INTERNATIONAL REVIEW PANEL REPORT FOR THE 2016 INTERNATIONAL FISHERIES STOCK ASSESSMENT WORKSHOP

28 November – 2 December 2016, UCT

[A Dunn[1], M Haddon[2], A M Parma[3], A E Punt[4]]

Introduction

The Panel recognised the very high quality of the research presented at the 2016 International Fisheries Stock Assessment Review Workshop. This included research on Southern African hake and sardine, and whether it is possible to detect the effects of small pelagic fishery closures around islands on indicators of penguin population growth rate. The Panel thanked the workshop participants for their hard work preparing and presenting the workshop papers, for the extra analyses undertaken during the workshop, and for the informative input provided during discussions.

This report starts with observations from the Panel on some general issues for the species / programmes reviewed, and then focuses on addressing questions posed by DAFF (MARAM/IWS/DEC16/General/5), a more detailed technical review and finally recommendations concerning each. The recommendations are annotated by their priorities (H, M, L).

Summary of general issues

Hake

The Panel reviewed the development of models that account for spatial structure and inter-specific predation and cannibalism. The output from the GeoPop model can be used to inform development of spatial models for *M. capensis* and *M. paradoxus* off South Africa. Development of models for the entire *M. capensis* and *M. paradoxus* resources should consider hake in Namibia as well as South Africa. Unfortunately, to date this has proved to be infeasible owing to a lack of data for Namibia being shared. The Panel strongly recommends that efforts be made to allow assessment analysts to have access to all data from the entire southern African region to maximize the opportunities for progress on models that use all of the available information.

The development of the predation model for the Cape hakes is a major and challenging scientific endeavour. The 2016 IWS reviewed two potential predation models. The two models provide substantially different estimates for the current status of the *M. paradoxus* resource and the reasons for this need to be understood before a predation model can be used for management purposes. The Panel consequently identified several areas for additional investigation that should help to better understand the behaviour of the two predation models.

Sardine

The Operational Management Procedure (OMP) for the South African sardine will be revised during 2017. The Panel reviewed aspects of how candidate OMPs will be evaluated, with a focus on the operating models that represent the hypotheses regarding the population dynamics and stock structure of sardine off South Africa.

[1] MPI, Wellington, New Zealand
[2] CSIRO Oceans and Atmosphere Flagship, Australia
[3] Centro Nacional Patagonico, Chubut, Argentina
[4] University of Washington, Seattle, USA

Penguins

Detection of closure effects

There has been a declining trend in the numbers of African penguins at many colonies. An ongoing question is whether pelagic fishing near islands impacts penguin population growth rates negatively. The work conducted during 2016 focused on implementing the recommendations related to evaluating the statistical power to detect biologically meaningful impacts caused by the fishery that were developed by the 2015 International Workshop. The work was led by the Penguin Task Team (Mike Bergh, Doug Butterworth, Kevern Cochrane [chair], Taryn Morris, Richard Sherley and Henning Winker). Technical support for the Task Team was provided by Andrea Ross-Gillespie, without whom little progress would have been made. The Panel noted that it had not proved possible to fully implement the recommendations related to evaluating power made by the 2015 IWS. However, sufficient progress had been made that it should be possible to identify for which combinations of response variable and island it is possible to conclude there is a fishery effect, for which there is no fishery effect, and for which neither conclusion can as yet be reached. The power analysis should also assist management to identify the response variable and island combinations for which no conclusions could be drawn even given continued collection of the data concerned over 20 more years. The Panel developed an algorithm for synthesizing the results of the experiment conducted to date.

Penguin pressure model

The power analysis and associated island closure experiment cannot identify if the *primary* cause for the decline in the numbers of African penguins is due to fishing near offshore islands. The Panel recalled that several previous IWS Panels had highlighted the need to develop and implement a comprehensive research program that quantifies the core reasons for the reduction in penguin population numbers, and identifies potential mitigation measures.

The penguin pressure model provides a way to explore the implications of a large range of potential factors by combining a demographic simulation with the modelling of multiple pressures. It includes food availability and food competition by commercial fisheries, oil spills, predation by terrestrial and marine predators, and extreme climate events. While it can explore the implications of more potential factors than the model of Robinson et al. (2015), the conclusions that can be drawn from the penguin pressure model can only be considered in a strategic rather than a tactical sense.

The penguin pressure model may be improved by fitting it to the available observational data in an integrated framework, and using this to identify the sets of parameters that are able to mimic the available data. In its current version, components of the model have been fitted piecewise to available data, but no integrated statistical framework has been applied. A result is that the overall trend in penguin abundance indices is not adequately predicted by the model. Once fitted to the available data, the outputs of the penguin pressure model could be used to develop hypotheses on how different factors and processes may have impacted the state of the penguin population, and the uncertainties associated with such inferences, and to explore how different management options could be used to halt the population decline. When used in this context, the penguin pressure model would constitute an important tool in assisting with research planning, development of monitoring strategies, and investigating potential management interventions.

General considerations

The Panel once again reiterated the recommendation from past panels that for each species being reviewed, a summary document should be produced that provides an overview of the fishery, its history of exploitation, and briefly describes the data available. This “fishery description” document should be provided to the Panel well in advance of the review meeting, as it would help Panel members unfamiliar with South African fisheries and fisheries management techniques to become better prepared for the review. The work of the Panel was made more difficult this year by the large number of documents that were made available only just before the start of the workshop.

A. Hake

A.1 Assessment

A.1.1 Advise on the reliability of the medium term projections presented for catches, resource depletion, CPUE and fishing effort in the context of long-term planning for the industry.

The estimated status of the *M. paradoxus* stock is now less optimistic regarding current biomass relative to MSYL than the assessment on which OMP-2014 was based (MARAM/IWS/DEC16/Hake_Assess/P1). With a view towards reducing workload, the Panel had not been provided with the specifications for the new assessment so did not review the technical basis for the medium-term projections and hence cannot comment on whether the change in status is justified. The Panel was informed that the technical basis for the new projections will be evaluated by the Demersal Working Group. Nevertheless, the projections appear to have followed the process used for past projections that have formed the basis for OMP evaluations.

A.1.2 Comment on the suggestion that OMP revision be brought one year forward to allow for a possible TAC decrease in excess of 5% in 2018 in the interests of better catch and resource performance over the next 10 years

MARAM/IWS/DEC16/Hake_Assess/P2 provides projection results for OMP variants in which the TAC output from OMP-2014 is reduced by a further 5,000t (OMP-2015_5) or 10,000t (OMP-2015_10). These OMP variants lead to slightly faster recovery to MSYL, but only by one year in median terms. This is partially due to OMP-2015_5 and OMP-2015_10 leading to higher TACs after about 2023. It may therefore be possible to achieve faster rates of recovery to MSYL of *M. paradoxus* for some OMP variants, but account needs to be taken of the trade-off between the rate of recovery and the catch in the short-term and over the entire projection period. Without the ability to account for such a trade-off, bringing the OMP revision forward by a year is not likely to lead to important benefits

A.1.3 Prioritised research recommendations

A.1.3.1 (H). A “historical” analysis should be undertaken in which the results (time trajectories of biomass in absolute terms and relative to MSYL) are plotted for the reference sets of operating models proposed for the next update to the hake OMP as well as those on which OMP-2014 and earlier OMPs were based. Such “historical” analyses provide a way to understand the changes in estimated stock status due to changes to the assumptions on which the operating models are based, rather than changes due, for example, to estimates of incoming recruitment as further resource monitoring data become available.

A.2 Modelling predation

A.2.1 Review the two hake predation models presented, with a particular view towards identifying why (or suggesting approaches to determine why) they provide appreciably different estimates of current depletion (relative to pristine) for the *M. paradoxus* population.

The differences between the two hake predation models are difficult to evaluate because the models differ in a large number of respects. In particular, the models differ substantially in estimates of absolute biomass, especially for *M. paradoxus*, even when there is no predation. Thus, the comparison of the models should proceed by first selecting as many common assumptions as possible (minimally the no predation models should be based on the same reference case specifications). Only once the differences between the two predation models when there is no predation are understood (or eliminated), should attempts be made to compare "with predation" variants of those models. The comparison of the models "with predation" should start with the same or similar initial conditions (for 1916) because these can influence the subsequent dynamics substantially. Analyses in which the basal natural mortality rate for animals of age 15+ are set to same value for the two predation models may also assist these comparisons.

A.2.2 Prioritised recommendations

A.2.2.1 (H). The hake predation model in MARAM/IWS/DEC16/Hake_Pred/P2 includes an upper bound on predation mortality. The predicted diets should be constrained if predation mortality is constrained, but this is not the case at present. The "switch" behaviour evident for this model may be due to the complexity of the likelihood surface, and the Panel recommends that this be explored further using 'jittering'. Care should be taken that predation mortality does not hit its upper bound in the first year of the model (1916).

A.2.2.2 (H). The hake predation model in MARAM/IWS/DEC16/Hake_Pred/P1 should be extended to: (a) allow for time-varying amounts of hake in the diets of hake predators, (b) consider a constraint that the basal natural mortality rate decreases with age, and (c) consider iterating the calculation of natural mortality to avoid assuming that the total mortality rate for year y is the same as that for year y-1 when computing predation mortality (or consider shorter time-steps).

A.2.2.3 (H). The proportion of the consumption of hake by prey age-class for age 15+ hake in MARAM/IWS/DEC16/Hake_Pred/P2 is very high compared to that of age 14 hake. One reason for this is that there are more 15+ than 14 hake, but the full set of reasons needs to be explored.

A.2.2.4 (H). Plot the time-trajectories of the proportions of hake in the diets of predators by hake species and age-class. This information may assist in understanding the reasons for the differences between the two models.

A.2.2.5 (H). The preference function should be a function of length. However, this would complicate fitting the model to the diet data (currently predators and prey categorized by length) because it would be necessary to convert predator and prey numbers from age to length using age-length keys and this would lead to high variability in model predictions. MARAM/IWS/DEC16/Hake_Pred/P2 overcomes this problem by explicitly modelling populations by length as well as age. This problem could also be overcome by basing the diet data on predator ages rather than predator lengths so it only necessary to convert prey ages to prey lengths when constructing the likelihood function.

A.2.2.6 (M). Consider a sensitivity test in which the basal natural mortality rates are higher at low and old ages than at intermediate ages. The support for such a sensitivity test could be explored by examining age-composition information for longline-caught fish, ideally those caught during the early years of the longline fishery when untrawlable grounds were first intensively fished.

A.2.2.7 (M). The diet data for the west and south coast should be combined. This can be achieved by weighting the diet data by coast by a measure of the numerical abundance by age-class on the west and south coasts, e.g. from the results of the surveys. Alternatively, a spatial predation model could be developed.

A.2.2.8 (M). The Panel reiterates the recommendations from the 2013 and 2014 Panels, which are important to implement.

(a) Scale hake prey-by-species information upwards to account for unidentified hake prey. This applies primarily to *M. capensis* predators because they consume both hake species. However, ignoring unidentified hake will underestimate the proportion of hake in the diet of *M. paradoxus*.

(b) Plan, and then implement, a review of the sampling strategy for diet data given the results of the current model as well as other needs for diet data. This recommendation pertains to DAFF.

A.3 GeoPop

A.3.1 Provide brief comments on the papers presented applying the GeoPop approach to hake

MARAM/IWS/DEC16/HakeGeoPop/P1 and MARAM/IWS/DEC16/HakeGeoPop/P2 summarize the application of the GeoPop model to survey data for Namibia and South Africa. A key conclusion from MARAM/IWS/DEC16/HakeGeoPop/P2 is that there is one main nursery ground for *M. paradoxus* on the West Coast and a minor nursery ground on the South Coast, while MARAM/IWS/DEC16/HakeGeoPop/P1 identified three potential stocks of *M. capensis* off Namibia and South Africa. The results from fine-scale spatial models such as GeoPop can be used to provide parameter values and data for spatial stock assessment models for hake, ideally if they produce estimates of movement rates by age-class. However, the current version of GeoPop can provide only estimates of the relative abundance of age-classes spatially and not the rates of movement between regions that lead to that non-homogeneous spatial abundance. The results from GeoPop could be used and included in spatial modelling for hake in the short-term, as follows.

(a) Compare the spatial distributions by age from GeoPop with the raw survey data at the level of the spatial cells used in the spatial model (Rademeyer, 2013, 2014) to assess the extent to which GeoPop mimics the actual data.

(b) Compare the outputs from the spatial model with those from GeoPop to assess whether the spatial model is able to replicate the patterns in distribution even without formally including the GeoPop results into the likelihood function (or in the form of penalty terms) for the spatial model.

(c) Include the spatial distribution information from GeoPop into the likelihood function of the spatial model; it will then be necessary to downweight the spatial (but not trend) information from the survey data when formulating the likelihood function.

The Panel endorses the future work outlined in MARAM/IWS/DEC16/HakeGeoPop/P2, particularly (a) modelling growth and fishery removals more realistically, and (b) formulating GeoPop so that it outputs the probability of fish of a given age moving from each of the spatial cells in the spatial model to each of the other spatial cells. The latter may require formulating GeoPop as a dynamic rather than an equilibrium model.

B. Sardine

B.1 Stock structure

B.1.1 On the basis of the current preferred hypothesis (MARAM/IWS/DEC16/Sardine/P7), advise on how alternative hypotheses about the contribution of south coast spawning biomass to west coast recruitment are best constructed and weighted, taking account of: a) approaches involving fits to stock recruitment relationships; and b) ouput from the hydrodynamic individual–based model (IBM), and how the associated uncertainty is best quantified?

Several papers (e.g. MARAM/IWS/DEC16/Sardine/P6, P9, P10, and P11) provided ways to estimate the contribution of south coast spawning to west coast recruitment. The methods on which these papers were based differed, but all attempted to find the proportional contribution of south coast spawning to west coast recruitment (p) that leads to the best fits to the stock and recruitment data. These papers all concluded that the best estimate of p was likely to be large (close to 100% of south coast spawning contributing to recruitment on the west coast). However, the information in the data to support any particular value for p was very weak, highlighting the importance of the prior assigned to alternative values for p.

The Panel notes that the current hydrodynamic model that could be used to define a prior for p is based on several assumptions that need further exploration. The Panel supports the following proposed additional work on this hydrodynamic model:

- repeat the sardine IBM computations using the new 3-d hydrodynamic model (larger spatial coverage and nested models; 1/15° resolution and 100 vertical levels; 6-hourly wind, heat and salinity forcing);
- release particles during peak spawning periods (as inferred from GSI data) for fish off the west and south coasts; and
- include larval behavior (e.g. diurnal vertical movement (see Parada et al., 2008) or directed swimming, etc); larval feeding; and larval predation.

The Panel specifically highlights the importance of adding more biological realism to the IBM, for example in the form of accounting for diurnal vertical movement of larvae. The Panel notes further that quantifying some of the factors that might impact transportation rates such as feeding and predation rates has proved difficult in other regions, and hence difficult to include in IBMs.

The Panel was advised that the hydrodynamic model was unlikely to be updated before when its output would be needed to provide a prior to weighting values of p. Hence, in the short-term, the outputs by month from the current hydrodynamic model should be weighted by GSI values by month as this should (to a first approximation) weight estimates of transportation rates by the relative amount of spawning.

B.1.2 Advise whether other stock structure hypotheses need to be considered, taking account of the implications of the initial genetics results received very recently.

The stock structure hypothesis on which the assessment and OMP evaluation is proposed to be based involves a single stock (in the sense of a reproductively isolated unit) with two "components" (MARAM/IWS/DEC16/Sardine/P7). The genetic study currently underway, of which the Panel have only seen a presentation, suggests that there are three 'units' off South Africa inferred from analysis of genes under thermal selection, but that all three units are found off the west coast, off the south coast, and off the east coast, with no obvious link between unit and coast. This suggests there is no appreciable reproductive isolation among sardine across all three coasts, but that there appears to be a longshore cline for genes under thermal selective pressure. The Panel concluded that there is no need to change the current stock structure hypothesis.

B.1.3 Comment broadly on the implications of these various hypotheses concerning the need or otherwise for spatial management (i.e. for a directed sardine TAC split by area).

There is a need to consider OMP variants that include spatial management considerations given the stock structure hypothesis includes spatial components. The results of projections under OMP variants will likely depend on the contribution of south coast spawning to west coast recruitment (p), as well as other spatial aspects of the operating model such as the extent of west to south coast movement of age 1+ animals. The Panel does not recommend 'integrating' results across values for parameters that could have a substantial impact on the performance statistics (such as p). Such qualitative differences imply rather that the performance statistics for each such operating model should be considered separately.

The Panel recommends that OMP variants be developed that attempt to be 'adaptive' and hence able to respond to the further information that future data will provide towards resolving stock structure uncertainty. An example of such an OMP would involve modifying the ratio of the proportion of the catch taken on the west vs south coast based on spatial trends in recruitment estimates.

B.1.4 Prioritised recommendations.

B.1.4 (H). Analyses that estimate the contribution of south coast spawning to west coast recruitment should plot the observed and model-predicted time-series of recruitment because the analyses in MARAM/IWS/DEC16/Sardine/P11 suggest that even the best estimate for the parameter determining this contribution may not correspond to particularly good fits.

B.2 Proposed projection framework for OMP testing and telated matters

B.2.1 Advise on which of the various future movement hypotheses that have been put forward (e.g. MoveR, MoveB) need be considered further, and on their relative plausibilities/weighting.

MARAM/IWS/DEC16/Sardine/P12 outlines the alternatives to model movement of age 1+ animals from the west to the south coast. The Panel agrees that the MoveR hypothesis (implemented as the proportion of 1 year olds that move being drawn randomly from the proportions estimated by the model for 2006-2015) be considered further. The hypothesis may not lead to "regime-like" trends on the south coast. Thus, the data on sediment core samples for Namibia should be explored to assess whether they provide information on inter-decadal variability of biomass. A scenario with regime-shifts in movement should be developed to match such variation if such variation is detected.

The Panel was concerned that the relationship identified for the MoveB hypothesis (proportion of 1 year olds that move in November of year y is assumed to be related to the ratio of south to west biomass in November year y-1) is unlikely to reflect a causal relationship, but instead may result from higher south coast biomass arising through higher movement rates. However, the outputs for the projections based on the MoveR hypothesis should be plotted to check that the resulting patterns match those in Fig. 6 of MARAM/IWS/DEC16/Sardine/P12, because the pattern in Fig. 6 should be an emergent property of the model. If neither the MoveR hypothesis nor a hypothesis based on regime shifts in movement of 1+ animals from the west to the south coast, is able to produce a pattern similar to that of Figure 6 of MARAM/IWS/DEC16/Sardine/P12, it may be necessary to develop a variant of the MoveB hypothesis in which the proportion of 1 year olds that move in November of year *y* has a dome-shaped relationship with the ratio of south to west biomass in November year *y*-1.

In summary, the MoveR hypothesis should form the basis for OMP evaluations unless the analyses of the sediment core samples for Namibia suggest the need for a regime-shift model for movement or the outputs from projections using the MoveR hypothesis do not match the pattern evident in Fig. 6 of MARAM/IWS/DEC16/Sardine/P12.

B.2.2 Comment on how future recruitment is best generated for projections, taking into consideration: (a) which stock recruitment relationship(s) to use; (b) whether to fit them internally or externally to the assessment, and in the latter case how best to proceed in a Bayesian (MCMC) context; and (c) how best to generate residuals about that relationship.

(a) The reference case analysis should consider (roughly) three variants of the operating model of MARAM/IWS/DEC16/Sardine/P2 (i.e. different values for *p*) in which recruitment is related to spawning stock biomass according to a Hockey-Stick stock-recruitment relationship. The extent of variability in recruitment, σ_R, for each draw from the posterior should be set to 0.5, while the extent of auto-correlation in recruitment for each draw should be calculated by fitting an AR-1 model to the residuals by draw.

(b) The robustness tests that involve changing the form of the stock-recruitment relationship (Beverton-Holt, Generalized Ricker, and Shepherd should be considered) should be based on fitting an operating model with no stock-recruitment relationship (with log-recruitments being estimated rather than deviations about the stock-recruitment relationship). The parameters of the stock-recruitment relationship for each draw from the posterior should be based on either (i) fitting the stock-recruitment relationship to the estimates of stock and recruitment and taking the best estimates, (ii) fitting the stock-recruitment relationship to the estimates of stock and recruitment and generating values for the stock-recruitment relationship parameters from a multivariate normal distribution defined by inverting the Hessian matrix, or (iii) applying MCMC to the stock and recruitment data for each draw. Approach (iii) is much more computationally intensive than approaches (i) and (ii). Which approach is to be used should be determined for one of the reference case models (and some values for *p* / OMP variants) by comparing risk statistics among the approaches to determine whether one of the simpler approaches performs adequately. This would involve conducting projections based on the MCMC values for the stock-recruitment relationship parameters (to mimic approach (iii) and hence the ideal approach), and then using approaches (i) and (ii) to see if the risk statistics do not change appreciably, thus supporting the use of a simpler approach.

B.2.3 Provide comments on other aspects of the specifications proposed for OMP testing.

The Panel was informed that industry was interested that OMP variants be examined that consider moving the boundary between the south and west management areas. The Panel recommends:

1) the operating models should be based on an assumed "true" stock boundary at Cape Agulhas to reduce computational demands associated with conditioning operating models;
2) the analysts should work with the industry to identify how the spatial pattern of catches would likely change given a change to the boundary; and

3) it will be necessary to modify how abundance estimates are generated if the boundary is moved by a substantial extent from the line at Cape Agulhas.

The OMP testing should consider alternative values for the proportional contribution of south coast spawning to west coast recruitment (*p*). The Panel recommends considering values for *p* of 0 (no contribution of south coast spawning to west coast recruitment), a value based on the hydrodynamic model, and a larger value given that there are several sources of uncertainty that are not included in the hydrodynamic model. The Panel could not identify an objective basis for assigning weights to alternative values for *p*. Consequently, the Panel recommends that results be shown for each individual choice for *p*.

B.2.4 Advise on an appropriate risk specification, noting the comparison available of the productivity of South African sardine relative to other sardine stocks worldwide, and advising both how this might best be taken into account and how that analysis might be improved.

MARAM/IWS/DEC16/Sardine/P13 provided information on the proportion of sustainable yield to spawning biomass for sardine stocks for which data were available. However, the underlying data appear to be of questionable accuracy in some cases. Moreover, most sardine stocks (including that off South Africa) are not managed on the basis of a constant exploitation rate strategy. A potentially more informative question to answer would be how one might make the risks comparable among the different harvest strategies used in each example case. One possibility would be to compare their relative performance in terms of the particular fisheries' response to application of the harvest strategy in use. Possible measures of that performance could be how often the stock declines to low levels (pre-defined limits). It might also be useful to determine whether or not, in each example case, fishing in fact can have a meaningful influence on each stock.

The evaluation of OMP variants relies on the risk criterion that is used when tuning OMP variants. The probability of dropping below the current threshold used to compute risk (the average biomass from 1991 to 1994) is substantially higher under the current assessment than that on which OMP-14 was based, in particular in the absence of catches. It is necessary to fully understand why the change in the estimate of the average biomass under zero fishing ("K") and in risk has occurred. Reasons to explore include how maturity in the model is defined, but whether this is the key reason is unclear. Options to consider if the risk criterion is to be changed include defining the threshold as some proportion of the biomass at which expected recruitment declines and some proportion of K (as estimated for one form of stock-recruitment model).

Risk criteria have not been specified for the two component operating model. The Panel recommends that risk criteria be selected for each of the south and west coasts, but that the probability of being below the threshold need not be the same for the two coasts.

B.2.5 Prioritised recommendations

B.2.5.1 (M) Create plots of the proportion of each sampled fish according to its genetic unit after arranging the samples by longshore location, using the genetics data by season to assess whether there is more evidence for population structure, for example temporally.

C. Penguins

C.1 Island closure analysis

C.1.1. Do the analyses reported in direct response to recommendation A.2 from last year constitute an acceptable implementation of that recommendation for each of the response variables considered?

Recommendation A.2 from the 2015 International Workshop Panel outlined a set of 22 steps to follow to evaluate closure effects on penguins. MARAM/IWS/DEC16/Peng/BG2 provides a summary of the extent to which the 22 steps have been followed and MARAM/IWS/DEC16/Peng_Clos/P1a reports on the results of the analyses conducted to date. Results are reported for a subset of the analyses requested given the considerable time required to conduct the analyses. Two of the 2015 Panelists (Ana Parma and André Punt) were consulted inter-sessionally to clarify the details of the algorithm, which were not fully specified in the report of the 2015 IWS meeting. The Panel notes that a substantial amount of work has been undertaken and commends the Penguin Task Team and specifically Dr Andrea Ross Gillespie for their efforts to quantify the probability of concluding whether results under the closures will have sufficient power to detect within 20 years if fishing near islands will have a negative impact on penguin populations. It agrees that the calculations conducted follow the spirit of Recommendation A.2. Thus, while some additional analyses could be undertaken, the Panel concludes that the work undertaken satisfies the intent of Recommendation A.2.

C.1.2. Given the results from those analyses, advise to what extent the points you raised in recommendation A.3 might now be addressed; this is to assist the pertinent DAFF scientific working group which is required to make a recommendation on continuation of or adjustment to the current island closure programme at a meeting to be held on 7 December.

The Panel noted that of the five steps identified under Recommendation A.3 from the 2015 IWS, all but step 4 ("Use the results of the power analysis to assess whether there are values for λ (or δ) that are no longer plausible given current data (i.e., as the power to detect them, given the current stage of the experiment, is already very high.)") were completed.

The Panel identified an algorithm for interpreting the results of the experiment (Fig. 1).

1. Compute the (cumulative) probability that δ is less than the threshold (this is currently -0.1), using Equation 2 of MARAM/IWS/DEC16/Peng_Clos/P1a (denoted "X"). Given the conclusion in item C.1.5.1 below, it is necessary to apply only the closure estimator.
2. If X is greater than P_{min} for the response variable concerned (Fig. 1a, right panel, assuming P_{min} is 0.84 or less), it can be concluded that there is a fishery effect because δ is less than the threshold corresponding to a demographically important impact on the penguin growth rate. Stop.
3. If X is less than P_{min},
   - 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 data, i.e. no future data are simulated) to approximate the distribution of P(δ<threshold) for each value of δ, where the distribution is integrated across the operating models;
   – for each candidate value of δ, say δ^c, evaluate whether X is less than a chosen percentile of the simulated distribution of P(δ<threshold| δ^c) and reject δ^c if it is. Identify the maximum value of δ (denoted δ_{crit}; Fig. 1b) that can be rejected by this rule. It can then be concluded that the current data allow values for δ less than

δ_{crit} to be excluded because, given the power analysis, it should have been possible (with a given probability) to detect $\delta < \delta_{crit}$ if this is actually the case.

4. 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 8 of MARAM/IWS/DEC16/Peng_Clos/P1a) should be considered to decide whether to continue the experiment or not.

The Panel noted that a threshold based on population dynamics models had been selected for one of the response variables (fledging success) while the thresholds for the other response variables had been selected by an assumption of proportionality with reproductive success. Four papers (Boresma and Rebstock. 2009; Hennicke and Culik, 2005; Horswill et al., 2014; McClung et al., 2004) were made available during the workshop that showed relationships between predator mass and survival. The Panel recommended that they be considered by the Pelagic Working Group when they attempt to synthesize the results of the analyses in MARAM/IWS/DEC16/Peng_Clos/P1a.

The Panel noted that the analyses have not attempted to integrate information across the response variables. However, while chick condition and chick growth are likely correlated, chick condition/growth and fledgling success affect processes that are sequential in the life history of penguins, which means that a fishery effect on each of chick condition/growth and fledgling success in combination could lead to a biologically meaningful population effect. Moreover, increases in forage trip length due to fishery impacts may have negative consequences for adult survival.

C.1.3 Comment on the relative merits of fisheries assessment analyses (including this penguin case) being based on aggregated data in contrast to individual data utilising mixed effects models in a one-step process, as used in the draft paper by Sherley (and any possible further associated computation results), and advise whether this might necessitate revision of the conclusions otherwise to be drawn under questions 1) and 2) above.

MARAM/IWS/DEC16/Peng_Clos/P4 (updated based on additional analyses; Table 1) suggests that analysing disaggregated data can lead to different estimates of the impact of closures on chick condition as well as to more precise estimates. Table 2 explores some of the consequences of the impact of lower standard errors for these estimates on the power analyses and found them to be small. However, the analyses on which Table 2 were based did not generate disaggregated data as this would be prohibitive computationally.

The Panel explored whether the model fitted to disaggregated data results in adequate fits as quantified using posterior predictive p-values (Fig. 2). The histograms in Fig. 2 should be uniform (or close to uniform) if the model fit is adequate, which is the case. The Panel concludes there is no evidence from the posterior predictive p-values to suggest non-independence, at least for chick condition. The use of disaggregated data in an estimator would require that steps 1-4 of Table 2 of MARAM/IWS/DEC16/Peng_Clos/P1a be followed (see item C.1.5.2).

C.1.4 Advise on the extent to which the draft paper by Sherley addresses recommendation A.5.

Recommendation A.5 from the 2015 IWS involved fitting the operating model using Bayesian rather than likelihood-based methods. MARAM/IWS/DEC16/Peng_Clos/P2 outlines how models can be fitted to the data for example cases involving two response variables. The use of Bayesian methods has the advantage of avoiding the need to specify how to handle cases in which the point estimate of σ_α is zero. It is difficult to compare the results in MARAM/IWS/DEC16/Peng_Clos/P2 with those in MARAM/IWS/DEC16/Peng_Clos/P1a owing to differences in the way the data enter the analysis (aggregated vs disaggregated) and the error model for the data (log-normal vs normal) and the fact that MARAM/IWS/DEC16/Peng_Clos/P1a treated the closure effect as multiplicative while MARAM/IWS/DEC16/Peng_Clos/P2 treated it as additive. The approach outlined in MARAM/IWS/DEC16/Peng_Clos/P2 could form the basis for evaluating power along the lines of MARAM/IWS/ DEC16/Peng_Clos/P1a, but it would require implementing the steps in Table 2 of MARAM/IWS/ DEC16/Peng_Clos/P1a except for step 1, which would be replaced by the results of Bayesian analyses.

C.1.5 Recommendations and other conclusions

C.1.5.1 (*) Table 4 in MARAM/IWS/DEC16/Peng_Clos/P1a shows the values of $P_{\min}$ for different combinations of operating models and estimators (Closure and Catch). Examination of the results indicated that the Catch estimator resulted in values of $P_{\min}$ substantially lower than 0.5 (as would be expected for an unbiased estimator) whenever the operating model involved a closure effect (by itself or in combination with a catch effect). By contrast, the Closure estimator proved to be more robust to the choice of operating model, even when a pure catch-effect operating model was used. While this bias was taken into account in the power calculations by the use of $P_{\min}$, the interpretation of point estimates and confidence bounds based on the Catch estimator would be questionable given this bias.

C.1.5.2 (*) If the work in MARAM/IWS/DEC16/Peng_Clos/P2 is to be continued based on disaggregated data, it will be necessary to implement steps 2 – 7 of Table 2 of MARAM/IWS/DEC16/Peng_Clos/P1. In addition the following is recommended.

1. The estimator should be based on a multiplicative rather than additive model (i.e. Equation 1 of MARAM/IWS/DEC16/Peng_Clos/P2) with log-normal or gamma errors (for indices that can be negative, add a constant equal to the mean of the data).
2. The estimator based on Equation 2 of MARAM/IWS/DEC16/Peng_Clos/P1a should be replaced by the Bayesian estimate of the posterior probability that $P(\lambda<T)$ or $P(\delta<T)$.
3. Step 2 involves the generation of future data. Generation of future disaggregated data should account for any overdispersion and non-independence by sampling the error according to the random-effects structure assumed in the mixed-effect model assumed.
4. Step 4 may not be necessary for a Bayesian estimation method, but this needs to be checked.
5. Step 5 should be based on the Bayesian posterior distribution for the effect size.

C.1.5.3 (*) As noted in item 1.5.1, the Panel recommends that the Closure estimator form the basis for decision making in the short-term. However, this does not preclude use of the Catch estimator in the future. In principle, the Catch estimator could be modified by accounting for the bias associated with the catch-biomass correlation, but the resulting estimator will need to be more robust than the current Closure estimator (see, for example, Table 3).

C.1.5.4 (M) Construct posterior predictive checks for the chick survival analysis in MARAM/IWS/DEC16/Peng_Clos/P2. This could not be accomplished during the current workshop as simulation of data that are censored is not straightforward.

C.2 Pressure model

C.2.1 The pressure model claims results that inform on the impact of pelagic fishing close to Robben Island on penguins. Advise on the reliability of these conclusions, and

those from the Robinson et al. model, in the context of informing the recommendation to be made on 7 December that is referenced above.

The penguin pressure model is fundamentally a simulation model, with some relationships set based on external model fits. Thus, this model is best used to evaluate broad management policies in a strategic sense. The penguin pressure model considers relationships between anchovy biomass and penguin demographic variables, for example, chick survival and juvenile survival (MARAM/IWS/ DEC16/Peng_Press/P2), which is informative about some aspects of penguin population dynamics, but may fail to represent the net effect of anchovy biomass on the period between egg laying and reaching age 2 unless such relationships are available for all aspects of the chick and juvenile phase. Moreover, the penguin pressure model is unable to fit a key data source (numbers of female moulting penguins over time) adequately. In contrast, the Robinson et al. model considers the net effect of anchovy on reproductive success and survival to the end of the first calendar year of life in combination, and the data sets used were adequate for direct parameter estimation. The net effect of anchovy on reproductive success and first year survival is an output of the model and hence not directly observed. The Panel recommends that the penguin pressure model not be used for tactical purposes. Section C.2.2 outlines how the Panel recommends the penguin pressure model be used for scientific and management purposes.

C.2.2 To the extent that time may permit the issues concerned to be addressed, advise on the reliability of the pressure model to inform on other penguin-fishery management related issues, taking account of queries concerning this model that have been tabled and its authors' responses.

The penguin pressure model attempts to capture as many of the factors impacting penguin population dynamics as possible, with some parameter values based on external fits to data, other parameter values set based on expert opinion, and others "tuned" to fit the available data. Unlike the Robinson et al. model, the penguin pressure model includes factors and processes (e.g. climate change and floods) for which there are limited data. Such factors and processes may not be detectable on purely statistical grounds, but are appropriate when considering risk in a strategic sense. The Panel recommends that the penguin pressure model be reformulated as an integrated population model, ideally with multiple islands represented explicitly, and with the parameters estimated by fitting to the available data. The outputs of such an integrated population model could be used to examine the relative impact of different factors and processes on the future state of the population so as to identify management implications of each set of parameters (hypotheses) that are shown to be able to replicate the available data. The penguin pressure model could then constitute a key input into research planning, development of monitoring strategies, and perhaps management interventions.

References

Boresma, P.D. and G.A. Rebstock. 2009. Foraging distance affects reproductive success in Magellanic penguins. *Marine Ecology Progress Series* 375: 263-275.

Hennicke, J.C. and B.M. Culik. 2005. Foraging performance and reproductive success of Humboldt penguins in relation to prey availability. *Marine Ecology Progress Series* 296: 173-181.

Horswill, C., Matthiopoulos, J., Green, J.A., Meredith, M.P., Forcada, J., Peat, H., Preston, M., Tranthan, P.N. and N.T. Ratcliffee. 2014. Survival in macaroni penguins and the relative importance of different drivers: individual traits, predation pressure and environmental variability. *Journal of Animal Ecology* 83: 1057-1067.

McClung, M.R., Seddon, P.J., Massaro, M. and A.N. Setiawan. 2004. Nature-based tourism impacts on yellow-eyed penguins *Megadyptes antipodes*: does unregulated visitor access affect fledging weight and juvenile survival? *Biological Conservation* 119: 279-285.

Parada, C., Mullon, C., Roy, C., Fréon, P., Hutchings, L. and C.D. van der Lingen. 2008. Does vertical migratory behaviour retain fish larvae onshore in upwelling ecosystems? A modelling study of anchovy in the southern Benguela. *African Journal of Marine Science* 30: 437-452.

Rademeyer, R.A. 2013. An initial attempt at a spatially structured stock assessment for the South African hake resource including explicit movement. Paper MARAM/IWS/DEC13/Hake/P9.

Rademeyer, R.A. 2014. An initial attempt at a spatially structured stock assessment for the South African hake resource including movement based on a gravity model. Paper MARAM/IWS/DEC14/Hake/P5.

Robinson, W.M.L., Butterworth, D.S. and É.E. Plagányi. 2015. Quantifying the projected impact of the South African sardine fishery on the Robben Island penguin colony. *ICES Journal of Marine Science* 72: 1822-1833.

Table 1. Estimates (in log space) of effect size and standard error for chick condition based on analysing the 2004 and 2008-13 data using a Bayesian linear mixed model with no biomass term. Results are shown when the data are disaggregated and when they are aggregated.

Data Type	Island	Effect Size	SE/SD
Disagg	Dassen	0.003	0.03
	Robben	-0.12	0.03
Agg	Dassen	-0.04	0.11
	Robben	-0.06	0.10

Table 2. Integrated detection probabilities for 1, 10 and 20 years' of future simulated data. "Original" responds to run 13 of Table 6 of MARAM/IWS/DEC16/Peng Clos/P1a and "Adjusted SE" replaces the GLM estimates of variance with those from run (2) of MARAM/IWS/DEC16/PengClos/P4.

		GLM-bias-adjusted delta	**SE**	**1**	**10**	**20**
Dassen	Original	-0.08	0.22	0.85	0.95	0.97
	Adjusted SE	-0.08	0.10	0.72	0.86	0.90
Robben	Original	-0.13	0.20	0.88	0.95	0.96
	Adjusted SE	-0.13	0.09	0.77	0.86	0.91

Table 3. The values of P_{min} at which $P_i(\lambda_i < T)$ from Equation 2 is 0.5 when λ/δ is equal to the Threshold (here -0.1) (source: Table 4 of MARAM/DEC16/Peng_Close/P1a for Chick Growth). After each such grouping, the weighted P_{min} bias is listed for each year. P_{min} values for catch only and catch+closure OMs have received half the weight of those for closure only OMs, since the former are "double-counted" owing to the two correlation values.

			(a) Dassen Island					(b) Robben Island				
OM	EM	Cor	1	5	10	15	20	1	5	10	15	20
1. Catch	Catch	0.2	0.46	0.44	0.42	0.41	0.42	0.49	0.48	0.49	0.49	0.47
2. Catch	Catch	0.4	0.42	0.40	0.37	0.37	0.36	0.45	0.43	0.43	0.43	0.40
6. Closure	Catch	-	0.11	0.08	0.06	0.05	0.04	0.13	0.11	0.07	0.05	0.03
7. Ca+Cl	Catch	0.2	0.20	0.18	0.15	0.13	0.15	0.24	0.22	0.18	0.15	0.12
8. Ca+Cl	Catch	0.4	0.14	0.11	0.09	0.08	0.08	0.21	0.18	0.14	0.11	0.09
Catch EM P_{min} bias correction			0.24	0.21	0.19	0.18	0.18	0.28	0.26	0.23	0.21	0.19
3. Closure	Closure	-	0.51	0.48	0.47	0.48	0.50	0.50	0.50	0.51	0.51	0.49
4. Catch	Closure	0.2	0.61	0.63	0.65	0.67	0.69	0.60	0.59	0.63	0.66	0.66
5. Catch	Closure	0.4	0.62	0.64	0.67	0.68	0.69	0.58	0.61	0.61	0.63	0.63
9. Ca+Cl	Closure	0.2	0.57	0.58	0.60	0.61	0.63	0.54	0.56	0.57	0.57	0.58
10. Ca+Cl	Closure	0.4	0.56	0.55	0.56	0.55	0.57	0.55	0.55	0.57	0.57	0.58
Closure EM P_{min} bias correction			0.58	0.58	0.59	0.60	0.62	0.56	0.57	0.58	0.59	0.59

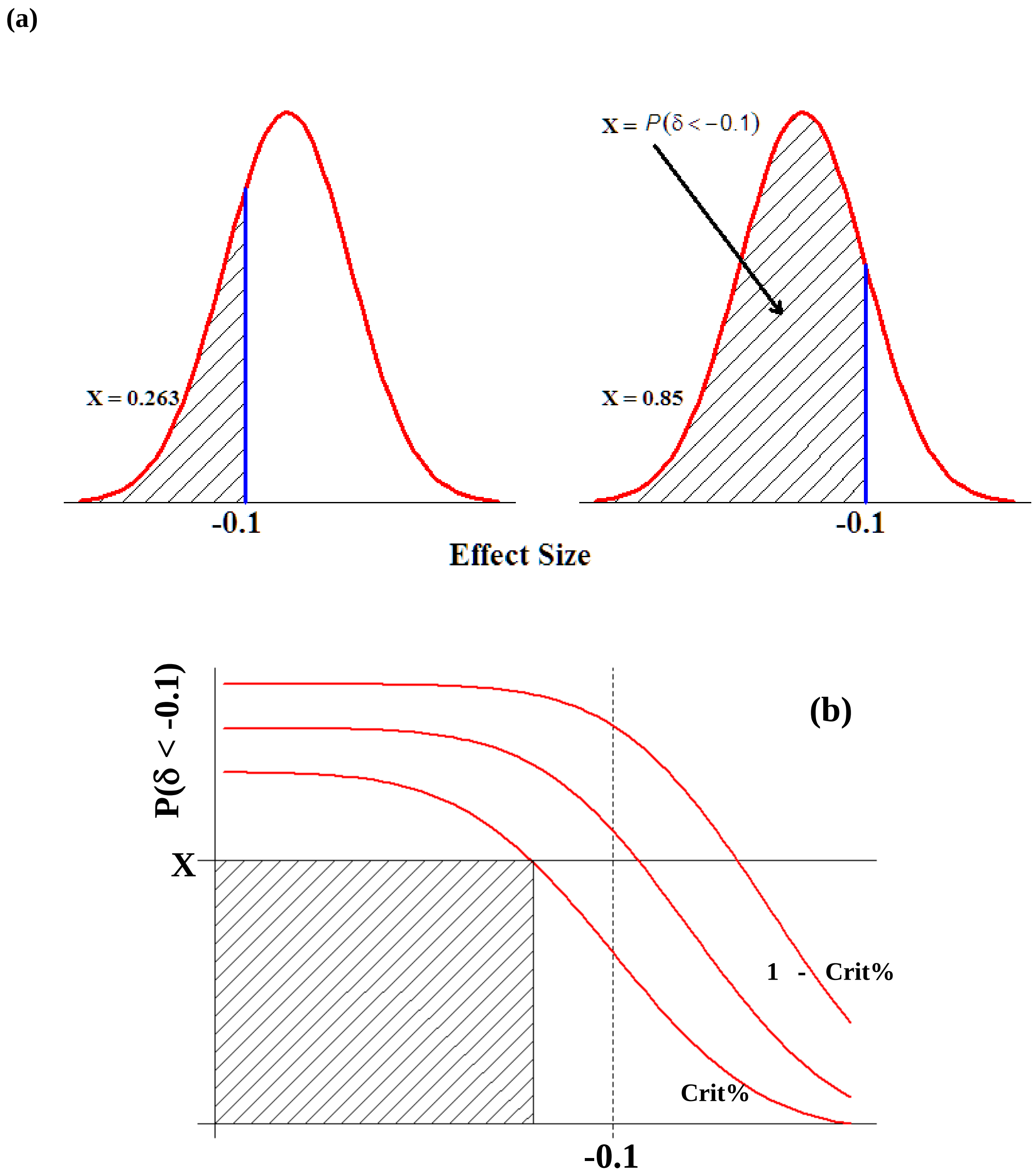

Figure 1. Two of the steps in the algorithm identified for interpreting the results of the fishing/closure experiment.

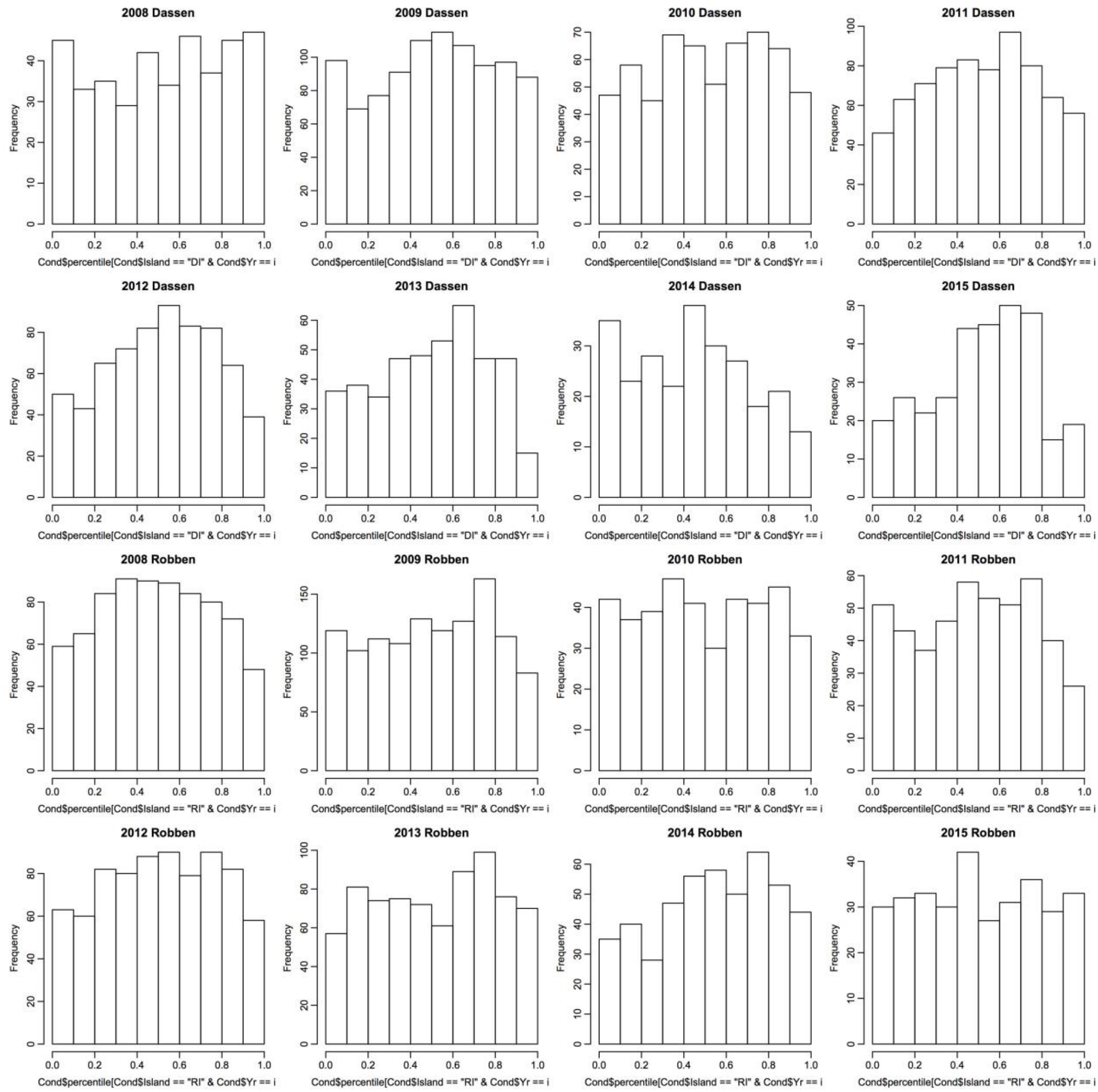

Figure 2. Posterior predictive p-values for each island x year combination from the fit to disaggregated data (2008-2015).