



QUANTIFISH
Quantitative Fisheries Science

Southern Bluefin Tuna Operating Model Progress

Darcy Webber, Paige Eveson, Rich Hillary



5 August 2024

Working Paper prepared for the Operating Model and Management Procedure Technical Meeting
1 September 2024, Taipei, Taiwan.

Quantifish Ltd
Tauranga, New Zealand

Copyright and disclaimer

To the extent permitted by law, all rights are reserved and no part of this publication covered by copyright may be reproduced or copied in any form or by any means except with the written permission of Quantifish.

Important disclaimer

Quantifish advises that the information contained in this publication comprises general statements based on scientific research. The reader is advised and needs to be aware that such information may be incomplete or unable to be used in any specific situation. No reliance or actions must therefore be made on that information without seeking prior expert professional, scientific, and technical advice. To the extent permitted by law, Quantifish (including its employees and consultants) excludes all liability to any person for any consequences, including but not limited to all losses, damages, costs, expenses, and any other compensation, arising directly or indirectly from using this publication (in part or in whole) and any information or material contained in it.

Contents

Introduction.....	1
Methods	1
Modifications to tagging likelihood.....	1
Modifications to the POP likelihood.....	3
Cohort slicing of LFs	5
Direct removal of catches	11
GMRF selectivity.....	17
Discussion.....	17
References.....	17

Introduction

The previous operating model (OM) specifications, code, and software presented challenges for communicating the population dynamics and statistical assumptions underpinning the southern bluefin tuna (SBT) model; addressing uncertainty within the OM grid; and revising and implementing alternative hypotheses in the stocks assessments and future management procedure evaluations. Upgrading to modern software should improve the flexibility, utility, and understanding of the SBT operating and assessment models for all Commission for the Conservation of Southern Bluefin Tuna (CCSBT) participants. Improvements to model structural and statistical procedures will potentially result in better presentation and understanding of historical, current, and future SBT stock status, its associated uncertainty, and management procedure (MP) performance.

The objectives of this project were to: develop a new OM in Template Model Builder (TMB) software; code modifications to the OM (to be decided by the OMMP Working Group to improve estimation efficiency and allow future flexibility in adding/removing complexity and features as needed); and to complete validation tests comparing estimates from the new OM implementation with the current AD model builder (ADMB) OM.

This paper explores some modifications to the TMB OM, specifically: changes to the tagging likelihood; modifications to the POP likelihood; cohort slicing of length frequencies (LFs), direct removal of catches, and time varying selectivity specified using Gaussian Markov random fields (GMRF). This paper aims to serve as a starting point for discussions during the OMMP technical meeting in Taiwan.

Methods

Modifications to tagging likelihood

The tagging likelihood in the previous OM included a large number of “H*” parameters corresponding to the harvest rate for tagged fish in the same time period (year) that tagging occurred. These parameters were required to account for incomplete mixing of tagged fish within the wider population in the year of tagging. However, these parameters were not well estimated by the model, particularly when using Bayesian inference.

To resolve this issue, an alternative approach was implemented in the new OM whereby the recaptures in the year of tagging were removed from the number of releases in that year, after accounting for non-reporting (i.e., recaptures divided by reporting rate to get estimate of actual number of tags recaptured, not just those reported). The probability of recapture in the year of tagging was then set to zero, and the probability of a tagged fish surviving to the next year with at least one tag in place no longer needs to take into account the harvest rate in that year, but rather is a function of natural mortality and tag shedding parameters only.

Mathematically, the adjusted numbers of tag releases (N^{adj}) from cohort k , by tagger group t , at age i is:

$$N_{k,t,i}^{adj} = N_{k,t,i} - \frac{R_{k,t,i,i}}{\lambda_{k-i+1,i}}$$

where $N_{k,t,i}$ is the original number of releases, $R_{k,t,i,i}$ is the observed number of recaptures from this set of releases in the same year as release (i.e., at age i), and $\lambda_{k-i+1,i}$ is the reporting rate for a fish of age i in the year of release ($k - i + 1$).

The modified probabilities of a fish from cohort k tagged by tagger group t at age i surviving the year of tagging with one tag still attached (S') or two tags still attached (S'') is:

$$S'_{k,t,i} = 2\xi_t \exp(-M_i - \Omega_t)$$

$$S''_{k,t,i} = \xi_t^2 \exp(-M_i - 2\Omega_t)$$

where M_i is natural mortality for a fish of age i , ξ_t is the probability of a tag shedding immediately for tagger group t , and Ω_t is the continuous rate of tag shedding for tagger group t . Note that these probabilities no longer need to account for a fish being caught in the year of release.

This change has a minor impact on parameter estimation since, in essence, it means the recaptures in the year of tagging are being fitted exactly, whereas in the previous version the H* parameters (for which there was one for every tagging event, i.e., tagger group, tagging cohort and release age) allowed for an almost but not perfect fit. This is because the data are multinomial and therefore not independent, so fitting the recaptures in the year of tagging perfectly can lead to a slightly worse fit to the recaptures in subsequent years. However, a comparison between the standardised residuals using the original ‘‘H*’’ likelihood and the modified ‘‘no-H*’’ version confirms that the differences are small (Figure 1).

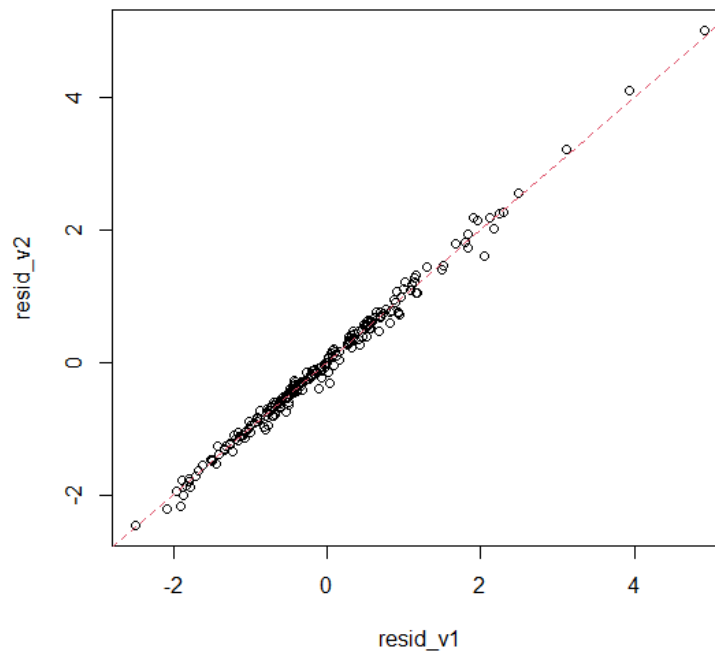


Figure 1: Comparison between standardised residuals for the tag recapture data between the previous ‘‘H*’’ version of the likelihood (v1) and the modified ‘‘no-H*’’ version of the likelihood (v2).

Two additional changes were made to simplify the code for the tagging likelihood, but these changes had no effect on parameter estimation. First, the previous code involved calculating a term called ‘‘tag_offset’’ that changed the absolute value of the likelihood but had no bearing on parameter estimation, so was removed. Second, a Dirichlet-multinomial (D-M) distribution was written for TMB since an ‘‘off-the-shelf’’ function did not already exist, and this was used rather than coding the distribution directly within the tagging likelihood.

Modifications to the POP likelihood

The current model for the POPs is purely age-based – this is not a problem for the juveniles in the comparison given the targeted sampling to obtain 3-year-olds, but for the adults in the comparison if we do not have a direct age estimate we infer the expected age from the distribution of length-at-age. We propose that, for cases where the adults in the comparison only have an observed length-at-capture, we account for the distribution of possible ages this adult *could* be given for its length. So, for comparisons of aged adults the POP probability remains as currently defined:

$$\mathbb{P}(\text{POP}|z_i, z_j) = \mathbb{I}(c_j < y_i < c_j + a) \sum_a \frac{2\varphi_{c_j, (y_i - c_j)}}{\text{TRO}_{c_j}}$$

The key adult covariates are $z_i = \{y, a\}$ the year and age at capture and for the juveniles $z_j = c_j$ it is the cohort of its birth. The variable TRO is the Total Reproductive Output – the actively reproducing adult population – and $\varphi_{y,a}$ is the relative reproductive output-at-age (*per capita* probability of being a successful parent). For the comparisons where the adult only has an observed length, we propose the following modification of the above age-based POP probability:

$$\mathbb{P}(\text{POP}|z_i, z_j) = \mathbb{I}(c_j < y_i < c_j + a) \sum_a \frac{2\varphi_{c_j, (y_i - c_j)}}{\text{TRO}_{c_j}} \pi(a|y_i, l_i)$$

The underlying dynamics of the probability are still age-based, but given an observed adult length l_i we integrate across all the possible ages this adult could be via the distribution of age-given-length, $\pi(a|y, l)$. This distribution has a time-dependence for two reasons: (1) the underlying distribution of length-at-age changes over time; and (2) the length distribution of sampled adults changes from year to year. This means the prior age distribution (the distribution of possible ages the adult could be before measuring its observed length) will change also. The distribution of age-given-length can be defined via Bayes' rule:

$$\pi(a|y, l) = \frac{\pi(l|y, a)\pi(a|y)}{\pi(l|y)}$$

We have already parameterised the distribution of length-at-age so we need to estimate the prior age distribution for each year of adult sampling. This can be done by making use of the observed length frequency of the genotyped adults, $O_{y,l}$. The predicted distribution of adult length in year y can be defined as follows:

$$\pi(l|y) = \sum_a \pi(l|y, a)\pi(a|y)$$

Assuming a multinomial distribution for the observed adult ages we can use maximum likelihood techniques to estimate the parameters of the prior age distribution-by-year (for this work we assumed a lognormal distribution for each year). All the code to do this part of the analysis is written in TMB and can be included in the wider suite of new R/TMB OM software. At this stage we have only used the adults genotyped via the SNP methodology, not those genotyped via microsatellites from many years ago, but ultimately, we will do this for all sampling years. Figure 2 shows the fits to the SNP-genotyped adult length distributions, and the associated prior age distributions are displayed in Figure 3.

An alternative method that was proposed at the OMMP14 meeting uses the abundance-at-age in the OM and the estimated selectivity of the Indonesian fleet. If we assume that the adult samples are taken proportional to their age availability in this fishery then the prior age distribution required to construct the distribution of age given length is proportional to $N_{y,a} \times S_{y,a}$ where the first term is the abundance-

at-age, and the second term is (time-varying) selectivity-at-age in the Indonesian fishery. The major argument in favour of this approach is that it captures more directly the influence of recruitment variation and juvenile and sub-adult fishing mortality on the distribution of ages in the spawning ground fishery. The major argument in favour of the approach using the length distribution in the samples themselves is that we get a more direct representation of what likely ages are appearing in the samples, not the Indonesian fishery (from which they are drawn but not randomly). Both methods have different advantageous features and within the new OM there will be a switch to choose between the two options when running the OM.

In the actual POP data, 94 of the POPs have an aged adult, with only 22 where length is the covariate, not age. We would not expect a major shift in the results based on this change, but it will likely have subtle effects that will be explored in the testing phase.

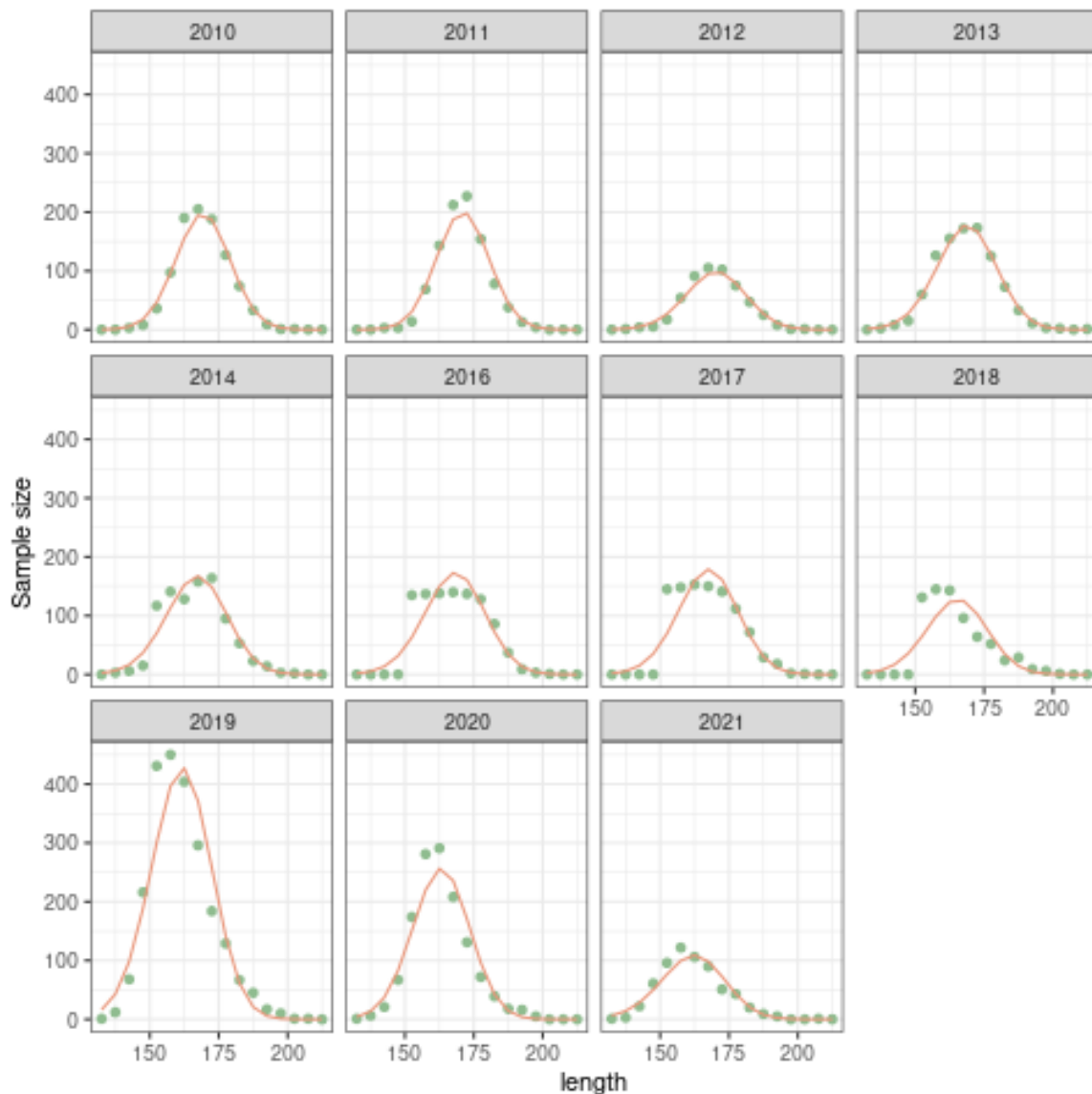


Figure 2: observed (green circles) and predicted (pink lines) length frequency of SNP genotyped adults.

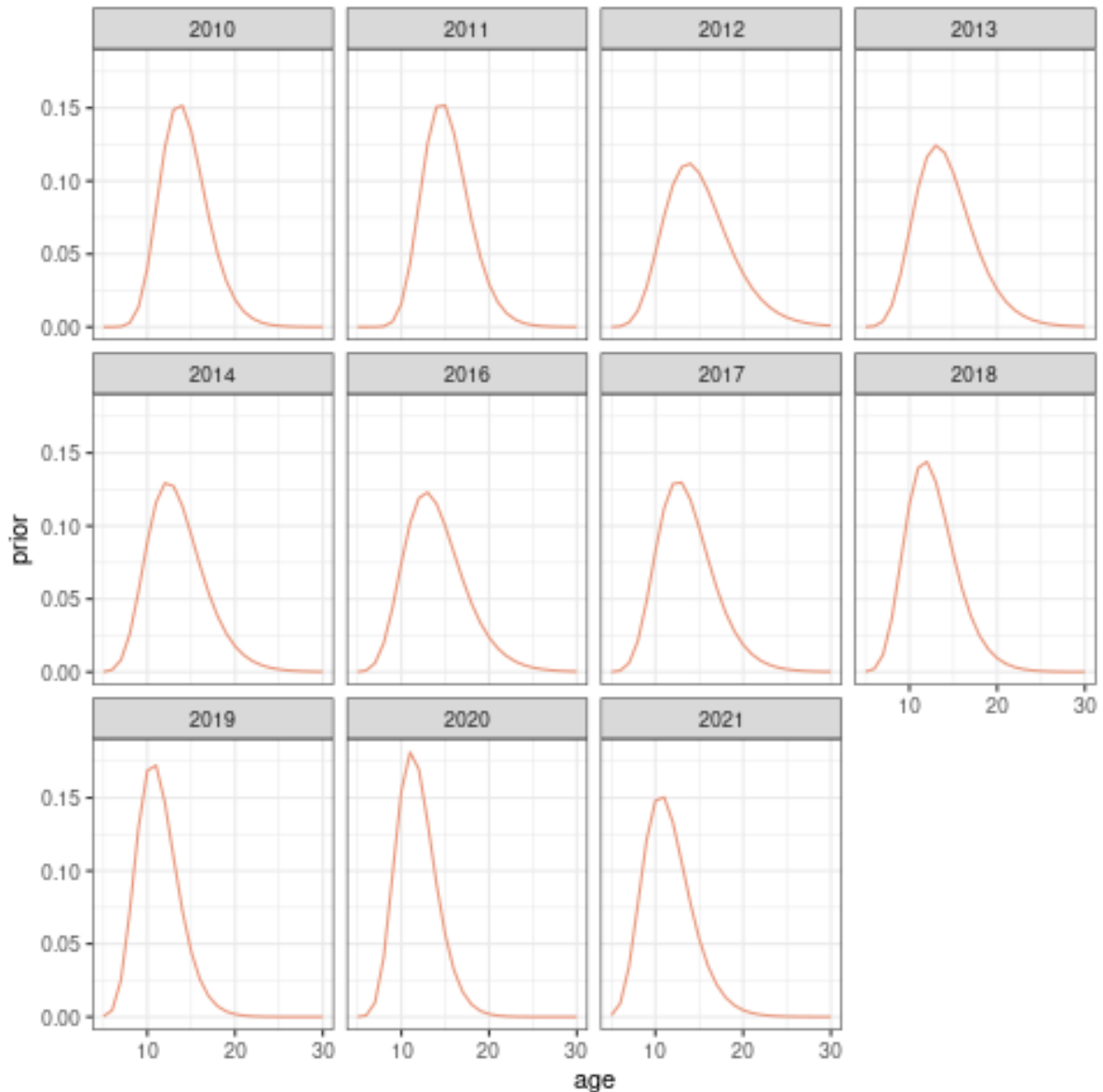


Figure 3: estimated prior age distribution-by-year for the SNP genotyped adults.

Cohort slicing of LFs

In the ADMB OM, time varying selectivity at age was estimated for the LL3 fishery, time invariant selectivity at age was estimated for the LL4 fishery (Figure 4), and the LFs for these two fisheries were fitted separately (Figure 5, Figure 6).

Code was written to cohort slice LFs into AFs for the LL3 and LL4 fisheries. This is done outside of the TMB model code (i.e., in R code using the function `get_sliced_afs` which is embedded in the function `get_data`). This code slices all four longline fisheries, but the user can choose to fit to the original LFs or the sliced AFs for each fishery.

In short, the process of cohort slicing for each LF involves taking the mean length at age for each year and season for each LF ($l_{y,s,a}$), finding the midpoints between each length at age (and appending zero and infinity at start and end), then cutting the LFs at these midpoints (Figure 7, Figure 8).

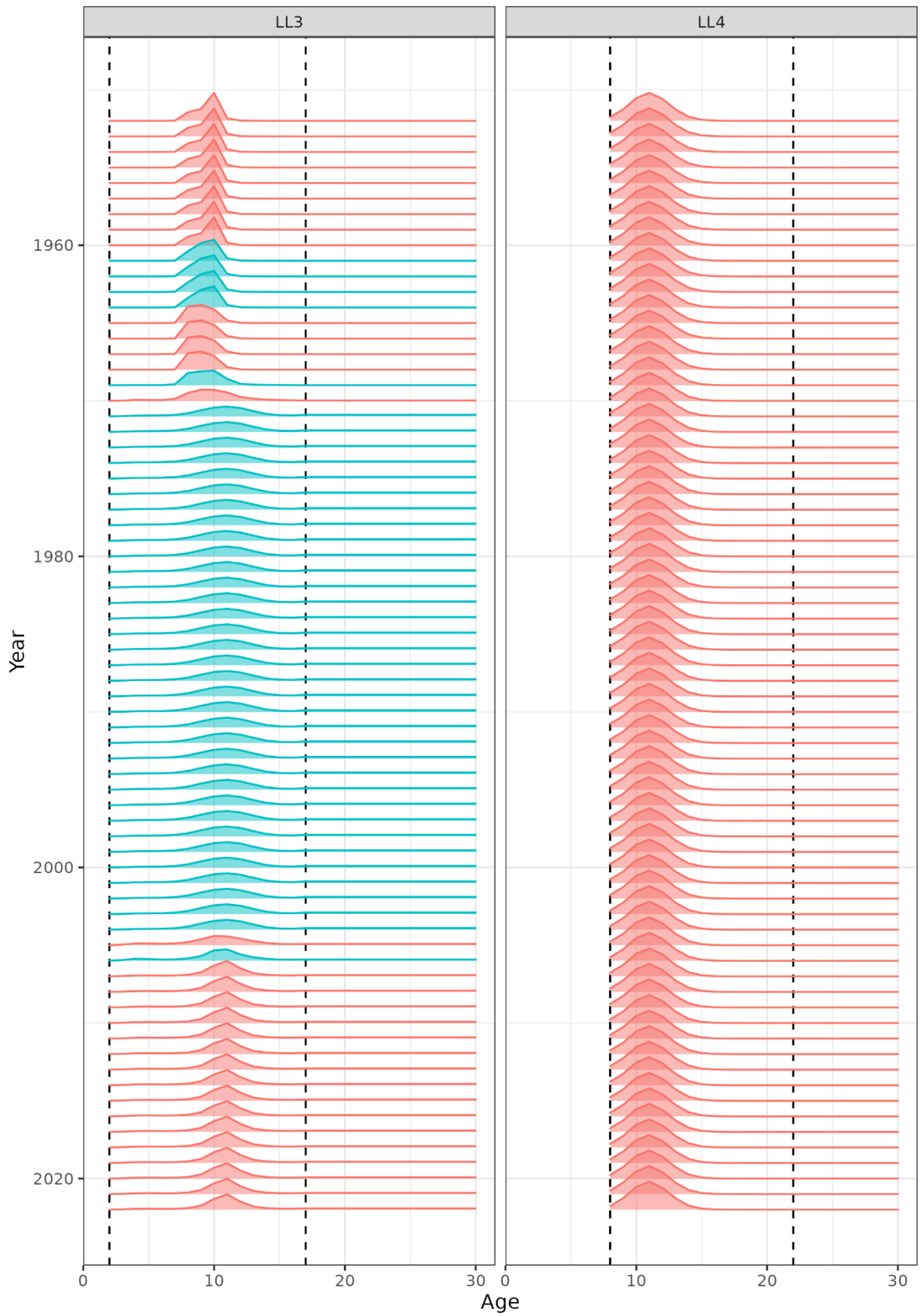


Figure 4: selectivity estimated by the previous OM for LL3 and LL4. The colours represent selectivity periods of that are assumed to be the same (i.e., there are nine periods of different selectivity for LL3 and LL4 selectivity is time invariant).

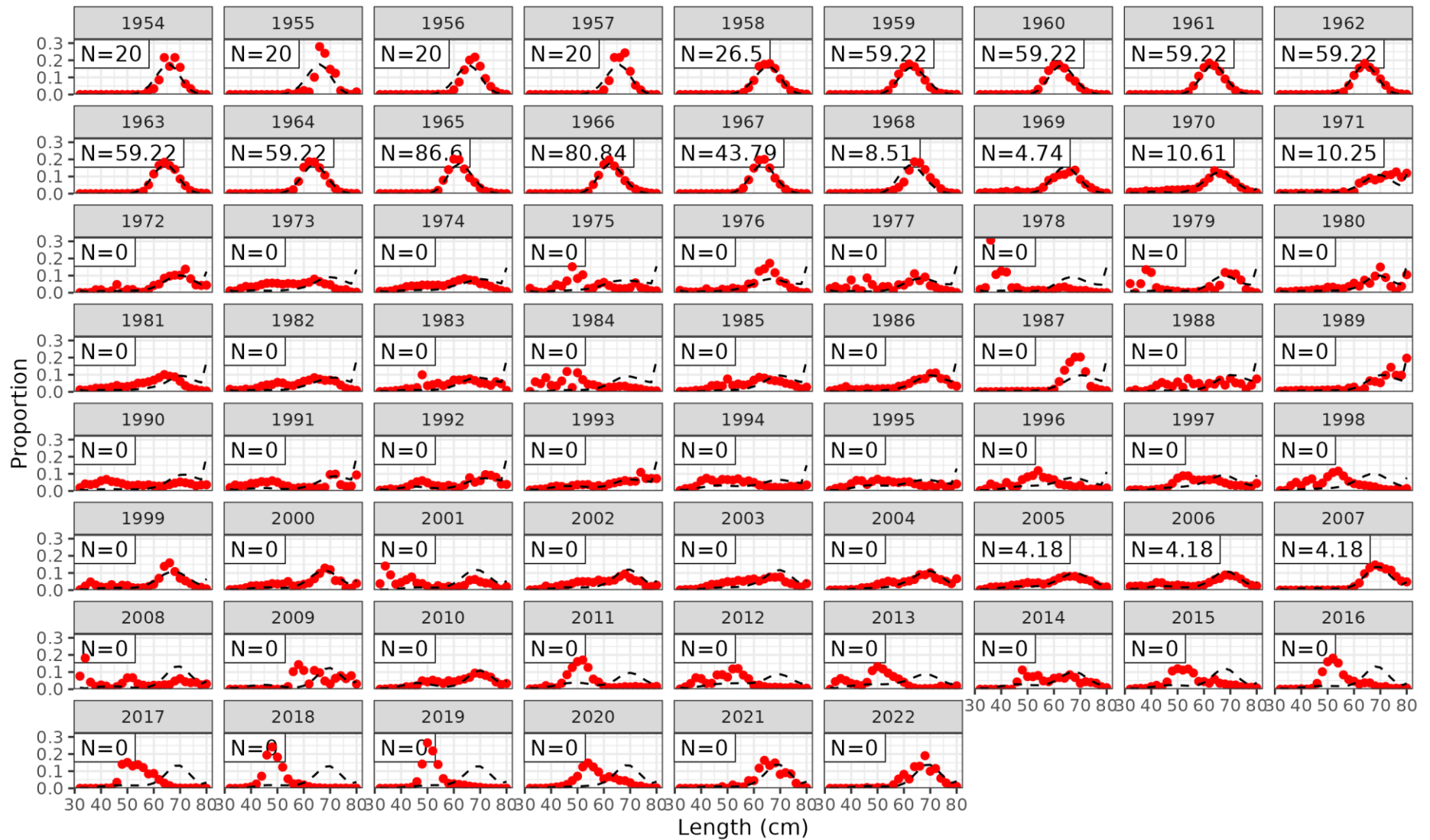


Figure 5: Observed LFs (red) and model fit (black) to the LL3 LFs. The effective sample size (N) is also shown for each year.

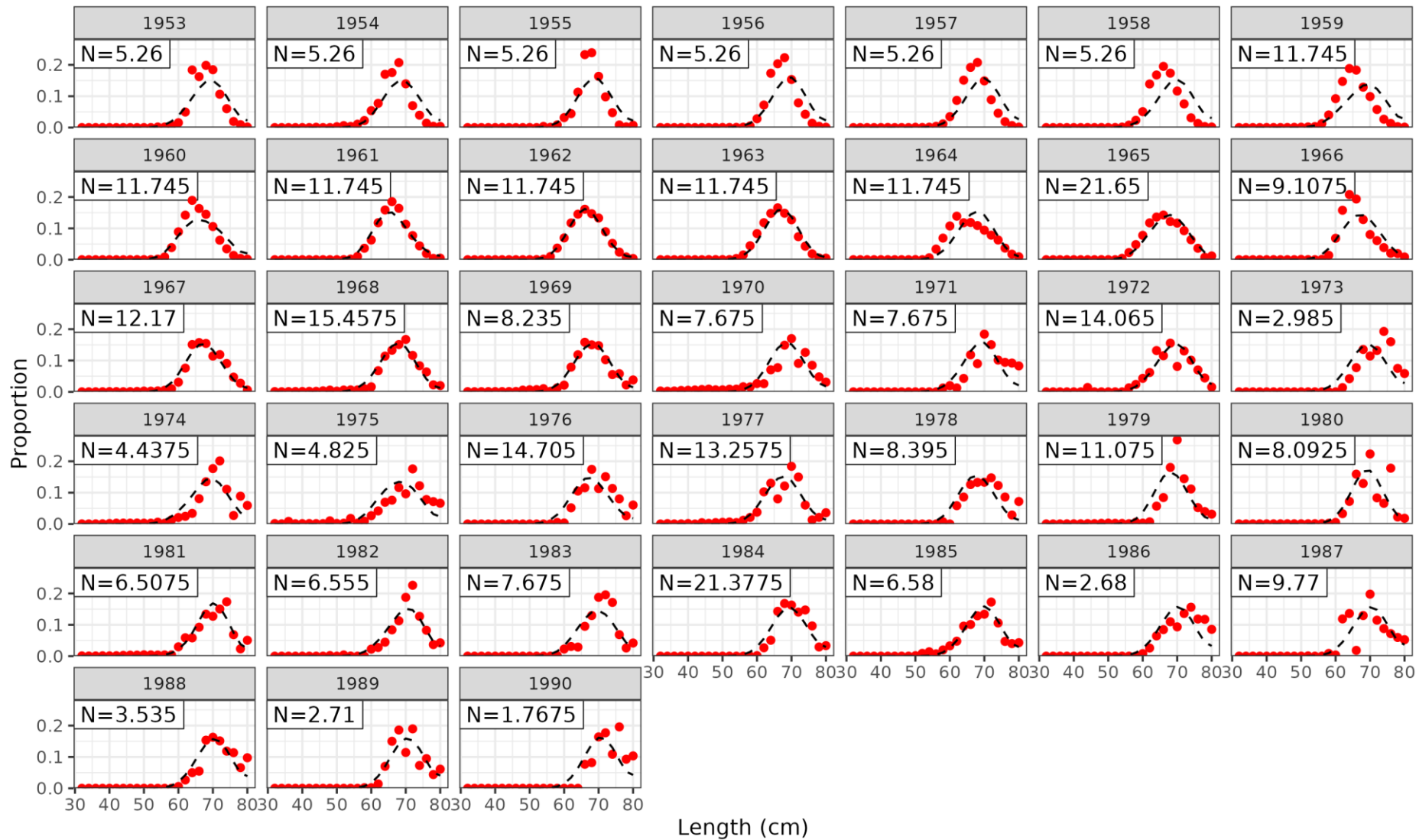


Figure 6: Observed LFs (red) and model fit (black) to the LL4 LFs. The effective sample size (N) is also shown for each year.

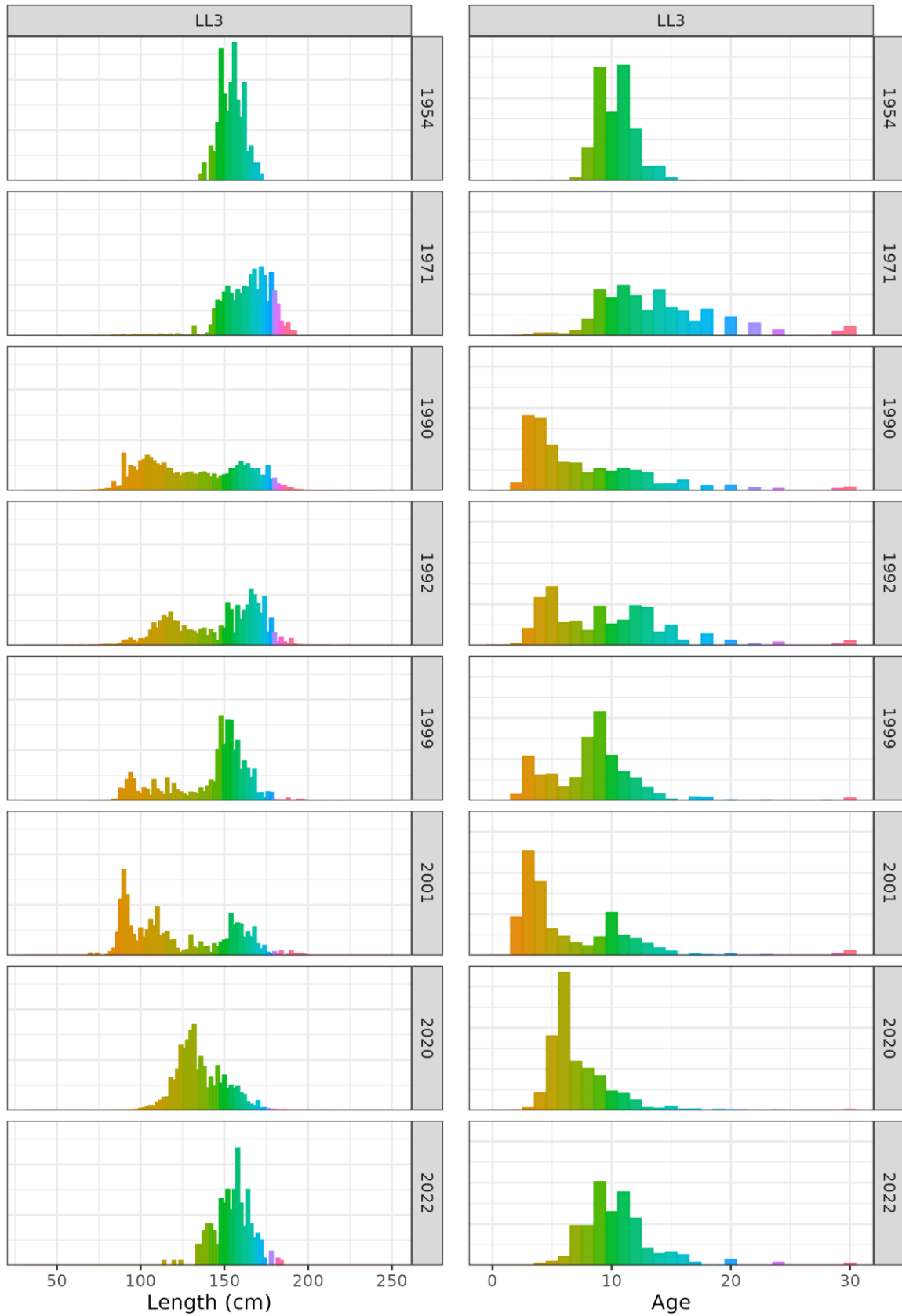


Figure 7: original LFs for a subset of years (left) and sliced LFs for those same years (right) for the LL3 fishery. Each colour represents an age.

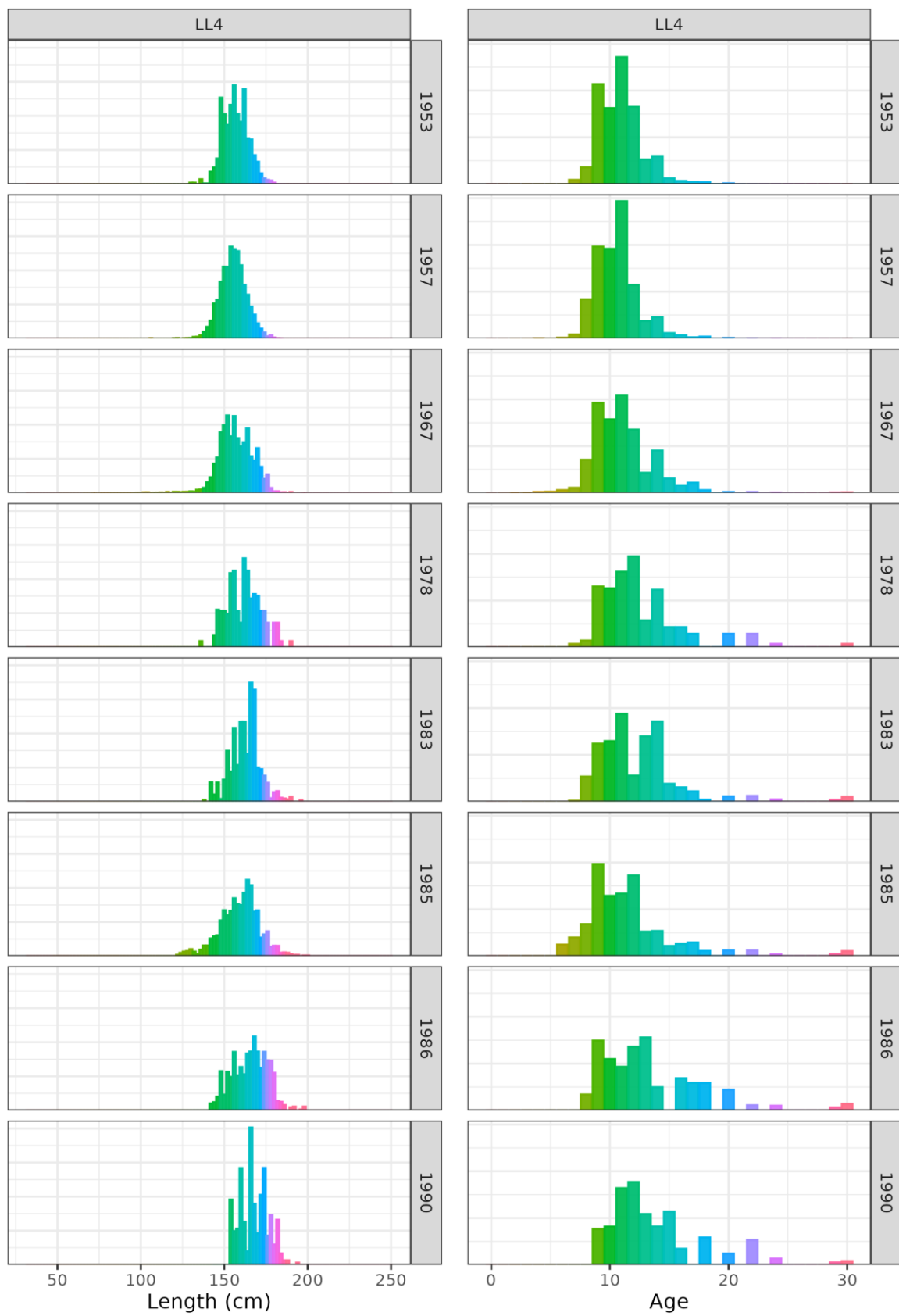


Figure 8: original LFs for a subset of years (left) and sliced LFs for those same years (right) for the LL4 fishery. Each colour represents an age.

Direct removal of catches

Typically, the catch for each year (y), season (s), and fishery (f), (see Table 1 for other variable definitions) is removed by estimating selectivity ogives and using the process outlined in Table 2. The code has been modified so that direct removals can (optionally) be specified for any fishery (all seasons and all years for the selected fishery, see Table 2). When specifying direct removals, the catch is removed from the model using the observed catch (tonnes) for each fishery ($C_{y,s,f}$) which is split proportionally by the age frequency (AF) or the cohort sliced LF ($A_{y,s,f,a}$) for each fishery.

An example of direct removal of the catch is provided below. In this example, the LL3 catch is all removed because there is an LF associated with the catch in every year. However, the LL4 catch is not all removed because in some years there is catch but no LF (Figure 9). Thus, as it is currently coded, when no AF or LF is available for a fishery in a year then the catch collapses to zero, even if there is catch in that year/season/fishery. This can be amended by padding out the missing years with interpolated AFs or LFs (for example, derive the average LF and use this for missing years).

Further to the example above, the predicted LFs for the LL3 and LL4 fisheries can be derived using $C_{y,s,f,a}^N$. Note that the model predicted LFs do not match the observed LFs exactly because the catches are modified by the age-length-key (to convert from catch at age to catch at length) which distorts the predictions (Figure 10, Figure 11). However, when direct removals are specified for a fishery with AFs the predicted AFs match the observed AFs exactly (e.g., Figure 12). Finally, when direct removals are specified for one or more fisheries, the implied selectivity for these fisheries can be derived from the harvest rates (e.g., Figure 13).

Table 1: Variables used throughout document.

Variable	Class	Description
a	Dimension	Age
y		Year
s		Season
f		Fishery
$N_{y,s,a}$	Derived quantity	Numbers at age in the population
$w_{y,f,a}$		Weight (tonnes) at age for each fishery
$S_{y,f,a}$		Selectivity at age for each fishery
M_a		Natural mortality at age
$H_{y,s,f,a}$		Harvest rate at age for each fishery
$C_{y,s,f,a}^N$		Catch (numbers) at age for each fishery
$C_{y,s,f}$	Covariate	Catch (tonnes) for each fishery
$A_{y,s,f,a}$		Proportion at age derived from an LF (i.e., cohort sliced LF) for each fishery

Table 2: Standard procedure for removal catch from an age structured stock assessment model versus direct removals.

Standard removals	Direct removals
$U_{y,s,f,a} = \frac{C_{y,s,f}}{\sum_a N_{y,s,a} S_{y,f,a} W_{y,f,a}}$	<p>Define the catch biomass (tonnes) as</p> $C_{y,s,f} = X_{y,s,f} \sum_a A_{y,s,f,a} W_{y,f,a}$ <p>where $X_{y,s,f}$ is the catch in numbers obtained by dividing the catch in biomass by the average weight:</p> $X_{y,s,f} = \frac{C_{y,s,f}}{\sum_a A_{y,s,f,a} W_{y,f,a}}$ <p>The catch at age in numbers can be calculated as</p> $C_{y,s,f,a}^N = X_{y,s,f} A_{y,s,f,a}$
$H_{y,s,f,a} = \frac{U_{y,s,f,a} S_{y,f,a}}{\sum_a N_{y,s,a} S_{y,f,a} W_{y,f,a}}$	$H_{y,s,f,a} = \frac{C_{y,s,f,a}^N}{N_{y,s,a}} = \frac{X_{y,s,f} A_{y,s,f,a}}{N_{y,s,a}} = \frac{N_{y,s,a}}{C_{y,s,f} \sum_a A_{y,s,f,a} W_{y,f,a}}$
<p>The catch from all fisheries is removed from the population using</p> $N_{y,s+1,a} = N_{y,s,a} \left(1 - \sum_f H_{y,s,f,a} \right) e^{-0.5M_a}$	
<p>The catch for deriving LFs and AFs is calculated as</p> $C_{y,s,f,a}^N = H_{y,s,f,a} N_{y,s,a}$ <p>And the catch biomass is</p> $C'_{y,s,f} = H_{y,s,f,a} N_{y,s,a} W_{y,f,a}$	

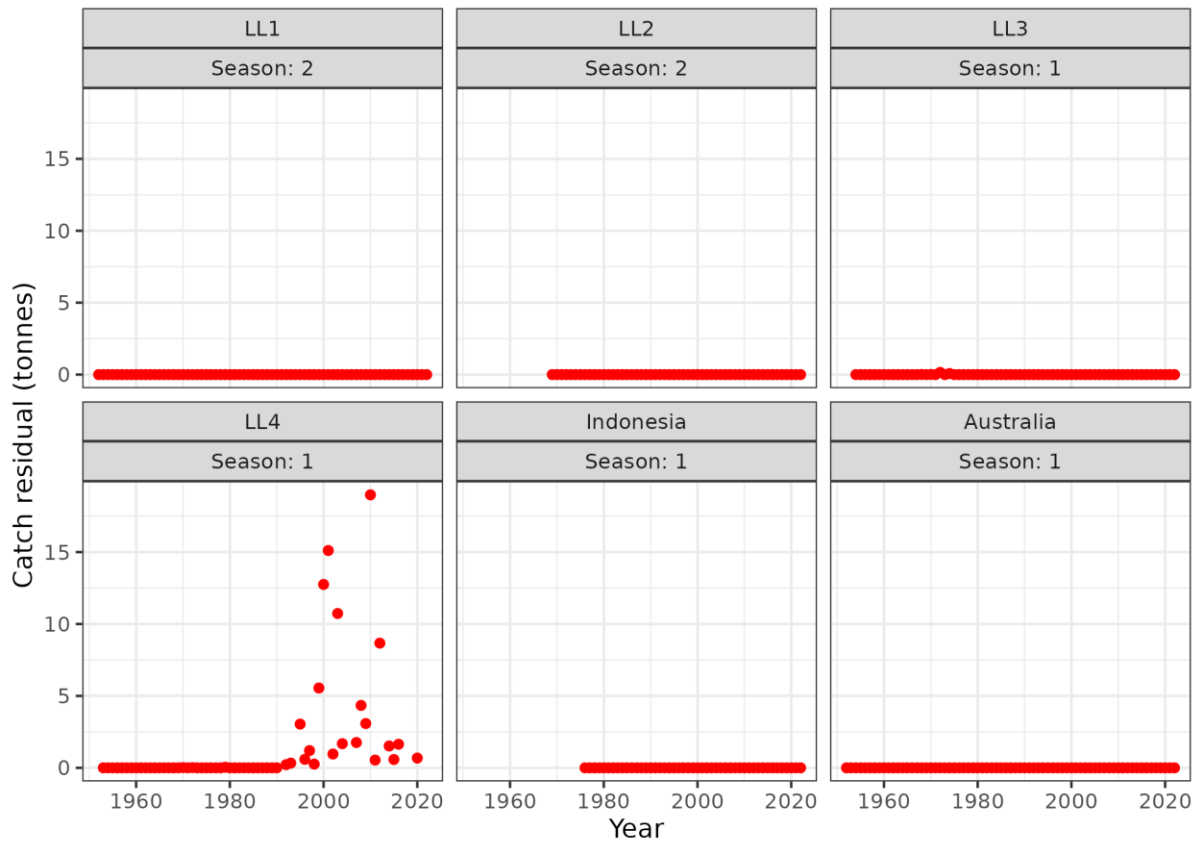


Figure 9: Catch residuals (input catch minus output catch, tonnes).

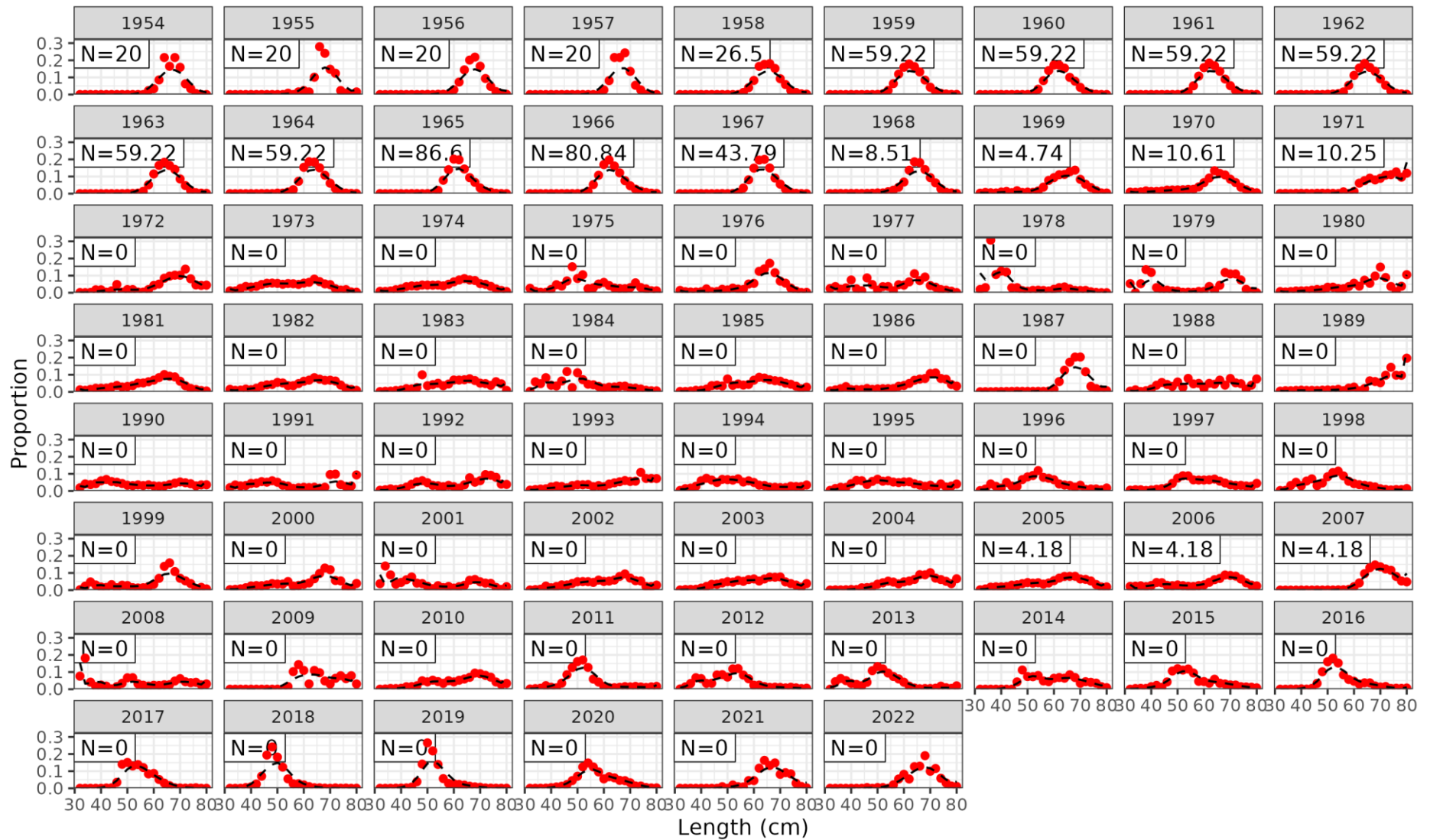


Figure 10: Observed LFs (red) and model fit (black) to the LL3 when treating as direct removals.

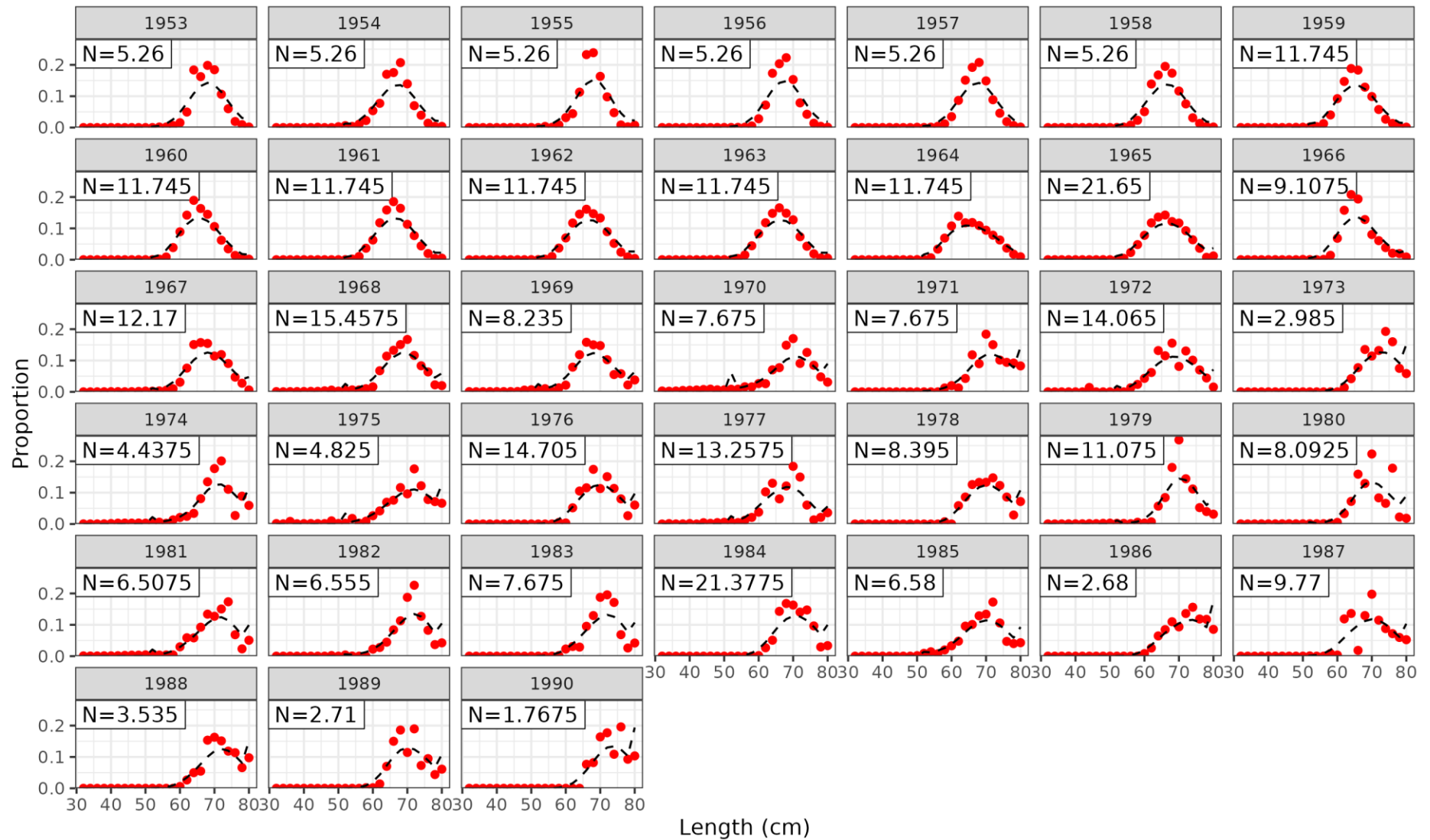


Figure 11: Observed LFs (red) and model fit (black) to the LL4 when treated as direct removals.

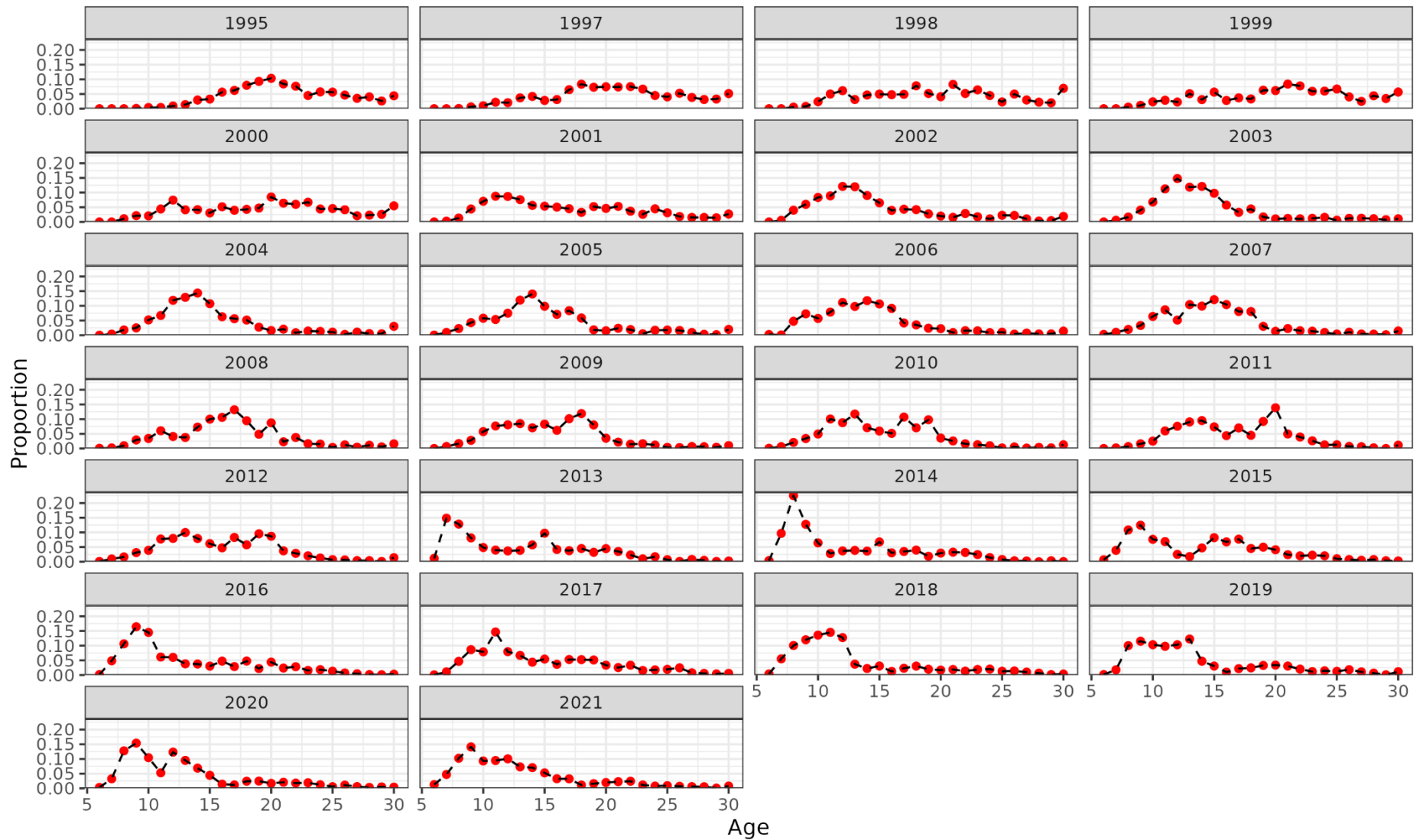


Figure 12: Observed AFs (red) and model fit (black) to the Indonesian fishery when treated as direct removals.

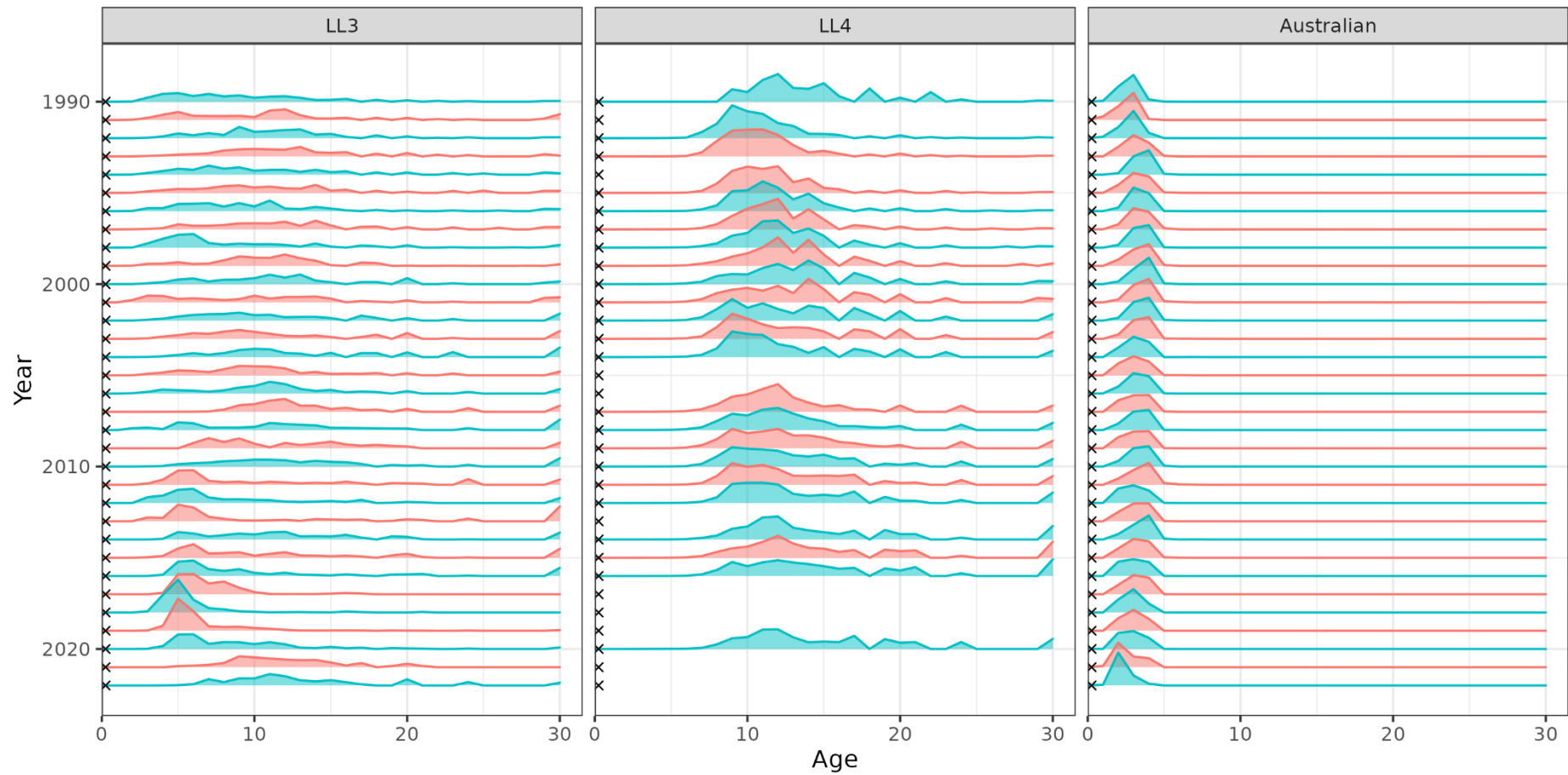


Figure 13: Implied selectivities (i.e., normalised harvest rates) for the LL3, LL4, and Australian surface fisheries when treated as direct removals.

GMRF selectivity

Following Cheng et al. (2023), time varying selectivity specified using GMRFs was added to the TMB code and can (optionally) be used to specify selectivity. When using this option for a fishery, the selectivity at age during each year is estimated (i.e., one parameter per age/year for the specified fishery). An additional four parameters can be estimated (for each fishery) including the partial selectivity correlation by age (ρ_f^{age}), the partial selectivity correlation by year (ρ_f^{year}), the partial selectivity correlation by cohort (ρ_f^{cohort}), and the variance of the GMRF process (σ_f^2). These parameters are only estimable if the selectivity parameters are treated as random effects or when using Bayesian inference.

In the current implementation, selectivity for each fishery is not estimated relative to an average (i.e., skeleton or prior) selectivity for that fishery, but instead as parameters in log-space locked up using the GMRF. An alternative approach is to estimate an average selectivity for each fishery, which is then adjusted by a selectivity deviate for each age/year and this deviate is locked up using a GMRF. The latter approach may be worth investigating in the future because there are divergent transition issues with the current selectivity specifications when using Bayesian inference and specifying GMRF selectivity. These issues have also hampered the estimation of the parameters defining the partial selectivity correlations and variance of the GMRF process.

Despite these issues, time varying selectivity that looks similar to that estimated by ADMB can be estimated when specifying the selectivity deviates and main effects and the partial selectivity correlations and variance of the GMRF process as fixed parameters (e.g., Figure 14 where the selectivity parameters were fixed at $\rho_f^{\text{age}} = \{0.5, 0.1, 0.5\}$, $\rho_f^{\text{year}} = 0.6$, $\rho_f^{\text{cohort}} = 0.001$, and $\sigma_f^2 = \{0.1, 0.2, 0.1\}$).

Discussion

This paper explores modifications to the OM that are in progress or have been implemented in the TMB version of the OM. Changes to the tagging likelihood have been implemented and tested; modifications to the POP likelihood is a work in progress; code has been developed to implement cohort slicing of the LFs for all fisheries with LFs; direct removal of catches has been coded, and time varying selectivity using GMRFs has been implemented. Other options have also been added to the TMB version of the OM (e.g., the option to fit to LFs/AFs using the Dirichlet-multinomial distribution) but were not explored in this paper.

References

Matthew LH. Cheng, James T. Thorson, James N. Ianelli, Curry J. Cunningham. (2023). Unlocking the triad of age, year, and cohort effects for stock assessment: Demonstration of a computationally efficient and reproducible framework using weight-at-age. *Fisheries Research* 266: 106755.

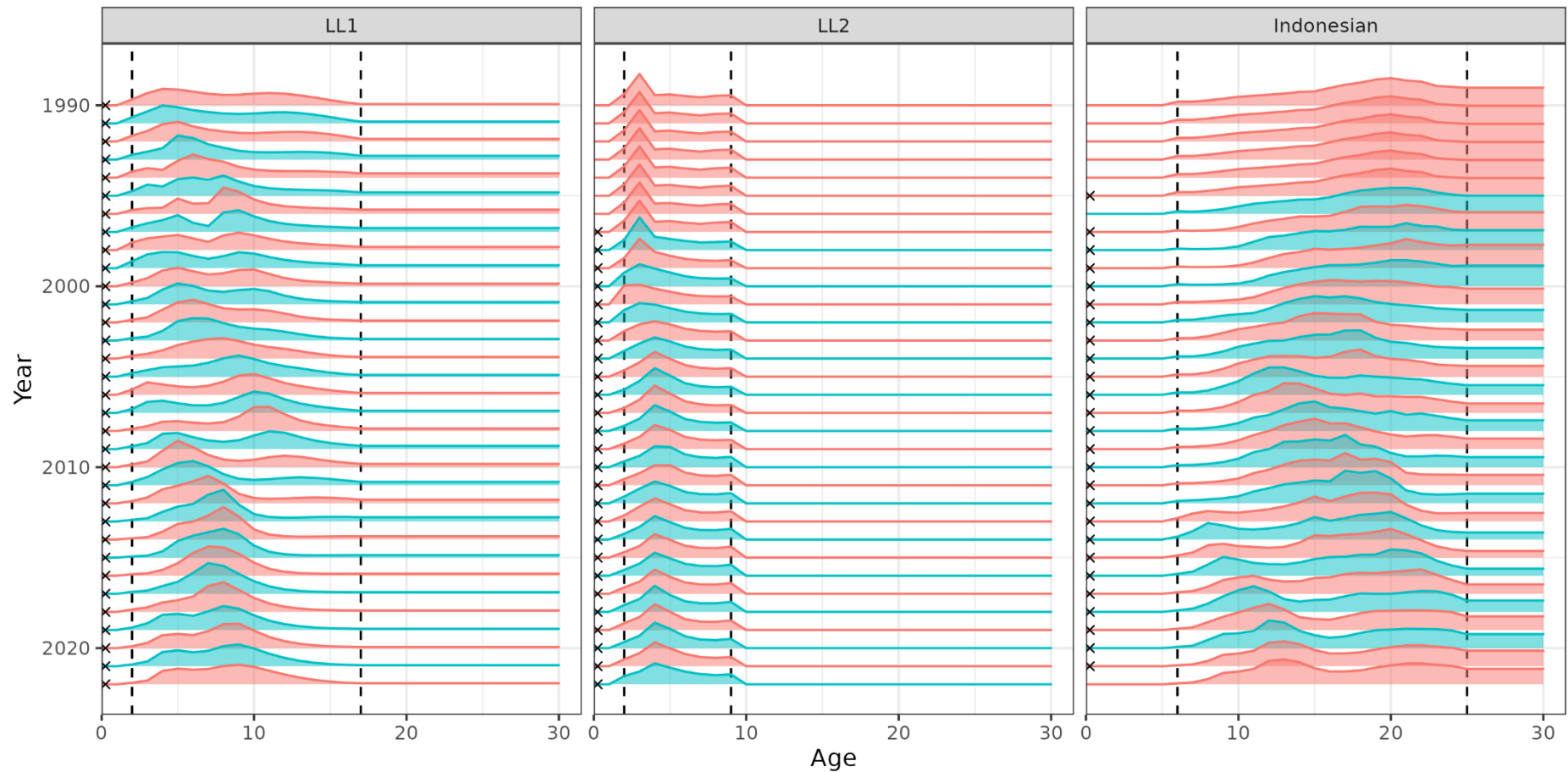


Figure 14: GMRF selectivities for the LL1, LL2, and Indonesian fisheries. The dashed vertical lines indicate the ages over which selectivities are estimated, the crosses to the left of the selectivity ogive each year within each panel indicate that LF or AF data are available during that year (e.g., there is no AF data in 1996 for the Indonesian fishery), and the colours represent selectivity blocks (i.e., the Indonesian selectivity in 2022 is assumed to be the same as in 2021).