Improved methods for length and age composition for New Zealand fish stocks

Alistair Dunn1, Darcy Webber2, Sophie Mormede3

1. Ocean Environmental Ltd.

2. Quantifish Ltd.

3. SoFish Consulting Ltd.

4 April 2022

# Summary

*Proposal to Fisheries New Zealand for funding to develop improved methods for estimating scaled length and age compositions for New Zealand fish stocks from length composition and paired age-length data.*

* Age composition data from commercial and recreational fisheries are an important input for informing assessments of fish stocks.
* Improved methods of estimation could help reduce costs, reduce uncertainty, and allow for the use of previously unused length and age data in stock assessments and other analyses.
* Newly developed methods for modelling length composition and paired age-length data will:
* provide improved estimates of length and age composition data for New Zealand fish stocks and reducing uncertainty, accounting for strata with low levels of length and/or otolith sampling, spatial-temporal variability in age and length, and within-season growth effects.
* Improved sampling efficiency, allowing for a lower cost for collecting and reading otoliths, the use of data sampled throughout the year, as well as giving efficient age composition estimates for stocks and years not previously available.
* This project will develop an integrated method to estimate scaled length and age frequency distributions to account for various covariates (month, year, spatial changes) and compare it to the standard method (i.e., Catch-at-age and CALA) using case studies with selected New Zealand stocks.

# Objectives

## Overall objective

To develop improved methods for estimating scaled length and age composition for New Zealand fish stocks from length composition and paired age-length data.

## Specific objectives

1. To evaluate, via model development and simulation, the relative accuracy and effectiveness of different methods (i.e., Bayesian categorical regression models versus standard scaled length frequency combined with age-length-keys) to estimate length and age compositions from randomly collected length frequencies and age-length observations.
2. To apply the methods to selected case studies to compare the resulting scaled length and age composition estimates and related levels of uncertainty.
3. To present the analyses, results, conclusions, and guidelines to the Fisheries New Zealand (FNZ) Statistics, Assessments, and Methods Working Group (SAMWG), and submit a draft FAR to FNZ.

# Background

### Overview of age composition methods used in New Zealand fisheries

Length and age composition data from commercial and recreational fisheries are important for informing integrated stock assessment models used in the assessment of fish stocks, as well as inform (i.e., via determining recruitment) the evaluation of semi-quantitative assessments.

In New Zealand, length and age compositions for fish stocks have typically been estimated by either:

1. randomly sampling the population or the catch and using these samples to infer the length and/or age composition (i.e., for length composition and direct ageing); or
2. randomly sampling the population or catch for length, systematically sampling the population or catch for a smaller set of otoliths for aging, developing an empirical age-length-key (ALK) from the sample of aged otoliths, and applying this ALK to the samples where only length is measured to convert into an age composition.

The purpose of an ALK is to empirically derive an age composition from length composition data and paired age-length data. Estimates of age composition for most stocks have used ALK methods, with some (e.g., snapper, hoki) using direct ageing. With these methods, bootstrap resampling is typically used to provide uncertainty estimates that are used to determine initial effective sample sizes.

Collection of otoliths and reading of their age can be expensive (depending on the difficulty of reading for a species and the cost of obtaining otolith samples), with the cost imposing a considerable constraint on the number of otoliths able to be collected and aged. Using empirical methods, the sampled otoliths should be representative of the length-at-age of fish in question, be spatially (i.e., to account for variability in length-at-age across locations) and temporally (i.e., to account for seasonal growth) representative, and have an adequate number of samples to fully describe the age-length relationship.

The number of otoliths required to be aged using ALK methods is dependent on the number of length bins / ages, the variability within these, and the required accuracy of the resulting age composition. Where inadequate or non-representative otolith sampling has taken place (i.e., there are many missing or few data points for an age, length bin, or strata), it is often not possible to reliably estimate the underlying age composition (Ailloud & Hoenig 2019). This results in age composition data for stocks in some years and/or fisheries being inadequate, and unable to be used in stock assessment (e.g., the 2017 sub-Antarctic non-spawning hoki fishery age-composition).

Due to the cost of reading otoliths and the difficulty of ensuring adequate age representation during otolith collection, it is ideal to have robust methods that make the best use of the available data to estimate the age composition. Categorical models of length at age composition modelling are likely to be a better method that makes greater use of available data and allow for more cost-effective sampling.

### Age-length-key (ALK) methods

There are a number of methods for empirically estimating age composition based on ALKs, for example, the standard/forward method (Fridriksson 1934), the inverse method (Clark 1981), and the forward-inverse method (Hoenig et al. 2002).

The standard (aka forward) method uses the distribution of age within each length bin to calculate age composition from length (length) frequencies. This approach has been found to be a statistically efficient estimation method for estimating age composition (Ailloud & Hoenig 2019). However, it relies on an ALK calculated from representative age-length data generated from adequate numbers of samples collected for each fishery, area, and/or season for a species. Moreover, this method cannot be used to combine ALKs across areas, within a season where there is a significant growth effect, or across seasons — as this can result in underestimated variance and introduce bias in the resulting age composition (Aanes & Vølstad 2015).

In general, almost all fishery age-compositions for New Zealand fisheries use the standard method (for example, the method in Bull & Dunn 2002) and typically assume negligible spatial or temporal bias.

### Categorical models to estimate length and age composition

A better approach for deriving age compositions from length composition data and paired age-length data is to estimate the relationship between age and length using a categorical model. This approach was initially developed by Rindoff & Lewy (2001) using continuation ratio-logits. This was updated by Berg & Kristensen (2012) to estimate age-specific abundance indices; by Berg et al. (2014) for the case of research surveys; and Correa et al. (2020) for Bering Sea Pacific cod. Berg & Kristensen (2012) used generalised additive models (GAMs) with continuation ratio-logits to model the probability of age given length and spatial covariates for surveys of cod, haddock, and herring in the North Sea. They found that their estimates from modelled spatial varying age-length relationship had higher internal and external consistency and performed (statistically) better than standard ALK approaches (Berg & Kristensen 2012; Berg et al. 2014). Further, Berg at al. (2014) used this model to estimate the between-age correlations and found a general pattern of increasing positive correlations with age using the approach, providing better estimates of effective sample size. This relationship was later used to Berg & Nielson (2016) to model the correlation structure of age composition data in state-space stock assessment models as a way of addressing the concerns of the use of the multinomial distribution raised by Francis (2014).

Thorson & Haltuch (2019) implemented a categorical spatial-temporal model to calculate biomass and age composition in VAST (Thorson 2019) for systematic surveys, using the approach of Berg et al. (2014), and showed that, in a simulation experiment, it resulted in a significant increase in the effective sample size over standard [survey] design-based methods. They concluded that categorical spatial-temporal models improve estimation of effective sample sizes, account for co-variates, and have improved statistical properties over standard survey design-based methods.

Modelling approaches that estimate the relationship between length and age to estimate age composition have not been considered in detail in the scientific literature. While Rindoff & Lewy (2001) and Berg & Kristensen (2012) considered the application to fishery catch-at-age, they focused on age-specific abundance from research surveys; Berg et al. (2014) and Thorson & Haltuch (2019) looked only at age-specific abundance from research surveys.

Since the studies of Rindoff & Lewy (2001), Berg & Kristensen (2012), and Berg et al. (2014), modelling software to implement spatial-temporal GAMs has improved considerably —for example see mgcv (Wood et al. 2016; Wood 2017), Stan (Carpenter et al. 2017), brms (Bürkner 2017, 2018), and VAST (Thorson 2019; Thorson & Haltuch 2019).

Modern categorical modelling approaches are likely to provide an improved alternative to bootstrap approaches for calculating age composition using ALKs by estimating age composition across fisheries, areas, and/or season for a species — avoiding the strict requirements in standard ALK methods for age-length samples from every area, time period; and more robustly accounting for covariates and in-season growth for a given species.

Development of such models is likely to provide length and age composition data for New Zealand species that have not previously been available due to non-representative sampling or poor sampling coverage, as well as more robustly accounting for spatial-temporal variability and in-season growth effects. Further, application of spatial-temporal models may allow improved sampling efficiency with a lower cost due to a reduced number of otoliths required, as well as giving age composition estimates for stocks and years not previously available.

# Methods

This project will implement and evaluate the relative accuracy and effectiveness of categorical models compared with standard ALK methods to estimate age compositions from randomly collected length frequencies and age-length observations. The approach will involve the:

1. Development of a categorical model (e.g., multinomial) that can predict the true length composition and associated uncertainty from samples of length
2. Scaling of the length composition derived in step 1 by the catch
3. Development of an ordinal model that can predict age given length using paired age-length data and covariates
4. Conversion of the scaled length composition derived in step 2 to an age composition using the model described in step 3

Evaluation of methods will be via a simulation study based on characteristics for a range of typical fish stocks within New Zealand. The methods will then be applied to selected case studies (i.e., hake and ling) to compare point and variance estimates produced by each method on the resulting scaled length or age compositions, and the effect of using length and age data collected outside of the time range selected for the standard method.

Guidance on the best method for estimating age composition from length and paired age-length observations using categorical models versus standard ALK methods have not been fully considered in the scientific literature or evaluated for cases that represent typical New Zealand fish stocks. The results from this analysis will be used to develop conclusions and general guidelines for the use of length and age-length observations in the estimation of scaled length and age composition for New Zealand fish stocks.

Results, conclusions, and guidelines from the study will be reported to the Fisheries New Zealand Statistics, Assessments, and Methods Working Group (SAMWG), and in a draft FAR to FNZ.

# Reporting requirements

## Research reporting

### Objectives 1-3

1. To submit to the MPI contracts manager and Fisheries New Zealand project scientist a draft Fishery Assessment Report, as specified in Research Reporting Form 7, by 30 June 2022.
2. To present the report detailed in 1 above to a meeting of the Fisheries New Zealand Statistics, Assessments, and Methods Working Group (SAMWG) by 30 October 2022.
3. To submit to the MPI contracts manager and Fisheries New Zealand project scientist a final Fishery Assessment Report, as specified in Research Reporting Form 7, by 30 October 2022.

## Data Reporting

To submit any data generated, collected or modified during the course of this project to the Fisheries New Zealand Research Data Manager by 30 December 2022.

# Price

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Objective** | **Milestone** | **Reporting**  **Requirement** | **Milestone Description** | **Due Date** | **Fixed Price (excl GST) (NZD)** |
| 1 | 1 | 1 | Update standard and implement categorical modelling approaches to estimate scaled length and age frequencies | 30 June 2022 | $30,520 |
| 1 | 2 | 1 | Evaluate, via simulation, the relative accuracy and effectiveness of different methods | 30 June 2022 | $13,940 |
| 2 | 3 | 1 | Apply the methods to case study examples (e.g., Chatham Rise & sub-Antarctic ling, and west coast South Island and sub-Antarctic hake, and a selected rock lobster stock) and compare the outcomes on the resulting length (all species) and age-frequencies (hake and ling) | 30 June 2022 | $5,140 |
| 3 | 4 | 1 | Submit a draft FAR (as per Reporting Form 6) to MPI | 30 June 2021 | $7,320 |
| 3 | 5 | 2 | Present methods and results to the Fisheries New Zealand Statistics, Assessments, and Methods Working Group (SAMWG) | 30 October 2022 | $3,660 |
| 3 | 6 | 3 | Submit a final FAR (as per Reporting Form 6) to MPI | 30 October 2022 | $2,180 |
| 3 | 7 | 4 | Submit data to the Fisheries New Zealand data management team | 30 December 2022 |  |
| **PROJECT TOTAL $NZD** | | | | | $62,760 |

# References

Aanes, S.; Vølstad, J.H. (2015). Efficient statistical estimators and sampling strategies for estimating the age composition of fish. Canadian Journal of Fisheries and Aquatic Sciences 72, 938–953. https://doi.org/10.1139/cjfas-2014-0408

Ailloud, L.E.; Hoenig, J.M. (2019). A general theory of age-length keys: combining the forward and inverse keys to estimate age composition from incomplete data. ICES Journal of Marine Science 76, 1515–1523. https://doi.org/10.1093/icesjms/fsz072

Berg, C.W.; Kristensen, K. (2012). Spatial age-length key modelling using continuation ratio logits. Fisheries Research 129–130, 119–126. https://doi.org/10.1016/j.fishres.2012.06.016

Berg, C.W.; Nielsen, A. (2016). Accounting for correlated observations in an age-based state-space stock assessment model. ICES J Mar Sci 73, 1788–1797. https://doi.org/10.1093/icesjms/fsw046

Berg, C.W.; Nielsen, A.; Kristensen, K. (2014). Evaluation of alternative age-based methods for estimating relative abundance from survey data in relation to assessment models. Fisheries Research 151, 91–99. https://doi.org/10.1016/j.fishres.2013.10.005

Bull, B.; Dunn, A. (2002). Catch-at-age: User manual v1.06.2002/09/12. NIWA Technical Report 114. NIWA, Wellington, New Zealand, 23 p.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. Journal of Statistical Software 80. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. The R Journal 10, 395–411.

Carpenter, B.; Gelman, A.; Hoffman, M.D.; Lee, D.; Goodrich, B.; Betancourt, M.; Brubaker, M.; Guo, J.; Li, P.; Riddell, A. (2017). STAN: A Probabilistic Programming Language. Journal of Statistical Software 76. https://doi.org/10.18637/jss.v076.i01

Clark, W.G. (1981). Restricted Least-Squares Estimates of Age Composition from Length Composition. Canadian Journal of Fisheries and Aquatic Sciences 38, 297–307. https://doi.org/10.1139/f81-041

Correa, G.M.; Ciannelli, L.; Barnett, L.A.K.; Kotwicki, S.; Fuentes, C. (2020). Improved estimation of age composition by accounting for spatiotemporal variability in somatic growth. Can. J. Fish. Aquat. Sci. 1810–1821. https://doi.org/10.1139/cjfas-2020-0166

Francis, R.I.C.C. (2014). Replacing the multinomial in stock assessment models: A first step. Fisheries Research 151, 70–84. https://doi.org/10.1016/j.fishres.2013.12.015

Fridriksson, A. (1934). On the calculation of age distribution within a stock of cod by means of relatively few age-determinations as a key to measurements on a large scale. Rapports et Procès-verbaux des Réunions. Conseil Permanent International pour l’Exploration de la Mer 86, 1–14.

Hoenig, J.M.; Hanumara, R.C.; Heisey, D.M. (2002). Generalizing Double and Triple Sampling for Repeated Surveys and Partial Verification. Biometrical Journal 44, 603–618. https://doi.org/10.1002/1521-4036(200207)44:5<603::AID-BIMJ603>3.0.CO;2-4

Rindorf, A.; Lewy, P. (2001). Analyses of length and age distributions using continuation-ratio logits. Canadian Journal of Fisheries and Aquatic Sciences 58, 1141–1152. https://doi.org/10.1139/f01-062

Thorson, J.T. (2019). Guidance for decisions using the Vector Autoregressive Spatio-Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fisheries Research 210, 143–161. https://doi.org/10.1016/j.fishres.2018.10.013

Thorson, J.T.; Haltuch, M.A. (2019). Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. Can. J. Fish. Aquat. Sci. 76, 401–414. https://doi.org/10.1139/cjfas-2018-0015

Wood, S.N. (2017). Generalized Additive Models: An Introduction with R, 2nd ed. Chapman and Hall/CRC.

Wood, S.N.; Pya, N.; Säfken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). Journal of the American Statistical Association 111, 1548–1575.