Fisheries New Zealand
Tini a Tangaroa

2026 Independent International Review of Orange Roughy Stock Assessment Modelling

New Zealand Fisheries Science Review 2026/02

Burch, P.
Dorn, M.W.
Martell, S.J.D.

ISSN 2382-2406 (online)

ISBN 978-1-997337-64-5 (online)

August 2026

Te Kāwanatanga o Aotearoa
New Zealand Government

Disclaimer

Requests for further copies should be directed to:

Fisheries Science Editor

Fisheries New Zealand

Ministry for Primary Industries

PO Box 2526

Wellington 6140

NEW ZEALAND

Email: Fisheries-Science.Editor@mpi.govt.nz

Telephone: 0800 00 83 33

This publication is also available on the Ministry for Primary Industries websites at:

http://www.mpi.govt.nz/news-and-resources/publications

http://fs.fish.govt.nz go to Document library/Research reports

Please cite this report as:

Burch, P.; Dorn, M.W.; Martell, S.J.D. (2026). 2026 Independent International Review of Orange Roughy Stock Assessment Modelling. *New Zealand Fisheries Science Review 2026/02*. 62 p.

TABLE OF CONTENTS

Contents

PLAIN LANGUAGE SUMMARY

Orange roughy are among the world's longest-lived fish, reaching 120–150 years old, making their stock assessment uniquely challenging. New Zealand manages several orange roughy stocks using acoustic surveys and age-structured population models. This independent international review met in May and June 2026 to evaluate the science behind these assessments, examining data quality, model assumptions, and uncertainty. The panel found that key parameters — especially natural mortality and acoustic survey catchability — are highly uncertain and that this uncertainty is not fully reflected in stock status conclusions. The review provides recommendations for improving acoustic survey design, ageing data collection, model structure, and approaches for data-poor stocks, with the goal of building a more robust science foundation for future management decisions. The panel recommends that New Zealand orange roughy assessments move toward a tiered structure in which four assessment approaches are employed as appropriate to each stock's data situation: CASAL for data-rich stocks with acoustic and age data; surplus production and stock reduction analysis for data-poor stocks and as cross-checks for data-rich stocks; and management procedure evaluation as the overarching framework for translating stock status uncertainty into management advice. This is consistent with international best practice for data-limited and parameter-uncertain fisheries and with the specific challenges documented throughout this review. This report is a summary of panel deliberations, with 26 recommendations for future work.

EXECUTIVE SUMMARY

Burch, P.[1]; Dorn, M.W.[2]; Martell, S.J.D.[3] (2026). 2026 Independent International Review of Orange Roughy Stock Assessment Modelling.

New Zealand Fisheries Science Review 2026/02. 62 p.

This review was commissioned by Fisheries New Zealand to provide independent international scrutiny of orange roughy stock assessment modelling in New Zealand. Orange roughy present a suite of biological characteristics that collectively make stock assessment more difficult than for almost any other commercially fished species. New Zealand orange roughy assessments have evolved through three broad phases from 1990, each driven by new data availability and the limitations of previous approaches. A three-person panel (Burch, Dorn, Martell) convened in Wellington in June 2026 following a virtual preparatory meeting. The panel reviewed four high-information stock assessments — East and South Chatham Rise (ESCR), Northwest Chatham Rise (NWCR), Challenger Plateau (ORH 7A), and Mid-East Coast (MEC) — examined data inputs and biological parameters, evaluated approaches for low-information stocks, and assessed reference points and harvest control rules. The panel issued 26 recommendations organised under six thematic headings.

Overall findings. The current assessment approach for New Zealand orange roughy is built on a solid foundation of quantitative acoustic surveys adapted to the unique challenges of orange roughy biology and behaviour, and a relatively simple assessment framework to estimate stock status and trends using those acoustic estimates as a primary source of information. When there were critical uncertainties in the assessment, such as the divergence in biomass between the 38 and 120 kHz echosounders for the ESCR stock, these uncertainties were carried forward into assessment results rather than hidden. New Zealand orange roughy assessments are broadly consistent with orange roughy assessments in other jurisdictions, such as Australia.

All four assessments apply a common Bayesian age-structured framework (CASAL) with natural mortality fixed at $M = 0.045\ \text{yr}^{-1}$ or estimated using an informative prior. An informative acoustic catchability prior centred at $q = 0.8$ (*CV* 19%) is also used. This consistency facilitates cross-stock comparison but also means that common structural assumptions propagate uniformly across all stocks. With around one generation of fisheries data available, estimates of virgin biomass (B_0) and stock productivity are weakly constrained by observations and strongly influenced by prior assumptions. The historical catch, acoustic catchability prior, the fixed *M* (or prior), and the assumed recruitment regime collectively determine most of the variation in estimated stock status.

A common finding across the stocks is that the available data are internally inconsistent in ways that exceed sampling error within any single assessment. Acoustic abundance indices and age composition data conflict across multiple stocks and assessment cycles; the practical response has been to exclude one data source rather than resolve the underlying inconsistency. More fundamentally, observed

---

[1] CSIRO

[2] University of Washington

[3] Sea State Inc

acoustic trajectories cannot be fully reconciled with documented removals under any plausible combination of M, q, and initial biomass, suggesting either that the observation model is mis-specified; due to changes in fish availability, survey coverage, or acoustic detection, or that the population dynamics model is missing important processes such as environmental drivers of recruitment, spatial redistribution, or unaccounted mortality. The panel considers this as a fundamental unresolved problem in New Zealand orange roughy stock assessment, one that cannot be addressed by refinements to the existing model structure alone.

Stock-specific findings

ESCR: the central unresolved issue is the divergence between the 38 kHz and 120 kHz acoustic biomass series since 2017, which drives the difference between stock status estimates of approximately 8% and 18% B_0. Both runs were correctly presented; resolving the divergence is the single most important near-term science priority for this stock. Age composition data were excluded from the base model due to unresolved instability and the extensive otolith collection is substantially under-utilised.

NWCR: the single acoustic biomass series cannot distinguish between competing hypotheses about productivity, recruitment and M. The apparent biomass increase at the Morgue feature most likely reflects progressive concentration of fish on a closed benthic protection area rather than genuine recruitment, and the availability component of the q prior warrants re-examination in light of this spatial shift. The ~35% B_0 estimate of current stock status should be interpreted cautiously given structural uncertainties not captured in the reported confidence intervals (i.e. M, q).

Challenger ORH 7A: The multi-area surveys where no detectable biomass was found in some areas raises questions about spatial availability of the different stock components; the rebuilding signal may partly reflect redistribution of fish rather than genuine recovery.

MEC: With only three acoustic surveys spanning 2013–2021, the empirical basis for tracking stock trajectory remains limited given the stock's slow dynamics. MEC is just above the boundary between high- and low-information stock categories.

Data inputs. The 38 kHz / 120 kHz acoustic divergence since 2017 is the dominant source of uncertainty in the data collection programme. Resolution will likely require concurrent dual-frequency surveys with standardised sphere calibration. The existing otolith collection (>321,000 otoliths, fewer than 4% aged) is a major underutilised scientific asset. An international review of the ageing protocol was undertaken (Horn et al. 2016). Existing ageing data used in assessments should be reviewed to *ensure that they are consistent with the new protocol.*

Biological parameters. Natural mortality is fixed at $M = 0.045\ yr^{-1}$ in some assessments with no uncertainty propagated to derived quantities including stock status and F_{MSY}. An informed prior on M should be developed from estimates of maximum age (Then et al. 2015, Hamel 2015) and a consistent process should be followed to evaluate whether to estimate M using an informative prior, or to fix M at the median of the prior. The simplified CASAL ageing error specification (normal distribution, $CV = 0.1$) should be replaced by a full age-reading error matrix constructed from an inter-reader double-read study (e.g. Punt et al., 2008), consistent with current Australian orange roughy assessment practice.

Low-information stocks and assessment framework. Applying the full CASAL framework to data-poor stocks (ORH 1, East Cape 2A) would produce posteriors dominated by priors, creating a false impression of the precision. Instead, the panel recommends using methods such as Stock Reduction Analysis (SRA) approach (Walters et al. 2006) as a more transparent alternative, potentially combined with hierarchical Bayesian modelling. Targeted acoustic surveys would transform these from catch-only assessments to model-based assessments with an absolute abundance anchor. More broadly, the panel recommends that New Zealand orange roughy assessments move toward a tiered structure: CASAL for data-rich stocks with acoustic and age data; surplus production models and SRA for data-poor stocks and as cross-checks; and Management Procedure evaluation (MP) as the overarching framework for translating stock status uncertainty into management advice. The Neubauer et al. (2026) ESCR management procedure evaluation should be completed and adopted, then extended to NWCR and ORH 7A.

Reference points. Reference points should be reformulated in terms of B_{MSY} and F_{MSY}, which are biologically meaningful and directly estimable under a re-parameterised model in which *MSY* and F_{MSY} are the leading parameters with B_0 and steepness derived. B_0 should be retained as a communication tool Assessments should explore the sensitivity of stock status conclusions to alternative hypotheses about the post-1980 recruitment, given evidence that the post-1980 reduction in recruitment success may reflect a persistent oceanographic regime shift, or a systematic change in q associated with the change in acoustic survey gear. A life-history informed prior on F_{MSY} from per-recruit analysis (approximately 0.02–0.03 yr^{-1}) should be incorporated independently of the stock-recruit relationship.

Recommendations. The panel recommends that New Zealand orange roughy assessments move toward a tiered structure in which four assessment approaches are employed as appropriate to each stock's data situation: CASAL for data-rich stocks with acoustic and age data; surplus production and stock reduction analysis for data-poor stocks and as cross-checks for data-rich stocks; and management procedure evaluation as the overarching framework for translating stock status uncertainty into management advice. This is consistent with international best practice for data-limited and parameter-uncertain fisheries and with the specific challenges documented throughout this review.

The panel makes 26 specific recommendations, emphasising that the most important issues (specifically, the unresolved acoustic frequency divergence, the fixed natural mortality, the narrow q prior) are structurally linked.

Recommendations 1–6 (acoustic survey methods) address the 38/120 kHz divergence, absorption coefficient standardisation, expansion of the q prior, availability and detectability framework, improved backscatter filtering, and exploring new aggregations. Recommendations 7–10 (biological parameters and ageing) address an informed M prior, age-reading error matrix, review of existing age data, maturity-at-age estimation, and sex-specific M. Recommendations 11–14 (age data collection and sampling design) address stratified sampling design, commercial catch ageing, separate selectivity curves for surveys and separate fisheries components, and fitting to conditional age-at-length data. Recommendations 15–20 (model structure and parameterisation) address *MSY* and F_{MSY} re-parameterisation, life-history and catch-history priors on F_{MSY} and *MSY*, hierarchical models for data-poor stocks, spatial feature distribution modelling and ESCR acoustic series partitioning. Recommendation 21 (assessment quality control) suggests a stock assessment checklist, and retrospective analysis with MCMC diagnostics. Recommendations 22–26 (reference points, and management procedures) address B_{MSY} and F_{MSY} reference points, completion of the Neubauer et al. (2026) MP evaluation, use of SRA as a precautionary lower bound, surplus production models for cross-checks and backup and spatiotemporal CPUE standardisation.

1. Introduction

1.1 Background and Remit

Orange roughy (*Hoplostethus atlanticus*) present some of the most difficult stock assessment challenges in commercial fisheries science. The species is characterised by extreme longevity (120–150+ years), very low natural mortality ($M \approx 0.045$ yr^{-1}), late maturation (age 25–35 years), and a generation time of approximately 41 years. The entire history of commercial fishing in New Zealand (which began in 1979) spans around one generation of the stock. These biological characteristics, combined with the species' tendency to form discrete spawning aggregations in deep water (750–1200 m), make it difficult to apply conventional trawl surveys and assessment approaches.

New Zealand manages orange roughy under the Quota Management System (QMS). Four principal stocks are subject to data-rich stock assessment: the East and South Chatham Rise (ESCR), the Northwest Chatham Rise (NWCR), the Challenger Plateau (ORH 7A), and the Mid-East Coast (MEC). The 2025 ESCR assessment indicated stock status in the range of 8–18% of unfished biomass (B_0), triggering statutory rebuilding requirements under the Fisheries Act 1996 and initiating a period of considerable scientific and industry scrutiny of the assessment methods and data underpinning that conclusion.

In this context, Fisheries New Zealand commissioned an independent international review of orange roughy stock assessment modelling. The review was convened in Wellington, New Zealand, from 8–12 June 2026, preceded by a virtual half-day preparatory meeting. The present report constitutes the formal output of that review, prepared by the panel in accordance with the Terms of Reference (ToR) dated 8 May 2026.

1.2 Terms of Reference

The primary objectives of the review, as specified in the ToR, were to:

- Develop recommendations that can inform a science work plan for data collection and stock assessments of orange roughy fisheries in New Zealand;
- Review and comment on recent high-information stock assessments (ESCR, NWCR, Challenger ORH 7A, and MEC), each of which employed somewhat different assessment approaches;
- Compare and contrast alternative approaches, including previous methodologies and surplus production models for assessments or management procedures;
- Review current and past data inputs (catch history, acoustics, ageing, biological parameters (especially natural mortality *M*), trawl surveys, CPUE, length data), including data that was available but not used; and
- To the extent possible, provide recommendations on the science work plan, best approaches for high- and low-information stocks, alternative assessment approaches, and data collection programmes.

Secondary objectives included examining current estimates of the B_{MSY} reference level and providing practical guidance on assessment approaches for low-information stocks (ORH 1, East Coast area 2A North).

The review was explicitly constrained in scope. The following were out of scope: the efficacy of past or present Fisheries New Zealand management measures or the QMS in general; recommended future management actions; relitigating past assessments except for illustrative purposes; and scaling of the catch history with respect to historical overruns.

1.3 Review Panel and Authorship

This report was authored jointly by the three members of the independent international review panel. All three panel members participated in the workshop sessions, the sequestered deliberation period, and the development of the conclusions and recommendations presented here.

Additional participants in the open sessions included Fisheries New Zealand scientists Matt Dunn, Sophie Mormede, Gretchen Skea (chair, Deepwater Working Group), Ian Tuck (chair, open sessions), and Pamela Mace. Matt Dunn presented the most recent orange roughy assessments, Phil Neubauer presented management procedure evaluation work, and Richard O'Driscoll and Tim Ryan presented acoustic survey methods and results. During the panel's sequestered writing period (Wednesday–Thursday), panel members could initiate contact with outside participants but outside participants could not initiate contact with the panel.

1.4 Declaration of No Conflict of Interest

All three panel members declare no conflict of interest with respect to this review. Each member has no financial interest in New Zealand orange roughy fisheries, no contractual relationship with any quota owner, industry body, or fishing company with interests in these fisheries, and no prior involvement in the New Zealand stock assessments or management decisions that are the subject of this review. Panel members' engagements are solely through contracts with Fisheries New Zealand to serve as independent international reviewers. The conclusions and recommendations in this report reflect the panel's collective independent scientific judgment.

1.5 Documents Reviewed

The panel reviewed the following primary documents in advance of and during the workshop.

- Cordue, P.L. (2014a). A management strategy evaluation for orange roughy. ISL Report.
- Cordue, P.L. (2014b). The 2014 orange roughy stock assessments. *New Zealand Fisheries Assessment Report 2014/50*.
- Deepwater Group Ltd. (2024). 2020 Stock Assessment Update of ORH 3B East and South Chatham Rise. *New Zealand Fisheries Assessment Report 2024/11*.
- Dunn, M.R.; A'mar, T.; Doonan, I, (2022). Assessment of the Mid-East Coast orange roughy stock for 2022. *New Zealand Fisheries Assessment Report 2022/38*. 77 p. https://fs.fish.govt.nz/Doc/25285/FAR-2022-38-Mid-East-Coast-Orange-Roughy-Stock-Assessment-2022-4308.pdf.ashx
- Dunn, M.R. (2024). Assessment of the Challenger Plateau orange roughy stock for 2024. *New Zealand Fisheries Assessment Report 2024/58.*
- Dunn, M.R.; Datta, S.; Doonan, I.J. (2025a). Stock assessment of Chatham Rise orange roughy in 2025. *New Zealand Fisheries Assessment Report 2025/46*.
- Dunn, M.R.; Doonan, I.J.; Anderson, O.F. (2025b). Additional analyses for orange roughy (ORH 3B). *New Zealand Fisheries Assessment Report 2025/18.*
- Fisheries New Zealand (2025a). Fisheries Assessment Plenary pages 907–919, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69783-Fisheries-Assessment-Plenary-May-2025-Volume-2-INTRODUCTION-ORANGE-ROUGHY-ORH

- Fisheries New Zealand (2025b). Fisheries Assessment Plenary pages 950–991, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/70159-Fisheries-Assessment-Plenary-May-2025-Volume-2-Orange-roughy-ORH-3B
- Fisheries New Zealand (2025c). Fisheries Assessment Plenary pages 920–926, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69831-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-NORTHERN-NORTH-ISLAND-ORH-1
- Fisheries New Zealand (2025d). Fisheries Assessment Plenary pages 992–1006, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69828-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-CHALLENGER-PLATEAU-ORH-7A
- Fisheries New Zealand (2025e). Fisheries Assessment Plenary pages 927–949, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69840-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-CAPE-RUNAWAY-TO-BANKS-PENINSULA-ORH-2A-2B-3A
- Ryan, T.; Tilney, R.L. (2025). Acoustic biomass surveys of orange roughy spawning aggregations in ORH 3B NWCR and ORH 3B ESCR Chatham Rise, June/July 2022. *New Zealand Fisheries Assessment Report 2025/23.* 71 p.
- Ryan, T.; Tilney, R.; Kunnath, H. (2025). Acoustic biomass surveys of orange roughy spawning aggregations in ORH 3B ESCR Chatham Rise, June–July 2024. *New Zealand Fisheries Assessment Report 2025/23*. 77 p.
- Terms of Reference for a Fully Independent International Review of Orange Roughy Stock Assessment Modelling. (ToR; 2026). Fisheries New Zealand, 8 May 2026.

In addition, copies of the following presentations were provided to the panel to assist with the review.

- Dunn, M.R., (2026). Stock assessment of orange roughy where CPUE is the only biomass index. Presentation for the orange roughy review, June 2026.
- Dunn, M.R., et al. (2026a). Stock assessment of orange roughy in ORH3B: East & South Chatham Rise, including a short history (why we ended up where we did). Presentation for the orange roughy review, May 2026.
- Dunn, M.R., et al. (2026b). Stock assessment of orange roughy in ORH3B. The 2025 assessment of Northwest Chatham Rise. Presentation for the orange roughy review, June 2026.
- Dunn, M.R., et al. (2026c). Stock assessment of orange roughy in ORH7A. The 2024 assessment. Presentation for the orange roughy review, June 2026.
- Dunn, M.R., et al. (2026d). Stock assessment of orange roughy for the Mid-East Coast. The 2022 assessment. Presentation for the orange roughy review, June 2026.
- Mace, P. (2026). A brief history of orange roughy (*Hoplostethus atlanticus*) fisheries in New Zealand. Presentation for the orange roughy review, May 2026.
- O'Driscoll, R. (2026). Acoustic surveys for orange roughy. Presentation for the orange roughy review, June 2026.
- Neubauer, P.; Kim, K.; Large, K.; Langley, A.; Tingley, G. (2026). Evaluating management procedures for ORH 3B ESCR, presentation to the Independent International Review of Orange Roughy Stock Assessment Modelling.

1.6 Structure of this Report

The report follows the structure of the Terms of Reference. Section 2 provides an overview of the New Zealand orange roughy assessment framework and its historical development. Sections 3 through 5 address the high-information stock assessments, data inputs, and biological parameters respectively. Section 6 addresses low-information stocks. Section 7 evaluates alternative assessment approaches including surplus production models and management procedure evaluations. Section 8 addresses reference points. Section 9 presents the panel's recommendations, organised by theme, in response to the Terms of Reference.

2. Overview of the New Zealand Orange Roughy Assessment Framework

2.1 Species Biology and Assessment Challenges

Orange roughy present a suite of biological characteristics that collectively make stock assessment more difficult than for almost any other commercially fished species. They are among the longest-lived vertebrates in marine systems, with lifespans exceeding 150 years and a generation time of approximately 41 years. Natural mortality is very low ($M \approx 0.045$ yr^{-1}), maturation is late (ages 25–35), and spawning may not occur annually. Fecundity is low relative to body size (approximately 40 000–60 000 large, energy-rich eggs). The consequence of this life history is that the population is rapidly depleted under exploitation, but recovers very slowly, and the entire commercial fisheries history in New Zealand spans only about one generation.

The species aggregates in dense spawning concentrations on seamounts and ridge features at depths of 750–1200 m during June–July. These aggregations are highly localised and predictable, making the fishery efficient but also highly vulnerable to local depletion. The aggregating behaviour also means that conventional catch-per-unit-effort (CPUE) indices are unreliable as abundance indicators: as fish density within an aggregation declines, catchability tends to increase (hyperstability), while at the scale of the broader stock, depletion of individual features can give the appearance of stability in aggregated CPUE even as the stock declines. For these reasons, CPUE has been rejected as an abundance index in all recent New Zealand orange roughy assessments.

Acoustic surveys have become the primary source of abundance information. However, orange roughy present specific acoustic challenges. Their wax-ester filled swim bladder produces a relatively low target strength, with values differing between frequencies (−52.0 dB at 38 kHz vs. −48.7 dB at 120 kHz). Fish aggregations extend very close to the seabed, within the acoustic dead zone of hull-mounted transducers, and a single large non-target species having an air-filled swim bladder (e.g., Johnson's cod) within the aggregation can materially inflate biomass estimates. Survey coverage is inherently partial; a prior on acoustic catchability (q) is therefore required, and the construction and parameterisation of that prior is an influential modelling choice in the assessment.

Age determination has also proven problematic. Orange roughy are aged from otolith growth zones, but individuals tend to segregate by sex, and potentially age and maturity within spawning aggregations, creating sampling bias issues. Ageing protocols have been revised over time (Horn et al. 2016), and the reliability of historical age-frequency data collected prior to protocol standardisation is uncertain. Despite a large existing otolith collection (over 321 000 otoliths catalogued), only a small fraction has been aged, and the utility of age data has been inconsistent across assessments.

2.2 Evolution of the Assessment Approach

New Zealand orange roughy assessments have evolved through three broad phases, each driven by new data availability and the limitations of previous approaches.

In the 1990s and early 2000s, age-structured models were developed that incorporated the available catch history, research trawl survey biomass indices, acoustic survey biomass indices, and age-frequency data. These models projected strong stock rebuilding following catch reductions. However, the expected recovery was not observed in subsequent acoustic surveys, leading to a loss of confidence in model-based assessments. The divergence between model predictions and observations was attributed to a combination of poorly estimated productivity, unreliable age data, and the inherent difficulty of estimating unfished biomass (B_0) from less than one generation of fisheries data and only declining trends in abundance.

In response, Fisheries New Zealand adopted a model-free ‘stock evaluation’ approach over the period of approximately 2007–2013, based principally on trends in acoustic biomass indices rather than formal model-based stock status estimation. While this approach reduced the risk of overoptimistic and poorly calibrated stock status estimates, it also precluded formal quantification of stock status relative to reference points and limited the ability to set catch levels.

The current assessment framework was established by Cordue (2014b), following extensive revision of the input data for East and South Chatham Rise (ESCR), in which a more restrictive data quality threshold reduced the quantity of data used. The Cordue (2014b) approach uses a Bayesian age-structured model implemented in CASAL, with the following key features: (i) CPUE is rejected as an abundance index; (ii) acoustic surveys provide the primary biomass signal, scaled via an informative prior on catchability ($q \approx 0.8$, representing approximately 80% of surveyed biomass being available to the survey); (iii) natural mortality is fixed at $M = 0.045$ yr^{-1}, consistent with longevity-based estimates; and (iv) recruitment (year-class strength, YCS) is estimated with log-normal, near-uniform priors. This framework was applied across multiple stocks through approximately 2020, and the resulting stock status estimates informed the harvest control rules established by the Cordue (2014a) management strategy evaluation. This framework was subsequently modified by Matt Dunn, motivated by concerns about overparameterisation in other orange roughy assessments when year class strengths were estimated (Stevenson et al. 2022), and was used in assessments from about 2021 to the present.

The 2023 ESCR assessment encountered serious difficulties: age-frequency data conflicted with the acoustic biomass indices and could not be readily fitted and the steep decline in acoustic indices could not be reconciled with a plausible population trajectory under the Cordue (2014b) framework. That assessment was ultimately not accepted for management advice. The 2025 ESCR assessment (Dunn et al. 2025a) responded by simplifying the model to an acoustics-only approach (with some age-data included in sensitivity runs), and presenting parallel model runs based on the two available acoustic frequency series (38 kHz and 120 kHz). The divergence between these acoustic biomass series in recent years was not resolved and represents the central unresolved technical issue in the current ESCR assessment.

3. Review of High-Information Stock Assessments

This section reviews the four stock assessments presented to the panel: the East and South Chatham Rise (ESCR), the Northwest Chatham Rise (NWCR), the Challenger Plateau (ORH 7A), and the Mid-East Coast (MEC). Each assessment shares the broad Cordue (2014b) framework of Bayesian age-structured modelling in CASAL with acoustic surveys as the primary abundance index, but differs in data availability, model configuration, and the specific challenges encountered. For each stock, the panel summarises the data used, the key modelling choices, the resulting stock status estimate, and the principal concerns arising from its review.

3.1 East and South Chatham Rise (ESCR)

3.1.1 Background and Data

The ESCR stock is the largest and most data-rich orange roughy stock in New Zealand, and the assessment reviewed here (Dunn et al. 2025a) is the most consequential. The stock is distributed across a series of discrete spawning features on the Chatham Rise at depths of approximately 800–1000 m. Commercial fishing began in 1979, and the catch history extends to 2024–25.

The primary abundance data are acoustic biomass estimates from two survey platforms and frequency channels: a 38 kHz hull-mounted system operated from fishing vessels (available from 2002), and a dual-frequency (38 and 120 kHz) system producing independent 38 and 120 kHz estimates from 2011 onward. Otolith-based age-frequency data have been collected since the 1980s, with a revised ageing protocol adopted after Horn et al. (2016). The 2025 assessment also had access to additional spawning-season otolith collections, though the vast majority of the over 321 000 catalogued otoliths remain unaged.

3.1.2 Model Structure and Key Assumptions

The 2025 ESCR assessment simplified the Cordue (2014b) framework considerably relative to previous applications. The primary model configurations were acoustics-only; age-frequency data were excluded from the base model runs, though sensitivity analyses incorporating age data were presented. The simplification was necessary because age-frequency data conflicted with the acoustic biomass indices in the 2023 assessment cycle and could not be reconciled with the acoustic time series under a plausible population trajectory.

Key structural assumptions in the 2025 assessment include: natural mortality fixed at $M = 0.045\ \text{yr}^{-1}$; a two-phase recruitment structure with a higher pre-1980 recruitment level (R_0) and a lower post-1980 level (R_1), with no stock-recruit relationship (steepness = 1 for within-phase recruitment); and an informative prior on acoustic catchability (q) centred at 0.8. Virgin biomass (B_0) is not directly estimated but is derived from R_0, M, and the assumed life history schedule.

Two parallel model runs were presented, differing only in which acoustic series was treated as the primary index: one using the 38 kHz series and one using the 120 kHz series. These two series have diverged materially since approximately 2017, with the 38 kHz series indicating higher recent biomass than the 120 kHz series. The cause of this divergence has not been resolved.

3.1.3 Stock Status

The 2025 assessment estimated current spawning stock biomass (*SSB*) at approximately 8–18% B_0, depending on which acoustic series was used: the 120 kHz model yielded the lower estimate (~8% B_0) and the 38 kHz model the higher (~18% B_0). Both estimates fall below the 20% B_0 soft limit, triggering rebuilding requirements under the Harvest Strategy Standard 2008. The assessment explicitly noted that projections are 'especially uncertain,' reflecting the unresolved acoustic divergence and the fundamental productivity uncertainty inherent in less than one generation of data.

The post-1980 recruitment shift hypothesis is estimated to represent a substantial reduction in productivity relative to the pre-fishing period. The magnitude of this shift strongly influences both the estimated B_0 and the projected rebuilding trajectory, and is itself poorly constrained by the available data.

3.1.4 Panel Findings

The central unresolved issue in the ESCR assessment is the divergence between the 38 kHz and 120 kHz acoustic series. This is not merely a technical issue: it determines whether the stock is estimated at approximately 8% or 18% B_0, a difference that affects the stringency of rebuilding requirements and the sustainable catch level. The panel finds that the 2025 assessment correctly identified this divergence as the dominant source of uncertainty and appropriately presented both model runs rather than forcing a single estimate.

However, the panel notes that presenting two parallel model runs without a basis for weighting or selecting between them leaves managers without clear guidance. Resolving the acoustic divergence is an important near-term science priority for this stock. This is discussed further in Section 4.

The decision to exclude age-frequency data from the base model is understandable given the instability observed in 2023, but represents a significant loss of biological information. The panel is concerned that the current otolith collection, one of the most extensive for any deepwater species, is being substantially under-utilised. A systematic investigation of the sources of age-data instability (sampling selectivity, spatial stratification, protocol changes) is warranted before age data are permanently set aside.

The use of process error terms in the acoustic likelihood for the 'Old Spawning Plume' and 'Combined' AOS series is a pragmatic response to unexplained variability, but risks absorbing a real biological signal as statistical noise. The panel recommends that the basis for process error parameterisation be more explicitly justified in future assessments.

Finally, the panel notes that the confounding of B_0 and post-1980 recruitment level is a fundamental structural problem for this stock, not an artefact of any particular modelling choice. With only about one generation of data, the model cannot simultaneously estimate the level of pre-fishing biomass and the magnitude of the subsequent productivity shift. This issue is discussed more broadly in the context of model parameterisation in Section 9.

3.2 Northwest Chatham Rise (NWCR)

3.2.1 Background and Data

The Northwest Chatham Rise stock occupies seamount and ridge features on the western and northern portions of the Chatham Rise, broadly adjacent to but spatially distinct from the ESCR stock. Commercial fishing began in the early 1980s. The stock is smaller than ESCR, with a virgin biomass estimated at approximately 59 000 t.

The available data are more limited than for ESCR. Acoustic surveys provide the primary abundance signal, but no usable age-frequency data were incorporated in the base model. An attempt was made to include age-frequency data from a 2022 'Morgue' sampling programme, but these data were inconsistent with the acoustic trend and implying implausible year-class patterns, and were excluded. The cause of the anomaly (sampling selectivity change, spatial stratification, or ageing error) was not resolved.

3.2.2 Model Structure and Key Assumptions

The NWCR assessment uses the Cordue (2014b) CASAL framework in an acoustics-only configuration, consistent with the simplified ESCR approach. Natural mortality is estimated with an informative prior (μ = 0.045, CV = 0.33) and the same informative prior on acoustic catchability ($q \approx$ 0.8, CV = 0.19) is also applied. The same two-phase recruitment structure assumed for ESCR (pre- and post-1980 levels) was explored, with no stock-recruit relationship within phases.

An important sensitivity analysis presented for NWCR compared two alternative model configurations that produced similarly good fits to the acoustic data but with very different biological interpretations: (i) a model estimating natural mortality (M free) and steepness h=0.75, and (ii) a model with time-varying recruitment (TV recruitment). Both achieved comparable likelihoods, illustrating the fundamental non-identifiability problem when fitting a single acoustic index with limited contrast. The time-varying recruitment model was ultimately rejected as biologically implausible as it implied an approximately four-fold increase in recruitment over the assessment period, but this rejection was based on biological judgment rather than statistical model selection. The q prior is doing substantial work in constraining the fitting of the acoustic time series.

3.2.3 Stock Status

The NWCR assessment estimated current spawning stock biomass at approximately 35% B_0, within the 30–50% B_0 target range. The two-phase recruitment model for NWCR estimated a ~4-fold post-1980 increase that was considered implausible and was not carried forward. Productivity in the final models with estimated M is near expected levels. The NWCR catch utilisation has been low relative to the catch limit: approximately 18% of the allowable catch was taken in 2023–24. This low utilisation does not primarily reflect reduced commercial interest in the stock; rather, it reflects a spatial shift in the distribution of the spawning aggregation. Morgue hill, the principal NWCR spawning feature, was closed to bottom trawling as a Benthic Protection Area in 2001. The closure prohibits trawl gear from contacting the seabed within the area. Assessment documentation indicates that the main spawning aggregation has shifted from the Graveyard hill, which remains open to fishing, to Morgue, which is acoustically observable but closed to directed fishing. Historically 60–85% of annual NWCR catch was taken during the spawning season; that fraction has fallen to approximately 20% in recent years, consistent with spawning biomass concentrating on a protected feature unavailable to the fishery.

3.2.4 Panel Findings

The NWCR assessment illustrates a general problem that applies across all orange roughy stocks: when the only abundance information is a single acoustic index time series, the model has very limited ability to distinguish between competing hypotheses about productivity, recruitment, and natural mortality. The similar fits obtained under the M-free and TV-recruitment configurations are a concrete demonstration of this non-identifiability. The panel acknowledges that rejecting the TV-recruitment model on plausibility grounds is reasonable, but notes that such judgments introduce subjectivity into the assessment process and are difficult to defend transparently. For example, a TV-recruitment model for the ESCR showing a recruitment decrease of similar magnitude was considered acceptable in the final model configuration.

The Morgue Benthic Protection Area provides additional ecological context for rejecting the TV recruitment model. The four-fold increase in recruitment implied by that model is not plausible as a genuine demographic event. A more parsimonious explanation for the increasing acoustic biomass signal at Morgue is that spawning fish are progressively concentrating on a protected feature where bottom trawl gear is prohibited from contacting the seabed. Morgue is fully observable acoustically but closed to directed fishing. If the Graveyard-to-Morgue shift has been occurring progressively since the 2001 closure, the acoustic trend at Morgue would show rising biomass without any genuine recruitment event. This spatial availability effect should be modelled explicitly (e.g., logit proportions used in Neubauer et al. 2026), and the q prior for NWCR warrants re-examination in light of the growing proportion of spawning biomass concentrating on a feature that is surveyed but not harvestable.

The exclusion of age data in the assessment, despite a sampling programme specifically designed to provide them, is a concern. Understanding why the 2022 Morgue age data produced anomalous results is important not just for NWCR but for interpreting age data reliability across all New Zealand orange roughy stocks. The panel recommends a dedicated investigation of this anomaly before the next assessment cycle.

The stock status estimate for NWCR (~35% B_0) should be interpreted with caution given the limited data constraining it. The estimate is largely determined by the acoustic trend and the q prior; the model has little independent information on B_0. If true catchability differs materially from the prior, status could be substantially different. This structural uncertainty is not fully captured in the reported confidence intervals.

3.3 Challenger Plateau (ORH 7A)

3.3.1 Background and Data

The Challenger Plateau stock (ORH 7A) is distributed across a series of seamounts and ridge features on the Challenger Plateau west of the North Island, at depths of approximately 800–1100 m. Commercial fishing began in the early 1980s. The fishery has changed spatially over time, extending east and south but contracting from the north, complicating the interpretation of abundance trends from any single survey area.

The primary abundance data are acoustic biomass estimates from three spatially distinct survey areas, each surveyed with varying frequency over the assessment period. The most recent assessment (Dunn 2024) incorporated acoustic time series from all three areas. No usable age-frequency data were included in the base model, consistent with the acoustics-only approach applied to ESCR and NWCR.

3.3.2 Model Structure and Key Assumptions

The Challenger ORH 7A assessment uses the same CASAL Bayesian age-structured framework as the other stocks, with M estimated using an informative prior, and an informative q prior. The key methodological question specific to this stock is how to combine acoustic estimates from three spatially distinct survey areas into a coherent population-level abundance index.

Two approaches were compared. The first ('combined series') sums the acoustic biomass estimates across all three areas and fits the aggregate as a single index with a shared $q = 0.8$. This approach assumes the surveys collectively observe approximately 80% of spawning biomass and that spatial coverage is consistent over time. The second approach ('3Series') treats the three areas as separate abundance indices; each linked to the same underlying population dynamics but with independently estimated catchabilities. This approach assumes that the proportion of stock available to each survey area does not change over time.

Both configurations were presented in the assessment, with the combined series approach used as the base model. The 3Series approach produced broadly similar stock status estimates but greater uncertainty, as the additional flexibility in catchability parameters allowed more variation in how the acoustic data are interpreted.

3.3.3 Stock Status

The Challenger ORH 7A stock is estimated to be rebuilding, with current spawning stock biomass in the vicinity of the lower end of the 30–50% B_0 target range. The Cordue (2014b) assessment estimated status at approximately 42% B_0; the 2024 assessment indicates the assessment estimates rebuilding from approximately 2000 to 2015, followed by moderate decline associated with increased catches in recent years, though with considerable uncertainty around the trajectory.

3.3.4 Panel Findings

The Challenger ORH 7A assessment presents a methodologically interesting case because the multi-area acoustic survey design explicitly raises the question of spatial availability: how much of the stock is observable by the survey, and does this change over time as the fishery and the population distribution evolve? The comparison of the combined-series and 3Series approaches is a useful sensitivity analysis, but the panel notes that neither approach resolves the underlying uncertainty about spatial availability.

The combined series assumption (q = 0.8, constant across areas and time) is particularly strong given the documented spatial expansion of the fishery. If fish have been progressively depleted from previously fished features and survey coverage has not tracked that expansion proportionally, the combined series could overstate recent abundance. Conversely, if fish are re-aggregating on previously fished features as the stock rebuilds, coverage could improve over time. The panel recommends that future assessments for this stock include a more explicit treatment of spatial availability, consistent with approach developed by Neubauer et al. (2026) for ESCR.

Relative to ESCR and NWCR, the Challenger ORH 7A assessment is less compromised in terms of internal data conflicts. The acoustic data are more consistent across areas than for ESCR, showing a rise to approximately 2015 followed by decline, and no major unresolved acoustic issue analogous to the 38/120 kHz divergence was identified. Acceptable fits to the age composition data were obtained. However, the time-varying availability issues described above means that the apparent coherence of the acoustic signal may partly reflect fish redistribution. The principal uncertainties (M, B_0, q, and spatial availability) are shared with all other orange roughy stocks and are discussed in Section 5.

3.4 Mid-East Coast (MEC)

3.4.1 Background and Data

The Mid-East Coast stock (ORH MEC) is distributed across a series of discrete features along the continental slope off the east coast of the North Island at depths of approximately 800–1000 m. The fishery is strictly feature-based: the surrounding seafloor is mostly unsuitable for bottom trawling, so all commercial catch is taken from a limited number of trawlable features. There is no background fishery between features and no option for effort to redistribute to open ground. Commercial fishing began in the mid-1980s.

A notable characteristic of the broader MEC area is the presence of substantial numbers of small orange roughy distributed outside the main spawning features on non-trawlable ground. Whether these small fish are uniquely a component of the MEC stock is unknown. The harvestable component is restricted to fish on trawlable features, but the full stock includes the broader distribution.

The MEC stock occupies an intermediate position in the data richness spectrum: richer than the low-information stocks (ORH 1, East Cape 2A) but considerably poorer than ESCR. Available data include the catch history, acoustic surveys conducted in 2013, 2017, and 2021, trawl surveys from 1992–2010, age-frequency samples collected in 1989–1991 and 2010, and length-frequency data. The most recent assessment is from 2022 (Dunn et al. 2022).

3.4.2 Model Structure and Key Assumptions

The MEC assessment model is older in its architecture than the other three assessments and was originally designed to integrate a richer suite of data types: catch history, trawl surveys, age-frequency compositions, acoustic surveys (2013, 2017, 2021), and length-frequency data. This more integrated approach reflected the data available from the early and mid-fishery period and the absence of a long acoustic time series to serve as a single dominant index.

Three acoustic surveys have been conducted (2013, 2017, 2021), providing a short time series of absolute abundance estimates (conditional on the prior for q). These surveys provide information on biomass scale and, in principle, stock trajectory. Trawl surveys and age-frequency data from the 1989–2010 period contribute information on the early depletion trajectory. The complex gear history, reflecting changes in vessel type, gear configuration, and the sequential exploitation of individual features on otherwise untrawlable ground, requires careful reconstruction of effective fishing mortality over time.

Natural mortality is fixed at $M = 0.045\ yr^{-1}$, and the model estimates year class strengths. Given the short acoustic time series (three surveys spanning 2013–2021), the q prior continues to play a dominant role, and the limited number of survey points constrains the ability of the model to track stock trajectory. The translation to spawning stock biomass depends largely on the assumed catchability. The interpretation of q for MEC is further complicated by the stock structure: if a substantial component of the stock is distributed on ground where availability to acoustic surveys is variable because of relatively large acoustic dead zones, the relationship between the acoustic estimate and total stock biomass is unclear.

3.4.3 Stock Status

In the most recent assessment, the MEC stock is estimated to be at 22.4% B_0, slightly above the soft limit. The stock is estimated to have been slowly rebuilding since 1994–95, though given the low productivity of the species and the limited number of survey observations, this result should be interpreted cautiously.

3.4.4 Panel Findings

The monitoring path forward for MEC is constrained by the untrawlable nature of most of the stock's range. Trawl surveys, which contributed to the historical model, cannot be meaningfully conducted across the broader MEC area. Acoustic surveys targeted at aggregations are the only practical method for tracking the spawning component of the stock. A broader-area acoustic transect programme could in principle characterise the distribution of small fish on non-trawlable ground, providing information on the non-harvestable stock component that the current model does not address; however, if smaller orange roughy are mixed in with other species this may not be possible. Biological sampling is similarly constrained: age and length data can only be collected from the commercial fishery operating on trawlable features. This concentration of all available monitoring on a small and spatially fixed set of features means that MEC monitoring is structurally more like an index site monitoring programme than a stock-wide survey, and the assessment model needs to reflect that limitation explicitly.

Given its data limitations and the structural complexity introduced by the feature-only fishery and the non-trawlable small-fish component, MEC is just above the boundary between the high- and low-information stock categories defined in the Terms of Reference.

3.5 Cross-Stock Comparison and Synthesis

Taken together, the four assessments reviewed reveal a consistent set of structural features, shared limitations, and stock-specific variations that have important implications for the future direction of the science programme.

All four assessments share the same fundamental framework: a Bayesian age-structured model in CASAL with natural mortality either fixed at $M = 0.045\ yr^{-1}$ or estimated using an informative prior, and an informative prior on acoustic catchability centred at $q = 0.8$. This consistency is a strength in that it enables direct comparison of stock status estimates across stocks on a common scale, but it also means that common structural assumptions, and their associated uncertainties, propagate uniformly across all assessments. If M is systematically biased, or if the $q = 0.8$ prior is poorly calibrated for a particular stock, all status estimates will be affected in the same direction.

Age-frequency data played a meaningful role in only one assessment (MEC), and even there the data are from the 1989–2010 period and contribute primarily to the early depletion trajectory rather than current stock status. The exclusion or anomalous behaviour of age data in ESCR and NWCR represents a substantial loss of biological information from a stock assessment programme that has invested heavily in otolith collection. Resolving the sources of age-data instability is a priority science need identified consistently across all four assessments. In addition, the treatment of the age data in the assessment does not adequately reflect the level of uncertainty (Pennington & Volstad 1994) leading to potential over-influence of this data on the outcome.

Finally, the cross-stock comparison reinforces the importance of a point that applies to all four assessments equally: with only about one generation of fisheries data, estimates of virgin biomass (B_0) and stock productivity are weakly constrained by observations and strongly influenced by structural assumptions. The q prior, the value of M, the cumulative fishery removals, and the assumed recruitment regime collectively determine most of the variation in estimated stock status. Data collection and modelling choices that reduce uncertainty in these parameters, rather than adding further complexity to already data-limited models, represent the highest-value investments for the future science programme.

A cross-cutting finding that the panel considers important to state explicitly is that the available data are internally inconsistent in ways that go beyond sampling error in any single assessment. Two manifestations of this inconsistency stand out. First, the acoustic abundance indices and the age composition data are in conflict across multiple stocks and assessment cycles: attempts to fit both simultaneously produce either anomalous model behaviour (ESCR 2023, NWCR 2022) or implausible biological trajectories, and the practical response has been to drop one data source rather than to reconcile the conflict. This is a symptom of a deeper problem; the two data streams are providing incompatible signals about the state of the stock, and excluding one does not resolve the underlying inconsistency, it merely defers it.

Second, a major unresolved problem is that the acoustic abundance indices for some stocks show trajectories that cannot be explained by the documented removals alone. Under any plausible combination of M, q, and initial biomass, the observed pattern of catches should produce a predictable depletion and partial recovery trajectory. The fact that observed acoustic indices deviate from this expected trajectory in ways that are not attributable to sampling error suggests that either the observation model is wrong; the indices are not measuring what the model assumes they measure, due to changes in availability, survey coverage, or acoustic detection, or the population dynamics model is missing important processes such as environmental drivers of recruitment, spatial redistribution of fish between surveyed and unsurveyed areas, or unaccounted sources of mortality. The panel considers this an unresolved problem in New Zealand orange roughy stock assessments, and one that cannot be addressed by refinements to the existing model structure alone. It motivates the investigation of alternative assessment approaches discussed in Section 7.

4. Data Inputs

The quality and interpretation of data inputs are the foundation of any stock assessment, and for orange roughy they present unusual challenges at every stage. This section reviews the principal data streams used in the four assessments: acoustic surveys, otolith ageing, catch history, and ancillary data including CPUE, trawl surveys, and length frequencies. The acoustic survey review draws heavily on O'Driscoll (2026), prepared specifically for this panel.

4.1 Stock Structure Assumptions

The panel did not specifically examine the many studies that inform the current stock structure assumptions for New Zealand orange roughy assessments as this information was not presented during the review and it noted that most of these studies were decades old. While there may be some benefit in the application of more sophisticated modern genetic tools, the panel considered that the current stock structure assumptions for the four high-information assessments (ESCR, NWCR, ORH 7A and MEC) were likely a reasonable compromise between large enough to minimise the amount of movement and small enough to capture the fisheries and stocks dynamics.

4.2 Acoustic Surveys

4.2.1 Survey Programme History and Technology

Acoustic surveys are the primary source of abundance information for New Zealand orange roughy and represent a substantial and sustained investment in fisheries monitoring. The survey programme has evolved through several phases. Single-frequency (38 kHz) towed acoustic systems were first used on research vessels from 1986, with early surveys of the north Chatham Rise spawning aggregations reported by Do & Coombs (1989). Prior to 2002, abundance estimates relied largely on these research vessel surveys and CPUE, neither of which proved sufficiently reliable or cost-effective for ongoing monitoring. From 2002, hull-mounted 38 kHz echo-sounder systems were deployed on commercial fishing vessels during dedicated spawning-season surveys, providing the first consistent acoustic time series for ESCR and other stocks. From 2011, dual-frequency (38 and 120 kHz) acoustic optical systems (AOS) were introduced on commercial fishing vessels; net-attached towed bodies that sample the acoustic field immediately ahead of the net. The AOS approach was developed to improve biomass estimation by sampling closer to the aggregation and reducing dead-zone losses (Ryan & Kloser 2016). Most recently, in 2021, a towed-body dual-frequency system deployed from a research vessel was added to the programme, offering improved coverage and reduced susceptibility to weather-related bubble attenuation.

4.2.2 Technical Challenges

Orange roughy present a series of well-documented acoustic survey challenges that are worth reviewing in detail because they bear directly on the interpretation of the survey time series and the 38/120 kHz divergence.

Target strength and sound absorption. The wax-ester swim bladder of orange roughy produces a relatively low and frequency-dependent target strength. At 38 kHz the target strength is approximately -52.0 dB, compared to approximately -48.7 dB at 120 kHz (using the Coombs & Barr (2007) with the intercept adjusted using Kloser et al. (2013)). This approximately 3 dB difference means that a 38 kHz survey will record lower raw backscatter per fish than a 120 kHz survey, and the conversion from backscatter to fish density is sensitive to the assumed target strength equation and its parameterisation.

A further source of uncertainty is the choice of sound absorption equation. Two equations are in common use; Francois & Garrison (1982) and Doonan et al. (2003), and they can produce vessel-based biomass estimates that differ by up to 30%. Neither equation is universally preferred and the choice is not always documented consistently across the New Zealand survey time series. Uncertainty in target strength and absorption together represent a significant contributor to uncertainty in absolute biomass estimates from both series.

Dead zone. Orange roughy sometimes aggregate in dense concentrations very close to the seabed, often within the bottom acoustic dead zone of hull-mounted transducers. The depth of the dead zone depends on the acoustic pulse length and transducer characteristics; but is typically 0.5–1 m above the seabed; on sloping ground this can exceed several metres. Empirical corrections following Ona & Mitson (1996) are applied, but the correction can exceed 40% on steep topography even with towed systems, representing a substantial and spatially variable source of uncertainty. Because fish may concentrate in exactly this region during spawning, a systematic fraction of the aggregation may be missed by hull-mounted surveys, especially when aggregations form over features such as hills. The AOS towed in mid water or the 2021 towed-body system reduce but do not eliminate this limitation.

Bubble attenuation and vessel motion. Hull-mounted transducers are susceptible to acoustic signal attenuation caused by air bubbles entrained beneath the hull in rough sea conditions. Orange roughy spawn in winter and surveys are conducted in conditions that regularly exceed the practical limits of hull-mounted systems (approximately 25 knots wind, 2 m swell). A hull correction factor of approximately 1.33 has been applied to the New Zealand orange roughy acoustic data to account for this effect, but the correction factor was derived from limited empirical data and its uncertainty is not fully propagated into biomass estimates. Vessel motion artefacts in hull-mounted systems can be corrected using the approach of Dunford (2005). If sea state conditions have changed systematically over the survey period, the hull correction factor may introduce a time-varying bias into the hull-mounted time series. Towed systems substantially reduce bubble attenuation by operating below the bubble layer, thus improving data quality.

Species identification and composition. While orange roughy are the dominant species in spawning aggregations, other species are present. A single large Johnson's cod (*Pseudophycis bachus*) can produce backscatter comparable to a substantial number of orange roughy, and misidentification or failure to exclude non-target species can inflate biomass estimates. Multi-frequency comparison, trawl verification, and underwater camera systems are used to improve species identification, but uncertainty remains, particularly at the edges of aggregations.

Survey availability and spatial coverage. Acoustic surveys observe fish that are present and available to the survey at the time and location of sampling. Orange roughy migrate to spawning features during a limited season (approximately June–July) and availability to the acoustic survey is very low outside the spawning window, because fish are dispersed rather than aggregated. Within the spawning season, survey coverage of the various spawning features is incomplete and has varied over the time series as the survey design has evolved. The q prior ($q = 0.8$) is intended to account for the fraction of spawning biomass not observed, but this prior is based on sparse historical data and expert judgment rather than empirical estimation, and its uncertainty is not fully represented in model outputs.

4.2.3 The 38 kHz / 120 kHz Divergence

The divergence between the 38 kHz and 120 kHz acoustic time series for ESCR since approximately 2017 is the most consequential unresolved technical issue in the current assessment programme. The two series agreed reasonably well in earlier years, lending confidence that both were measuring the same underlying phenomenon. Since 2017, the 38 kHz series has indicated higher and relatively stable biomass, while the 120 kHz series has declined steeply.

Several hypotheses have been advanced to explain the divergence, none of which has been conclusively demonstrated or ruled out. These include: changes in the target strength relationship at one or both frequencies over time (e.g., due to changes in fish size or condition); changes in the calibration of the 120 kHz system or the application of calibration corrections; changes in species composition within the survey area affecting the backscatter at one frequency more than the other; changes in the depth distribution or vertical positioning of fish relative to the seabed (affecting dead-zone losses differently at each frequency); and genuine biological signal reflecting changes in fish behaviour or vertical distribution that produce differential backscatter responses at the two frequencies within the same integrated sampling volume.

The panel notes that the 2025 assessment of ESCR and NWCR (Dunn et al. 2025a) correctly identified the divergence as unresolved and presented results under both series rather than selecting one. However, the absence of a framework for evaluating the relative reliability of the two series leaves assessment users without guidance on which model run should inform management decisions. This is not a problem that will resolve itself with the passage of time or the accumulation of additional survey years under the current design. Dedicated research effort is required.

The central diagnostic question is whether the divergence reflects a gear problem or a change in fish behaviour, and these two classes of explanation have fundamentally different implications for how the series should be treated. A gear problem, such as a systematic change in the calibration of the 120 kHz system, or a change in deployment geometry, would mean that one series is producing biased estimates of true backscatter, and the appropriate response is to identify and correct the bias, or to eliminate the affected series. A change in fish behaviour; such as a change in the vertical distribution of fish relative to the seabed, a change in swim bladder state or fish orientation affecting target strength differently at the two frequencies, or a change in species composition within survey strata; would mean that both systems are accurately measuring backscatter but the relationship between backscatter and fish density has changed. In that case, both series may be internally consistent but neither can be interpreted without updating the target strength or availability assumptions. The two scenarios require different scientific responses and different management implications, yet the current evidence base does not allow the panel to distinguish between them.

Resolving this divergence between the 38 and 120 kHz surveys should be the highest-priority near-term science investment for the ESCR stock. Until this research is completed and its conclusions accepted, the panel considers the ESCR stock status estimate to be subject to a fundamental acoustic uncertainty that cannot be resolved within the current modelling framework.

4.2.4 Panel Advice on the Path Forward

Given the central role of acoustic surveys in all New Zealand orange roughy assessments and the unresolved 38/120 kHz divergence, the panel offers the following advice to guide the near-term acoustic research. The recommendations are sequenced by urgency and do not require waiting for later steps before beginning earlier ones.

An in-situ comparison of the current dual-frequency acoustic survey array with another comparable unit would ideally determine whether the issues are equipment related or biological. Pending resolution of the divergence, the panel recommends that assessments continue to present both model runs in full rather than selecting one series or averaging across them without a basis for weights. The range between the two runs should be communicated explicitly as unresolved acoustic uncertainty; not a probability distribution on stock status, and management advice should be framed accordingly. Any interim weighting scheme applied to the two series should be based on documented technical evidence.

4.2.5 Development of pixel-by-pixel dB-differencing filter

Given the low target strength of orange roughy, development of tools to more reliably exclude fish with swim-bladders for echo-integration are needed. Developing a pixel-by-pixel dB-differencing filter to exclude backscatter not consistent with orange roughy is recommended, based on per-cell comparisons between the 38 kHz and 120 kHz echosounders. This approach would provide a more granular tool for screening for contamination by fish with swim-bladders than current methods.

4.2.6 Survey Design: Availability and Detectability

A recurring source of confusion in the interpretation of New Zealand orange roughy acoustic surveys is the conflation of two conceptually distinct components of acoustic catchability: availability and detectability. Separating these components is essential for understanding what the acoustic index actually measures, how it should be incorporated into a stock assessment model, and how its uncertainty should be represented.

Availability is the proportion of the stock that is spatially and temporally present at the surveyed features during the survey window. It is driven by behaviour; what fraction of the population aggregates at known spawning features versus remaining dispersed across the broader range, by the completeness of feature coverage (surveyed versus unsurveyed), and by the timing of the survey relative to peak aggregation. Availability is inherently a population-level quantity: it describes where the fish are, not whether the acoustic system can see them. It is also likely to be the largest and most variable component of q. Inter-annual variation in the fraction of the stock that aggregates at the survey site, variation in survey timing relative to the spawning peak, and slow changes in spatial distribution driven by long-term depletion or environmental change can all alter availability without any change in true abundance or any change in the performance of the acoustic system.

Detectability is the proportion of fish that are present at the surveyed location and time that are actually detected by the acoustic system. It is a function of the survey instrument and its deployment (e.g., hull mounted versus towed transducer): dead-zone losses for fish very close to the seabed, target strength uncertainty and its dependence on fish size, sound absorption, bubble attenuation in rough weather, species composition of backscatter, and the spatial completeness of transect coverage within the aggregation footprint.

The 19% *CV* on the New Zealand q prior captures uncertainty in target strength; one component of detectability, and nothing else. It does not capture dead-zone variability, transect coverage uncertainty, vessel mounted versus towed transducer, or any component of availability. The prior is therefore too narrow relative to the true uncertainty in the acoustic-to-biomass conversion.

The requirement that fish must form a detectable acoustic plume further blurs the boundary between availability and detectability. If orange roughy are present at a feature but aggregating at a density below the threshold required to produce a coherent acoustic return, they are formally available; they are at the right place at the right time, but functionally undetectable. The survey returns near-zero not because the fish are absent but because the aggregation density is insufficient. This density-dependent detectability creates a non-linear relationship between true abundance and the acoustic index: at high abundance, fish aggregate densely and are reliably detected; at low abundance, fish may be present but dispersed, and the survey underestimates them in a way that is not proportional to their true number. This non-linearity is consequential precisely when accurate measurement matters most; at low stock sizes approaching limit reference points. A stock that has declined to the point where spawning aggregations no longer reliably plume will appear to have declined further than it has, potentially triggering more severe management responses or, conversely, appearing to recover rapidly as abundance crosses the threshold for a plume to form. Neither the current q prior nor the CASAL model structure accounts for density-dependent detectability and the current practice of omitting survey estimates where no fish were found is appropriate.

The panel recommends the variance on the prior on q be increased to account for the variability in availability and detectability, not just target strength. This could be done using a similar approach as the Australian eastern zone assessment that is described below.

4.2.7 Development of an acoustic prior

Australian and New Zealand approaches to acoustic survey design and uncertainty quantification differ in an important respect. In the Australian eastern zone orange roughy assessment, acoustic catchability (q) is constructed as an explicit three-component product: uncertainty in target strength, uncertainty in the proportion of the stock visible to the survey (availability), and a random survey error term (Burch & Sutton 2026). Each component is estimated or bounded from empirical data or literature, and the resulting uncertainty in q is propagated directly into the stock assessment. In contrast, the New Zealand approach applies a prior on the aggregate q (approximately 0.8) without decomposing its components. This makes it difficult to evaluate whether the prior is consistent with the available data or whether its uncertainty is being adequately represented.

4.3 Otolith Ageing

4.3.1 Collection and Utilisation

New Zealand has the most extensive orange roughy otolith collections in the world, with over 321 000 otoliths catalogued from spawning-season and fisheries sampling programmes. This represents a substantial and sustained investment in biological data collection that has the potential to provide strong constraints on model estimates of year-class strength, selectivity, and productivity. In practice, only 10 277 otoliths have been aged, representing less than 4% of the collection. The gap between collection and utilisation reflects both the labour intensity of otolith ageing, and reductions in the use of aged otoliths because of a change in the ageing protocol (Horn et al. 2016).

The cost of ageing orange roughy otoliths is substantial, due to the labour required in fixing and sectioning and then estimating the number of annuli. This expense is why the gap between the size of the otolith archive and the number of aged specimens is so large, and it has direct implications for how any future ageing programme is designed. Ageing every collected otolith is not feasible; the objective must be to age a strategically selected subsample that provides the maximum assessment value per dollar spent. This means the design of the ageing subsample; which otoliths to prioritise, from which features, seasons, years, separately for survey and fisheries components with different selectivity.

Incorporating age data into assessment models raises a modelling question that the current framework does not address: the selectivity relationship between the acoustic survey and the commercial fishery is likely to be different, yet assessments have not routinely estimated separate selectivity curves for each data source. The acoustic survey during spawning observes fish that are aggregated and acoustically detectable, with a selectivity determined by fish behaviour and aggregation geometry. The commercial trawl catch selects fish based on location and season, gear configuration, depth, and the dive response of fish to approaching gear. Observer samples may differ further depending on when and where observers sample within the fishing operation. Fitting a single selectivity function to age compositions from sources with genuinely different selectivity processes introduces internal contradictions that the model resolves by distorting other parameter estimates. Estimating separate survey and fishery selectivity curves, as is standard practice in multi-fleet age-structured assessments, should be explored as part of the next generation of New Zealand orange roughy models.

4.3.2 Sources of Instability

The failure of age-frequency data to produce stable model behaviour in the 2023 ESCR assessment, and the anomalous results from 2022 NWCR Morgue ageing, point to one or more of the following underlying problems. First, sampling selectivity: spawning aggregations are known to segregate by sex, size, and possibly age, meaning that samples collected from commercial catches or targeted biological sampling may not represent the age structure of the stock as a whole. If selectivity changes over time as the fishery targets different features or depths, apparent changes in the age distribution may reflect sampling artefacts rather than genuine population dynamics.

There is direct evidence of spatial and temporal stratification of spawning orange roughy, with the assessment recording that different spawning features support different age structures, and that these differ across years. Dunn et al. (2024) noted the apparent loss of fish aged 30–40 years from the ESCR spawning plume samples and a corresponding gain of those ages at the Morgue feature, and concluded this may reflect fish from different age classes spawning on different features in different years. The 2025 assessment (Dunn et al. 2025a) further noted that a substantial proportion of old fish present at the Old Spawning Plume in 1984 were absent in 2003 but returned in 2012. Dunn et al. (2025a) treated age frequencies separately by feature (Old Spawning Plume, Rekohu, the Crack) precisely because pooling across features would obscure genuine compositional differences. If sampling effort is not consistently stratified across features from year to year, the resulting age-frequency data will reflect changes in where sampling occurred as much as changes in the population, producing model instability that is difficult to diagnose or correct.

The statistical distribution assumed for the age composition likelihood may not be appropriate. All current New Zealand orange roughy models use the multinomial distribution for age composition data. The multinomial treats each aged fish as an independent observation, implying that the full nominal sample size is informative. In practice, fish sampled from a single tow or within a single spawning aggregation are highly correlated in age: a set of tows on one feature on one day is effectively a single cluster of observations, not a sample of independently drawn individuals. When the effective sample size is much smaller than the nominal sample size input into the assessment, the multinomial likelihood overweighs the age composition data relative to other data streams, causing the model to fit age composition residuals at the expense of other data inputs. While the age data were downweighted within the assessment, it is unclear whether this ad hoc correction is sufficient. This is a plausible contributor to the instability observed in the 2023 ESCR assessment and the 2022 NWCR Morgue data. Alternative approaches are available and in routine use in North Pacific groundfish assessments: the Dirichlet-multinomial distribution, which estimates an overdispersion parameter that scales effective sample size to be consistent with observed residual variance; and iterative reweighting methods such as Francis (2011), which adjust composition weights empirically so that the assumed and realised variability are self-consistent. Either approach could be implemented within the CASAL framework and should be evaluated before age composition data are re-introduced into the base models.

A related and practical approach is to bin age composition data into broader age groups before fitting. With a species that may live beyond 120 years and typical sample sizes of a few hundred aged fish, the expected count per individual age class across the full age range is much less than one for most age classes. The resulting composition vector may include many zeros and ones, which are the least informative possible observations and introduce a systematic bias: individual year-class counts dominated by sampling noise are treated by the multinomial likelihood as carrying precise information about year-class strength. Aggregating into broader age bins; for example, 10 or 15-year groups, chosen to align with biologically meaningful stages such as pre-recruit, young adult, mature adult, and old adult, concentrates the available observations into a manageable number of cells each with adequate expected counts, removes the false precision of year-class-level fitting, and makes the multinomial approximation substantially more defensible. Binning and alternative composition likelihoods (Dirichlet-multinomial, Francis reweighting) are complementary approaches to the same underlying problem of sparse composition data and should be evaluated in combination. Annual recruitment deviates should not be estimated where the composition data are binned, because the data cannot support cohort-level inference. The panel suggests estimating recruitment in blocks aligned to the age bins (e.g., decadal deviates) with a penalty or hierarchical prior on the block deviates.

As noted above, the panel finds that the existing otolith collection is a substantial underutilised asset. Systematic ageing of a well-designed stratified subsample, using the current protocol and with rigorous quality control, could substantially improve the information content of future assessments. The ageing programme should consider which otoliths to prioritise, from which features, seasons, years, separately for survey and fisheries components with different selectivity. Age frequencies by each season, year, and survey /fisheries component should be large enough to capture the age structure of the sampled population. Sampling should be distributed across as many hauls as possible to address strong intra-trawl correlations (Pennington & Volstad 1994). Australia targets the collection and ageing of 1000 otoliths from each acoustic optical survey. While all New Zealand orange roughy assessments would benefit from additional age frequency data, the panel recommends prioritising the ESCR, as this is the largest and most important New Zealand orange roughy stock.

Conventional methods of otolith ageing (i.e., counting annuli in sectioned otoliths) are labour intensive and hence expensive. Preliminary work undertaken in Australia using computer vision models to estimate fish age from sectioned otoliths was not successful for fish age > 40 years. An alternative approach that was also explored by Australia was to estimate fish age from otolith weight and other covariates including sex, length and year of sampling using a generalised linear model (GLM). This produced ageing error estimates around twice the magnitude of annuli counting. While this approach was considered promising by Australia's Southern and Eastern Scalefish and Shark Resource Advisory Group (SESSFRAG) in 2025, SESSFRAG felt the level of uncertainty in age estimates was too high and that a project should be developed to improve precision by incorporating additional morphometric information such as otolith width and otolith weight. If this approach can obtain sufficiently accurate and precise estimates of age it may provide substantial cost savings.

4.4 Catch History

The catch history for New Zealand orange roughy stocks is generally well-documented, supported by mandatory catch reporting under the QMS. Annual commercial landings are available from the onset of each fishery and area-specific breakdowns can be reconstructed from logbook data. Some uncertainty remains in the early years of each fishery (late 1970s to mid-1980s), when reporting requirements were less stringent and some catch was misreported by area. Historical catch overruns are out of scope for this review per the Terms of Reference, but the catch history is the primary driver of estimated historical depletion and uncertainty in early catches propagates into B_0 and current stock status estimates. Future assessments should include sensitivity analyses to alternative early catch history assumptions.

The complex gear history, reflecting changes in vessel type, gear configuration, and the sequential exploitation of individual features on otherwise untrawlable ground, requires careful reconstruction of effective fishing mortality over time. This heterogeneity is likely not to be adequately captured by a single selectivity curve. When data from sources with genuinely different selectivities are fitted with a common curve, the model cannot distinguish selectivity misspecification from genuine population change, and the resulting parameter estimates absorb the mismatch as spurious recruitment or mortality signals. This is a plausible contributor to the difficulty of integrating multiple data types in orange roughy assessments, and may help explain why recent assessments have retreated from the full integrated data complement. Resolving the selectivity problem; either by estimating gear-specific selectivity curves or by restricting integration to data sources with demonstrably comparable selectivity, is a prerequisite for a reliable integrated assessment. Misspecification of the multinomial sample size for the age composition likelihoods could also be a source of tension as the time series has grown.

4.5 Ancillary Data

4.5.1 Catch Per Unit Effort

CPUE from commercial fishing operations has been excluded from all recent New Zealand orange roughy assessments on the grounds that hyperdepletion and hyperstability render it uninformative or misleading as a stock-level abundance index. Both effects are well documented for this species and the panel concurs with the exclusion. While there is a large volume of logbook and biological data for the New Zealand orange roughy fisheries, the CPUE record lacks the temporal contrast needed for reliable standardisation: directed effort has varied greatly across years and features due to quota changes, closures, and economic conditions, and many feature-year combinations are poorly represented or absent from the time series. The catch distribution can also be extremely skewed, with 10–20% of annual catch potentially coming from a single tow on a dense undisturbed aggregation, which means standard delta-GLM or log-linear standardisation methods cannot reliably extract a stock signal even when applied to otherwise well-specified models.

Two further compounding factors are the sequential disturbance effect and skipper targeting skill. Analysis of New Zealand orange roughy fishing records has demonstrated a quantifiable depression of catch rates after previous towing activity on the same feature; an effect that was only detectable when actual net position (corrected for wire angle and depth) was used as the spatial reference rather than vessel GPS position. Net position offset is therefore a necessary covariate in any spatially explicit standardisation. Vessel identity should be treated as a random effect. Since logbook records identify vessel but not individual skipper, vessel is the only available proxy for targeting skill. Catching orange roughy from aggregations forming over complex underwater features requires considerable expertise, and targeting precision rather than fish density is likely a primary driver of the extreme catch rate variability observed in this fishery.

The panel does not recommend reintroducing CPUE as a direct abundance index for the four data rich stocks, but considers two limited uses worth pursuing. First, when the directed fishery ceases or falls to near-zero effort despite an open quota, this economic shut-down reveals that fish density has fallen below the profitable fishing threshold. Years of negligible directed effort provide floor constraints on stock density independent of the hyperdepletion problem, and the panel notes that this approach can be explored for stocks where effort is near zero for extended periods. Second, spatially explicit estimators, such as the Walters (2003) spatial CPUE methodology and spatiotemporal models (e.g., sdmTMB, VAST) should be considered for data-poor New Zealand orange roughy stocks, with skipper random effects treated as a model component, to determine whether a standardised index can be constructed.

4.5.2 Trawl Surveys

Trawl surveys were conducted on the Chatham Rise and other areas in the 1990s and 2000s and contributed to some earlier assessments (and still do for ORH 7A and MEC). Their utility is limited by the same aggregating behaviour that confounds CPUE: trawl catchability is highly variable and poorly understood for a species that forms discrete dense aggregations, and trawl surveys have a small spatial footprint relative to the distribution of the stock. The panel concurs that acoustic surveys are the preferred abundance index for this species and that trawl survey data are most useful as a source of biological samples (age, length, weight, maturity) or a potential recruitment index, rather than as a direct abundance index of adult orange roughy for model fitting.

4.5.3 Length Frequencies

Length-frequency data have been collected from commercial catches and research surveys over much of the fisheries history. Their utility for stock assessment is constrained by the slow growth of orange roughy: the species does not exhibit the well-defined length modes that make length-frequency analysis informative for faster-growing species. Nonetheless, length data can provide useful information on the presence or absence of recruits and on changes in mean size over time that may reflect shifts in exploitation pattern or year-class success. Length data may also assist in the estimation of selectivity ogives for fisheries components. These data are currently underutilised in the assessment models and could merit further attention.

One opportunity that the panel considers worth investigating is the potential to derive a recruitment index from the length-frequency data collected during trawl surveys. Trawl surveys have sampled orange roughy across a broader size range than the spawning-aggregation acoustic surveys, and the length-frequency distributions from these surveys may contain a detectable signal from small, recently recruited fish that are not yet part of the spawning aggregation and therefore not captured by the acoustic surveys. If small fish can be reliably identified in the trawl length-frequency data, the time series of their relative abundance could serve as a leading indicator of future recruitment to the spawning stock, providing a data stream that is currently absent from all New Zealand orange roughy assessment models. The panel recommends that the available trawl survey length-frequency time series be systematically reviewed with this question in mind, and that the feasibility of constructing a recruitment index be evaluated.

4.5.4 Observer and Biological Sampling Data

Fisheries observers have been deployed on orange roughy vessels and provide records of catch composition, discards, and biological samples. Observer coverage rates and sampling protocols have varied over the history of the programme. The panel notes that observer data have been under-utilised in recent assessments and that a review of observer records may yield additional information that could inform both the acoustic surveys and the biological parameter estimates used in the models.

5. Biological Parameters

The biological parameters used in orange roughy stock assessment models provide some of the primary determinants of stock status, sustainable yield, and rebuilding trajectory. For a species with the life history of orange roughy, where only about one generation of data is available, parameter assumptions may carry more weight than observed data in driving model outputs. This section reviews the four most influential biological parameters: natural mortality, maturity, recruitment regime and the stock-recruit relationship, and acoustic catchability.

5.1 Natural Mortality

5.1.1 Current Practice

Natural mortality is fixed at $M = 0.045\ yr^{-1}$ in several base models. This value was established by Doonan (1994) based on longevity-based estimators and has been used in New Zealand orange roughy assessments for over two decades. The value is broadly consistent with estimates from other orange roughy populations internationally.

Fixing *M* removes one degree of freedom from the estimation problem but has consequence: all estimates of B_0, sustainable yield, and stock status are conditional on assumptions about *M* being correct. *M* is strongly confounded with R_0; a higher *M* produces lower equilibrium biomass per recruit and therefore a higher implied R_0 to match observed biomass, while a lower *M* does the reverse. Because the data span only about one generation, the model cannot resolve this confounding from observations alone. If *M* is fixed, the uncertainty in *M* is not propagated into reported stock status confidence intervals, which understates true uncertainty in management advice. Nevertheless, it must be acknowledged that it is not always possible to estimate *M* reliably in stock assessment models, even when an informative prior is used.

5.1.2 Attempts to Estimate *M*

Cordue (2014b) explored estimating *M* within the MCMC sensitivity analyses, using a strongly informative prior (normal distribution, mean 0.045, *CV* 15%). The sensitivity run produced posterior estimates centred near the prior mean, indicating that the data provide little additional information about *M* beyond the prior. This is not surprising given the confounding with R_0 described above: in a one-way-trip data environment, the model cannot distinguish between a stock with low *M* and high virgin biomass and one with higher *M* and lower virgin biomass, because both could produce the observed depletion trajectory.

In the NWCR sensitivity analyses, estimating *M* or allowing time-varying recruitment produced similar fits to the acoustic data but very different biological interpretations. The workshop document noted that ‘estimating any more than two or three parameters produces essentially the same fit,’ illustrating the fundamental non-identifiability of the model under the available data. When the likelihood surface is flat with respect to *M*, the posterior is dominated by the prior mean and its variance.

Close-kin mark-recapture (CKMR) was presented to the workshop by panel member Paul Burch as a potential method for estimating population size and survival rates; and therefore *M* because catches are known, independently of acoustic surveys and model assumptions. CKMR uses genetic relationships between sampled individuals: parent-offspring pairs (POPs) or half-sibling pairs (HSPs) identified in a tissue sample to provide an absolute abundance estimator, because the probability of observing a kin-pair is related to population size and adult survival (Bravington et al. 2016a). The method has been successfully applied to provide formal TAC advice for southern bluefin tuna (Bravington et al. 2016b) and school shark (Thomson et al. 2020, 2025). A recent CKMR design study for key Southern and Eastern Scalefish and Shark Fishery (SESSF) stocks (Thomson et al. 2026) found that the Australian eastern zone orange roughy stock would be a poor CKMR candidate for two specific reasons: first, the coincidence of fishing selectivity and maturity means that sampled adults are drawn from the same narrow age range as potential parents, making POPs virtually undetectable in practice; and second, the 16–17 year lag in the spawning biomass year with highest statistical precision means a CKMR study initiated now would not deliver precise abundance estimates for at least a decade. These constraints apply equally to New Zealand orange roughy given the same life-history and fisheries characteristics. The panel accepts that CKMR does not currently offer a practical path to independent *M* estimation or absolute biomass for New Zealand orange roughy, though the method should be revisited if genotyping costs fall substantially or sampling designs that circumvent the selectivity-maturity overlap can be developed.

A more tractable path to an informed *M* prior is through meta-analytic life-history estimators. Hamel (2015) developed a method for combining multiple such estimators; relating *M* to longevity, growth rate, body size, and reproductive effort, into a coherent Bayesian prior. Then et al. (2015) provided an updated cross-species dataset of *M* and life-history parameters spanning over 200 fish species, and evaluated the predictive performance of a range of *M* estimators. They concluded that *M* estimates based on maximum age (*Amax*) alone performed as well as or better than multi-parameter estimators, and recommended an updated Hoenig non-linear least squares estimator: $M = 4.899 \times Amax^{-0w916}$. However, Then et al. (2015) did not apply a consistent transformation across the alternative model forms they compared. In the presence of the substantial heteroscedasticity expected in the observed *M–Amax* relationship; both in observation error and process error, all models should be fitted under a log transformation. Re-evaluating the Then et al. (2015) data by fitting the one-parameter *Amax* model under a log-log transformation with slope constrained to −1 (as in Hamel 2015), the point estimate becomes $M = 5.4/Amax$. For orange roughy with a maximum observed age of approximately 150 years, this yields $M \approx 0.036$ yr^{-1}; somewhat below the fixed value of 0.045 yr^{-1} but within the plausible range. Hamel & Cope (2022) further refined *M* estimation by appropriately accounting for sources of error in both *Amax* and *M*, and recommend a prior defined as a lognormal distribution with median 5.4/*Amax* and log-scale standard deviation of 0.31. This prior is directly applicable to orange roughy and is consistent with the approach already widely adopted in US West Coast rockfish assessments.

5.1.3 Comparison with Australian Practice

The panel notes that the Australian eastern zone orange roughy assessment estimates *M* within the model using a prior derived from the posterior of New Zealand stock assessments (Cordue 2014b), with a prior mean of ~0.037 yr^{-1}. The Australian approach propagates uncertainty in *M* through the full MCMC posterior, resulting in wider but more realistic confidence intervals on stock status. The panel notes that estimating *M* in the Australian model increased overall uncertainty, which is precisely the point: the New Zealand practice of fixing *M* suppresses a major source of uncertainty rather than acknowledging it. The Australian experience with *M* treatment has now been systematically evaluated across a broader range of stocks. Punt et al. (2026) reviewed how *M* specifications have evolved across stocks in the SESSF and re-analysed ten stocks under three frameworks: (a) pre-specification based on indirect longevity methods; (b) estimation without a prior; and (c) estimation with a prior informed by longevity data. Assessment performance was evaluated using likelihood profiles, retrospective analysis (Mohn's ρ), and hindcast predictive skill. The study concluded that no single approach dominated across all stocks and criteria, but the weight of evidence across diagnostics supported estimation with a prior as the preferred default. Pre-specification was found to carry appreciable risk of bias when the population dynamics or observation model is substantially mis-specified, because there is no mechanism within the model to accommodate discrepancy between the assumed M and the information content of the data. The recommended best practice for data-rich and data-moderate SESSF assessments is to estimate M using a prior based on longevity information (e.g., the Hamel–Cope method or equivalent), with pre-specification reserved for situations where diagnostics indicate model misspecification or the estimated M diverges markedly from the prior. Where pre-specification is adopted in the presence of apparent misspecification, sensitivity runs and decision tables should be used to characterise the consequences of alternative M assumptions, or better yet conditional on the assumed prior variance for M.

5.2 Maturity

A number of approaches have been used to model maturity in orange roughy assessments. A common practice from Cordue (2014b) is to set maturity equal to selectivity and estimate the curve, but this confounds two process that should be kept separate. The opaque-to-translucent transition zone in orange roughy otoliths is associated with first spawning and provides a direct biological marker of maturity that is independent of the fish's reproductive condition at the time of capture. The large existing otolith collection, the majority of which was sourced from spawning aggregations where the transition zone was recorded, provides a substantial and largely untapped dataset for fitting a maturity ogive from transition-zone observations. The panel recommends that a carefully selected subsample of the existing collection be used to estimate a logistic maturity curve for each stock as a function of estimated age. The resulting empirical maturity ogive should be fixed in CASAL. Rather than coupling maturity and selectivity, the panel recommends estimating selectivity as a penalised offset from the fixed maturity curve to link two curves without compromising the empirical maturity estimates. Alternatively, selectivity could be estimated independently if there is sufficient age composition data and acceptable results can be obtained, i.e., low cryptic biomass.

5.3 Recruitment Regime and the Stock-Recruit Relationship

5.3.1 The Two-Phase Recruitment Structure

Two assessments (ESCR and NWCR) explored the use of a two-phase recruitment structure. The final ESCR assessment model estimated a pre-1980 recruitment level (R_0, corresponding to the unfished state) and a reduced post-1980 level (R_1), with the change point coinciding with the onset of commercial fishing. The NWCR assessment estimated a substantial increase in recruitment post-1980 that was deemed implausible and not carried forward. No stock-recruit relationship is assumed within either phase (steepness = 1 within phase), meaning recruitment is treated as a constant in each period rather than a function of spawning stock biomass. This structure was introduced because the acoustic time series showed a sustained decline that could not be reconciled with expected stock rebuilding under the earlier, simpler models.

The magnitude of the post-1980 recruitment reduction is a critical and a poorly constrained parameter. The ratio R_1/R_0 determines how fast the stock can rebuild under reduced fishing mortality. If R_1 is close to R_0, the stock has good recovery potential; if R_1 is substantially lower, rebuilding will be very slow regardless of catch reductions. The 2025 ESCR assessment estimated the post-1980 recruitment shift as substantial, but this estimate is weakly constrained by observations and sensitive to the assumed value of M and the q prior. The three parameters (M, q, and R_1/R_0) are jointly non-identifiable from the available acoustic time series alone, suggesting reconsideration of post-1980 recruitment shift approach in future ESCR assessments, and whether other approaches, such as estimating decadal recruitment blocks would be more robust.

5.3.2 Steepness and the Beverton-Holt Relationship

Steepness (h) describes the resilience of recruitment to reductions in spawning stock biomass: $h = 1$ implies no relationship (recruitment is independent of SSB greater than 20% SSB_0), while lower values imply recruitment declines as SSB declines, reducing recovery potential. The Cordue (2014a) MSE estimated steepness at approximately 0.6 across a range of scenarios, but with a very wide credible interval (0.31–0.95), reflecting the near-total lack of contrast in spawning biomass over which a stock-recruit relationship could be estimated. Given the wide uncertainty in h, projections presented with a single steepness assumption substantially understate the uncertainty in rebuilding timelines.

5.3.3 Confounding of B_0, Steepness, and the Post-1980 Recruitment Shift

A fundamental structural problem in ESCR orange roughy assessment is the near-complete confounding of three key parameters: virgin biomass (B_0), steepness (h), and the post-1980 recruitment level (R_1). All three influence the predicted stock trajectory in similar ways over the available data range, and only about one generations' data cannot resolve the confounding. A model with high B_0, high steepness, and moderate R_1 may produce virtually the same fit to the acoustic time series as one with lower B_0, lower steepness, and a different R_1. The result is that reported stock status (%B_0) is highly sensitive to assumptions that are effectively unconstrained by the data.

One approach to reducing this confounding, explored in the context of other data-limited assessments (Martell et al. 2008), is to re-parameterise the model so that maximum sustainable yield (MSY) and the fishing mortality at MSY (F_{MSY}) are the leading estimated parameters, with B_0 and steepness derived from them rather than estimated directly. This re-parameterisation is algebraically equivalent but computationally advantageous when the data are informative about productive capacity (what level of yield is sustainable) but not about the components (B_0 and h) that together determine it. The panel recommends this approach be explored for orange roughy assessments and is discussed further in Section 9.

A further and practically important advantage of the MSY and F_{MSY} parameterisation is that it enables the construction of an empirically grounded prior on R_0 via a prior on MSY. Under the current parameterisation, R_0 is estimated with a relatively vague prior because virgin recruitment is not directly observable and any prior is difficult to justify from data. MSY, by contrast, is informed by the catch history: the observed ranges of historical catch can provide an informative prior on MSY that is grounded in observable quantities. Through the algebraic relationship between MSY, F_{MSY}, M, and the life history schedule, a prior on MSY implies a prior on R_0 without requiring independent information about virgin biomass or steepness. This re-parameterisation provides a method of incorporating catch history information into the prior structure of the model, rather than treating R_0 as a free parameter constrained only by a diffuse prior.

The chain of inference extends one step further to acoustic catchability. If a prior on MSY implies a prior on R_0, and R_0 together with M and the life history schedule determines B_0, then given the observed acoustic index; which estimates $q \times B$, a prior on B_0 implies a corresponding constraint on q. The catch history and the acoustic time series are therefore jointly informative about q under this parameterisation, rather than q being an independent expert-judgment input fixed universally at 0.8. Stocks with different catch histories and different acoustic index levels would yield different implied q posteriors, which is more defensible than applying a single universal prior to all stocks regardless of their individual data.

This integration of catch history, acoustic surveys, and life history through the *MSY and* F_{MSY} re-parameterisation represents a coherent and fully Bayesian approach to the parameter confounding problem that the current framework addresses only partially through independent priors on R_0, *q*, and *M*.

5.4 Acoustic Catchability

5.4.1 The *q* Prior

Acoustic catchability (*q*) translates observed acoustic backscatter into an estimate of spawning stock biomass: *SSB* = acoustic index / *q*. A prior of $q \approx 0.8$ is used in all four New Zealand assessments, representing the expert judgment that acoustic surveys observe approximately 80% of available spawning biomass. This prior was established by Cordue (2014b) for ESCR, NWCR and ORH 7A based on expert assessment of survey coverage, the proportion of the stock expected to be present at surveyed features during the survey window (from historical but sparse data), and the known limitations of hull-mounted acoustic systems (dead zone, bubble attenuation, species identification). For MEC, Cordue (2014b) used the same approach, but set the prior to $q = 0.6$ because only a single aggregation in a small area was surveyed; the most recent assessment used $q = 0.8$ (Dunn et al. 2022).

The *q* prior is the second most influential scalar parameter in the assessment after *M*. Because *SSB* is directly proportional to $1/q$, a 20% change in *q* produces a 20% change in estimated *SSB* and therefore in $\%B_0$. The prior is applied as an informative Bayesian prior in CASAL, but its construction is not decomposed into identifiable components. The 0.8 value has remained unchanged since Cordue (2014b) despite changes in survey methodology (introduction of dual-frequency systems, towed-body surveys) and changes in the spatial distribution of the fisheries and the surveyed features.

5.4.2 Uncertainty in *q* and the Australian Comparison

A critical limitation of the New Zealand *q* prior is that its *CV* of approximately 19% reflects uncertainty in target strength alone. Target strength uncertainty (arising from the Coombs & Barr (2007) and Kloser et al. (2013) relationships and the concurrent length frequency sampling needed to apply them) is one component of *q*, but not the dominant one. Because *q* effectively converts acoustic backscatter into an absolute estimate of spawning stock biomass, its uncertainty should encompass all sources of uncertainty in that conversion: target strength uncertainty, availability uncertainty (what fraction of the spawning stock is present at the surveyed features during the survey window), spatial coverage uncertainty (what fraction of the feature area is traversed), and species composition uncertainty. By omitting these components, the New Zealand *q* prior substantially understates the true uncertainty in absolute biomass and gives the CASAL model false precision in estimating % B_0. A prior that accurately reflected all sources of uncertainty in the acoustic-to-biomass conversion would be considerably wider, with direct consequences for the confidence with which stock status can be determined.

New Zealand and Australian practices differ meaningfully in *q* construction. The Australian eastern zone assessment decomposes *q* into three explicit components: uncertainty in target strength, uncertainty in the proportion of the stock visible to the survey (availability), and a random survey error term (Burch & Sutton 2026). Each component is bounded from empirical data or literature, and the combined uncertainty is propagated through the assessment. This approach is more transparent, more defensible in review, and more likely to capture the true uncertainty in *q* than a single aggregated prior whose *CV* reflects only one of these components.

The panel notes that adopting a decomposed *q* prior in New Zealand assessments would require empirical work to characterise each component. Target strength uncertainty could be bounded from the existing literature and from multi-frequency comparison data. Survey availability is the hardest component to quantify, but the spatial modelling work of Dunn et al. (2025a) and the Neubauer et at. (2026) MP operating models provide a framework for thinking about this explicitly. The panel recommends that the next round of assessments include a formal review and decomposition of the q prior, with particular attention to whether the 0.8 value remains appropriate given changes in survey methodology and the unresolved acoustic frequency divergence.

5.4.3 Interaction between *q*, *M,* and Stock Status

It is important to recognise that *q* and *M* are not independent in their effect on stock status. A lower *q* (the survey misses more of the stock) implies a higher true *SSB* for the same observed acoustic index. A lower *M* implies slower productivity and a higher equilibrium biomass per recruit. Both parameters shift estimated stock status in the same direction, and both are subject to substantial prior-dominated uncertainty. The joint prior on (*M, q*) therefore determines a large fraction of the posterior distribution on % B_0, and the sensitivity of stock status to joint perturbations of these parameters is considerably larger than sensitivity to either parameter alone. Future assessments should present joint sensitivity analyses across plausible ranges of both *M* and *q* rather than treating each parameter independently.

6. Low-Information Stocks

6.1 Overview and Scope

Two of the stocks within the ToR; ORH 1 (East Coast North Island) and the East Cape area 2A, have very limited data relative to the Chatham Rise and Challenger Plateau stocks. No accepted acoustic survey exists for either stock. Age composition data are sparse or absent. For these stocks, the assessment problem is fundamentally different from the data-rich case: the question is not which model structure best fits the available data, but whether the available data are sufficient to support any model-based stock status estimate at all, and if not, what alternative approaches can provide defensible management advice.

6.2 Hierarchical Bayesian Modelling

The panel's primary recommendation for low-information stocks is the adoption of a hierarchical Bayesian model structure that explicitly shares information across stocks. All New Zealand orange roughy stocks belong to the same species and share the same life history. Biological parameters (*M*, growth schedule, maturity ogive, acoustic availability, and the broad range of productive capacity) are not independent quantities estimated separately for each stock. Data-rich stocks (ESCR, NWCR, ORH 7A) contribute strongly to the hyperprior; data-poor stocks borrow strength from it when their own data are insufficient to independently constrain parameters. The degree of shrinkage toward the common distribution is determined by the data; stocks with more information deviate more freely from the hyperprior than stocks with little information.

There would be practical benefits for New Zealand orange roughy assessment with this approach. *M* is currently fixed at 0.045 for several stocks by assumption. A hierarchical model would instead estimate a hyperprior distribution for *M* informed by all stocks simultaneously, with each stock's estimated *M* shrunk toward the common distribution in proportion to data scarcity. For data-rich stocks with age composition data, estimated *M* may deviate from 0.045 if the data support it; for data-poor stocks, *M* would be constrained near the among-stock mean while correctly propagating the uncertainty due to data scarcity. Similarly, the *q* prior (currently fixed at 0.8 for all stocks) could be treated as a stock-specific quantity drawn from a common hyperprior, allowing among-stock variation in acoustic availability to be estimated rather than assumed away. Productivity parameters: virgin recruitment, the stock-recruit relationship, and the hypothesised post-1980 recruitment shift, could share a hyperprior informed by the stocks with the most age and recruitment data. Selectivity-at-age curves and annual recruitment deviations (year-class strengths) are additional candidates for hierarchical treatment: selectivity curves are unlikely to differ substantially across stocks that share the same gear and aggregation behaviour, and pooling year-class strength estimates across stocks would provide a more powerful test of whether episodic recruitment events exist and whether they are a stock-specific or an overall phenomenon.

Hierarchical Bayesian models of this type are routinely used in North Pacific groundfish assessments, particularly for rockfish complexes where multiple stocks share similar life history characteristics but vary in data availability. They are computationally feasible within a Bayesian MCMC framework using Stan or TMB, and could in principle be implemented within an extended CASAL2 model. The initial investment in model development is non-trivial, but the scientific benefits (e.g., uncertainty propagation for data-poor stocks, partial estimation of *M* and *q* rather than fixed assumption, and formal use of the information contained in the data-rich stocks) would outweigh the costs. The panel recommends this as a priority methodological development for the upcoming assessment cycles. For ORH 1 and the East

Cape stocks specifically, the approach of Edwards et al. (2022), simultaneously fitting multiple local stock assessment models with shared biological parameters, offers a computationally accessible entry point to hierarchical modelling that does not require a full Stan or TMB implementation, and could be evaluated as part of the exploratory modelling work.

6.3 Simpler Alternatives: Stock Reduction Analysis

Pending development of a hierarchical model, two simpler approaches offer more transparent alternatives to running a full CASAL model on sparse data. The first is an explicit stock reduction analysis (SRA): given a prior on *MSY* (informed by the catch history lower bound and a life-history productivity upper bound, as recommended in Section 9), sample trajectories of stock biomass consistent with the catch history and the assumed biological parameters, and report the distribution of current depletion directly (Walters et al. 2006).

The second is a minimum population estimate (MPE) (Walters et al. 2006). These methods estimate the minimum biomass consistent with the observed catch history under assumed biology, providing a conservative lower bound on stock size rather than a point estimate of current status. For management purposes, a defensible lower bound is often more useful than an imprecise central estimate: if the lower bound exceeds the limit reference point, management can proceed with reasonable confidence; if it falls below, precautionary reductions are warranted regardless of uncertainty about the central estimate. MPE-style approaches are particularly appropriate for stocks where the primary policy question is whether the stock is above or below a reference point, rather than the precise level of depletion.

6.4 Data Requirements for Improved Assessments

The most direct path to improved stock status information for ORH 1 and East Cape 2A is data collection. Acoustic surveys, even a single well-designed survey, would transform these from catch-only assessments to model-based assessments with an absolute abundance anchor. The panel acknowledges that surveys of these stocks are more logistically challenging than ESCR or NWCR surveys; but the absence of survey data is not a fixed condition. A targeted acoustic reconnaissance survey, even if imprecise, would provide information that is currently entirely absent. The question of whether acoustic surveys are feasible for these stocks (e.g., what spawning features exist, whether fish aggregate to detectable densities, and what vessel time would be required) should be evaluated.

6.5 Metapopulation Approaches and the Closure Question

An alternative modelling framework that has been explored for New Zealand orange roughy is to treat the stock as a metapopulation; a collection of semi-discrete spawning feature subpopulations connected by adult movement. Under this framework, observed biomass declines at individual features that are not fully explained by local catch could be attributed to net emigration to other features rather than to depletion, potentially producing less pessimistic stock status estimates than a single closed-population model. This framing is ecologically consistent with the genetic evidence that New Zealand orange roughy constitute a single panmictic population: spawning features are aggregation sites, not reproductively isolated units, and redistribution among features as local densities change is biologically plausible. A questionable assumption, however, is that it assumes that all sub populations are known and the population is closed.

One potentially productive research application would be to fix *M* at its current assumed value of 0.045 and estimate net emigration to the unobserved portion of the stock area as a free parameter (McAllister & Kirchner 2002). With *M* fixed, any residual decline in observed biomass beyond what catch alone can explain must be attributed either to the fixed *M* term or to unobserved losses, and the latter can be parameterised as a net emigration rate.

A further complication for any spatially explicit model is that slow changes in physical oceanography may drive spatial expansion or contraction of orange roughy aggregations independently of fishing pressure. If the physical environment is shifting on decadal timescales, the carrying capacity and habitat suitability of individual spawning features may be changing, causing fish to redistribute toward more favourable locations or contract away from marginal ones regardless of local depletion state. Under these conditions, neither q nor B_0 is stationary: q changes as the survey samples a shifting spatial distribution, and B_0 ceases to be a fixed historical reference point because the unfished equilibrium distribution may no longer exist in the current environment.

7. Alternative Assessment Approaches

7.1 Context and Motivation

The Terms of Reference explicitly ask the panel to compare and contrast alternative assessment approaches, including previous methodologies and surplus production models for assessments or management procedures. This section addresses that question. The current New Zealand framework applies a full age-structured Bayesian model (CASAL) to each stock individually, producing annual estimates of spawning stock biomass as a percentage of B_0, and comparing those estimates to fixed reference points. This framework explicitly represents the life history of a long-lived species, incorporates age composition information when available, and produces a posterior distribution on stock status that can in principle be compared to management thresholds. However, the preceding sections of this report have documented a series of limitations in how the framework is applied (i.e., parameter confounding, fixed M, a narrow prior on q) that call into question whether the complexity of an age-structured model is warranted. Alternative approaches that are simpler, more transparent, or explicitly designed for decision-making under uncertainty deserve consideration.

7.2 Surplus Production Models

Surplus production models (Schaefer 1954, Fox 1970) describe stock dynamics in terms of net biomass change — the difference between production and harvest — without requiring explicit age structure, growth schedules, or recruitment functions. They require only a catch time series and an abundance index, and can be fitted in a Bayesian state-space framework (e.g., JABBA, CMSY) that propagates process error and observation error through to stock status estimates. For species where age structure data are limited or unreliable, production models can provide a more straightforward uncertainty representation than an age-structured model that is formally complex but empirically under-determined.

For New Zealand orange roughy, the primary limitation of production models is the non-equilibrium dynamics inherent in a one-way-trip fishery with a recruitment regime shift. Standard production models assume that the stock-production relationship is stationary; that the same MSY and B_{MSY} apply throughout the fisheries history. For orange roughy with a hypothesised post-1980 recruitment regime shift, this assumption is questionable: the productive capacity of the stock in the pre-recruitment-shift period may not be the same as in the post-shift period, and a single surplus production curve fitted to the full catch and index history will conflate these two regimes. A time-varying production model, or one that explicitly allows for a step change in carrying capacity, could address this limitation, but at the cost of additional parameters that the data may not support.

Despite this limitation, the panel considers that surplus production models have a useful role in the New Zealand orange roughy assessment programme as a complement to the age-structured framework rather than a replacement for it. Fitting a production model to the same data as the CASAL model provides an independent cross-check on stock status. If the two frameworks produce broadly consistent conclusions, confidence in both is strengthened. If they diverge materially, the source of the divergence is itself informative — it may point to structural assumptions in the age-structured model (such as the recruitment regime parameterisation or the fixed M) that are driving the result rather than the data. Production models are also more directly interpretable for non-specialist audiences and provide a useful communication tool alongside the more complex CASAL outputs.

7.3 The Minimum Population Estimate Approach

The minimum population estimate (MPE) approach using stochastic stock reduction analysis (Walters et al. 2006), takes a different perspective on the assessment problem. Rather than attempting to estimate the full posterior distribution on current biomass, the MPE asks a more conservative question: what is the smallest population that is consistent with the observed catch history, given assumed biology and the constraint that the stock was not driven to extinction? This minimum estimate provides a precautionary lower bound on stock size that is robust to parameter uncertainty — any population larger than the MPE is also consistent with the data, but the MPE itself cannot be ruled out.

The MPE is derived by running an ensemble of stock trajectory simulations across a range of assumed productivity parameters (*MSY*, F_{MSY}, B_0 or equivalent), retaining only those trajectories in which the stock survives (i.e., biomass does not reach zero) and identifying the trajectory with the lowest current biomass among the survivors. This provides a model-ensemble lower bound that does not depend on a single set of parameter assumptions. The approach is particularly suited to stocks with one-way-trip data where the likelihood surface is flat and the posterior is dominated by the prior: by searching over a broad range of parameter space rather than conditioning on a single model, the MPE avoids the false precision of a single posterior mode.

The panel endorses the MPE approach as a useful and scientifically defensible tool for stocks where the primary management question is precautionary: is there any plausible parameter combination under which the stock could be at or below the limit reference point? If the MPE itself is above the limit reference point, the answer is no, and management can proceed with confidence. If the MPE falls below the limit reference point, precautionary action is warranted regardless of where the central CASAL estimate sits. Used alongside the CASAL model, the MPE provides the lower tail of the plausible stock status distribution that the CASAL posterior, conditioned on its priors, may not adequately represent.

7.4 Management Procedures and Management Strategy Evaluation

The most sophisticated alternative framework available for New Zealand orange roughy is formal Management Procedure evaluation (MP) within a management strategy evaluation (MSE) framework. The Cordue (2014a) MSE established the current reference points and a previous harvest control rule for ORH 3B by testing candidate strategies against an ensemble of age-structured operating models representing uncertainty in key parameters. That work demonstrated that target reference points (30–50% B_0) combined with the current HCR provide acceptable performance across a range of plausible stock states, and it remains the scientific foundation for current management.

The Neubauer et al., (2026) management procedure evaluation for ESCR extends this approach by incorporating the unresolved acoustic frequency divergence as a source of operating model uncertainty. By conditioning operating models on both the 38 kHz and 120 kHz acoustic series and evaluating candidate MPs against both sets of conditioning, the evaluation tests whether the current HCR is robust to the acoustic uncertainty that has been the central unresolved issue in recent assessments. The panel regards this as a methodologically important development: rather than attempting to resolve which acoustic series is correct before making a management decision, the MP evaluation asks which management actions perform acceptably across both scenarios. This reframes the management problem in a way that is more compatible with the actual state of knowledge.

The panel endorses the MP evaluation framework. Several features of the orange roughy assessment problem make it well-suited to this approach. The major sources of uncertainty — *M*, *q*, the acoustic series divergence, the recruitment regime — are not resolvable in the near term. An MP evaluation framework accepts these uncertainties explicitly, encodes them in the operating model ensemble, and tests candidate HCRs against the full range. Management advice can then be stated in terms that are robust to the unresolved uncertainties rather than conditional on a single model's parameter estimates.

Transitioning fully to an MP-based management framework requires that the operating model ensemble be agreed in advance by stakeholders and scientists, that the performance metrics (biomass targets, catch stability, probability of limit breach) be specified through a collaborative process, and that the evaluation be updated periodically as new data accumulate. This is a more resource-intensive process than annual CASAL model runs, but it is also more transparent, more robust to model uncertainty, and more likely to produce management advice that is durable across assessment cycles. The panel recommends the Neubauer et al. (2026) ESCR MP evaluation be completed and, subject to satisfactory results, extended to NWCR and ORH 7A.

7.5 Comparison of Approaches

The three frameworks — age-structured CASAL, surplus production, and MP evaluation — are not mutually exclusive and serve different purposes within the assessment programme. The age-structured model provides the most detailed biological representation and is the appropriate tool when age composition and acoustic data are both available and informative. The surplus production model provides a simpler, more transparent cross-check and is useful for communication and for stocks where age data add little information. The SRA provides a precautionary lower bound that is robust to parameter uncertainty and is particularly important for stocks at or near reference points. The MP evaluation framework uses all of the above as operating models and asks which management actions are robust across their disagreements; it is the appropriate framework for translating uncertain stock status estimates into management advice that is consistent with the evidence.

The panel's recommendation is that New Zealand orange roughy assessments move toward a tiered structure in which all four approaches are employed as appropriate to each stock's data situation: CASAL for data-rich stocks with acoustic and age data; surplus production and SRA for data-poor stocks and as cross-checks for data-rich stocks; and MP evaluation as the overarching framework for translating stock status uncertainty into management advice. This is consistent with international best practice for data-limited and parameter-uncertain fisheries and with the specific challenges documented throughout this review.

8. Reference Points and Harvest Control Rules

8.1 Current Reference Point Framework

New Zealand fisheries operational guidelines require that stock assessments identify target and limit reference points and evaluate stock status relative to those points. For orange roughy, the conventional approach expresses current spawning stock biomass as a fraction of estimated unfished biomass (B/B_0) and compares this to reference points derived from simulations conducted by Cordue (2014a) for ORH 3B. The current target reference point is in the range of 30–50% B_0, and the limit reference point is 20% B_0. The harvest control rule reduces catch when biomass declines below the target, with more severe reductions below the soft limit. Similar harvest strategies and reference points are used in the Australian fisheries management system, and in U.S West Coast and Alaska fisheries.

This framework has operated satisfactorily for the well-studied stocks, but the panel notes several structural issues with how reference points are defined and applied across the New Zealand orange roughy complex. These are discussed in the subsections below.

8.2 The B_0 Benchmark and Its Limitations

Expressing stock status as B/B_0 presupposes that B_0, the unfished equilibrium spawning biomass, is both estimable and a meaningful biological benchmark. For ESCR orange roughy, neither condition is straightforwardly satisfied. As discussed in Section 5, B_0 is estimated from the model's stock-recruit relationship parameters (R_0 and steepness h) in combination with life-history schedules. With one-way-trip data and a hypothesised recruitment regime shift around 1980, B_0 represents an unfished state that may never have been observed by the fisheries and whose estimation is dominated by the prior on steepness rather than the data. A stock status determination of, say, 25% B_0 carries a precision that is inherited from the steepness prior, not from empirical information.

A deeper concern is whether B_0 under pre-1980 recruitment conditions is the appropriate reference point. If the post-1980 reduction in recruitment success is a persistent regime shift driven by oceanographic change rather than fishing effects, then the unfished biomass achievable under current productivity conditions is substantially lower than B_0 estimated from the full life history. Comparing current biomass to the old B_0 would then systematically understate the degree of stock rebuilding possible under current environmental conditions. If the assessment continues to include a post-1980 reduction in recruitment, the panel recommends that assessments explore sensitivity of stock status conclusions to alternative B_0 definitions, including a conditional B_0 computed under post-1980 recruitment parameters.

8.3 B_{MSY} -Based Reference Points

An alternative to B/B_0-based reference points is to express stock status relative to B_{MSY} — the biomass expected at maximum sustainable yield. B_{MSY} is the reference point embedded in international agreements (e.g., the UN Fish Stocks Agreement) and is the natural target when the management objective is to maintain or rebuild stocks to levels that support sustainable maximum yield. For a Schaefer-type surplus production model, $B_{MSY} = 0.5\ B_0$; for a Beverton-Holt stock-recruit relationship with low steepness (as implied by orange roughy's life history), B_{MSY}/B_0 is typically in the range 0.3–0.5, depending on steepness. The current New Zealand target of 30–50% B_0 is therefore broadly consistent with a B_{MSY}-based target, but the correspondence is implicit rather than explicit.

Making the B_{MSY} target explicit has several advantages. First, it anchors the reference point to a biologically meaningful objective — maximising sustainable yield — rather than a fraction of an uncertain B_0 estimate. Second, it facilitates comparison across stocks and with international assessments that use B_{MSY} as a standard benchmark. Third, it is consistent with the re-parameterisation of the assessment model recommended in Section 9, under which *MSY* and F_{MSY} are leading parameters and B_{MSY} is a derived quantity. The panel suggests that Fisheries New Zealand explicitly define reference points in terms of B_{MSY} and F_{MSY}, with B_0 used as a communication tool.

8.4 Harvest Rate Reference Points and the U Notation

New Zealand orange roughy assessments have historically expressed exploitation as a fishing mortality rate *F* (annual instantaneous rate), but some analyses and management documents use an annual harvest rate *U* (the fraction of vulnerable biomass removed per year, equivalent to $F/Z(1 - e^{-Z})$ for a simplified model). The distinction matters for communication and for comparing assessments across stocks: *F* and *U* diverge significantly at higher exploitation rates, and mixing the two notations in the same document or cross-stock comparison creates confusion. The panel recommends that all New Zealand orange roughy assessments and summary documents adopt a consistent notation and define the harvest rate metric clearly, including specifying whether *F* is instantaneous or annual and whether *U* is computed on spawning biomass or vulnerable biomass.

The choice of F_{MSY} as the harvest rate reference point is also consequential for orange roughy given the species' life history. With $M = 0.045$ yr^{-1}, F_{MSY} from per-recruit analysis is approximately 0.02–0.03 yr^{-1}, implying *MSY* harvest rates of roughly 2–3% of vulnerable biomass per year. These are very low by the standards of most commercial fisheries, and they mean that even modest overfishing — say $F = 1.5 \times F_{MSY}$ — represents an absolute rate of exploitation that remains low but that cannot be sustained without stock decline given the species' slow productivity. The harvest control rule must therefore be calibrated to keep *F* in a very narrow range, and the uncertainty in F_{MSY} itself must be propagated into the control rule design.

8.5 Life-History Cross-Checks on F_{MSY}

The panel has previously noted (Section 5.3.1–3) that F_{MSY} derived from the fitted CASAL model inherits confounding from B_0, steepness, and *q*. An independent cross-check can be constructed from life-history parameters alone, using per-recruit analysis. Given estimates of natural mortality (*M*), the growth schedule, and the maturity ogive, per-recruit analysis yields the fishing mortality rate at which yield per recruit is maximised (F_{max}) and related rates ($F_{0.1}$, $F_{35\%}$) that are commonly used as proxies for F_{MSY}. For orange roughy with $M = 0.045$ yr^{-1}, these calculations consistently place F_{MSY} in the range 0.02–0.03 yr^{-1}, regardless of the stock-recruit relationship parameters.

This life-history F_{MSY} estimate is valuable precisely because it is independent of the data-limited stock-recruit relationship. It should be reported alongside the model-estimated F_{MSY} in each assessment, and material divergence between the two should trigger scrutiny of the model parameterisation. More formally, the per-recruit F_{MSY} range can be incorporated as an informed prior on F_{MSY} under the re-parameterised model structure recommended in Section 9. This provides a prior that is anchored to observable life-history quantities and that propagates uncertainty in *M* realistically; if *M* is uncertain, F_{MSY} is uncertain, and both uncertainties should be represented in the posterior on stock status. An equilibrium age-structured model implementing this re-parameterisation using the R software is provided in Appendix 4.

8.6 Non-Stationarity and the Validity of Fixed Reference Points

A fundamental assumption of the current reference point framework is that B_0 and B_{MSY} are stationary quantities, that the same unfished biomass and maximum sustainable yield apply throughout the assessment period and into the future. For ESCR orange roughy, this assumption is questionable on several grounds. Although an assumption, the post-1980 recruitment shift represents a step change in productivity, and there is no basis for assuming that the pre-1980 productivity level will return. If the shift is real, and was driven by a regime change in ocean circulation or prey availability rather than fishing effects on spawning stock size, then the stock is operating on a permanently lower production function, and reference points derived from pre-1980 productivity are not achievable targets but artefacts of a historical regime that no longer exists.

If the assessment continues to include a post-1980 reduction in recruitment, the panel recommends that the sensitivity of reference point calculations to the post-1980 productivity regime be explicitly evaluated in each assessment. This could take the form of computing B_0 and B_{MSY} under a model conditioned on post-1980 recruitment data only, or under a time-varying productivity model that allows the stock-recruit relationship to shift. The resulting alternative reference points should be presented alongside the standard estimates to bound the uncertainty in the management target. In the longer term, the transition to an MP-based management framework (Section 7.4) provides the most robust approach to managing under non-stationary productivity, because the operating model ensemble can explicitly include alternative productivity scenarios and the harvest control rule can be tested against all of them.

9. Recommendations

9.1 Overview

This section consolidates the panel's recommendations arising from the review. Recommendations are organised thematically under six headings: acoustic survey methods; biological parameters and ageing; age data collection and sampling design; model structure and parameterisation; assessment quality control; and reference points and management procedures. Recommendations are numbered sequentially for ease of reference in the science work plan process.

The panel emphasises that the recommendations below are not independent. Several of the most important issues (i.e., the unresolved acoustic frequency divergence, the fixed natural mortality, the narrow *q* prior) are structurally linked. Progress on any one of them without corresponding progress on the others will deliver incomplete benefit. The panel therefore recommends that Fisheries New Zealand treat the recommendations as a package and plan to address the linked issues in a coordinated sequence.

9.2 Acoustic Survey Methods

Recommendation 1 — Resolve the 38 kHz / 120 kHz divergence as a priority. The divergence between the two acoustic frequencies since 2017 is the single most important unresolved technical issue in the New Zealand orange roughy assessment programme. The panel recommends a structured programme of concurrent dual-frequency surveys, at a minimum for ESCR and NWCR, with standardised sphere calibration before and after each survey leg. Comparisons should be made at the level of individual echo-integration intervals and across surveys to characterise whether the divergence is spatially uniform or localised, and whether it has changed over time. A field comparison between echosounders used during the survey and independent 38 and 120 kHz systems should also be conducted to determine whether the divergence has an equipment origin. Until the divergence is resolved, both frequency series should be carried as separate scenarios in stock assessments, with management advice tested for robustness to both (see Recommendation 23).

Recommendation 2 — Standardise acoustic absorption coefficient equations. All New Zealand orange roughy surveys should use a consistently specified absorption equation, with the choice (Francois & Garrison 1982 vs. Doonan et al. 2003) documented and applied uniformly. Where historical surveys used different equations, a correction factor should be computed and applied to place all years on a common basis. The magnitude of the equation effect (up to 30% at survey depths) is large relative to the management decisions being made and should not be left as an unquantified source of inter-annual variation.

Recommendation 3 — Develop an explicit availability and detectability framework. The assumption that fish must aggregate (plume) at a density exceeding a detection threshold before they are acoustically visible introduces a non-linear and potentially density-dependent component into *q*. Research should be directed at quantifying the relationship between local density, pluming behaviour, and acoustic detectability. Where commercial vessel operations overlap with survey timing, quantify and correct for any reduction in survey *q* attributable to fishing disturbance.

Recommendation 4 — Expand the *q* prior to include availability and detectability uncertainty. The current *q* prior (mean 0.8, *CV* 19%) captures only target strength uncertainty. Availability, the fraction of the spawning stock that is present in the survey stratum and forming aggregation plumes during the survey window, and coverage, (i.e., the fraction of the surveyed area that is acoustically sampled,) are likely to be as uncertain as TS and perhaps more so. The panel recommends that all three components be quantified separately and combined to produce a composite *q* prior with a *CV* substantially wider than the current 19%. In the interim, a sensitivity analysis using a broader q prior (CV 30–50%) and range of q means should accompany each stock assessment.

Recommendation 5 — Develop a pixel-by-pixel dB-differencing filter to exclude backscatter not consistent with orange roughy, based on per-cell comparisons between the 38 kHz and 120 kHz echosounders; this approach would provide a more granular tool for screening for contamination by fish with swim-bladders than current methods.

Recommendation 6 — Reserve survey time to explore for new aggregations. Each acoustic survey should include a dedicated allocation of time to explore areas adjacent to known spawning features for previously uncharacterised orange roughy aggregations. Discoveries of new aggregations would revise upward the effective survey coverage and could meaningfully change stock status conclusions, particularly for stocks where spatial coverage of the assessment area is incomplete.

9.3 Biological Parameters and Ageing

Recommendation 7 — Develop an informed prior on natural mortality. Natural mortality is fixed at $M = 0.045\ yr^{-1}$ in most New Zealand orange roughy assessments, with no uncertainty propagated to derived quantities. The panel recommends adopting the Hamel & Cope (2022) lognormal prior: median 5.4/*Amax* and log-scale standard deviation of 0.31, where *Amax* is the maximum observed age. This prior is derived by fitting the one-parameter Hoenig model under a log-log transformation (correcting the inconsistent transformation in Then et al. 2015), accounts for uncertainty in both *Amax* and *M*, and is the current standard in US West Coast groundfish assessments. For orange roughy with *Amax* ≈ 150 years, the prior median is approximately $0.036\ yr^{-1}$, somewhat below the current fixed value of $0.045\ yr^{-1}$. The prior should be incorporated into the CASAL model and the resulting posterior uncertainty in *M* propagated through to stock status and reference point estimates. Prior sensitivity should be evaluated using both *Amax* = 150 years (current assumption) and the maximum age in the available otolith collection aged under the Horn et al. (2016) protocol (Recommendations 11–12).

Recommendation 8 — Replace the simplified ageing error specification with an age-reading error matrix. The current CASAL specification uses a normal distribution with $CV = 0.1$ as a proxy for ageing error. This symmetric approximation does not capture inter-reader variability, systematic directional bias at older ages, or the structural differences between age classes where the otolith zone count is unambiguous and those where it is not. Australian orange roughy assessments use a full ageing error matrix (Punt et al. 2008), and the New Zealand programme should adopt the same approach. Software for applying this approach is available at https://github.com/pfmc-assessments/AgeingError . Construction of the matrix requires a double-read inter-reader study on a subsample of the existing otolith collection (Recommendation 11 below). Until the matrix is available, assessments should include a sensitivity run with a wider ageing error specification to bound the effect on estimated age compositions.

Recommendation 9 — Estimate a maturity-at-age curve for each stock from otolith transition zone data. The opaque-to-translucent transition zone in orange roughy otoliths is associated with first spawning and provides a direct biological marker of maturity that is independent of the fish's reproductive condition at the time of capture. The large existing otolith collection, the majority of which was sourced from spawning aggregations where maturity staging was recorded, provides a substantial and largely untapped dataset for fitting a maturity ogive from transition-zone observations. The panel recommends that a subsample of the existing collection be examined specifically to score presence or absence of the transition zone, and that a logistic maturity curve be fitted to these data as a function of estimated age. The resulting empirical maturity ogive should replace the fixed schedule currently assumed in CASAL, and uncertainty in maturity-at-age should be propagated through per-recruit calculations of F_{MSY} (Recommendation 16). This work is complementary to the stratified sampling design of Recommendation 11.

Recommendation 10 — Investigate sex-specific natural mortality. Although the standard model assumes a common *M* for both sexes, there are biological reasons to expect sex differences in longevity. If otolith samples show a progressive increase in the proportion of females among the oldest age classes, this pattern is consistent with lower *M* in females than in males. The panel recommends that the existing age-structure data be examined for sex-ratio trends across age classes, and that if such a trend is present, a model incorporating sex-specific *M* be developed and evaluated as a sensitivity. This analysis requires that sexing be recorded on a sufficient proportion of aged otoliths. It should also be noted if sex-specific fishing mortality is occurring, it would bias estimates of sex-specific *M*, which highlights the limitations of simple empirical analyses. A model-based approach, in which sex-specific selectivities are estimated along with sex-specific *M*, as with the most recent Australian orange roughy assessment, may be needed to fully evaluate these interrelated processes.

9.4 Age Data Collection and Sampling Design

Recommendation 11 — Develop a stratified otolith sampling design. The existing otolith collection (>321 000 otoliths, fewer than 4% aged) is a major underutilised scientific asset. The sampling design that generated this collection was not always spatially stratified by spawning feature, leading to documented instabilities in estimated age compositions across years (ESCR 2023, NWCR 2022). The panel recommends development of a formal sampling design with spatial stratification proportional to acoustic biomass by spawning feature, applied consistently in all future sampling programmes. Sampling effort should be spread as broadly as possible across tows within each stratum, with the same principle applied to selection of otoliths for ageing from each tow, to avoid pseudo-replication and to maximise the coverage of the age composition estimate. Ageing effort should be allocated to produce a well-designed subsample from the existing collection under this design before the next assessment, prioritising the ESCR stock, to maximise the value of already-collected material.

Recommendation 12 — Implement routine age sampling from commercial catches. The current biological sampling programme is concentrated on commercial fishing by the vessel also undertaking spawning aggregation acoustic surveys. Observer-collected otoliths from commercial trawl operations represent an independent, year-round data stream on harvested age structure that is currently unused in assessments. The panel recommends establishing a routine programme to age observer-collected otoliths and incorporate the resulting commercial catch age compositions into future assessments. Where fishing occurs on both spawning aggregations and at other locations, otolith samples should cover both fishing modes so that age selectivity can be estimated separately for each. This provides a check on selectivity assumptions and broadens the representativeness of the age composition data beyond the spawning season.

Recommendation 13 — Estimate separate survey and fisheries selectivity curves, parameterised relative to maturity. Most current assessment models apply a single selectivity-at-age function across both the acoustic survey and the commercial fishery, despite these operations sometimes having different size and age selectivity. Future model structures should estimate separate survey and fisheries selectivity curves, consistent with standard practice in multi-fleet age-structured assessments. This requires the separate commercial age composition data produced by Recommendation 12. Since maturity and selectivity are distinct biological processes, they should not share a common parameterisation, as when assuming selectivity equals maturity. Selectivity curve parameters should be specified as offsets from the maturity curve, enforcing the biological constraint that selectivity cannot greatly precede maturity and ensuring that the biological information about the maturity schedule propagates coherently to selectivity.

Recommendation 14 — Explore fitting conditional age-at-length and length composition data in assessment models. Conditional age-at-length (CAAL) analysis allows otolith age readings to be used alongside length composition data without requiring a fully pre-specified growth model, and makes more efficient use of age-length data collected from the same sampling events. The panel recommends that the feasibility of fitting CAAL observations and length frequencies within the CASAL framework be evaluated, and that trial model runs including these data be presented alongside the base model in future assessments. This approach has the potential to extract additional information from existing biological collections without requiring new sampling.

9.5 Model Structure and Parameterisation

Recommendation 15 — Re-parameterise assessment models with *MSY* and F_{MSY} as leading parameters. Under the current CASAL parameterisation, the model estimates R_0 (virgin recruitment) and steepness h, with *MSY* and F_{MSY} derived as functions of these and other parameters. With one-way-trip fishery data, R_0 and h are highly confounded and neither can be estimated independently of the stock-recruit relationship; however, there is information about *MSY*. Re-parameterising the model so that *MSY* and F_{MSY} are the directly estimated quantities avoids forcing the model to resolve a confound that the data cannot support, and aligns the model output directly with management quantities. An informed prior on F_{MSY} can then be constructed from per-recruit analysis (see Recommendation 16), and an empirically grounded prior on *MSY* can be derived from the catch history (Martell et al. 2008).

Recommendation 16 — Incorporate a life-history informed prior on F_{MSY}. Per-recruit analysis using the orange roughy life-history schedule (M = 5.4/max age, observed growth, observed maturity ogive) places F_{MSY} in the range of approximately 0.02–0.03 yr^{-1}, independently of the stock-recruit relationship. This range should be used to construct an informative prior on F_{MSY} under the re-parameterised model of Recommendation 15. Because this prior is derived from life-history data rather than from B_0 and steepness estimates, it is independent of the confounding that affects model-derived F_{MSY}, and provides a useful cross-check on the latter. Uncertainty in M should be propagated through the per-recruit analysis to produce a distribution over the F_{MSY} range rather than a point estimate.

Recommendation 17 — Incorporate a catch-history informed prior on *MSY*. The historical catch record provides a non-trivial lower bound on *MSY*: the maximum sustained catch observed while the stock was not driven to zero cannot exceed *MSY*. 'Maximum sustained catch' is intended as the highest catch level maintained over a period sufficiently long relative to the stock's response time without an observed decline in the abundance index; for orange roughy the panel suggests a window of at least one to two decades. The panel acknowledges that for New Zealand orange roughy stocks such a period may not exist, in which case the bound is uninformative and should be reported as such. A uniform prior on *MSY* bounded below by the maximum observed sustained yield and above by a life-history productivity ceiling provides an empirically grounded prior that replaces the current diffuse prior on R_0. Together with the F_{MSY} prior of Recommendation 16, these two priors jointly constrain the model's productive capacity without imposing an arbitrary point estimate on the R_0/steepness confound, or an informative prior on q.

Recommendation 18 — Explore hierarchical Bayesian model structures for low-information stocks. All New Zealand orange roughy stocks share the same species biology, and parameters such as M, growth, acoustic catchability, selectivity, year-class strength, and post-1980 recruitment dynamics are not independent across stocks — they are draws from a common among-stock distribution. A hierarchical model estimates this hyperprior distribution jointly across all stocks, allowing data-rich stocks (ESCR, NWCR, ORH 7A) to inform the hyperpriors from which lower-information stocks (ORH 1, East Cape 2A) borrow strength. Hierarchical models of this type are in routine use for Pacific groundfish assemblages and are feasible within a Bayesian MCMC framework such as CASAL2 or Stan.

Recommendation 19 — Evaluate the logit-transformed proportions method for modelling acoustic biomass distribution across features. The management procedure evaluation by Neubauer et al. (2026) uses logit-transformed proportions to model the changing spatial distribution of acoustic survey biomass across spawning features. This approach makes use of all available survey data and addresses a recognised problem in orange roughy assessments: that changes in which features are surveyed, and when, can confound biomass trends with distributional shifts. The panel recommends that this spatial modelling method be evaluated for incorporation into the primary stock assessment models, where it could help stabilise the acoustic time series and reduce sensitivity of the assessment to incomplete or uneven feature coverage.

Recommendation 20 — Evaluate the ESCR acoustic time series partitioning. For the ESCR assessment, there is an unresolved tension between the acoustic survey estimates from 2002–2010, which covered only a portion of the stock area and could not be screened using dB differencing, and those from 2011–2024. The transition from hull mounted to a towed transducer also corresponds to the same time period of the unresolved tension in the acoustic series and may represent a systematic change in q and warrant a split in the time series. Model runs in which the earlier time series is down-weighted or excluded should be systematically evaluated and presented alongside the base model. A model conditioned only on the 2011–2024 acoustic time series plus the full catch history provides a self-contained check on whether conclusions are sensitive to the earlier, more uncertain biomass estimates. For example, is there still evidence of reduced recruitment post-1980 when only the 2011–2024 acoustic time series is used?

9.6 Assessment Quality Control

Recommendation 21 — Develop a checklist of required elements for stock assessments, conduct retrospective analysis, and report MCMC diagnostics. A retrospective analysis (peeling the terminal year of data sequentially) should be conducted on the final model for each stock assessment to detect systematic retrospective bias in stock status estimates, which can indicate model misspecification or data inconsistency. MCMC convergence diagnostics — including R-hat statistics, effective sample sizes, trace plots, and posterior predictive checks — should be reported for the final model run in every assessment document. A standardised checklist of required items should be developed and applied consistently across all orange roughy stock assessments, so that reviewers and managers can readily verify that quality standards have been met.

9.7 Reference Points, Management Procedures

Recommendation 22 — Define reference points explicitly in terms of B_{MSY} and F_{MSY}. The current framework expresses stock status as B/B_0 and compares to fixed fractions of B_0. The panel recommends that reference points be reformulated in terms of B_{MSY} and F_{MSY}, which are biologically meaningful quantities (maximum sustainable yield) and are directly estimable under the re-parameterised model of Recommendation 15. B_0 should be retained as a communication tool. Where B_0 is used, assessments should report sensitivity of stock status conclusions to the assumed productivity regime (pre- vs. post-1980).

Recommendation 23 — Complete and adopt the Neubauer et al. (2026) MP evaluation for ESCR. The management procedure evaluation for the East Coast South Island (ESCR) stock, incorporating the 38 kHz / 120 kHz acoustic divergence as an operating model uncertainty an important near-term modelling deliverable. The panel recommends that completion of this work be prioritised and, subject to satisfactory results, adopt the resulting management procedure for ESCR in the next assessment review. If possible, the framework should then be extended to NWCR and ORH 7A.

Recommendation 24 — Employ Stock Reduction Analysis as a precautionary lower bound. For stocks with limited or uninformative data, the SRA approach (Walters et al. 2006) should be computed alongside the assessment models central estimate and reported in assessment documents. The SRA provides the lower tail of the plausible stock status distribution that a single-model CASAL posterior, conditioned on its priors, may underrepresent.

Recommendation 25 — Routinely fit surplus production models as cross-checks. For stocks with a usable abundance index and catch history, a state-space surplus production model (JABBA or equivalent) should be fitted and reported alongside the CASAL model as a standard cross-check. Material divergence between the two should trigger investigation of the source of disagreement rather than being resolved by expert judgment. For stocks where CASAL cannot be fitted reliably due to data limitations, production models may be the most reliable estimator of stock status.

Recommendation 26 — Explore the application of spatiotemporal CPUE standardisation for data-poor stocks. For stocks with limited or no acoustic survey time series, CPUE data from commercial logbooks offer a potentially useful relative abundance index. However, nominal CPUE is confounded by changes in targeting behaviour, vessel efficiency, and spatial distribution. For data-poor orange roughy stocks, the panel recommends that CPUE standardisation be performed using spatiotemporal models (e.g., sdmTMB), which explicitly account for spatial autocorrelation and changes in the spatial footprint of the fisheries. Standardised CPUE indices produced by this approach should be evaluated as auxiliary data inputs in surplus production models (Recommendation 25) and as a cross-check on assessment conclusions.

10 References

Bravington, M.V.; Grewe, P.M.; Davies, C.R. (2016a). Absolute abundance of southern bluefin tuna estimated by close-kin mark-recapture. *Nature Communications*. https://doi.org/10.1038/ncomms13162

Bravington, M.V.; Skaug, H.J.; Anderson, E.C. (2016b). Close-kin mark recapture. *Statistical Science, 31*, 259–274.

Burch, P.; Sutton, C. (2026). Eastern Zone Orange Roughy (*Hoplostethus atlanticus*) stock assessment based on data up to 2024. https://doi.org/10.25919/czp7-cx76

Coombs, R.F.; Barr, R. (2007). In situ measurements of orange roughy (*Hoplostethus atlanticus*) target strength. *ICES Journal of Marine Science*, *64*(6): 1220–1234.

Cordue, P.L. (2014a). A Management Strategy Evaluation for orange roughy, ISL Client Report for Deepwater Group Ltd, p. 42.

Cordue, P.L. (2014b). The 2014 orange roughy stock assessments. *New Zealand Fisheries Assessment Report*, *50*. 135 p.

Deepwater Group Ltd. (2024). 2020 Stock Assessment Update of ORH 3B East and South Chatham Rise. *New Zealand Fisheries Assessment Report 2024/11*.

Doonan, I.J. (1994). Life History Parameters of Orange Roughy: Estimates for 1994. *New Zealand Fisheries Assessment Research Document 94/19*.

Doonan, I.J.; Coombs, R.F.; McClatchie, S. (2003). The absorption of sound in seawater in relation to the estimation of deep-water fish biomass. *ICES Journal of Marine Science*, *60*(5), pp.1047–1055.

Dunford, A.J. (2005). Correcting echo-integration data for transducer motion. *Journal of the Acoustical Society of America 118 (4)*: 2121–2123. https://doi.org/10.1121/1.2005927

Dunn, M.R. (2024). Assessment of the Challenger Plateau orange roughy stock for 2024. *New Zealand Fisheries Assessment Report 2024/58*.

Dunn, M.R. (2026). Stock assessment of orange roughy where CPUE is the only biomass index. Presentation for the orange roughy review, June 2026.

Dunn, M.R.; A'mar, T.; Doonan, I. (2022). Assessment of the Mid-East Coast orange roughy stock for 2022. *New Zealand Fisheries Assessment Report 2022/38*. 77 p. https://fs.fish.govt.nz/Doc/25285/FAR-2022-38-Mid-East-Coast-Orange-Roughy-Stock-Assessment-2022-4308.pdf.ashx

Dunn, M.R.; Datta, S.; Doonan, I.J. (2025a). Stock assessment of Chatham Rise orange roughy in 2025. *New Zealand Fisheries Assessment Report 2025/46*.

Dunn, M.R.; Doonan, I.J.; Anderson, O.F. (2025b). Additional analyses for orange roughy (ORH 3B). *New Zealand Fisheries Assessment Report 2025/18*.

Dunn, M.R., et al., (2026a). Stock assessment of orange roughy in ORH3B: East & South Chatham Rise, including a short history (why we ended up where we did). Presentation for the orange roughy review, May 2026.

Dunn, M.R., et al., (2026b). Stock assessment of orange roughy in ORH3B. The 2025 assessment of Northwest Chatham Rise. Presentation for the orange roughy review, June 2026.

Dunn, M.R., et al., (2026c). Stock assessment of orange roughy in ORH7A. The 2024 assessment. Presentation for the orange roughy review, June 2026.

Dunn, M.R., et al., (2026d). Stock assessment of orange roughy for the Mid-East Coast. The 2022 assessment. Presentation for the orange roughy review, June 2026.

Do, M.A.; Coombs, R.F. (1989). Acoustic measurements of the population of orange roughy (*Hoplostethus atlanticus*) on the North Chatham Rise, New Zealand, in winter 1986. *New Zealand Journal of Marine and Freshwater Research*, 23: 225–237.

Edwards, C.T.; Roux, M.J.; Clark, M.R. (2022). Spatially integrated modelling of data-limited orange roughy (*Hoplostethus atlanticus*) using environmental covariates. *Canadian Journal of Fisheries and Aquatic Sciences*, *79*(3): 448–457.

Fisheries New Zealand. (2025a). Fisheries Assessment Plenary pages 907–919, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New

Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69783-Fisheries-Assessment-Plenary-May-2025-Volume-2-INTRODUCTION-ORANGE-ROUGHY-ORH

Fisheries New Zealand. (2025b). Fisheries Assessment Plenary pages 950–991, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/70159-Fisheries-Assessment-Plenary-May-2025-Volume-2-Orange-roughy-ORH-3B

Fisheries New Zealand. (2025c). Fisheries Assessment Plenary pages 920–926, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69831-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-NORTHERN-NORTH-ISLAND-ORH-1

Fisheries New Zealand. (2025d). Fisheries Assessment Plenary pages 992–1006, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69828-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-CHALLENGER-PLATEAU-ORH-7A

Fisheries New Zealand. (2025e). Fisheries Assessment Plenary pages 927–949, May 2025: stock assessments and stock status. Compiled by the Fisheries Science Team, Fisheries New Zealand, Wellington, New Zealand. 1955 p. https://www.mpi.govt.nz/dmsdocument/69840-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-CAPE-RUNAWAY-TO-BANKS-PENINSULA-ORH-2A-2B-3A

Francis, R.C., (2011). Data weighting in statistical fisheries stock assessment models. *Canadian Journal of Fisheries and Aquatic Sciences*, *68*(6):1124–1138.

Francois, R.E.; Garrison, G.R. (1982). Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption. *The Journal of the Acoustical Society of America*, *72*(6): 1879–1890.

Fox Jr, W.W. (1970). An exponential surplus-yield model for optimizing exploited fish populations. *Transactions of the American Fisheries Society*, *99*(1): 80–88.

Hamel, O.S. (2015). A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. *ICES Journal of Marine Science*, *72*(1): 62–69.

Hamel, O.S.; Cope, J.M. (2022). Development and considerations for application of a longevity-based prior for the natural mortality rate. *Fisheries Research*, *256*, p.106477.

Horn, P.L.; Tracey, D.M.; Doonan, I.J.; Krusic-Golub, K. (2016). Age determination protocol for orange roughy (*Hoplostethus atlanticus*). *New Zealand Fisheries Assessment Report 2016/03*. 30 p.

Kloser, R.J.; Macaulay, G.J.; Ryan, T.E.; Lewis, M. (2013). Identification and target strength of orange roughy (*Hoplostethus atlanticus*) measured in situ. *The Journal of the Acoustical Society of America*, *134*(1): 97–108.

Ona, E.; Mitson, R.B. (1996). Acoustic sampling and signal processing near the seabed: the deadzone revisited. *ICES Journal of Marine Science*, *53*(4): 677–690.

Mace, P. (2026). A brief history of orange roughy (*Hoplostethus atlanticus*) fisheries in New Zealand. Presentation for the orange roughy review, May 2026.

Martell, S.J.D.; Pine W.E.; Walters C.J. (2008). Parameterizing age-structured models from a fisheries management perspective. *Canadian Journal of Fisheries and Aquatic Sciences*. 65(8): 1586-1600. https://doi.org/10.1139/F08-055

McAllister, M.; Kirchner, C. (2002). Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. *Bulletin of Marine Science*, 70, 499-540

Neubauer, P.; Kim, K.; Large, K.; Langley, A.; Tingley, G. (2026). Evaluating management procedures for ORH 3B ESCR, presentation to the Independent International Review of Orange Roughy Stock Assessment Modelling.

O'Driscoll, R. (2026). Acoustic surveys for orange roughy. Presentation for the orange roughy review, June 2026.

Pennington, M.; Vølstad, J.H. (1994). Assessing the effect of intra-haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics*: 725–732.

Punt, A.E.; Burch, P.; Privitera-Johnson, K.M.; Bessell-Browne, P.; Tuck, G.N. (2026). Comparison of treatments of natural mortality in integrated assessment models, with reference to fish stocks off southeast Australia: Towards best practices for estimation. *Fisheries Research*, 293: 107631. https://doi.org/10.1016/j.fishres.2025.107631

Punt, A.E.; Smith, D.C.; Krusic-Golub, K.; Robertson, S. (2008). Quantifying age-reading error for use in fisheries stock assessments, with application to species in Australia's southern and eastern scalefish and shark fishery. *Canadian Journal of Fisheries and Aquatic Sciences*, 65(9): 1991–2005.

Ryan, T.E.; Kloser, R.J. (2016). Improved estimates of orange roughy biomass using an acoustic-optical system in commercial trawlnets. *ICES Journal of Marine Science*, *73*(8): 2112–2124.

Ryan, T.; Tilney, R.L. (2025). Acoustic biomass surveys of orange roughy spawning aggregations in ORH 3B NWCR and ORH 3B ESCR Chatham Rise, June/July 2022. *New Zealand Fisheries Assessment Report 2025/23*. 71 p.

Ryan, T.; Tilney, R.; Kunnath, H. (2025). Acoustic biomass surveys of orange roughy spawning aggregations in ORH 3B ESCR Chatham Rise, June–July 2024. *New Zealand Fisheries Assessment Report 2025/23*. 77 p.

Schaefer, M.B. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Bulletin of the Inter-American Tropical Tuna Commission 1(2)*: 25–56.

Stephenson, F.; Biggerstaff, A.; Dunn, M.; A'mar, T.; Doonan, I.; Rowden, A.; Geange, S. (2022). A 2022 orange roughy stock assessment for Louisville Ridge, West Norfolk Ridge, Lord Howe Rise, and Northwest Challenger Plateau. South Pacific Regional Fisheries Management Organisation, 10th Meeting of the Scientific Committee, Seoul, Korea, 26-30 September 2022. SC10-DW01_rev1. 37 p.

Terms of Reference for a Fully Independent International Review of Orange Roughy Stock Assessment Modelling. (ToR; 2026). Fisheries New Zealand, 8 May 2026.

Then, A.Y.; Hoenig, J.M.; Hall, N.G.; Hewitt, D.A. (2015). Evaluating the predictive performance of empirical estimators of natural mortality rate using information on over 200 fish species. *ICES Journal of Marine Science*, *72*(1): 82–92.

Thomson, R.; Bravington, M.; Lloyd-Jones, L.; Baylis, S.; Feutry, P.; Gunasekera, R. (2025). Close-Kin Mark-Recapture data-driven models for school shark in the SESSF. Hobart, CSIRO https://doi.org/10.25919/5w0g-yk12

Thomson, R.B.; Klaer, K.; Tremblay-Boyer, L.; Bessell-Browne, P.; Bayliss, S.; Burch, P.; Dell, J.; Bravington, M.V. (2026). Close-Kin Mark-Recapture design for key and rebuilding stocks in the SESSF. May 2026, CSIRO, Australia.

Thomson, R.B.; Bravington, M.; Feutry, P.; Gunasekera, R.; Grewe, P. (2020). Close kin mark recapture for School Shark in the SESSF (No. FRDC Project No. 2014/024). CSIRO Oceans and Atmosphere, Hobart.

Walters, C. (2003). Folly and fantasy in the analysis of spatial catch rate data. *Canadian Journal of Fisheries and Aquatic Sciences*, 60: 1433-1436.

Walters, C.J.; Martell, S.J.D.; Korman, J. (2006). A stochastic approach to stock reduction analysis. *Canadian Journal of Fisheries and Aquatic Sciences*, 63: 212–223.

APPENDIX 1: WORKSHOP AGENDA

Pre-workshop virtual half-day (20 May 2026)

The pre-workshop virtual half-day was held on 20 May 2026. Presentations were restricted to providing background context; questions and comments were limited to panel members unfamiliar with New Zealand orange roughy assessments.

History of orange roughy fisheries in New Zealand: context and facts (Pamela Mace)

The latest ESCR assessment (Matt Dunn)

Changes from previous assessments (Matt Dunn)

Main workshop: Agenda (8–12 June 2026)

Date	Topic	Presenter(s)
Monday morning (8 June)	Remaining high-information stock assessments: (ii) NWCR (iii) Challenger ORH 7A	Matt Dunn Discussion facilitated by Chair
Monday afternoon (8 June)	(iv) Mid east coast (MEC), plus further discussion of (i) ESCR Discussion topics: • Data not used and why • Differing aspects of these four assessments • Which features make the most sense • Recommendations	Facilitated by Chair
Tuesday morning (9 June)	• Low-information stocks: ORH 1, East Cape 2A • Discussion of pros and cons of acoustic surveys • Other data inputs and biological parameters; assumed productivity • Alternative approaches for conducting these assessments: recommendations	Matt Dunn Richard O'Driscoll / other acoustic experts (possibly virtual) Discussion facilitated by Chair
Tuesday afternoon (9 June)	Presentation and discussion on surplus production models for assessments and Management Procedure Evaluations	Phil Neubauer Discussion facilitated by Chair

Wednesday–Thursday (10–11 June)	International panel sequestered: further discussion and drafting of report with conclusions and recommendations	Panel only. Panel may contact outside participants by email or phone; outside participants may not initiate contact with the panel.
Friday morning (12 June)	International panel presents draft conclusions and recommendations: • Best approaches for high(ish)-information and low-information stocks • Alternative assessment approaches • Refinements to existing data collection programmes • Alternative data collection programmes • Alternative biological parameters	Facilitated by Chair
Friday afternoon (12 June)	International panel modifies / finalises report	

APPENDIX 2: Terms of Reference for the Orange Roughy stock Assessment review

Ministry for Primary Industries
Manatū Ahu Matua

Fisheries New Zealand
Tini a Tangaroa

Fisheries New Zealand Terms of Reference for a Fully Independent International Review of Orange Roughy Stock Assessment Modelling

1. Background

Orange roughy (*Hoplostethus atlanticus*) is a unique species, with no analogue in the world of commercial fisheries. They live mostly in depths of 750-1200 m, extending down to at least 1500 m. They live to about 120-150 years of age, with one otolith read as 230-245 years old. They grow slowly with a maximum size of about 45 cm in New Zealand waters. They do not mature until age 25-35 and may not spawn every year once mature. They produce a relatively small number of eggs (an average of 40 000–60 000 eggs) that are large (2-3 mm in diameter) by fish standards.

Orange roughy was not commercially fished until 1979 in New Zealand waters, with most fisheries in other parts of the world starting in the 1980s or 1990s. New Zealand, Australia and, to a lesser extent, some other parts of the world (e.g. Namibia, Chile, Southern Indian Ocean, and NE Atlantic countries) have conducted considerable research, yet have barely covered a single generation of the species' life history (in New Zealand, the estimated average generation time for the largest stock is 41 years).

Analysis of CPUE data has largely failed due to hyperdepletion and suspected susceptibility to disturbance by fishing activities, which appear to break up aggregations for days or weeks. Research trawl surveys have also largely failed, due to the aggregating nature of the species, particularly during spawning, but also when feeding.

Acoustic surveys of various kinds may also only work in specific circumstances. For example, fishing vessel hull-mounted acoustic surveys do not work on hills and knolls due to a large shadow zone near the seabed where acoustic signals cannot be detected and where orange roughy often congregate, particularly when disturbed. Acoustic estimates are also complicated by orange roughy having a swim bladder filled with lipids rather than gas, which means orange roughy have a low target strength relative to other species with gas-filled bladders at similar depths. It only takes a small amount of gas-filled bladder species to erroneously inflate estimates of orange roughy biomass, even when using towed body acoustics.

Estimation of orange roughy age frequencies have also proven problematic for two reasons. First, orange roughy often tend to segregate by sex and possibly size on the spawning grounds with females on the bottom of spawning plumes and males on the top. Different spawning and non-spawning areas also tend to have different size (and age) distributions, which makes it difficult to obtain representative sampling. Second, their longevity and the agreed ageing technique (cutting thin otolith sections through the longest axis and mounting on slides, as agreed during a workshop involving New Zealand, Australia and Chile) makes ageing expensive, so relatively few samples have been aged.

The productivity of orange roughy is amongst the lowest of commercially fished species. It took almost 10 years after the start of the fisheries in New Zealand to definitively demonstrate that early growth is exceptionally slow. It had previously been assumed that growth was average (K=0.2) and natural mortality, M, was low (0.1) to average (0.2). More recently, growth (K) has been estimated to be slower (0.059) and natural mortality (M) assumed to be lower (0.045); however, even lower values of M have sometimes been used or advocated.

2. Participants

The independent international review panel experts are:

Paul Burch – CSIRO, Australia: has conducted stock assessments for orange roughy in Australia

Martin Dorn – University of Washington, USA: has conducted stock assessments for other long-lived species

Sean Cox – Consultant, Vancouver, Canada: has extensive stock assessments experience

Steve Martell – Consultant, Seattle, USA: has extensive stock assessment experience

Additional participants

Additional people will participate in the open sessions of the workshop. These will include:

Matt Dunn, who has been the lead research scientist carrying out most of the recent (and many historical) orange roughy stock assessments

Sophie Mormede, who has previously been involved in independent review of orange roughy stock assessments

Gretchen Skea, who is the chair of the Deepwater Working Group

Ian Tuck, who currently chairs the stock assessment plenary meetings, and will chair the open sessions of this meeting

Pamela Mace, who has worked on orange roughy for many years in various capacities including chairing the DWWG and various orange roughy plenary meetings

Phil Neubauer, who will present an outline of an orange roughy management procedure using production models

Richard O'Driscoll, who will participate in discussions about orange roughy acoustics

Other DWWG members: open

3. Terms of Reference

Primary Objectives

The International Review Panel will:

a) Develop recommendations that can inform a science work plan for data collection and stock assessments of orange roughy fisheries in New Zealand
b) Review and comment on recent high information stock assessments (ESCR, NWCR, Challenger ORH 7A, and Mid-east Coast MEC), each of which used somewhat different assumptions
c) Compare and contrast alternative approaches, including previous methodologies, and surplus production models for assessments or Management Procedures
d) Review current and past data inputs (catch history, acoustics, ageing, biological parameters (especially *M*), trawl surveys, CPUE, length data, including how these are reviewed and agreed through review processes
e) To the extent possible, provide recommendations on:
    i. Establishing a science work plan for assessing orange roughy fisheries
    ii. Best approaches for high(ish)-information stocks and for low information stocks
    iii. Alternative types of assessment approaches
    iv. Refinements to existing data collection programmes, for example, the use and frequency of data collection including acoustic surveys as well as observer collected data.
    v. Alternative data collection programmes
    vi. Alternative biological parameters

Secondary Objectives

a) Examine the current estimates of the B_{MSY} level for orange roughy stocks.
b) Provide advice on practical approaches for assessments of low information orange roughy stocks (ORH 1, East Coast area 2A)

Out of scope

- The efficacy of past and present FNZ orange roughy fisheries management measures themselves, or the Quota Management System in general
- Recommended future management actions

- Relitigating past assessments, except for discussing these for illustrative purposes on the way forward
- Scaling of the catch history (previous overruns)

Outputs

The International Review Panel will produce a report with recommendations addressing the Terms of Reference.

4. Background documents

The following documents will be useful to the workshop:

Primary (●) and secondary (○) documents

● Pages 907-919 in Fisheries New Zealand (2025*). Fisheries Assessment Plenary*● Pages 950-991 in Fisheries New Zealand (2025*).* As above.
https://www.mpi.govt.nz/dmsdocument/70159-Fisheries-Assessment-Plenary-May-2025-Volume-2-Orange-roughy-ORH-3B

● Dunn, M.R.; Datta, S.; Doonan, I.J. (2025). Stock assessment of Chatham Rise orange roughy in 2025. New Zealand Fisheries Assessment Report 2025/46. 59 p.
https://fs.fish.govt.nz/Doc/26145/FAR-2025-46-Stock-assessment-of-Chatham-Rise-orange-roughy-in-2025.pdf.ashx

● Dunn, M.R.; Doonan, I.J.; Anderson, O.F. (2025). Additional analyses for orange roughy (ORH 3B). New Zealand Fisheries Assessment Report 2025/18. 43 p.
https://fs.fish.govt.nz/Doc/25870/FAR-2025-18-Additional-analyses-for-orange-roughy-ORH3B.pdf.ashx

○ Pages 920-926 in Fisheries New Zealand (2025*).* As above.
https://www.mpi.govt.nz/dmsdocument/69831-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-NORTHERN-NORTH-ISLAND-ORH-1

● Pages 992-1006 in Fisheries New Zealand (2025*).* As above.
https://www.mpi.govt.nz/dmsdocument/69828-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-CHALLENGER-PLATEAU-ORH-7A

○ Dunn, M.R. 1 (2024). Assessment of the Challenger Plateau orange roughy stock for 2024. New Zealand Fisheries Assessment Report 2024/58. 42 p.
https://fs.fish.govt.nz/Doc/25788/FAR-2024-58-Assessment-of-the-Challenger-Plateau-orange-roughy-stock-for-2024.pdf.ashx

● Pages 927-949 in Fisheries New Zealand (2025*).* As above.
https://www.mpi.govt.nz/dmsdocument/69840-Fisheries-Assessment-Plenary-May-2025-Volume-2-ORANGE-ROUGHY-CAPE-RUNAWAY-TO-BANKS-PENINSULA-ORH-2A-2B-3A

○ Dunn, M.R.1 ; A'mar, T.1 ; Doonan, I.1 (2022).

4. Rules of participation

Relevant aspects of the “Membership and Protocols for all Science Working Groups in 2025” will be followed for this review. Participants must commit to:

- participating appropriately in discussions;
- maintaining confidentiality of presentations, discussions and deliberations;
- adopting a constructive approach;
- avoiding repetition of earlier deliberations;
- facilitating an atmosphere of honesty, openness and trust;
- respecting the role of the Chair and the Panel; and
- listening to the views of others, and treating them with respect

5. Timetable

The workshop is scheduled for 8-12 June 2026 and will be held at Charles Ferguson House, MPI, Wellington, New Zealand. Ian Tuck, FNZ will chair the open sessions.

Workshop format

1. An initial half day virtual meeting with an introduction on the history of orange roughy fisheries in New Zealand (just the facts, not opinions) and Matt Dunn presenting the ESCR assessment as one example of a previous assessment including model changes (what changed from previous assessments that showed stock size increasing), then questions to be addressed in the main workshop.
2. An in-person 2-day meeting with the international reviewers, with further presentations and verbal contributions, including from other participants.
3. Two-day period during which the international reviewers develop their conclusions and recommendations independently of the wider group.
4. Half day in-person meeting for the international reviewers to report back on draft conclusions to the wider group.
5. Final wrap up of independent review report, taking on board feedback on conclusions if reviewers consider appropriate.

APPENDIX 3: List of Attendees

Review Panel
Paul Burch (CSIRO)
Martin Dorn (University of Washington)
Steve Martell (Sea State Inc)

Organisers
Pamela Mace (Fisheries New Zealand)
Gretchen Skea (Fisheries New Zealand)
Ian Tuck (Fisheries New Zealand)

Other Fisheries New Zealand Staff
James Andrew (Fisheries New Zealand)
Alexander Arkhipkin (Fisheries New Zealand)
Andy Biggerstaff (Fisheries New Zealand)
Ian Blixt (Fisheries New Zealand)
Rich Ford (Fisheries New Zealand
Bruce Hartill (Fisheries New Zealand)
Alexandra Northmore (Fisheries New Zealand)
Marine Pomarède (Fisheries New Zealand)
Rob Tinkler (Fisheries New Zealand)

Fisheries New Zealand Research Providers
Matt Dunn (ESNZ)
Sophie Mormede (Independent Orange Roughy Reviewer)
Phil Neubauer (Dragonfly Data Science)
Richard O'Driscoll (ESNZ)
Tim Ryan (CSIRO)

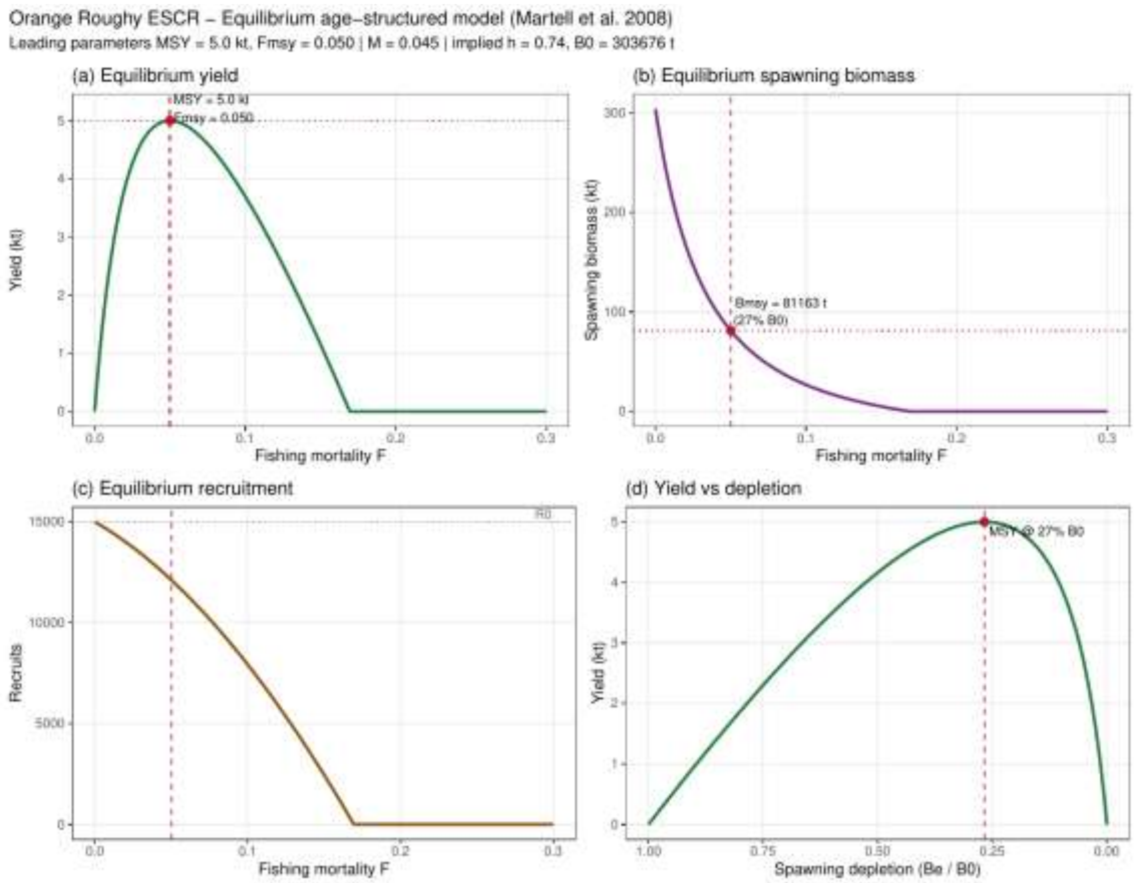
Other Participants
Aaron Irving (Deepwater Industry Council)
Geoff Tingley (Deepwater Industry Council)
Paul Starr (Deepwater Industry Council)
Tanayaz Arnaudora (Deepwater Industry Council)
Andy Smith (Independent Consultant)
Dion Iorns (Talleys)
Barry Weeber (ECO)
Jack Fenaughty (Sanfords)
Inge Wisselink (Deepwater Industry Council)

APPENDIX 4: *MSY* Re-parameterisation of the Equilibrium Age-Structured Model

This appendix presents an example of how the standard age-structured production model can be re-parameterised so that *MSY* and F_{MSY} are the leading quantities of interest, with B_0 and steepness (h) derived analytically from the age-schedule information on growth, maturity, and natural mortality. In the conventional parameterisation, B_0 and steepness are specified directly; however, these parameters are often confounded when fitting to one-way-trip data such as those available for orange roughy (Martell et al. 2008). By re-parameterising with *MSY* and F_{MSY} as the primary inputs, the model is anchored to quantities that can be informed by cross-stock comparisons and meta-analytic priors, while $B_0$and steepness emerge as derived outputs once the age-schedule biology is specified. The derivation uses the Goodyear compensation ratio (κ) computed analytically from the incidence functions and their derivatives at $F = F_{MSY}$, from which steepness is obtained as $h = \kappa/(4 + \kappa)$ and $B_0 = R_0 \times \phi^E$. This approach is particularly relevant to orange roughy given the one-way-trip nature of most stock trajectories and the difficulty of estimating B_0 and h independently from the available data.

Figure A4.1 — Equilibrium relationships

Figure A4.1. Equilibrium age-structured model for the ESCR orange roughy stock, parameterised with MSY = 5,000 t and F_{MSY} = 0.05 yr^{-1} as leading quantities (Martell et al. 2008). Panels show (a) equilibrium yield vs. fishing mortality, (b) equilibrium spawning biomass vs. fishing mortality, (c) equilibrium recruitment vs. fishing mortality, and (d) yield vs. spawning depletion (B^E/B_0). The red dashed line marks F_{MSY}; the implied steepness h and unfished spawning biomass B_0 are derived outputs from the biology schedule.



R code — 06_martell2008_eqm.R

The following R code reproduces the equilibrium analysis and Figure A4.1. Biology parameters (*M*, growth, maturity, selectivity) are taken from the ESCR assessment (Cordue 2014b parameter set). Dependencies: ggplot2, dplyr, tidyr, patchwork.

```
# ==============================================================================
# 06_martell2008_eqm.R
# Equilibrium age-structured model for Orange Roughy (ESCR)
# Leading parameters: MSY and Fmsy   (Martell, Pine & Walters 2008, CJFAS 65)
#
# Method ("incidence functions"):
#   1. Build unfished and fished survivorship-at-age (la, lz)
#   2. Compute per-recruit incidence functions:
#        phi.E  = spawning biomass per recruit, unfished
#        phi.e(F) = spawning biomass per recruit, fished
#        phi.q(F) = yield (weight) per recruit, fished  (Baranov)
#   3. At F = Fmsy, the condition dYe/dF = 0 gives the Goodyear
#      compensation ratio (kappa) ANALYTICALLY from the incidence
#      functions and their derivatives. Steepness h = kappa/(4+kappa).
#   4. MSY = Re(Fmsy) * phi.q(Fmsy)  ->  solve for R0, hence B0.
#   5. Trace equilibrium yield / SSB / recruitment over F.
#
# Biology (user-specified):
#   M     = 0.045  (fixed)
#   mat:  a50 = 25, a95 = 35     (Francis & Horn)
#   sel:  offset 5 yr younger -> a50 = 20, a95 = 30
# ==============================================================================

suppressMessages({library(ggplot2); library(dplyr); library(tidyr); library(patchwork)})
theme_set(theme_bw(base_size = 11) + theme(panel.grid.minor = element_blank()))

# ---- LEADING PARAMETERS (the two that drive everything) ---------------------
MSY  <- 5000      # maximum sustainable yield (tonnes)
Fmsy <- 0.05      # fishing mortality at MSY (yr^-1)

# ---- Fixed biology -----------------------------------------------------------
M      <- 0.045              # natural mortality (fixed)
A      <- 100L               # plus-group age
ages   <- 1:A

# von Bertalanffy growth + weight-length (ORH NZ; Cordue 2014 set)
Linf   <- 42.3; K <- 0.059; t0 <- 0.0
a_wl   <- 0.0258e-3; b_wl <- 3.17
La     <- Linf * (1 - exp(-K * (ages - t0)))
wa     <- a_wl * La^b_wl                       # weight-at-age (kg)

# Maturity (Francis & Horn): a50 = 25, a95 = 35
mat50  <- 25; mat95 <- 35
mat_a  <- 1 / (1 + exp(-log(19) / (mat95 - mat50) * (ages - mat50)))

# Selectivity: offset 5 yr younger -> a50 = 20, a95 = 30
sel50  <- mat50 - 5; sel95 <- mat95 - 5
va     <- 1 / (1 + exp(-log(19) / (sel95 - sel50) * (ages - sel50)))

# Spawning biomass per recruit weighting (fecundity proxy = mature weight)
fa     <- mat_a * wa

# ==============================================================================
# INCIDENCE FUNCTIONS
# ==============================================================================

# Unfished survivorship-at-age (la), with plus group
surv_unfished <- function() {
  la <- numeric(A); la[1] <- 1
  for (a in 2:A) la[a] <- la[a - 1] * exp(-M)
  la[A] <- la[A - 1] * exp(-M) / (1 - exp(-M))   # plus-group
  la
```

```
}
la    <- surv_unfished()
phi.E <- sum(la * fa)                            # SSB per recruit, unfished
phi.B0_perR <- phi.E                             # (same: mature-weight weighting)

# Fished incidence functions at a given F
incidence <- function(Fmort) {
  Za <- M + Fmort * va
  Sa <- exp(-Za)
  lz <- numeric(A); lz[1] <- 1
  for (a in 2:A) lz[a] <- lz[a - 1] * Sa[a - 1]
  lz[A] <- lz[A - 1] * Sa[A - 1] / (1 - Sa[A])   # plus-group
  # spawning biomass per recruit (fished)
  phi.e <- sum(lz * fa)
  # yield per recruit (Baranov)
  qa    <- wa * (va * Fmort / Za) * (1 - exp(-Za))
  phi.q <- sum(lz * qa)
  list(phi.e = phi.e, phi.q = phi.q, lz = lz)
}

# Numerical derivative helper
dF <- 1e-5
ddF <- function(f, Fmort) (f(Fmort + dF) - f(Fmort - dF)) / (2 * dF)

# ============================================================================
# SOLVE FOR COMPENSATION RATIO (kappa), R0, B0 FROM MSY & Fmsy
# ============================================================================
# At Fmsy:  dYe/dF = 0,  Ye = MSY.
# Ye(F) = Re(F) * phi.q(F),  Re(F) = R0 (kappa - phi.E/phi.e) / (kappa - 1)
# Setting dYe/dF = 0 and solving for kappa gives:
#
#   kappa = (phi.E/phi.e) * [ 1 - (phi.e' * phi.q) / (phi.e * phi.q') ]
#
# evaluated at Fmsy.   (phi.e' < 0, phi.q' > 0 near Fmsy  =>  kappa > 1)

inc_msy <- incidence(Fmsy)
phi.e_msy <- inc_msy$phi.e
phi.q_msy <- inc_msy$phi.q
phi.e_prime <- ddF(function(x) incidence(x)$phi.e, Fmsy)
phi.q_prime <- ddF(function(x) incidence(x)$phi.q, Fmsy)

kappa <- (phi.E / phi.e_msy) * (1 - (phi.e_prime * phi.q_msy) /
                                    (phi.e_msy * phi.q_prime))
h     <- kappa / (4 + kappa)                     # Beverton-Holt steepness

# Equilibrium recruitment at Fmsy, then R0, B0
Re_msy <- MSY / phi.q_msy
R0     <- Re_msy * (kappa - 1) / (kappa - phi.E / phi.e_msy)
B0     <- R0 * phi.E                              # unfished spawning biomass

# Beverton-Holt parameters (Goodyear)
so   <- kappa / phi.E                             # max recruits per unit SSB
beta <- (kappa - 1) / (R0 * phi.E)

# Equilibrium recruitment as a function of F
Re_of_F <- function(Fmort) {
  phi.e <- incidence(Fmort)$phi.e
  R <- (so * phi.e - 1) / (beta * phi.e)
  pmax(R, 0)
}

cat("================ DERIVED LEADING-PARAMETER MODEL ================\n")
cat(sprintf("Inputs:   MSY = %.0f t   Fmsy = %.4f   M = %.3f\n", MSY, Fmsy, M))
cat(sprintf("phi.E (SSB/recruit, unfished) = %.4f kg\n", phi.E))
cat(sprintf("Compensation ratio kappa      = %.3f\n", kappa))
cat(sprintf("Implied steepness h           = %.3f\n", h))
cat(sprintf("Unfished recruitment R0       = %.0f\n", R0))
```

```r
cat(sprintf("Unfished spawning biomass B0  = %.0f t\n", B0))

# ==============================================================================
# EQUILIBRIUM CURVES OVER F
# ==============================================================================
Fgrid <- seq(0, 0.30, length.out = 400)
eq <- data.frame(F = Fgrid) %>%
  rowwise() %>%
  mutate(
    phi.e = incidence(F)$phi.e,
    phi.q = incidence(F)$phi.q,
    Re    = Re_of_F(F),
    Be    = Re * phi.e,           # equilibrium spawning biomass
    Ye    = Re * phi.q,           # equilibrium yield
    Dep   = Be / B0
  ) %>%
  ungroup()

# Reference points
Bmsy   <- Re_msy * phi.e_msy
dep_msy <- Bmsy / B0

cat(sprintf("Bmsy                          = %.0f t  (%.1f%% B0)\n",
            Bmsy, 100 * dep_msy))

# Self-check: numerical max of Ye should equal MSY at Fmsy
i_peak <- which.max(eq$Ye)
cat(sprintf("Self-check: numeric peak yield = %.1f t at F = %.4f\n",
            eq$Ye[i_peak], eq$F[i_peak]))

saveRDS(list(eq = eq, MSY = MSY, Fmsy = Fmsy, B0 = B0, R0 = R0,
             Bmsy = Bmsy, dep_msy = dep_msy, kappa = kappa, h = h,
             la = la, wa = wa, mat_a = mat_a, va = va, ages = ages),
        "martell2008_eqm.rds")

# ==============================================================================
# FIGURE 1 — equilibrium relationships (Martell et al. 2008 style)
# ==============================================================================
acc <- "#b2182b"   # reference-point accent

# Panel A: equilibrium yield vs F
pA <- ggplot(eq, aes(F, Ye / 1000)) +
  geom_line(linewidth = 1, colour = "#1b7837") +
  geom_vline(xintercept = Fmsy, linetype = "dashed", colour = acc) +
  geom_hline(yintercept = MSY / 1000, linetype = "dotted", colour = acc) +
  annotate("point", x = Fmsy, y = MSY / 1000, colour = acc, size = 2.5) +
  annotate("text", x = Fmsy, y = MSY / 1000 * 1.04,
           label = sprintf("MSY = %.1f kt\nFmsy = %.3f", MSY/1000, Fmsy),
           hjust = -0.05, size = 3) +
  labs(title = "(a) Equilibrium yield", x = "Fishing mortality F",
       y = "Yield (kt)")

# Panel B: equilibrium spawning biomass vs F
pB <- ggplot(eq, aes(F, Be / 1000)) +
  geom_line(linewidth = 1, colour = "#762a83") +
  geom_vline(xintercept = Fmsy, linetype = "dashed", colour = acc) +
  geom_hline(yintercept = Bmsy / 1000, linetype = "dotted", colour = acc) +
  annotate("point", x = Fmsy, y = Bmsy / 1000, colour = acc, size = 2.5) +
  annotate("text", x = Fmsy, y = Bmsy / 1000,
           label = sprintf("Bmsy = %.0f t\n(%.0f%% B0)", Bmsy, 100*dep_msy),
           hjust = -0.05, vjust = -0.2, size = 3) +
  labs(title = "(b) Equilibrium spawning biomass", x = "Fishing mortality F",
       y = "Spawning biomass (kt)")

# Panel C: equilibrium recruitment vs F
pC <- ggplot(eq, aes(F, Re)) +
  geom_line(linewidth = 1, colour = "#8c510a") +
```

```
  geom_vline(xintercept = Fmsy, linetype = "dashed", colour = acc) +
  geom_hline(yintercept = R0, linetype = "dotted", colour = "grey50") +
  annotate("text", x = max(Fgrid), y = R0, label = "R0",
           hjust = 1.1, vjust = -0.3, size = 3, colour = "grey40") +
  labs(title = "(c) Equilibrium recruitment", x = "Fishing mortality F",
       y = "Recruits")

# Panel D: yield vs spawning depletion (Be/B0)
pD <- ggplot(eq, aes(Dep, Ye / 1000)) +
  geom_line(linewidth = 1, colour = "#1b7837") +
  geom_vline(xintercept = dep_msy, linetype = "dashed", colour = acc) +
  annotate("point", x = dep_msy, y = MSY / 1000, colour = acc, size = 2.5) +
  annotate("text", x = dep_msy, y = MSY / 1000,
           label = sprintf("MSY @ %.0f%% B0", 100 * dep_msy),
           hjust = -0.05, vjust = 1.5, size = 3) +
  scale_x_reverse() +
  labs(title = "(d) Yield vs depletion",
       x = "Spawning depletion (Be / B0)", y = "Yield (kt)")

fig1 <- (pA | pB) / (pC | pD) +
  plot_annotation(
    title = "Orange Roughy ESCR — Equilibrium age-structured model (Martell et al. 2008)",
    subtitle = sprintf("Leading parameters MSY = %.1f kt, Fmsy = %.3f | M = %.3f | implied h = 
%.2f, B0 = %.0f t",
                       MSY/1000, Fmsy, M, h, B0))

ggsave("martell2008_figure1.pdf", fig1, width = 11, height = 8.5)
cat("\nSaved: martell2008_figure1.pdf\n")
```