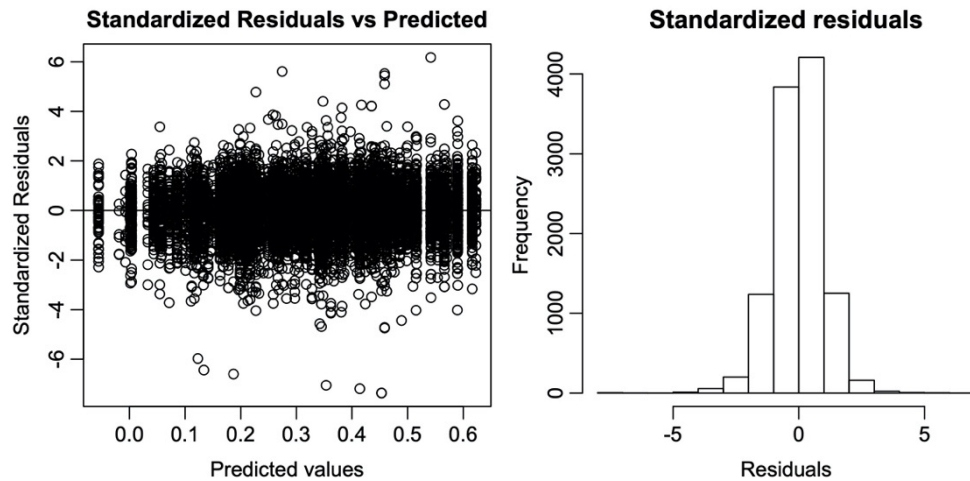

Figure A4.2. Standardised residual versus predicted (left) and histogram of standardised residuals (right) for the best fitting linear mixed model (Gaussian error, identity link, M1, Table 1) used to assess the effect of the fishing closures on chick condition at Dassen Island and Robben Island.
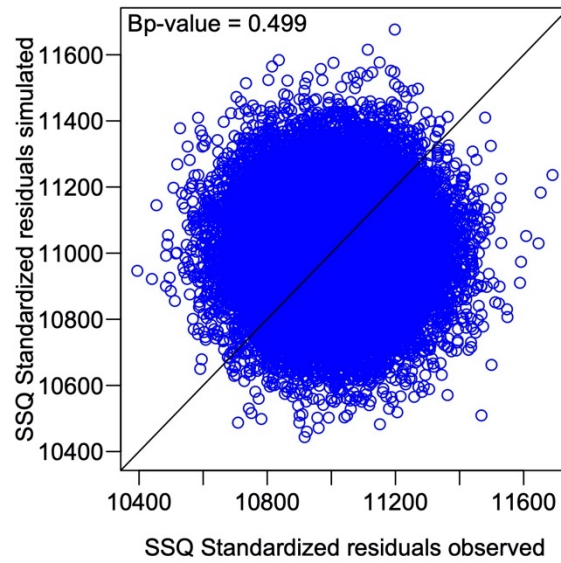
Figure A4.3. Posterior predictive check of the model adequacy for the best fitting model (M1, Table 1) assessing the effect of the fishing closure on chick condition at Dassen Island and Robben Island, plotting simulated versus observed sums-of-squares discrepancies based on Standardized residuals. The Bayesian p-value = 0.499.
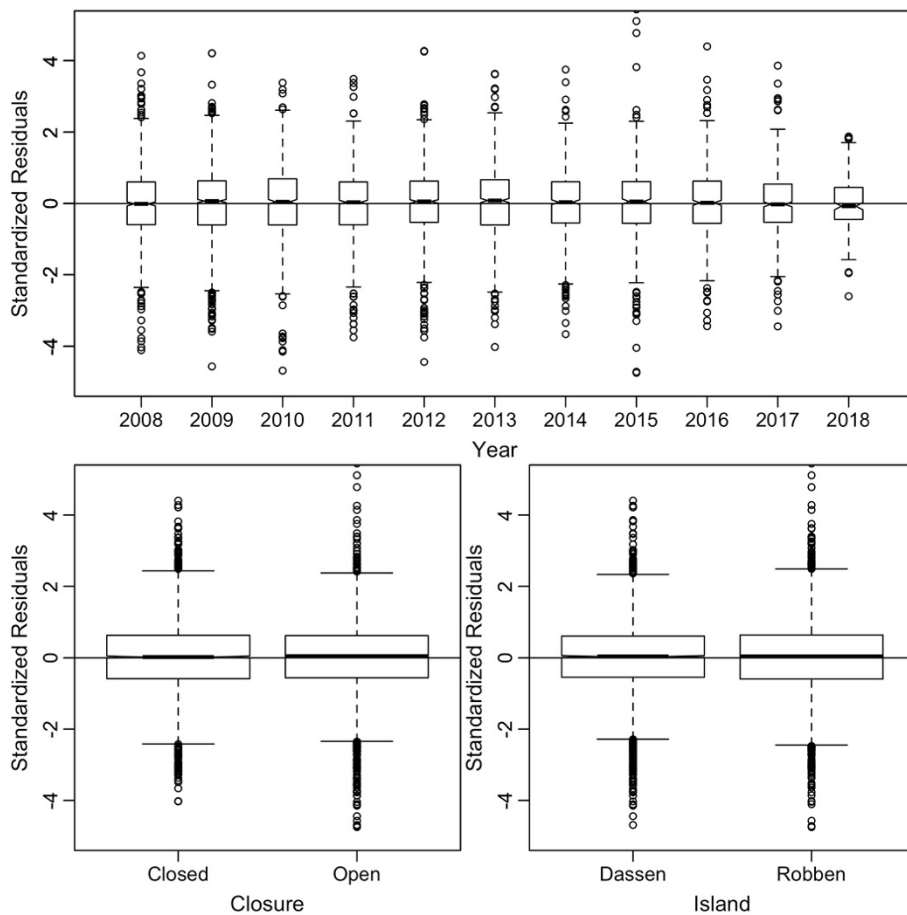
Figure A4.4. Additional model validation plots for the best fitting model (M1, Table 1) assessing the effect of the fishing closure on chick condition at Dassen Island and Robben Island. Standardized residuals are plotted against Year (top, 2008 to 2018), Closure status (bottom left, "Closed" or "Open" to fishing) and Island (bottom right, Dassen Island and Robben Island).

19
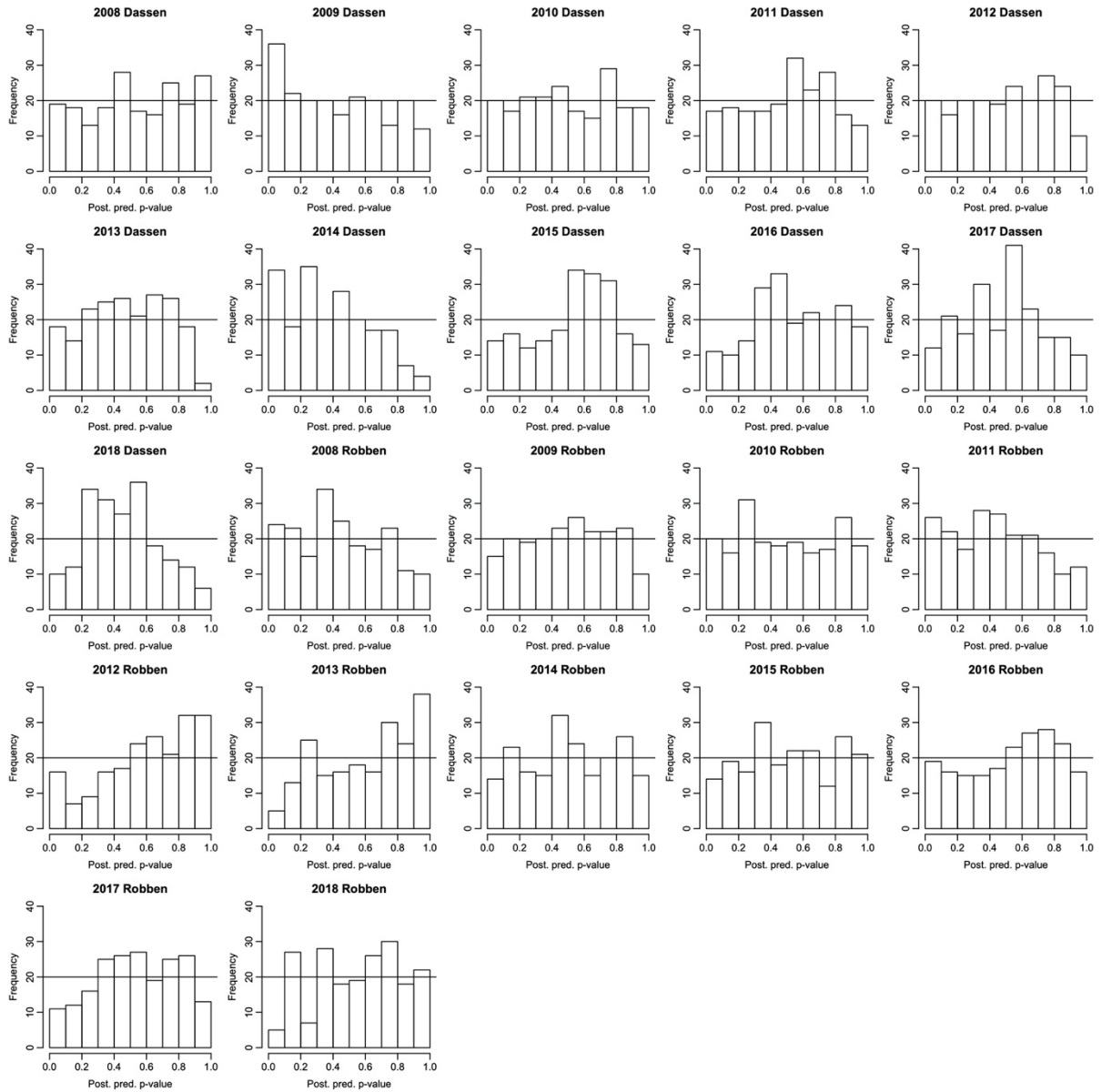
Figure A4.5. Posterior predictive p-values (based on 200 samples for each Island-Year case) from the best fitting model (M5, Table 1) for penguin chick condition by year and island in the Western Cape, 2008–2018. The horizontal line shows the expected frequency based on a uniform distribution of posterior predictive p-values.

Maximum distance travelled, Eastern Cape
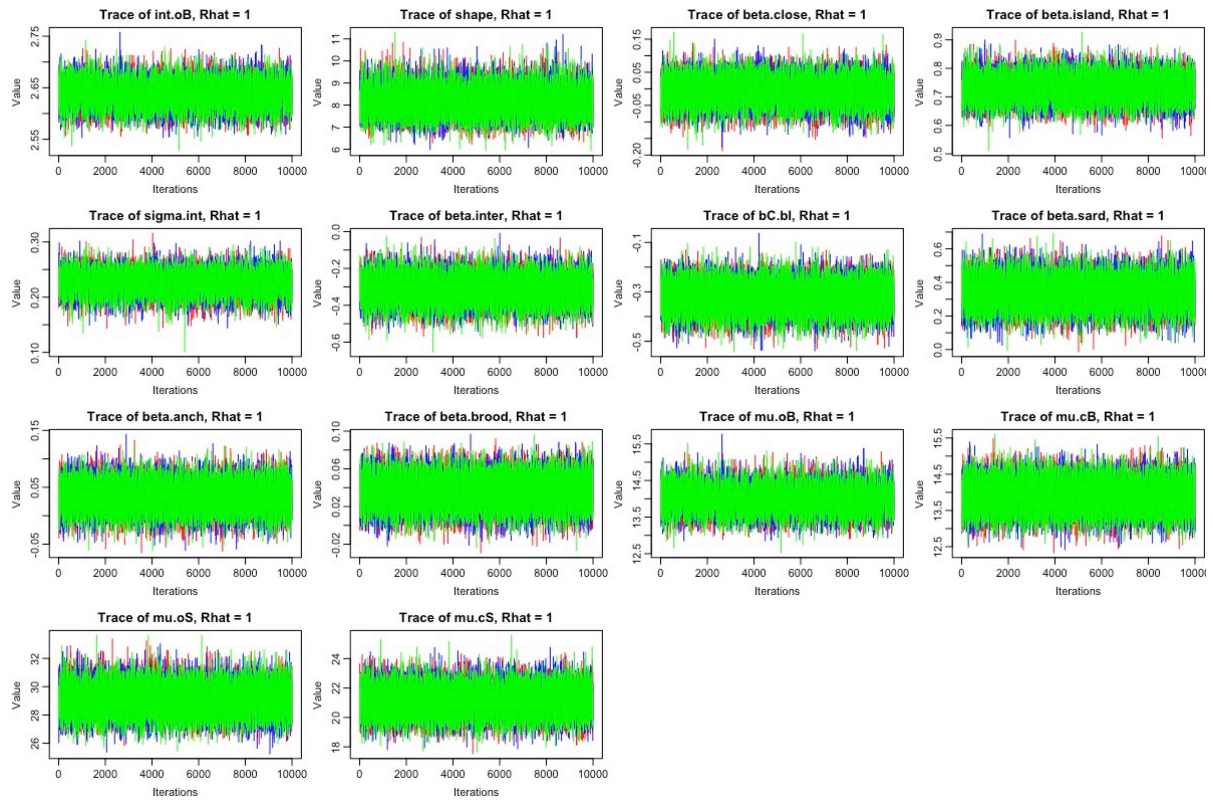
Figure A4.6. Trace plots used to assess convergence of the three MCMC chains used to estimate parameters in the best fitting Eastern Cape Chick Maximum Distance model (M6, Table 2).
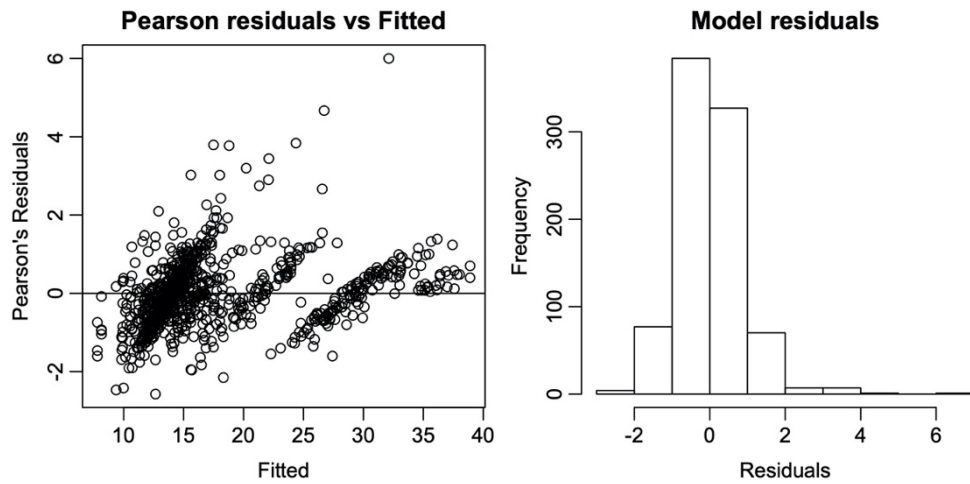
Figure A4.7. Pearson's residuals versus fitted (left) and histogram of Pearson residuals (right) for the best fitting generalised linear mixed model (Gamma errors, log link, M6, Table 2) used to assess the effect of the fishing closures on maximum distance travelled at Bird Island and St. Croix Island
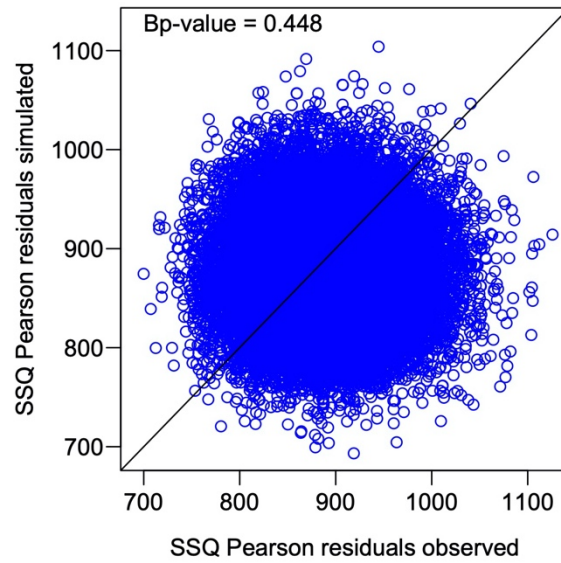
Figure A4.8. Posterior predictive check of the model adequacy for the best fitting model (M6, Table 2) assessing the effect of the fishing closure on Maximum Foraging Distance at Bird Island and St. Croix Island, plotting simulated versus observed sums-of-squares discrepancies based on Pearson residuals. The Bayesian p-value = 0.448.
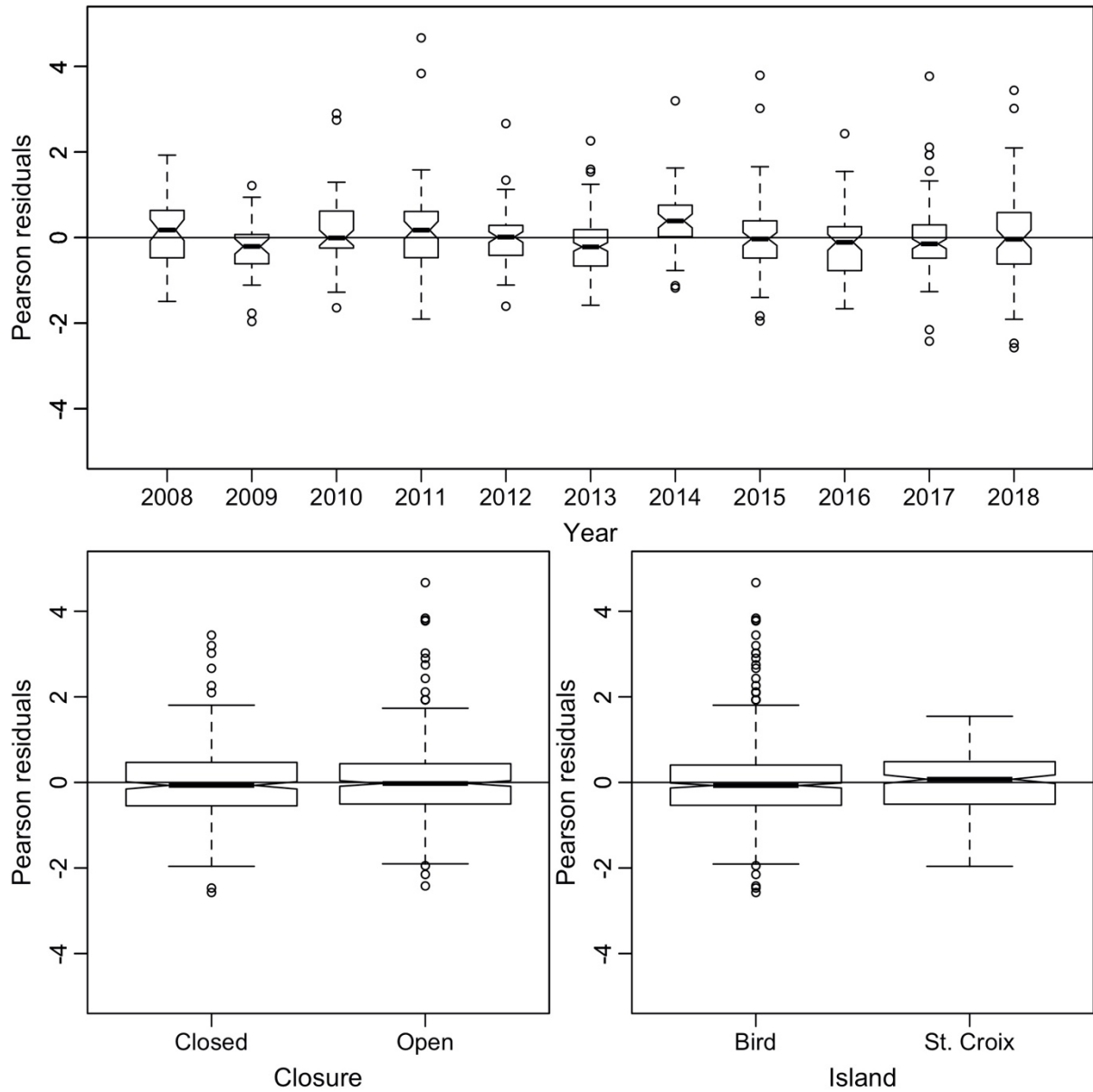
Figure A4.9. Additional model validation plots for the best fitting model (M6, Table 2) assessing the effect of the fishing closure on Maximum Foraging Distance at Bird Island and St. Croix Island. Pearson residuals are plotted against Year (top, 2008 to 2018), Closure status (bottom left, "Closed" or "Open" to fishing) and Island (bottom right, Dassen Island and Robben Island).
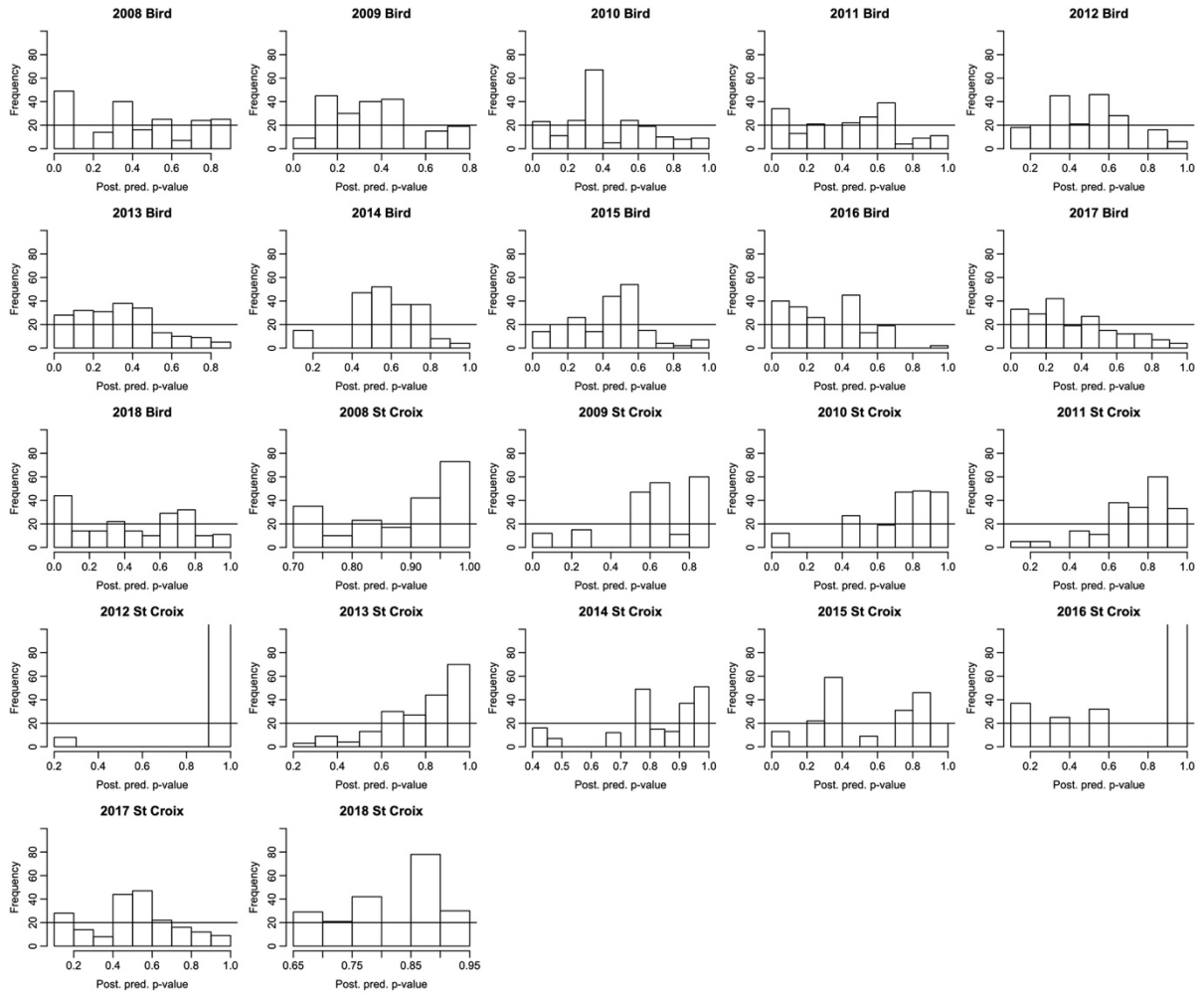
Figure A4.10. Posterior predictive p-values (based on 200 samples for each Island-Year case) from the best fitting model (M6, Table 2) for Maximum Foraging Distance by year and island in the Eastern Cape, 2008–2018. The horizontal line shows the expected frequency based on a uniform distribution of posterior predictive p-values.
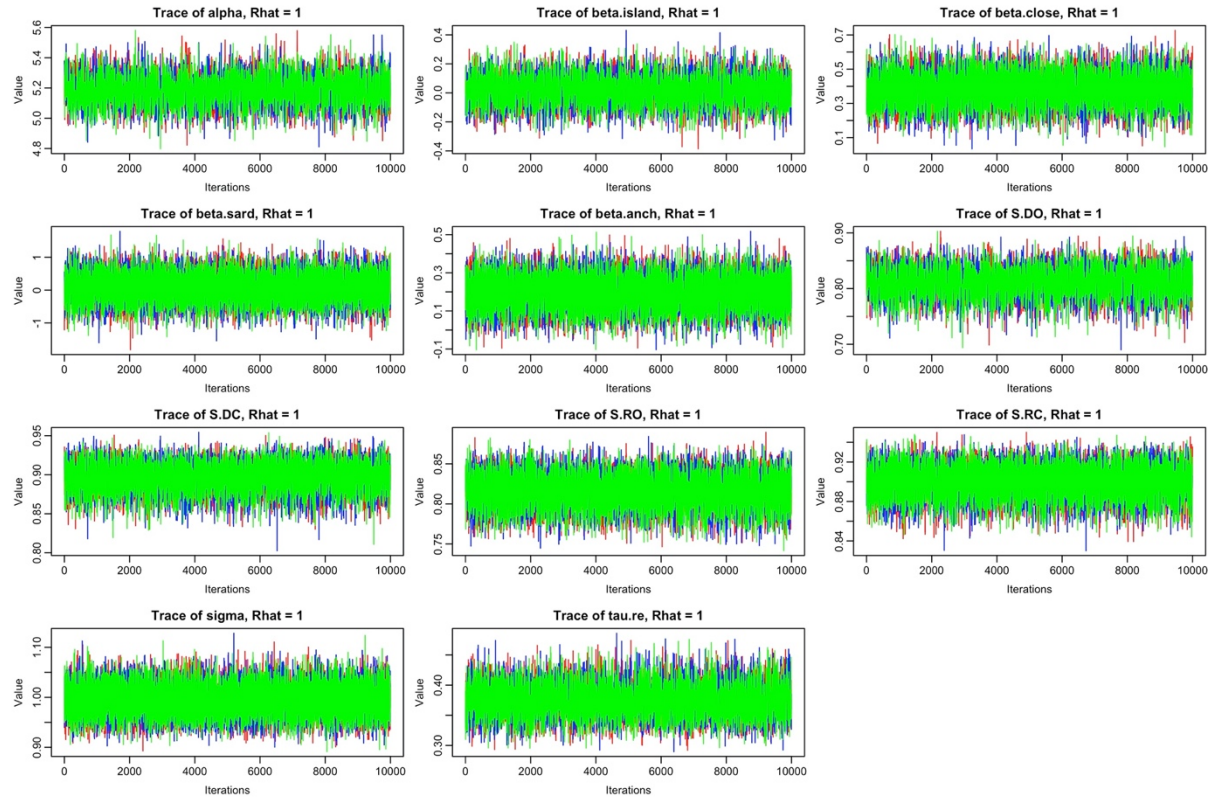
Chick Survival, Western Cape

Figure A4.11. Trace plots used to assess convergence of the three MCMC chains used to estimate parameters in the best fitting Western Cape Chick Survival model (M1, Table 3).
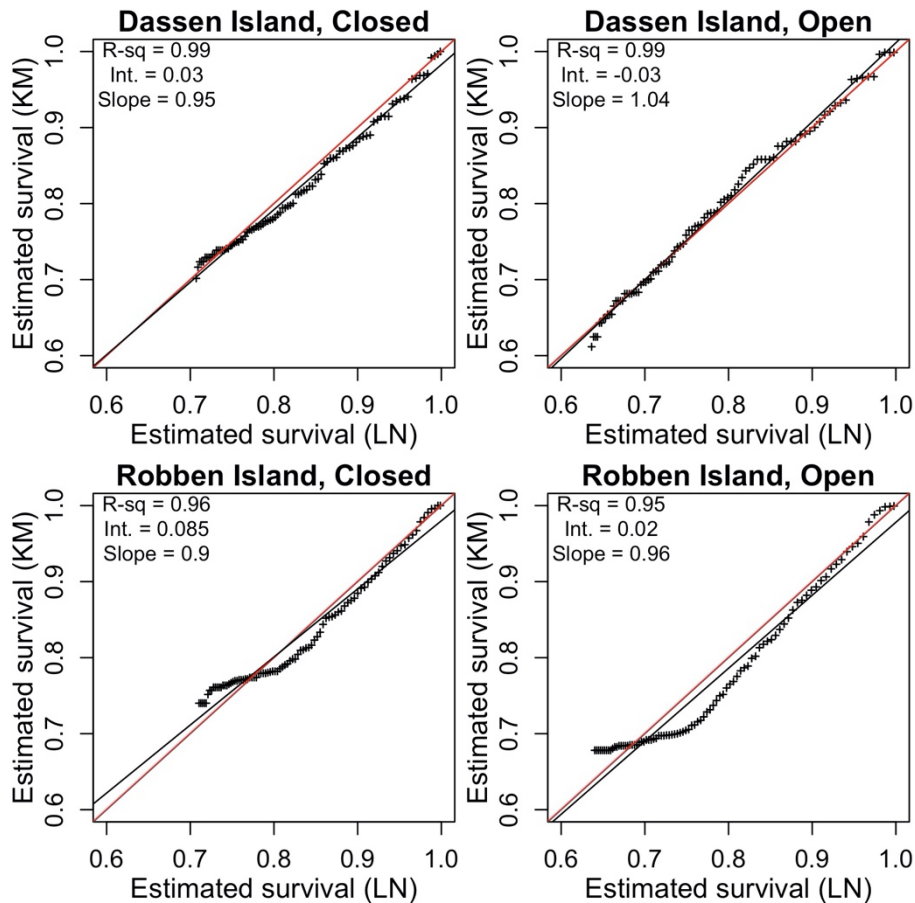
25

Figure A4.12. Probability (P-P) plots for Chick Survival at Dassen Island (top) and Robben Island (bottom) during years that were Closed (left) and Open (right) to fishing. Black points (+) show the relationship between the estimated survival probability ($\hat{S}(t)$) at $t = 1, ..., 90$ days based on the non-parametric Kaplan-Meier (KM) model and the estimated survival probability based on the parametric log-normal (LN) model ($S(t; \hat{\theta}_n)$), where $n =$ the number of chicks sampled at each island in each closure regime. Both the Kaplan-Meier and log-normal model used here for the validation plots had no frailty term included. If the specified hazard function fits well to the underlying mortality, the plotted curves should approximate a straight line in these plots. The black lines show linear regression fits to each data subset, with the adjusted R^2 (R-sq), intercept (Int.) and slope estimates of the model shown in the top-left hand corner of each panel. The red lines show perfect 1-1 fits (intercept = 0, slope = 1) for reference.
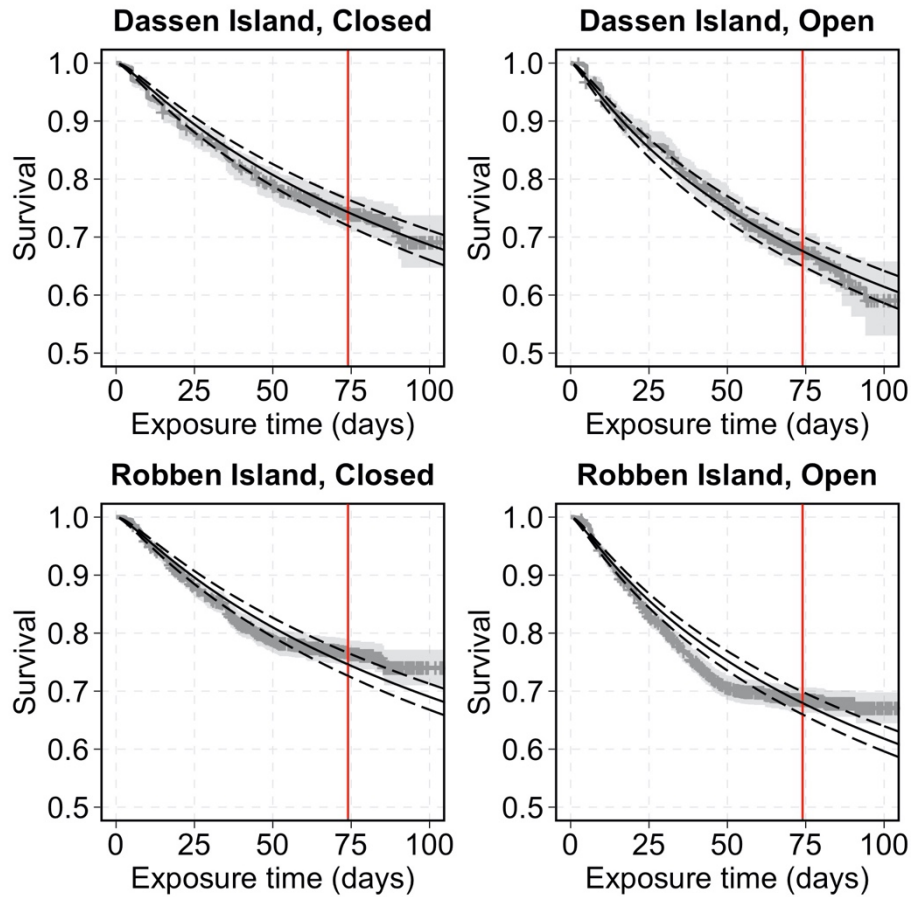
Figure A4.13. Model validation plots for Chick Survival at Dassen Island (top) and Robben Island (bottom) during years that were Closed (left) and Open (right) to fishing. Panels show the comparison of the non-parametric Kaplan-Meier (KM) estimate of survival (grey points, +) and its 95% confidence intervals (grey polygons) and the predicted survival rates (solid black curves) and 95% credible intervals (black dashed curves) based on a model with a log-normal hazard function and no shared frailty term. The vertical red line marks time = 74 days, the age at which the predicted chick survival is compared between islands and closure statuses in the results section of this document and elsewhere (Sherley et al. 2013, 2015, 2018, 2019). Crucially, the predictions from the log-normal model and the KM estimate (which is derived only from the observations) are not credibly different at 74 days, which indicates adequate model fit to predict chick survival at time = 74 days.