WORKSHOP ON CATCH FORECAST FROM BIASED ASSESSMENTS (WKFORBIAS; outputs from 2019 meeting)

VOLUME 2 | ISSUE 28

ICES SCIENTIFIC REPORTS

RAPPORTS
SCIENTIFIQUES DU CIEM



ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

The material in this report may be reused for non-commercial purposes using the recommended citation. ICES may only grant usage rights of information, data, images, graphs, etc. of which it has ownership. For other third-party material cited in this report, you must contact the original copyright holder for permission. For citation of datasets or use of data to be included in other databases, please refer to the latest ICES data policy on ICES website. All extracts must be acknowledged. For other reproduction requests please contact the General Secretary.

This document is the product of an expert group under the auspices of the International Council for the Exploration of the Sea and does not necessarily represent the view of the Council.

ISSN number: 2618-1371 I © 2020 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 2 | Issue 28

WORKSHOP ON CATCH FORECAST FROM BIASED ASSESSMENTS (WKFORBIAS; outputs from 2019 meeting)

Recommended format for purpose of citation:

ICES. 2020. Workshop on Catch Forecast from Biased Assessments (WKFORBIAS; outputs from 2019 meeting).
ICES Scientific Reports. 2:28. 38 pp. http://doi.org/10.17895/ices.pub.5997

Editors

Larry Alade • Christopher Legault

Authors

Valerio Bartolino • Robert Boenish • Liz Brooks • Noel Cadigan • Steven Cadrin • Santiago Cervino Ghislain Chouinard • Kiersten Curti • Mikael van Deurs • Daniel Howell • Christopher Legault • Tanja Miethe • Colin Millar • Kotaro Ono • Andrea Perreault • Pia Schuchert • Nicola Walker • Jonathan White • Laura Wise

ICES International Council for the Exploration of the Sea
CIEM Conseil International pour l'Exploration de la Mer

Contents

i Executive summary ii
ii Expert group information iii
1 Introduction 1
2 Summary of Workshop Papers and Discussions 2
3 WKFORBIAS Conclusions 22
4 WKFORBIAS Recommendations and Decision Tree 24
5 WKFORBIAS Action List: Guidelines for investigating a retrospective pattern 27
6 References 30
Annex 1: List of participants 31
Annex 2: WKFORBIAS Summary Draft Report 32
Annex 3: Resolution 37

i Executive summary

The workshop on catch forecasts from biased assessments, WKFORBIAS, met to address and develop general guidelines for dealing with the issue of retrospective patterns in stock assessments. Sixteen working papers were presented during the workshop, addressing one or more of the terms of references. The Workshop reaffirmed previous recommendations that retrospective analysis should always be conducted as a diagnostic to examine the internal consistency of an analytical stock assessment. The Mohn's rho statistic that compares estimates from assessments with recent years of data removed to estimates from the current assessment is the standard tool for retrospective analysis. Examination of recent category 1 and 2 ICES stock assessments indicates that a majority of the assessments does not exhibit strong retrospective patterns that require changes to standard management advice. Across the wide range of stocks examined, no obvious explanatory variables, such as model type, location, fishery type, or biological trait, separated stocks with and without strong retrospective patterns. By comparison, both the magnitude and the proportion of stocks with retrospective patterns were greater in the Northeast US than observed for ICES stocks.

For the stock assessments that do show strong retrospective patterns, the first step was to identify what constitutes a strong retrospective pattern then a decision tree was developed to help expert groups determine a course of action. A number of general recommendations from WKFORBIAS include: 1) when evaluating a retrospective pattern, the consistency of the pattern is of primary importance; 2) a large Mohn's rho statistic driven by one outlier should not be treated in the same manner as a consistent directional retrospective pattern; 3) retrospective patterns should be viewed as one of many diagnostics to be used in determining whether to use an assessment for management advice or not; 4) a strong consistent retrospective pattern can be the basis for adjusting catch advice or downgrading the level of an assessment. For stocks that exhibit a strong retrospective pattern where SSB is consistently overestimated and F is consistently underestimated, it is recommended to adjust catch advice. For less common retrospective patterns, where SSB consistently increases with additional years of data, then adjusting catch advice is not recommended. However, in the event an assessment is being downgraded, strong consideration should be given to maintaining models with robust trends even if the scale of the assessment is uncertain (i.e. downgrading from a category 1 to a category 2 assessment); 5) Management Strategy Evaluation can potentially be a useful tool for examining the robustness of harvest control rules to different magnitudes of retrospective pattern and could be useful for situations exhibiting strong retrospective patterns over multiple assessments. However, simulating specific retrospective patterns is challenging as demonstrated by the Rose approach. A complete set of recommendations are provided in the report including a compiled action list for consideration by expert groups to evaluate possible sources of the retrospective pattern.

ii Expert group information

Expert group name	Workshop on Catch Forecast from Biased Assessments (WKFORBIAS)
Expert group cycle	Annual
Year cycle started	2019
Reporting year in cycle	1/1
Chairs	Larry Alade, USA
	Christopher Legault, USA
Meeting venue and dates	11-15 November 2019, Woods Hole, USA (20 participants)

1 Introduction

The workshop on catch forecasts from biased assessments, WKFORBIAS, met in Woods Hole, MA, United States, on 11-15 November 2019. There were a number of objectives for the workshop including:

1. Examine the extent and the magnitude of retrospective bias identified in ICES category 1 and 2 stock assessments;
2. Determine the acceptability of such assessment; and
3. Investigate potential correction of population metrics using Mohn's rho statistics.

A retrospective "bias[1]" is a pattern of systematic change in the assessment estimate of population size or other assessment-derived quantities that occur when additional years of data are added to the assessment (Mohn, 1999; Legault, 2009; Hurtado-Ferro, 2015). There are two forms of retrospective pattern: historical and within-model. The historical retrospective describes the results of a final assessment for successive number of years, which can exhibit consistent pattern of overestimating or underestimating assessment values from one year to the next. Such inconsistency may arise from changes in data, assessment model or assessment model formulation. The historical retrospective pattern has direct influence on management choices in the past because of its dependency on the information available at the time. In contrast, the within-model retrospective pattern is a measure of internal model consistency based on identical data stream, model type and/or formulation that peels off recent years of data in successive model runs. The only change for the within-model retrospective pattern is the number of years of data included in the model. For the purpose of the report, the primary focus is on the within-model retrospective issue.

The issue of retrospective pattern in stock assessments has previously been identified, as early as in the 1980s (Butterworth, 1981; Sinclair *et al*., 1991; Parma 1993), and it is not unique to any region, fishery type, biological characteristics or modelling framework. Further, the ICES Methods Working Group (MGWG), has been addressing the topic since 1991. While the retrospective issue is not a new one to ICES, there is no formal process in place to address the issue in a consistent manner. As such, this workshop was designed with the intent of developing a working framework to provide guidance to analytical stock assessments with retrospective patterns. This report summarizes a wide range of analyses conducted during WKFORBIAS along with conclusions and recommendations agreed upon in the November 2019 workshop meeting. While WKFORBIAS was unable to provide a resolution to the retrospective problem, a set of recommendations and a decision-tree chart was developed to help experts determine a course of action when an assessment shows a strong within-model retrospective pattern.

[1] The term 'retrospective pattern' is considered more appropriate and is used in this report to describe this characteristic.

2 Summary of Workshop Papers and Discussions

A total of 16 working papers were presented during the workshop. The terms of reference A-E were addressed by 5, 5, 3, 5, and 7 different working papers, respectively, with some working papers addressing more than one term of reference.

Retrospective Pattern in age-structured fish stock assessments: Mohn's Rho estimates from 2018 and 2019 ICES assessments.

Ghislain Chouinard and Colin Millar

Terms of Reference: A

Abstract

A number of ICES stock assessments have shown consistent retrospective patterns in the estimates of population metrics from successive assessments. The patterns in the estimates of SSB (sometimes referred to as bias) from these assessments can be positive (overestimates – population abundance successively revised downwards) or negative (underestimates – population abundance successively revised upwards). These patterns will result in catch forecasts that lead to fishing mortality being either consistently higher or lower than the actual target. Causes of the retrospective patterns are various and often difficult to identify.

In 2018 and 2019, ICES expert groups were asked to calculate the Mohn's rho statistic as a diagnostic of retrospective patterns. These results have been collated and used with an existing rule of thumb to identify fish stocks assessed by ICES that suffer from significant retrospective patterns. Results indicate that 15 to 20% of age-structured stock assessments have significant retrospective patterns based on the rule of thumb from Hurtado-Ferro *et al.* (2015).

Summary and Discussions

Hurtado-Ferro *et al.* (2015) found that non-zero values of Mohn's rho can be generated with no model specification. Rule of thumb for significant retrospective pattern:

- Long-lived stocks: Mohn's rho higher than 0.2 or lower than -0.15
- Short-lived stocks: Mohn's rho higher than 0.3 or lower than -0.22

2018-2019 data for 74 out of 84 category 1 and 2 age-structured stock assessments indicated that there were more overestimates than underestimates. Retrospective patterns are present in stocks from all ecoregions with a higher number of stocks in the Western water region. There was no specific pattern or trend based on assessment method. Qualitative evaluation of Mohn's rho for 19 and 14 age-structured stock assessments suffered from retrospective patterns in 2018 and 2019 respectively. Of the 74 stocks evaluated, Mohn's rho improved for eight stocks, four stocks got worse and seven stocks did not change between 2018 and 2019.

Brooks and Legault (2016) proposed a 'rule of thumb' that specified to adjust using Mohn's rho when Mohn's rho adjusted estimates are outside the 5 and 95% bootstrapped estimates of SSB and F. The study indicated that adjustment for Mohn's rho in those cases was better than ignoring the retrospective pattern but did not eliminate the retrospective pattern entirely. In 12% of the cases, the adjustment was actually in the wrong direction. Workshop participants felt that Mohn's rho was a simple diagnostic to help diagnose stocks with retrospective patterns but closer examination of retrospective plots are useful to gain fuller understanding and detect

emerging issues in an assessment. Given the potential for error, although small, identifying the important population processes that vary over time when addressing retrospective patterns should be a priority (Szuwalski *et al.*, 2017).

Summary of Retrospective Patterns from the most recent Management Track Assessments in the Northeast United States

Chris Legault

Terms of reference: A

Abstract

As summary of recent stock assessments in the Northeast United states were examined for Mohn's rho for spawning stock biomass. In this region, a classification of Mohn's rho is based on a rule of thumb developed by Brooks and Legault (2016) that specifies that a major retrospective pattern occur when either of the rho-adjusted terminal assessment year estimates are outside the 5−95% confidence interval of SSB or F. Mohn's rho from the recent assessments suggests that retrospective patterns are common in this region. Mohn's rho estimates tend to scale much larger compared to those from the 2018 and 2019 ICES stock assessments. The vast majority of the Mohn's rho for SSB were positive and there were no common pattern detected by region, life history, or stock assessment model type. Generally, there were no strong signals of improvement in Mohn's rho. The increase number of rejected assessments in recent years indicate that retrospective patterns are getting worse.

Summary and Discussions

This paper examines retrospective patterns for many but not all of the stock assessments for the northeast United States in recent years. As indicated previously, a stock is considered to have a major retrospective pattern following criteria by Brooks and Legault (2016). It was concluded that while not all stocks assessments exhibit a strong retrospective pattern, it is common in the region. The vast majority of Mohn's rho for estimates of spawning stock biomass are positive. There has been no strong signal of improvement for retrospective diagnostics and it has led to an increased number of rejected assessments, indicating the issue is getting worse. Similar to ICES findings, there is no pattern detected in Mohn's rho by ecoregion, life history, and stock assessment method.

Plenary Discussion

Workshop participants participated in an open discussion on the merits and application of Mohn's rho as a metric to characterize within-model retrospective patterns. The measure of Mohn's rho is based on the comparison of model terminal year estimate of each assessment peel relative to the estimates from a full assessment. One of the properties of Mohn's rho is that values are bounded by -1 and infinity, leading to an asymmetric rule of thumb. Mohn's rho is dependent on the number of peels (5 vs. 7 years). The ICES standard is 5 peels and the Northeast US standard is 7 peels. Miller and Legault (2017) found the value of Mohn's rho often stabilized after 5 peels. Models with short time-series can cause problems for large numbers of peels and may even result in a different model. It was also suggested that the number of peels could be conditioned on life history (i.e. plus group in the age composition). Directionality of retrospective peels is important. Some workshop participants pointed out that Mohn's rho does not appear to have good statistical properties for cases that are not directionally consistent. For instance, individual year peels with large retrospective pattern in both directions could present the appearance of

reasonable retrospective diagnostics based on Mohn's rho. Hence, it is recommended that such cases should be scrutinized carefully.

Workshop participants were asked to comment on the utility of Mohn's rho. The following bullets capture the consensus among the group:

- Mohn's rho is one of many model diagnostics. Solving the retro does not imply the model is adequate for providing catch advice nor should the model necessarily be rejected if there is a notable Mohn's rho.
- Some of the workshop participants indicated that terminal year estimates based on Mohn's rho outside the confidence limit is an indication of uncertainty underestimation in the stock assessment. Precision estimates assume that the model application is correct. So in cases when rho adjusted estimates fall outside of the confidence limits suggest that the model error is greater than imprecision.
- Often, there are other model diagnostics for consideration when evaluating the merits of a model for catch advice including patterns in residuals, poor fit to indices, Log-likelihood, etc. It is advised that analysts should not only focus on retrospective patterns but also take a closer look at some of the model assumptions and biological data. For instance, a strong retro could be due to just data conflicts in the model, warranting further investigation in data inputs.
- Workshop participants felt that Mohn's rho is sufficient for measuring the magnitude of retrospective patterns. It is also considered good for when the peels are directionally consistent. However, in cases when the peels are not unidirectional (i.e. mix of both positives and negatives), each peel should be evaluated a bit more closely.

Retrospective Forecasting – Evaluating the performance of stock projections for New England Stocks

Liz Brooks and Chris Legault (Published in Canadian Journal of Aquatic Sciences in 2016)

Terms of reference: B

Abstract

Projections are used to explore scenarios for catch advice and rebuilding and are an important tool for sustainably managing fisheries. We tested each projection specification for 12 groundfish stocks in the Northwest Atlantic to identify sources of bias and evaluate techniques for reducing bias. Projections were made from assessments using virtual population analysis (VPA) with 1–7 years of recent data removed from the full time-series and were then compared with results from a VPA assessment on the full time-series of data. The main source of bias in projections was the assessment model estimates of the numbers-at-age in the terminal model year+1(Na,t+1). Recruitment was responsible for more bias in projections beyond 3 years, when population numbers begin to be dominated by cohorts that were statistically generated. Retrospective analysis was performed and several adjustment factors to reduce bias were tested. Even after adjusting for bias, the remaining bias in projections was non-negligible. The direction of bias generally resulted in projected spawning stock biomass (SSB) and catch being overestimated, and the bias in catch was nearly always larger than in SSB. Scientists need to clearly communicate the direction and magnitude of this bias, managers need to consider this additional uncertainty when specifying future catch limits, and both scientists and managers need to develop more robust control rules so that objectives are achieved.

Summary and Discussions

This study uses estimates from the most recent assessment as 'true values' to evaluate the performance of retrospective forecasting and not simulation-based. Generally, when retrospective pattern is not accounted for in the forecast, projections generally performed poorly. Starting abundance-at-age was considered the most important factor responsible for the most retrospective pattern in the first two years of the projections. After 3-4 years, projected recruitment tends to account for more retrospective pattern because catch and SSB in the short term depend on current cohorts where longer term projections depend on recruitment which is typically assumed based on previous patterns. To account for a 'major' retrospective pattern in the projections, adjustments based on Mohn's rho were applied to starting abundance-at-age. It should be noted that rho adjusted SSB performed well in the projections as well or better than rho adjustment based on abundances-at-age. Scale of retrospective pattern reduction was achieved for a number of peels seemed to taper between 5 and 7, suggesting that the current approach to use rho for SSB based on 7-year peel should be maintained. Although rho adjusted projections resulted in less retrospective pattern than not rho adjusting, there was still some residual retrospective pattern in the projections, suggesting that rho adjustment helps but is not enough. It should be noted that uncertainty characterized from VPA bootstrap is not always sufficient to characterize all uncertainty and therefore it raises the issue that both retrospective pattern and uncertainty will need to be addressed. One suggestion is to develop a modified Harvest Control Rule by 1.) Apply the Brooks and Legault (2016) rule of thumb, 2.) Rho adjust if there is indication of retrospective pattern, 3.) Generate lognormal distribution with a high CV around the result of step 1 (uncertainty) and provide advice for a percentile less than the median to account for residual retrospective pattern. The study also explored the utility of simpler models such as the aggregated biomass dynamic models and concluded that simpler models did not resolve retrospective pattern due to instability in reference points estimates, carrying capacity and initial conditions. Overall, this study highlighted the challenges with retrospective patterns as it affects SSB and catch, including rebuilding estimates.

All models (not just VPA) are susceptible to retrospective patterns. The current approach to examining and testing for retrospective patterns is appropriate, but we need to consider 'edge effect' for rule of thumb. Clearly, rho adjustment performs better than alternatives tested, but does not completely solve the retrospective pattern problem. It is recommended to incorporate additional 'unmodelled' uncertainty in future HCRs. Future investigations to improve forecast include consideration for the number of years for average condition (3 years vs. 5 years). This could be useful especially for stocks with strong biological trends. Recent recruitment for forecasting was recommended to be helpful in longer-term forecast, especially for stocks with strong recent low recruitment. It was also recommended to consider short-term projections (≤3 years) as long-term projections tend to be very suspicious. This study demonstrates the value of evaluating performance with real data and recommend that models intended to inform management should be rigorously tested preferably with real data but can also be accomplished with simulated data.

The Fallacy and Consequence of Retrospective 'Bias' and 'Correction' in Stock Assessment

Steve Cadrin

Terms of Reference: A

Abstract

Contemporary estimates of abundance or mortality from stock assessment models usually change when a new year of data are added. For some stock assessments, there is a pattern to this retrospective inconsistency in which changes are in the same direction. Such patterns suggest

that the estimation model is misspecified. Unfortunately, the average retrospective inconsistency is commonly referred to as a 'bias', and estimates of stock size and fishing mortality are often 'corrected' for this perceived bias. The logic implied by retrospective 'bias' and 'correction' is that retrospective estimates from the updated model are true, and deviations from the updated estimates represent bias that can be corrected. The logical fallacy is exposed when another new year of data is added to the model, and the previous estimates that were assumed to be true are now considered to be biased. Simulation studies have shown that the scale and direction of retrospective inconsistencies do not reflect true bias. Several independent estimates of stock size also suggest that retrospective corrections were in the wrong direction. The terminology of retrospective 'bias' and 'correction' and the implicit logical fallacy are misleading to fishery managers, who are compelled to make decisions based on best available scientific information.

Summary and Discussions

Workshop participants debated this alternative perspective on the use of the retrospective terminology as it relates to retrospective adjustment, performance and management implications. Contrary to the term retrospective bias, the term retrospective inconsistency is being proposed and it is said to occur when contemporary estimates of abundance or mortality changes with new addition of data. The logical fallacy in the term retrospective bias implies that retrospective estimates from an updated model are true and deviations from the updated estimates (i.e. retrospective peels) represent bias that can be corrected, questioning the notion of retrospective inconsistency as a measure of bias. An alternative perspective is that the inconsistent pattern demonstrates that bias exists, even if the truth is not the final assessment or known at all. Deroba (2014) suggests that a retrospective pattern does not necessarily indicate the direction of parameter bias but proclaims that when application of Mohn's adjustment to accurate estimates of biomass, subsequent spawning biomass and yields tend to be near target levels. Further, Hurtado-Ferro (2015) demonstrates through simulations that the scale and the direction of retrospective inconsistencies do not reflect true bias. As such, it was pointed out during the workshop that a retrospective adjustment should be considered carefully as there are some conditions under which a Mohn's rho adjustment may not be suitable for application. For instance, a retrospective adjustment in the wrong direction can result from inconsistencies between model estimates and independent estimates of stock size, suggesting a possible internal data conflict in the model among other potential reasons. In the event of a retrospective-adjusted stock estimate that results in a lower estimate of stock size than observed catch is also an indication that a retrospective adjustment could be in the wrong direction. However, if a new model is accurate, retrospective adjustment appears to be appropriate. The 2018 Atlantic herring assessment is an example where the rho adjusted 2016 values was right in line with the 2018 update model estimate (Deroba, 2018).

Relative to the performance of retrospective adjustment, the summary asserts that even after adjusting for retrospective pattern, the remaining retrospective pattern in projections was non-negligible (Brooks and Legault, 2016). It was also pointed out that retrospective adjustment made to starting abundance in projections has not been sufficient to prevent overfishing for most stocks, but when used in conjunction with another or with alternative control rule that reduced the target harvest rate as biomass fell below the target (Wiedenmann and Jensen, 2019).

In conclusion, this paper suggests that the terminology of retrospective 'bias' and 'correction' and the implicit logical fallacy are misleading to fishery managers, who are compelled to make decision based on best available science. If a retrospective inconsistency is treated as diagnostic and an indicator of scientific uncertainty, retrospective patterns can be used to improve models and consequences of continued retrospective patterns can be more appropriately communicated

to managers. Retrospective adjustment can be performed well for meeting management objectives, but the adjustment should be considered a precautionary approach to managing in the context of scientific uncertainty. Finally, the summary suggests that adjustment and catch advice may be more consistent over time, but conclusions made from the adjustment are not always more accurate.

A simple Method for Longitudinal Retrospective Analyses: Moving window rho

Robert Boenish

Terms of Reference: B and Cii

Abstract

Atlantic cod (*Gadus morhua*) in the Gulf of Maine (GoM), USA once served as a regionally important fishery, but have been serially depleted to <5% of historic spawning stock biomass. Recent management efforts to rebuild GoM cod have largely failed and model diagnostics remain poor. The stock assessment has listed retrospective patterns as a major concern for many years, leading to the simultaneous use of two working models- one with ramped and fixed mortality (M-RAMP and M = 0.2, respectively). Past debate on retrospective analysis has focused on whether or not to 'correct' retrospective patterns using fixed year approaches. These approaches necessarily only focus on the most recent year's retrospective patterns (e.g. 5 or 7 years). However, assessment models in this region are fit on many decades of data, thus understanding retrospective patterns over a longer timeframe is warranted. We introduce a new moving window approach to describe temporal changes in retrospective patterns, "moving window Rho". Results suggest retrospective patterns shifted from a relatively stable state (updated SSB revised upward) to a separate stable state (updated SSB revised downward) in the 1990s. These results mirror other data that suggest a regional state change took place during this time. This work is an important step towards more accurate catch accounting, and will have immediate and overdue implications for cod management and the ways in which scientists may view the 'retrospective problem' in fisheries. Evaluation of longitudinal retrospective patterns may signal need for when alternative modelling approaches or specifications are warranted.

Summary and Discussions

This study investigated longitudinal patterns of retrospective patterns. A clarification identified by the workshop is that the longitudinal approach is based on a moving average and not a moving window because the long-term peels conducted without a change the terminal year. It was also noted that every retrospective peel did not run into convergence issues despite the formulation of fishery selectivity blocks in the model. Overall, this approach was considered useful to evaluate systematic and temporal issues that may be a source of retrospective pattern in the model. For example, northern cod showed strong retrospective patterns when the stock collapsed, suggesting that there may be a spatial factor due to greater possibility of the stock being outside the survey area when the stock is small, suggesting a change in availability to the survey. Other factors identified included a change in natural mortality and unreported catches.

Comparing the Rose Approach to Rho-Adjusted Catch Advice: Does the Type of Accounting for a Retrospective Matter

Chris Legault

Terms of reference: D and E

Abstract

A common practice in the northeast US to account for a strong retrospective pattern is to modify the terminal year population estimates using a rho-adjustment before providing catch advice. This typically results in lower catch advice for stocks that exhibit the usual retrospective pattern of updated biomass being adjusted down (indicated by a positive value of Mohn's rho). A recently proposed alternative creates an ensemble of models that each addresses the retrospective pattern in different ways, for example by increasing catch or natural mortality in recent years. The name Rose Approach derives from the Shakespeare line about any other flower smelling as sweet because the actual cause of the retrospective pattern is not attempted to be identified. Instead, the ensemble of possible models is used to provide a range of catch advice and some measure of central tendency from the ensemble is used for the overall catch advice. The two approaches are compared for a few stocks to demonstrate the pros and cons of each approach.

Summary and Discussions

The Rose approach is based on a model ensemble framework that uses revised data or assumptions to resolve model retrospective patterns. One of the advantages of this method is that is has the potential to address model misspecification while accounting for blaming the "wrong suspect" (i.e. wrong model assumptions or data). The challenge with the Rose approach is the difficulty in determining the rules ahead of time of which models to exclude in the ensemble based on model diagnostics. However, sensitivity analyses can be viewed as a potential approach for making such determination. Often, models that do not sufficiently meet diagnostics standards and are often dropped with 100% of the weighting placed on the desired base run. It was suggested during the workshop that perturbation analyses can be done to determine effectively when rho is zero. A potential approach is through a local influence analyses to put bounds on the problem (Cadigan and Ferro, 2005). It was noted during the workshop that Mohn's rho in the northeast U.S. is greater than observed for ICES stocks and no patterns were observed in the retrospective results suggesting the source of problem. Workshop participants questioned at what point is the rho statistics a sign of model misspecification to reject the model. A suggestion was to consider rho adjustment to harvest control rules as opposed to the typical within assessment rho adjustment. It was further debated that model rejections often lead to the use of simple index approaches that ignore a lot of vital information and require ad-hoc decisions that can lead to higher catch advice than the rejected model. It was pointed out that there are a lot of benefits to having a model even if there are diagnostic issues. As such, it was suggested that one could think of retrospective pattern as uncertainty. When you have more uncertainty, advice should be more precautionary. One recommendation for ICES stocks is to perhaps reconsider the use of category 2 stocks for trends. The frequent use of Category 3 in place of rejected assessments results in the only use of one source of data (survey data) and ignores all other sources of information.

Accounting for retrospective Patterns in a State-Space Stock Assessment

Noel Cadigan and Andrea Perreault

Terms of Reference: E

Abstract

We investigate using process errors to account for retrospective patterns in state-space stock assessment models (SSSMs). Retrospective patterns are caused by a misspecification of some aspect of the stock assessment model, and process errors are included in SSSMs for this purpose. An advantage of this approach is that the retrospective adjustment is made to the model formulation rather than the model output, and the process errors can then be forecasted to potentially provide more reliable future catch advice. Accounting for retrospective pattern using process errors essentially turns the pattern into an uncertainty. Advice from such a SSSM should have higher uncertainty to acknowledge that there is uncertainty in population processes in addition to unknown model parameters. However, how process errors are modelled can have important impacts on the efficacy of a SSSM when there is model misspecification. Process errors are usually included in the population dynamics model and may not lead to a better stock assessment if the misspecification is in the observation model (e.g. change in q). In addition, process errors may be confounded with variations in fishing mortality in practice. Hence, more investigation of how process errors should be included in SSSMs is required. We will simulation test this using a variety of model misspecifications. We illustrate concepts using three case studies (3LNO American plaice, 3NO cod and 3Ps cod) where we use correlated process errors to reduce retrospective patterns.

Summary and Discussions

This analysis uses process errors to account for retrospective patterns in state-space stock assessment models. In cases when model misspecification is the source of retrospective pattern, process errors in SSSM are designed for this purpose. Retrospective adjustment is made to the model in the formulation rather than the model output, and process errors can then be forecasted to potentially provide more reliable future catch advice. The three case studies (3LNO American plaice, 3NO cod and 3Ps cod) illustrated in this study demonstrate the potential to reduce retrospective patterns using correlated process errors in state-space stock assessment model. However, it is easy to go wrong. Process errors tend to be confounded with mortality (particularly for an assessment with a single survey), or with misspecification of the observation model such as occasional shifts in catchability. It was recommended at the workshop that simulation testing should help with providing guidance on the issue of confounding in the model's process errors. The recommended strategy was to start with the most parsimonious model and fix aspects of the model that are problematic. Concerns were raised about trying to account for structural model misspecification with stochastic process error. Immigration/emigration could be interpreted in recruitment and could potentially lead to inappropriate application of process errors and potentially misleading interpretations. Preliminary simulations were explored that looked at models with no process errors. The simulations considered different fishing and natural mortality scenarios. When no process error in the underlying process, preliminary results only show differences in F among the different M scenario (i.e. when M is mis-specified). It was recommended that additional simulation work be conducted to fully understand the behaviour of process errors relative to reducing retrospective patterns.

Retrospective Bias and Management Strategy Evaluation (MSE): Case Study of North Sea Saithe

Daniel Howell and Daisuke Goto

Terms of Reference: D and E

Abstract

Retrospective bias in assessments used for quota setting is, obviously, a source of error in both the assessment and the management of a stock. Management Strategy Evaluation (or more specifically Harvest Control Evaluation) is an accepted tool for analysing the impacts of errors of management performance and designing management strategies that are robust to the errors in the assessment. One can use these tools to evaluate the impact of consistent retrospective patterns on any given HCR for a particular stock, and go further and attempt to identify HCRs that would be robust to a given assumed level of retrospective bias. We note that this analysis can be simplified for a HCR where the quota is a function of the terminal year estimates. Where there is a slope-based HCR the analysis would be more involved, and this is not attempted here.

We use the recently conducted MSE for North Sea saithe as a base. To this we add different levels of retrospective bias and analyse the implications for the performance of the fishery. Preliminary results suggest that adding unaccounted retrospective bias leads to little change in the long-term average yield, but can lead to increased levels of risk and interannual catch variability. For the N. Sea saithe, we found that a consistent overestimation of the stock by over around 20% pushed the risk above the ICES acceptable threshold, but that at lower levels of bias the HCR remained precautionary. This result is obviously stock-specific, the recruitment function and the basis for the HCR (top of the MSY curve vs. lying on the acceptable risk threshold) are likely to be the key stock specific drivers.

We will continue this work, and aim to expand the analysis to consider what would be an optimal HCRs that gave good yields while being robust to different levels of retrospective bias. We will also aim to coordinate with CEFAS to compare the results with North Sea cod in order to begin to get an idea of the interspecies variability.

Summary and Discussions

Considering retrospectives as a source of error, this study recommends the use of Management Strategy Evaluations (MSE) to deal with the sources of error. The question remains about how much unaccounted bias is the harvest control rule robust to. The MSE approach recommends adding bias on estimated fishing mortality and abundances to evaluate if the HCR is robust to errors. Based on the case study on North Sea saithe, HCR is robust up to 18% bias in abundance and 24% bias in fishing mortality. In essence, it matters where the bias is placed. One of the lessons learned is that long-term catch tends to be robust and risk increases with increases in bias and more variability in the catch. It was pointed out during the workshop that consistent retrospective patterns are easy to fix and model within the MSE framework, but inconsistent ones are more problematic. In this case study, it is known that at some point the retrospective pattern is going to stop but is a source of error that cannot be anticipated in advance as to when it will stop. Correlation between stock size and retrospective patterns tend to be higher with higher stock sizes. It was speculated that there is something within the model mechanics that generates this pattern. If a stock is increasing and one year of data shows a decline, the model does not respond immediately to the change in direction with just one year of data. This suggests the model has some inertia. But by the time enough years have been added to convince the model to change direction, the model had overestimated the peak. The study recommends that adjusting the retrospective pattern in the HCR instead of the model makes the cost apparent. The issue is to

demonstrate how to account for the actual retrospective pattern among other sources of uncertainty.

Quantifying and Accounting for retrospective Bias in the Assessment of North Sea Cod

Nicola Walker, Gwladys Lambert, Simon Fischer and José De Oliveira

Terms of Reference: D and E

Abstract

In recent years, assessments of North Sea cod have shown a downscaling of SSB and upward revision of fishing mortality (F). This is caused by lower catch rates of older fish in the surveys compared to commercial catches and has resulted in large cuts to TAC and advice in recent years.

Mohn's rho provides a metric for quantifying the level of retrospective bias in stock assessments. However, in the case of North Sea cod, three levels of retrospective bias can be considered: (1) a simple data peel, (2) a data peel that accounts for revisions to modelled or smoothed data sources and (3) retrospective bias associated with changes to the assessment procedure. We quantify and compare Mohn's rho for each level.

Management strategy evaluations (MSE) can account for retrospective bias in assessments by adjusting the control parameters of harvest control rules (HCRs) to ensure they meet their objectives, even when assessments are biased. For cod, four operating models were tested earlier this year, and it was found that the combination of $F_{target} = F_{MSY}$ and $B_{trigger} = MSY\ B_{trigger}$ (termed A*) was precautionary in the long term, even when the assessment was biased or contained assumptions that did not agree with the underlying Operating model (ICES, 2019).

Here, we employ the same MSE framework to assess the suitability of using Mohn's rho to adjust catch forecasts for North Sea cod directly. We determine the differences in F from applying HCR A* to both operating model and management procedure and examine if and how these differences correlate with Mohn's rho. The analysis is applied with two operating models: one conditioned on the current assessment and another that includes year effects in the surveys. Finally, we repeat the analysis using an HCR that was shown not precautionary for the second OM and examine whether adjusting the catch forecast can make it precautionary.

Summary and Discussions

This analysis examines three approaches to quantifying retrospective pattern for North Sea cod including 1.) Performing analytical peels by removing one year of data from assessment inputs at a time 2.) peeling off selected raw data by re-evaluating maturity smoothers, survey GAMS and external estimates of natural mortality 3.) Evaluating historical peels based on past advice sheets that incorporate changes to the assessment procedure. Analytical estimates of rho from raw data peels showed the highest levels of retrospective pattern. This suggests that the true level of retrospective pattern in the assessment procedure for North Sea Cod will likely be underestimated using peels only from the assessment inputs. Revisions to the tuning indices resulted in the largest increase in retrospective pattern, despite the low Mohn's rho estimated for the indices themselves. Aside from maturity at age 1 and the IBTS Q3 tuning indices for age 4, values of Mohn's rho calculated for the assessment inputs were low ($< \pm 5\%$). However, this only considers the last 10 years of the time-series. Retrospective peels of maturity show notable changes in the trend further back in time while the tuning indices demonstrate some larger differences at the beginning of the time-series.

Another aspect of the study considered two operating models in an MSE framework evaluating Mohn's rho: a baseline operating model following the current ICES assessment and an alternative operating model that includes survey year effects. Mohn's rho for fishing mortality, recruitment and SSB were built into the MSE to evaluate the performance of using rho to adjust catch projections. For the purpose of the workshop, the operating models were used to demonstrate how retrospective pattern can be accounted for in the MSE by evaluating the difference between fishing mortality in the management plan forecast and fishing mortality realized in the operating models with Mohn's rho for fishing mortality only. Preliminary conclusions indicate that Mohn's rho may correlate with the differences between target F and realized F and can be accounted for by an additional alpha parameter in the simulations when there is a persistent bias, but otherwise suggested that the alpha parameter is too blunt for dealing with interannual variation. Further, this work demonstrates that MSE can account for retrospective pattern in assessments by adjusting control parameters. The retrospective pattern observed in recent assessment of North Sea cod disappeared as the MSEs progressed, even when the OM included survey year effects that were not modelled in the management procedure. It was recommended that future work should continue to improve the modelling of retrospective patterns in MSEs.

Retrospective Bias in Some Short-lived North Sea Stocks

Mikael Van Deurs

Terms of Reference: D and E

Abstract

Retrospective bias in the assessment of three short-lived species (one sprat and two sandeel) in the North Sea were examined among a suite of other model diagnostics to characterize the magnitude of bias in the assessment and implications on forecast. Rho adjustments on assessment estimates were examined by accounting for the life history and longevity of these species. A Management Strategy Evaluation was conducted to analyse the impacts of errors of management performance that are robust to the errors in the assessment. A number of recommendations were developed as general guidelines to consider for dealing with bias in the assessments for short-lived species. 1) Retrospective analyses should be examined for both SSB and recruitment to account for the longevity of these species. Sandeel has a life span of 3+ while Sandeel is approximately 4+. 2) It is important not to judge Mohn's rho alone. As seen in Sandeel 1r, Mohn's rho was high while other diagnostics were reasonable. 3) MSE is important to determine if Mohn's rho is too large. If Mohn's rho is within the uncertainty envelope of the MSE or assessment model, analyst should employ expert judgement to determine the acceptability of the assessment. Analyst should keep in mind that the estimation model used in the MSE could also be prone to retrospective pattern and therefore there may be circularity issues with the interpretability of the results from the simulation. 4) It is important to decide on a suite of useful diagnostics plots for the expert group to help with providing guidance with decision about the quality of the model. 5.) Sometime Mohn's rho could be changing over time and therefore care should be taking for adjusting Mohn's rho when forecasting the future. 6.) Consider adopting a category 2 assessment rather than downgrading to category 3 and above or rejecting an assessment for cases when the analytical assessment suffers from a strong retrospective pattern, but other model diagnostics appear reasonable with a believable trend.

Summary and Discussions

This study explores the notion that if consideration is given to a suite of model diagnostics instead of just Mohn's rho when judging the sufficiency of an analytical model for advice, it may change our perception about the acceptability of an assessment. Three short-lived North Sea stocks (Sprat and two Sandeels stocks) were used as a case study for this analysis. A number of interventions to the model formulation were considered, including the introduction of a power model for survey to resolve retrospective patterns for these stocks. The results do show a fair amount of Mohn's rho above Hurtado-Ferro's rule of thumb in an MSE without model misspecification. Additionally, it was pointed out that rho statistics alone were less informative than series of peels. Evaluation of residual patterns show some patterning but nothing substantial or alarming. One of the lessons learned from this study is not to judge a model solely based on Mohn's rho. Sandeel illustrates an example of a stock assessment with high Mohn's rho but other diagnostics were reasonable. For short-lived species, it was suggested that first year rho for recruitment can be used for dealing with recent recruitment estimates, especially for large year classes. It was also recommended to use MSE to determine if Mohn's rho is too large. If rho is within what was observed in the MSE or within the uncertainty envelope of the assessment model, it should already be accounted for within the bounds of uncertainty and may not be problematic. However, one should keep in mind that when applying MSE's, estimation models used within the MSE simulation may exhibit retrospective patterns too and should be carefully vetted. Due to occasional non-stationarity of rho, it is recommended to be careful to adjust with Mohn's rho (which track past patterns), when forecasting the future. Rho adjustment should not only rely on Mohn's rho statistics but should be done on a stock-specific basis (i.e. consideration for life history characteristics and stock dynamics).

Irish Sea Cod Retrospective Bias

Pia Schuchert and Mathieu Lundy

Terms of Reference: D and E

Abstract

The assessment of Irish Sea cod uses a statistical catch-at-age model (ASAP). Isolated, sporadic recruitment events in recent years, most notably in 2013, contributed to a marked stock size increase from a historically low stock size. However, abundance of cod is still very low and displays a very steep age profile with an age 7+ group.

Recent assessments show a strong retrospective bias with the SSB being revised downwards and F upwards. Mohn's Rho for the 2019 assessment was around 90% for SSB and F and hence the stock was downgraded by ICES to a category 3 stock.

In an attempt to reduce the bias, various options were explored by the working group, such as further investigations of survey catchability and changing assumptions of natural mortality. A recent study suggests that there is a migratory pattern out of the Irish Sea and mixing with the adjacent Celtic Sea stock. Including this assumption into the assessment reduced the Mohn's Rho, but the bias is still evident.

Here we present an MSE (management strategy evaluation) to further the understanding of the bias. MSE was performed in a hindcast manner, i.e. the MSE was projected to 2025 starting with assessments for years 2013, 2014, 2015, 2016, 2017 and 2018 - the stock-structure up to the starting year are regarded as perfectly known.

4 management strategies were then explored

- Catches at FMSY (= $F_{MSYupper}$)
- Catches at $F_{MSYlower}$

- Catches following the 2 over 3 rule (as done for category 3 stocks)
- Known catches (up to 2018), TAC (2019 and 2020) and zero thereafter.

SSB, F and the probability of the stock falling below biological reference points was explored. A similar metric to Mohn's Rho was calculated between the forecasts and the actual "known" stock structure in 2018. The assessment and forecasts are shown to be strongly dependent on rare stochastic recruitment events and mortality reducing cohort size in the following years when fish do not appear in the older age classes. Risk metrics of the unsustainable low stock size are high for all assessment years under the MSY scenarios, even for years projecting a high future SSB.

Simulation through an MSE procedure could provide a framework future catch advice by providing more precautionary F reference points. For stocks with high retrospective bias, MSE may provide an alternative precautionary framework on which to base forecasts whilst utilizing all sources of information and stocks in present condition.

Summary and Discussions

The stock was benchmarked in 2017 using ASAP as stock assessment model (Legault and Restrepo, 1999). The assessment is tuned to four different surveys with age-0 recruitment in the formulation. There are some notable contradictions in the survey trends particularly in the recruitment time-series. The ASAP model exhibited strong retrospective patterns (SSB rho = 0.9) and was downgraded to a category 3 assessment. Three MSE scenarios were evaluated in a4a based on the following scenarios, including 1) A scenario with known catches 2) a precautionary approach using BRPs from the last benchmark and 3) the ICES index-based 2 over 3 approach. Trends in estimated SSB consistently increased in all cases despite the strong retrospective patterns. Due to the consistent trends in SSB, it was recommended that the stock may be a good candidate for category 2 assessment. Recruitment for the stock has been very low and there may be an M or migration issue. Spatial analyses based on temperature seems to indicate that young fish may be more adaptable to handle warmer temperatures while older fish are emigrating to colder suitable habitats, which could possibly be source of the unexplained bias in the model. To better understand the retrospective diagnostics for the stock, it was recommended that age-specific Mohn's rho be calculated to help further investigate where things may be going wrong in the model.

Summarizing the Scope of Work and the Targeted TOR for Southern hake Stock Assessment with GADGET

Santiago Cerviño

Terms of Reference: B

Abstract

Southern hake assessment is a length-based model developed in GADGET (https://github.com/Hafro/gadget). The retrospective analyses in the most recent assessment indicate a strong Mohn's rho in recruitment (-1.05), fishing mortality (0.31) and SSB (-0.45). This assessment provides a case study at the workshop to explore and test alternative model assumptions and configurations to resolve the retrospective bias in the assessment. Exploratory work to be considered at the workshop include: 1.) Examine changes in biological parameters based on updated life history analyses. Such as alternative assumptions in natural mortality, including externally estimated age dependent M, changes in growth parameters and Sex-dependent model configuration to account for sexual dimorphism. 2.) Explore changes in fishery selection pattern

by exploring alternative time blocks and fleet disaggregation and 3.) Changes in catchability assumptions based on density-dependent assumptions I place of the existing linear catchability assumptions in the calibration model.

Summary and Discussions

This analysis explores and tests alternative model configurations with the objective of categorizing potential causes of retrospective pattern in the model. 22 different model formulations were evaluated based on changes in growth, natural mortality, fishery selectivity blocks, catchability and a leave one data out analysis. When catch and size distribution data were dropped, there were convergence issues with the model. Although there were large reduction in Mohn's rho with the exclusion of survey data, the lack of survey information to anchor the fishery dependent was not considered desirable but indicates that there may be some internal data conflicts causing some of the unexplained retrospective pattern in the model. Other model interventions such as changes in M, K, catchability and selectivity blocks in the fishery showed no improvement in retrospective pattern. At the workshop it was recommended to look at data themselves for conflicts. A simple evaluation is to look at total Z from catch curves. If Z is decreasing from the fishery catch but surveys imply that Z is increasing, then it shows a data conflict in the raw data. Follow-up analyses during the workshop revealed that there were conflicting signals in the surveys, especially for middle sized fish as well as possibility of missing recent catch. Mohn's rho was reduced by increasing catch since 2010 by 40%. A request was made to WKFORBIAS to recommend the stock for an interbenchmark. WKFORBIAS felt that the workshop was not the appropriate venue for such recommendation and that this will have to go through the ICES process for calling for an interbenchmark. Further, WKFORBIAS felt that it is hard to justify changes in the survey, but missing catch is likely the culprit. Some jitter analyses were considered to check for model stability but found different solutions with different starting points with all having worse likelihood than the original solution. Retrospective patterns were not looked at following the jitter analyses and was recommended to be explored in the future.

Successes and Failures in the Daily Fight to Stock Assessment biases: Experience from an ICES assessment Working Group.

Valerio Bartolino

Terms of Reference: A and B

Abstract

In its essence, stock assessment aims to provide unbiased estimates of the abundance and demographic structure of fish stocks. This is routinely done by ICES assessment working groups, including HAWG among those.

Despite the long experience, availability of high quality time-series and availability of state of the art assessment models, fighting against biases remains one of the main challenges of the herring, sprat and sandeel stocks assessed by HAWG across the western Baltic, the North Sea, and the Celtic Sea ecoregions.

This wants to be a brief summary of some recent patterns in the retrospective analyses of HAWG stocks, including solutions that have been found by the group, cases that remained unsolved and new emergent issues.

Summary and Discussions

This summary provides an overview of some of the challenges and lessons learned by the herring assessment working group (HAWG) in attempt to reconcile the issue of retrospective patterns. A survey was sent out to the HAWG stock assessors inquiring information about retrospective patterns during the last five years in the assessment or forecast. Some of the survey questions included which parameter (F or SSB) exhibited the retro, how did the retrospective patterns relate to the uncertainty envelope from the model estimates, what was the history of the retrospective pattern (i.e. gradual or sudden) and if there were any trends, whether there has been any success in identifying the cause of the retrospective pattern and finally, if there has been a benchmark to try to resolve the retrospective pattern. Several potential reasons for the retrospective patterns were discussed and offered some insights in attempt to reconcile the retrospective issue. Some of the reasons identified by the WG include:

1. Understanding the link between catchability and high densities. The variability in survey observations does not always relate to cohort strength and also there is the issue of density-dependent mortality as a potential factor. Retrospective patterns were identified to be affected by strong year classes detected in the survey, especially for short-lived stocks. For the short-lived stocks, the retrospective patterns typically show up initially in recruitment but then quickly transfers to spawning stock biomass.
2. Stock containment and catchability of older ages. Changes in survey availability that does not adequately capture the entire stock is another potential factor for unexplained retrospective patterns. Some approaches in the past is to down-weight some of the years via CV, but again this is a very subjective approach. Binding selective pattern or variance parameters across ages has known to have big impacts on retrospective patterns and model fits.
3. Introduction of selectivity blocks typically has helped but sometimes leads to increased retrospective patterns.
4. Seasonality in catches - Moving of small catches between quarters resulted in a few cases with no impact on model results but tend to alleviate Mohn's rho. While this is an unconventional approach to resolving retrospective patterns, it is likely that selectivity was not well estimated for the small catches in their original designated season.
5. Natural mortality – Using multispecies models to provide external estimates of M relies on models that do not receive the same amount of scrutiny as single stock assessments. However, it is noted that changes in the multispecies model inputs could have considerable impacts on the scaling of M.
6. Retrospective patterns that are transient – During the 2014 benchmark for Celtic Sea herring, a more flexible model was adopted (SAM like) but was not able to resolve the retrospective patterns. In 2015, an interbenchmark resulted in a change in the model to ASAP. While initial analyses indicated a reduction in retrospective pattern, a change on one age group resulted in the increase in the retrospective pattern and remains unresolved.
7. Trends in growth cause problems for short-lived stocks. There have been questions by the WG of how to capture such trends in the forecasts. At the moment, there has been no analytical approach for doing this other than just to reduce the number of years used in the averages.

It was noted in WKFORBIAS that some retrospectives just affect the last few years and some impact the whole time-series. The number of years to consider in a retrospective pattern is important to capture where the issues may arise in the model. It was noted that a 5- or 7-year retrospective peel would simply not capture the issue.

Clarification on the Calculation of Mohn's rho for Iberian Sardine during WGHANSA2018.

Laura Wise and Hugo Mendes

Terms of Reference: A and E

Abstract

In 2018, ICES assessment working groups were requested to report for potential retrospective problems in category 1 and 2 age-based fish stock assessments. The suggested procedure on how to quantify and assess the direction of the parameter bias uses an approach based on the Mohn's rho value. During WGHANSA 2018, and for the Iberian sardine assessment, the Mohn's rho was estimated at SSB = 0.39, F = -0.18 and Recruitment = 0.49. The retrospective pattern in these key parameters was considered to be above a sensitive threshold prompting an interbenchmark to address this issue and a recommendation to further investigate the severity of the stock retrospective pattern during WKFORBIAS. While investigating the causes that lead to the bias it was identified a mistake in the methodology for the Mohn´s rho estimation. With calculations corrected, the Iberian sardine stock assessments revealed a much lower Mohn's rho values for the key parameters (SSB = -0.02, F = 0.02, R = 0.04), suggesting no retrospective problems for this stock. This analysis shows that minor alterations in the methodology and interim year assumptions can lead to differences in the 5-year average of Mohn´s rho values, which was the ICES proposed method to assess for retrospective problems.

Summary and Discussions

In Stock Synthesis, the fleet ID sign for the catch at age for the terminal year of each peel was inadvertently set incorrectly. Instead of setting the fleet ID to a negative when the observation is processed and its expected value and log likelihood is calculated but does not contribute to the total log likelihood, the fleet ID sign was set to positive. The resulting reduction in Mohn's rho from the adjustment was discussed at WKFORBIAS. The adopted adjustment to the model ID was considered appropriate and should be the basis for future retrospective analyses.

Retrospective Bias Thoughts, Examples, Ideas and Solutions

Jonathan White, Hans Gerritsen, Colm Lordon and Michaël Gras

Terms of Reference: B and Ci

Abstract

Retrospective bias patterns are apparent in many assessment of fisheries stocks. Bias comprises a shift in the estimated values of variables with the addition of a new incremental time step (usually year) in the data. Retrospective bias takes two forms, Analytical and Advice. Analytical bias results from a single data series, by estimating variables with reducing time steps. This often forms part of the quality control checking of an assessment. Advice bias is a shift in estimated variables in the scientific advice that has been provided to management. The later may include minor shifts with applied data and includes the assumptions made for the intermittent year while the former results from a single data series. The potential causes of bias are explored. The indicator Mohn's Rho may be considered useful in instances of bias commonly occurs in the

same direction, however it will not show issues within assessments where positive and negative annual signals are averaged out in the calculation of the indicator. Examples of Mohn's rho and bias are provided indicating potential causes, with discussion on solutions and a hit list of proposed recommendations in reducing assessment and advice bias.

Summary and Discussions

Retrospective pattern Prevention 101: Ideally, the best solution for dealing with retrospective pattern, is to avoid it in the first instance. If possible, develop/ choose an assessment that does not show retrospective pattern.

Consideration: the two types of retrospective patterns

"Advice" retrospective pattern

- Differences among time-series of variables (F, SSB, Recruitment) in the given advice
- Based on different datasets (which may include differences in the historical period), abrupt changes following benchmarks (data, model choice and settings)
- Also known as historical retrospectives

"Analytical" retrospective pattern

- Differences among time-series of variables (F, SSB, Recruitment) in an assessment
- Based on a single dataset
- Also known as within-model retrospectives

Retrospective Pattern in Review

- Are estimates outside confidence intervals – are they reasonable
- Mohn's rho, as applied, is an averaging, and therefore maybe low if positive and negative patterns balance
- A low Mohn's rho, may be hiding more notable patterns, for instance within age classes
- Recruitment
    - In many instances is erratic,
    - Forecasting of recruitment is not simplistic and rarely are we able to estimate with any reasonable accuracy
    - Recruitment – not critical? As it's not a direct measure
- Pattern on F and biomass (population (stock) numbers) is the critical side
- With a short time-series it could be difficult to determine the relative pattern.

Retrospective Pattern Review – Causes

- Large and sudden changes in survey or catch trends can give rise to retrospective pattern
- In cases of short time-series, an additional, new year's data will have a larger weighting on retrospective pattern and Mohn's rho estimate
- Differences based upon model uncertainty XSA retrospective *vs*. SAM retrospective may be very different, with SAM incorporating process error and using this to account for some of the pattern
- Data:
    - Poor catch reporting
    - Broad stroke discard estimates
    - Broad stroke application of age profiles to catch
    - Expected age class catch "missing" from commercial

    - Poor agreement among survey/ commercial indices and catch
- Migration
- Natural mortality – unaccounted for/ trends (real and/or estimated) – where can a model put mortality, we have no real estimate of it
- Spawning stress, egg/ juvenile survival, probably not parasite or disease

Retrospective pattern – A way forward

Investigate which stocks do not have notable retrospective pattern, and why:

- Haddock Iceland (5)
- Cod Iceland (5)
- Sole assessments (Celtic Sea)
- Northern stocks
-

What they have in common:

- Good contrast in age classes – more age classes for models to "lock on" to
- "Good data" – clean catch, sampling, raising, discarding, indices, age-length profiles

Solutions to retrospective pattern – compile a 'hit list'

- Come up with assessments that don't have retrospective pattern, therefore no need to correct
- Assess on single identified stocks not 'multiple' disaggregated stocks (herring 6a7bc)
- Forensically work through data to figure out what data are causing retrospective pattern
- Consider their credibility/ provenance of data causing retrospective pattern
- Single (tuning) fleet runs – leave all but one out runs, see how each tuning fleet affects retrospective pattern.
- Do tuning fleets indicate the same trends?
- Truncate time-series.
- Smooth noise that is not certain.
- Examine at catchability in surveys.
- Transfer of age profiles to unsampled catch – spatial and temporal.
- Raisings in catch data relative to discards.
- Cut time-series down, including tuning fleets to allow step changes.
- Testing with MSEs.

If assessment adjustments are to be applied

- Are they credible or fudge factors?
- Keep them simple (there is already a heavy overhead on Stock Coordinators)

When Model-based Stock Assessments rejected for use in Management and what Happens.

Punt et al. (Published in Fisheries research in 2020)

Presented by Chris Legault

Terms of Reference: Cii

Abstract

Model-based stock assessments form a key component of the management advice for fish and invertebrate stocks worldwide. It is important for such assessments to be peer-reviewed and to pass scientific scrutiny before they can be used to inform management decision making. While it is desirable for management decisions to be based on quantitative assessments that use as much of the available data as possible, this is not always the case. A proposed assessment may be found to be unsatisfactory during the peer-review process (even if it utilizes all of the available data), leading to decisions being made using simpler approaches. This paper provides a synthesis across seven jurisdictions of the types of diagnostic statistics and plots that can be used to evaluate whether a proposed assessment is 'best available science', summarizes several cases where a proposed assessment was not accepted for use in management, and how jurisdictions are able to provide management advice when a stock assessment is 'rejected.' The paper concludes with recommended general practices for reducing subjectivity when deciding whether to accept an assessment and how to provide advice when a proposed assessment is rejected.

Summary and Discussions

There are different experiences around the world why models are rejected for use in management. The peer review process is a very important step in the stock assessment process and often involves subjective decisions and there are no standard rules for model acceptance/rejection. This study examines stock assessment processes for seven jurisdictions, compiles case studies, looked for commonalities and provided tentative guidance on best practices. The reality is that rules for rejections are not stated or standardized. A number of reasons for rejecting a model include but are not limited to errors in basic data, unreasonable parameter estimates and model outputs, inability to determine scale of fishing mortality and biomass, extreme sensitivity to changes in parameter, conflicts in data and poor retrospective diagnostics. When a model is rejected, the fallback position is sometimes pre-determined and other times it is ad-hoc. Other fallback positions include accepting status quo advice or last agreed model with data update. Downgrading to simpler models or immediate work to improve model have been some other identified fall back positions. It is unlikely that there will ever be a set of rules that determine whether a stock assessment, particularly a complex stock assessment based on many assumptions and datasets, should be accepted or rejected that can be applied automatically (and it is not clear that such rules would even be desirable). However, the review offers some thoughts on best practices including:

1. Increased use of external peer reviewers
2. More holistic evaluation across model diagnostics for acceptance/rejection instead of focusing on single diagnostic
3. Partial acceptance/rejection instead of all-or-nothing approach
4. Identification of metrics and thresholds to summarize diagnostics
    a. Failing a threshold would not necessarily mean overall model is rejected
    b. Provides common basis for review and increase consistency
5. Develop standards for model output

6. Document recommendations from peer review and subsequent responses
7. Create easily accessible and machine readable stock assessment results
8. Develop fall-back positions in advance.

3 WKFORBIAS Conclusions

1. Retrospective patterns are an indication of inconsistency in the data or model, or both.
2. Identifying the important population processes that vary over time when addressing retrospective patterns should be a priority.
3. Mohn's rho is one of many model diagnostics. Removing the retrospective pattern does not imply the model is always sufficient for providing catch advice nor should the model necessarily be rejected if there is a notable Mohn's rho.
4. Mohn's rho is sufficient for measuring the magnitude of retrospective patterns, but more suitable for when the peels are directionally consistent. However, in cases when the peels are not unidirectional (i.e. mix of both positives and negatives), each peel should be evaluated closely.
5. In the event a model exhibits a major retrospective pattern (Brooks and Legault, 2016), Mohn's rho adjustment to the terminal year estimate performs better than not adjusting. However, Rho adjustment does not completely resolve the retrospective pattern problem. It is recommended to incorporate additional 'unmodeled' uncertainty in future HCRs.
6. Mohn's rho adjustment should not only rely only on the magnitude of Mohn's rho, but should be done on a stock-specific basis (i.e. consideration for life-history characteristics and stock dynamics, stock assessment model, to determine the criteria for adjustment). Additional investigation is warranted to further define adjustment approach(es) and criteria to account for differing systems, life-history types and model applications
7. When forecasting, it is recommended to use recent recruitment trends in longer-term forecast, especially for stocks with recent low recruitment, when low recruitment is expected to continue in the projection period.
8. Short-term forecast is highly recommended as long-term forecasts tend to be less reliable.
9. The terminology of retrospective bias and correction and the implicit logical fallacy are misleading to fishery managers, who are compelled to make decisions based on the best available science. If a retrospective inconsistency is treated as a diagnostic and an indicator of scientific uncertainty, retrospective adjustment can be performed for meeting management objectives. The adjustment should be considered precautionary to managing in the context of scientific uncertainty.
10. The moving window approach used to describe temporal changes in retrospective patterns offers a promising approach to evaluate systematic issues that may be a source of the retrospective pattern in the model.
11. The application of state space models demonstrates the potential to reduce retrospective pattern by modelling correlated process errors. However, based on this workshop, this approach comes with caveats. Process errors tend to be confounded with mortality (particularly for an assessment with a single survey), or with misspecification of the observation model (i.e. occasional shifts in catchability).
12. The Rose approach which uses a model ensemble framework offers a promising approach to addressing retrospective patterns. One advantage of the method is the potential to address model misspecification while avoiding the assignment of blame to the wrong model assumption or data. The drawback to this method is the difficulty in determining the rules ahead of time for which models to include or exclude in the ensemble approach based on model diagnostics. Further, this approach could be time consuming but it was suggested at WKFORBIAS that methods such as local influence analyses could be explored as a mechanism for determining which models to include in the ensemble.

13. Management Strategy Evaluations (MSEs) offer a potential approach to evaluate the impact of consistent retrospective patterns on any given HCR for a particular stock. However, simulating specific retrospective patterns (e.g. developing operating models that produce a similar Mohn's rho in terminal years of an estimation model that persists during projection years) is challenging and involves an understanding of model misspecification, the nature of the misspecification, multiple plausible misspecifications (as in the Rose approach) and continued misspecification during the projection period.

4 WKFORBIAS Recommendations and Decision Tree

1. Always check for the presence of retrospective pattern as a diagnostic tool. The retrospective analysis should extend as far back into the time-series as possible, so that each retrospective analysis has similar information available for parameter estimation (e.g. time-series of stock indices and age compositions are long enough to be informative).
2. A retrospective pattern should not be interpreted as a bias because the true values are not known (only in simulation are true values known); instead it should be called a retrospective pattern. Likewise, any changes made should be called a retrospective adjustment (as opposed to a correction).
3. Determine if a stock assessment exhibits a major or minor retrospective pattern. Complete 5-year analytical peels. For stocks that do not have thresholds defined through stock-specific simulation analysis (see research recommendations below), a major retrospective pattern would be indicated by rho >0.2 or <-0.15 for long-lived stocks, >0.3 or <-0.22 for short- lived stocks, or 2/3 or 3/5 years with peels outside of the confidence intervals following Hurtado-Ferro *et al.* (2015) (See decision tree chart below). Mohn's rho should not be evaluated in isolation; instead WGs should consider additional diagnostics such as, but not limited to, residual patterning, convergence diagnostics, fits to indices, correlations between estimated parameters, parameter boundary issues, likelihood profiling, and an extended series of retrospective peels (to examine patterns in the context of stock trends or changes in data, model assumptions, fisheries). Examine retrospective analysis of several estimates (Spawning Stock Biomass, Fishing Mortality and abundance-at-age).
4. In the event a model exhibits a major retrospective pattern, consult the retro action list (see below) to evaluate possible causes and modelling resolutions.
5. When a major retrospective pattern results in estimates of SSB being consistently being overestimated and F is consistently underestimated, it is recommended to adjusting catch advice (Figure 4.1).
6. Make use of decision tree (Figure 4.1) as guidance for identifying a mechanism for providing advice.
7. If there is a major retrospective pattern with robust temporal trends in the assessment quantities across the full time-series derived from an age-structured assessment but only the scale is uncertain, then an assessment should be downgraded to a category 2 assessment instead of a category 3 assessment if possible. This preserves the use of composition data to inform population trends.
8. It was discussed during WKFORBIAS that it has not been common practice to downgrade an assessment to category 2 due to limited guidance on how to provide advice based on category 2 assessment. Instead, majority of ICES assessments that are downgraded tend to downgrade to a category 3 assessment, resulting in a loss of important demographic data. As such, WKFORBIAS recommends that the basis for a category 2 assessment advice needs to be revisited and if necessary, further improved for future applications.
9. When designing an MSE, for each candidate HCR, varying levels of retrospective inconsistency can be simulated through model-misspecification to evaluate the level of retrospective pattern to which the HCR is precautionary. This can then be compared to the retrospective pattern in the assessments to evaluate if the HCR remains precautionary. However, such MSEs require an understanding of plausible sources of retrospective patterns.

10. Continue research on appropriate diagnostics for classifying the causes of a retrospective pattern.
11. Continue work on the use of state-space models for minimizing retrospective patterns through the incorporation of process error (independent and auto-correlated). Process error may not be able to reconcile model misspecification that involves major structural error.
12. Continue to develop a framework/rubric for determining the appropriate rule of thumb for Mohn's rho for a particular stock (i.e. determine what level of retrospective pattern is acceptable). It should be noted that stock specific rule-of-thumb should account for the effect of data quality/quantity, species life history and modelling approach. Any change in these sources (future assessments) could potentially lead to revising the rule-of-thumb for that stock of given interest. This implies that the rule-of-thumb might not be a static and should be re-examined whenever possible in the event of big changes occurring in data collection, or model.
13. Continue to explore suitability of potential methods and criteria for applying a retrospective adjustment (apply some precautionary reduction (e.g. 75% or 50%) to the resulting catch advice). ICES may want to consider forming a study group for these analyses
14. Continue to improve the modelling of retrospective patterns in MSEs . WKFORBIAS also recommends evaluating the potential for using Eqsim to produce a similar evaluation of the degree of additional bias that the F_{MSY}/B_{pa} combination is robust to. Where there are retrospective patterns in the recent assessments, consideration should be given to including this in the MSE simulations. This would ensure that the resulting HCR is optimized to use in an assessment model with this level of retrospective pattern. However, such MSEs require an understanding of plausible sources of retrospective patterns.
15. WKFORBIAS recommends to replace the figure of the historical retrospective in the ICES Advices on fishing opportunities with a figure of the analytical retrospective which appears more appropriate for the section "Quality of the assessment". Because clients and readers of the ICES advice may still have an interest in the historical retrospective, a simple text with a permanent link to the historical retrospective should be included.

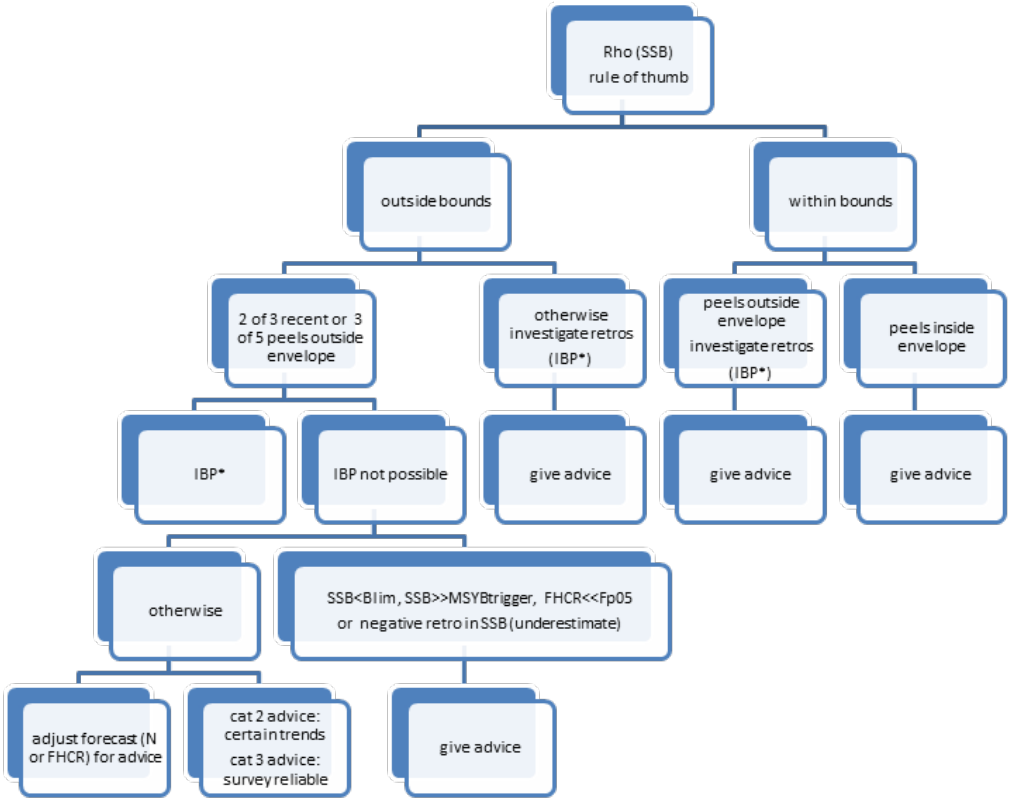


Figure 4.1. Decision tree for handling assessments with retrospective patterns. In the case of an Interbenchmark Protocol (IBP), repeat the process of checking for retrospective patterns as outlined in this decision tree.

5 WKFORBIAS Action List: Guidelines for investigating a retrospective pattern

When an assessment model has a retrospective pattern, it is recommended to generate a list of potential causes, ability to resolve with the data, and the anticipated effect:

- When retrospective peels are directionally consistent, the direction of Mohn's rho on SSB could suggest the direction that the causative mechanism is operating.
- Consider whether a stock specific table could be put in the advice sheet about which of these causes are more/less likely (probably only a subset of this table for any given stock)
- Look at stock specific data inputs to decide which rows belong in the table

Example (red font indicates longer-term activity)

General Model diagnostics and sensitivity analyses:

- Look at relative F (catch biomass/index biomass) vs catch curve Z
- Look at time-series of residuals
- Explore fits to one survey at a time or one fleet at a time
- Explore truncating data series (perhaps drop poorly known historical) or treat "bad data" year as missing
- Look at moving window of rho to identify changes in the pattern, possibly relate it to data or structural issues

Retrospective Cause	What to look at in Data?	How to resolve Data issue?	Model resolution?	What was attempted/ other comments?
Change in M	Disease prevalence Condition factor Lack of consistency in age comp	Tagging data Predation data Catch curve in survey (when F~0) High resolution data	Estimate year where M changes Estimate age(s) where M changes Likelihood profile over M breakpoints *Fix M in age/year blocks to remove retro*	..
Migration	Lack of consistency in age comp	Tagging data Otolith trace element Genetics Parasites	Spatial model Add to M?	Do fish come back? Does it require changing stock area definition
Stock mixing	Lack of consistency in age comp	Tagging data Otolith trace element Genetics Parasites	Spatial model	Requires benchmark to redefine "stock" Does it require changing stock area definition What is temporal pattern of mixing
Missing landings	Lack of consistency in age comp	Improve 'traceability' from boat to dock to dealer to market Increase Port sampling Recreational sampling	Increase landings data to remove retro Try to estimate landings scalar	
Missing discards	Look for "funky" patterns (spatial, species comp) Lack of consistency in age comp	Increase observer coverage/ EM Explore discard mortality assumption	Increase discards data to remove retro Try to estimate discard scalar	
Survey change	Vessel/gear change Availability change (thermal/habitat/salinity) Time-series gaps, vessel swap Timing of survey	Do calibration study	Estimate changes in q Allow survey selectivity to change? Estimate conversion factors between vessels/gear Start new time-series (or partition it)	
Fishery operation change	Look at management changes (min size, closed areas) Look at number of vessels	…	Add selectivity block/new fleet/catchability	

Retrospective Cause	What to look at in Data?	How to resolve Data issue?	Model resolution?	What was attempted/ other comments?
	Changes in gear (e.g. bycatch reduction) Look at spatial effort and targeting Changes in fleet		Spatial/temporal component (fleets as areas)	
Ageing error	Missing ages in age comp Look at age-length key for abberations	Validation and QA/QC tests Inter reader comparisons	Include ageing error in model Move to length based model	
Recruitment	Examine retro pattern across younger ages Evaluate noise in data for recruitment age	Initiate recruitment survey Look at maturity data: Is there skipped spawning? Staging errors?	If recr. age is too noisy, specify older recr age Don't look at rho in SSB if maturity is suspect Autocorrelated recruitment PE *Include environmental covariate in SR (difficult to forecast, likely transient*	Maybe don't react to retro if it is strong yc passing through. However, continue to monitor retrospective patterns pattern, especially when the strong YC begins to recruit into the fishery.
Biological Inputs	Check growth eqn (length-age and L-W); change in seasonality sampling? Check maturity eqn Look at sensitivity of smoother Look for retro on modeled inputs	Revise sampling	Try integrated model	Short term fix:

6 References

Brooks, E. N., and Legault, C. M. 2016. Retrospective forecasting – evaluating performance of stock projections for New England groundfish stocks. Canadian Journal of Fisheries and Aquatic Sciences, 73: 935–950.

Deroba, J.J., 2014. Evaluating the consequences of adjusting fish stock assessment estimates of biomass for retrospective patterns using Mohn's rho. North American Journal of Fisheries Management, 34: 380–390.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., Ono, K., Vert-Pre, K. A., Whitten, A. R., Punt, A. E. 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 72: 99–110.

Legault, C. M. 2009. Report of the Retrospective Working Group, 14–16 January 2008, Woods Hole, Mass. US Department of Commerce Northeast Fisheries Science Center Reference Document 09-01.

Legault, C. M., and Restrepo, V. R. 1999. A Flexible Forward Age-Structured Assessment Program. ICCAT Collected Volume of Scientific Papers, 49: 246–253.

Miller, T. J., and Legault, C. M. 2017. Statistical behavior of retrospective patterns and their effects on estimation of stock and harvest status. Fisheries Research, 186: 109-120.

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56: 473–488.

Punt, A. E., Tuck, G. N., Day, J., Canales, C. M., Cope, J. M., de Moor, C. L., De Oliveira, J. A. A., Dickey-Collas, M., Elvarsson, B., Haltuch, M. A., Hamel, O. S., Hicks, A. C., Legault, C. M., Lynch, P. D., and Wilberg, M. J. 2020. When are model-based stock assessments rejected for use in management and what happens then? Fisheries Research 224

Szuwalski, C. S., Ianelli, J. N., and Punt, A. E. 2018. Reducing retrospective patterns in stock assessment and impacts on management performance. ICES Journal of Marine Science, 75: 596–609.

Wiedenmann, J., and Jensen, O. P. 2018. Scientific uncertainty in assessment estimates for New England groundfish stocks and its impact on achieving target catches. Canadian Journal of Fisheries and Aquatic Sciences. 75: 342-356.

Wiedenmann, J., and Jensen, O. P. 2019. Could recent overfishing of New England groundfish have been prevented? A retrospective evaluation of alternative management strategies. Canadian Journal of Fisheries and Aquatic Sciences, 76: 1006-1018.

Annex 1: List of participants

Institute	Country (of institute)	Email
Fisheries and Marine Institute of Memorial University of Newfoundland	Canada	Andrea.perreault@mi.mun.ca
Northeast Fisheries Science Center	United States	Chris.legault@noaa.gov
International Council for the Seas	Denmark	Colin.millar@ices.dk
Institute of Marine Research	Norway	Daniel.howell@hi.no
International Council for the Seas	Denmark	ghislain@ices.dk
Marine Institute	Ireland	Jonathan.white@marine.ie
Northeast Fisheries Science Center	United States	Kiersten.curti@noaa.gov
Institute of Marine Research	Norway	Kotaro.ono@hi.no
Northeast Fisheries Science Center	United States	Larry.alade@noaa.gov
Instituto Portuges de Mar e da Atmosfera	Portugal	lwise@ipma.pt
Northeast Fisheries Science Center	United States	Liz.brooks@noaa.gov
Technical University of Denmark (DTU AQUA)	Denmark	mvd@aqua.dtu.dk
Centre for Environment, Fisheries and Aquaculture Science (Cefas)	United Kingdom	nicola.walker@cefas.co.uk
Fisheries and Marine Institute of Memorial University of Newfoundland	Canada	Noel.cadigan@mi.mun.ca
Agri-food and Biosciences Institute (AFBI)	United Kingdom	Pia.schuchert@afbini.gov.uk
Environmental Defense Fund	United States	rboenish@edf.org
Centre Oceanografico de Vigo	Spain	Santiago.cervino@ieo.es
SMAST – Dept. of Fisheries Oceanography	United States	scadrin@umassd.edu
Marine Scotland Science, Marine Laboratory	United Kingdom	t.miethe@marlab.ac.uk
Dept. of Aqautic Resources, Sweedish University of Agricultural Sciences	Sweden	valerio.bartolino@slu.se

Annex 2: WKFORBIAS Summary Draft Report

WKFORBIAS – Workshop on Catch Forecast from Biased Assessments
11-15 November 2019, Woods Hole, MA, USA
Chairs: Larry Alade (USA) and Chris Legault (USA)

Executive Summary

The workshop on catch forecasts from biased assessments, WKFORBIAS, met to address and develop general guidelines for dealing with the issue of retrospective patterns in stock assessments. A total of 16 working papers were presented during the workshop, addressing one or more of the terms of references. The Workshop reaffirmed previous recommendations that retrospective analysis should always be conducted as a diagnostic to examine the internal consistency of an analytical stock assessment. The Mohn's rho statistic that compares estimates from assessments with recent years of data removed to estimates from the current assessment is the standard tool for retrospective analysis. Examination of recent category 1 and 2 ICES stock assessments indicate that a majority of the assessments do not exhibit strong retrospective patterns that require changes to standard management advice. Across the wide range of stocks examined, no obvious explanatory variables, such as model type, location, fishery type, or biological trait, separated stocks with and without strong retrospective patterns. By comparison, both the magnitude and the proportion of stocks with retrospective patterns were greater in the Northeast US than observed for ICES stocks.

For the stock assessments that do show strong retrospective patterns, the first step was to identify what constitutes a strong retrospective pattern then a decision tree was developed to help expert groups determine a course of action. A number of general recommendations from WKFORBIAS include: 1) When evaluating a retrospective pattern, the consistency of the pattern is of primary importance. 2) A large Mohn's rho statistic driven by one outlier should not be treated in the same manner as a consistent directional retrospective pattern. 3) Retrospective pattern should be viewed as one of many diagnostics to be used in determining whether to use an assessment for management advice or not. 4) A strong consistent retrospective pattern can be the basis for adjusting catch advice or downgrading the level of an assessment. For stocks that exhibit a strong retrospective pattern where SSB is consistently overestimated and F is consistently underestimated, it is recommended to adjust catch advice. For less common retrospective patterns, where SSB consistently increases with additional years of data, then adjusting catch advice is not recommended. However, in the event an assessment is being downgraded, strong consideration should be given to maintaining models with robust trends even if the scale of the assessment is uncertain (i.e. downgrading from a category 1 to a category 2 assessment). 5) Management Strategy Evaluation can potentially be a useful tool for examining the robustness of harvest control rules to different magnitudes of retrospective pattern and could be useful for situations exhibiting strong retrospective patterns over multiple assessments. However, simulating specific retrospective patterns is challenging as demonstrated by the Rose approach. A complete set of recommendations are provided in the report including a compiled action list for consideration by expert groups to evaluate possible sources of the retrospective pattern.

WKFORBIAS Workshop Conclusions

1. Retrospective patterns are an indication of inconsistency in the data or model, or both.
2. Identifying the important population processes that vary over time when addressing retrospective patterns should be a priority.
3. Mohn's rho is one of many model diagnostics. Removing the retrospective pattern does not imply the model is always sufficient for providing catch advice nor should the model necessarily be rejected if there is a notable Mohn's rho.
4. Mohn's rho is sufficient for measuring the magnitude of retrospective patterns, but more suitable for when the peels are directionally consistent. However, in cases when the peels are not unidirectional (i.e. mix of both positives and negatives), each peel should be evaluated closely.
5. In the event a model exhibits a major retrospective pattern (Brooks and Legault, 2016), Mohn's rho adjustment to the terminal year estimate performs better than not adjusting. However, Rho adjustment does not completely resolve the retrospective pattern problem. It is recommended to incorporate additional 'unmodelled' uncertainty in future HCRs.
6. Mohn's rho adjustment should not only rely only on the magnitude of Mohn's rho, but should be done on a stock-specific basis (i.e. consideration for life-history characteristics and stock dynamics, stock assessment model, to determine the criteria for adjustment). Additional investigation is warranted to further define adjustment approach(es) and criteria to account for differing systems, life-history types and model applications
7. When forecasting, it is recommended to use recent recruitment trends in longer-term forecast, especially for stocks with recent low recruitment, when low recruitment is expected to continue in the projection period.
8. Short-term forecast is highly recommended as long-term forecasts tend to be less reliable.
9. The terminology of retrospective bias and correction and the implicit logical fallacy are misleading to fishery managers, who are compelled to make decisions based on the best available science. If a retrospective inconsistency is treated as a diagnostic and an indicator of scientific uncertainty, retrospective adjustment can be performed for meeting management objectives. The adjustment should be considered precautionary to managing in the context of scientific uncertainty.
10. The moving window approach used to describe temporal changes in retrospective patterns offers a promising approach to evaluate systematic issues that may be a source of the retrospective pattern in the model.
11. The application of state space models demonstrates the potential to reduce retrospective pattern by modelling correlated process errors. However, based on this workshop, this approach comes with caveats. Process errors tend to be confounded with mortality (particularly for an assessment with a single survey), or with misspecification of the observation model (i.e. occasional shifts in catchability).
12. The Rose approach which uses a model ensemble framework offers a promising approach to addressing retrospective patterns. One advantage of the method is the potential to address model misspecification while avoiding the assignment of blame to the wrong model assumption or data. The drawback to this method is the difficulty in determining the rules ahead of time for which models to include or exclude in the ensemble approach based on model diagnostics. Further, this approach could be time consuming but it was suggested at WKFORBIAS that methods such as local influence analyses could be explored as a mechanism for determining which models to include in the ensemble.
13. Management Strategy Evaluations (MSEs) offer a potential approach to evaluate the impact of consistent retrospective patterns on any given HCR for a particular stock. However, simulating specific retrospective patterns (e.g. developing operating models that

produce a similar Mohn's rho in terminal years of an estimation model that persists during projection years) is challenging and involves an understanding of model misspecification, the nature of the misspecification, multiple plausible misspecifications (as in the Rose approach) and continued misspecification during the projection period.

WKFORBIAS Workshop Recommendations and Decision Tree

1. Always check for the presence of retrospective pattern as a diagnostic tool. The retrospective analysis should extend as far back into the time-series as possible, so that each retrospective analysis has similar information available for parameter estimation (e.g. time-series of stock indices and age compositions are long enough to be informative).
2. A retrospective pattern should not be interpreted as a bias because the true values are not known (only in simulation are true values known); instead it should be called a retrospective pattern. Likewise, any changes made should be called a retrospective adjustment (as opposed to a correction).
3. Determine if a stock assessment exhibits a major or minor retrospective pattern. Complete 5-year analytical peels. For stocks that do not have thresholds defined through stock-specific simulation analysis (see research recommendations below), a major retrospective pattern would be indicated by rho >0.2 or <-0.15 for long-lived stocks, >0.3 or <-0.22 for short- lived stocks, or 2/3 or 3/5 years with peels outside of the confidence intervals following Hurtado-Ferro *et al.* (2015) (See decision tree chart below). Mohn's rho should not be evaluated in isolation; instead WGs should consider additional diagnostics such as, but not limited to, residual patterning, convergence diagnostics, fits to indices, correlations between estimated parameters, parameter boundary issues, likelihood profiling, and an extended series of retrospective peels (to examine patterns in the context of stock trends or changes in data, model assumptions, fisheries). Examine retrospective analysis of several estimates (Spawning Stock Biomass, Fishing Mortality and abundance-at-age).
4. In the event a model exhibits a major retrospective pattern, consult the retro action list (see below) to evaluate possible causes and modelling resolutions.
5. When a major retrospective pattern results in estimates of SSB being consistently being overestimated and F is consistently underestimated, it is recommended to adjusting catch advice (Figure A2.1).
6. Make use of decision tree (Figure A2.1) as guidance for identifying a mechanism for providing advice.
7. If there is a major retrospective pattern with robust temporal trends in the assessment quantities across the full time-series derived from an age-structured assessment but only the scale is uncertain, then an assessment should be downgraded to a category 2 assessment instead of a category 3 assessment if possible. This preserves the use of composition data to inform population trends.
8. It was discussed during WKFORBIAS that it has not been common practice to downgrade an assessment to category 2 due to limited guidance on how to provide advice based on category 2 assessment. Instead, majority of ICES assessments that are downgraded tend to downgrade to a category 3 assessment, resulting in a loss of important demographic data. As such, WKFORBIAS recommends that the basis for a category 2 assessment advice needs to be revisited and if necessary, further improved for future applications.
9. When designing an MSE, for each candidate HCR, varying levels of retrospective inconsistency can be simulated through model-misspecification to evaluate the level of retrospective pattern to which the HCR is precautionary. This can then be compared to the s retrospective pattern in the assessments to evaluate if the HCR remains precautionary.

However, such MSEs require an understanding of plausible sources of retrospective patterns.

10. Continue research on appropriate diagnostics for classifying the causes of a retrospective pattern.
11. Continue work on the use of state-space models for minimizing retrospective patterns through the incorporation of process error (independent and auto-correlated). Process error may not be able to reconcile model misspecification that involves major structural error.
12. Continue to develop a framework/rubric for determining the appropriate rule of thumb for Mohn's rho for a particular stock (i.e. determine what level of retrospective pattern is acceptable). It should be noted that stock specific rule-of-thumb should account for the effect of data quality/quantity, species life history and modelling approach. Any change in these sources (future assessments) could potentially lead to revising the rule-of-thumb for that stock of given interest. This implies that the rule-of-thumb might not be a static and should be re-examined whenever possible in the event of big changes occurring in data collection, or model.
13. Continue to explore suitability of potential methods and criteria for applying a retrospective adjustment (apply some precautionary reduction (e.g. 75% or 50%) to the resulting catch advice, apply Mohn's rho?). ICES may want to consider forming a study group for these analyses
14. Continue to improve the modelling of retrospective patterns in MSEs . WKFORBIAS also recommends evaluating the potential for using Eqsim to produce a similar evaluation of the degree of additional bias that the F_{MSY}/B_{pa} combination is robust to. Where there are retrospective patterns in the recent assessments, consideration should be given to including this in the MSE simulations. This would ensure that the resulting HCR is optimized to use in an assessment model with this level of retrospective pattern. However, such MSEs require an understanding of plausible sources of retrospective patterns.
15. WKFORBIAS recommends to replace the figure of the historical retrospective in the ICES Advices on fishing opportunities with a figure of the analytical retrospective which appears more appropriate for the section "Quality of the assessment". Because clients and readers of the ICES advice may still have an interest in the historical retrospective, a simple text with a permanent link to the historical retrospective should be included.

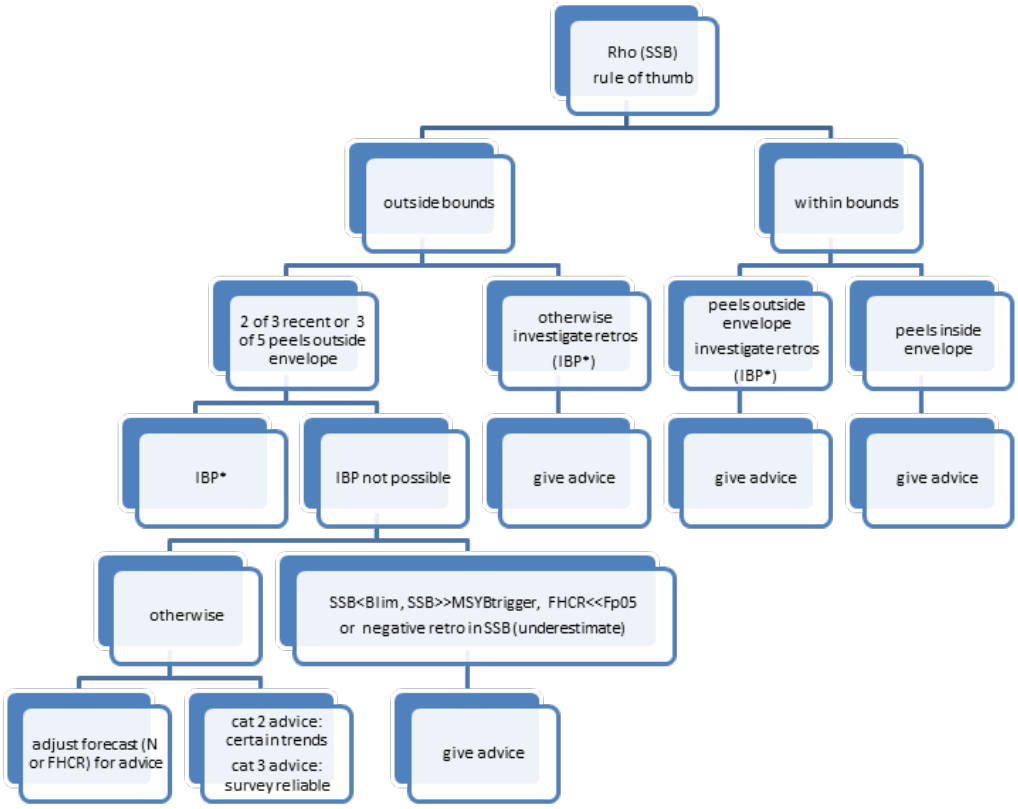


Figure A2.1. Decision tree for handling assessments with retrospective patterns. In the case of an Interbenchmark Protocol (IBP), repeat the process of checking for retrospective patterns as outlined in this decision tree.

Annex 3: Resolution

A prior version of this resolution was approved by ACOM in March but the Workshop was postponed and the ToRs have since then been modified.

2018/2/FRSG29 The **Workshop on catch forecasts from biased assessments** (WKFORBIAS), chaired by Larry Alade (USA) and Chris Legault (USA) will meet at Woods Hole, USA from 11–15 November 2019 to:

a) Document the extent and magnitude of bias as identified in ICES category 1 and 2 stock assessments through the retrospective analysis performed by the ICES expert groups in 2018
b) Categorize the potential causes for bias as identified in ICES stock assessments with respect to factors like stock longevity, quality of input data (catch and survey), model assumptions and environmental changes
c) Develop criteria for characterizing a major or minor bias through retrospective analysis as it pertains to:
    i) acceptability of an assessment,
    ii) consideration for correction (using the results of the retrospective analysis) for stock status determination and catch forecast, while taking into consideration other model diagnostics from the assessments.
d) Investigate through simulations the performance of using Mohn's rho to correct biased assessments for the purpose of making catch projections or, alternatively, to directly correct catch forecasts, particularly as it relates to the short and long-term tradeoffs in achieving MSY targets.
e) In cases where Mohn's rho can be used, develop criteria for its use including a description of the methodology for correcting population metrics (for example, indicate whether the correction should be applied to spawning stock biomass or numbers at age) or the catch forecasts
    i) Develop standard methods for the presentation of bias corrected values in advice sheets.
    ii) In cases where Mohn's rho should not be used, recommend other approaches.

WKFORBIAS will report by 6 January 2020 for the attention of ACOM.

Supporting information

Priority	High. A consistent approach in the methodology for catch projections when category 1 and 2 age or length-based stock assessments display bias is required.
Scientific justification	Provision of credible advice on fishing opportunities requires that advice is based on the best available unbiased estimates of population metrics. In some of the ICES assessments, there is evidence of some bias in category 1 and 2 age or length-based assessments and the approach followed in providing advice has varied between stocks. In some cases, the assessment is rejected with advice delayed and a benchmark (or interbenchmark) is organized to resolve the issue while in other cases the values are either used as is to provide advice in the category 1-2 framework or the population estimates are used as indicative of trends in a category 3 frameword. In this context, it is desired that a more consistent approach be defined when there is evidence of bias in these assessments. It has been proposed that a measure of bias in the form of the Mohn's rho be used to correct population

	estimates for the purpose of catch projections however it is considered that the approach needs to be thoroughly examined to determine whether it should be used or if ome other approach should be followed. The worshop will focus on first documenting the extent of the problem (ToR a) based on data produced by assessment working groups during 2018 as per a generic ToR on the calculation of Mohn's rho that was incorporated in their work deliverables. ToR b would be addressed by compiling the causes identified for retrospective bias by the assessment working groups. Investigation of these causes for some stocks would be desirable but is not the main focus of the workshop. Based on existing litterature, studies, general rules should be determined for accepting or rejecting stock assessments with retrospective bias (ToR c). It is expected that for some of the stocks where a significant bias was identified in 2018, some simulations could be conducted to determine whether a correction using Mohn's rho would have been likely to reduce the bias in catch projections (ToR d). In those cases, the methodology that should be used should be thoroughly described. If not, then other approaches should be considered (ToR e).
Resource requirements	Some support will be required from the ICES Secretariat.
Participants	We expect 10-20 attendees. It is expected that some of the participants would prepare case studies.
Secretariat facilities	None
Financial	
Linkages to advisory committees	ACOM
Linkages to other committees or groups	All stock assessment Working Groups
Linkages to other organizations	Relevant to all national laboratories contributing to ICES advice on fishing opportunities