

Southern Bluefin Tuna Operating Model Progress

D'Arcy Webber, Paige Eveson, Rich Hillary

Working Paper prepared for the 14th Operating Model and Management Procedure Technical Meeting, CCSBT, 20-24 June 2024, Seattle, USA.

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 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 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), and direct removal of catches. This paper aims to serve as a starting point for discussions during the OMMP14 meeting in Seattle.

Methods

Modifications to tagging likelihood

The tagging likelihood in the previous OM involved a large number of “H*” parameters corresponding to the harvest rate on 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.

This change has a small but insignificant 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 quite perfect fit (since natural mortality is constant across tagging cohorts and tagger groups).

A couple of additional changes were made to simplify the code for the tagging likelihood, but that have 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” one 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 for the adults in the comparison we only have its observed length-at-capture, we account for the distribution of possible ages this adult *could* be given its length. So, for comparisons of aged adults the POP probability remains as currently defined:

$$\mathbb{P}(POP | z_i, z_j) = \mathbb{I}(c_j < y_i < c_j + a_i) \frac{2\varphi_{c_j, a_i - (y_i - c_j)}}{TRO_{c_j}}$$

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:

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

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) given the length distribution of sampled adults changes from year to year which 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:

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

so 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 1 shows the fits to the SNP-genotyped adult length distributions, and the associated prior age distributions are displayed in Figure 2.

In the actual POP data, 94 of the POPs have an aged adult, with only 22 where length is the covariate, not age. So, 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.

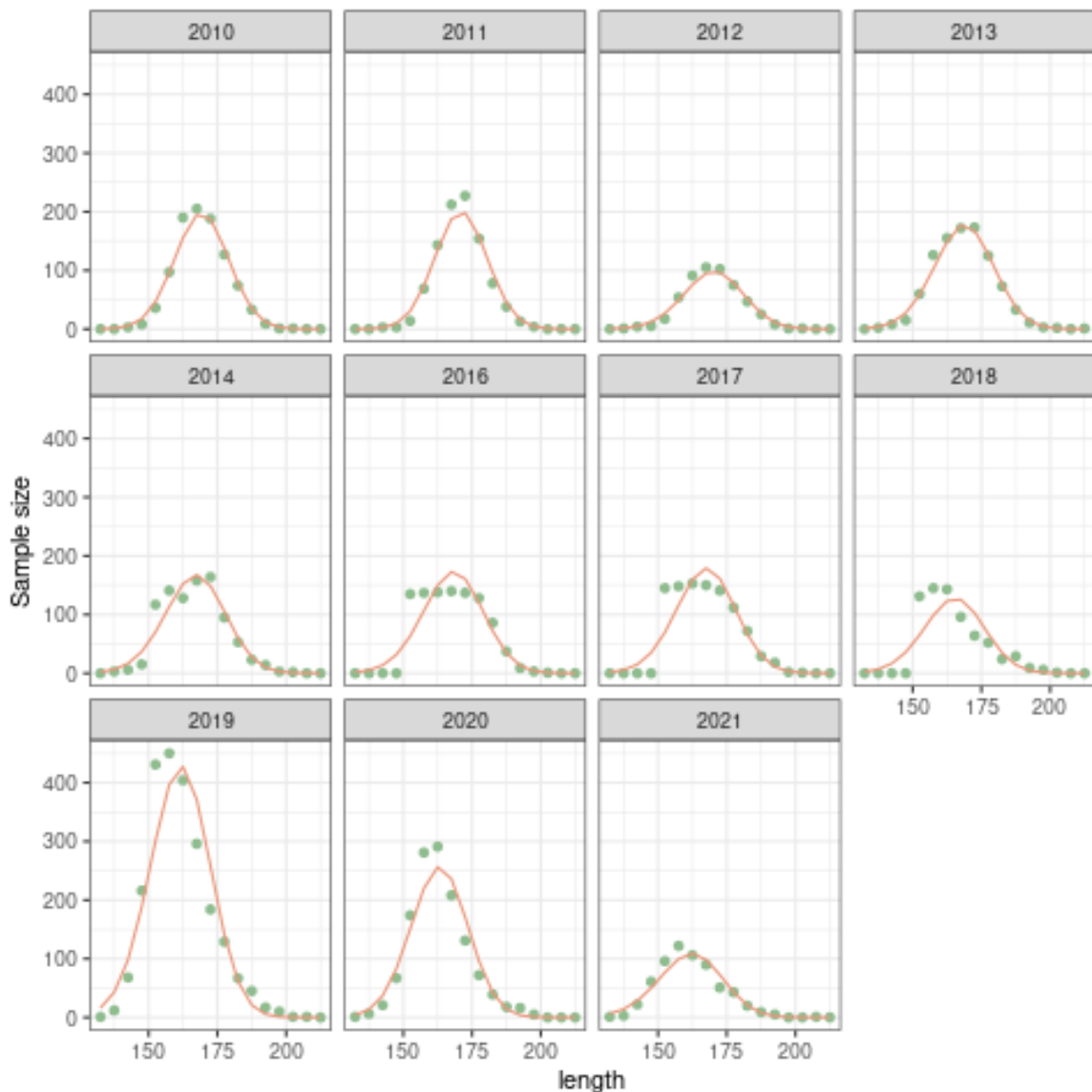


Figure 1: observed (green circles) and predicted (pink lines) length frequency of SNP genotyped adults.

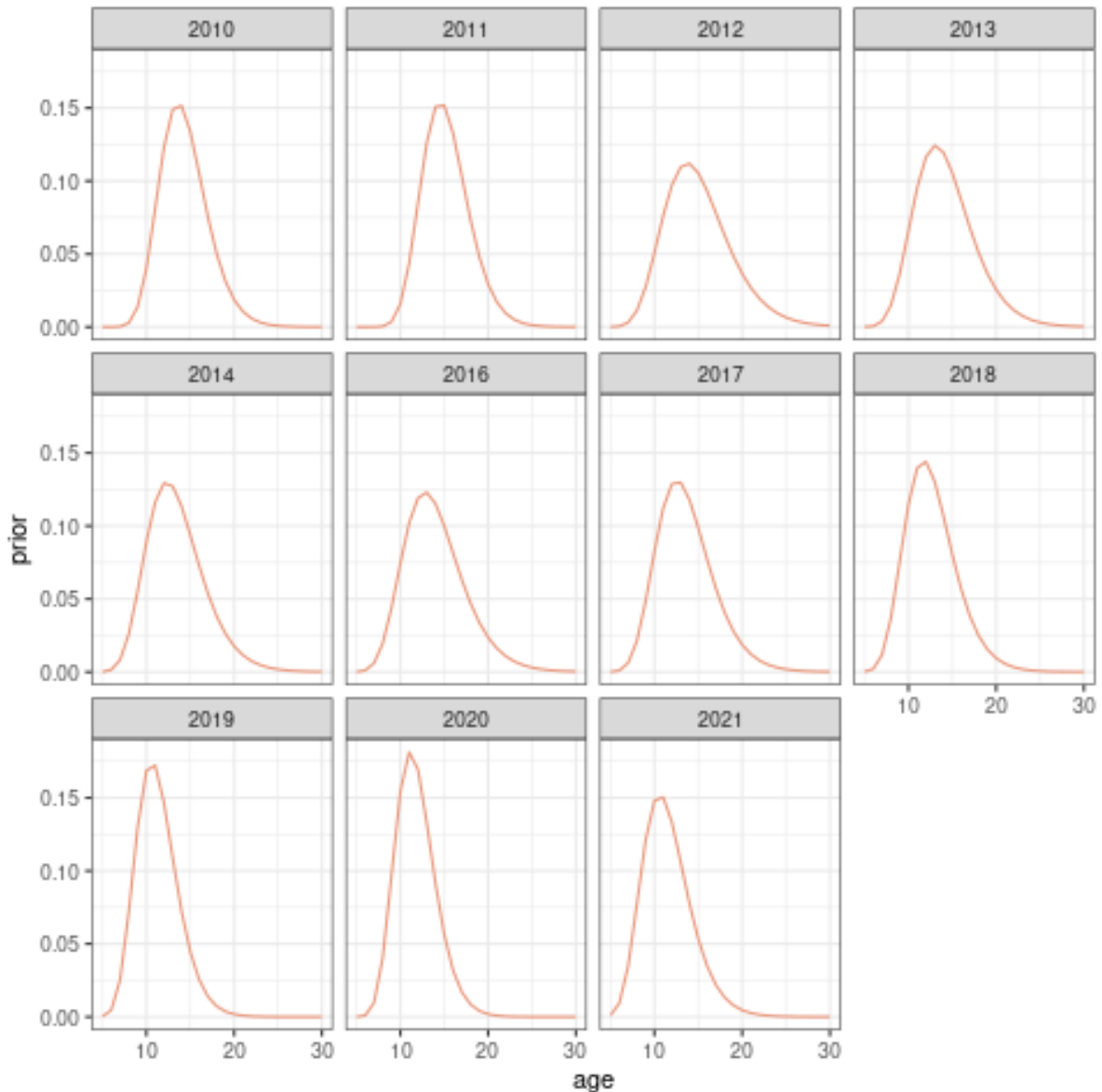


Figure 2: estimated prior age distribution-by-year for the SNP genotyped adults.

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 3), and the LFs for these two fisheries were fitted to separately (Figure 4, Figure 5).

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 actually 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 6, Figure 7).

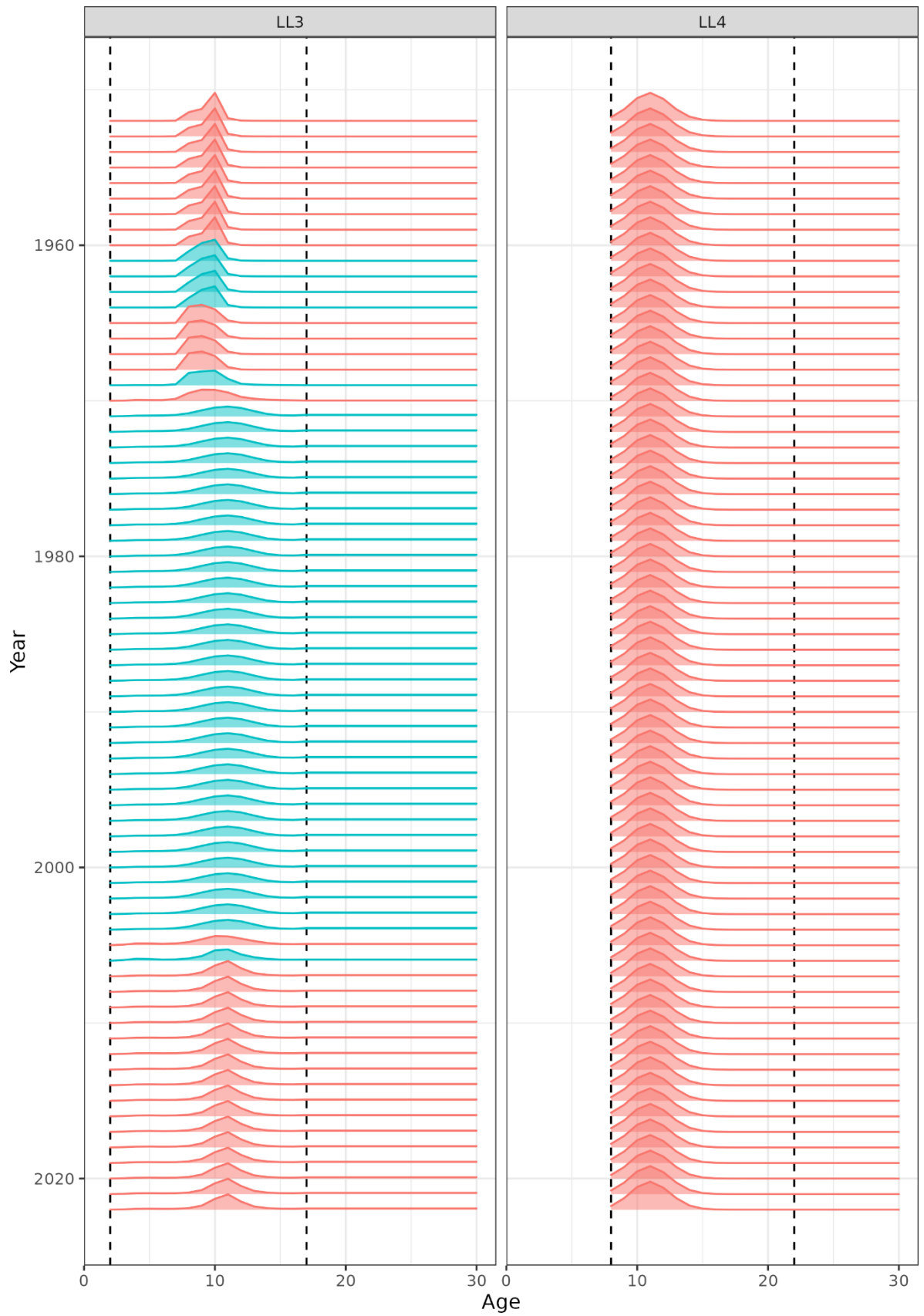


Figure 3: selectivity estimated by the previous OM for LL3 and LL4. The colours represent selectivity periods of that are assumed to be the same (i.e., there are nine periods of different selectivity for LL3 and LL4 selectivity is time invariant).

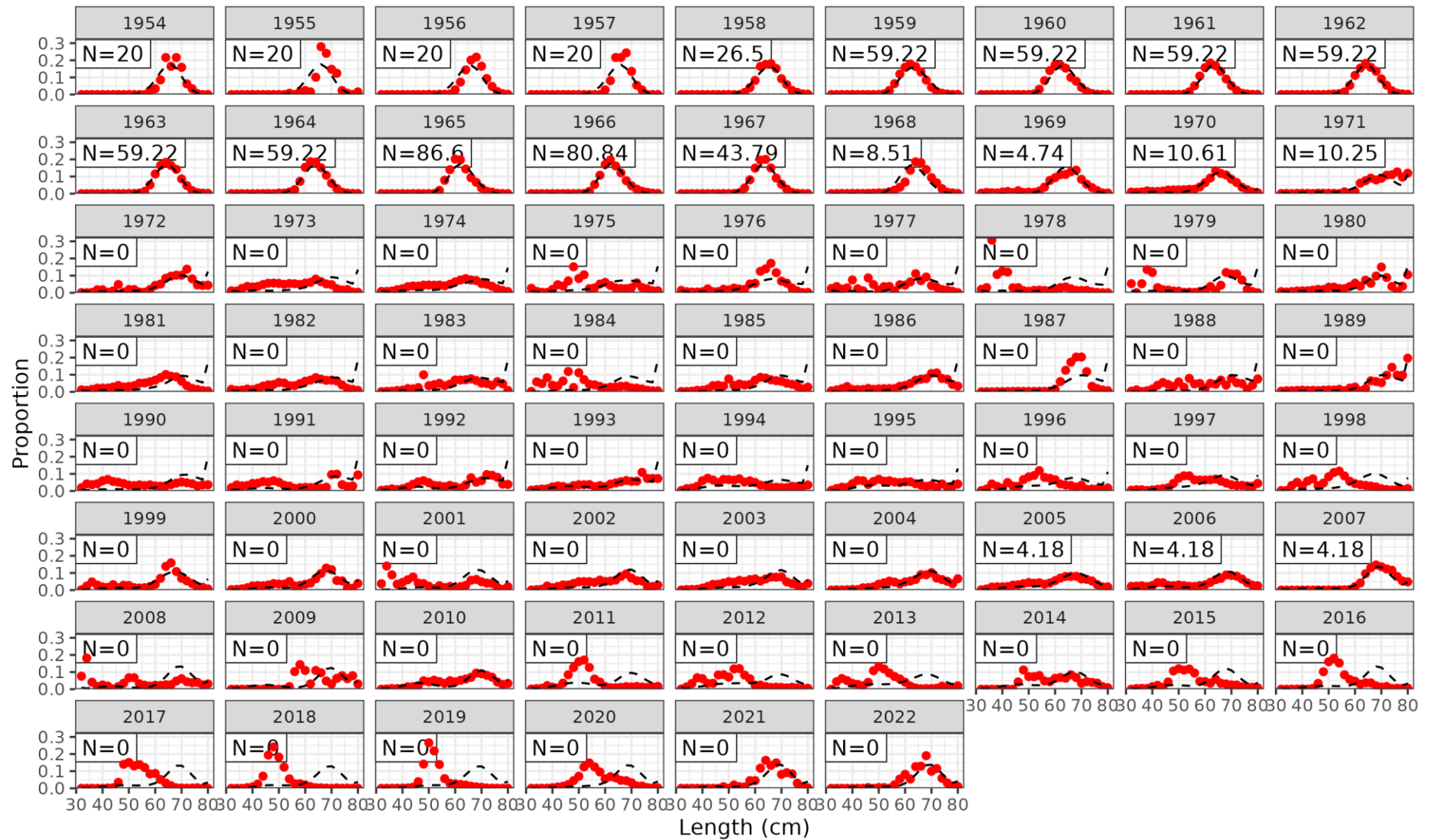


Figure 4: Observed LFs (red) and model fit (black) to the LL3 LFs. The effective sample size (N) is also shown for each year.

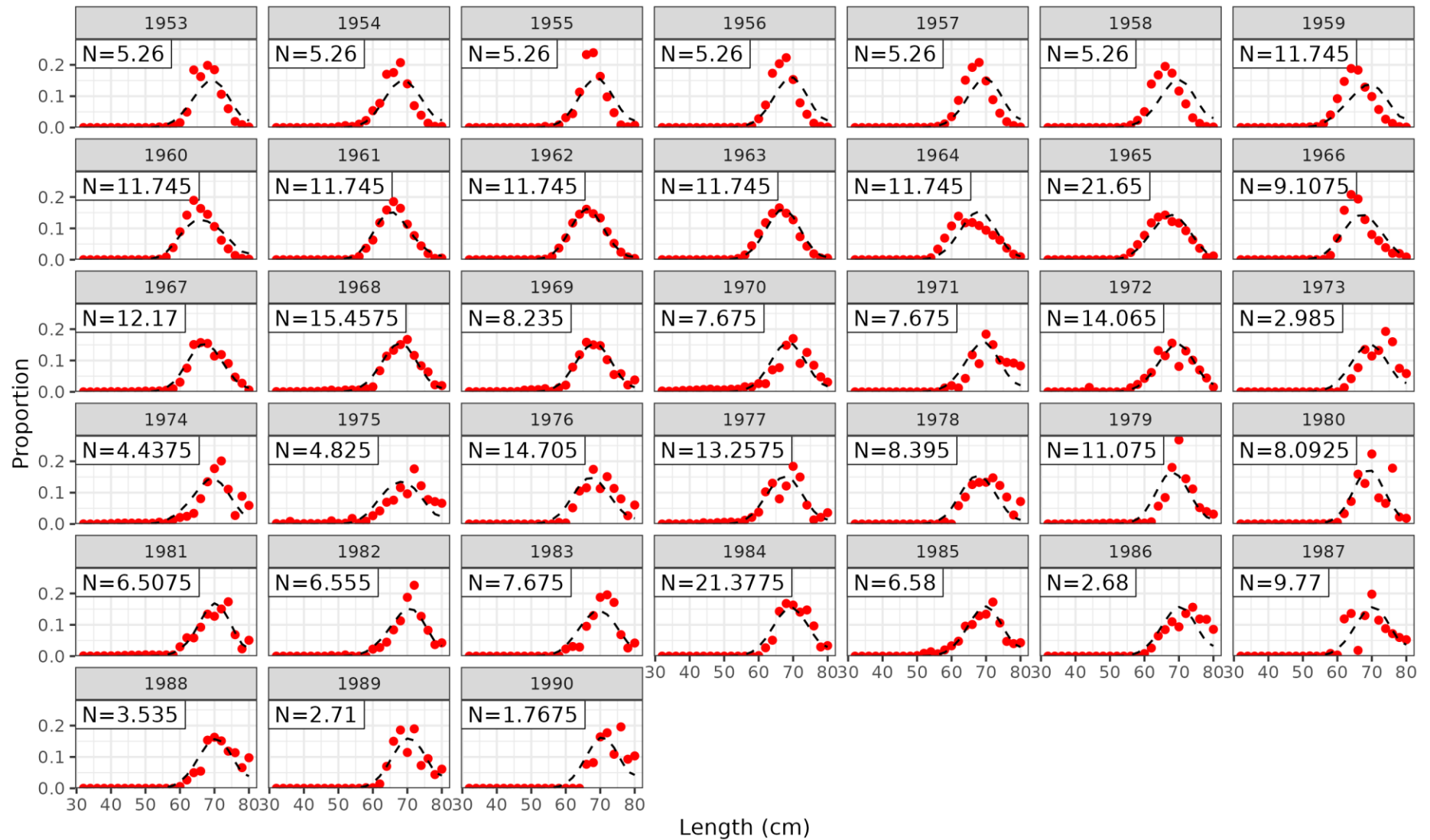


Figure 5: Observed LFs (red) and model fit (black) to the LL4 LFs. The effective sample size (N) is also shown for each year.

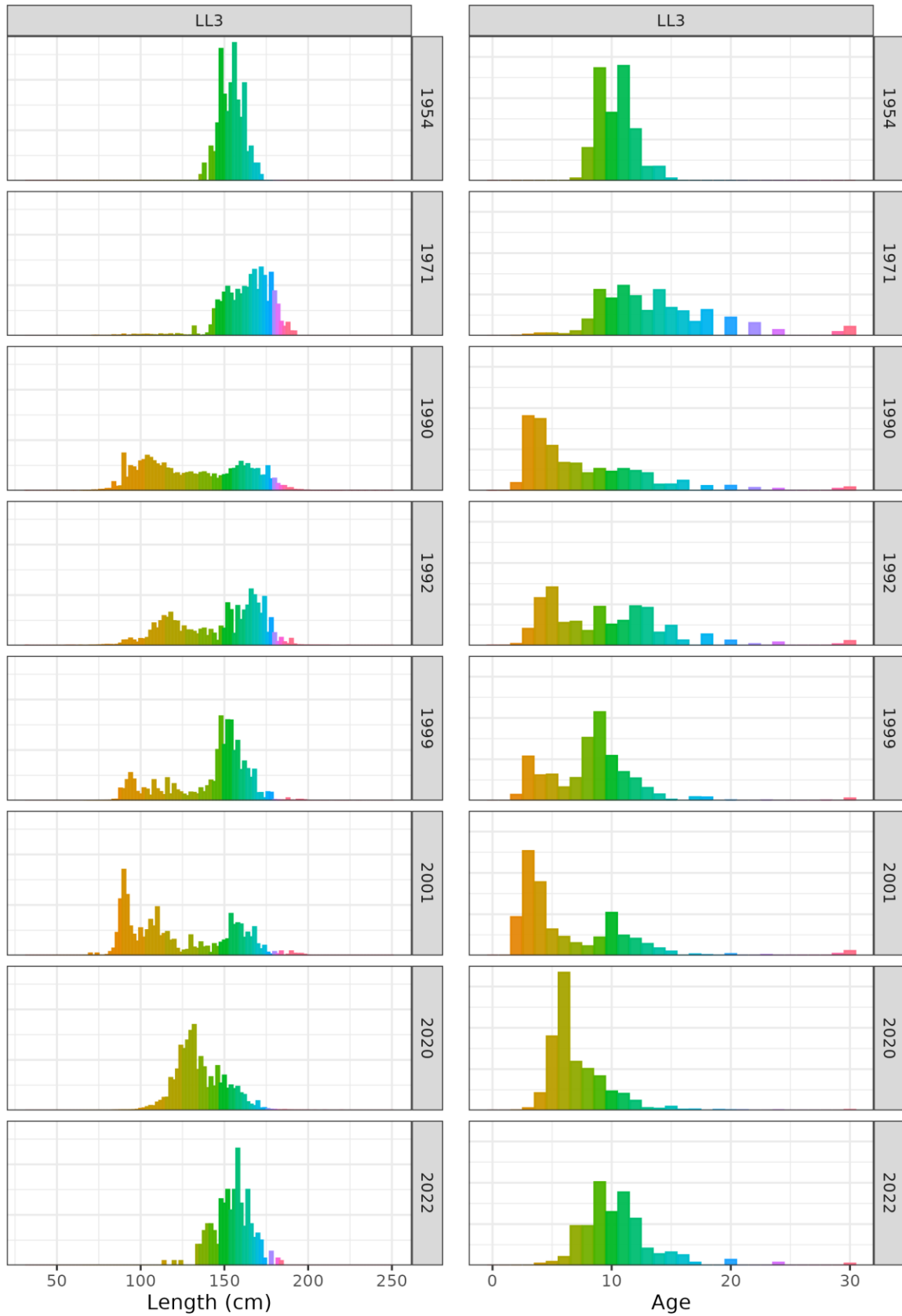


Figure 6: original LFs for a subset of years (left) and sliced LFs for those same years (right) for the LL3 fishery. Each colour represents an age.

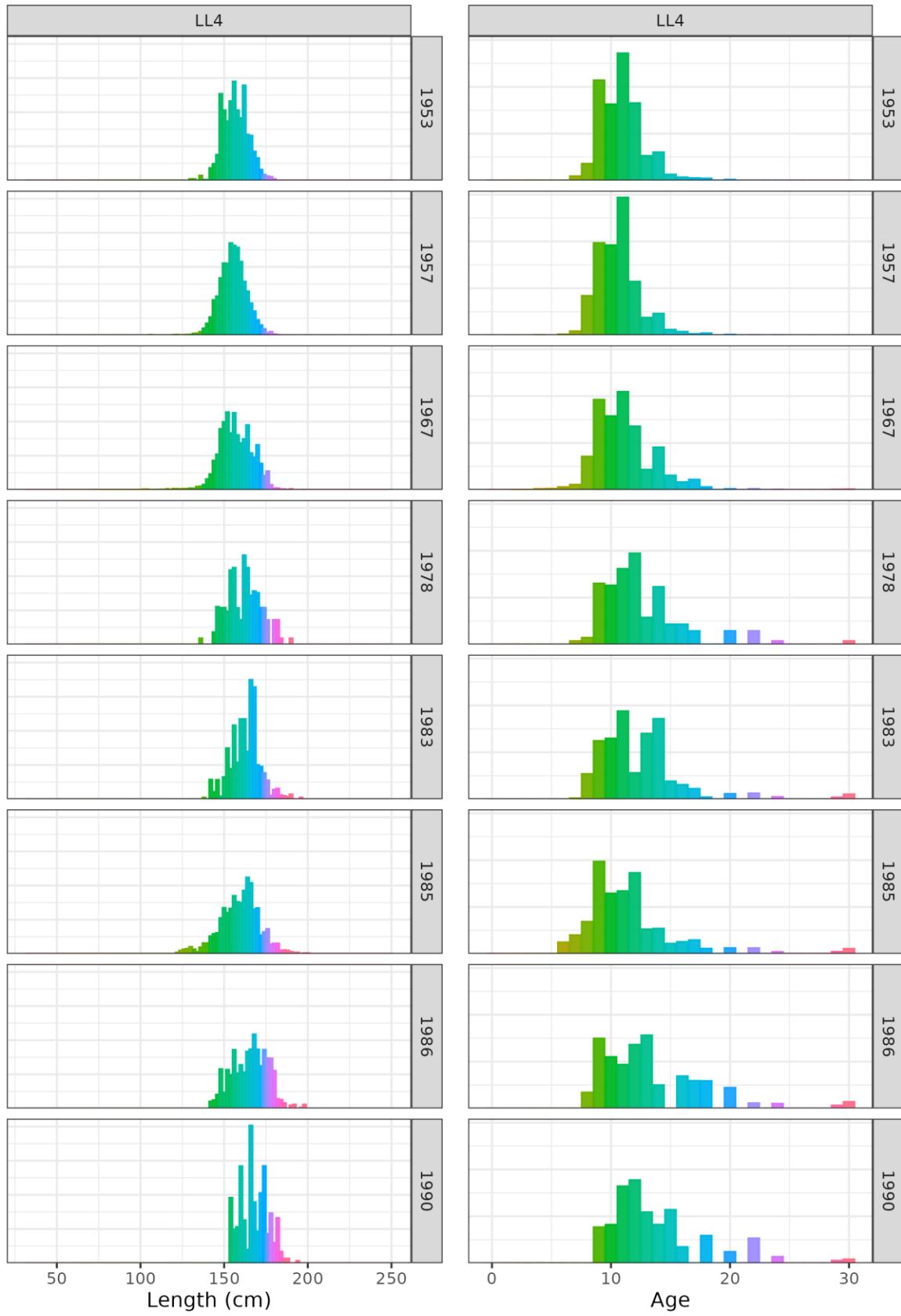


Figure 7: original LFs for a subset of years (left) and sliced LFs for those same years (right) for the LL4 fishery. Each colour represents an age.

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 further re-structured 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 AF or the sliced AF ($A_{y,s,f,a}$) 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 8). 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 could be amended by combining the LL3 and LL4 fisheries (this code change will be done in future updates to the TMB code).

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 a little (Figure 9, Figure 10). However, when a fishery with AFs is specified to use as direct removals then the predicted AFs match the observed AFs exactly (Figure 11).

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
$A_{y,s,f,a}$	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> $C_{y,s,f,a}^N = X_{y,s,f} A_{y,s,f,a}$
$H_{y,s,f,a} = \frac{U_{y,s,f,a} S_{y,f,a}}{\sum_a N_{y,s,a} S_{y,f,a} W_{y,f,a}}$	$H_{y,s,f,a} = \frac{C_{y,s,f,a}^N}{N_{y,s,a}}$ $= \frac{X_{y,s,f} A_{y,s,f,a}}{N_{y,s,a}}$ $= \frac{C_{y,s,f} A_{y,s,f,a}}{N_{y,s,a} \sum_a A_{y,s,f,a} W_{y,f,a}}$
<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> $C'_{y,s,f} = H_{y,s,f,a} N_{y,s,a} W_{y,f,a}$	

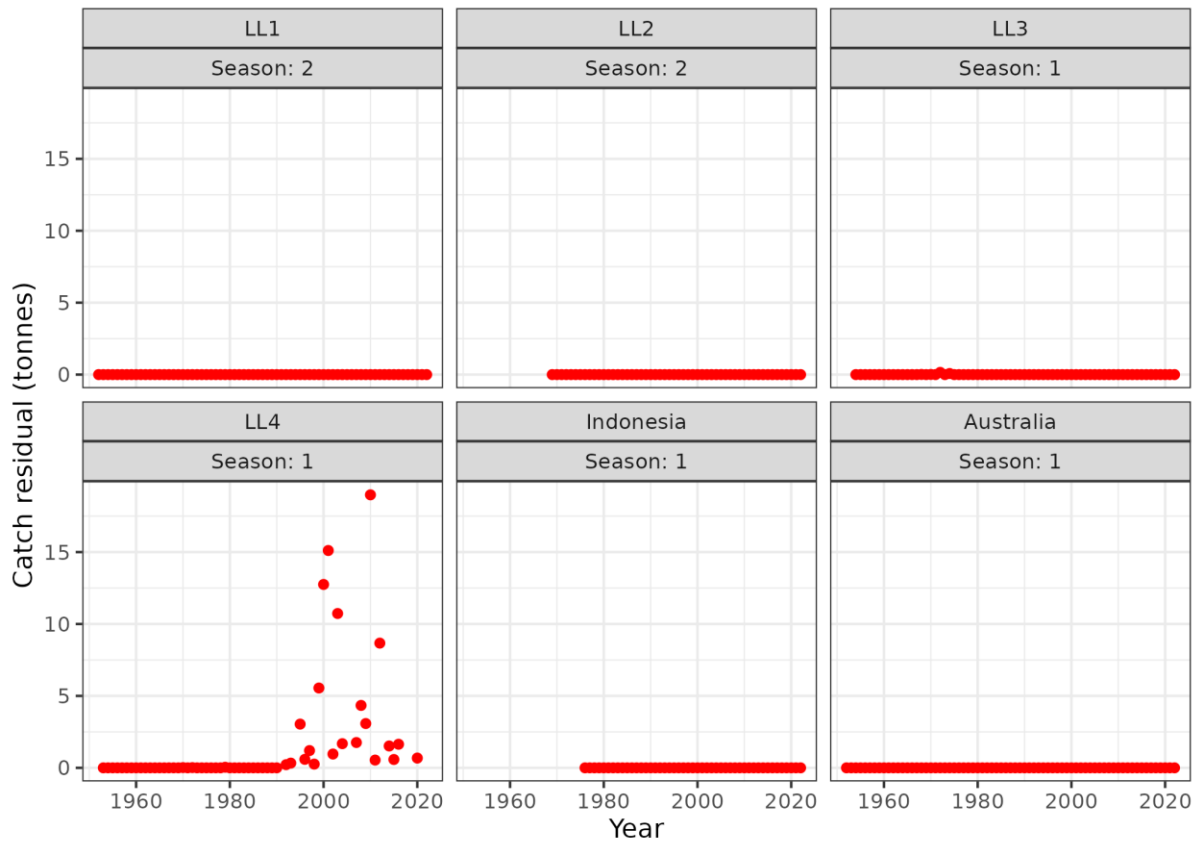


Figure 8: Catch residuals (input catch minus output catch, tonnes).

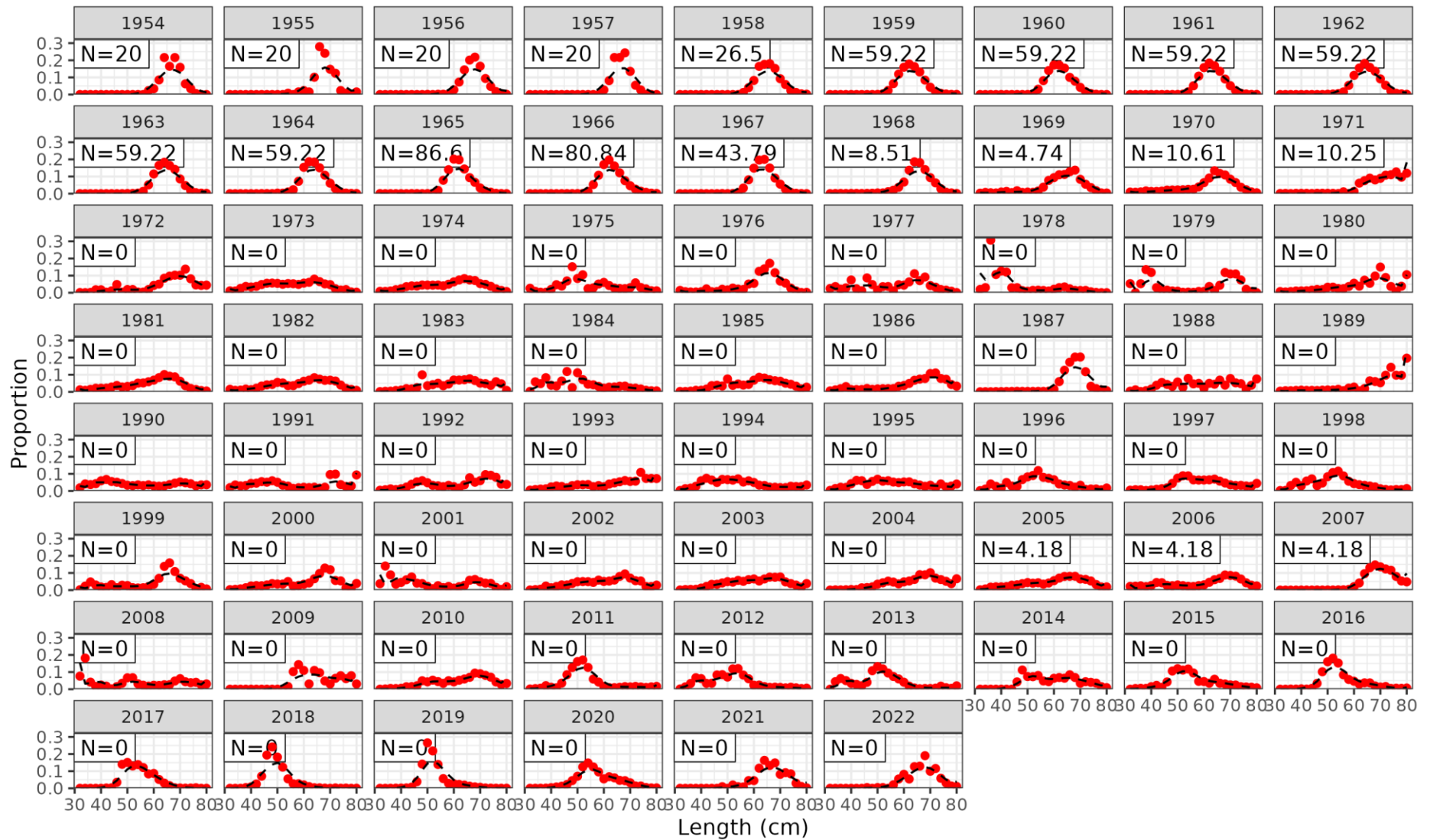


Figure 9: Observed LFs (red) and model fit (black) to the LL3 when treating as direct removals.

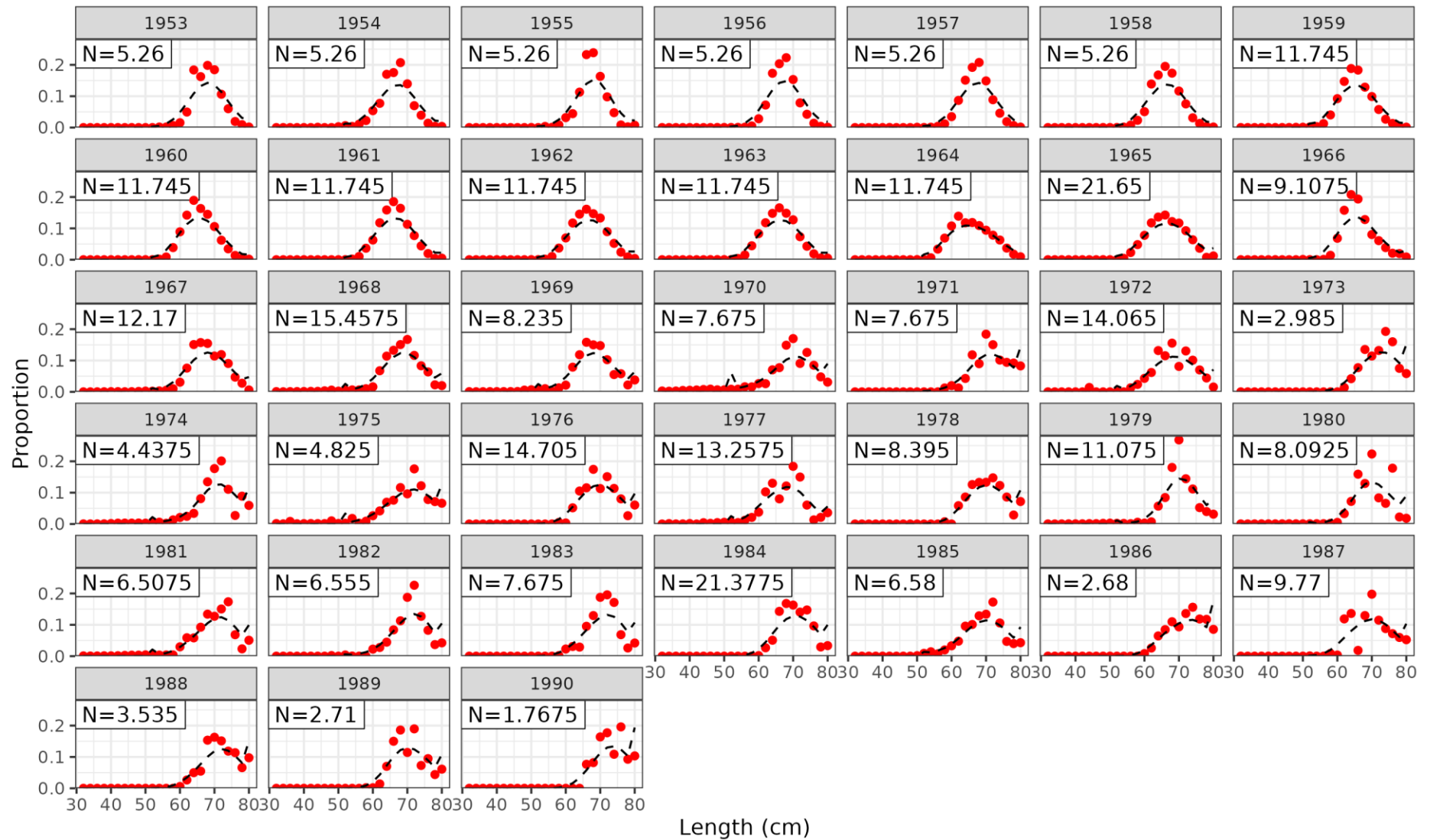


Figure 10: Observed LFs (red) and model fit (black) to the LL4 when treated as direct removals.

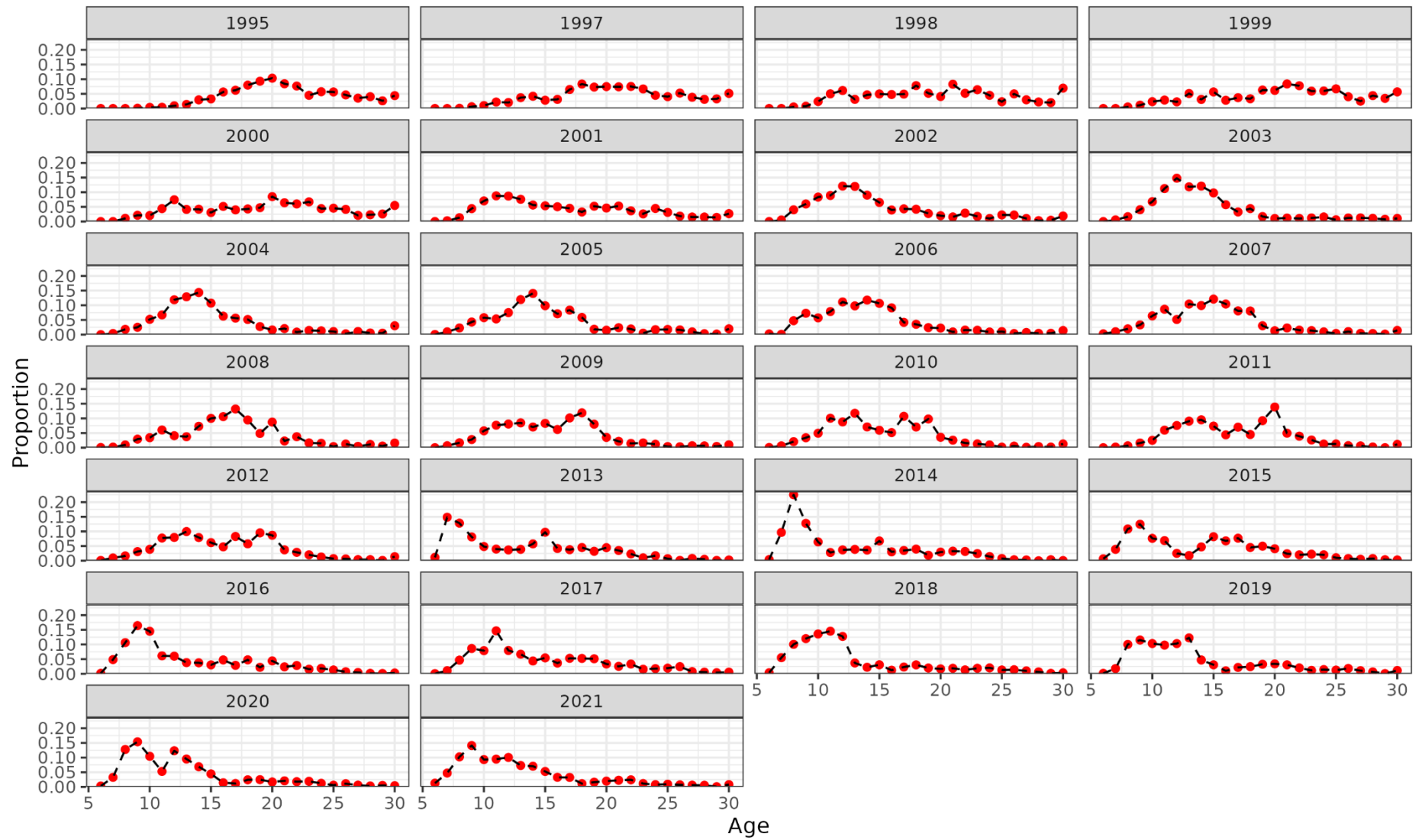


Figure 11: Observed AFs (red) and model fit (black) to the Indonesian fishery when treated as direct removals.

Discussion

This paper explores some 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; and direct removal of catches has been coded but requires some minor modifications going forward. 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.

Future code changes to the TMB OM will include: implementing Gaussian Markov random field (GMRF) selectivity (this has been coded but is still in its testing phase); one step ahead (OSA) residuals (this has been explored using the R package, but this will need to be integrated into the TMB code for future use), and potentially the option for specifying length based natural mortality (M).