Fisheries New Zealand
Tini a Tangaroa

Development of models for estimating age and length compositions from length and age observations

New Zealand Fisheries Assessment Report 2024/29

D.N. Webber
A. Dunn
S. Mormede

ISSN 1179-5352 (online)
ISBN 978-1-991308-01-6 (online)

July 2024

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:

Webber, D.N.; Dunn, A.; Mormede, S. (2024). Development of models for estimating age and length compositions from length and age observations. *New Zealand Fisheries Assessment Report 2024/29.* 79 p.

TABLE OF CONTENTS

EXECUTIVE SUMMARY 1

1. INTRODUCTION 2

1.1 A review of age-length keys 4

2. METHODS 4

2.1 Length composition models 4
2.1.1 Red rock lobster in CRA 2 5
2.1.2 Hake in HAK 7 and ling in LIN 3&4 6

2.2 Age conditional on length models 9

2.3 Estimating the scaled length compositions for red rock lobster, hake, and ling 11

2.4 Estimating the scaled age compositions using the scaled length composition and the age-at-length model 12

2.5 Empirical estimation of the scaled age compositions using ALKs 12

3. RESULTS 12

3.1 Length composition models 12
3.1.1 Red rock lobster in CRA 2 12
3.1.2 Hake in HAK 7 and ling in LIN 3&4 13

3.2 Estimates of the scaled length compositions 14
3.2.1 Red rock lobster in CRA 2 14

3.3 Age conditional on length models 14
3.3.1 Hake in HAK 7 and ling in LIN 3&4 14

3.4 Estimates of the scaled age compositions for hake 30

4. DISCUSSION 33

4.1 Scaled length compositions 33

4.2 Age-at-length and scaled age compositions 33

5. POTENTIAL RESEARCH 34

6. ACKNOWLEDGEMENTS 34

7. REFERENCES 35

8. APPENDIX 1: TABLES AND FIGURES FOR THE LENGTH COMPOSITION MODELS 37

8.1 Red rock lobster in CRA 2 37

8.2 Hake in HAK 7 and ling in LIN 3&4 48

9. APPENDIX 2: TABLES AND FIGURES FOR THE SCALED LENGTH COMPOSITIONS 51

9.1 Red rock lobster in CRA 2 51

10. APPENDIX 3: TABLES AND FIGURES FOR THE AGE AT LENGTH ORDINAL MODELS 62

10.1 Hake in HAK 7 62

10.2 Ling in LIN 3&4 70

11. APPENDIX 4: TABLES AND FIGURES FOR THE EMPIRICAL SCALED AGE COMPOSITIONS USING ALKS 76

11.1 Hake in HAK 7 and ling in LIN 3&4 76

Plain language summary

Integrated age and length structured stock assessment models require age and/or length composition data from catch or surveys to inform selectivity ogives and year class strengths. We develop model-based approaches using categorical and ordinal models that can include spatial and temporal variability and help account for missing data and/or non-representative data to generate estimates of age and length compositions. The results show that these approaches can improve the estimates of age compositions when compared with those using empirical methods.

EXECUTIVE SUMMARY

Webber, D.N.[1]; Dunn, A.[2]; Mormede, S.[3] (2024). Development of models for estimating age compositions from length and age observations.

***New Zealand Fisheries Assessment Report 2024/29.* 79 p.**

Integrated age and length structured stock assessment models use age and/or length composition data, either from the commercial catch or surveys, as observations to inform selectivity ogives and relative year class strengths. These data provide key information to allow the estimation of changes in demographic structure over time.

Scaled age compositions for fish stocks are typically derived using empirical methods where length observations and fish otoliths are randomly sampled from the catch, with the length samples used to estimate the scaled length composition and then multiplied by an age-length key (ALK) to produce the scaled age compositions. However, the collection of otoliths and subsequent reading for age can be expensive and the sampled otoliths need to be collected annually, be representative of the length-at-age of the population, be spatially representative (to account for variability in length-at-age across locations) and temporally representative (to account for seasonal growth), and have an adequate number of samples to fully describe the age-length relationship.

Model based approaches can better deal with spatial and temporal variability and infer parameters (with appropriate levels of uncertainty) for subsets of the parameter space where there may be little data available. Methods that model length frequency using categorical models were developed with dependent variables that include parameters for year, season, and location. Ordinal models were also developed to estimate the age composition given the scaled length composition. By leveraging the relationships between samples across years, seasons, and space, the estimates of parameters where the data are sparse or missing can be obtained with appropriate estimates of uncertainty.

Specifically, (i) categorial models were developed to estimate length frequency and associated uncertainty from samples of length with appropriate covariates; (ii) these were then used to estimate scaled length compositions; and (iii) ordinal models were developed to estimate age given length from samples of paired age-length data and with appropriate covariates. These scaled length composition and age given length models may then be combined to derive the scaled age composition.

In this project, software and the intensive computer resources required to fit multinomial categorical models meant that the length frequency data were not well fitted, and issues arose in developing models that can perform adequately for data sets with large numbers of sparse observations. Aggregation of length data allowed preliminary models to be fitted, but these had poor Markov chain Monte Carlo (MCMC) performance for hake and ling, but were adequate for rock lobster. Further work on modifying the assumption of multinomial distributions to Poisson decomposed multinomial distributions suggests that this may perform better and be able to be used to provide scaled length compositions for hake and ling, but it has not yet been applied to this case.

Results show that this approach generates equivalent estimates to empirical scaled length frequencies where the length sampling was adequate, and the age given length models can predict equivalent age-length models to ALKs. In cases where some strata, time periods, or years are missing, these models can be used to generate age compositions, but only if the run time and MCMC mixing of length frequency models can be improved. Sample size estimates for the length frequency composition appear to be better estimated that those calculated using bootstrapping; however, additional model development is required to be able to adequately model the length frequency data.

[1] Quantifish Ltd, New Zealand.
[2] Ocean Environmental Ltd, New Zealand.
[3] soFish Ltd, New Zealand.

1. INTRODUCTION

Integrated age and length structured stock assessment models use age and/or length composition data, either from the commercial catch or surveys, as observations to inform selectivity ogives and relative year class strengths. These data provide key information to allow the estimation of fishery removals and changes in the demographic structure over time.

Scaled age compositions for fish stocks are typically derived using empirical methods by either (i) randomly sampling the population or the catch and using these samples to directly infer the age composition (direct ageing) or (ii) randomly sampling the population for catch lengths, sampling a smaller set of aged otoliths to develop an empirical age-length key (ALK), and then applying the ALK to the scaled length composition to produce the age composition.

With the ALK method, the random length observations and fish otoliths are sampled either as a part of resource survey sampling, from commercial operations by observers on fishing vessels, or from collecting market (or fish shed) samples. These samples are then used to estimate the scaled length composition and multiplied by the ALK to produce the scaled age compositions.

Estimates of age composition for many New Zealand stocks use the ALK method although a limited number of deepwater and most inshore stocks use direct ageing. With both methods, bootstrap resampling has typically been used to provide uncertainty estimates and these are used to estimate initial (and relative) effective sample sizes for the age composition data when used in stock assessments (Bull & Dunn 2002) or other analyses. In the cases that use direct ageing, age composition analyses do not combine information across years or sampling methods and can have issues related to poor sampling of individual strata or spatial temporal non-representativeness that may bias the resulting age composition estimates.

There are a number of methods available to empirically produce scaled 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). In general, almost all fishery age compositions derived from ALKs for New Zealand fisheries use the standard method (for example, the method of Bull & Dunn 2002) and typically assume negligible spatial or temporal bias.

The collection of otoliths and subsequent reading for age can be expensive (depending on preparation required to read an otolith, the time taken to read each otolith, and the cost of obtaining the initial otolith samples), with the cost often imposing a constraint on the number of otoliths able to be aged as a part of a sampling programme. Using empirical methods, the sampled otoliths need to be collected annually, be representative of the length-at-age of the population, be spatially representative (to account for variability in length-at-age across locations) and temporally representative (to account for seasonal growth), and have an adequate number of samples to fully describe the age-length relationship. Where in-season growth is considerable and age compositions are required for different seasonal periods, multiple age-length keys may need to be obtained to adequately estimate age compositions from the length data.

Although otoliths are often collected at random within tows/hauls from surveys and fishing events (albeit only random at length in some cases), these are unlikely to be random with respect to location or time within season. Further, subsampling of the otoliths to be aged typically uses small numbers, that can inadvertently concentrate the samples from a subset of locations or part of the season. Hence the ages used in ALKs may contain spatial or temporal bias, and the patterns of bias may change over time (Ailloud & Hoenig 2019). In addition, sampling may not collect otoliths from critical areas or seasons of the population or from a fishery in some years. This can lead to data being discarded or unable to be used (e.g., Chatham Rise hake where the number of otoliths collected have been inadequate in recent years to construct a suitable ALK) and can result in wasted resources (e.g., significant proportions of data may have to be discarded if it is deemed that there are not enough data in a specific strata or year to adequately estimate the length or age composition). Current empirical methods using ALKs are also likely to be statistically inefficient and estimates of uncertainty may not be adequate.

Model based approaches can better deal with spatial and temporal variability and infer parameters (with appropriate levels of uncertainty) for subsets of the parameter space where there may be few data available. These approaches can allow estimates of age compositions for areas and years where a sample of length observations or otoliths may be inadequate. In this case, methods were developed to model length using categorical models with dependent variables that include parameters for year, season, and location. Separately, ordinal models were fitted to paired age and length data and these models can be used instead of an ALK to estimate the age compositions. By leveraging the relationships between samples across years, seasons, and space, the estimates of parameters where the data are sparse or missing can be obtained with appropriate estimates of uncertainty.

This increase in efficiency and the use of the samples across years, seasons, and areas allows better use of the age and length samples, potentially lowering the number of samples required to obtain similar levels of precision as empirical methods as well as allowing for the inclusion of spatial-temporal and other effects into the age composition estimates. While outside the scope of this study, such models may also be used to evaluate changes in the spatial-temporal distributions over time, allowing for consideration of climate or other environmental changes in length and age composition observations, as well as similar changes in growth productivity parameters that are estimated using these data.

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. Developments have used continuation ratio-logits (Rindorf & Lewy 2001), and the approach 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 age given length 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 by 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 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 covariates, and have improved statistical properties over standard survey design-based methods.

Since the studies of Rindoff & Lewy (2001), Berg & Kristensen (2012), and Berg et al. (2014), modelling software to implement ordinal regressions have improved considerably — for example *mgcv* (Wood et al. 2016, Wood 2017), *Stan* (Carpenter et al. 2017), and *brms* (Bürkner 2017, 2018).

In this report, ordinal regression was used to estimate age composition estimates given length and age samples, with length composition data estimated using categorical models using *brms*. While there has been some development using ordinal regression for age compositions (Rindorf & Lewy 2001, Berg & Kristensen 2012, Babyn et al. 2021), we revise and extend the method and use Bayesian inference.

Specifically, we (i) develop categorial models to estimate length composition (i.e., length frequency estimates) and associated uncertainty from samples of length with appropriate covariates; (ii) develop ordinal models to estimate age composition given length from samples of paired age-length data and with appropriate covariates; and hence (iii) investigate methods for combining these data to derive scaled age compositions from these models.

1.1 A review of age-length keys

In empirical ALK age composition methods, the ALK is used to translate the length composition data into an age composition, typically applied to a fleet, area, and season in each year. An ALK can be defined as an $a \times \ell$ joint distribution matrix where each row corresponds to an age a and each column to a length class ℓ. The ALK specifies the relative proportion of fish of length ℓ that are age a and can be written as:

$$M_{a,\ell} = P(A_i = a | Y_i = \ell) = \phi(Z_{a,\ell})$$

where $P(A_i = a | Y_i = \ell)$ is the conditional probability that a randomly selected fish i is of age a given that it is in length class ℓ. This conditional probability is usually determined as the empirical distribution of sampled ages and is equivalent to $\phi(Z_{a,\ell})$ which is the probability density that a fish of length ℓ is age a, i.e., $P(A_i = a | Y_i = \ell)$ is the proportion of fish at length ℓ with age a. The ALK is applied to length composition data to derive proportions in the catch-at-age observations:

$$P_a = \sum_{\ell} \left(\frac{M_{a,\ell} Q_\ell}{\sum_{a'} M_{a,\ell}} \right) \text{ where } \sum_a P_a = 1$$

where P_a is typically then used as an observation of the age composition in stock assessment models.

Within New Zealand, sampling uncertainty and associated effective sample sizes are usually estimated using either bootstrapping (Bull & Dunn 2002), or by approximating the sampling error from the number of sampled hauls or number of measurements (Langley 2021). Methods using bootstrapping use the sampled length composition and resample individual length samples from each haul and trip within strata to calculate uncertainty (Bull & Dunn 2002), and the consequent effective sample sizes with uncertainty in the ALK are ignored.

2. METHODS

Bayesian inference for the length composition categorical models and the age-length ordinal models was done using the R package *brms* (Bürkner 2017). The *brms* package is a wrapper for generalised linear mixed effects models (GLMMs) using Stan (Stan Development Team 2020) with Markov chain Monte Carlo (MCMC) using the Hamiltonian Monte Carlo (HMC) algorithm. Model selection was done using standard Bayesian MCMC diagnostics including the leave-one-out information criteria (LOO IC, see Vehtari et al. 2017) and Bayesian R^2. Using *brms* allows for flexible model specification and enables the approach to be easily implemented for other fish stocks.

2.1 Length composition models

Three example stocks were used to develop and compare methods for modelling length frequencies and the subsequent scaled length compositions: red rock lobster (*Jasus edwardsii*) in CRA 2 (Roberts et al. 2021); hake (*Merluccius australis*) in HAK 7 (Dunn et al. 2023); and ling (*Genypterus blacodes*) in LIN 3&4 (Mormede et al. 2022). Categorical models were used to fit the length data for each stock, with the specific approach slightly different in each case due to the sample collection design and explanatory factors for each species.

For a given fishing year, month, area, and haul, the counts of the number of fish measured were generated and arranged as vectors by length bin. As there were often a few either over or under a reasonable size range, the length frequencies were either truncated or aggregated into the first or final bin, respectively. Note that this occurs prior to scaling to obtain the population or survey length composition and the catch weight is not considered during model fitting. However, once the length frequencies have been fitted, the catch is used to rescale the model predicted length composition to obtain the overall length composition.

2.1.1 Red rock lobster in CRA 2

Data used for the red rock lobster fishery in CRA 2 used the information from Roberts et al. (2021) for the years 1993–2020 from either the observer catch sampling or voluntary logbook programmes.

In developing the method to estimate length frequency categorical models from catch sampling and logbook data, we note the generative process for how these data are collected. For the red rock lobster fishery, this is:

a. a fisher/vessel goes out on a trip on a particular day, month, and fishing year;
b. during that trip the fisher lifts one or more pots within a region or statistical area; and
c. the fisher or observer measures and records the length and sex of lobsters within these pots.

From this process, there are three important components to the sampling:

a. the number of trips (days) sampled and the temporal distribution of these trips (i.e., in what months did these trips occur);
b. the number of pots lifted during each trip and the spatial distribution of these pots between regions or statistical areas; and
c. the number of lobsters measured.

To capture the variability in these components and provide reasonable estimates of the resulting length compositions and the associated uncertainty, models should ideally be fitted to data at the level that the data are collected (by event or in this case, by haul (i)). However, initial testing showed that models using the data at the haul level were not feasible, with the number of data points required to be fitted resulting in models that took too long to run. Instead, the data were aggregated to the trip (j) level and the models were fitted to trip level data. The assumption is that the data collected at the trip level are an adequate random sample of the individual pot hauls and that the effect of spatial and temporal patterns do not occur at the scale of an individual trip.

For a given fishing year (y), a trip stratum was defined as the combination of month (m), statistical area (a), data source (catch sampling or logbook) (d), and trip (j). In each trip stratum, counts of the number of lobsters measured by sex category (s) and length bin (l) were generated and arranged as vectors by length bin, i.e.:

$$n_{y,m,a,d,j,l} = \boldsymbol{n}_{y,m,a,d,j} \in [0, \infty]$$

In the notation below, the sex category subscript was dropped because the models were fitted separately for each sex category. There are often a few lobsters either over or under a reasonable size range, so each end of the length frequency was aggregated into the first or final bin, respectively.

Hence, the predictor term (η) for each length bin (l), fishing year (y), month (m), and statistical area (a) was:

Equation 1: $$\eta_{y,m,a,l} = \boldsymbol{\eta}_{y,m,a} = \boldsymbol{X\beta} + \boldsymbol{Z\mu} = \beta_l^0 + X^y\beta_l^y + X^m\beta_l^m + X^a\beta_l^a + Z^{y,a}\mu_l^{y,a}$$

In this equation, $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are the coefficients at the population level (analogous to fixed effects) and group level (analogous to random effects), respectively, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the corresponding design matrices, β_l^0 is the intercept for each length, β_l^y are coefficients accounting for each fishing year, β_l^m are coefficients accounting for each month, β_l^a are coefficients for each statistical area, $\mu_l^{y,a}$ are coefficients for an interaction term by fishing year and statistical area. The fishing year, month, and statistical area terms are treated as population-level effects and the fishing year by statistical area interaction term is treated as a group-level effect.

The model predicts the proportion ($\boldsymbol{p}$) within each tail width (TW) bin (l) by fishing year, month, and statistical area:

Equation 2:

$$\boldsymbol{p}_{y,m,a} = \text{logit}(\boldsymbol{\eta}_{y,m,a}) \text{ where } \boldsymbol{p}_{y,m,a} \in (0,1) \text{ and } \sum \boldsymbol{p}_{y,m,a} = 1 \;\forall y, m, a$$

Counts of individuals within each TW bin at the trip level were fitted using a multinomial distribution:

Equation 3:

$$\boldsymbol{n}_{y,m,a,d,j} \sim \text{multinomial}(n, \boldsymbol{p}_{y,m,a})$$

and the following priors were used:

Equation 4:

$$\begin{aligned}
&\beta_l^0 \sim N(0, 10^2)\\
&\beta_l^y \sim N(0, 1^2)\\
&\beta_l^m \sim N(0, 1^2)\\
&\beta_l^a \sim N(0, 1^2)\\
&\mu_l^{y,a} \sim N\left(0, \left(\sigma_l^{y,a}\right)^2\right)\\
&\sigma_l^{y,a} \sim \text{Student}(3, 0, 2.5)
\end{aligned}$$

Initial model fits were done using variational inference, which provides an approximation of the posterior distribution. The approximate posterior distributions were used to explore different model structures and ensure that the model was not severely ill-conditioned. Full Bayesian inference was then done using the Hamiltonian Monte Carlo (HMC) algorithm. Initially, MCMC mixing was poor, but adequate MCMC mixing within a reasonable time frame was achieved by:

a. removing early model years with too few data;
b. using QR[4] decomposition of the data;
c. applying priors (Equation 4); and
d. specifying initial values for all model parameters (using the median obtained using variational inference).

2.1.2 Hake in HAK 7 and ling in LIN 3&4

Length data were available for hake in HAK 7 and ling in LIN 3&4 by fishing year (y), area (i.e., sub-region or stratum, denoted by m), trip (i), and haul (j), for all fish measured by sex (s) and length (cm). For both case studies, data from the bottom trawl commercial fleet and the *Tangaroa* resource surveys were used. In developing the method for estimating length composition categorical models for hake and ling, we note the generative process for how these data are collected. For a given time (t), in a location (s), and haul (i), counts of the number of fish measured by length bin (l) are sampled from all fish caught in that haul, this can be represented as:

[4] In linear algebra, a QR decomposition, also known as a QR factorisation or QU factorisation, is a decomposition of a matrix A into a product $A = QR$ of an orthogonal matrix Q and an upper triangular matrix R.

$$n_{t,s,i,l} \sim \phi_\pi(\pi_{t,s,i,l})$$

where $n_{t,s,i,l}$ is the number of sampled fish, ϕ_π is an error distribution describing each sample from a haul, and $\pi_{t,s,i,l}$ is the true number of fish in that haul. The true number of fish in a haul is drawn from the available population:

$$\pi_{t,s,i,l} \sim \phi_\mu(\lambda_{t,s,l})$$

where ϕ_μ is an error distribution describing each haul from the available population and $\lambda_{t,s,l}$ is the true number of fish by length bin (l) within the available population at any given time (t) and location (s). By combining the two error structures, these counts could be fitted using a multinomial distribution:

$$\boldsymbol{n}_{t,s,i} \sim \text{multinomial}(\boldsymbol{\lambda}_{t,s})$$

where the expected value of each length bin (l) can be derived using linear regression:

$$E(\boldsymbol{Y}|\boldsymbol{X}) = \ln\left(\frac{\mu}{1-\mu}\right) = g^{-1}(\boldsymbol{X\beta}) = \beta_l^0 + X^t\beta_l^t + X^s\beta_l^s + \cdots$$

where β_l^0 is the intercept for each length, β_l^t are coefficients accounting for time, β_l^s are coefficients accounting for space and can be extended to represent other relevant explanatory variables (e.g., vessel, sampling programme, etc.) as required.

These data were arranged as length frequencies for a given fishing year (y), area (m), and trip strata (i), i.e., all fish measured by sex (s) and length bin (l) were arranged as vectors using 1 cm length bins (l) such that:

$$n_{y,m,i,j,l} = \boldsymbol{n}_{y,m,i,j} \in [0,\infty].$$

Note that the sex category subscript was dropped in this notation because the sexes were stacked (i.e., the proportions-at-sex and -length were modelled as a single vector so that the sum of the proportions across sexes was one). As for red lobster, there were a few fish under or over length ranges, so each end of the length frequency was either aggregated into a minus or plus group as the first or final bin, or the data were omitted.

For both hake and ling, the number of observations at the haul level were sparse (small numbers of fish measured in each haul). Initial testing showed that models using the data at the sampling event level were not feasible, with the number of data points required to be fitted resulting in models that took too long to run. Instead, the data were aggregated to a 'trip stratum' (j) where data were aggregated to year, month, and strata, within each trip. This assumed that the data collected at the trip stratum level were an adequate random sample of the individual hauls and that the effect of spatial and temporal patterns do not occur at the scale of an individual trip stratum. The strata used for each stock (hake and ling) was based on the strata from previous analyses. For hake these were the three strata: deep, north shallow, and south shallow from Horn (2011). For ling, the four strata were: scampi (*Metanephrops challengeri*) target tows, coast (longitude ≤ 174°E), north rise (latitude < 43.55°S, longitude > 174°E), and south rise (latitude ≥ 43.55°S, longitude > 174°E) from Horn (2005).

For hake, data from all observed hauls from the bottom trawl commercial fishery were used for years 1991–2020 (see Saunders et al. 2021, Dunn et al. 2023).

In addition, all measured fish under 25 cm were omitted as these were very infrequent. Females over 130 cm were aggregated into the 130 cm length bin, and males over 120 cm were aggregated into the 120 cm length bin as few fish of these lengths were observed. Where there were no data for a given length bin, that length bin was assigned a zero. Data were limited to the West Coast of the South Island (WCSI) during the period June to September (see Dunn et al. 2023) with the age data from Saunders et al. (2021).

For ling, data from all observed hauls from the bottom trawl commercial fishery were used for the years 1990–2019 (Saunders et al. 2021, Mormede et al. 2022). In addition, all measured fish under 29 cm were omitted as these were very infrequent. Both males and females over 165 cm were aggregated into the 165 cm length bin as few fish of these lengths were observed. Where there were no data for a given length bin, that length bin was assigned a zero. Data were limited to the Chatham Rise in LIN 3&4 (Mormede et al. 2022) with the age and length data for LIN 3&4 as reported by Saunders et al. (2021).

For both models, counts of individuals within each length bin at the trip level were fitted using a multinomial regression model:

Equation 5:
$$\boldsymbol{Y}_{y,m,i} | x_i \sim \text{multinomial}(\boldsymbol{\mu}_{y,m,i})$$

with expectation vector $\boldsymbol{\mu}_i = \left(\mu_{1,i}, \ldots, \mu_{C,i}\right)^T$. Using component-wise coefficient vectors β_C for components $c = 1, \ldots, C$, the linear predictor $\eta_{c,i}$ is used to define $\mu_{c,i}$ as:

Equation 6:
$$\eta_{c,i} = x_i^T \beta_c$$

where for one reference component $\tilde{c}$, all elements of β are equal to 0 ($\beta_{\tilde{c}} = 0$) to guarantee identifiability as a consequence of the degrees of freedom reduced to $C - 1$.

The predictor term (η) for each length bin (l) during fishing year (y), area (m), and trip (i) was:

Equation 7:
$$\eta_{y,m,i,j,l} = \boldsymbol{\eta}_{y,m,i,j} = \boldsymbol{X\beta} + \boldsymbol{Z\mu} = \beta_l^0 + X^y \beta_l^y + X^m \beta_l^m + Z^i \mu_l^i + Z^{i,j} \mu_l^{i,j}.$$

As for red rock lobster above, similar equations were developed for the length composition data. Here, $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ are the coefficients at the population level (analogous to fixed-effects) and group level (analogous to random-effects), respectively, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are the corresponding design matrices, β_l^0 is the intercept for each length, β_l^y are coefficients accounting for each fishing year, β_l^m are coefficients accounting for each area, and $\mu_l^{i,j}$ are coefficients for an interaction term by trip and haul. The fishing year and stratum terms are treated as population-level effects and the trip and haul terms are treated as group-level effects.

The model then predicted the proportion within each length bin by fishing year, area, trip, and haul:

Equation 8:
$$\boldsymbol{p}_{y,m,i,j} = \text{logit}\left(\boldsymbol{\eta}_{y,m,i,j}\right) \text{ where } \boldsymbol{p}_{y,m,i,j} \in (0,1) \text{ and } \sum \boldsymbol{p}_{y,m,i,j} = 1 \ \forall y, m, i, j$$

and the following priors were used:

Equation 9:

$$
\begin{aligned}
&\beta_l^0 \sim N(0, 10^2)\\
&\beta_l^y \sim N(0, 1^2)\\
&\beta_l^m \sim N(0, 1^2)\\
&\mu_l^i \sim N\left(0, \left(\sigma_l^i\right)^2\right)\\
&\mu_l^{i,j} \sim N\left(0, \left(\sigma_l^{i,j}\right)^2\right)\\
&\sigma_l^i \sim \text{Student}(3, 0, 2.5)\\
&\sigma_l^{i,j} \sim \text{Student}(3, 0, 2.5)
\end{aligned}
$$

Initial model fits were done using variational inference, which provided an approximation of the posterior distribution, to explore different model structures and ensure that models were not severely ill-conditioned.

2.2 Age conditional on length models

For hake and ling, ordinal models were used to predict the distribution-at-age for a given length using aged otoliths. Age data for red rock lobster were not available and hence the age composition was not estimated.

The response variable was denoted y, a density function as f, and μ was used to refer to the main model parameter, which is usually the mean of the response distribution or some closely related quantity. In a regression framework, μ is not estimated directly but computed as $\mu = g(\eta)$, where η is a predictor term and g is the response function (inverse of the link function).

For ordinal models, y is one of the categories $1, \dots, K$. The intercepts of ordinal models are called thresholds and are denoted as τ_k, with $k \in \{1, \dots, K-1\}$, whereas η does not contain a fixed effects intercept. Note that the applied link functions g are technically distribution functions $\mathbb{R} \to [0, 1]$ and usually the standard normal distribution is used. The density of the *cumulative* family (implementing the most basic ordinal model) is given by

$$f(y) = g(\tau_{y+1} - \eta) - g(\tau_{y+1} - \eta).$$

The densities of the stopping ratio (*sratio*) and continuation ratio (*cratio*) families are given by

$$f(y) = g(\tau_{y+1} - \eta) \prod_{k=1}^{y} \left(1 - g(\tau_y - \eta)\right),$$

and

$$f(y) = \left(1 - g(\eta - \tau_{y+1})\right) \prod_{k=1}^{y} (\eta - \tau_k),$$

respectively. Both families are equivalent for symmetric link functions such as logit or probit. The *cumulative* and *sratio* models use $\tau - \eta$, whereas *cratio* uses $\eta - \tau$. This is done to ensure that larger values of η increase the probability of higher response categories. The linear predictor η can be generalised to also depend on the category k for a subset of predictors. This then leads to category specific effects.

It is not clear what specific method previous studies have used (Rindorf & Lewy 2001, Berg & Kristensen 2012, Babyn et al. 2021), but we note that *cratio* may not be the optimal choice as each age is not linked across years due to them belonging to different year classes. We suggest the *cumulative* may be the best choice. Initial testing suggested that the results from the different approaches were essentially the same, but that model convergence was more likely in the *cumulative* model for these case studies.

Separate models were fitted for females and males. The dependent variable in all models was age treated as an ordered categorical variable. Models were explored with k threshold values (τ_k), or with k different thresholds for each fishing year ($\tau_{k,t}$). In *brms,* this is specified using:

age | threshold(gr = fyear) ~ model terms.

Some ages were missing in some years (see Table 3.1 and Table A3.2). This meant that different thresholds could not be estimated each fishing year because *brms* requires that either the same estimated thresholds are used for all years or the same number of thresholds be estimated for all years. Three options are available to solve this problem:

1. fit separate models to data for each fishing year (this will use different thresholds for all years);
2. fit a single model to all data and use the same thresholds for all years; or
3. fit a single model to all data and use different thresholds for all years, add a single fish for each missing age/year, and impute the length for each individual that was added.

Option 1 is analogous to the conventional way of deriving age-length keys. There are two problems with this approach. First, if an age is missing from a fishing year, age five for example, then prediction of an age composition from a length composition will result in no fish of age five in the age frequency (the lengths will be assigned to age four, age six, and all of the other ages but never age five). Second, the ability to predict for years for which data are limited is lost.

Option 2 resolves both of the problems outlined in Option 1. However, if the thresholds are not allowed to vary by fishing year, and the age frequencies differ substantially by fishing year, then there may not be enough degrees of freedom to be able to adequately fit the model to the data.

Option 3 attempts to solve all of the above problems. However, care must be taken that the addition of imputed data does not lead to unintended consequences. A suitable method for imputing the length for missing ages is required, and here we impute the length for each missing age record by fitting a von Bertalanffy growth curve to the data for each sex and using the mean value from these models to predict length given the missing age.

All three options were tested in this report. Note that, following the preparation of this report, a solution for this issue was identified in the Stan forums (Koblinger 2023) by including an observation with zero weight for missing ages in the data when using *brms*. In *brms,* this is specified using:

age | weights(wt) ~ model terms

The following explanatory variables were considered:

- length – the length (cm) of each individual fish for which an otolith was collected. This is the continuous length (e.g., not necessarily rounded to the nearest integer);
- fyear – the fishing year that the otolith was collected;
- cyear – a continuous version of fyear;
- month – the month of the year that the otolith was collected;
- origin – the origin of the otolith (fleet or method of otolith collection, for example from the fisheries observers or resource surveys);
- stratum – the stratum used to scale up length samples. For hake these were the three strata: deep, north shallow, and south shallow from Horn (2011). For ling, these were the four strata: scampi target tows, coast (longitude ≤ 174°E), north rise (latitude < 43.55°S, longitude > 174°E), and south rise (latitude ≥ 43.55°S, longitude > 174°E) from Horn (2011);
- x – the longitude of the location of the haul where the otolith was collected; and
- y – the latitude of the location of the haul where the otolith was collected.

These variables were included as either 1- or 2-dimensional splines, or as factors. For example:

- s(length, k = 3) represents a 1-dimensional spline for length using a basis dimension of 3;
- t2(length, cyear, k = c(6, *nYear*)) represents a 2-dimensional tensor product smooth (Bürkner & Vuorre 2019) by length using a basis dimension of 6, and by fishing year (*nYear*, as a continuous variable) using a basis dimension for the number of fishing years; and
- month represents a population-level effect.

The model that would be most similar to the method that applies an ALK can be defined as:

age | threshold(gr = fyear) ~ s(length, by = fyear)

and then extended to allow for spatial-temporal effects with the addition of time (i.e., month) and spatial terms along with interactions if required. The use of ordinal regression models using *brms* is described by Bürkner & Vuorre (2019) where further details of the *brms* implementation of ordinal models can be found.

2.3 Estimating the scaled length compositions for red rock lobster, hake, and ling

The overall length compositions (by year and sex) can be determined by scaling the predicted length frequency, derived from the models developed in Section 2.1, to the total catch for red rock lobster, hake, and ling.

The catch (tonnes) for each fishing year, month, and area was estimated from catch and effort data for the respective case studies (red rock lobster, hake, and ling; Roberts et al. 2021, Mormede et al. 2022, Dunn et al. 2023) and denoted $C_{y,m,a}$.

In each of the case studies, the fitted models above were used to generate the predicted length frequency by year (y) in each trip stratum for each sex, length bin, and sample from the posterior (i):

Equation 10:
$$\boldsymbol{p}_{y,m,a,i} \sim \text{multinomial}\left(1, \boldsymbol{\mu}_{y,m,a}\right) \text{ where } \sum \boldsymbol{p}_{y,m,a,i} = 1 \ \forall y, m, a, i.$$

The expected weight (in tonnes) of a fish at each length was calculated as:

Equation 11:
$$w_{s,l} = \alpha_s L_s^{\beta_s}$$

where α_s and β_s are parameters of the length-weight relationship for each sex. The mean weight associated with the predicted length frequency for each year, month, area, and iteration was calculated as:

Equation 12:
$$\tilde{w}_{y,m,a,i} = \frac{1}{n}\sum_{s,l} \boldsymbol{p}_{y,m,a,i}\boldsymbol{w}.$$

And the expected number of fish was calculated as:

Equation 13:
$$\tilde{n}_{y,m,a,i} = \frac{C_{y,m,a}}{\tilde{w}_{y,m,a,i}}.$$

Finally, the scaled length compositions were derived using:

Equation 14:
$$\boldsymbol{P}_{y,m,a,i} = \tilde{n}_{y,m,a,i}\boldsymbol{p}_{y,m,a,i}.$$

The mean and SD by fishing year, sex, and length bin were calculated from the posterior distribution and then aggregated by posterior sample (i), fishing year (y), sex (s), and length bin (l) then converted to proportions:

Equation 15:
$$P_{y,s,l} = \boldsymbol{P}_{y,s} \text{ and } \sigma_{y,s,l} = \boldsymbol{\sigma}_{y,s}$$

The SD can be used to determine the relative weighting of the data set in stock assessment.

2.4 Estimating the scaled age compositions using the scaled length composition and the age-at-length model

The predicted scaled age composition (Y_a) is the product of the matrix from the ordinal age-at-length model ($A_{a,l}$) and the scaled predicted length composition vector (P_l) determined above:

Equation 16:
$$\boldsymbol{Y} = Y_a = A_{a,l}P_l = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_n \end{bmatrix}.$$

This gives the resulting scaled age composition, equivalent to the output of the empirical ALK method and the estimated age composition. The estimated sampling error from both methods can therefore be directly compared. The predicted scaled age composition takes on the dimensions of the matrix from the ordinal age-at-length model and the scaled predicted length composition vector; for example, if the age-at-length model was stratified by year (y) and area (a) and the scaled length composition was stratified by year (y) and month (m) then the resulting scaled age composition would be $\boldsymbol{Y}_{y,m,a}$.

2.5 Empirical estimation of the scaled age compositions using ALKs

For both hake in HAK 7 and ling in LIN 3&4, empirical scaled age frequencies were estimated by scaling observed sex and length frequencies in each tow to the catch from that tow, then aggregating over all tows and scaling to a stratum catch using empirical methods and uncertainty derived from bootstrapping (see Bull & Dunn 2002 for methods). The total aggregated length frequencies were calculated by summing over strata to generate a scaled length composition. Scaled age compositions were then calculated by multiplying the scaled length composition with an annual sex specific age-length key. Effective sample sizes for each of the annual age compositions (n) were then derived using a non-linear regression of $\log(\mathrm{CV}_i) \sim \log(p_i)$, where the coefficient of variation (CV) was calculated for each proportion (p_i) from 300 bootstraps. Specific methods for the calculations for hake are given by Dunn et al. (2023) and for ling by Mormede et al. (2022) and Saunders et al. (2021).

3. RESULTS

3.1 Length composition models

3.1.1 Red rock lobster in CRA 2

Length composition models were estimated for red rock lobster in CRA 2. The first three fishing years (1986, 1990, and 1991) were dropped from the CRA 2 model because there were too few vessels, days, trips, pots, and lobsters sampled by the catch sampling programme and no logbook samples for these years (Table A1.1 in Appendix 1) and, during these years, there were no samples in rock lobster statistical areas 905 and 906 (Table A1.2) and only the spring/summer (SS) season was sampled.

In CRA 2, there were few male and immature female lobsters under 36 mm tail width (TW) measured and, therefore, all lobsters from 30 to 36 mm TW were added to the [36, 38) length bin. There were few mature females under 42 mm TW, so all mature females from 30 to 42 mm were added to the [42, 44) bin. Likewise, there were few immature females over 70 mm TW, so all immature females greater than 70 mm TW were added to the [68, 70) bin. In the past, dropping fishing years and aggregating TW bins was usually done after the empirical length frequencies (LFs) were produced but before they were inputted into the stock assessment (Starr et al. 2021). However, initial model runs that included all years and attempted to fit to all TW bins took a long time to run and resulted in inadequate MCMC mixing (not shown). When these years and TW bins were dropped, MCMC mixing was adequate for all model runs and the run time was reasonable (Appendix 2).

MCMC mixing for the CRA 2 length composition models was acceptable for all three sex categories (male, immature female, and mature female; Figure A1.1, Figure A1.2, and Figure A1.3).

For all three sex categories, the model intercept parameters increased from low values for the smallest TW bins, peaked near the minimum legal size (MLS), then began to decline (Figure A1.4). The fishing year population level effects (fixed effects) for each TW bin varied by year, with some obvious correlation among TW bins (Figure A1.5). For example, the fishing year coefficients were very similar for the 75- and 77-mm TW bins for males (Figure A1.5). The population-level effects for month suggested some strong seasonal differences (Figure A1.6). The population-level effects for statistical area were different which was not surprising, given that these statistical areas are geographically distinct (Figure A1.7).

The conditional effect of a particular explanatory variable within a model is the effect of that variable conditioned on the remaining explanatory variables fixed at a reference value (the reference value used for all models was the parameter associated with the smallest TW bin, the first fishing year, April, and statistical area 905). Conditional effects for fishing year illustrate how the underlying LFs (i.e., the mean across all months and statistical areas) have changed over time (Figure A1.8). The conditional effects for month were surprisingly similar (Figure A1.9). The conditional effects for statistical area in CRA 2 were unusual in that non-adjacent areas were more similar (i.e., statistical areas 906 and 908 were similar to each other, and statistical areas 905 and 907 were similar to each other; Figure A1.10).

3.1.2 Hake in HAK 7 and ling in LIN 3&4

Length composition models were estimated for hake in HAK 7 and ling in LIN 3&4 using the same methods. The first fishing year (1990) was dropped from the hake model because the initial sampling indicated a pattern that caused incomplete mixing in the MCMC and a number of divergent transitions. This suggests that the MCMC algorithm would need to be tweaked to improve the diagnostics if this model were to be applied to a real-world application. Numbers of samples by year for each of hake and ling are given in Table A1.3.

For hake, there were few male and female hake under 25 cm and these were excluded from the analysis. Likewise, there were few males over 120 cm and females over 130 cm, so these were aggregated into a plus group at 120 and 130 cm, respectively. Similarly for ling, there were few male and female ling under 29 cm and these were excluded from the analysis, and there were few males over 150 cm and females over 165 cm, so these were aggregated into a plus group at 150 and 165 cm, respectively. The amount of data outside the ranges specified for hake and ling was very small but consideration of the influence of these data on the analysis is required; however, this consideration also applies to when an empirical ALK is used.

MCMC mixing in both the hake (Figure A1.11) and ling (Figure A1.12) models was poor. Even when the 1990 year was dropped from the hake model, the MCMCs had poor mixing and a number of divergent transitions, suggesting poor convergence. Further development is required to resolve these issues. While aggregation of the data to the trip stratum level was required to enable the models to run, the effect of spatial and temporal patterns at smaller scales would need to be evaluated for cases when aggregation to trip stratum results in an adequate estimate.

3.2 Estimates of the scaled length compositions

3.2.1 Red rock lobster in CRA 2

Scaled length compositions were calculated using the length composition categorical models, by scaling the samples up to the total catch, as described above in Section 2.3. Resulting scaled length frequencies for rock lobster are shown in Figure A2.10.

Model fits to these data are not presented at the level to which they were fitted (by trip). Instead, for each Quota Management Area (QMA) the data were aggregated to the two levels described below with model predictions made at these aggregated levels:

1. By fishing year, month, data source, and length bin separately for each statistical area and sex category

Equation 17:

$$\boldsymbol{n}_{y,m,a,d,s} = \sum_{j} \boldsymbol{n}_{y,m,a,d,j,s}$$

2. By fishing year, season, and length bin separately for each region and sex category

Equation 18:

$$\boldsymbol{n}_{r,y,t,s} \sum_{a \in r, m \in t, d, j} \boldsymbol{n}_{y,m,a,d,j,s}$$

3.3 Age conditional on length models

3.3.1 Hake in HAK 7 and ling in LIN 3&4

For hake, length for missing ages was imputed using the von Bertalanffy growth model (von Bertalanffy 1938) for each sex to predict length given the missing age. The von Bertalanffy model fits were inspected visually and deemed acceptable for hake (Figure A3.1, Figure A3.2. The number of missing ages imputed was substantial in some years, specifically for those older ages less frequently observed. The numbers of ages for each length for hake are given in Table 3.1 (females) and Table A3.2 (males).

For hake, the suite of model runs is presented for females (Table 1) and males (Table 2). The run time of these models ranged from about 20 minutes to over five hours, with the larger number of basis dimensions for the length splines having the greatest effect on run time. None of the model runs had any divergent transitions. The worst performing model for females (HF19) and males (HM19) included length as a linear term. All other model runs included length as a spline with a range of different basis dimensions (3–20). Splines with higher basis dimensions produced models that performed better according to LOO IC (leave-one-out information criteria); however, Bayesian R^2 suggested that the improvements to the deviance explained were minimal with higher basis dimensions (for females, models with the same structure but a basis dimension of 3, 6, 9, and 20 had improved LOO ICs but the mean Bayesian R^2 was similar for these models with, for example, female values of 0.621, 0.621, 0.622, and 0.623, respectively).

Models that included thresholds that varied by fishing year performed better, according to LOO IC and Bayesian R^2, than those without. Models without thresholds that differed by fishing year (i.e., HF18 and HM18) had relatively poor model fits compared to models with a similar structure that did include year varying thresholds (e.g., HF7 and HF11, HM7 and HM11, Figure 1, Figure 2). It is clear that the thresholds are important in allowing a better model fit and the thresholds changed substantially among years (Figure 3, Figure 4).

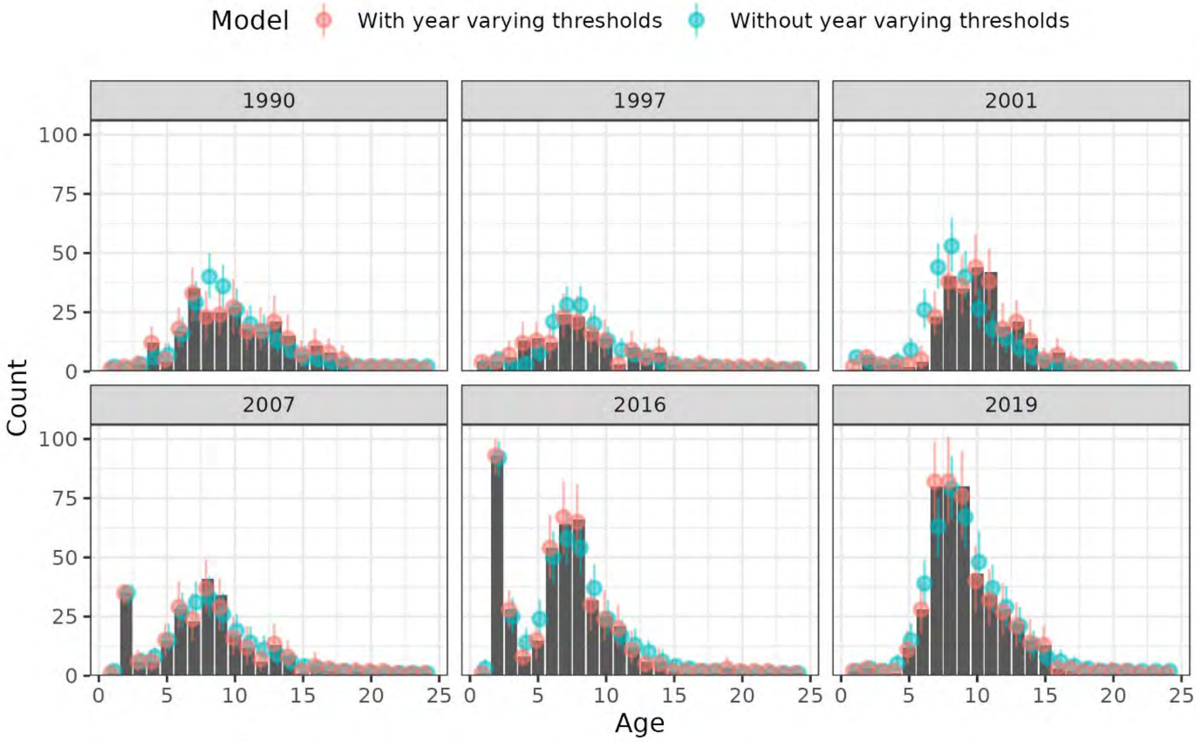

Figure 1: The distribution of the data (bars) and the posterior predictive distribution (median as points and 90% credible interval as lines) for the female hake models with (HF11) and without (HF18) year varying thresholds (see Table 1).

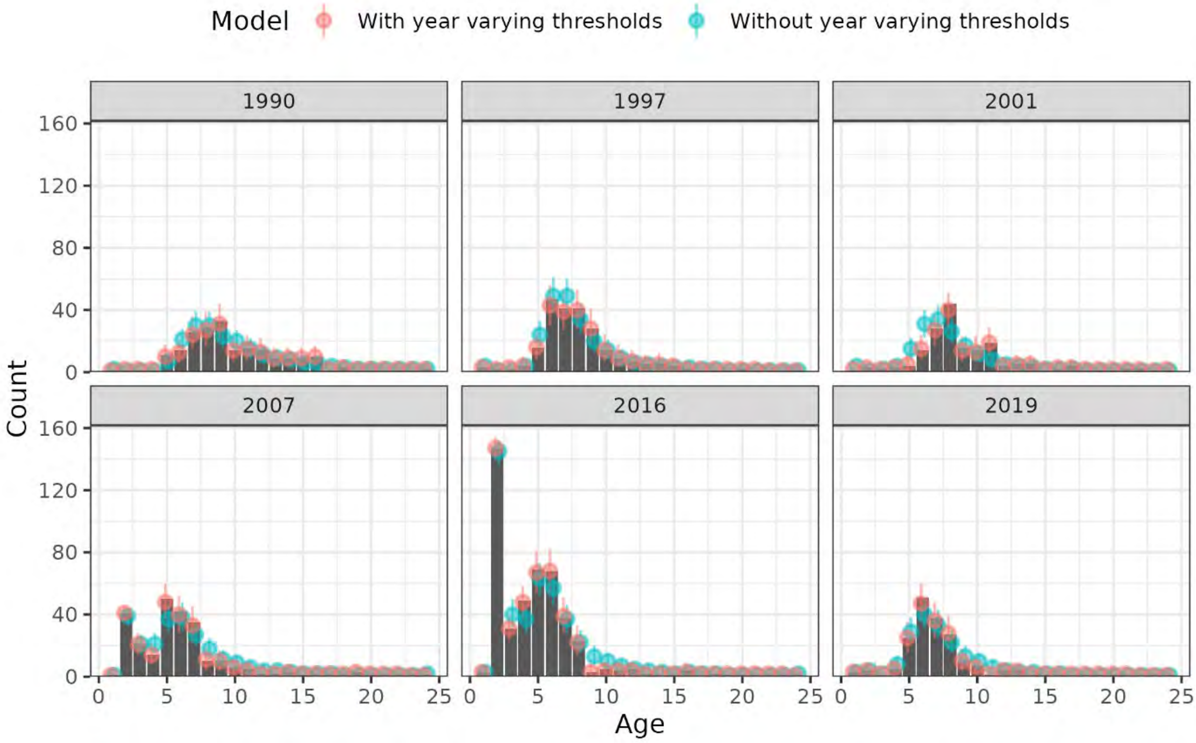

Figure 2: The distribution of the data (bars) and the posterior predictive distribution (median as points and 90% credible interval as lines) including for the male hake models with (HF11) and without (HF18) year varying thresholds (see Table 2).

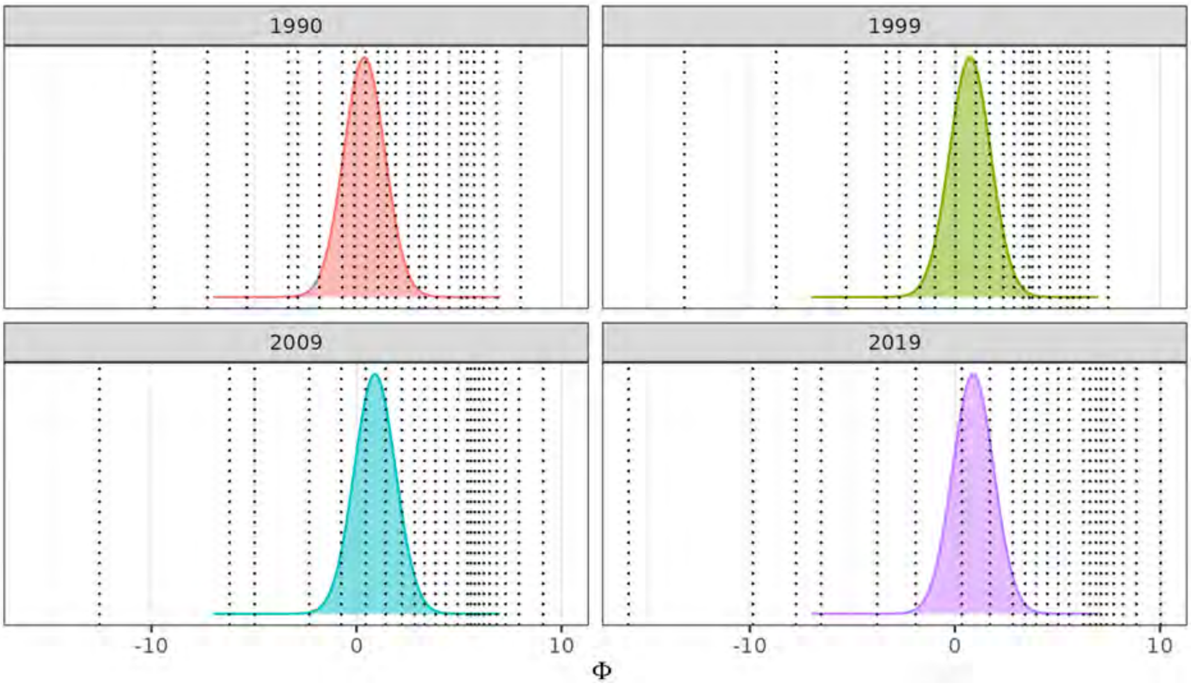

Figure 3: Distribution of predicted age (for average length cm) and thresholds for the selected female hake model (HF11) for a subset of years (see Table 1).

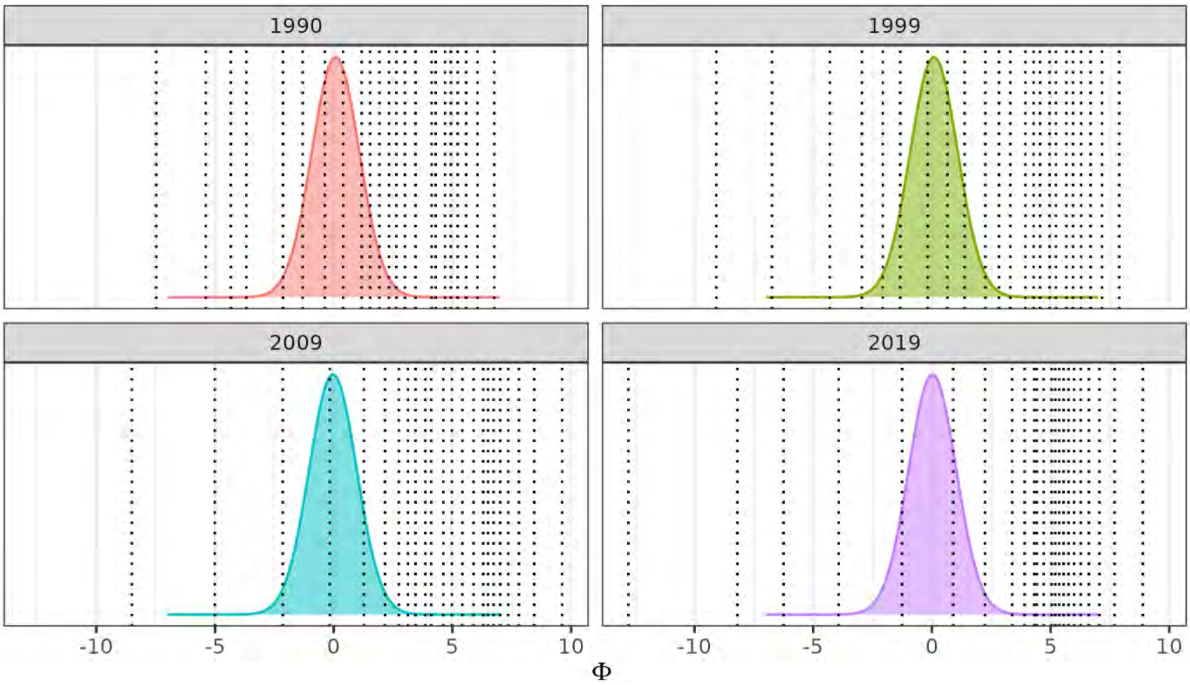

Figure 4: Distribution of predicted age (for average length cm) and thresholds for the selected male hake model (HM11) for a subset of years (see Table 2).

The two best model runs (HF1 and HM1), according to LOO IC and Bayesian R^2, both included a 2-dimensional tensor product smooth by length using a basis dimension of 6 and by fishing year (continuous) using a basis dimension of 26. The model run for males that had the best LOO IC had a slightly lower Bayesian R^2 when compared to the model for males with the second-best LOO IC (compare HM1 and HM2). However, models with 2-dimensional tensor product smooths by length and fishing year appeared to degrade fits to the data in some years. For example, compare HF4 and HF7 (Figure 5, Figure 6) or HM5 with HM7 (Figure 7, Figure 8).

Conditional effects plots were used to explore the distribution of lengths estimated for each age. When the length term was assumed to be linear (HF19 and HM19) the distribution of lengths for each age approximately followed a normal distribution (Figure 9, Figure 10). However, when a smooth term was used instead (e.g., HF4 and HM5) the distribution of lengths for each age was more variable and had more complex shapes for each age (Figure 11, Figure 12). The inclusion of a spatial smooth term provided a small improvement in model fit (e.g., models HF2 and HM2, see Figure A3.5, Figure A3.10), but the effect was small and we hence ignore the spatial terms in the development of the approach in this report. In practise, evaluation of the improvement of additional explanatory terms would be recommended to identify the importance of such covariates on model outcomes.

Despite these issues, HF4 and HM5 were selected as the two models to proceed with because the MCMC mixing for the female and male models was excellent (Figure A3.3, Figure A3.4) and the inclusion of space and month did not improve model fits substantially.

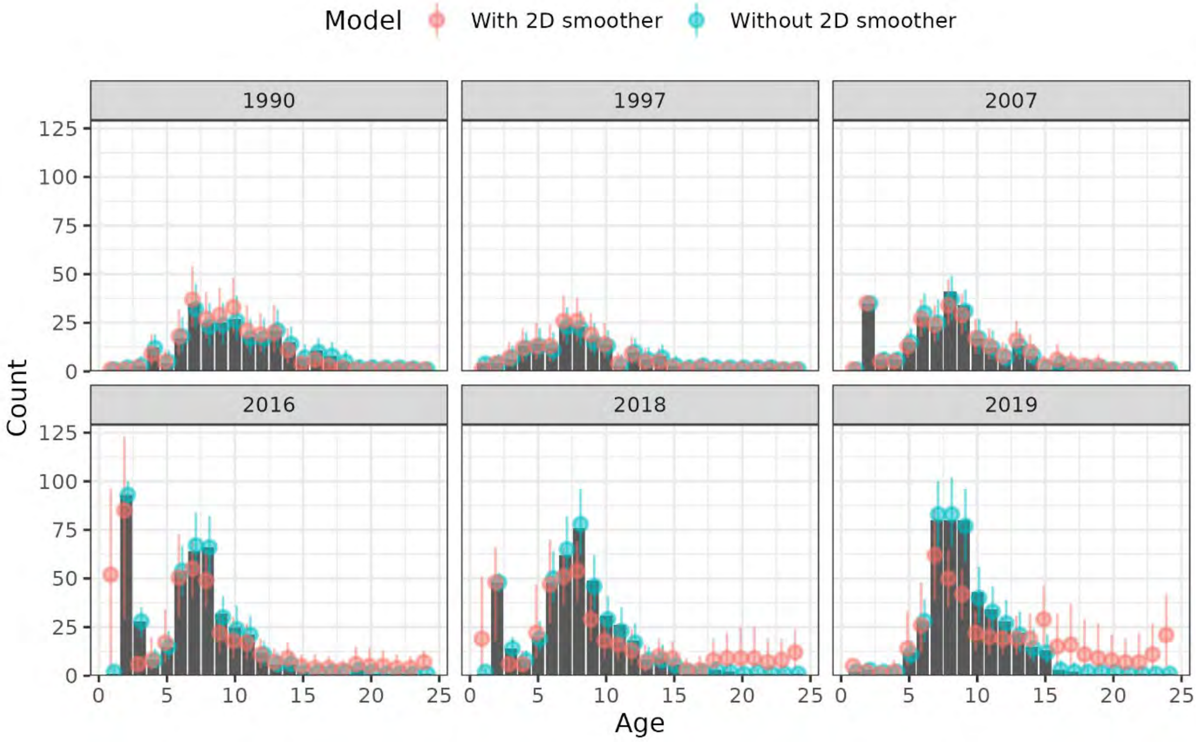

Figure 5: The distribution of the data (bars) and the posterior predictive distribution (median as points and 90% credible interval as lines) for the female hake models with (HF4, red) and without (HF7, blue) 2-dimensional tensor product smooths (see Table 1).

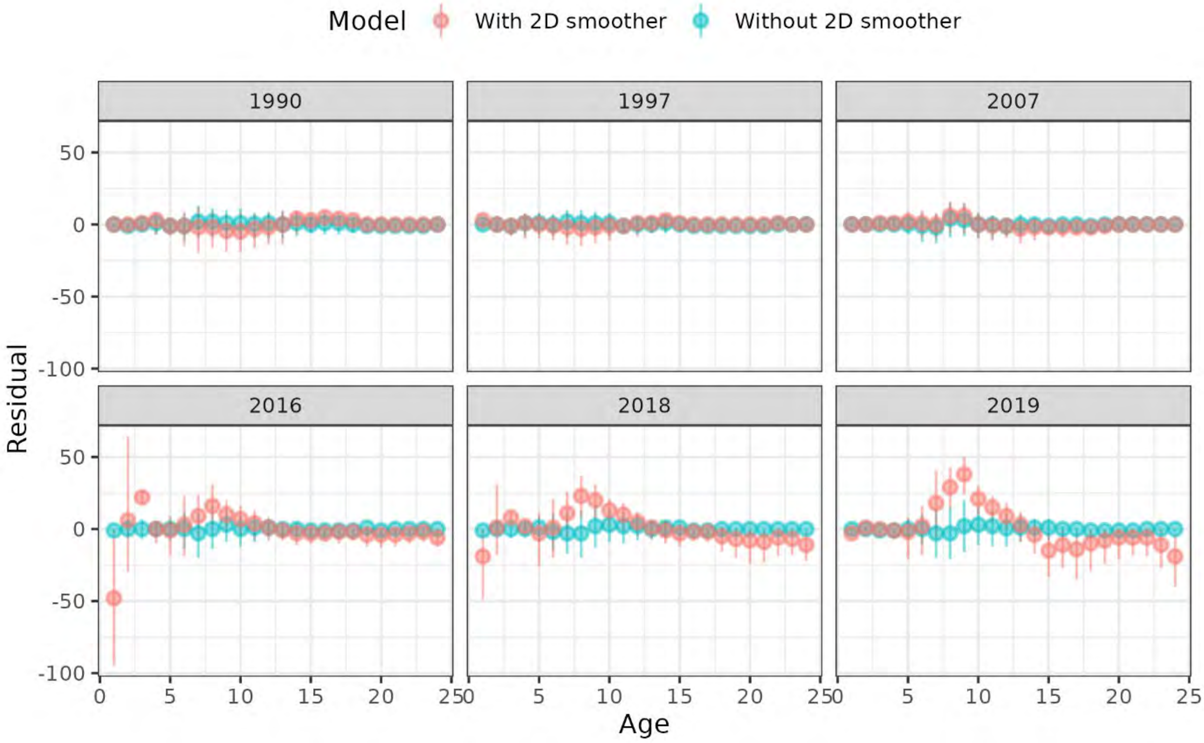

Figure 6: Residuals (median as points and 90% credible interval as lines) for the female hake models with (HF4, red) and without (HF7, blue) 2-dimensional tensor product smooths (see Table 1).

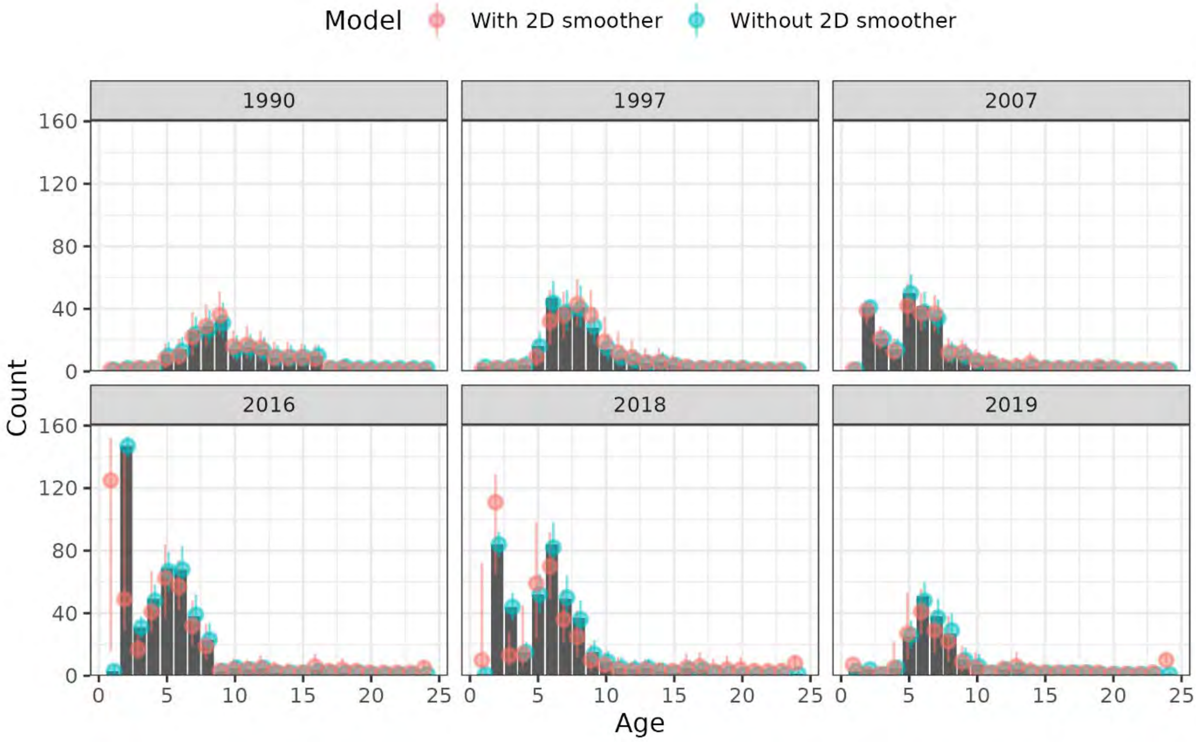

Figure 7: The distribution of the data (bars) and the posterior predictive distribution (median as points and 90% credible interval as lines) for the male hake models with (HM5, red) and without (HM7, blue) 2-dimensional tensor product smooths (see Table 2).

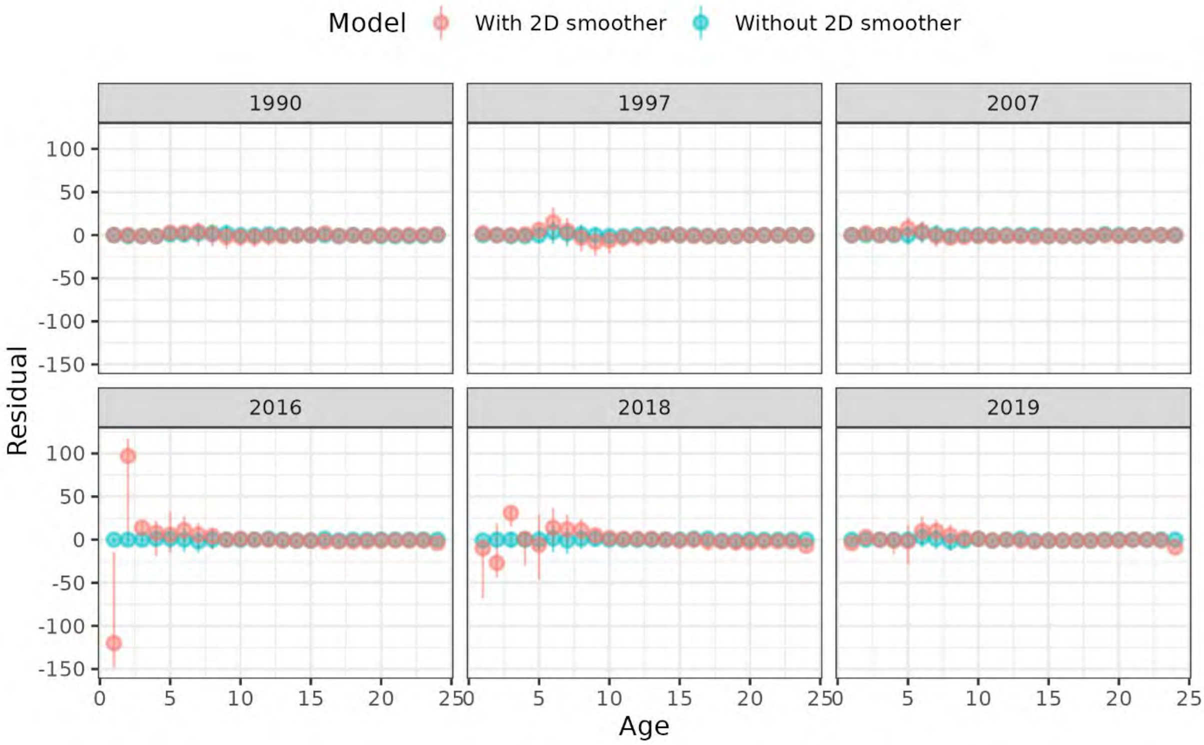

Figure 8: Residuals (median as points and 90% credible interval as lines) for the male hake models with (HM5, red) and without (HM7, blue) 2-dimensional tensor product smooths (see Table 2).

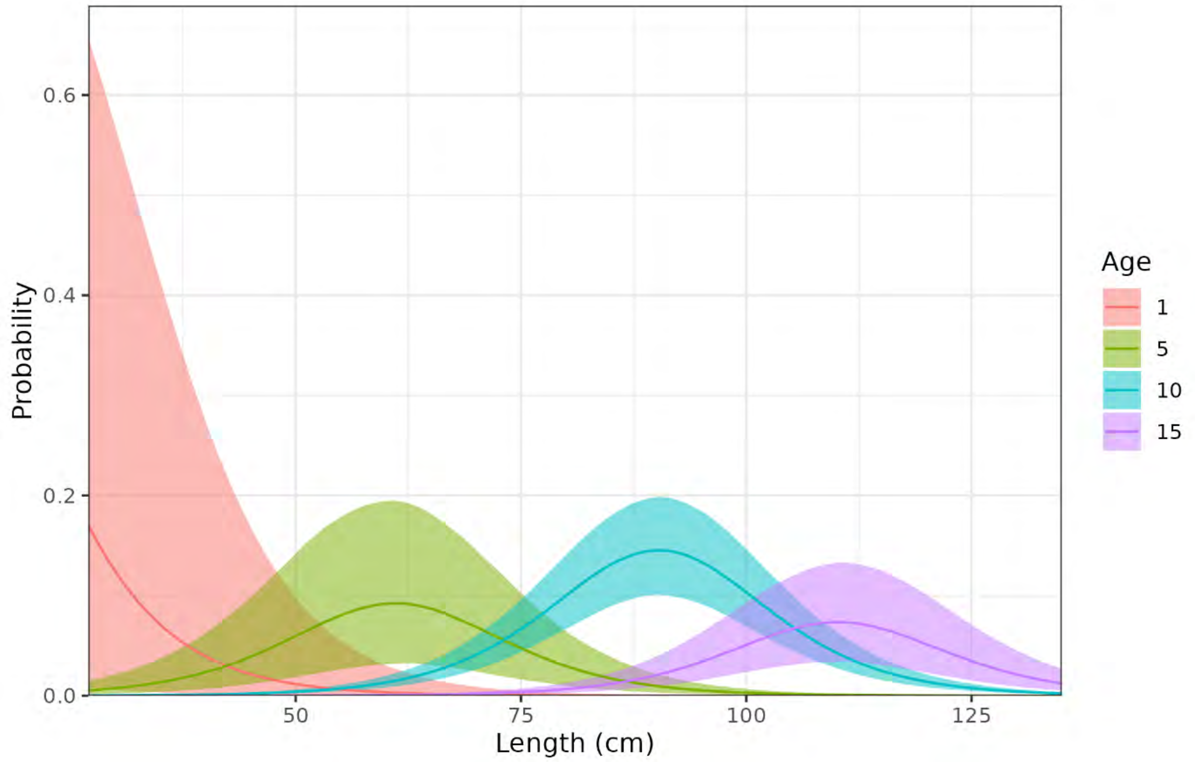

Figure 9: Conditional effect of length for a subset of ages for HAK females for the selected model (HF19) (see Table 1).

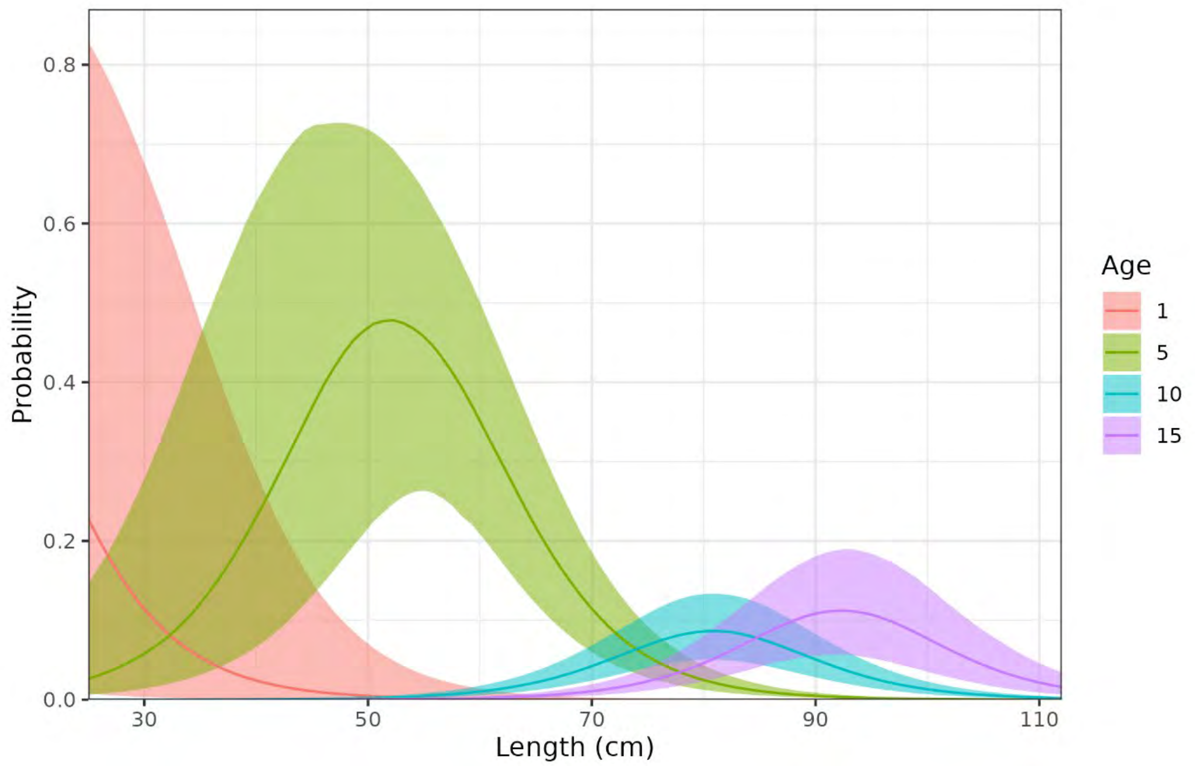

Figure 10: Conditional effect of length for a subset of ages for HAK males for the selected model (HM19) (see Table 2).

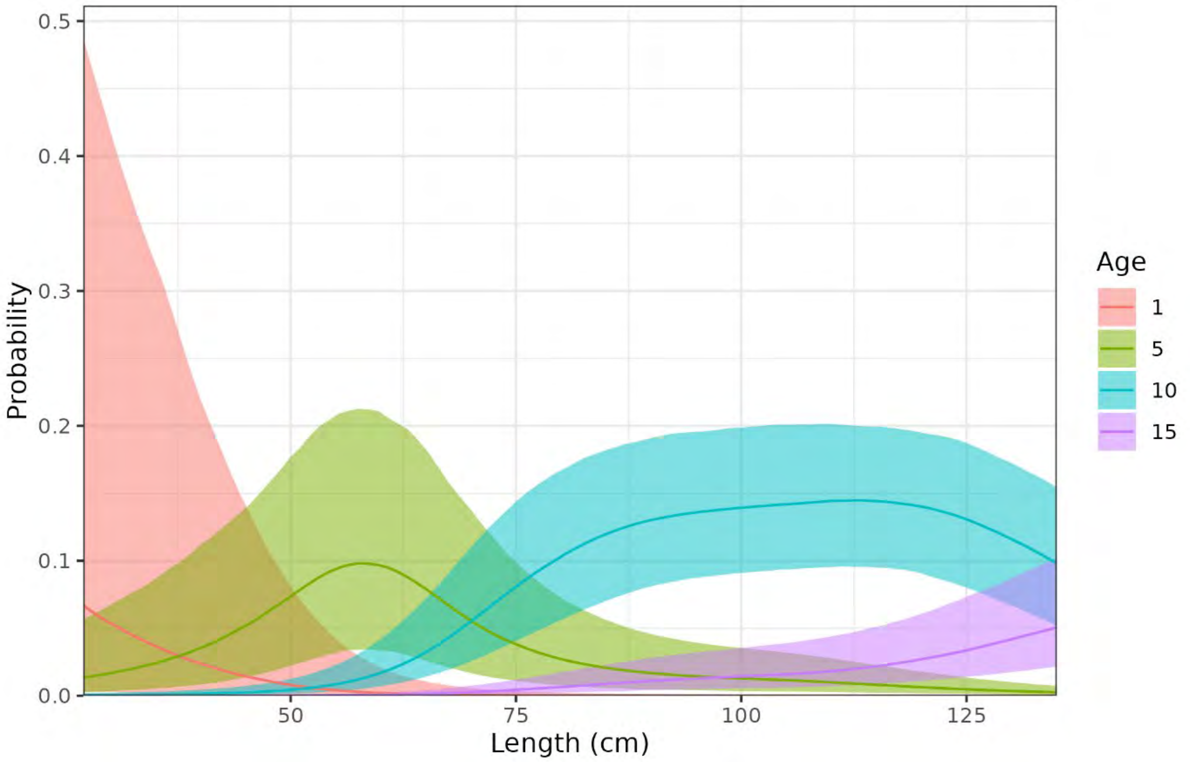

Figure 11: Conditional effect of length for a subset of ages for HAK females for the selected model (HF4) (see Table 1).

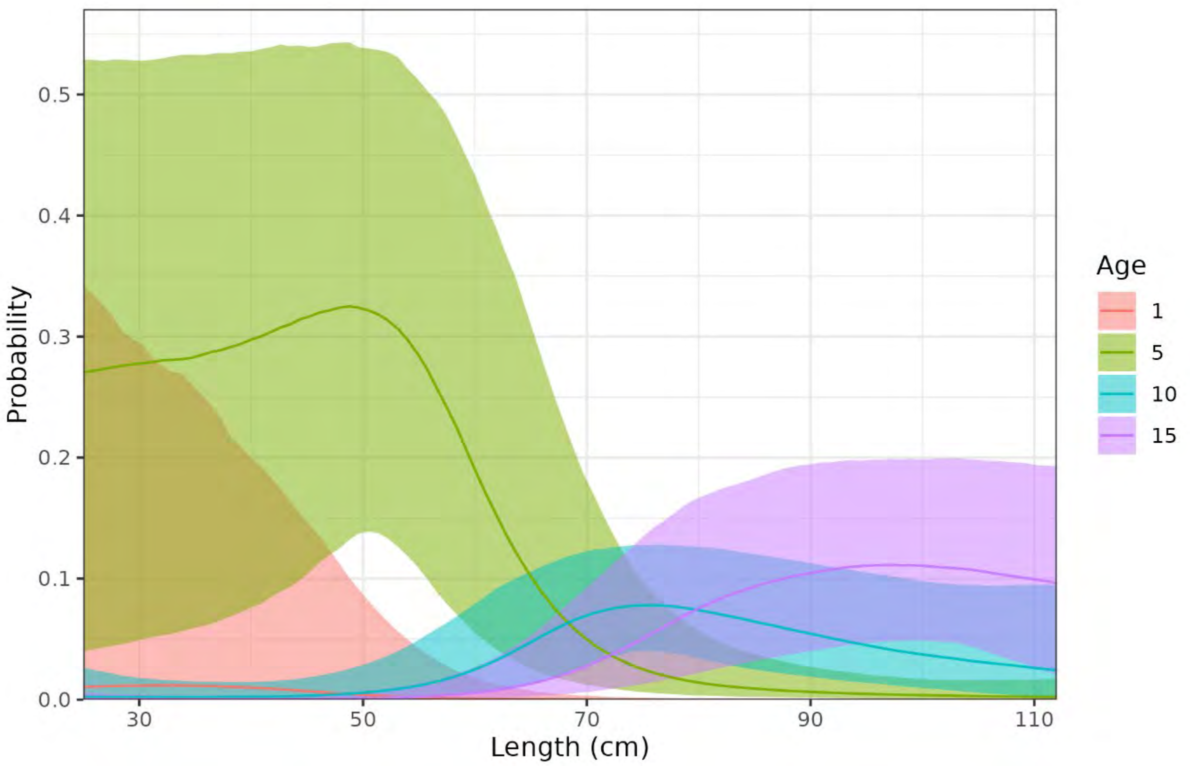

Figure 12: Conditional effect of length for a subset of ages HAK males for the selected model (HM5) (see Table 2).

Table 1: HAK female age-conditional-on-length model statistics for models fitted to the ordinal data: ΔELPD (expected log-predictive density, higher is better), LOO IC (leave-one-out information criteria, lower is better), P LOO (the effective number of parameters), Bayesian R^2 (higher is better), the number of divergent transitions, and time taken for each model run (HH:MM). The values in the ΔELPD column were computed by making pairwise comparisons between each model and the model with the largest ELPD. For this reason, the ΔELPD column will always have the value zero in the first row (i.e., the difference between the preferred model and itself) and negative values in subsequent rows for the remaining models. The models highlighted in blue are the chosen model structure and the grey indicate the preferred model (HF4), which was the model used for deriving scaled age compositions in Section 3.4, and the models in green are the models with (HF11) and without separate thresholds by fishing year (HF18).

		ΔELPD		LOO IC		P LOO		Bayesian R^2			
Label	Model structure	Mean	SD	Mean	SE	Mean	SE	Mean	SE	Divergent	Time
HF1	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + s(x, y) disc ~ 0 + fyear	0	0	28 918	196	595.3	15.3	0.697	0.003	0	01:38
HF2	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month + s(x, y) disc ~ 0 + fyear	0	2	28 919	196	596.5	15.2	0.697	0.003	0	02:05
HF3	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month disc ~ 0 + fyear	-7	5	28 933	196	588.2	15.1	0.696	0.003	0	01:34
HF4	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) disc ~ 0 + fyear	-8	5	28 934	196	583.7	15.1	0.696	0.003	0	01:03
HF5	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + stratum disc ~ 0 + fyear	-13	5	28 945	196	588.7	15.5	0.696	0.003	0	01:00
HF6	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26))	-49	29	29 016	191	623.0	15.7	0.703	0.003	0	02:31
HF7	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) disc ~ 0 + fyear	-331	36	29 581	205	622.2	15.5	0.615	0.004	0	02:16
HF8	age | thres(gr = fyear) ~ s(length, k = 20, by = fyear)	-379	33	29 677	192	599.1	13.9	0.623	0.004	0	05:20
HF9	age | thres(gr = fyear) ~ s(length, k = 9, by = fyear)	-384	33	29 686	193	589.7	13.8	0.622	0.004	0	02:33
HF10	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + s(x, y)	-385	32	29 688	192	589.2	13.7	0.623	0.004	0	01:30
HF11	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear)	-395	33	29 708	193	578.7	13.6	0.621	0.004	0	01:23
HF12	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + stratum	-402	33	29 722	193	581.3	13.5	0.621	0.004	0	00:58
HF13	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + month	-416	33	29 751	193	580.5	13.5	0.620	0.004	0	01:14
HF14	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + origin	-427	35	29 772	198	581.2	13.3	0.615	0.004	0	01:16
HF15	age | thres(gr = fyear) ~ s(length, k = 3, by = fyear)	-446	32	29 809	192	559.9	13.3	0.621	0.004	0	00:37
HF16	age | thres(gr = fyear) ~ s(length, k = 3)	-534	36	29 986	191	525.9	13.2	0.620	0.004	0	00:17
HF17	age | thres(gr = fyear) ~ t2(length, k = 6, by = fyear)	-557	36	30 033	194	598.9	14.6	0.623	0.004	0	00:46
HF18	age ~ s(length, k = 6, by = fyear)	-905	58	30 729	194	158.5	11.3	0.603	0.004	0	01:23
HF19	age | thres(gr = fyear) ~ length	-1 544	51	32 006	177	562.5	15.5	0.551	0.006	0	00:29

Table 2: HAK male age-conditional-on-length model statistics for models fitted to the ordinal data: ΔELPD (higher is better), LOO IC (lower is better), P LOO (the effective number of parameters), Bayesian R^2 (higher is better), the number of divergent transitions, and time taken for each model run (HH:MM). The values in the ΔELPD column were computed by making pairwise comparisons between each model and the model with the largest ELPD. For this reason, the ΔELPD column will always have the value zero in the first row (i.e., the difference between the preferred model and itself) and negative values in subsequent rows for the remaining models. The models highlighted in grey indicate the preferred model (HM5), which was the model used for deriving scaled age compositions in Section 3.4, and the models in green are the model with (HM11) and without separate thresholds by fishing year (HM18).

		ΔELPD		LOO IC		P LOO		Bayesian R^2			
Label	Model structure	Mean	SD	Mean	SE	Mean	SE	Mean	SE	Divergent	Time
HM1	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + s(x, y) disc ~ 0 + fyear	0	0	24011	198	619.7	17.7	0.695	0.004	0	01:13
HM2	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month + s(x, y) disc ~ 0 + fyear	-4	3	24019	198	624.7	18.2	0.696	0.004	0	01:54
HM3	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + stratum disc ~ 0 + fyear	-18	7	24048	198	608.9	17.4	0.693	0.004	0	00:56
HM4	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month disc ~ 0 + fyear	-21	7	24053	199	612.3	17.9	0.693	0.004	0	01:08
HM5	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) disc ~ 0 + fyear	-25	7	24061	199	611.4	18.0	0.692	0.004	0	01:19
HM6	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26))	-253	43	24518	208	656.4	27.7	0.689	0.004	0	01:03
HM7	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) disc ~ 0 + fyear	-260	37	24530	203	615.6	16.0	0.607	0.005	0	01:12
HM8	age | thres(gr = fyear) ~ s(length, k = 20, by = fyear)	-298	35	24608	191	594.0	14.5	0.619	0.005	0	05:02
HM9	age | thres(gr = fyear) ~ s(length, k = 9, by = fyear)	-308	35	24626	191	585.5	14.5	0.616	0.005	0	01:13
HM10	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + s(x, y)	-309	35	24629	192	590.7	14.1	0.615	0.005	0	00:54
HM11	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear)	-329	35	24669	192	580.3	14.4	0.612	0.005	0	00:57
HM12	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + stratum	-336	36	24683	193	582.5	14.3	0.611	0.005	0	00:48
HM13	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + origin	-382	38	24776	199	583.5	14.0	0.605	0.005	0	01:04
HM14	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + month	-388	36	24788	193	584.8	14.2	0.608	0.005	0	00:53
HM15	age | thres(gr = fyear) ~ s(length, k = 3, by = fyear)	-440	38	24890	195	566.2	13.8	0.609	0.005	0	00:32
HM16	age | thres(gr = fyear) ~ s(length, k = 3)	-514	39	25039	192	533.7	13.8	0.599	0.005	0	00:17
HM17	age | thres(gr = fyear) ~ t2(length, k = 6, by = fyear)	-516	41	25043	201	611.5	15.7	0.609	0.005	0	00:42
HM18	age ~ s(length, k = 6, by = fyear)	-820	64	25651	200	190.1	18.0	0.585	0.006	0	01:57
HM19	age | thres(gr = fyear) ~ length	-1598	53	27207	180	564.8	16.3	0.475	0.008	0	00:22

For ling, length for missing ages was imputed using the von Bertalanffy growth model (von Bertalanffy 1938) for each sex to predict length given the missing age. The von Bertalanffy model fits were inspected visually and deemed acceptable for ling (Figure A3.11, Figure A3.12). The number of missing ages imputed was substantial in some years, specifically for those older ages less frequently observed. The numbers of ages for each length for ling are given in Table A3.3 (females) and Table A3.4 (males).

For ling, the suite of model runs is presented for females (Table 3) and males (Table 4). As with hake, the run time of these models ranged from about 20 minutes to over five hours, with the larger number of basis dimensions for the length splines having the greatest effect on run time. Some of the model runs had divergent transitions implying that they should not be used to make inferences. A single female ling model had one divergent transition for the model with a smooth spline of space (latitude and longitude, LF8). All male ling models that included a tensor spline of space (latitude and longitude) had divergent transitions (i.e., models LM1 to LM4).

As with hake, models that included thresholds that varied by fishing year performed better, according to LOO IC and Bayesian R^2, than those without. The worst performing model for females and males included length as a linear term (LF19 and LM19); all other model runs included length as a spline with a range of different basis dimensions. Splines with higher basis dimensions produced models that performed better according to LOO IC; however, Bayesian R^2 suggested that the improvements to the deviance explained were minimal with higher basis dimensions (e.g., for females, models with the same structure but a basis dimension of 3, 6, 9, and 20 had improving LOO ICs but the mean Bayesian R^2 was similar for these models with, for example, female values of 0.805, 0.817, 0.818, and 0.818, respectively).

The best model runs for females and males (LF1 and LM1), according to LOO IC and Bayesian R^2, both included a 2-dimensional tensor product smooth by length using a basis dimension of 6 and by fishing year (continuous) using a basis dimension of 26. For females, the model structure selected by LOO IC included a spatial smooth term (LF1). However, as for hake, the 2-dimensional tensor product smooth by length and fishing year also resulted in poor fits to the observed age data (Figure 13, Figure 14). The best female ling model without a 2-dimensional tensor product smooth by length and fishing year provided better fits when assessed visually (LF8, Figure 13, Figure 14). For males, the first four model runs that had the best LOO ICs (LM1 to LM4) all had divergent transitions so could not be used for model inference. The model run for males that had the best LOO IC with no divergent transitions (LM5) also had a spatial smooth term but did not include a 2-dimensional tensor product smooth by length and fishing year so this model was compared to the next best model that did include a 2-dimensional tensor product smooth by length and fishing year (LM6, Figure 15, Figure 16).

For the best models according to LOO IC without divergent transitions and without a 2-dimensional tensor product smooth by length and fishing year, the MCMC mixing for female (LF8) and male (LM5) models was excellent (Figure A3.13, Figure A3.14). Conditional effects plots show the distribution of lengths estimated for each age for these models (Figure 17, Figure 18).

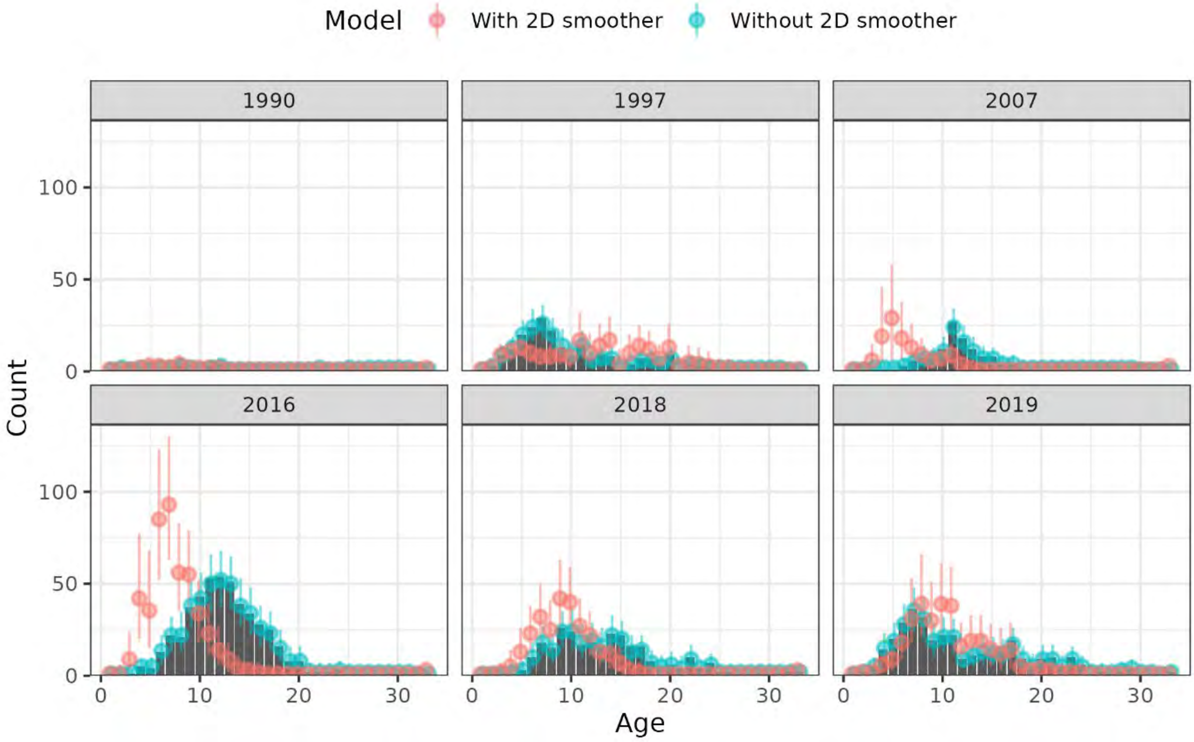

Figure 13: The distribution of the data (bars) and the posterior predictive distribution (median as points and 90% credible interval as lines) for the female ling models with (LF1, red) and without (LF8, blue) 2-dimensional tensor product smooths (see Table 3).

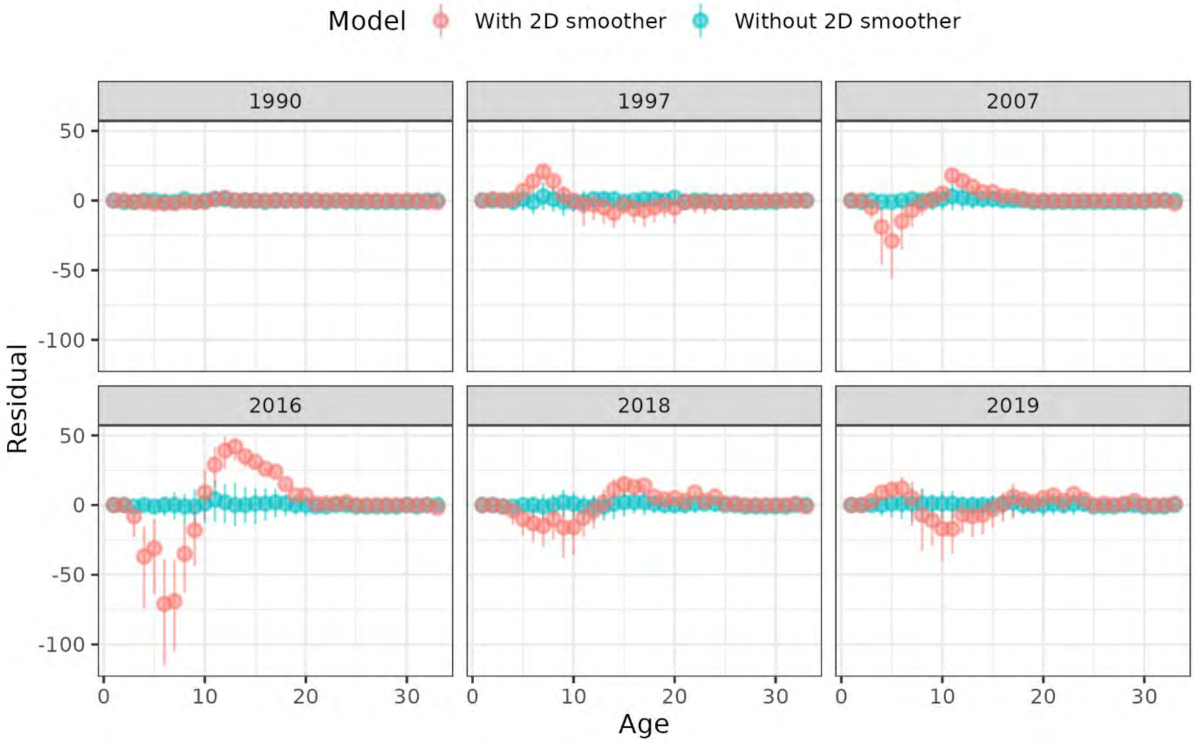

Figure 14: Residuals (median as points and 90% credible interval as lines) for the female ling models with (LF1, red) and without (LF8, blue) 2-dimensional tensor product smooths (see Table 3).

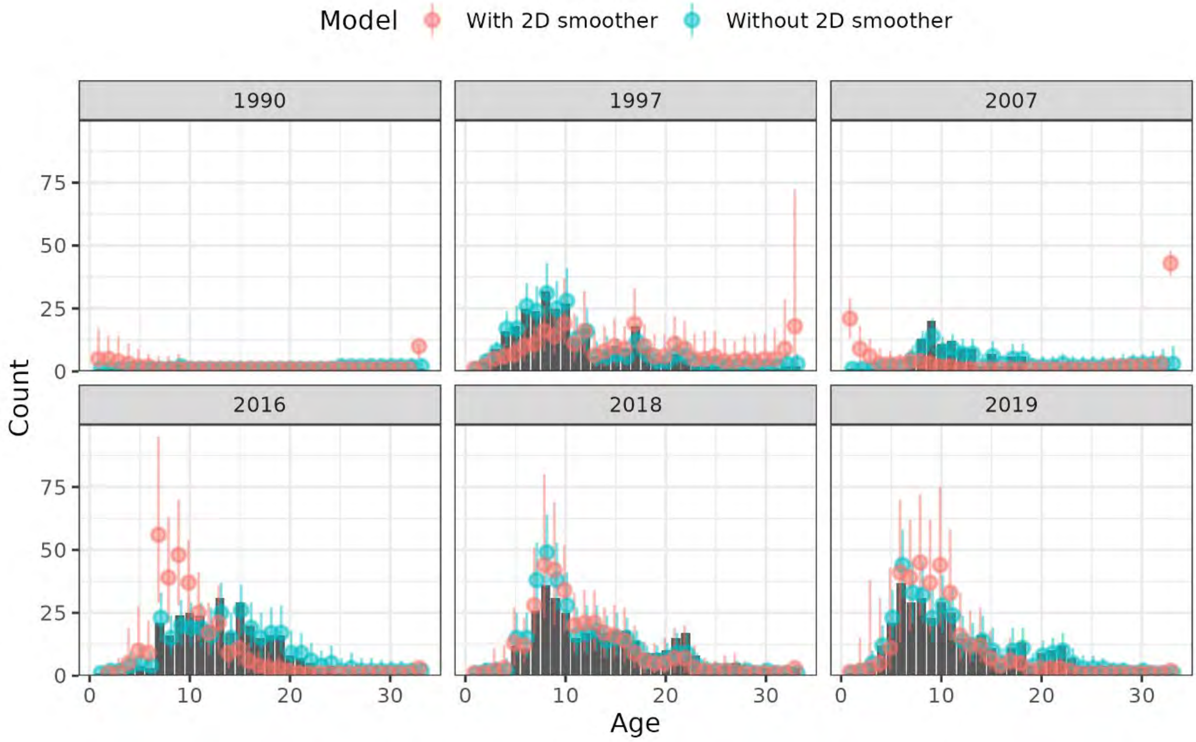

Figure 15: The distribution of the data (bars) and the posterior predictive distribution (median as points and 90% credible interval as lines) for the male ling models with (LM6, red) and without (LM7, blue) 2-dimensional tensor product smooths (see Table 4).

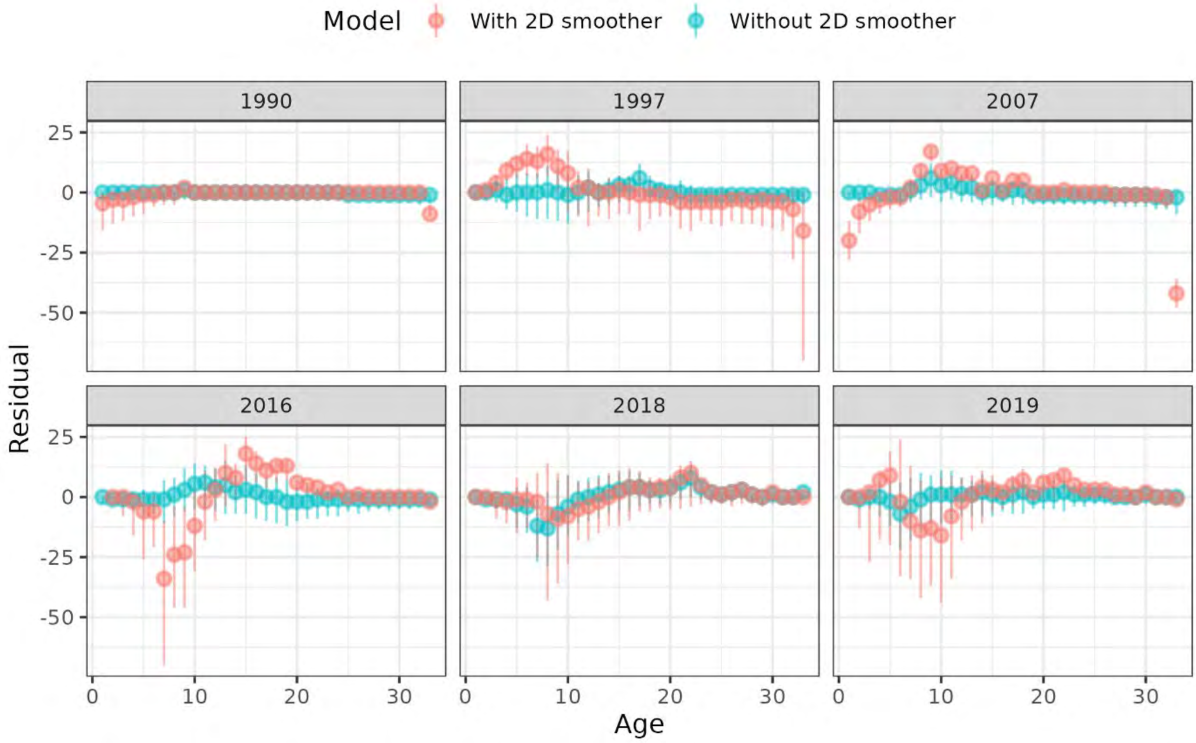

Figure 16: Residuals (median as points and 90% credible interval as lines) for the male ling models with (LM5, red) and without (LM7, blue) 2-dimensional tensor product smooths (see Table 4).

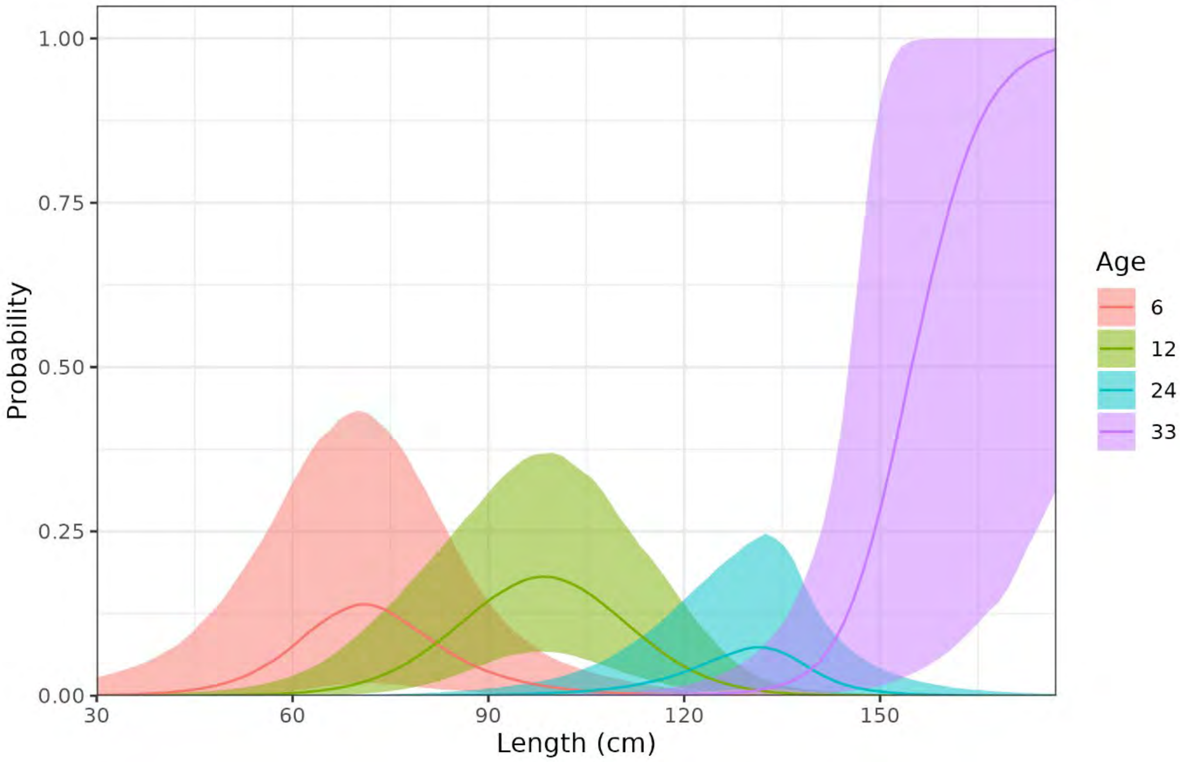

Figure 17: Conditional effect of length for a subset of ages LIN females for the selected model (LF8).

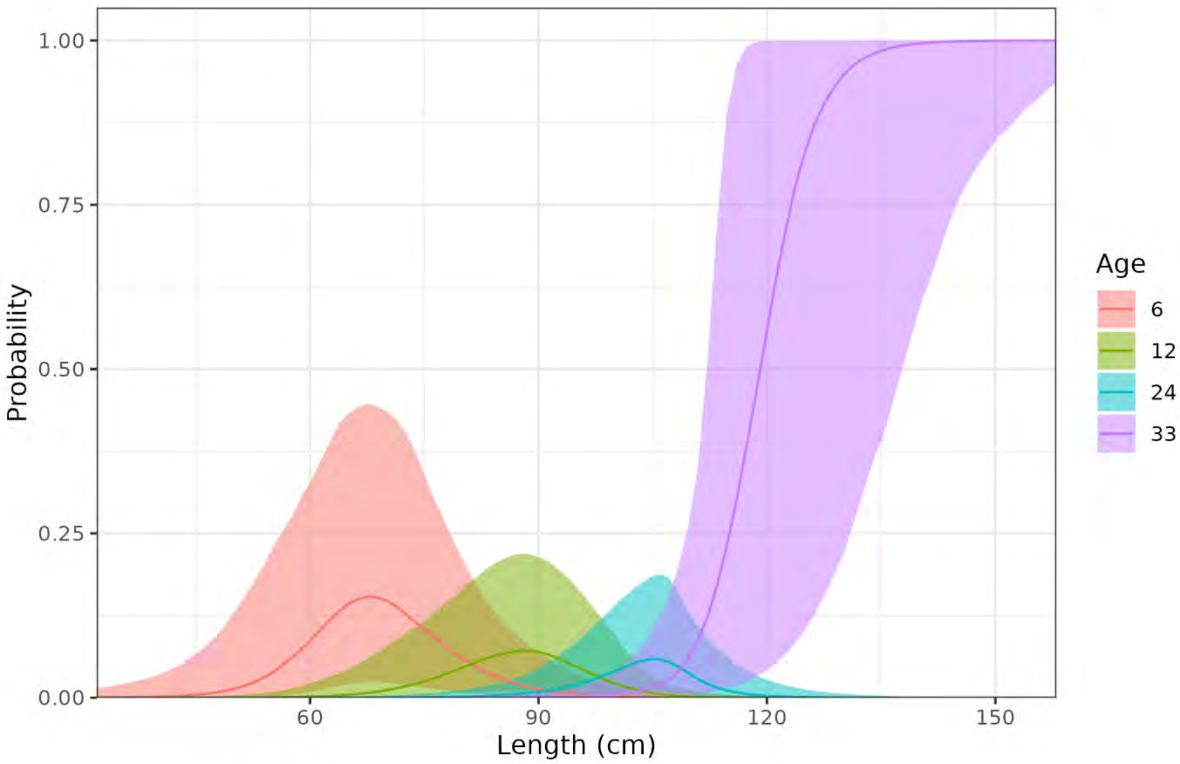

Figure 18: Conditional effect of length for a subset of ages LIN males for the selected model (LM5).

Table 3: LIN female age-conditional-on-length model statistics for models fitted to the ordinal data: ΔELPD (higher is better), LOO IC (lower is better), P LOO (the effective number of parameters), Bayesian R^2 (higher is better), the number of divergent transitions, and time taken for each model run (HH:MM). The values in the ΔELPD column were computed by making pairwise comparisons between each model and the model with the largest ELPD. For this reason, the ΔELPD column will always have the value zero in the first row (i.e., the difference between the preferred model and itself) and negative values in subsequent rows for the remaining models. The models highlighted in grey indicate the models with (LF1) and without (LF8) 2-dimensional tensor product smooths.

		ΔELPD		LOO IC		P LOO		Bayesian R^2			
Label	Model structure	Mean	SD	Mean	SE	Mean	SE	Mean	SE	Divergent	Time
LF1	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + s(x, y) disc ~ 0 + fyear	0	0	28 208	190	842.3	19.1	0.854	0.001	0	02:15
LF2	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month + s(x, y) disc ~ 0 + fyear	-3	8	28 214	190	854.3	19.1	0.854	0.001	0	02:15
LF3	age ~ s(length, k = 6, by = fyear)	-130	44	28 468	184	141.9	5.3	0.813	0.002	0	00:53
LF4	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month disc ~ 0 + fyear	-137	19	28 482	189	832.8	18.6	0.849	0.001	0	01:35
LF5	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) disc ~ 0 + fyear	-174	19	28 555	189	818.5	18.6	0.848	0.001	0	01:10
LF6	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + stratum disc ~ 0 + fyear	-227	23	28 662	192	808.1	17.9	0.842	0.002	0	02:02
LF7	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26))	-277	49	28 762	194	826.0	24.4	0.845	0.002	0	01:43
LF8	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + s(x, y)	-358	28	28 925	174	784.5	15.9	0.827	0.002	1	01:08
LF9	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) disc ~ 0 + fyear	-417	33	29 043	192	842.5	18.5	0.816	0.002	0	01:30
LF10	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + origin	-429	32	29 066	176	781.1	16.1	0.829	0.002	0	00:58
LF11	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + month	-514	33	29 236	176	783.7	15.7	0.817	0.003	0	01:06
LF12	age | thres(gr = fyear) ~ s(length, k = 20, by = fyear)	-522	33	29 252	174	785.9	15.8	0.818	0.002	0	05:19
LF13	age | thres(gr = fyear) ~ s(length, k = 9, by = fyear)	-527	33	29 263	174	784.1	16.0	0.818	0.002	0	02:26
LF14	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear)	-533	33	29 274	174	778.5	15.9	0.817	0.002	0	00:54
LF15	age | thres(gr = fyear) ~ s(length, k = 3)	-549	35	29 305	175	709.6	14.7	0.811	0.003	0	00:19
LF16	age | thres(gr = fyear) ~ s(length, k = 3, by = fyear)	-549	33	29 306	174	748.2	15.1	0.815	0.002	0	00:29
LF17	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + stratum	-607	41	29 422	184	777.0	15.4	0.805	0.003	0	01:33
LF18	age | thres(gr = fyear) ~ t2(length, k = 6, by = fyear)	-659	38	29 525	180	819.0	21.0	0.821	0.002	0	00:37
LF19	age | thres(gr = fyear) ~ length	-1 586	49	31 380	150	760.2	18.0	0.775	0.004	0	00:33

Table 4: LIN male age-conditional-on-length model statistics for models fitted to the ordinal data: ΔELPD (higher is better), LOO IC (lower is better), P LOO (the effective number of parameters), Bayesian R^2 (higher is better), the number of divergent transitions, and time taken for each model run (HH:MM). The values in the ΔELPD column were computed by making pairwise comparisons between each model and the model with the largest ELPD. For this reason, the ΔELPD column will always have the value zero in the first row (i.e., the difference between the preferred model and itself) and negative values in subsequent rows for the remaining models. The models highlighted in grey indicate the models with (LM6) and without (LM5) 2-dimensional tensor product smooths.

		ΔELPD		LOO IC		P LOO		Bayesian R^2			
Label	Model structure	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Divergent	Time
LM1	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month disc ~ 0 + fyear	0	0	30287	189	565.2	14.0	0.757	0.002	2044	13:20
LM2	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + month + s(x, y) disc ~ 0 + fyear	-15	30	30317	191	891.8	18.5	0.774	0.002	4	02:57
LM3	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + s(x, y) disc ~ 0 + fyear	-22	30	30330	191	886.0	18.5	0.773	0.002	5	02:56
LM4	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) disc ~ 0 + fyear	-120	22	30528	189	677.7	14.5	0.752	0.003	200	16:32
LM5	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + s(x, y)	-301	43	30889	177	793.9	15.2	0.746	0.003	0	01:23
LM6	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26))	-321	28	30929	183	818.5	17.4	0.752	0.003	0	01:19
LM7	age | thres(gr = fyear) ~ t2(length, cyear, k = c(6, 26)) + stratum disc ~ 0 + fyear	-332	32	30951	190	815.2	17.4	0.736	0.003	0	01:25
LM8	age ~ s(length, k = 6, by = fyear)	-368	48	31023	180	135.3	6.4	0.712	0.004	0	01:03
LM9	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) disc ~ 0 + fyear	-476	36	31238	191	833.5	17.4	0.719	0.003	0	02:37
LM10	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + origin	-504	37	31295	180	792.2	15.4	0.730	0.003	0	01:43
LM11	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + stratum	-558	43	31404	188	792.4	15.3	0.706	0.004	0	01:08
LM12	age | thres(gr = fyear) ~ s(length, k = 20, by = fyear)	-563	38	31412	178	803.9	15.5	0.722	0.003	1	05:38
LM13	age | thres(gr = fyear) ~ s(length, k = 9, by = fyear)	-569	38	31424	178	795.1	15.3	0.721	0.003	0	02:03
LM14	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear) + month	-578	37	31443	180	794.8	15.0	0.721	0.004	0	01:08
LM15	age | thres(gr = fyear) ~ s(length, k = 6, by = fyear)	-580	38	31447	178	784.1	15.1	0.720	0.003	0	01:06
LM16	age | thres(gr = fyear) ~ s(length, k = 3)	-595	40	31476	177	717.0	14.1	0.712	0.003	0	00:20
LM17	age | thres(gr = fyear) ~ s(length, k = 3, by = fyear)	-604	39	31494	178	764.2	14.7	0.718	0.003	0	00:31
LM18	age | thres(gr = fyear) ~ t2(length, k = 6, by = fyear)	-801	47	31889	187	861.6	21.5	0.718	0.003	0	00:31
LM19	age | thres(gr = fyear) ~ length	-2271	62	34830	148	763.1	16.9	0.593	0.008	0	00:25

In all models the most important explanatory variables were length and fishing year, but other potential explanatory variables included the month, origin, latitude, and longitude of the sample. The spline s(length, k = 3) tended to be better than s(length), and three basis dimensions were adequate for these data—potentially as the distribution of length-at-age was relatively smooth with a single mode. The ‘shape’ using higher values for the number of basis dimensions looked very similar and hence offered no significant improvement to the models. For both the hake and ling males and females, the LOO IC diagnostics indicated that the cumulative (“logit”) distribution was better than *cratio* or *sratio*.

Note that in developing the ordinal age-at-length models, the use of the cumulative(“probit”) distribution failed (potentially for computational reasons). In addition, attempts to use k-fold cross validation were not successful because when data are dropped, all data for some ages are also dropped. However, this issue could be solved by adding a single data point for each age class and assigning zero weight to each of these data points.

3.4 Estimates of the scaled age compositions for hake

Estimates of scaled age compositions were generated for hake only and for a subset of years (2012–2019) because the length composition models were very slow. Two contrasting ordinal models were used to estimate the age composition (HF4 and HM7). The female model (HF4) was selected because it included a 2-dimensional tensor product smooth that was shown to perform poorly when fitting to the age observations while the selected male model (HM7) did not include this term. Furthermore, simpler model were selected because these models only included terms for length and year (no additional terms like month or stratum were used) so that the ordinal model strata matched the length composition strata. Future implementations should consider not aggregating the length compositions before applying the ordinal model (leaving the length compositions stratified by year, month, and stratum). The scaled length compositions were very similar for both females and males (Figure 19, Figure 20). However, there were some differences in the scaled age compositions, particularly for females (Figure 19). Reasons for these differences may include: minor differences in the scaled length compositions; the addition of imputed age samples (using the von Bertalanffy growth curves) in the ordinal models; and/or the lack of fit to the paired age-length samples in some years (e.g., see 2019 in Figure 1).

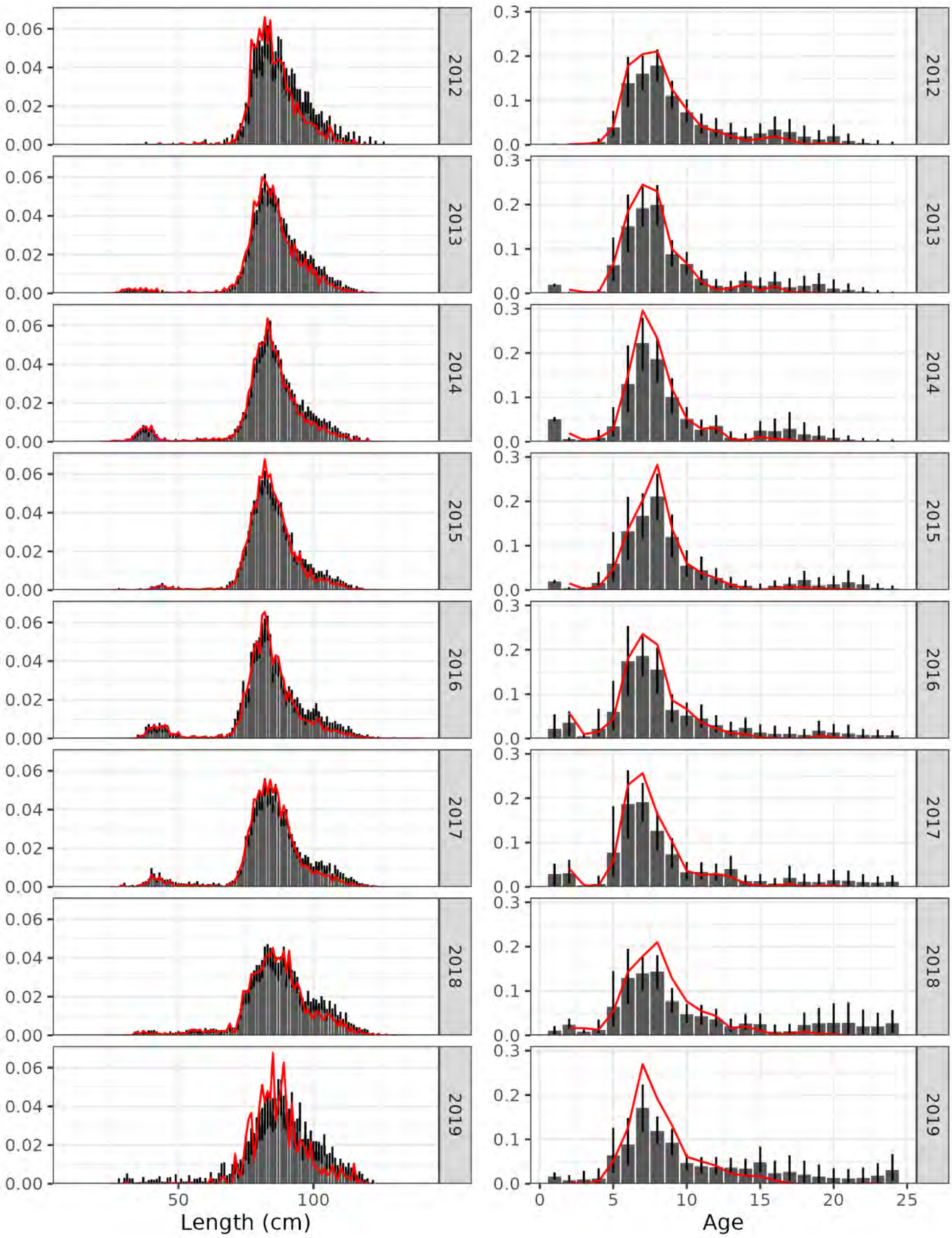

Figure 19: Estimates of the scaled length and age compositions for female hake (using HF4). The red line is the empirical ALK approach.

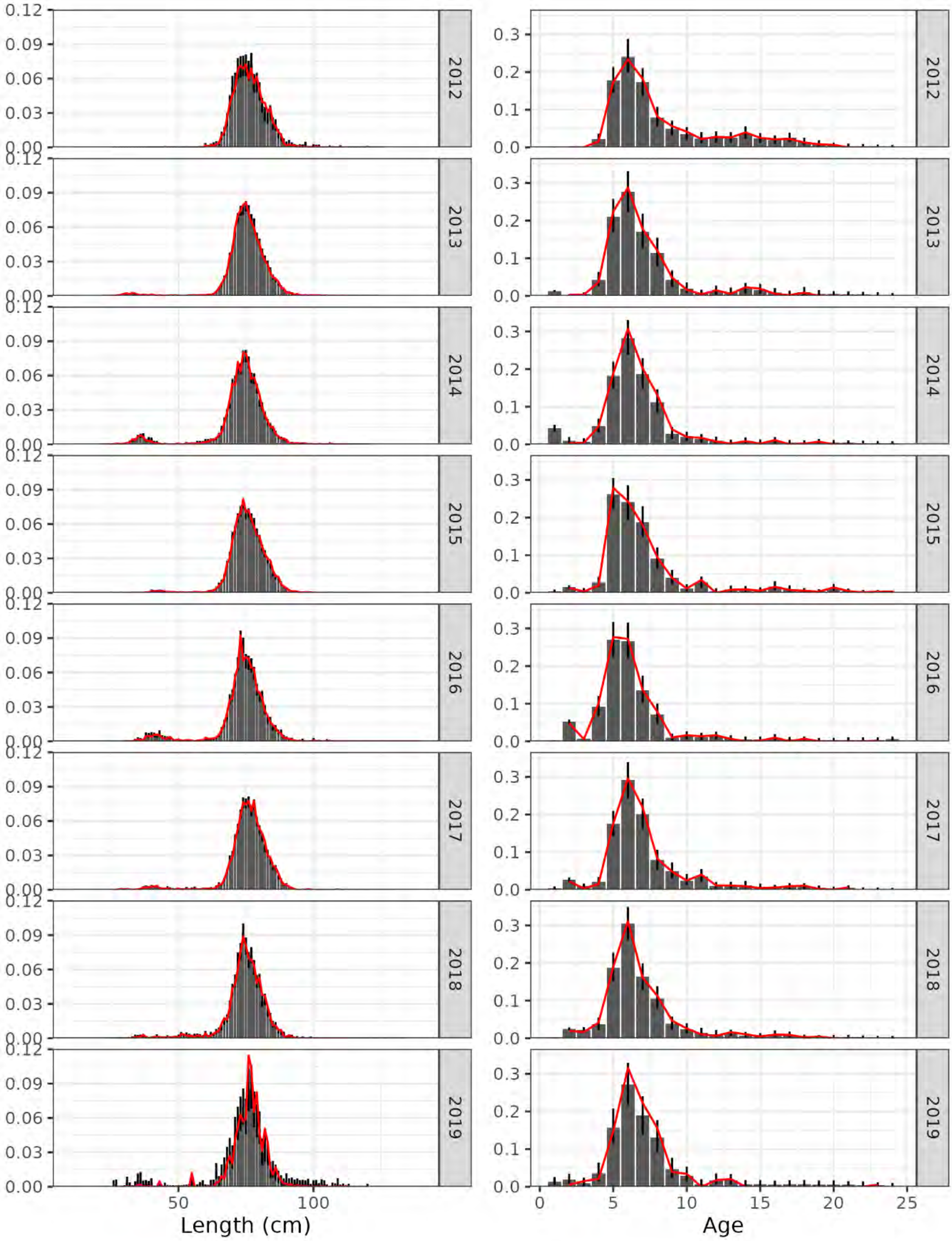

Figure 20: Estimates of the scaled length and age compositions for male hake (using HM7). The red line is the empirical ALK approach.

4. DISCUSSION

4.1 Scaled length compositions

The analysis of length frequency compositions generally gave similar results to the empirical methods when data were adequate to allow the estimation using empirical approaches. However, computational requirements did mean that the development of the categorical models to generate scaled length compositions was difficult and required some data aggregation and substantive model tweaking to produce robust results.

For red rock lobster, the assumption that trip stratum level is an adequate random sample of the individual hauls and that the effect of spatial and temporal patterns does not occur at the scale of an individual trip stratum is plausible and likely—especially as trips were short (usually a single day) and the sampling of the lengths in observer catch sampling and voluntary logbook programmes were spatially and temporally constrained to small areas and time.

For hake, ling, and more generally deepwater fish stocks, this is less likely to be ideal. Individual deepwater trips tend to target multiple species, in a wide range of areas and over different times of day and across seasons. While this can be aggregated into strata for modelling of the length frequencies, the increased levels of disaggregation rapidly lead to computer resource requirements that make evaluation of these models slow and difficult. However, in the hake and ling case studies, the results were similar and hence still may still be adequate for the estimation of scaled length composition estimates. More comprehensive testing and evaluation of the difference between empirical approaches and model-based approaches to the generation of scaled length compositions is required to fully resolve these difficulties and determine the best assumptions for an adequate estimation of the scaled length frequency.

Although the implementation of these methods is ‘resource hungry’, considerable development has been made to evaluate if these could be implemented. We note that the major problem with these models is the development of the length frequency categorical model requiring assumptions of aggregation of data to larger spatial scales (for example, at the trip-stratum level) rather than using the event-level sample data as in empirical methods such as used by catch-at-age (Bull & Dunn 2002).

A potential improvement is to substitute the Poisson decomposed multinomial distribution for the multinomial in the development of the categorical models for length composition, as this distribution is less computationally intensive.

4.2 Age-at-length and scaled age compositions

Ordinal models of the age-at-length relationship were reasonably successful, taking account of the relationship between ages and lengths over time, as well as accounting for sparse otolith samples that can occur in specific locations and times. Model run times for these models was reasonable and able to be implemented easily using *brms*. These are effectively the same as a highly spatially and temporally stratified ALK while resolving issues of missing or incomplete data from specific strata or temporal periods.

Where there were sparse observations within a stratum or year, the modelled results are more likely to return reasonable results than those provided by empirical methods. However, aggregation to a trip-stratum level was required to fit the length composition models using available computer resources, and hence scaled age compositions may not be adequate. This is a limitation of the approach until more efficient algorithms for fitting the categorical models using multinomial approaches are developed.

The development of modelling approach to investigate the effect of covariates on the age length relationship is recommended. Even if empirical ALKs are used, evaluation of appropriate stratification and identification of the effect of missing data or non-representative samples can be difficult using empirical methods. The development of ordinal models, such as presented here, allow such effects to be formally evaluated.

Determination of scaled age compositions requires an alternative approach, as computational overheads make this unlikely with current computing power. Future development may be best focused on developing hybrid methods. Here, empirical (or frequentist) methods could be used to estimate the scaled length frequency, to which an ordinal age-at-length model would then be applied to produce the scaled age composition. This approach would combine empirical (or frequentist) methods with Bayesian modelling frameworks to reduce time and memory requirements, while still allowing inference between strata and years to minimise the effect of sparse data.

5. POTENTIAL RESEARCH

The development of the multinomial and ordinal approaches to estimating age compositions from length- and age-data requires additional development before it can be directly used. Future research considerations should include:

- Further exploration of the Poisson decomposed multinomial distribution, instead of the multinomial distribution, for modelling length composition as it is much faster.
- Development of hybrid approaches to account for the complexities in modelling length composition, while including the ordinal models for estimating age compositions. This may also include frequentist approaches (e.g., *mgcv* with random effects, see Wood 2017, or template model builder, *TMB*) to model the scaled length composition data.
- Different treatments of time (e.g., Julian date rather than fishing year and month).
- Further exploration of spatial effects including spatial effects that vary over time in relationship to environmental variables (e.g., to account for climate change).
- Consider recoding the age conditional length models in *TMB* to make use of its efficient Gaussian Markov random field (GMRF) algorithm for modelling age structure through time.
- Consider bespoke *Stan* or *TMB* code to integrate length composition and age conditional on length models. This would significantly speed up the process of scaling the length compositions by the catch and then predicting the age compositions. This would also enable the use of more advanced features (e.g., GMRF approaches; Cheng et al. 2023).

The methods would also require development of appropriate diagnostic tools and evaluation of how much missing data can be imputed before the estimates become unreliable. K-fold validation (with appropriate choices of samples to drop) would be helpful in evaluating the robustness of the methods. The evaluation of different stratification approaches (e.g., using seasons or months, spatial strata or continuous surfaces, gear types) should also be evaluated in specific cases and may lead to development of general guidelines for analysis.

6. ACKNOWLEDGEMENTS

We thank the Fisheries New Zealand Data Management Team for the data extracts and additional information used in these analyses and their assistance in its interpretation. We thank members of the Statistics, Assessment and Methodology Working Group for their peer review and helpful discussion throughout. We also thank Phillip Neubauer (Dragonfly Data Science) who provided helpful discussions and comments on the methods developed within this report. This project was funded by Fisheries New Zealand under contract SEA2021-28.

7. REFERENCES

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

Babyn, J.; Varkey, D.; Regular, P.; Ings, D.; Mills Flemming, J. (2021). A gaussian field approach to generating spatial age length keys. *Fisheries Research 240*: 105956. https://doi.org/10.1016/j.fishres.2021.105956

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 Journal of Marine Science 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. 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.

Bürkner, P.-C.; Vuorre, M. (2019). Ordinal Regression Models in Psychology: A Tutorial. *Advances in Methods and Practices in Psychological Science 2*: 77–101. https://doi.org/10.1177/2515245918823199

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

Cheng, M.LH.; Thorson, J.T.; Ianelli, J.N.; Cunningham, C.J. (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. https://doi.org/10.1016/j.fishres.2023.106755

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. *Canadian Journal of Fisheries and Aquatic Sciences 77(11)*: 1810–1821. https://doi.org/10.1139/cjfas-2020-0166

Dunn, A.; Mormede, S.; Webber, D.N. (2023). Descriptive analysis and model inputs for the 2022 stock assessment of hake (*Merluccius australis*) off the west coast South Island (HAK 7) for the 2020–21 fishing year. *New Zealand Fisheries Assessment Report 2023/44*. 56 p.

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

Horn, P.L. (2005). A review of the stock structure of ling (*Genypterus blacodes*) in New Zealand waters. *New Zealand Fisheries Assessment Report 2005/59*. 41 p.

Horn, P.L. (2011). Stock assessment of hake (*Merluccius australis*) off the west coast of South Island (HAK 7) for the 2010–11 fishing year. *New Zealand Fisheries Assessment Report 2011/33*. 46 p.

Koblinger, B. (2023). Cumulative ordinal model in *brms* without any observations of one category. *https://discourse.mc-stan.org*.

Langley, A.D. (2021). Stock assessment of snapper in SNA 8 for 2021. *New Zealand Fisheries Assessment Report 2021/38*. 86 p.

Mormede, S.; Dunn, A.; Webber, D.N. (2022). Descriptive analysis of ling (*Genypterus blacodes*) on the Chatham Rise (LIN 3&4) up to 2020-21 and inputs for the 2022 stock assessment. *New Zealand Fisheries Assessment Report 2022/64*. 81 p.

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

Roberts, J.; Starr, P.J.; Webber, D.N.; Rudd, M.B.; Pons, M.; Goeden, Z. (2023). Data for the 2022 stock assessment of red rock lobsters (*Jasus edwardsii*) in CRA 2. New Zealand Fisheries Assessment Report 2023/42. Ministry for Primary Industries, 91 p.

Saunders, R.; Hart, A.C.; Horn, P.L.; Sutton, C.P. (2021). Catch-at-age for hake (*Merluccius australis*) and ling (*Genypterus blacodes*) for the 2018–19 fishing year and from a research trawl survey in 2020, and a summary of the available data sets from the New Zealand EEZ. *New Zealand Fisheries Assessment Report 2021/15*. 97 p.

Stan Development Team (2020). Stan Modelling Language: User’s Guide and Reference Manual Version 2.24. Stan Governing Body.

Starr, P.J.; Webber, D.N.; Rudd, M.B.; Pons, M.; Roberts, J. (2021). Data for the 2020 stock assessment of red rock lobsters (*Jasus edwardsii*) in CRA 4. *New Zealand Fisheries Assessment Report 2021/56.* 80 p.

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

Vehtari, A.; Gelman, A.; Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing 27*: 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

von Bertalanffy, L. (1938). A quantitative theory of organic growth (Inquiries on growth laws. II). *Human Biology 10*: 181–213.

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.

8. APPENDIX 1: TABLES AND FIGURES FOR THE LENGTH COMPOSITION MODELS

8.1 Red rock lobster in CRA 2

Table A1.1: Number of vessels, skippers, trips, days sampled, pots lifted, and lobsters measured in the CRA 2 catch sampling and logbook programmes each fishing year. The fishing years indicated in grey were not included in the categorical length models.

Fishing year	Catch sampling					Logbook					
	vessels	days	trips	pots	lobsters	vessels	skippers	days	trips	pots	lobsters
1986	2	3	3	159	562	–	–	–	–	–	–
1990	2	3	3	127	364	–	–	–	–	–	–
1991	1	2	2	115	918	–	–	–	–	–	–
1993	–	–	–	–	–	26	23	253	1 915	4 494	18 092
1994	–	–	–	–	–	30	25	254	1 613	4 307	18 133
1995	–	–	–	–	–	26	21	176	1 048	2 809	11 657
1996	–	–	–	–	–	21	19	187	902	2 508	12 522
1997	–	–	–	–	–	19	16	169	766	2 111	9 224
1998	–	–	–	–	–	21	17	191	781	2 199	9 172
1999	12	12	12	917	3 073	20	17	153	949	2 642	10 323
2000	13	13	13	1 002	3 022	16	16	198	950	2 568	9 186
2001	10	11	11	859	3 014	13	13	268	743	1 787	5 893
2002	12	13	14	724	2 274	18	18	267	1 197	2 738	8 269
2003	12	12	12	737	1 898	13	13	289	979	2 440	7 445
2004	12	12	12	530	1 353	14	14	319	1 144	2 730	8 532
2005	10	12	12	752	2 303	18	18	301	1 442	3 563	10 748
2006	9	12	12	772	1 964	17	17	251	1 057	2 765	8 836
2007	11	12	12	828	2 237	14	13	245	1 088	2 678	9 008
2008	7	12	12	874	2 568	15	15	257	1 281	3 302	10 415
2009	8	11	14	881	2 769	15	14	240	1 195	2 861	8 662
2010	12	13	13	783	1 926	16	16	274	1 471	3 518	12 003
2011	12	14	17	857	2 175	17	19	252	1 357	3 177	8 831
2012	11	13	13	835	2 465	17	17	256	1 502	3 696	11 636
2013	9	10	12	864	2 868	19	18	284	1 585	3 853	11 976
2014	8	12	12	882	2 666	15	16	266	1 373	3 147	8 016
2015	8	13	13	768	1 450	16	17	274	1 342	2 921	6 315
2016	9	13	13	763	2 005	15	16	223	859	1 829	3 658
2017	9	14	14	873	2 652	14	14	235	928	1 971	4 234
2018	9	11	12	965	2 785	11	11	176	453	1 230	4 374
2019	–	–	–	–	–	13	14	196	470	1 318	4 273
2020	10	16	19	1 551	5 174	15	15	174	526	1 490	5 072

Table A1.2: Number of vessels and trips by statistical area in either the CRA 2 catch sampling or logbook programmes each fishing year. The fishing years indicated in grey were not included in the categorical length models.

Fishing year	905		906		907		908	
	vessels	trips	vessels	trips	vessels	trips	vessels	trips
1986	–	–	–	–	1	2	1	1
1990	–	–	–	–	–	–	2	3
1991	–	–	–	–	–	–	1	2
1993	3	80	15	1 187	3	191	5	457
1994	8	271	15	805	6	262	6	275
1995	1	43	12	635	6	194	7	176
1996	2	43	9	537	7	154	3	168
1997	1	21	10	431	6	181	3	133
1998	3	102	11	454	4	67	3	158
1999	4	89	11	499	6	231	3	142
2000	3	63	11	534	5	189	3	177
2001	2	71	8	382	4	167	4	134
2002	3	90	10	704	5	195	5	222
2003	2	91	8	582	3	159	4	159
2004	2	97	10	811	3	34	5	214
2005	1	70	11	872	4	236	5	276
2006	1	30	9	606	3	186	6	247
2007	2	3	8	504	2	183	8	410
2008	1	22	9	628	2	144	5	499
2009	1	20	8	577	3	200	6	412
2010	3	96	9	717	5	185	7	486
2011	2	132	11	722	3	185	5	335
2012	2	91	8	699	4	208	5	517
2013	1	118	9	680	4	304	6	495
2014	1	93	9	728	7	361	3	203
2015	2	162	8	565	5	332	4	296
2016	2	119	6	294	6	301	4	158
2017	2	143	7	309	4	220	6	270
2018	2	30	5	206	4	141	4	88
2019	1	66	4	136	3	127	5	141
2020	1	59	6	189	4	146	6	151

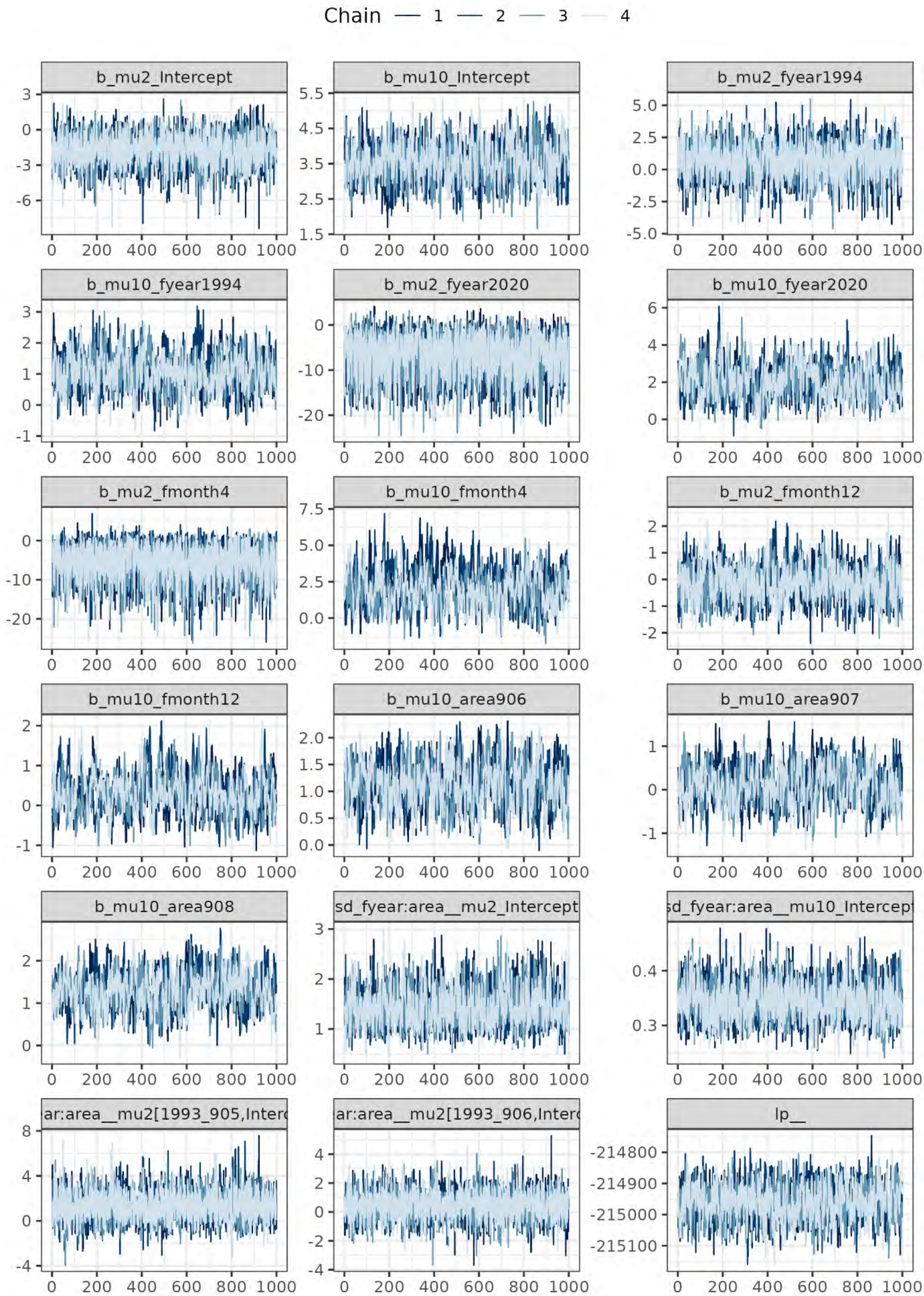

Figure A1.1: CRA 2 male length frequency categorical models MCMC trace plots.

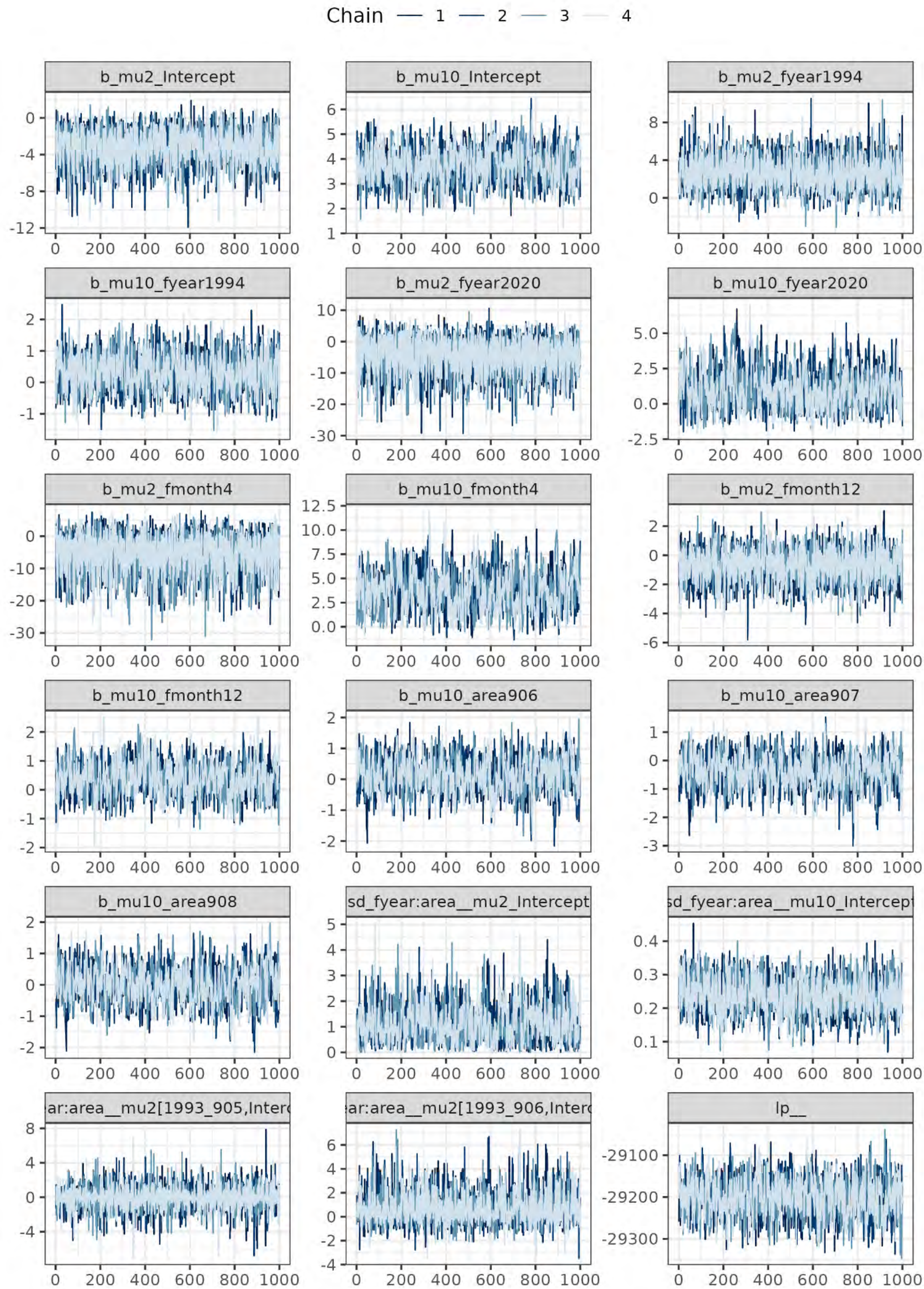

Figure A1.2: CRA 2 immature female length frequency categorical models MCMC trace plots.

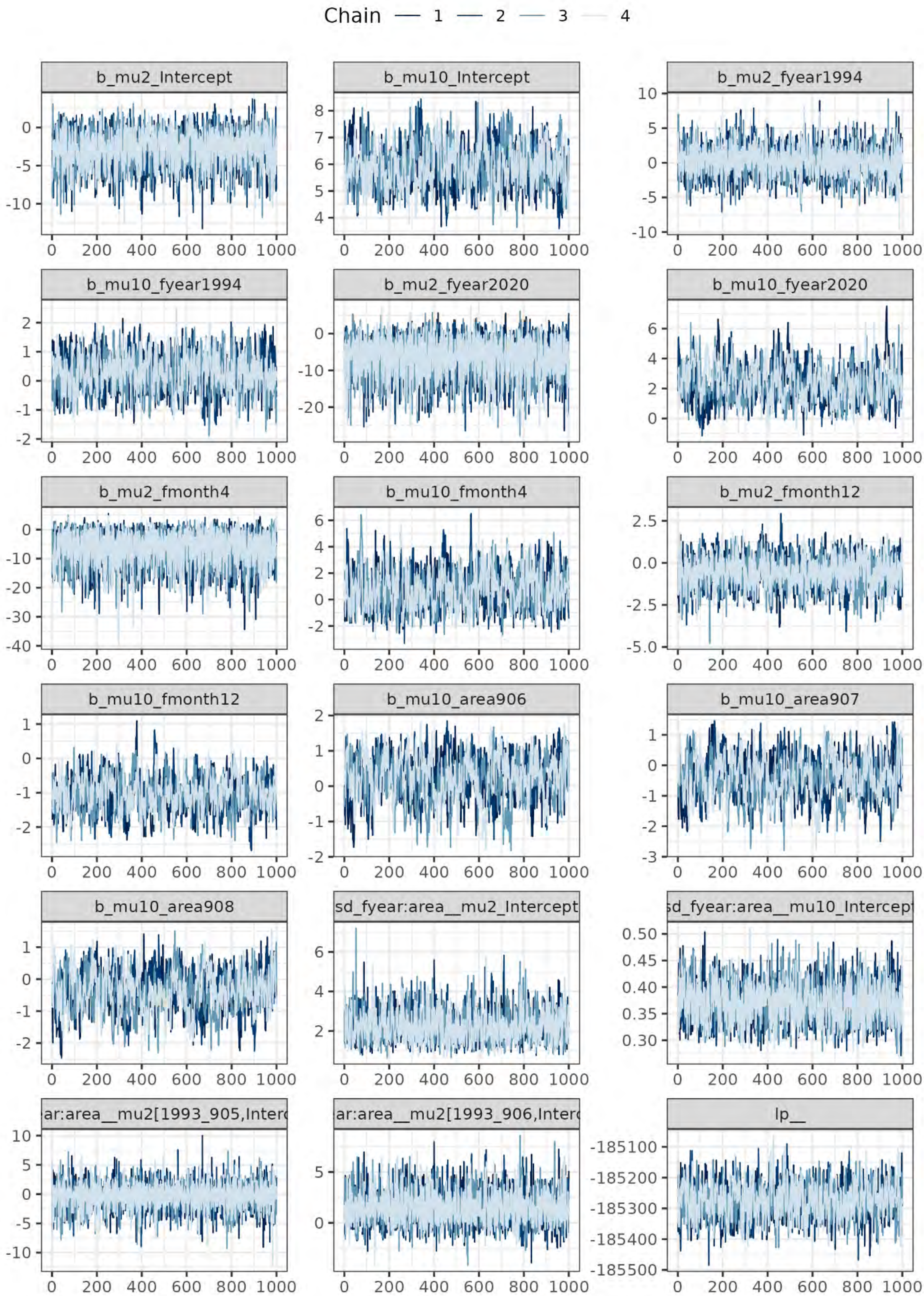

Figure A1.3: CRA 2 mature female length frequency categorical models MCMC trace plots.

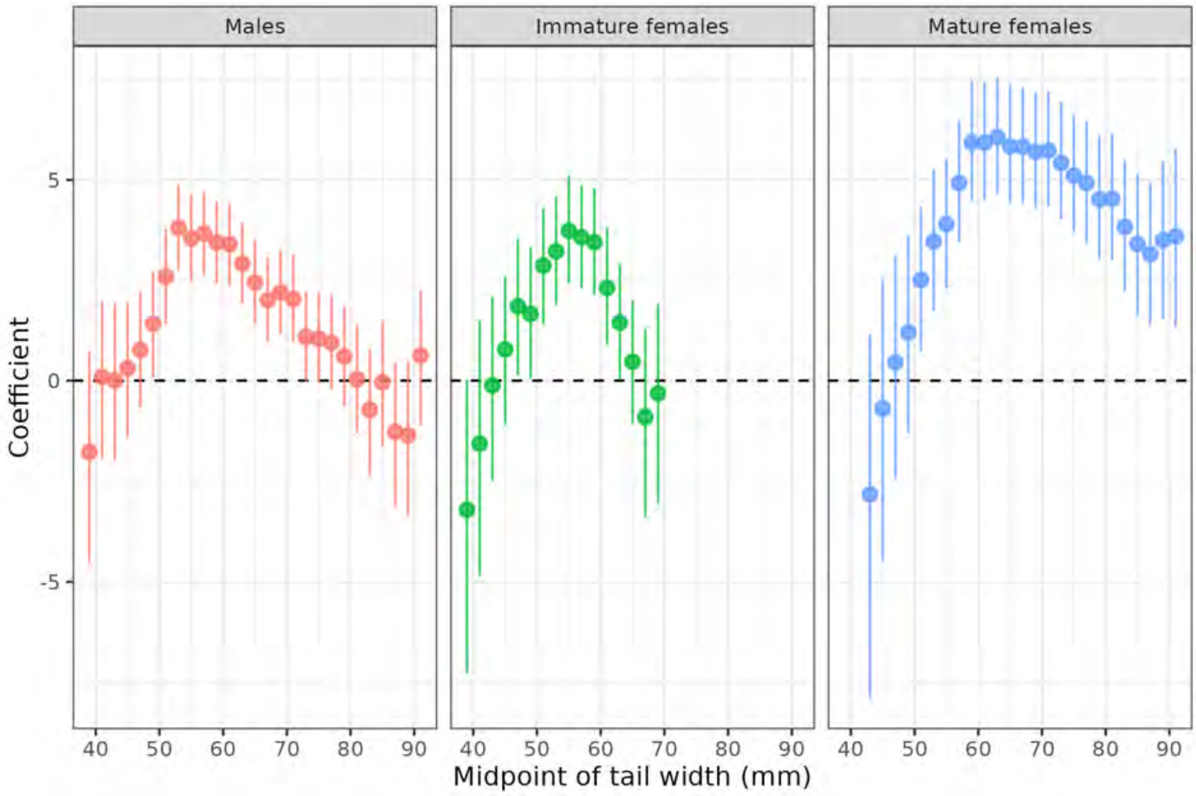

Figure A1.4: The population level effect of intercept by TW bin for each sex category, showing the mean and 95% credible intervals.

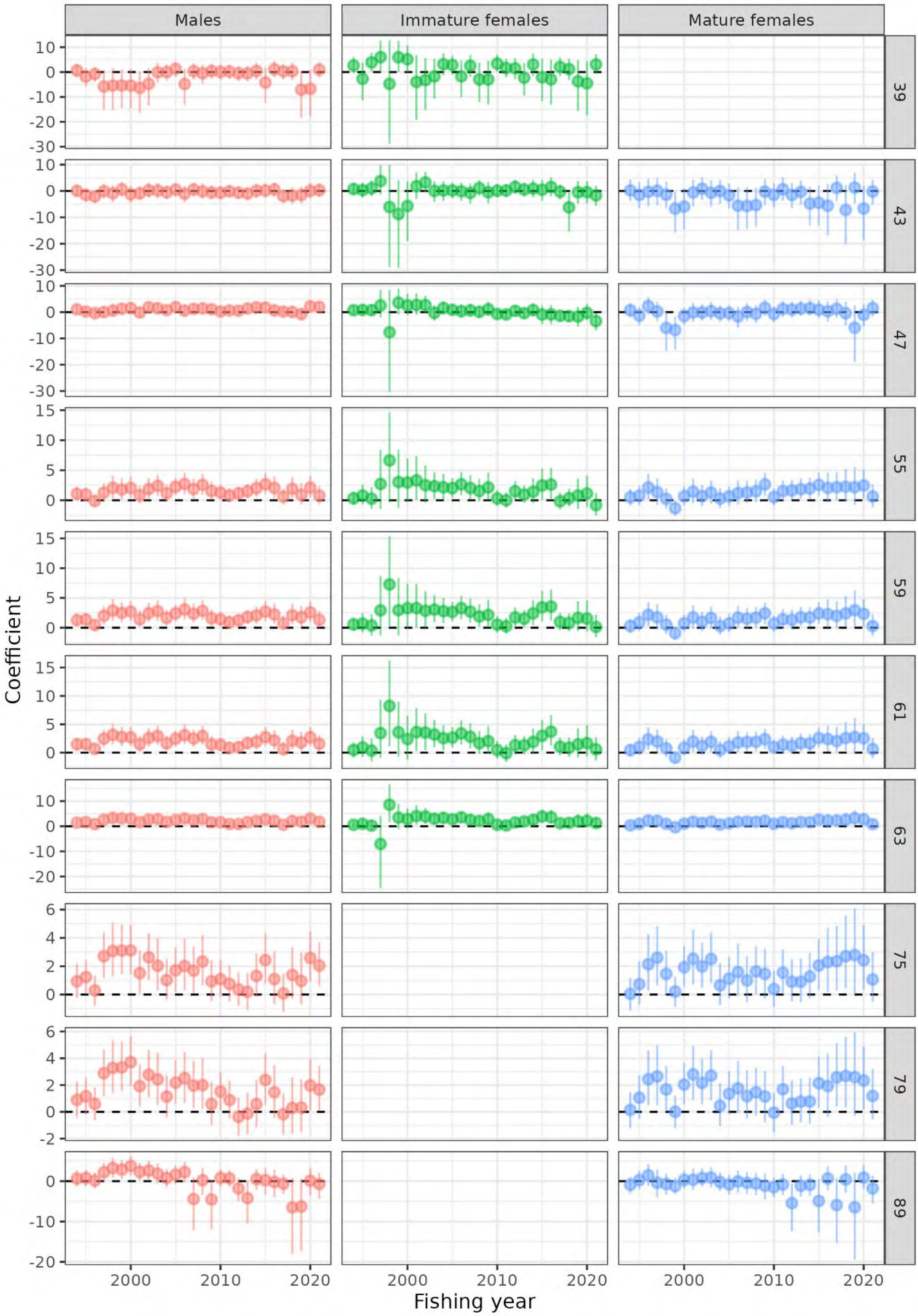

Figure A1.5: The population level effect of fishing year by TW bin for each sex category, showing the mean and 95% credible intervals.

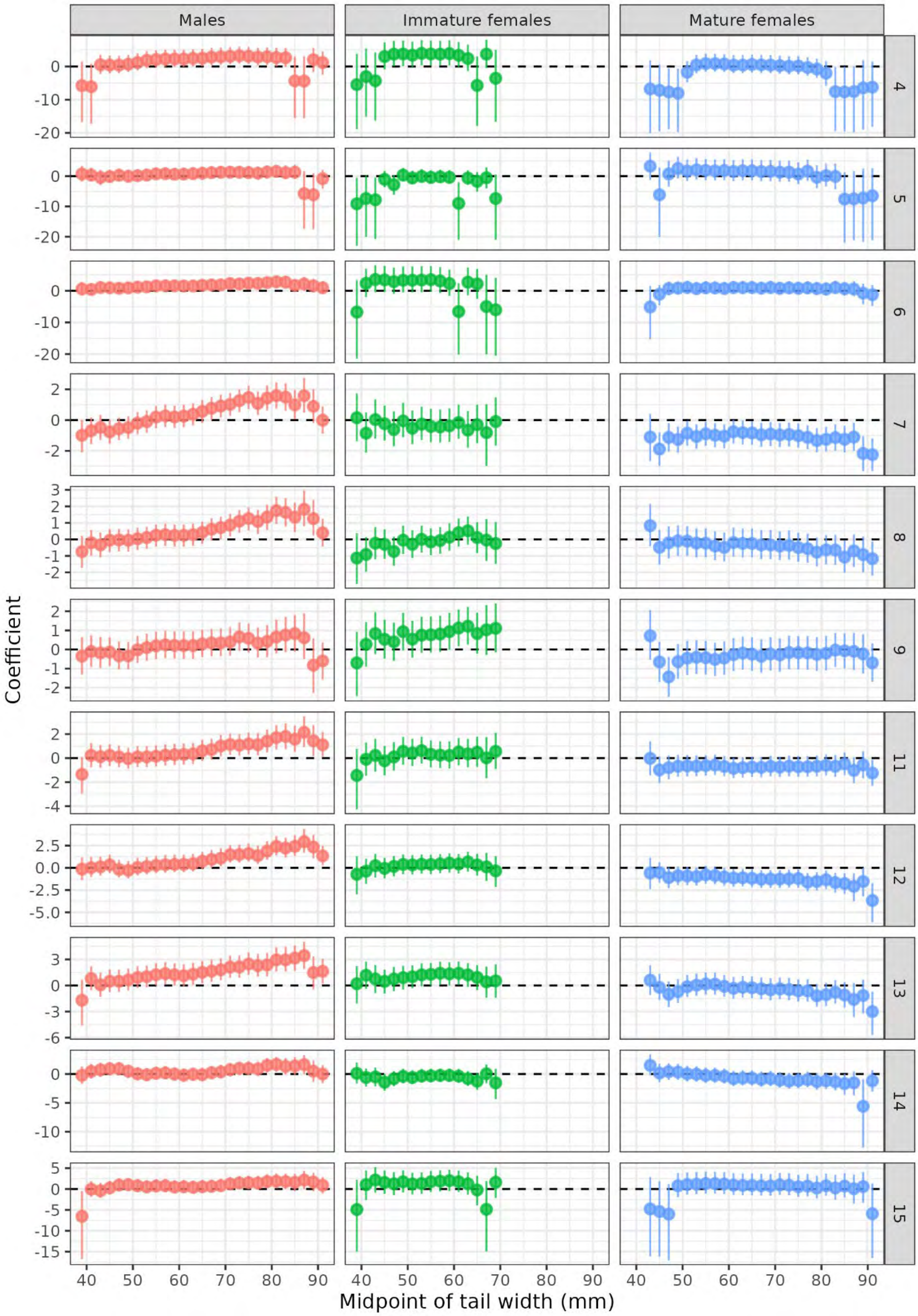

Figure A1.6: The population level effect of fishing month (4 = April, 15 = March) by TW bin for each sex category, showing the mean and 95% credible intervals.

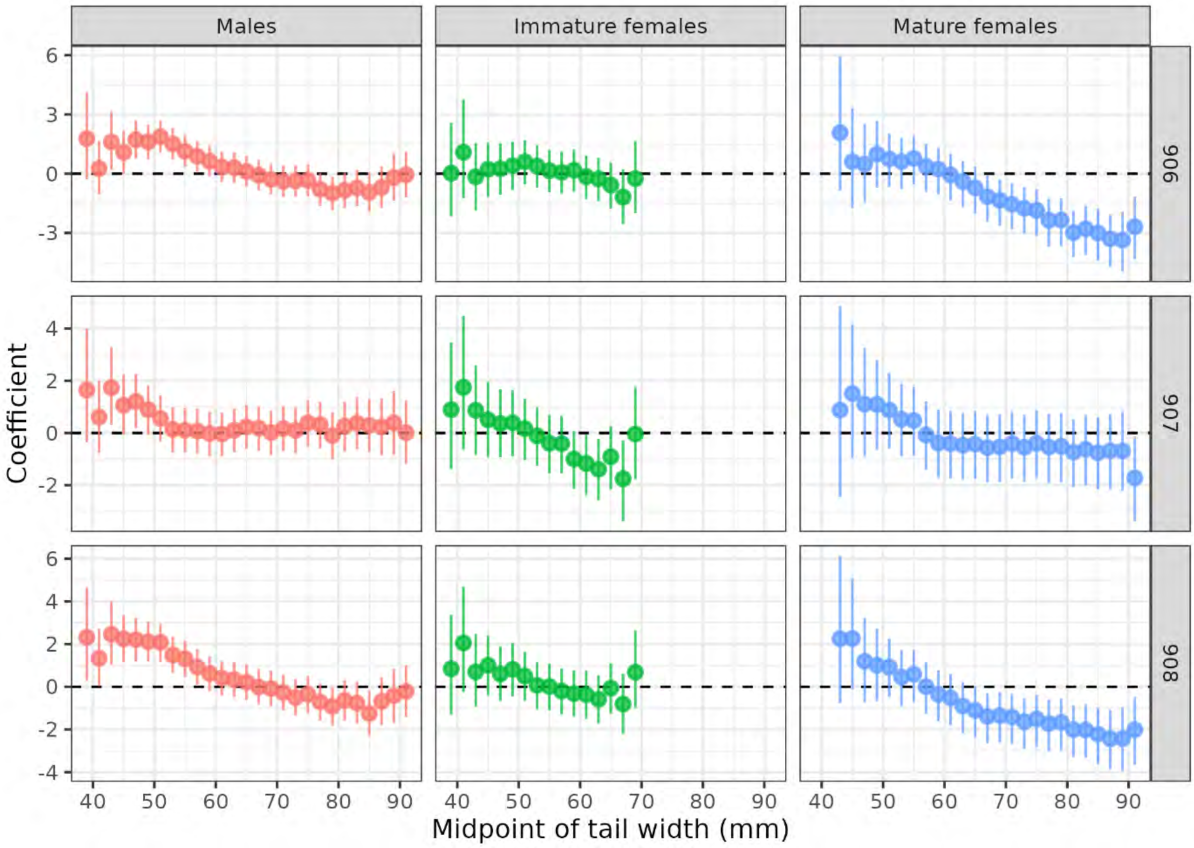

Figure A1.7: The population level effect of statistical area by TW bin for each sex category, showing the mean and 95% credible intervals. Statistical area 905 is not shown as it is the reference level.

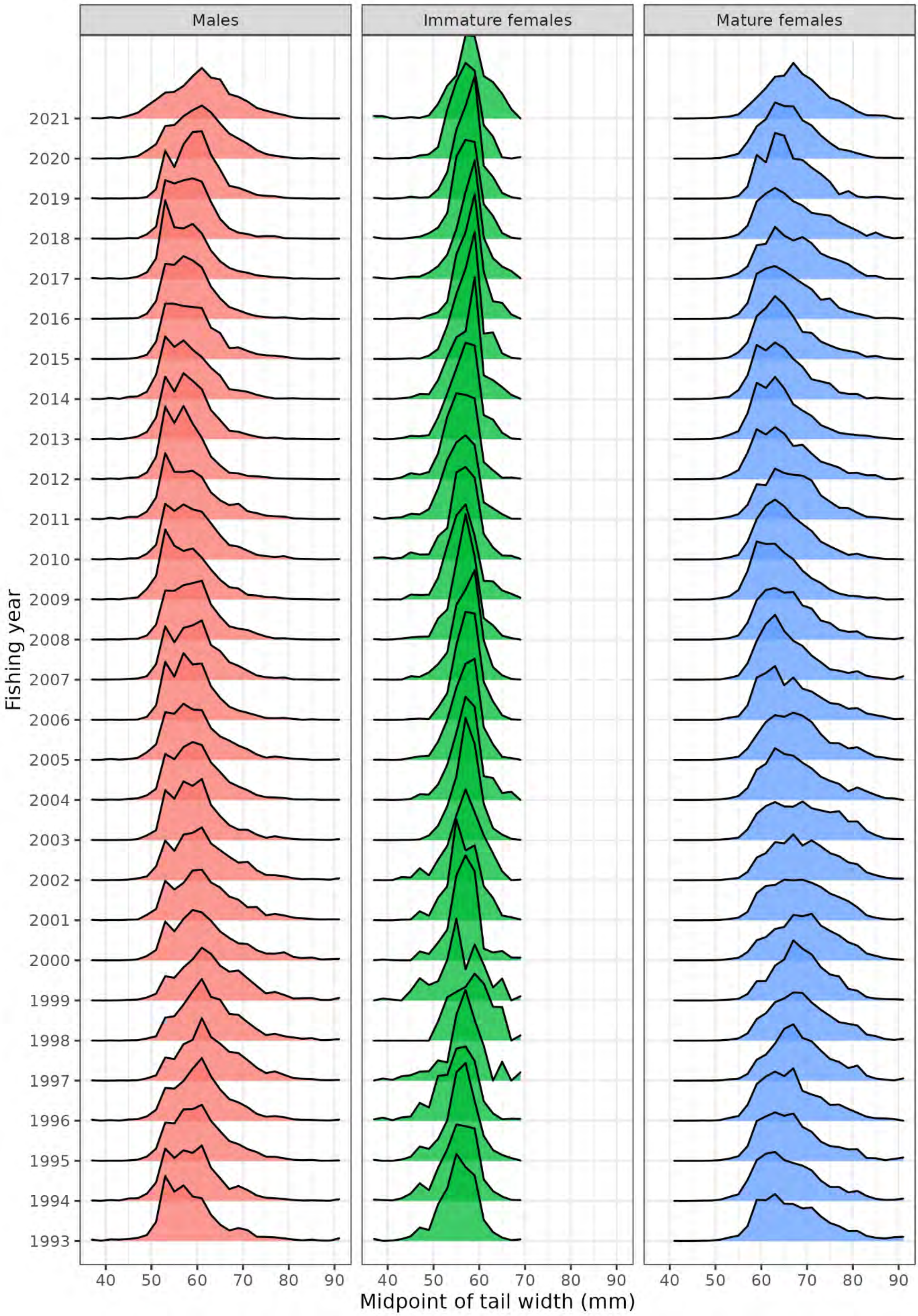

Figure A1.8: The conditional effect of fishing year by TW bin for each sex category.

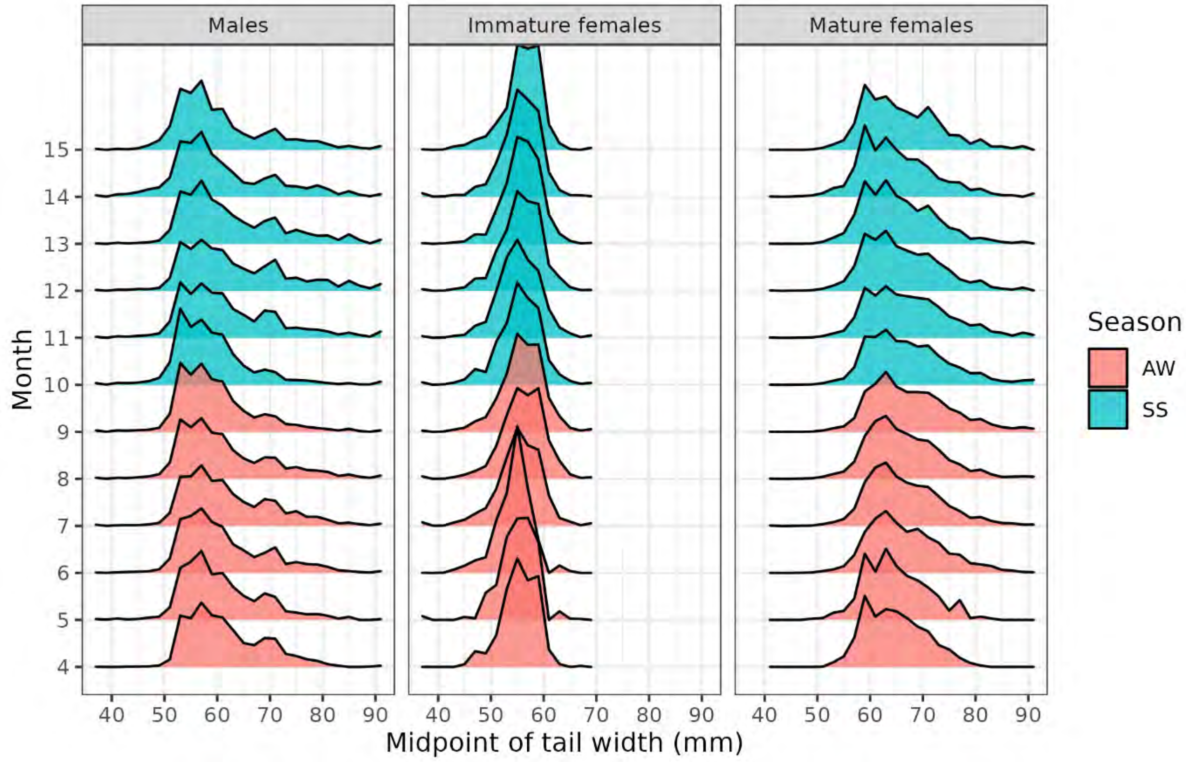

Figure A1.9 The conditional effect of fishing month (4 = April, 15 = March) by TW bin for each sex category and season (AW = autumn/winter, SS = spring/summer).

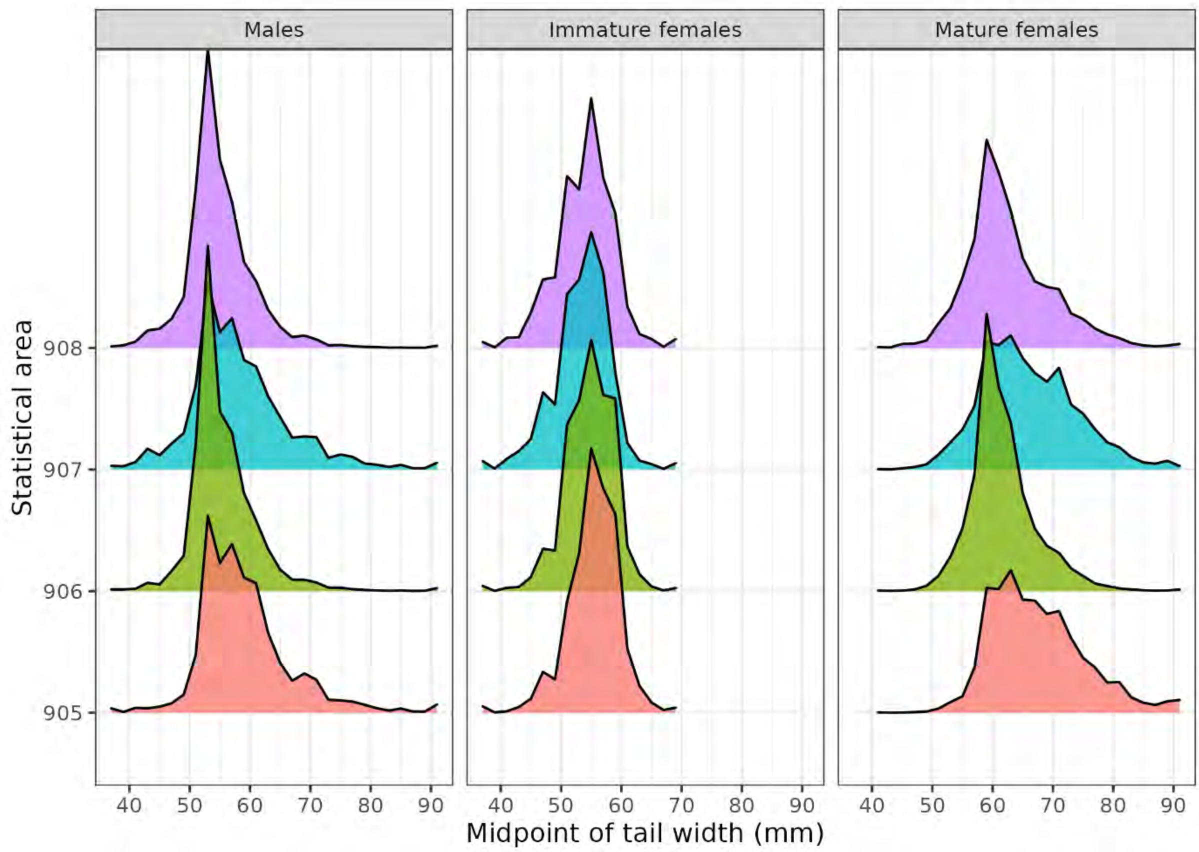

Figure A1.10: The conditional effect of statistical area by TW bin for each sex category.

8.2 Hake in HAK 7 and ling in LIN 3&4

Table A1.3: Number of vessels, skippers, trips, days sampled, sets, and fish measured in the observer catch sampling programme each fishing year for hake and ling. The fishing years indicated in grey were not included in the categorical length models.

Fishing year	Hake					Ling				
	vessels	days	trips	sets	fish	vessels	days	trips	sets	fish
1986	1	1	1	1	49	1	5	1	5	420
1987	8	28	8	47	2 589					
1988	11	41	11	57	3 239					
1989	7	44	7	115	4 100	2	4	2	7	34
1990	6	52	6	71	1 645	4	35	4	44	1 074
1991	11	57	11	168	5 035	7	65	7	97	2 546
1992	12	58	12	147	4 593	7	65	7	155	4 966
1993	11	55	13	124	5 368	14	63	14	100	1 553
1994	15	61	17	228	4 022	12	49	13	65	880
1995	9	67	10	171	10 001	9	79	11	127	1 551
1996	13	73	13	209	6 909	9	59	10	87	1 039
1997	11	71	12	284	8 216	11	69	11	134	1 233
1998	15	66	15	271	5 827	19	199	21	578	11 618
1999	16	71	16	340	10 546	17	142	18	328	5 841
2000	17	86	19	386	7 277	14	129	16	260	2 515
2001	21	76	26	322	6 803	21	122	23	185	2 859
2002	14	68	15	492	7 973	18	124	22	223	3 234
2003	13	78	13	414	7 129	16	91	18	153	1 585
2004	15	71	16	503	8 354	12	127	13	230	2 138
2005	12	83	13	266	9 229	12	127	12	272	3 020
2006	15	106	20	419	20 830	14	89	16	164	2 822
2007	13	50	14	134	9 386	13	82	18	135	1 886
2008	11	81	14	148	9 167	12	108	18	304	5 862
2009	14	91	16	183	11 964	14	124	16	235	3 373
2010	10	48	12	79	5 145	7	58	10	71	2 322
2011	11	62	13	104	5 648	7	71	9	114	1 892
2012	16	67	20	143	6 096	12	115	19	146	3 956
2013	17	121	47	518	24 278	14	125	24	156	2 788
2014	14	117	36	368	19 192	11	87	17	133	2 470
2015	18	114	41	505	28 087	10	66	16	89	2 195
2016	17	95	33	251	12 179	7	48	8	58	1 039
2017	21	113	33	411	21 265	14	108	18	134	3 136
2018	20	121	33	405	11 322	9	77	13	100	2 000
2019	16	61	23	157	3 249	15	101	23	133	2 471
2020	15	73	23	207	9 944	10	101	15	145	3 593
2021	11	61	13	132	5 188	6	30	9	32	626

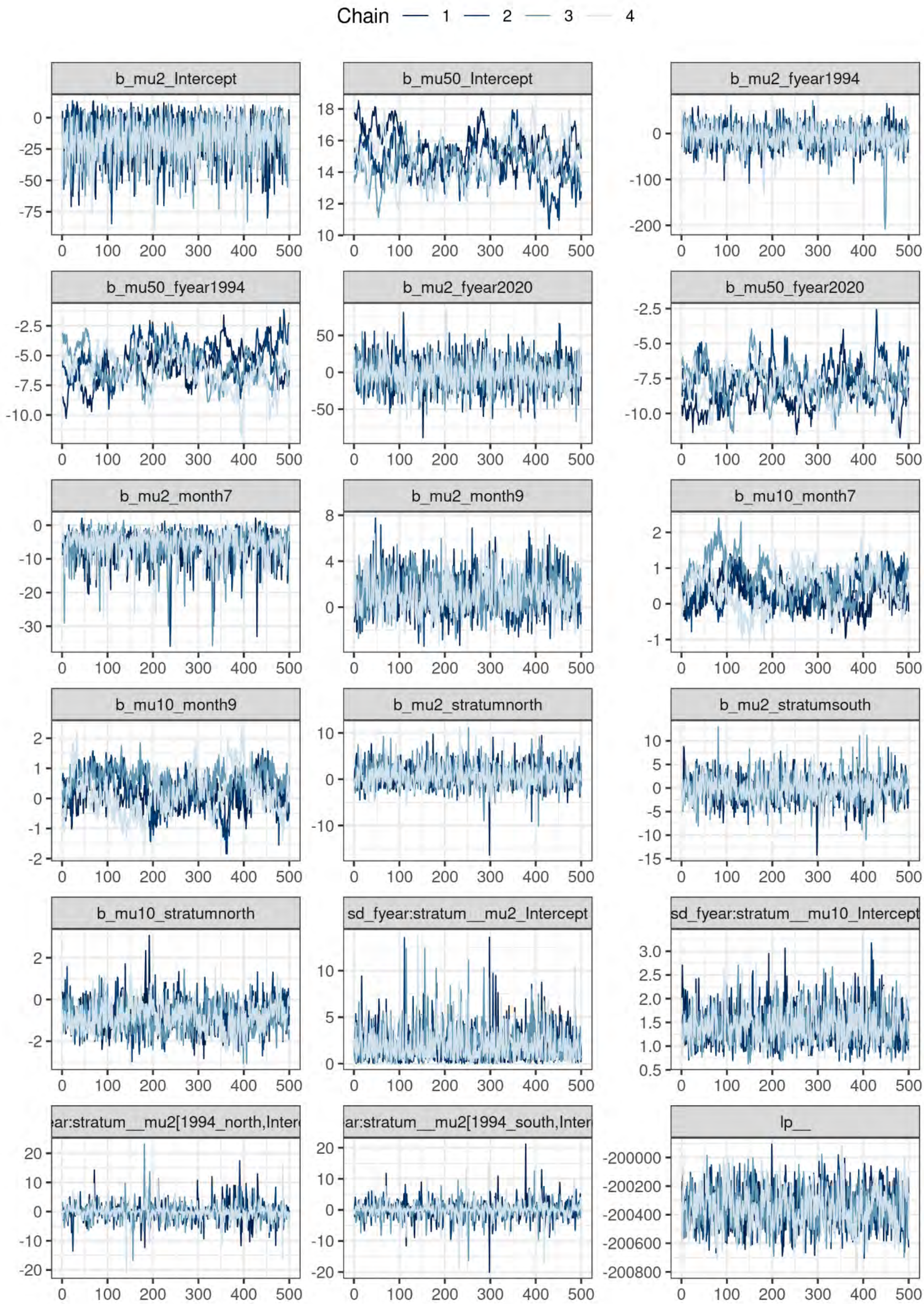

Figure A1.11: HAK LFs: MCMC trace plots for a random selection of parameters from the best female length frequency categorical model.

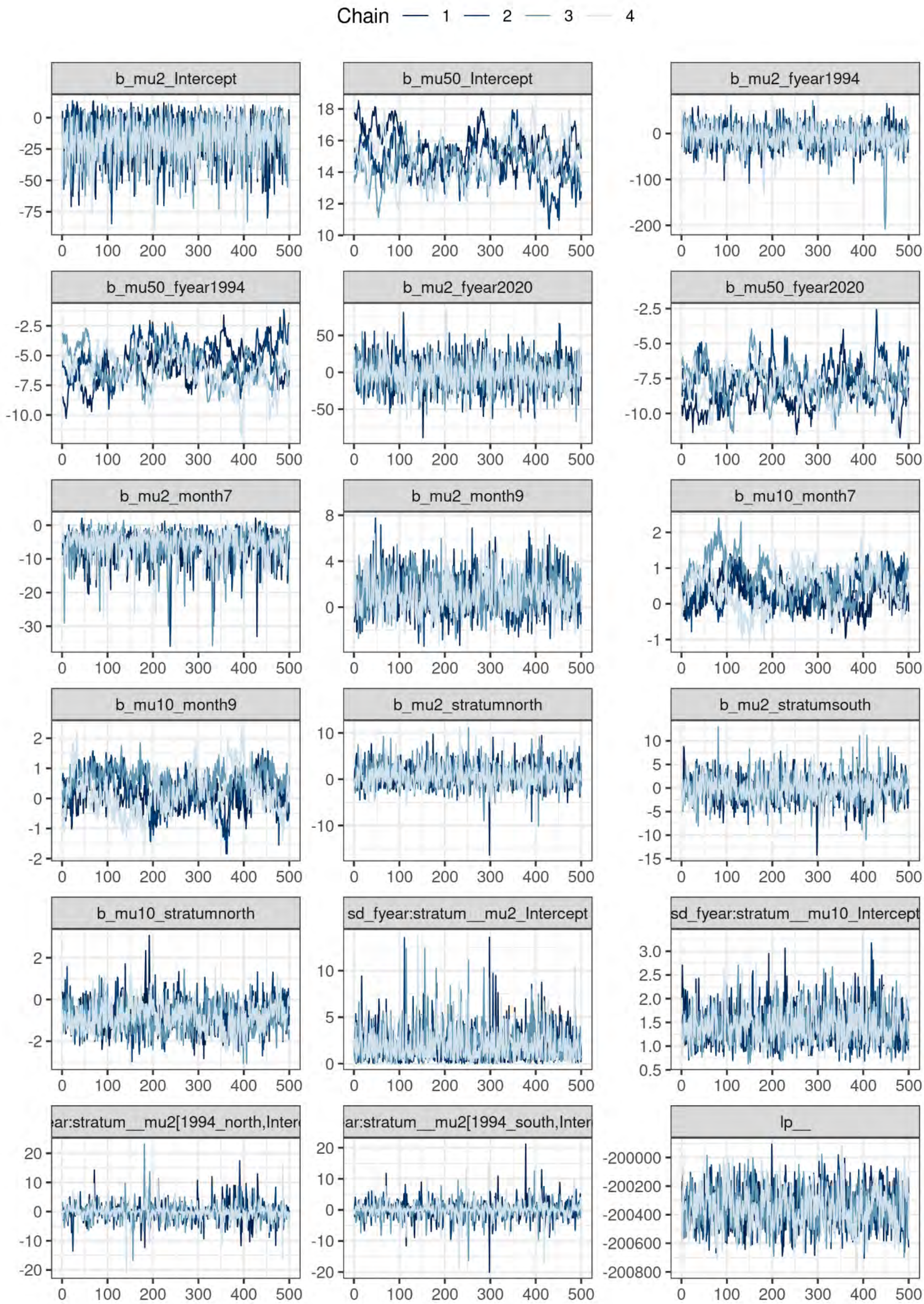

Figure A1.12: LIN LFs: MCMC trace plots for a random selection of parameters from the best female length frequency categorical model.

9. APPENDIX 2: TABLES AND FIGURES FOR THE SCALED LENGTH COMPOSITIONS

9.1 Red rock lobster in CRA 2

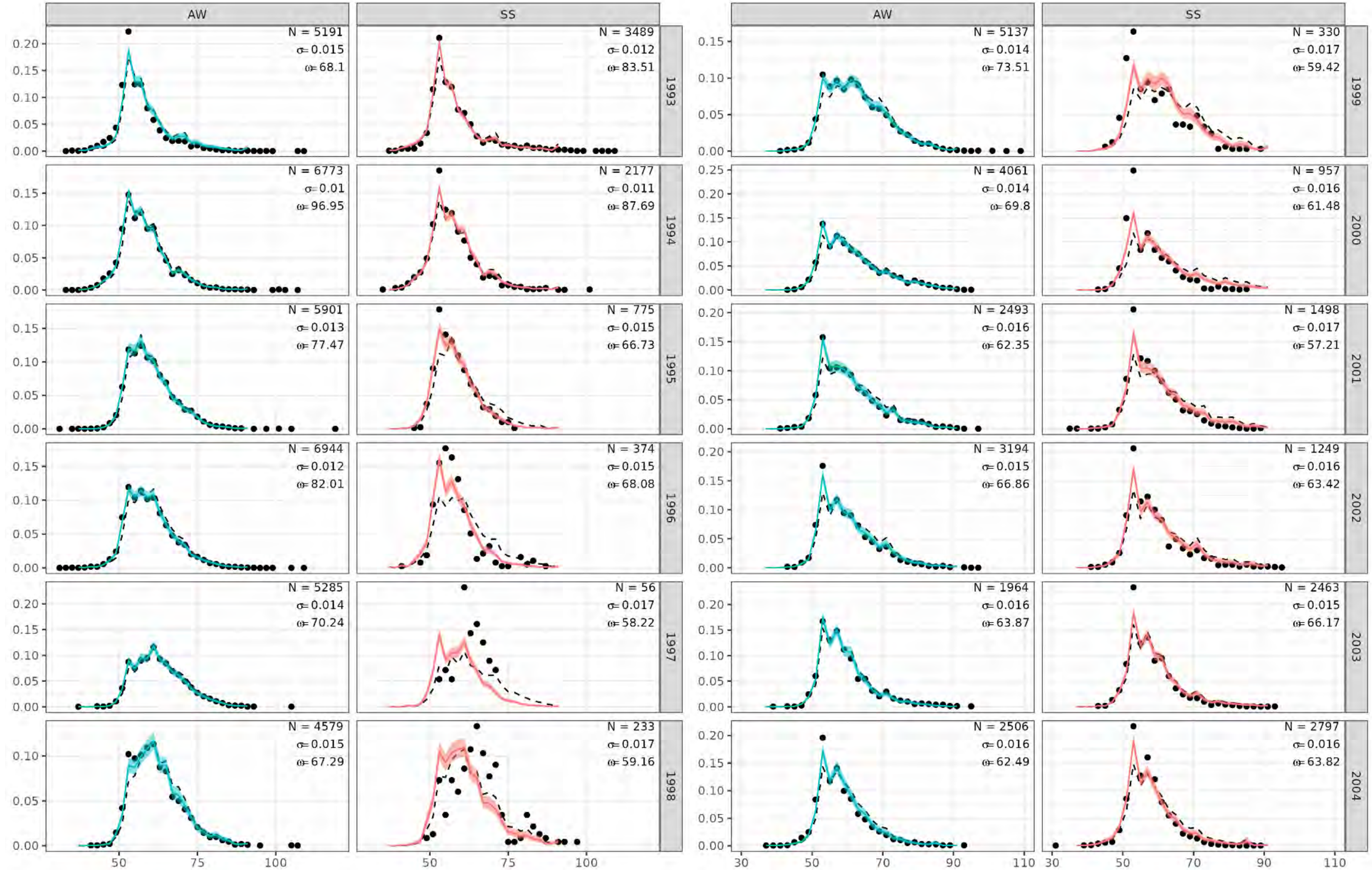

Figure A2.1: Males 1993 to 2004: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

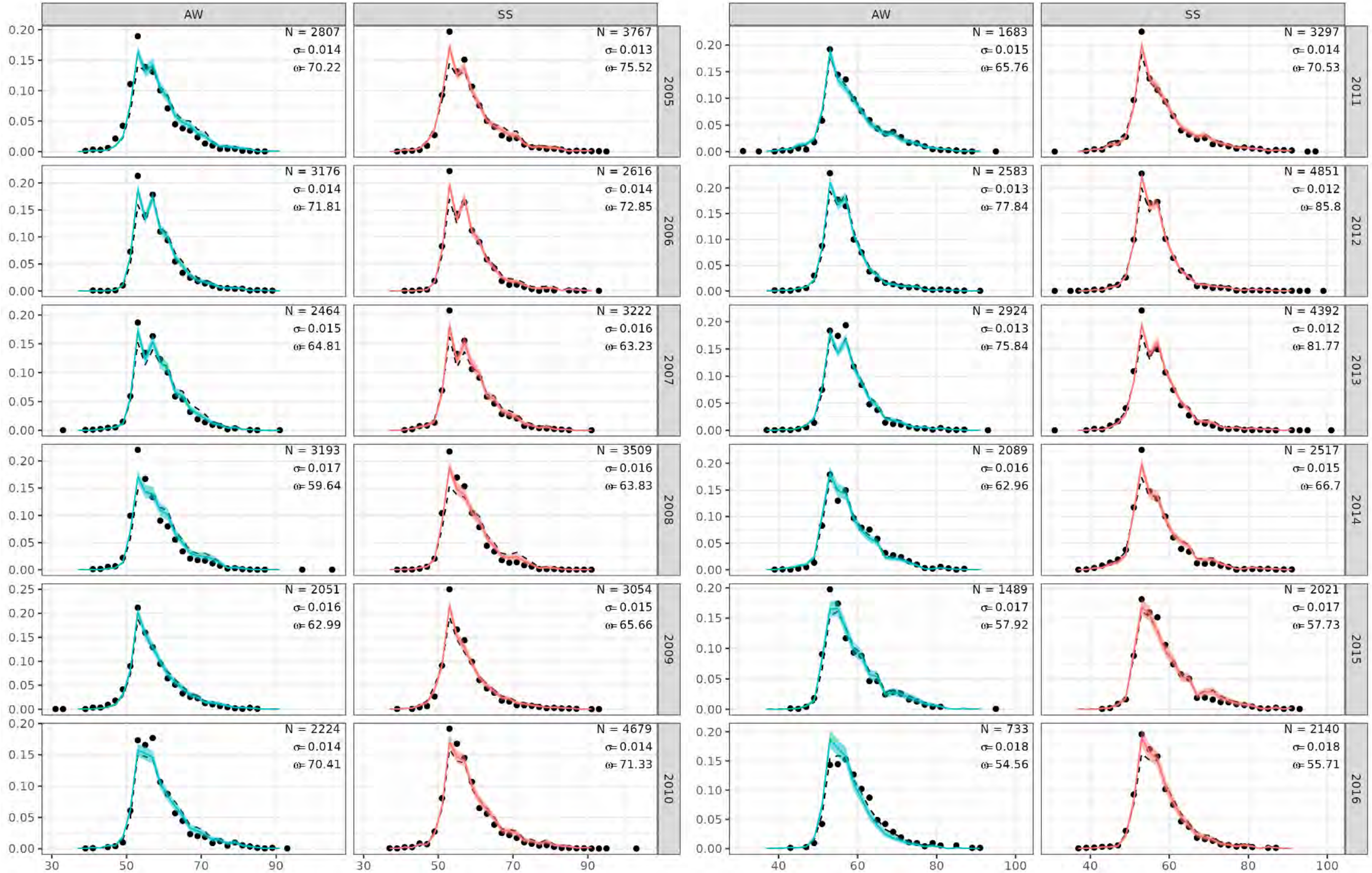

Figure A2.2: Males 2005 to 2016: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

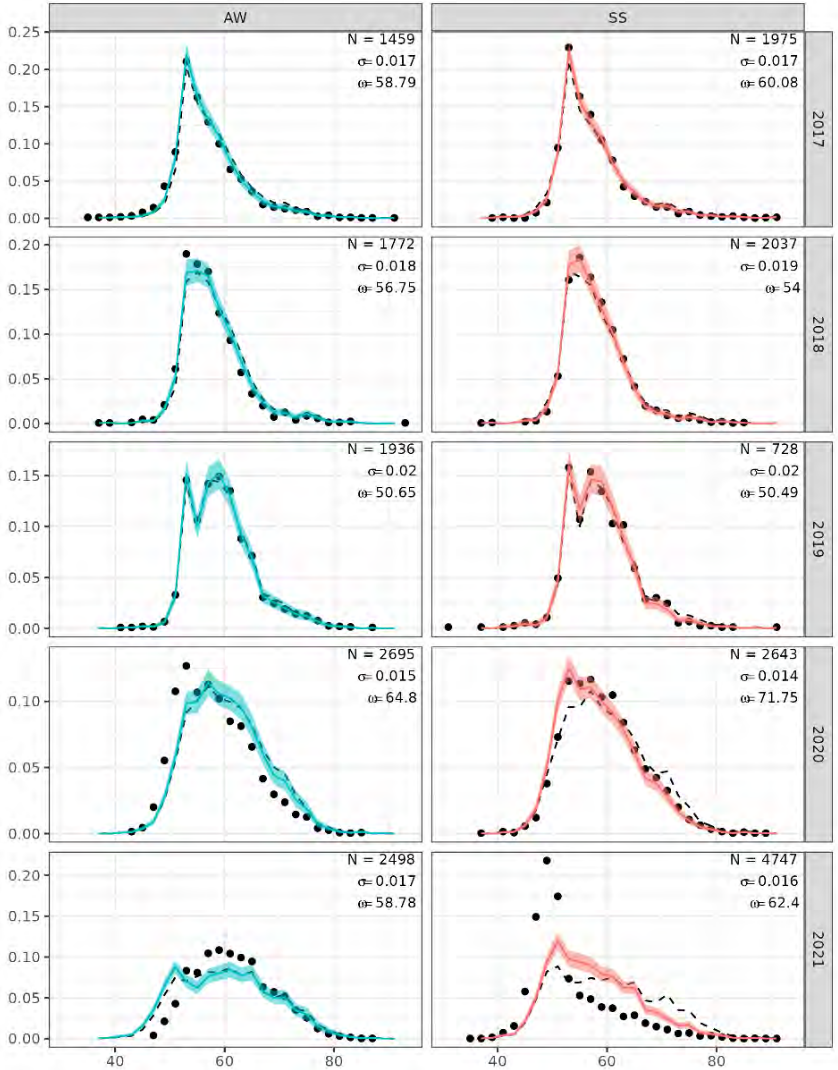

Figure A2.3: Males 2017 to 2021: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

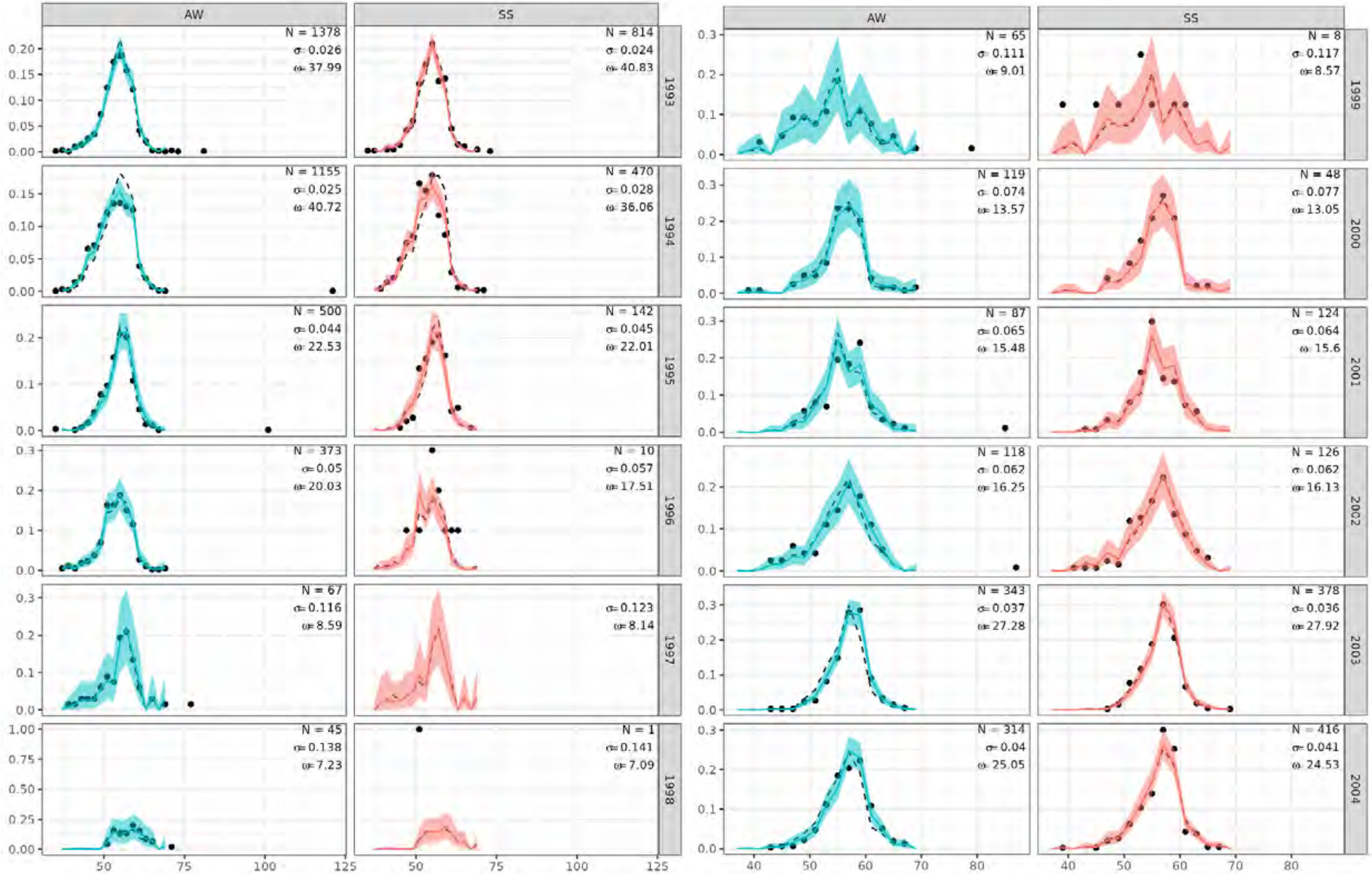

Figure A2.4: Immature females 1993 to 2004: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

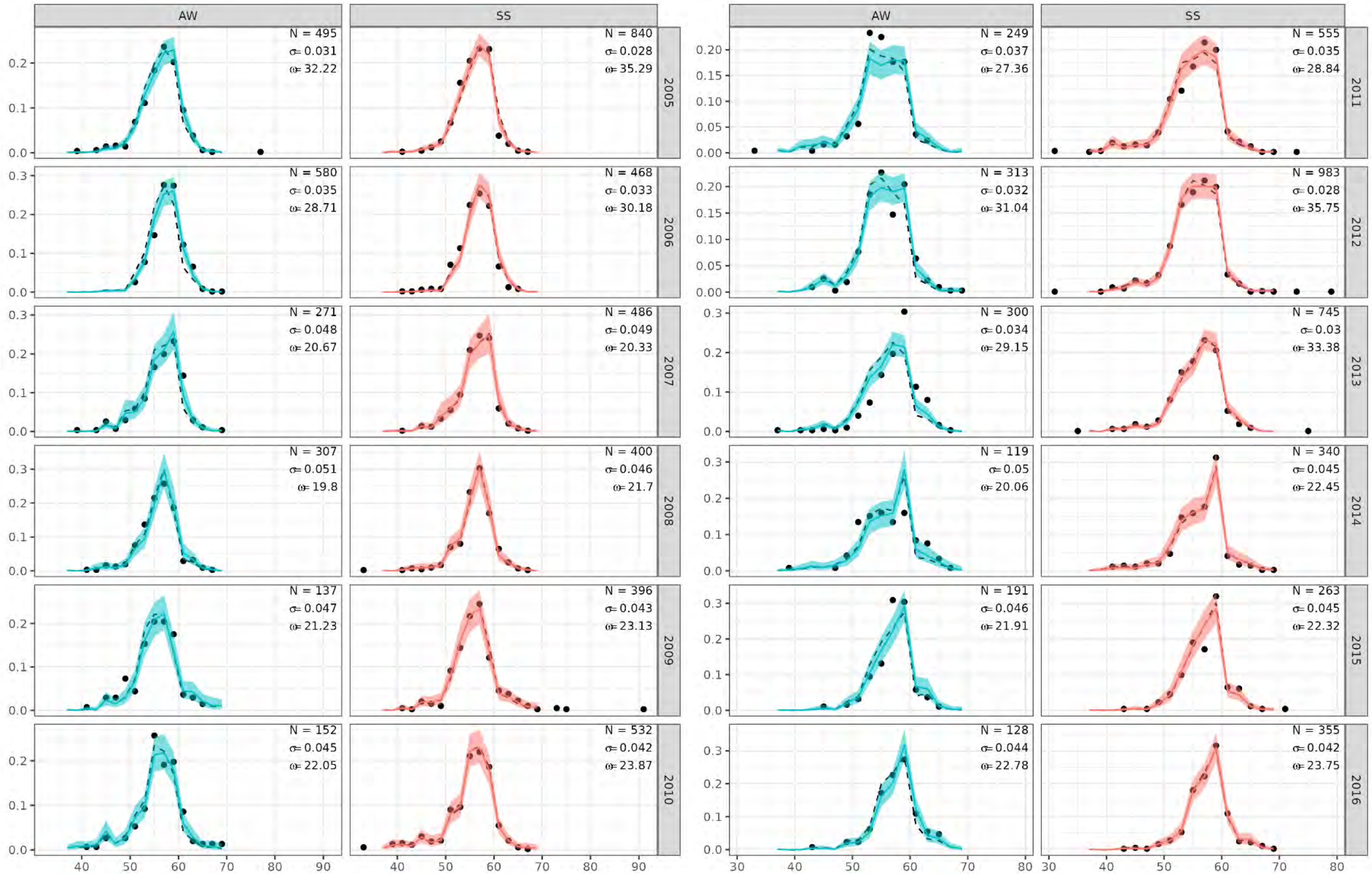

Figure A2.5: Immature females 2005 to 2016: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

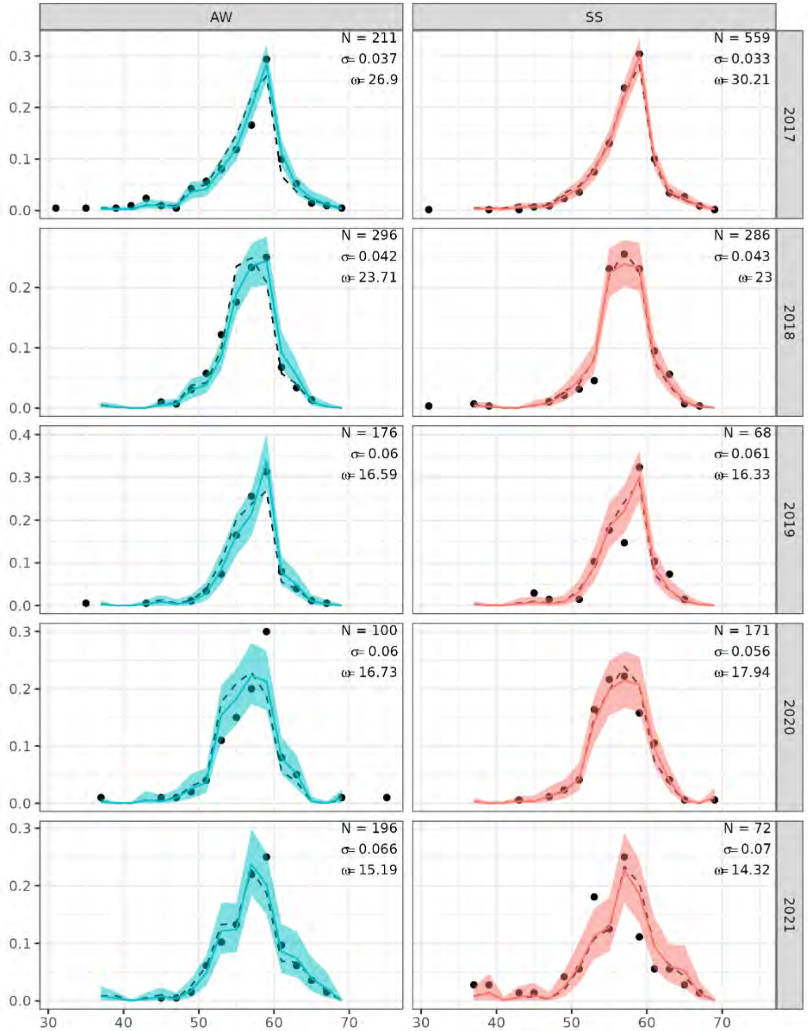

Figure A2.6: Immature females 2017 to 2021: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

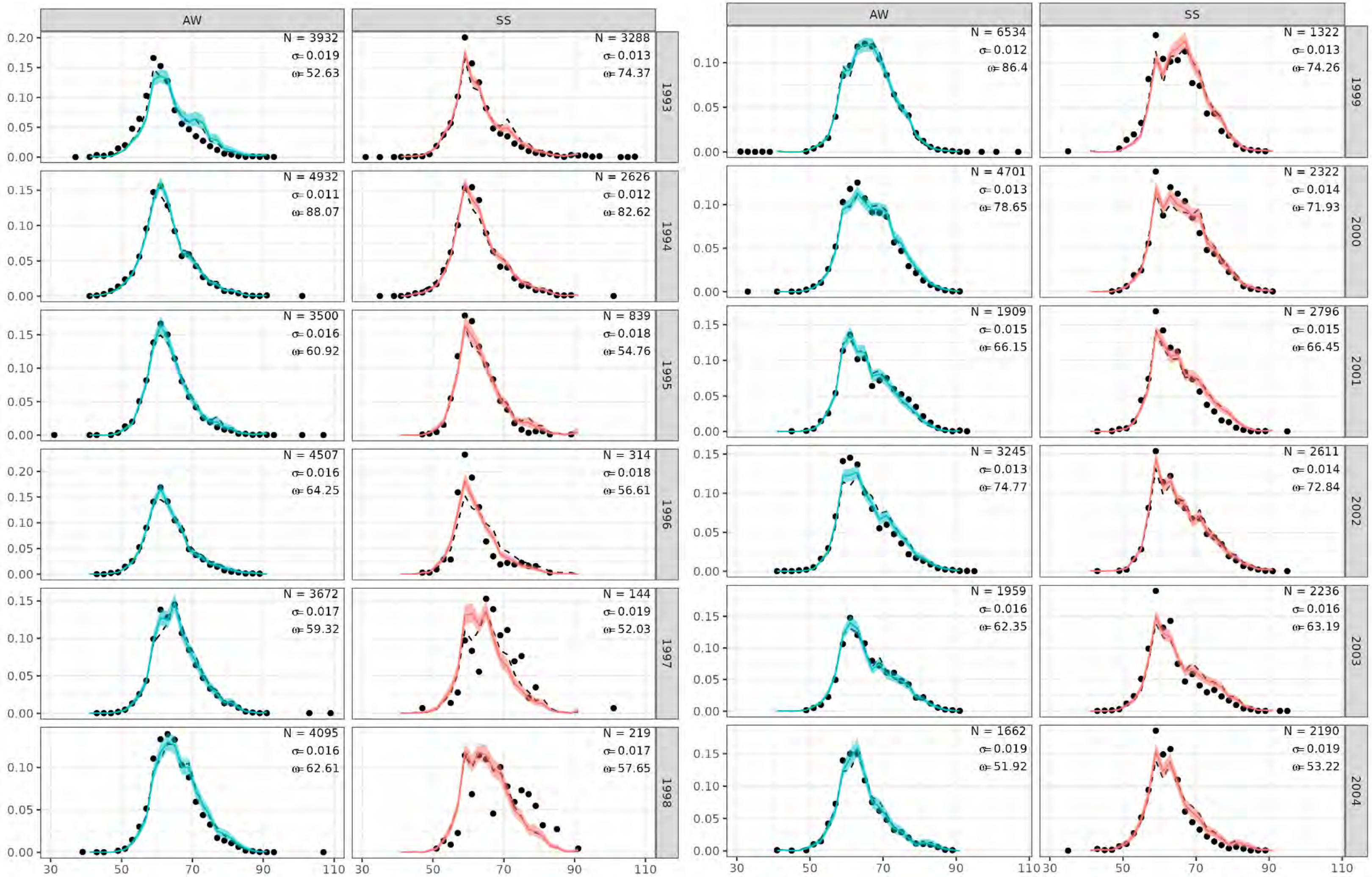

Figure A2.7: Mature females 1993 to 2004: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

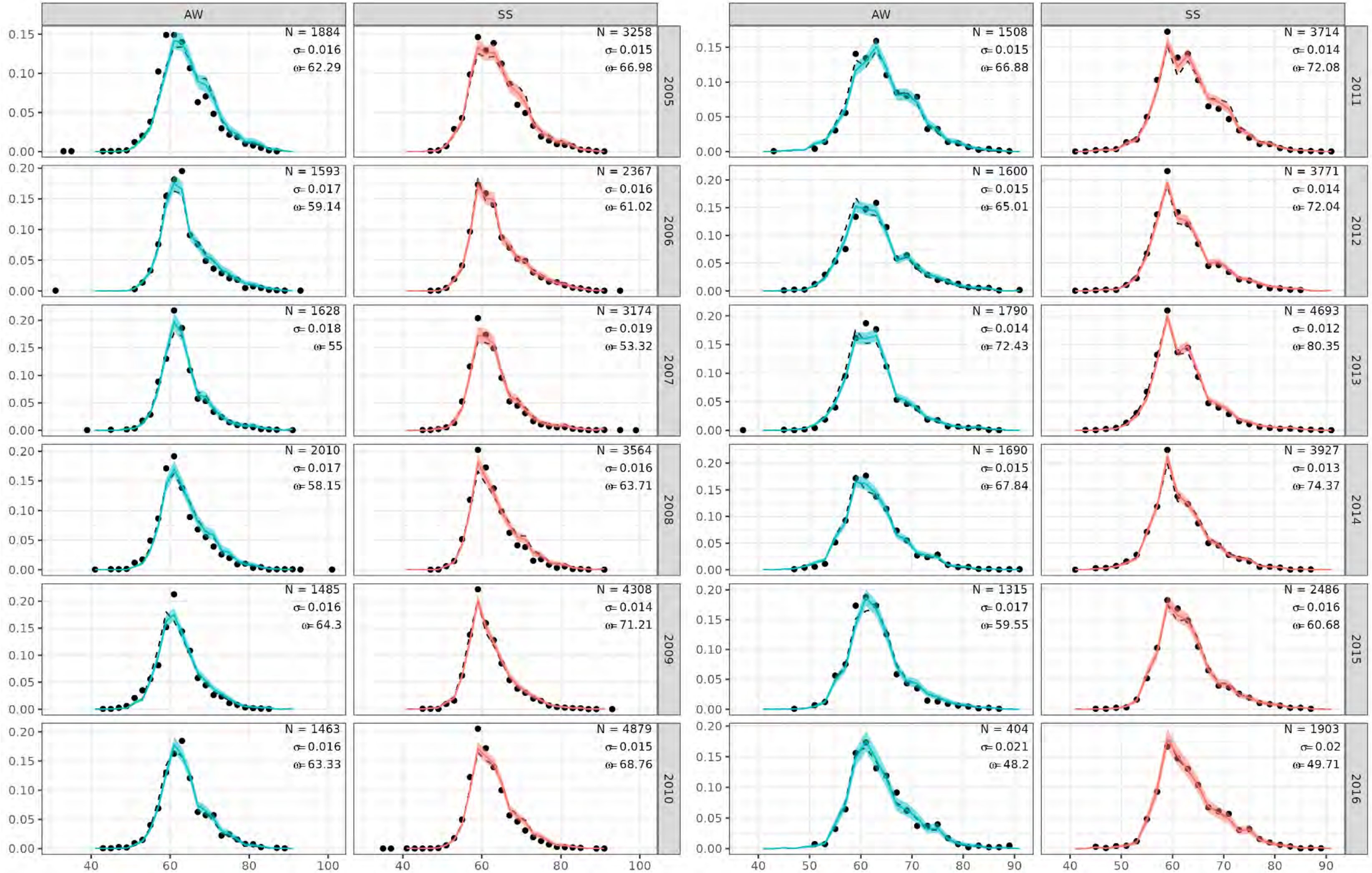

Figure A2.8: Mature females 2005 to 2016: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

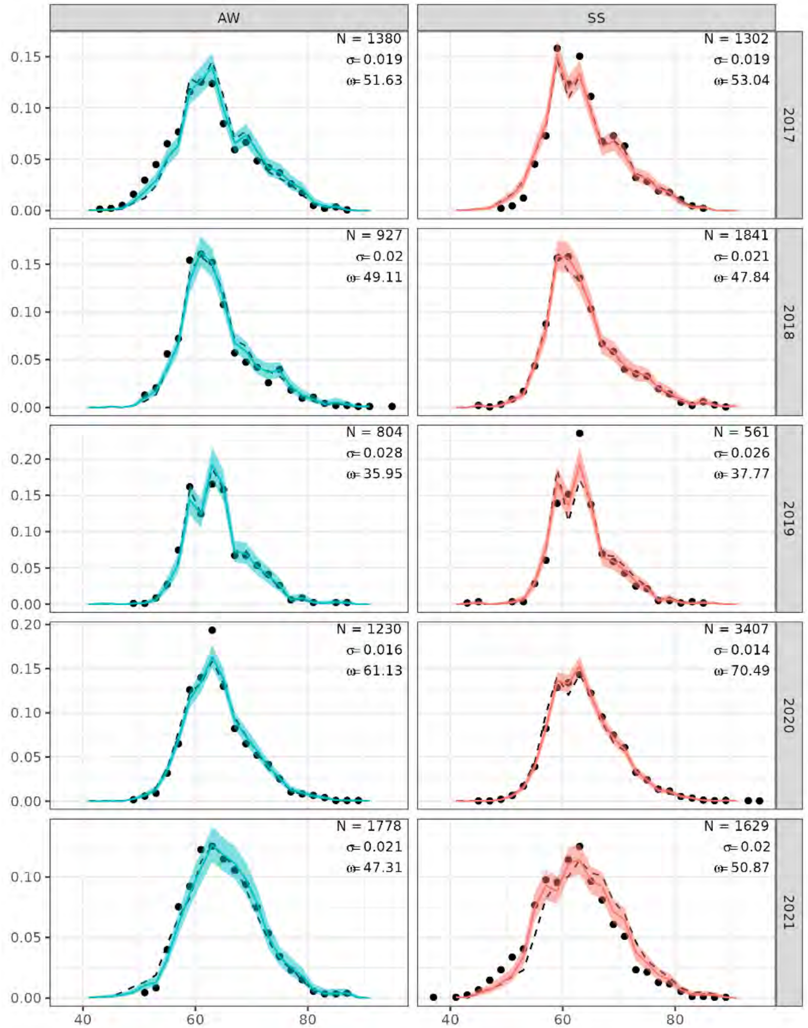

Figure A2.9: Mature females 2017 to 2021: scaled LF distributions by fishing year, season, and TW bin where the solid line is the mean and the shaded region is the 95% credible interval of the posterior distribution of the mean. The aggregated data (points) and unscaled predictions of the LF (dashed line) for the same stratum are also shown. Each panel also includes text which provides: the number of lobsters measured; the SD of the predicted LF for each stratum; and the data set weight for each stratum.

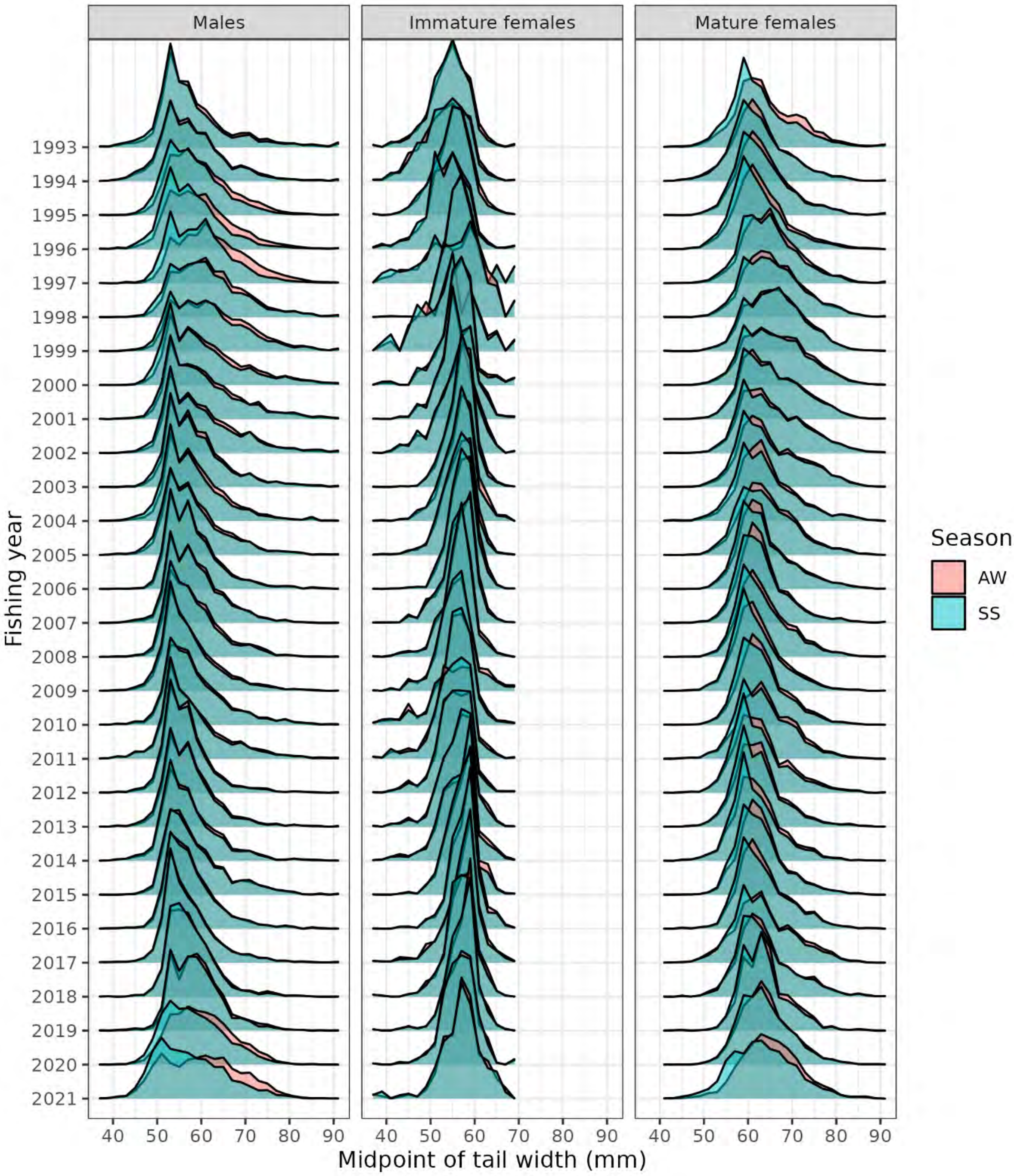

Figure A2.10: CRA 2 standardised length frequency distributions of the catch, by sex category, fishing year, and season (AW = autumn/winter, SS = spring/summer).

10. APPENDIX 3: TABLES AND FIGURES FOR THE AGE AT LENGTH ORDINAL MODELS

10.1 Hake in HAK 7

Table A3.1: HAK females: count of the number of aged female fish by age and fishing year.

Fishing year	Age																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total
1990	–	–	3	12	4	17	35	25	25	27	18	17	21	15	7	11	8	5	–	1	–	–	1	–	252
1991	–	2	3	24	24	30	37	60	38	32	21	15	12	10	7	4	4	2	–	–	1	–	–	–	326
1992	–	7	2	2	2	12	23	39	40	28	23	13	13	13	7	5	1	1	1	–	–	–	–	–	232
1993	8	11	2	1	2	5	7	13	15	14	8	12	8	6	5	6	–	1	2	–	–	–	1	–	127
1997	4	4	6	13	14	12	24	23	17	14	3	10	6	8	3	1	2	1	1	1	–	2	1	–	170
1998	3	5	4	8	31	21	36	29	27	18	9	9	6	2	3	3	3	3	2	–	–	1	–	2	225
1999	3	3	5	7	4	13	16	30	36	25	22	14	10	5	2	5	7	5	2	1	1	1	2	2	221
2000	–	2	4	4	3	6	15	34	46	33	27	24	22	11	8	4	6	6	2	3	–	–	–	–	260
2001	–	6	3	3	2	4	23	40	35	44	42	19	21	14	5	8	3	–	1	–	–	1	–	1	275
2002	–	5	5	8	1	2	17	33	51	52	61	49	40	22	18	8	4	5	2	2	2	2	–	–	389
2003	2	11	9	23	14	21	26	23	18	15	24	16	13	3	2	5	4	1	–	1	2	1	–	–	234
2005	1	17	8	10	41	62	28	23	25	24	10	9	8	2	2	1	–	–	–	–	–	–	–	–	271
2006	–	34	6	6	17	26	49	29	37	26	12	21	8	6	4	2	–	2	–	–	–	–	–	–	285
2007	1	35	6	6	15	28	23	41	34	17	12	6	13	8	1	4	2	1	2	–	–	–	–	–	255
2008	–	9	8	7	18	37	34	38	20	18	10	19	6	8	4	3	1	–	–	–	1	–	–	–	241
2009	6	13	4	27	32	37	36	30	23	17	17	13	11	6	–	–	–	–	–	–	–	–	–	–	272
2010	–	–	–	2	12	32	20	7	6	8	3	3	2	1	–	–	1	–	1	–	–	–	–	–	98
2011	–	–	3	1	15	59	74	62	44	15	10	15	7	6	6	6	2	1	–	–	1	–	–	–	327
2012	1	8	12	15	16	55	65	75	49	31	17	12	8	5	6	7	4	2	1	2	–	–	1	–	392
2013	26	19	10	6	25	49	69	85	47	37	17	8	6	11	5	6	2	2	2	–	1	–	–	–	433
2014	12	10	1	2	9	40	77	72	48	27	13	13	2	1	5	3	2	–	1	–	–	–	–	–	338
2015	5	6	2	9	17	35	45	68	51	29	25	14	6	3	1	3	4	5	1	1	1	–	–	–	331
2016	–	93	28	8	15	54	64	66	32	25	21	12	6	6	2	1	–	–	3	1	–	–	–	–	437
2017	–	13	3	3	19	60	68	56	41	19	21	18	20	4	3	1	3	–	1	1	–	–	–	–	354
2018	–	48	14	8	20	48	62	76	49	31	26	18	8	9	6	1	1	3	2	1	–	–	–	–	431
2019	2	3	1	1	12	28	80	80	80	43	34	28	22	15	13	3	2	–	–	–	–	1	1	–	449
Total	74	364	152	216	384	793	1 053	1157	934	669	506	407	305	200	125	101	66	46	27	15	10	9	7	5	7 625

Table A3.2: HAK males: Count of the number of aged male fish by age and fishing year.

Fishing year	Age																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total
1990	–	–	–	1	11	14	26	29	33	14	15	14	9	9	9	10	–	3	1	–	–	–	–	2	200
1991	1	–	–	1	8	14	27	39	56	35	18	25	12	11	11	8	4	3	2	1	–	–	–	5	281
1992	–	2	3	2	4	15	23	30	23	20	21	4	10	5	1	6	1	1	1	–	1	–	–	3	176
1993	5	–	3	3	3	8	5	18	18	28	10	17	4	6	6	7	2	4	–	2	1	–	2	–	152
1997	3	2	2	4	16	47	41	41	28	14	9	7	5	6	4	2	1	1	1	2	–	–	–	1	237
1998	2	2	7	19	28	30	35	20	26	10	10	8	10	6	4	7	4	–	3	2	2	1	–	–	236
1999	2	1	5	8	11	15	37	33	30	34	18	12	13	5	3	4	2	3	–	1	1	–	–	2	240
2000	1	1	2	3	3	8	21	35	32	45	27	11	12	4	11	5	3	3	4	5	1	1	2	4	244
2001	–	3	2	3	4	14	28	44	16	14	19	4	5	5	1	2	4	1	1	–	2	2	–	1	175
2002	–	3	5	11	23	15	8	8	7	1	1	2	2	–	–	2	–	–	–	–	–	1	–	–	89
2003	–	7	7	21	38	24	38	30	14	16	12	6	4	2	–	–	–	2	–	2	1	–	1	–	225
2005	1	15	5	23	66	48	29	19	9	5	3	–	2	3	–	2	2	–	–	1	–	–	–	–	233
2006	–	19	6	4	63	50	33	15	11	22	6	9	5	7	6	–	2	1	–	–	–	1	–	–	260
2007	–	41	21	14	50	42	35	10	10	6	5	2	2	3	–	–	1	–	3	–	–	–	–	–	245
2008	–	18	4	19	29	41	30	28	28	21	21	11	18	16	4	5	5	–	1	–	–	–	–	1	300
2009	9	14	26	39	43	31	22	14	11	15	8	16	9	9	7	4	–	–	–	–	–	–	–	–	277
2010	–	1	2	19	16	23	8	–	3	3	4	4	5	5	2	1	–	1	1	–	–	–	–	–	98
2011	–	–	2	29	38	40	44	11	12	10	12	7	7	4	5	4	3	–	2	3	–	–	–	–	233
2012	1	17	9	18	57	75	59	29	19	14	9	10	10	16	10	8	11	6	4	2	–	–	–	–	384
2013	45	16	10	18	50	59	39	30	12	5	1	3	2	5	4	2	1	3	–	–	–	–	–	–	305
2014	13	3	1	14	55	76	53	37	10	7	5	3	1	3	1	3	–	1	2	–	–	1	–	–	289
2015	7	9	2	11	73	57	46	28	15	4	11	–	3	3	2	7	4	2	1	6	2	1	1	1	296
2016	3	147	31	49	69	68	38	23	3	5	4	5	2	–	1	4	–	2	–	–	–	–	1	–	455
2017	2	12	7	12	62	88	63	27	18	9	15	4	4	5	3	2	4	4	2	1	3	–	–	–	347
2018	–	84	44	15	52	84	49	37	15	9	5	4	5	3	2	5	4	1	1	–	–	1	–	–	420
2019	3	4	1	5	25	51	38	27	10	7	–	4	4	–	–	–	–	–	–	–	–	–	1	–	180
Total	98	421	207	365	897	1 037	875	662	469	373	269	192	165	141	97	100	58	42	30	28	14	9	8	20	6 577

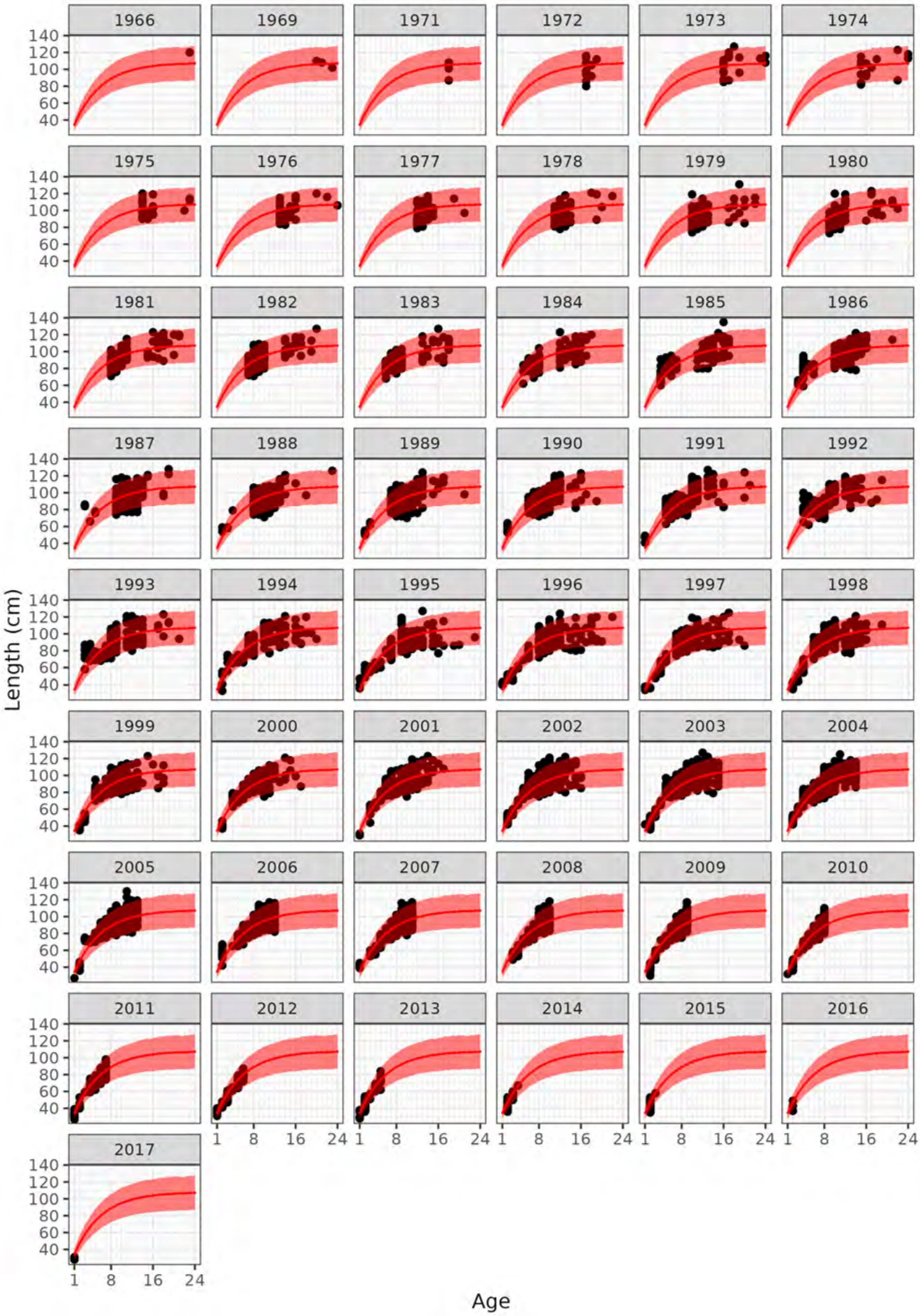

Figure A3.1: HAK females: length at age data (points) by year class for females and the posterior predictive distribution of a von Bertalanffy fitted to the data (red line is the median and shaded region is the 95% credible interval).

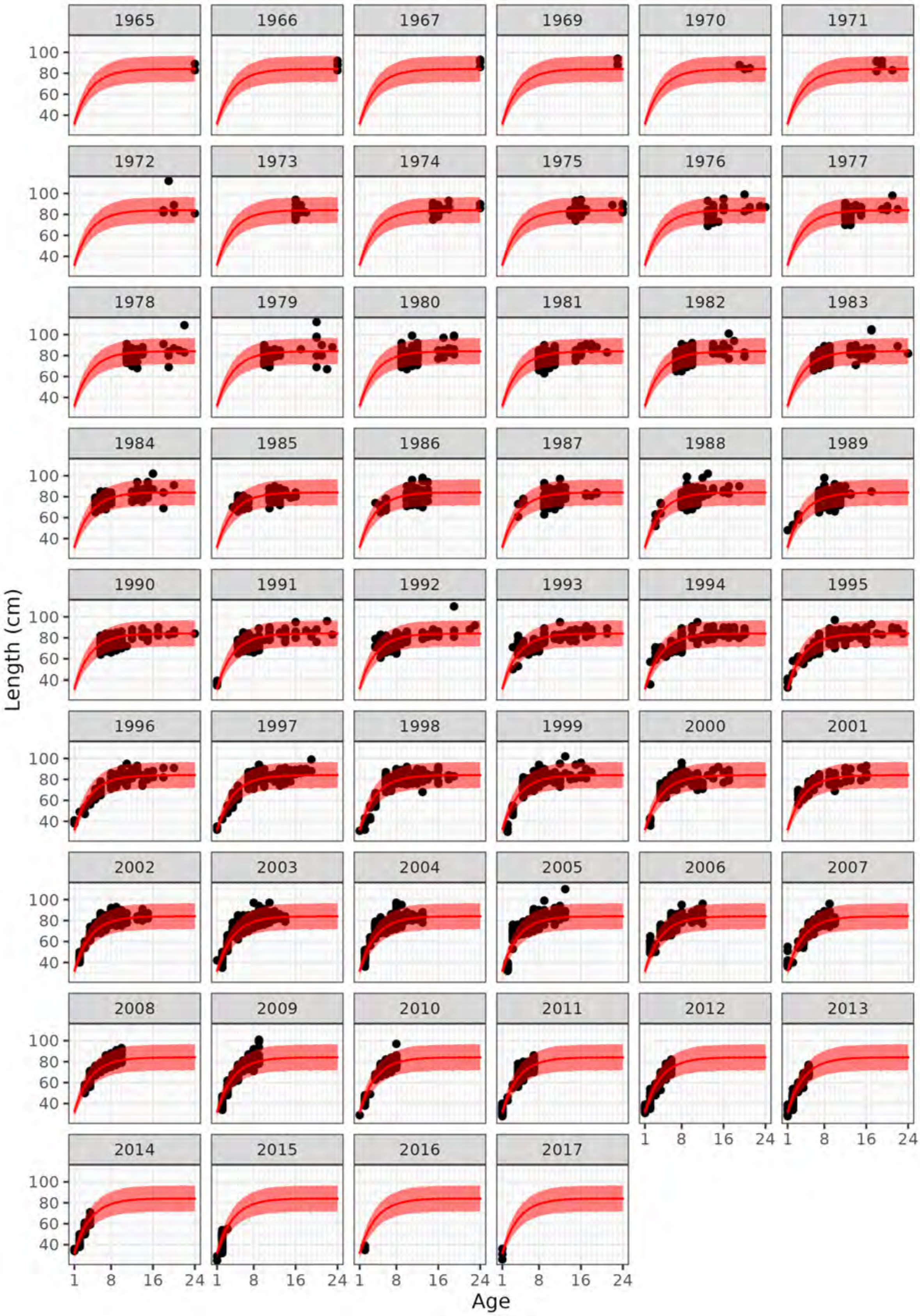

Figure A3.2: HAK males: length at age data (points) by year class for males and the posterior predictive distribution of a von Bertalanffy fitted to the data (red line is the median and shaded region is the 95% credible interval).

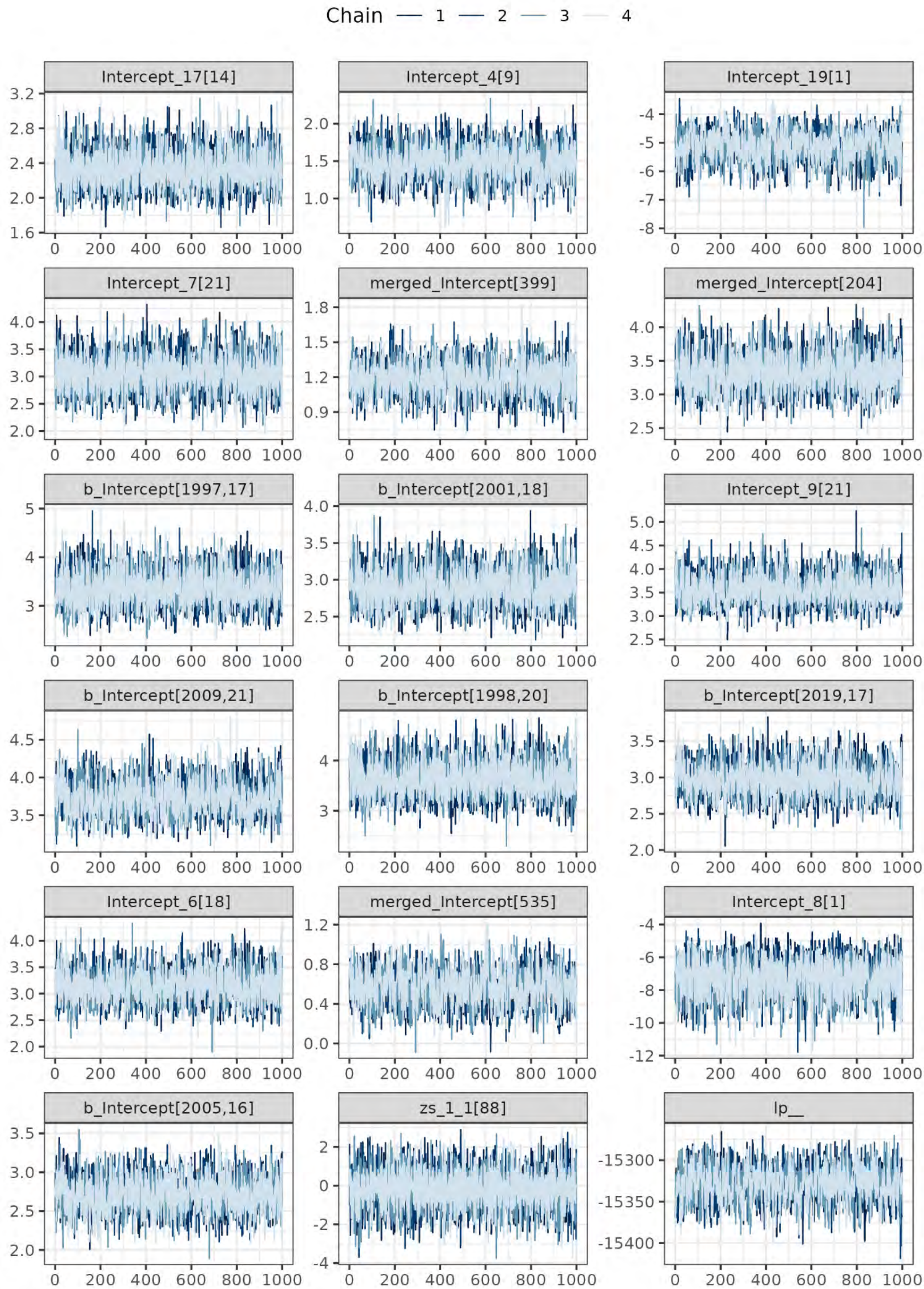

Figure A3.3: HAK females: MCMC trace plots for a random selection of parameters from the selected female model (HF4).

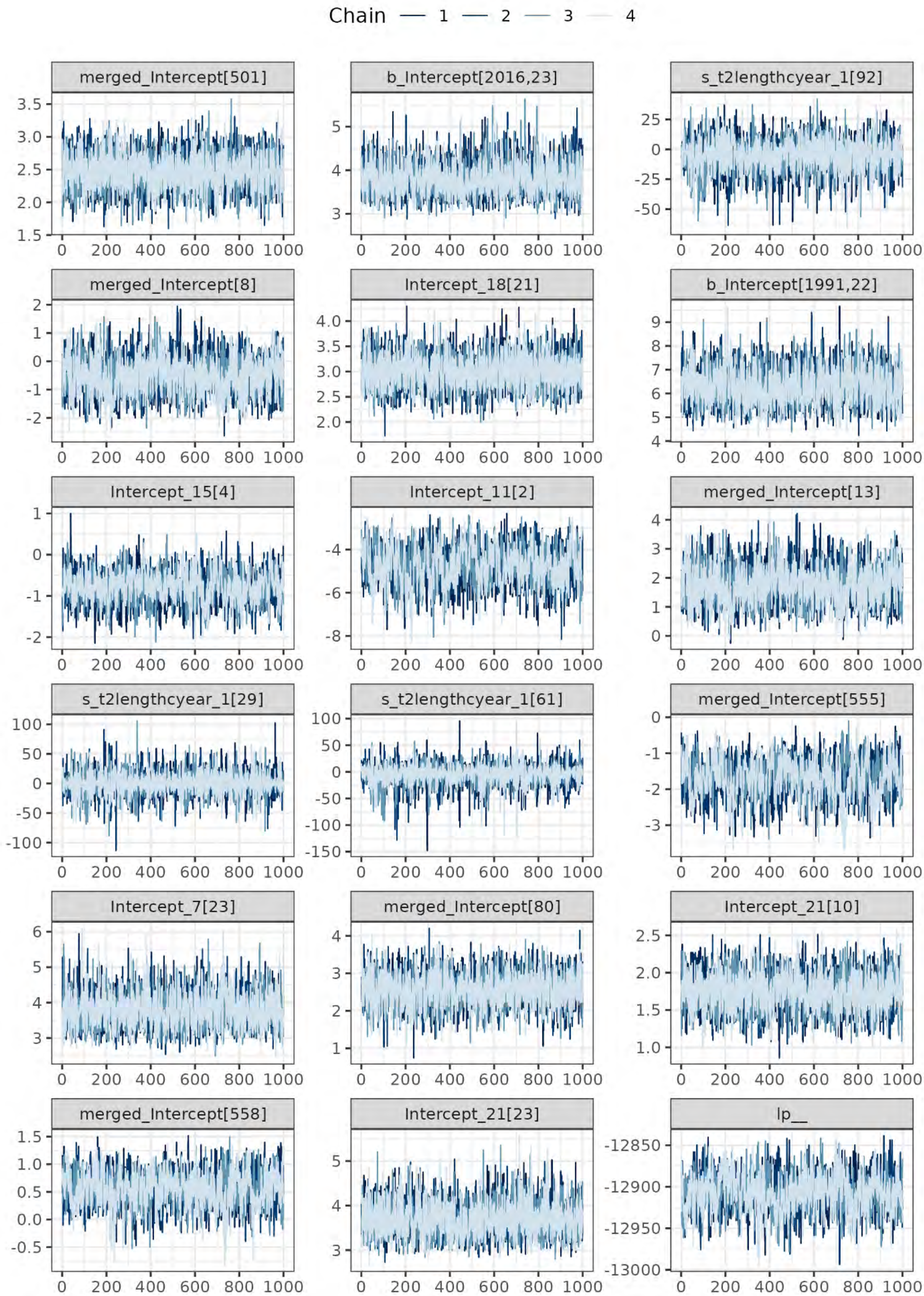

Figure A3.4: HAK males: MCMC trace plots for a random selection of parameters from the selected male model (HM5).

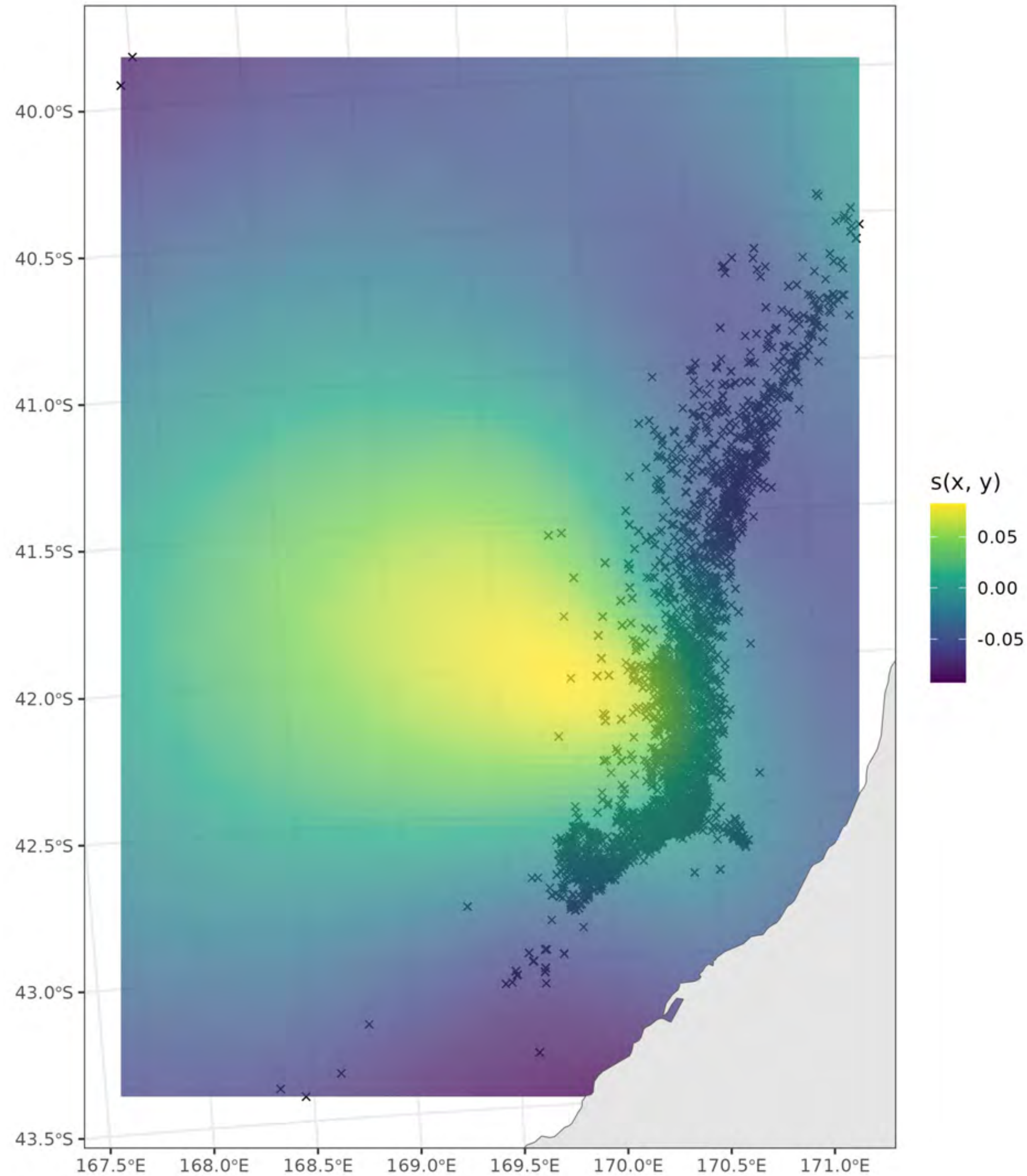

Figure A3.5: HAK female: conditional smooth from the selected model (HF2) and the data points (dark crosses).

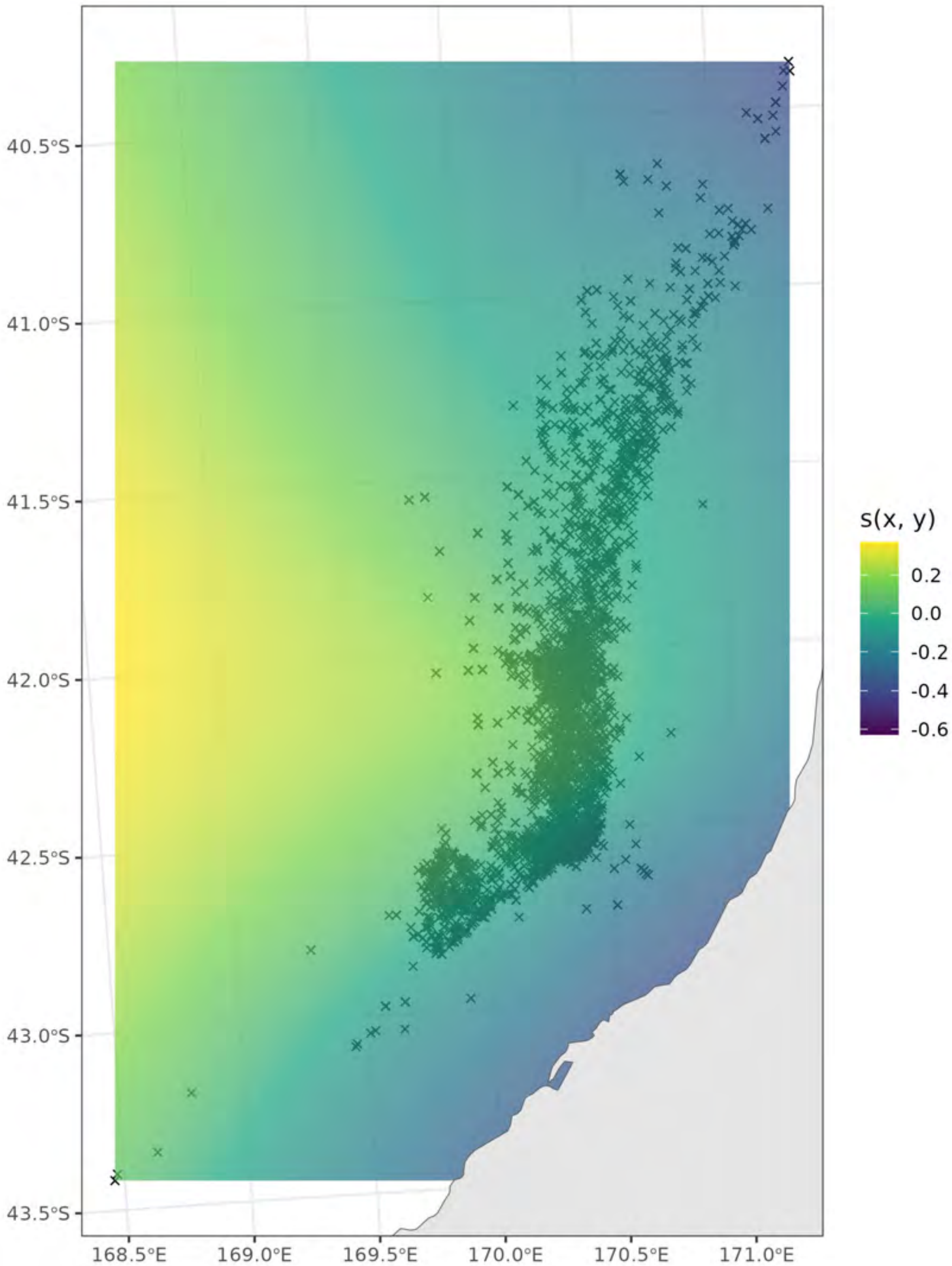

Figure A3.6 HAK male: conditional smooth from the selected model (HM2) and the data points (dark crosses).

10.2 Ling in LIN 3&4

Table A3.3: LIN females: count of the number of aged female fish by age and fishing year.

Fishing year	Age 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	Total
1990	–	–	–	–	–	1	–	3	–	1	3	4	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	13
1991	–	–	–	1	6	6	5	6	5	–	4	7	5	5	5	6	1	1	2	1	1	1	–	1	–	1	1	1	–	–	–	–	–	72
1992	4	21	14	15	12	22	20	22	11	4	12	16	20	12	16	11	6	9	4	5	6	3	2	3	1	5	–	1	1	–	–	–	1	279
1993	–	–	2	1	1	–	1	1	1	–	–	1	–	1	–	–	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	10
1997	–	4	18	14	31	17	9	9	12	11	10	1	5	5	5	4	6	6	7	6	2	3	4	1	1	2	–	–	–	–	–	–	2	195
1998	–	3	9	12	20	24	29	22	13	8	14	7	9	8	1	4	7	7	4	9	–	4	3	2	1	–	–	1	1	–	–	1	–	223
1999	–	11	16	18	21	27	29	18	23	28	8	15	12	10	5	3	6	6	5	2	3	4	1	1	1	1	–	1	–	–	–	–	–	275
2000	1	10	26	42	31	24	39	19	19	9	21	13	12	9	14	7	2	4	3	5	2	3	3	2	2	1	–	1	–	–	–	–	–	324
2001	5	20	18	45	32	26	29	20	19	21	23	20	16	8	9	3	4	6	1	2	6	1	2	4	–	–	1	–	–	–	–	–	–	341
2002	–	–	2	–	9	11	23	18	19	17	19	27	23	15	16	15	11	14	5	7	9	1	10	7	7	3	4	1	5	1	–	2	–	301
2003	–	–	–	–	8	5	22	16	18	20	26	28	32	25	10	13	14	9	4	6	4	2	6	8	4	2	5	1	–	1	1	–	1	291
2005	1	11	36	24	32	32	27	25	22	15	9	8	15	8	7	5	4	3	2	3	2	1	–	–	–	2	2	2	–	–	1	–	1	300
2006	–	–	–	–	1	–	5	8	23	24	25	28	25	16	12	11	5	6	1	6	4	–	1	1	1	1	–	1	–	–	–	–	–	205
2007	–	–	–	2	–	3	10	15	22	30	26	21	16	13	9	11	5	4	5	2	1	1	1	–	–	1	–	1	–	–	–	–	–	199
2008	–	–	–	–	1	3	6	7	7	12	27	18	12	8	8	4	4	2	1	–	–	–	1	–	–	–	–	–	–	–	–	–	–	121
2009	–	–	–	1	–	10	7	17	22	36	32	30	32	22	19	18	10	7	6	5	–	4	2	–	–	–	1	1	–	–	–	–	1	283
2010	–	–	–	–	1	12	14	20	23	28	36	58	72	28	44	21	16	11	5	12	5	5	4	3	–	3	–	1	1	–	–	1	–	424
2011	2	12	25	28	20	28	22	20	26	27	20	12	23	12	7	3	9	6	5	2	4	1	–	1	–	–	–	–	–	–	–	–	–	315
2012	–	12	27	31	25	23	27	17	21	14	22	17	16	10	5	12	9	5	5	4	4	3	–	1	–	1	–	–	–	–	–	–	–	311
2013	–	–	–	–	–	1	3	8	13	21	19	32	39	30	24	14	18	7	7	5	4	–	2	–	–	1	1	1	–	–	–	–	–	250
2014	–	–	–	1	4	3	6	10	13	12	8	11	18	11	19	15	16	4	7	4	3	1	2	3	–	1	–	–	–	–	–	–	–	172
2015	–	–	–	15	21	26	25	28	37	21	11	12	16	14	11	9	7	4	2	2	1	2	1	–	1	–	–	–	1	–	–	–	–	267
2016	–	–	–	5	4	13	22	21	37	43	53	54	51	39	34	28	25	16	8	8	2	2	2	3	–	1	1	1	–	–	1	–	–	474
2017	–	4	3	16	13	26	23	42	38	48	33	33	47	36	27	33	28	12	12	13	12	2	5	2	–	–	–	1	–	–	–	–	1	510
2018	–	–	–	–	3	10	18	15	26	24	17	18	16	23	22	16	16	8	5	6	4	10	4	7	2	2	1	1	–	–	1	2	2	279
2019	–	2	5	14	20	29	36	32	19	22	22	9	11	12	12	13	19	9	5	9	10	5	10	6	1	2	1	2	4	–	1	–	2	344
Total	13	110	201	285	316	382	457	439	489	496	500	500	544	380	341	279	248	166	111	124	89	60	66	56	22	30	18	19	13	2	5	6	11	6 778

Table A3.4: LIN males: Count of the number of aged male fish by age and fishing year.

Fishing year	Age 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	Total
1990	–	–	–	–	–	–	1	–	3	1	1	1	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	8
1991	–	2	3	2	–	9	2	3	4	6	5	5	12	2	3	4	2	3	5	–	1	–	–	1	1	1	–	1	–	–	–	–	–	77
1992	1	23	16	19	23	25	17	8	8	17	20	13	14	11	3	4	1	3	2	1	–	2	3	1	–	1	–	–	1	–	–	1	–	238
1993	–	–	–	–	2	–	–	2	–	–	–	–	–	1	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	6
1997	–	2	5	15	24	21	22	21	14	9	13	10	9	10	8	7	5	8	7	10	1	–	1	1	–	1	1	–	–	–	1	1	2	229
1998	1	4	9	16	18	25	24	32	25	27	13	17	6	8	10	9	18	9	5	4	7	5	2	2	2	–	–	–	1	1	1	2	2	305
1999	1	12	21	23	27	33	26	34	23	30	19	15	14	6	7	4	6	8	6	10	–	4	8	–	1	–	2	–	–	–	–	–	1	341
2000	2	9	30	40	35	20	12	22	23	22	14	10	10	8	5	9	3	4	5	10	2	9	5	2	4	1	1	1	–	–	1	1	–	320
2001	6	17	18	38	37	28	23	24	19	16	15	14	3	4	7	10	7	3	3	2	1	4	4	2	4	1	–	–	–	–	–	1	1	312
2002	–	–	–	3	9	17	16	21	19	21	24	19	18	15	11	2	2	13	8	7	7	8	2	8	5	2	6	1	1	2	1	–	2	270
2003	–	–	–	1	8	23	25	30	23	19	18	30	27	29	15	13	12	7	11	11	1	5	9	9	5	3	1	4	1	1	2	1	–	344
2005	–	10	43	33	22	28	33	24	14	9	16	14	9	9	8	6	2	–	4	4	1	–	4	–	1	3	1	1	–	–	–	–	1	300
2006	–	–	–	–	1	5	13	24	25	30	29	22	22	21	11	12	11	5	4	4	1	1	4	–	1	–	2	–	1	2	–	–	1	252
2007	–	–	–	–	1	4	25	16	31	24	22	21	12	10	11	8	6	1	5	4	2	1	4	2	3	–	2	3	1	–	–	1	2	222
2008	–	–	–	–	–	1	6	13	20	11	12	9	9	2	7	2	6	6	1	1	1	2	1	1	1	1	1	–	–	–	–	–	–	114
2009	–	–	–	–	–	3	12	16	23	18	32	33	21	20	19	21	14	12	6	2	6	–	1	4	1	2	2	2	2	–	–	–	1	273
2010	–	–	–	–	2	2	8	20	21	24	38	45	35	13	17	15	10	6	4	6	3	4	5	4	3	1	1	–	4	–	1	1	2	295
2011	1	8	20	19	22	33	31	17	25	18	12	18	23	20	11	3	2	10	8	3	2	–	3	1	1	1	–	1	1	1	–	–	–	315
2012	2	15	22	16	31	24	14	22	15	15	13	4	10	14	9	12	11	5	2	3	3	3	1	–	–	–	1	–	–	1	1	1	1	271
2013	–	–	–	–	1	1	9	10	9	11	11	7	7	5	7	8	8	4	1	2	–	2	–	1	–	–	–	1	–	–	–	–	1	106
2014	–	–	–	1	1	4	4	7	11	7	10	7	11	7	11	11	11	16	7	10	5	6	2	2	2	1	2	–	1	–	–	–	–	157
2015	–	–	–	8	14	24	22	25	20	14	9	13	6	10	10	6	8	9	5	4	2	2	–	2	3	–	–	–	–	–	–	–	–	216
2016	–	–	–	3	4	3	21	16	24	25	24	20	31	17	29	20	15	16	16	8	7	5	3	4	–	2	1	–	1	–	–	–	–	315
2017	–	4	5	15	30	44	46	54	39	38	52	35	35	33	36	33	21	21	22	28	21	12	10	10	6	3	1	3	1	2	1	1	4	666
2018	–	–	–	1	12	11	26	36	31	25	15	17	19	17	18	18	14	9	9	10	15	17	8	4	3	4	5	2	–	3	–	–	3	352
2019	–	–	5	12	21	37	29	31	23	29	25	15	13	16	11	5	11	13	4	9	10	12	7	4	4	4	2	2	–	3	–	–	–	357
Total	14	106	197	265	345	425	467	528	492	466	462	414	376	309	284	242	206	191	151	153	99	104	87	65	51	32	32	22	16	16	9	11	24	6 661

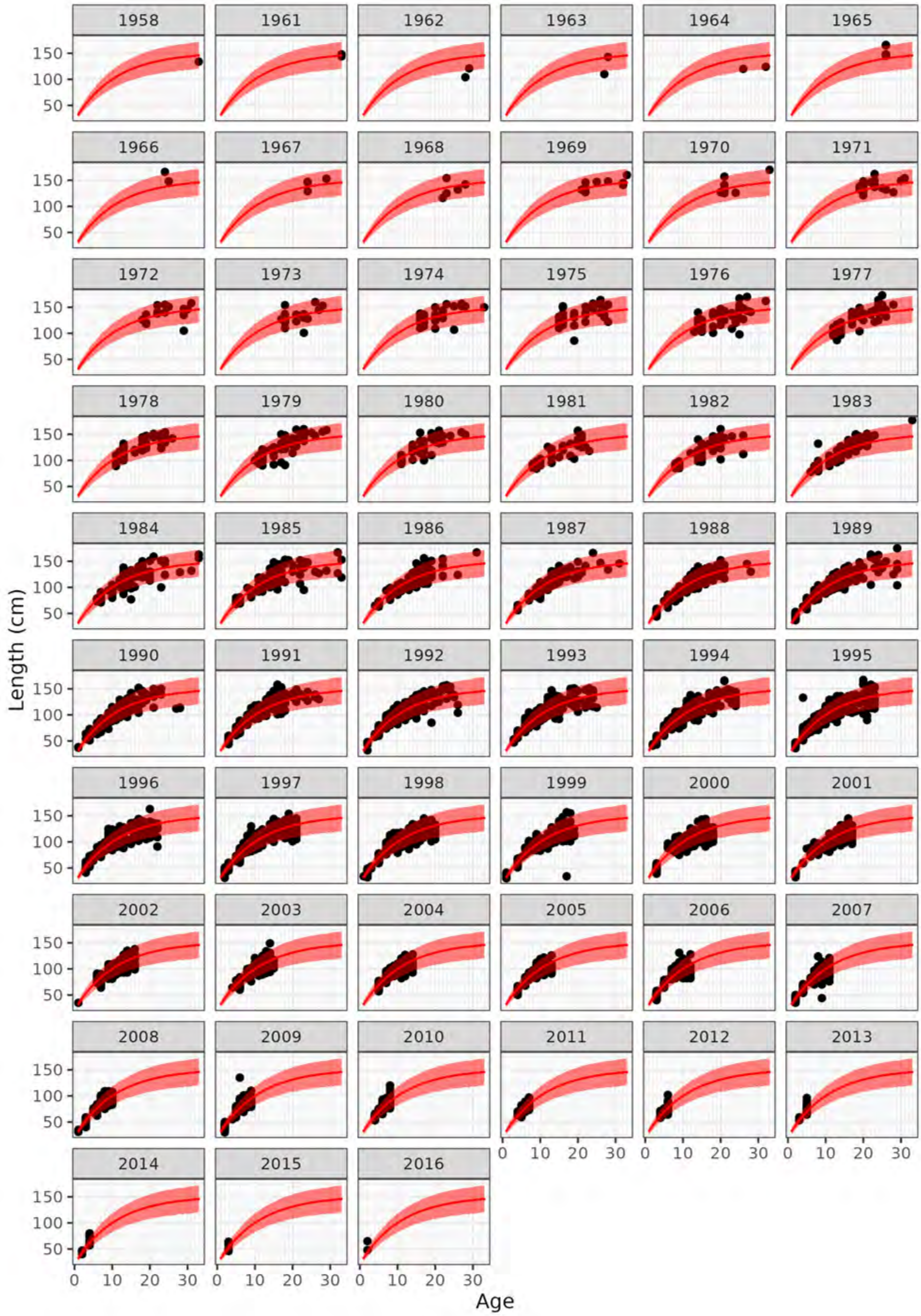

Figure A3.11: LIN females: length at age data (points) by year class for females and the posterior predictive distribution of a von Bertalanffy fitted to the data (red line is the median and shaded region is the 95% credible interval).

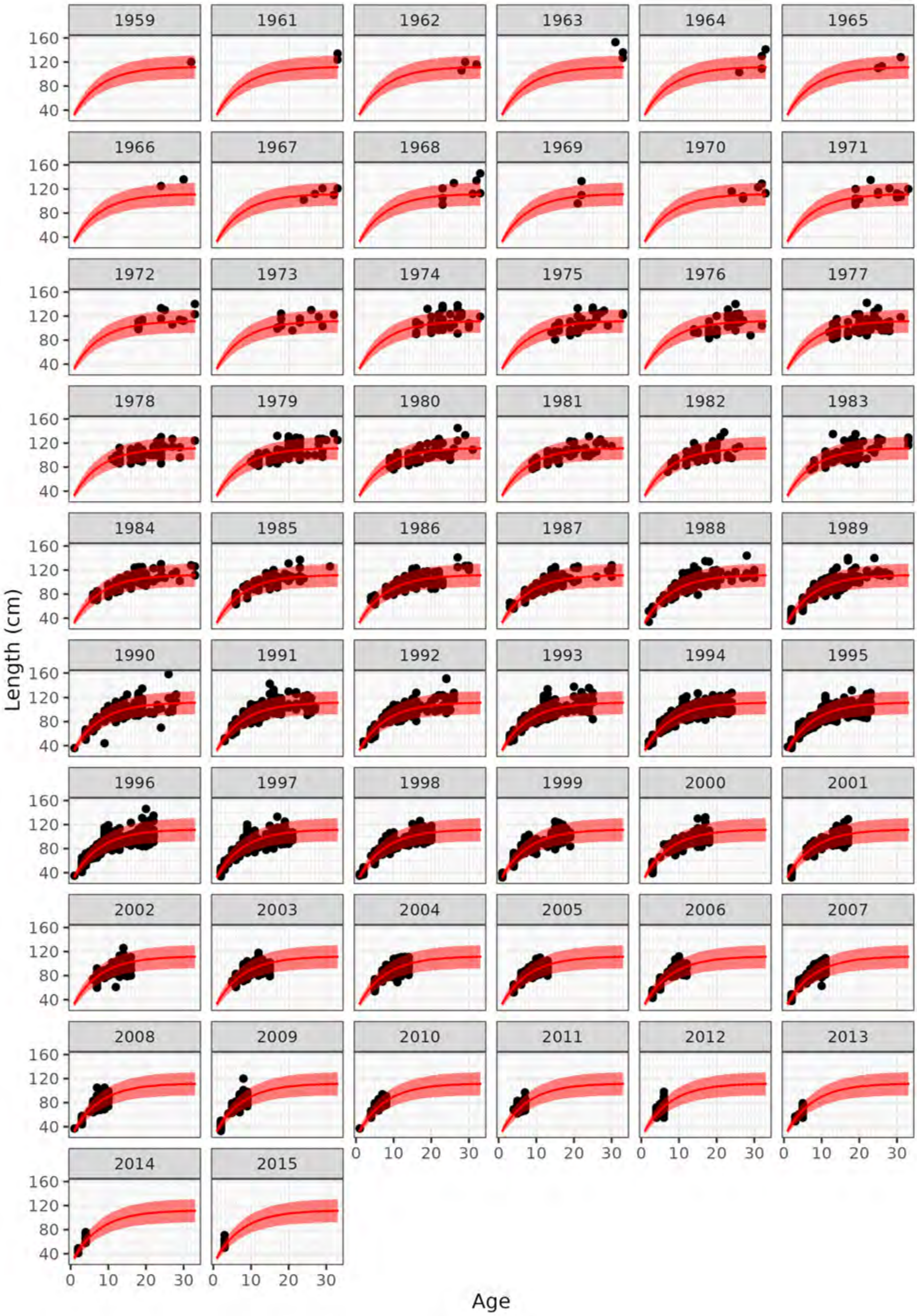

Figure A3.12: LIN males: length at age data (points) by year class for males and the posterior predictive distribution of a von Bertalanffy fitted to the data (red line is the median and shaded region is the 95% credible interval).

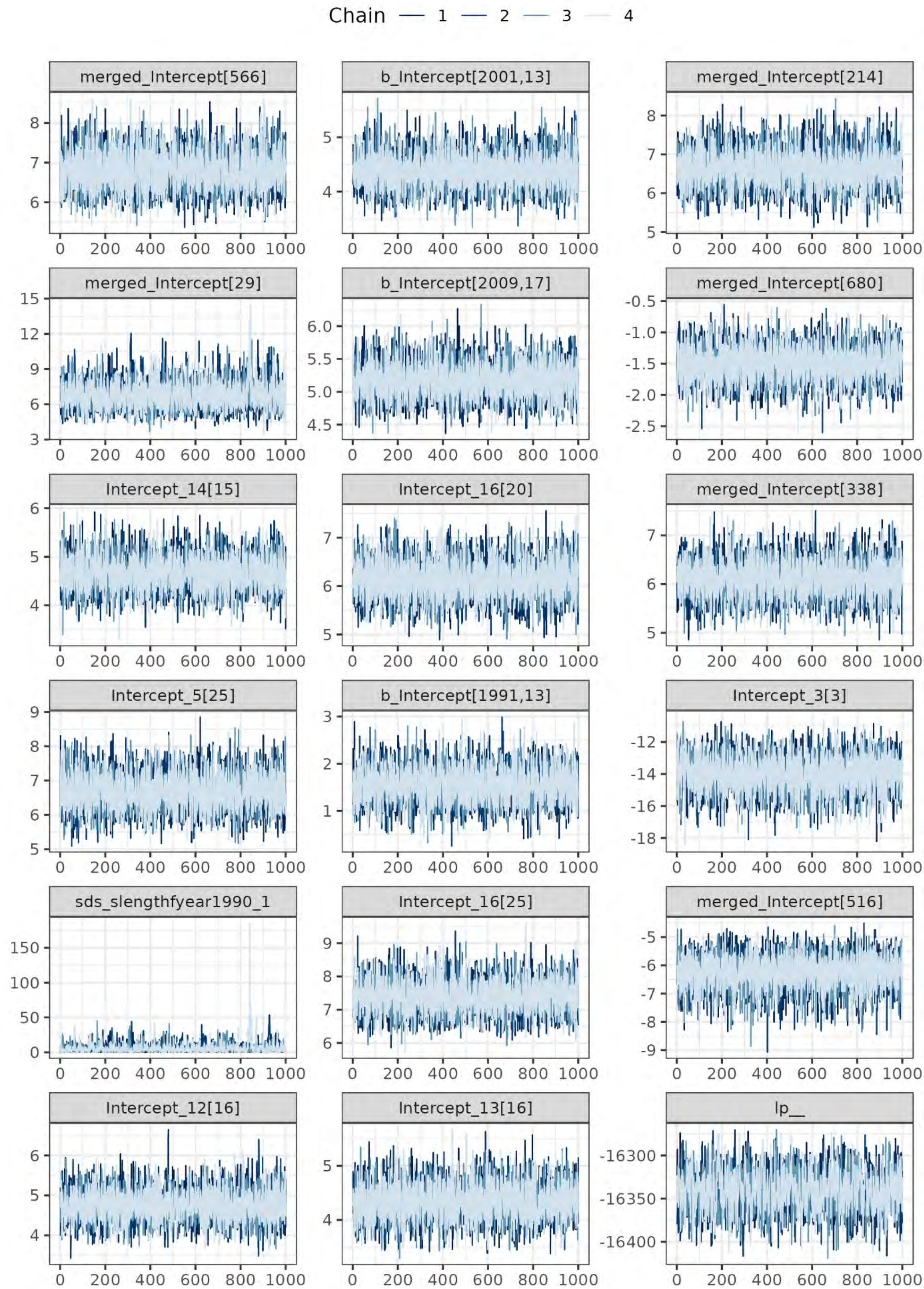

Figure A3.13: LIN females: MCMC trace plots for a random selection of parameters from the selected female model (LF8).

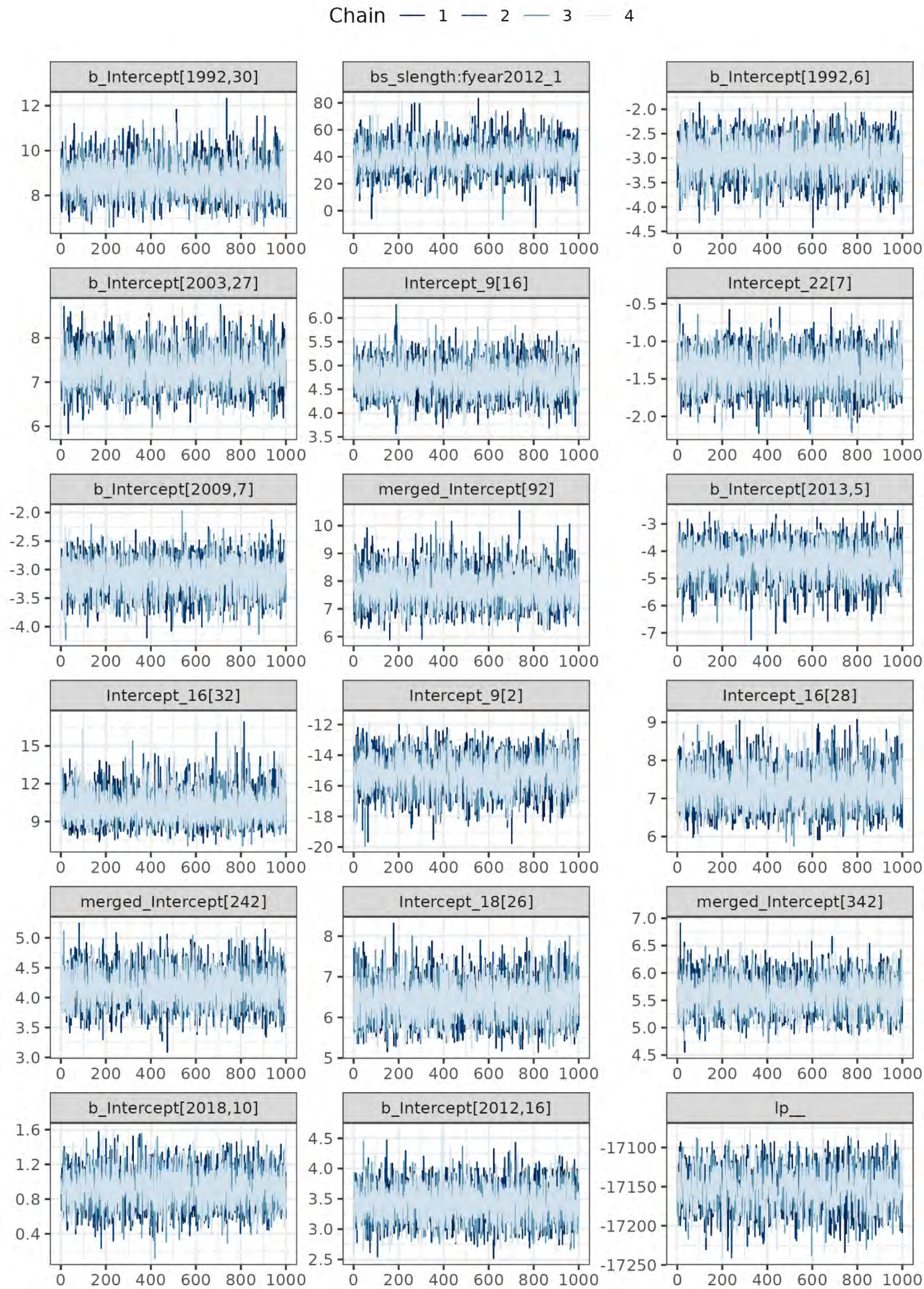

Figure A3.14: LIN males: MCMC trace plots for a random selection of parameters from the selected male model (LM5).

11. APPENDIX 4: TABLES AND FIGURES FOR THE EMPIRICAL SCALED AGE COMPOSITIONS USING ALKS

11.1 Hake in HAK 7 and ling in LIN 3&4

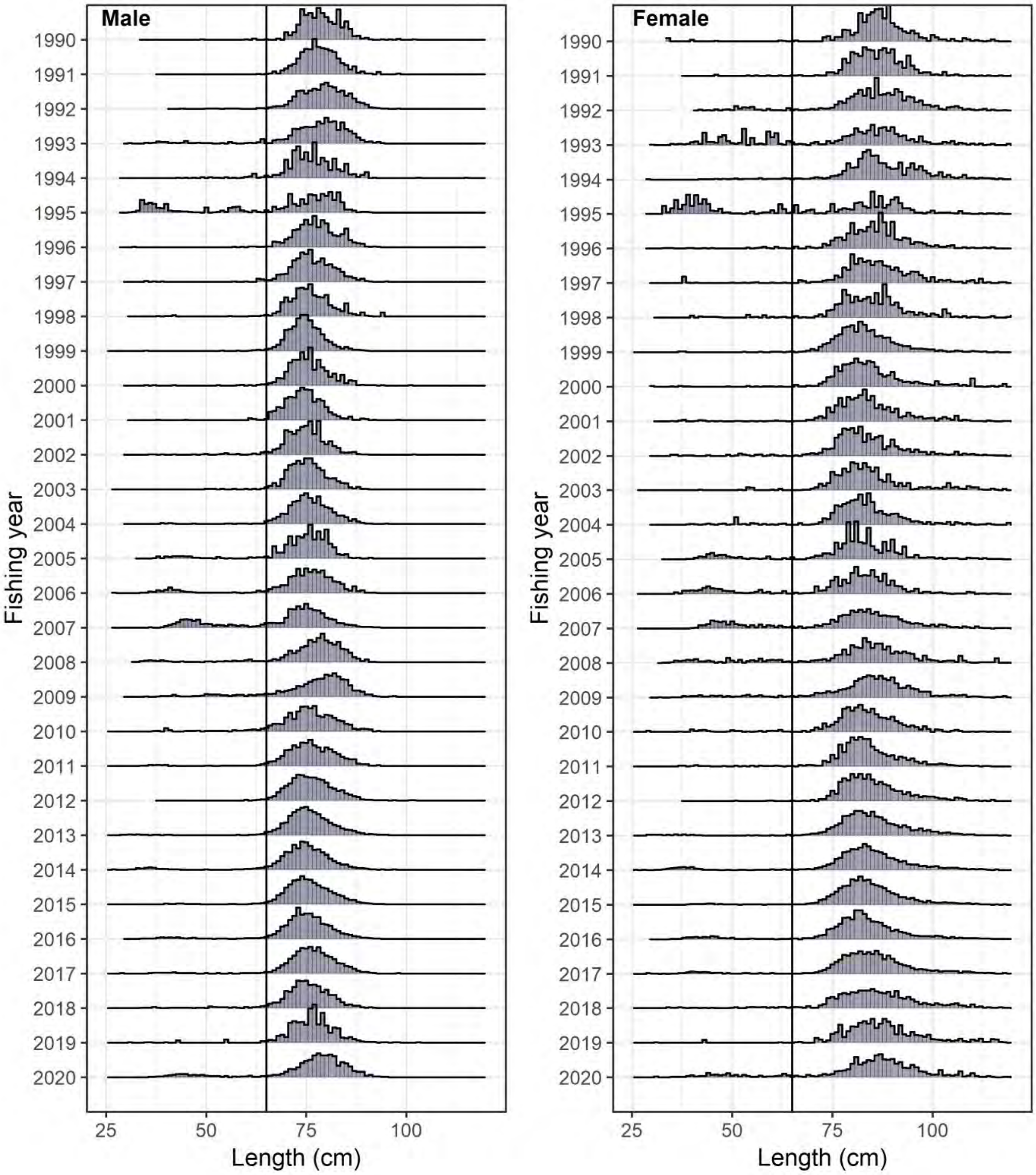

Figure A4.1: Observed scaled length frequencies for hake on the west coast South Island for 1989-90 to 2020-21. The solid vertical line indicates the cut-off for juvenile hake with lengths less than 65 cm for males and 69 cm for females.

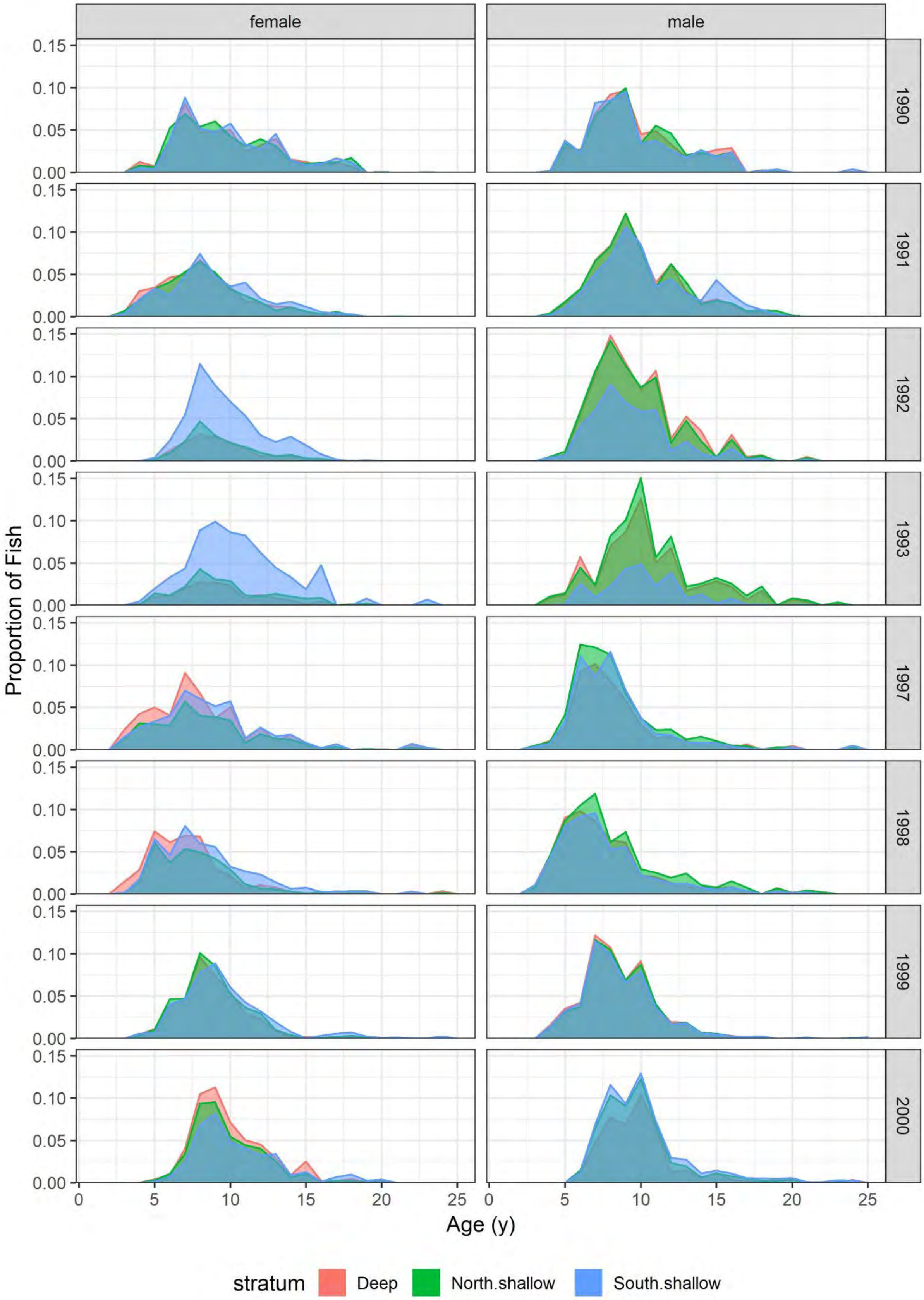

Figure A4.2: Estimated proportions at age (for ages 5+) for the deep, north shallow, and south shallow strata for hake on the west coast South Island, 1990–2000.

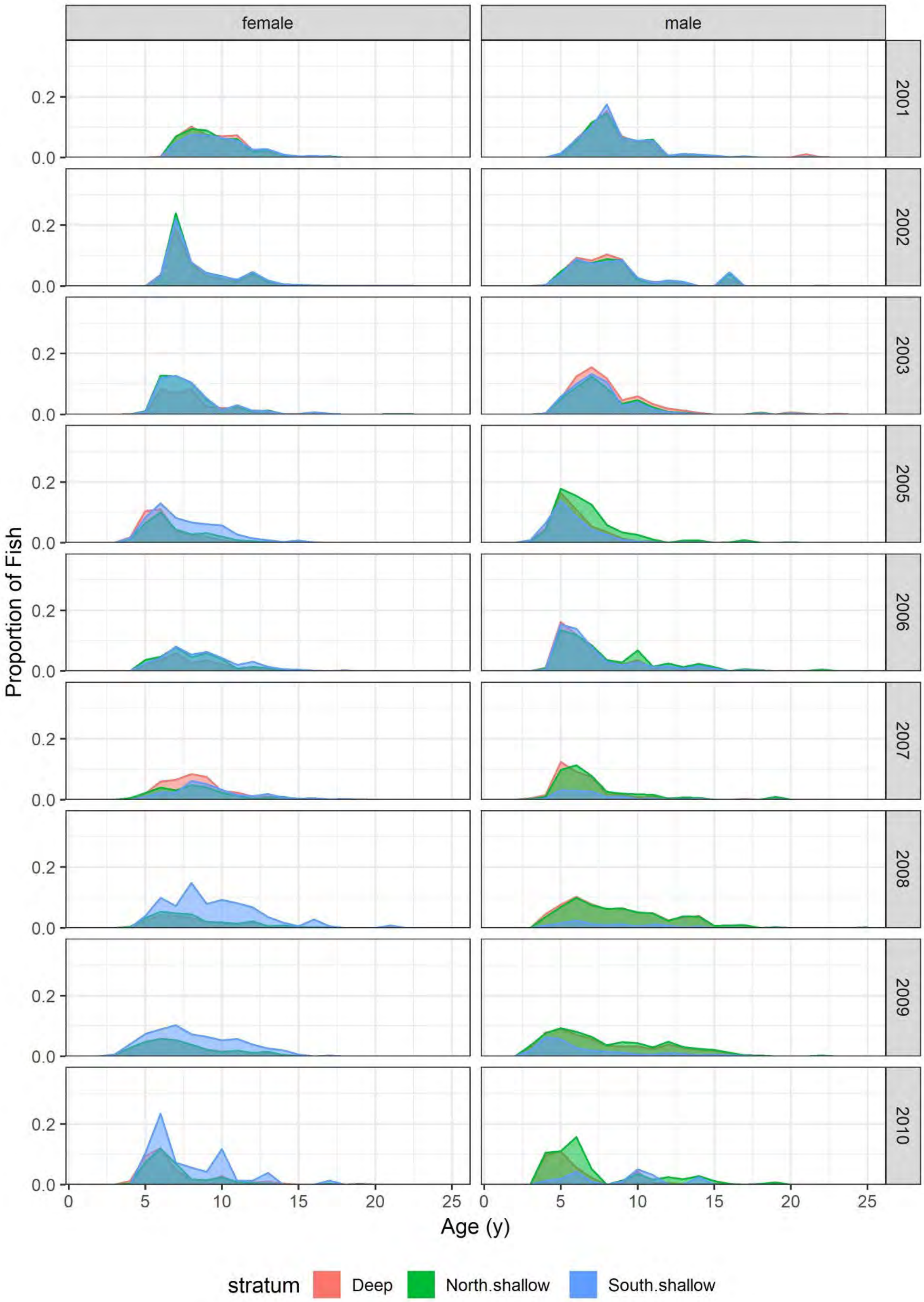

Figure A4.3: Estimated proportions at age (for ages 5+) for the deep, north shallow, and south shallow strata for hake on the west coast South Island, 2001–2010.

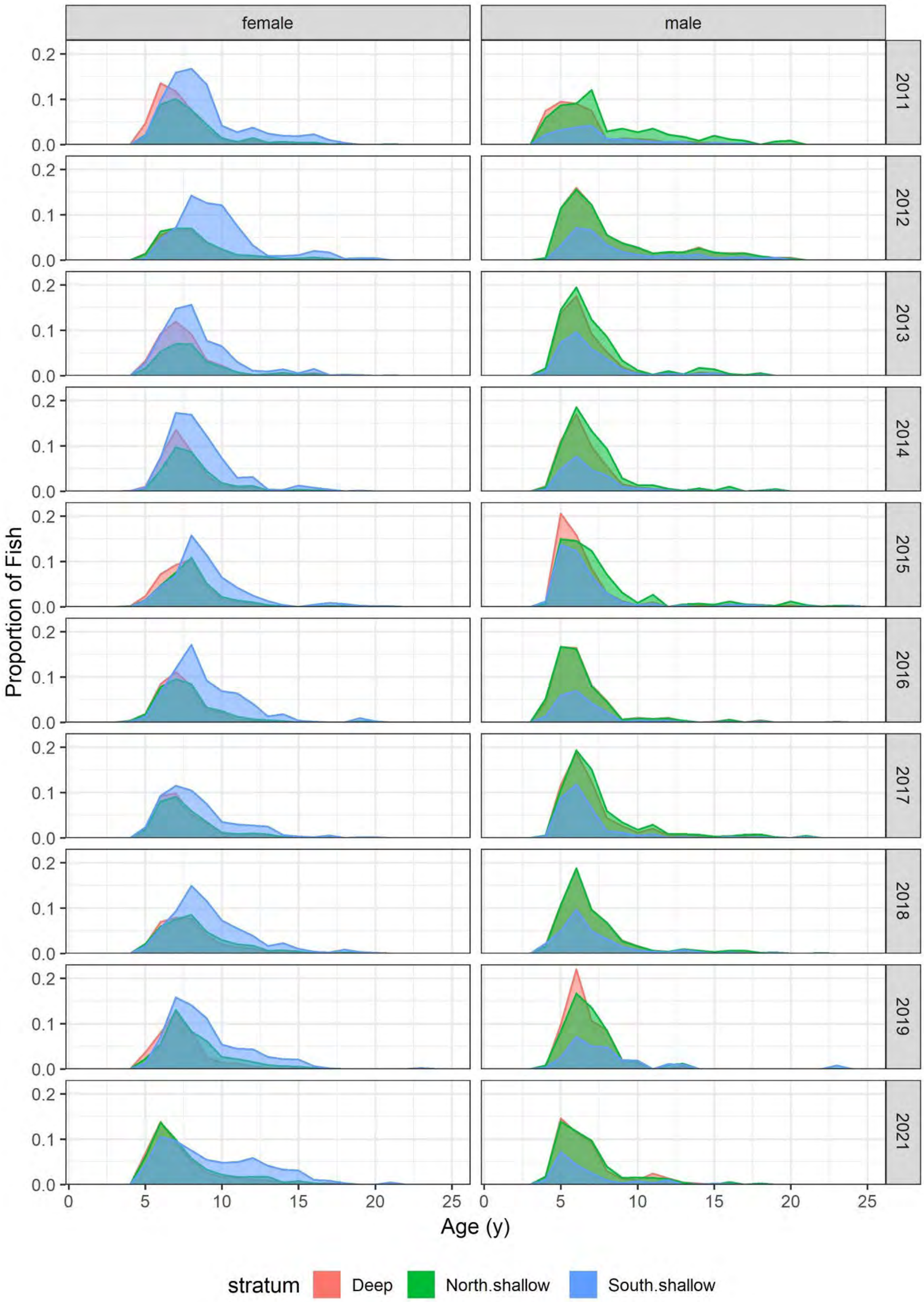

Figure A4.4; Estimated proportions at age (for ages 5+) for the deep, north shallow, and south shallow strata for hake on the west coast South Island, 2011–2021.