Thesis for the Degree of Master of Science

A length-based model for Korean chub mackerel (*Scomber japonicus*) stock

by

Jinwoo Gim

Department of Marine Biology

The Graduate School

Pukyong National University

August 2019

A length-based model for Korean chub mackerel (*Scomber japonicus*) stock [체장기반모델을 이용한 한국 고등어 （*Scomber japonicus*）자원평가]

Advisor: Prof. Saang-Yoon Hyun

By
Jinwoo Gim

A thesis submitted in partial fulfillment of the requirements
for the degree of

Master of Science

in Department of Marine Biology, The Graduate School,
Pukyong National University

August 2019

A length-based model for Korean chub mackerel
(*Scomber japonicus*) stock

A thesis
by
Jinwoo Gim

Approved by:

(Chairman) Dr. Chul Woong Oh

jaebong

(Member) Dr. Jae Bong Lee

Saangyoon Hyun

(Member) Dr. Saang -Yoon Hyun

August 23, 2019

Contents

List of figures

List of tables

A length-based model for Korean chub mackerel (*Scomber japonicus*) stock

김진우

부경대학교 대학원 해양생물학과

요 약

수산자원평가를 위한 연령구조모델 (age-structured model)은 잉여생산량모델 보다 개체군에 대한 더 구체적인 정보를 제공한다는 장점이 있다. 연령구조모델을 사용하기 위해서는 어업 또는 과학조사를 통한 개체의 시계열 (time-series) 연령자료가 요구된다. 하지만 연령구조모델의 사용은 연령자료의 부족이나 부재로 인해 제한되는 경우가 많다. 하지만 어떤 어종의 연령자료 대신 시계열 체장자료가 주어진 경우 자원평가를 위한 체장기반모델 (size-based model)의 적용을 고려할 수 있다. 따라서 본 연구에서는 Quinn 의 체장기반모델 (Quinn et al., 1998)을 변형하고, 이를 연령자료는 제한되어 있으나 시계열 체장자료는 확보된 한국 고등어 (*Scomber japonicus*) 자료에 적용함으로써 한국 고등어 개체군에 대한 자원평가를 수행하였다. 자료는 대형선망어업으로부터 얻어진 어획량 및 체장자료 (2000~2017 년)와 단위노력당 어획량 (CPUE)자료 (1996~2017 년)를 사용하였다. 본 연구에서 변형한 Quinn 의 모델은 개체군에 대한 가상의 연령구조를 가정하여, 개체군의 연령별 시간별 크기를 각 코호트 별로 추정할 수 있게 한다는 장점이 있다. Quinn 의 모델에 대한 변형은 크게 세가지 측면에서

이루어졌다. 먼저, Quinn 모델에서 잔차제곱합 (residual sum of squares)으로 표현된 목적함수를 음의 로그가능도함수 (negative log-likelihood function)로서 표현하였고, 이를 위하여 연도별 체장자료와 어획량자료는 각각 다항분포와 로그-정규분포를 따른다고 가정하였다. 둘째로, Quinn 의 모델에서 입력값으로 취급했던 순간자연사망률에 사전분포를 적용하여 추정하였다. 셋째로, AD model builder (ADMB)를 사용하여 모수들의 점추정치 및 추정치들의 불확실성을 계산하였다. 결과로서 추정된 연도별 자원량은 1.06×10^{6} ~ 2.26×10^{6} MT 의 범위에서 나타났으며, 추정된 연도별 순간어획사망률은 0.11~0.31 year^{-1} 의 범위를 보였고 순간자연사망률은 0.11 year^{-1} 로서 추정되었다.

Abstract

When data are not available on the ages of fish sampled by a fishery or survey, it is necessary to resort to a size-based model, especially when time series data about fish body sizes are provided. Assessment of Korean chub mackerel (*Scomber japonicus*) stock was one such case. Building upon Quinn's size-based model (Quinn et al., 1998), I assessed Korean chub mackerel stock. The merits of Quinn's size-based model lie in constructing an 'imaginary' age structure for a fish population, and using this to estimate the year and age population sizes of cohorts. The data used were yield and lengths of fish caught by a large purse seine fishery from 2000 - 2017, and catch-per-unit-effort from 1996 - 2017. I extended Quinn's model in three ways. First, I modified the objective function, which was the residual sum of squares in Quinn's model, into a negative log-likelihood function in which the distributions of the annual length data and yield data were assumed to follow a multinomial and a log-normal distributions, respectively. Then I applied a prior distribution for natural mortality as opposed to using a fixed value as in Quinn's model. Finally, using AD model builder, I estimated both the point estimates of the parameters and the uncertainty in those estimates. Estimates of annual biomass (including recruits) ranged from 1.06×10^6 MT to 2.26×10^6 MT. Estimates of fishing mortality rates were in the range of 0.11 - 0.31 year^{-1}, while an estimate of natural mortality was 0.11 year^{-1}.

1. Introduction

The vast majority of contemporary stock assessments that attempt to reconstruct population biomass are based on age-structured models, which are more informative than surplus production models (Punt et al., 2013). For an age-structured model, time series data about ages are key. However, these data are often limited in small scale fisheries, or in fisheries in developing countries. Under these circumstances, it is necessary to resort to size-based models, especially when time series data about the body sizes of fish caught by a fishery or survey are available. An assessment of Korean chub mackerel (*Scomber japonicus*) stock was such a case.

As a coastal-pelagic species, chub mackerel is a major commercial species in the Pacific Northwest. Because of its schooling behavior, it has mainly been caught by purse-seine nets, and is sometimes mixed with spotted mackerel (*Scomber australasicus*). In 2017, China, Japan, Russia and South Korea caught approximately 670,000 tons of fish, including chub mackerel and spotted mackerel (NPFC 2019). Amongst these countries, South Korea caught 103,870 MT of chub mackerel and 11,390 MT of spotted mackerel.

Due to a lack of long-term age data about the chub mackerel population in the Northwest pacific, there have been several attempts at stock assessments using length data (Hiyama et al., 2002; Choi et al., 2004; Wang et al., 2014). However, those studies did not use the length data directly, but converted the data into age information using a 'length-age key'. Also, they did not show the uncertainties of the estimates.

The purpose of our study was to apply a length-based model to Korean chub mackerel stock without using a 'length-age key'. Specifically, this study aimed to infer population parameters such as recruitment, growth, mortality and gear-selectivity.

As a length-based approach to assessing fish stocks, Cohen and Fishman (1980) developed a model of a stochastic growth process using time steps. If the initial distribution of body lengths in a cohort is assumed to be Gaussian, incorporating growth increments using the von Bertalanffy (LVB) growth model, the subsequent distributions will remain Gaussian throughout the lifetime of the cohort (Cohen and Fishman 1980). Based on the Cohen-Fishman model, Deriso and Parma developed a length-based model which incorporates stochastic growth, recruitment and mortality, and determined the probability distribution of the abundance and catch as a function of length

(Deriso and Parma 1988; Parma and Deriso 1990). As an extension of the Deriso and Parma model, Quinn et al. (1998) developed another length-based model. Quinn et al. (1998) assumed that length frequency distributions follow a discrete distribution. Under this model, the length distribution for a cohort, excepting recruits, becomes a mixture of Gaussian components, whereas the Deriso and Parma model confines it to being Gaussian or lognormal. This enhancement provided for more general selectivity and mortality representations and was therefore better for understanding fishing gear selectivity and mortality mechanisms. Quinn et al. (1998) used their length-based model in the case of target species which are difficult to age; however, the method can be widely applied in a situation in which age data are limited. There have been few studies that have attempted to extend Quinn's model. In this study, I modified Quinn's length-based model, and applied it to Korean chub mackerel stock.

2. Materials and Methods

2.1. Data

The length frequency (LF) and catch-per-unit-effort (CPUE) data for Korean chub mackerel were provided by the Korean National Institute of Fisheries Science (KNIFS). The KNIFS measured fork lengths from commercial large purse-seine (LPS) fisheries from January 2000 to December 2017. The LF data were classified into length classes of 1cm intervals and were aggregated annually. KNIFS has collected sample yields annually, and has sampled fishing efforts by selecting more than 80% of LPS fishing ships targeting chub mackerel. KNIFS calculated the annual CPUE (in MT/haul) for chub mackerel by dividing the sampled yields by the sampled fishing efforts from 1996 to 2017. The annual yield (in MT) of chub mackerel was provided by Statistics Korea (KOSTAT) from 1996 to 2017. In South Korea, around 90% of chub mackerel yield has been caught by LPS fisheries. CPUE and LF data from LPS fisheries were assumed to represent those of all chub mackerel fisheries. In addition, KNIFS measured the length and weight of chub mackerel from 2005 to 2017.

2.2. Length-based model

The length-based model developed by Quinn et al. (1998) set the cohort age from the recruitment age r to a maximum age of A, and assumed that the distribution of fish lengths at recruitment followed a discrete normal distribution. I assume that the age of recruitment is one year, $r=1$ and A is specified as six years for the Korean chub mackerel population. I followed the equations of Quinn's model, that describe how the distribution of the length frequency of a cohort propagates over time by stochastic growth and size-selective mortality (Appendix 1). The parameters involved are summarized in Appendix 2.

However, not all parameters in the model could be estimated. Values were therefore taken from the literature for μ_r, σ_r and L_∞ (Choi et al., 2000; Kim et al., 2018). Since the LF data were sampled from the LPS fisheries, the gear selectivity function of the LPS was assumed to be logistic (Equation. A.3.). The number of estimated parameters in this model was 28 ($N_{1996,r} - N_{2017,r}$, q, κ, σ_L, $L_{50\%}$, γ, M). The parameters were estimated using AD model builder (Founier et al., 2012).

2.3. Length-weight relationship

The allometric length-weight relationship was used to convert the fork length of an individual fish to its weight (Equation. A.16). Estimates of α and β were calculated from additional estimations using the length-weight data (2005-2017). First, the parameters (α, β) were estimated from the aggregated data (2005-2017) and applied to the entire year (2000-2017) (Fig. 1). Then the length-weight data were divided by year and the annual length-weight parameters α_j, β_j (j =2005, 2006, $\cdots$, 2017) were estimated (Fig. 2). Since the annual data did not cover the range from 2000 to 2004, the values α_j and β_j for this range were estimated from the aggregated data.

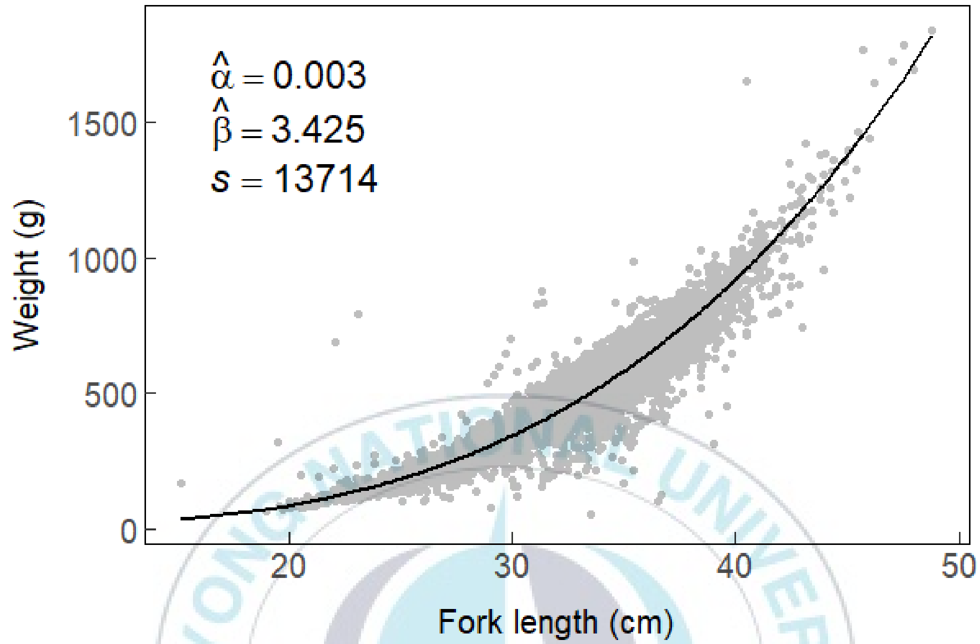
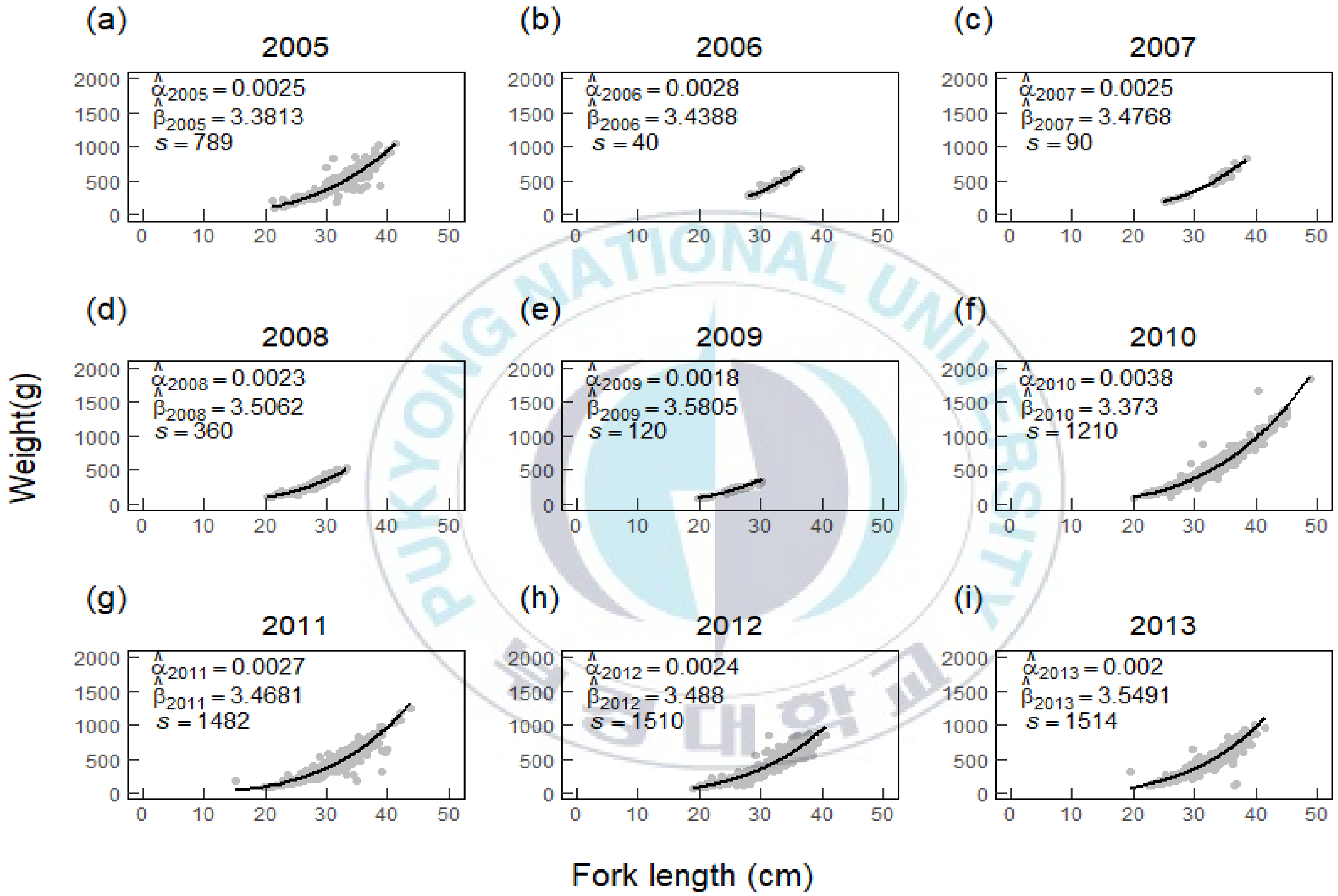


Fig. 1. Fitted length-weight relationship ($\text{weight} = \alpha \cdot \text{length}^{\beta}$) of the aggregated length-weight data (2005-2017) from Korean chub mackerel. The points denote data and the solid curve denotes the fitted curve ($\hat{\alpha} = 0.003$, $\hat{\beta} = 3.425$). s denotes the sample size. The vertical axis shows weight in grams of individuals and the horizontal axis denotes fork lengths in centimeters.


(a)
2005
$\hat{\alpha}_{2005} = 0.0025$
$\hat{\beta}_{2005} = 3.3813$
$s = 789$
(b)
2006
$\hat{\alpha}_{2006} = 0.0028$
$\hat{\beta}_{2006} = 3.4388$
$s = 40$
(c)
2007
$\hat{\alpha}_{2007} = 0.0025$
$\hat{\beta}_{2007} = 3.4768$
$s = 90$
(d)
2008
$\hat{\alpha}_{2008} = 0.0023$
$\hat{\beta}_{2008} = 3.5062$
$s = 360$
(e)
2009
$\hat{\alpha}_{2009} = 0.0018$
$\hat{\beta}_{2009} = 3.5805$
$s = 120$
(f)
2010
$\hat{\alpha}_{2010} = 0.0038$
$\hat{\beta}_{2010} = 3.373$
$s = 1210$
(g)
2011
$\hat{\alpha}_{2011} = 0.0027$
$\hat{\beta}_{2011} = 3.4681$
$s = 1482$
(h)
2012
$\hat{\alpha}_{2012} = 0.0024$
$\hat{\beta}_{2012} = 3.488$
$s = 1510$
(i)
2013
$\hat{\alpha}_{2013} = 0.002$
$\hat{\beta}_{2013} = 3.5491$
$s = 1514$
0 500 1000 1500 2000
0 10 20 30 40 50
Weight(g)
Fork length (cm)

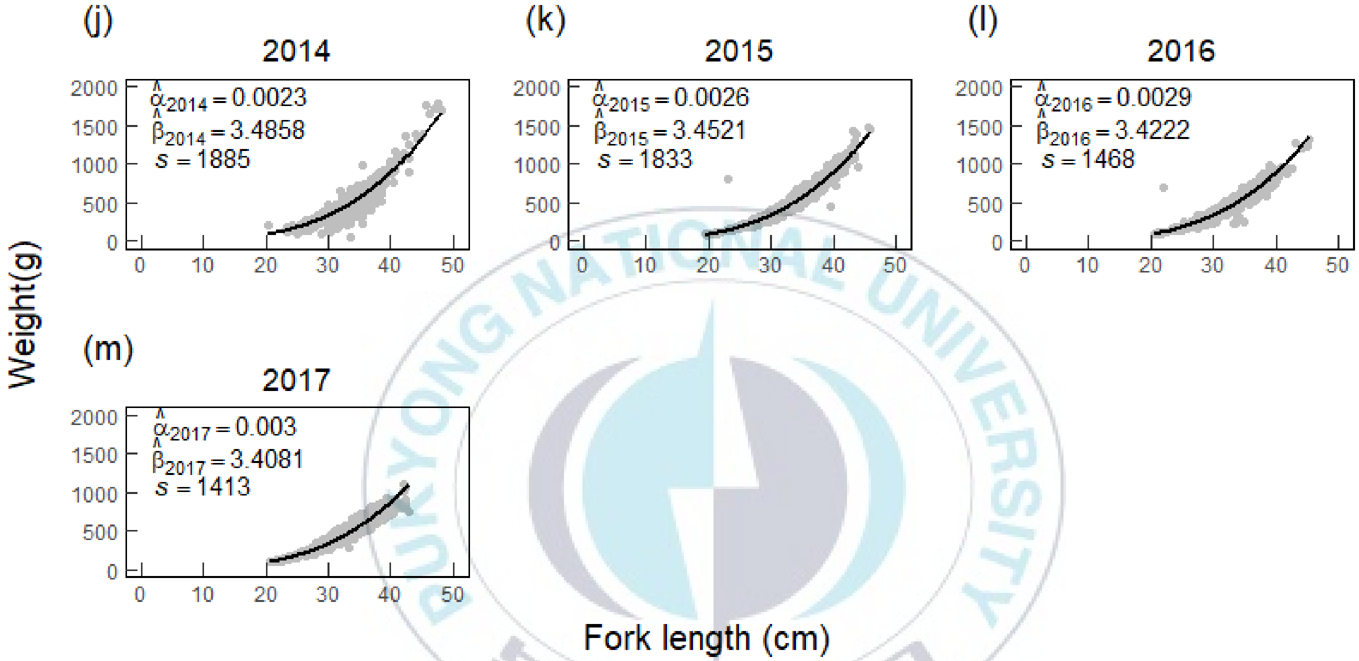


Fig. 2. Fitted length-weight relationships ($\text{weight} = \alpha_j \cdot \text{length}^{\beta_j}$) from the annual length-weight data (2005-2017) from Korean chub mackerel. The points denote the data and the solid curves denote the annual fitted

length-weight relationships. s denotes the sample sizes. The vertical axis shows weight in grams of individuals, and the horizontal axis denotes fork length in centimeters.

2.4. Objective functions

Parameters were estimated using multinomial maximum likelihood for the LF data and log-normal maximum likelihood for the observed yield. Let i denote the number of length classes ($i = 1,\ 2,\ \cdots,\ 42$) in the LF data and j denote the year ($j = 2000,\ 2001,\ \cdots,\ 2017$). Suppose that a number of fish caught in year j are selected and their fork length measured, and let o_{ij} denote the probability of how many caught fish are assigned to the i-th length class in year j. Let m_{ij} be the observed length frequency of the i-th class in year j and n_j be the observed total length frequencies in year j. Assuming that the m_i in year j follows a multinomial distribution ($\underset{\sim}{m}_j \sim Multinomial\left(n_j, \underset{\sim}{o}_j\right)$),

$$\Pr\left(\underset{\sim}{m}_j\right) = \binom{n_j}{m_{1,j}, m_{2,j}, \ldots m_{42,j}} o_{1,j}{}^{m_{1,j}} o_{2,j}{}^{m_{2,j}} \cdots o_{42,j}{}^{m_{42,j}}. \qquad (1)$$

Further, assume that the length frequencies of the catch in each year are independent. Thus, when the time range is from 2000 to 2017,

$$\Pr\left(\underset{\sim}{m}_{2000},\underset{\sim}{m}_{2001},\cdots,\underset{\sim}{m}_{2017}\right)=\Pr\left(\underset{\sim}{m}_{2000}\right)\cdot\Pr\left(\underset{\sim}{m}_{2001}\right)\cdot\cdots\cdot\Pr\left(\underset{\sim}{m}_{2017}\right).$$

Thus, as a multinomial log-likelihood function,

$$l\left(\underset{\sim}{o}_{2000},\underset{\sim}{o}_{2001},\cdots,\underset{\sim}{o}_{2017}\right)=\sum_{j=2000}^{2017}\log\left(\Gamma\left(n_j+1.0\right)\right)-\sum_{i=1}^{42}\sum_{j=2000}^{2017}\log\left(\Gamma\left(m_{ij}+1.0\right)\right)+\sum_{i=1}^{42}\sum_{j=2000}^{2017}m_{ij}\cdot\log o_{ij}\,. \quad (2)$$

where o_{ij} is the derived parameters,

$$\hat{o}_{ij}=\frac{\hat{C}_j\left(x_i\right)}{\sum_{i=1}^{42}\hat{C}_j\left(x_i\right)}, \quad (3)$$

where $\hat{C}_j\left(x_i\right)$ is the predicted number of fish caught in the i-th length class

x_i in year j. Furthermore, $\hat{C}_j\left(x_i\right)$ is linked with the estimates of parameters in the model.

$$\hat{C}_j(x_i) = \hat{N}_j(x_i)\cdot\frac{\hat{q}\cdot Effort_y\cdot\hat{S}\left(x_i\right)}{M+\hat{q}\cdot Effort_y\cdot\hat{S}\left(x_i\right)}\cdot\left(1-\exp\left(-M-\hat{q}\cdot Effort_y\cdot\hat{S}\left(x_i\right)\right)\right) \qquad (4)$$

where $\hat{N}_j\left(x_i\right)$ is a predicted abundance in the i-th length class x_i in year j, $Effort_j$ is the fishing effort in year j: $Effort_j = Y_j / CPUE_j$, $\hat{S}\left(x_i\right)$ is a predicted gear selectivity at x_i: $\hat{S}\left(x_i\right)=1\Big/\left[\left(1+\exp\left(-\hat{\gamma}\left(x_i-\hat{L}_{50\%}\right)\right)\right)\right]$.

Let Y_j denote the observed yield in year j. From Equation (A.24), the predicted yield in year j is $\hat{Y}_j=\sum_{i=1}^{42}\hat{C}_j(x_i)\cdot W\left(x_i\right)$. Then, assume a multiplicative error model:

$$Y_j=\hat{Y}_j\cdot\exp(\varepsilon_j),\ \text{where}\ \varepsilon_j \sim N\left(0,{\sigma_Y}^2\right) \qquad (5)$$

where ε_j is a stochastic error term, which is assumed to follow a normal distribution with a mean of zero and variance ${\sigma_Y}^2$. Taking the logarithm of both sides of Equation (5) results in $\log Y_j = \log \hat{Y}_j + \varepsilon_j$. Then, the distribution of $\log Y_j$ follows a normal distribution whose variance is a constant over time: $\log Y_j \sim N(\log \hat{Y}_j, {\sigma_Y}^2)$. Thus, as the normal log-likelihood,

$$l\left(\log \hat{Y}_j, {\sigma_Y}^2\right) = -\frac{d}{2}\log 2\pi - \frac{d}{2}\log {\sigma_Y}^2 - \frac{1}{2{\sigma_Y}^2}\sum_{j=2000}^{2017}\left(\log Y_j - \log \hat{Y}_j\right)^2 \quad (6)$$

where $d\ (=18)$ is the length of the observed yield (2000-2017). However, the maximum likelihood estimator (MLE) of ${\sigma_Y}^2$ can be analytically solved:

$$MLE \text{ of } {\sigma_Y}^2 = \frac{1}{d}\cdot\sum_{j=2000}^{2017}\left[\log Y_j - \log \hat{Y}_j\right]^2.$$

Thus, Equation (6) can be rewritten:

$$l\left(\log\hat{Y}_j\right) = -\frac{d}{2}\log\left[\frac{1}{d}\cdot\sum_{j=2000}^{2017}\left[\log Y_j - \log\hat{Y}_j\right]^2\right] - \frac{d}{2}\log 2\pi - \frac{d}{2}. \tag{7}$$

Assume a prior with lognormal distribution for M, and instantaneous natural mortality as a constant over time, that is, $M \sim \mathrm{lognormal}(\mu_M, {\sigma_M}^2)$. Thus, as the log-normal log-likelihood function,

$$l\left(M\right) = -\frac{1}{2}\cdot\log 2\pi - \log M - \log\sigma_M - \frac{\left(\log M - \mu_M\right)^2}{2\cdot{\sigma_M}^2}. \tag{8}$$

However, μ_M and σ_M can be represented using the mode and coefficient of variance (CV) of the prior distribution ($\mu_M = \log\left(\mathrm{mode}_M\cdot\left({CV_M}^2 + 1.0\right)\right)$, $\sigma_M = \sqrt{\log\left({CV_M}^2 + 1.0\right)}$).

An objective function J was constructed as a combination of log-likelihood functions. The parameters were estimated using the multinomial log-likelihood $l\left(o_{ij}\right)$ linked with the LF data m_{ij}, the normal log-likelihood

$l\left(\log \hat{Y}_j\right)$ linked with the observed yield Y_j and the lognormal log-likelihood for M :

$$J = -1.0 \cdot \left[\lambda_1 \cdot l\left(\underset{\sim}{\rho}_{2000}, \underset{\sim}{\rho}_{2001}, \cdots, \underset{\sim}{\rho}_{2017} \right) + \lambda_2 \cdot l\left(\log \hat{Y}_j\right) + l\left(M\right) \right] \qquad (9)$$

where λ_1 and λ_2 were weighting terms for the LF data and the yield data respectively, and were assigned the values λ_1=0.05 and λ_2=10.0 after repeated estimations adjusting fitted values for the data.

2.5. Estimation

ADMB software (Founier et al., 2012) was used to estimate the parameters which minimize the objective function value J using numerical differentiation, and its TPL (ADMB code) is shown in Appendix 3 and 4. It was numerically differentiated with respect to the free parameters to obtain the parameter estimates.

Parameter estimations were performed with different assumptions, and the best fit determined based on AIC:

$$AIC = 2\times k - 2\times\left\{ l\left(\underset{\sim}{\hat{o}}_{2000}, \underset{\sim}{\hat{o}}_{2001}, \cdots, \underset{\sim}{\hat{o}}_{2017}\right) + l\left(\log \underset{\sim}{\hat{Y}}\right)\right\}, \qquad (10)$$

where $k\ (=28)$ is the number of free parameters, $l\left(\underset{\sim}{\hat{o}}_{2000}, \underset{\sim}{\hat{o}}_{2001}, \cdots, \underset{\sim}{\hat{o}}_{2017}\right)$ and $l\left(\log \underset{\sim}{\hat{Y}}\right)$ were the maximized values of the likelihood functions linked with the data.

3. Results

By applying the length-weight data in two ways: one with $\hat{\alpha}$ and $\hat{\beta}$ from the aggregated length-weight data (Case I) and the other with annual $\hat{\alpha}_j$ and $\hat{\beta}_j$ from the annual length-weight data (Case II), the best fit in each case was found and the results compared.

By minimizing the objective function J, the annual LF data (2000-2017) and yield data (2000-2017) were fitted with their corresponding model values linked with the estimates of the parameters. The predicted values for the yield and LF data were fitted well by the best fit in both Case I and Case II (Fig. 3, Fig. 4, Fig 5).

M was unusual among the parameters to be estimated, because I assumed a prior distribution for M. Changing Mode_M and CV_M as input values, the free parameters (including M) were estimated, and the best fit, where AIC was the lowest in both cases, was found. In Case I, the minimum AIC (36551.6) was obtained at $\hat{M}$ = 0.13 year^{-1} (Fig. 6). In Case II, the minimum AIC (36541.6) was obtained at $\hat{M}$ = 0.11 year^{-1} (Fig. 6). Interestingly, the trend of the AIC for $\hat{M}$ varied with A. For A=6, as shown in Fig. 6, the graph was concave in shape. Thus, the lowest point (minimum AIC) could be specified. For A=5, however, the graph showed a monotonically increasing trend and $\hat{M}$ for minimum AIC was nearly zero. For A=7 or more, the estimations failed. In both cases, the best fit was shown when A=6. In the model, the year of the first recruit is determined by A. When A=6, if I trace back the last age-class in 2000 (the first year of LF data), the recruits will start from 1995. However, the recruits from 1995 could be ignored and recruits estimated from 1996, because the LF data in 2000 poorly support the recruitment in 1995. Consequently, the number of estimated parameters was 28 ($N_{1996,r} - N_{2017,r}$, q, κ, σ_L, $L_{50\%}$, γ, M).

In Case I and Case II, the fitted values with best fit were almost the same (Fig. 3, Fig. 4, Fig. 5), but the predicted annual biomass, recruitment, and fishing mortality rates differed in value, with similar trends. In Case I and Case II, the predicted annual average (2000-2017) biomasses were 2.03×10^6 MT and 1.74×10^6 MT (Fig. 7), and the estimated annual average (1996-2017) recruitments in weight were 137.97×10^3 MT and 110.76×10^3 MT (Fig. 8).

The instantaneous fishing mortality rates in average (2000-2017) were 0.17 year^{-1} and 0.19 year^{-1} (Fig. 9). The estimates of the parameters and the standard error in each case are shown in Tables 1 and 2. The relative standard errors among the estimates of annual recruits were the largest in $\hat{N}_{2014,r}$ and the smallest in $\hat{N}_{1996,r}$ in both cases (Fig. 8).

Table 1. Estimates of parameters and standard error (SE) of the estimates at the best fit using $\hat{\alpha}$ and $\hat{\beta}$ from the aggregated length-weight data (Case I). $N_{1996,r}$ to $N_{2017,r}$ denote the annual recruitment parameters. q (year$^{-1}\cdot$ haul^{-1}) denotes the catchability coefficient. κ (year^{-1}) denotes the growth parameter. $L_{50\%}$ (cm) and γ (cm^{-1}) denote the parameters in the selectivity function. M (year^{-1}) denotes the natural mortality and σ_L (cm) denotes the deviation of stochastic error term in the growth equation.

Parameters	Estimates	SE	Parameters	Estimates	SE
$\log N_{1996,r}$	21.59	0.48	$\log N_{2010,r}$	21.81	0.54
$\log N_{1997,r}$	20.83	0.51	$\log N_{2011,r}$	21.81	0.53
$\log N_{1998,r}$	21.90	0.49	$\log N_{2012,r}$	21.73	0.53
$\log N_{1999,r}$	21.73	0.51	$\log N_{2013,r}$	20.99	0.70
$\log N_{2000,r}$	21.74	0.50	$\log N_{2014,r}$	19.08	2.42
$\log N_{2001,r}$	21.49	0.53	$\log N_{2015,r}$	21.44	0.53
$\log N_{2002,r}$	21.53	0.52	$\log N_{2016,r}$	21.34	0.57
$\log N_{2003,r}$	21.69	0.52	$\log N_{2017,r}$	18.77	2.31
$\log N_{2004,r}$	21.71	0.51	$\log q$	-10.75	0.40
$\log N_{2005,r}$	21.54	0.55	$\log \kappa$	-2.26	0.03
$\log N_{2006,r}$	21.78	0.54	$\log L_{50\%}$	3.33	0.01
$\log N_{2007,r}$	21.85	0.54	γ	0.46	0.02
$\log N_{2008,r}$	21.94	0.54	$\log M$	-2.01	0.51
$\log N_{2009,r}$	21.85	0.55	σ_L	0.99	0.04

Table 2. Estimates of parameters and standard error (SE) of the estimates at the best fit using $\hat{\alpha}_j$ and $\hat{\beta}_j$ from annual length-weight data (Case II). $N_{1999,r}$ to $N_{2017,r}$ denote the annual recruitment parameters. q (year^{-1} · haul^{-1}) denotes the catchability coefficient. κ (year^{-1}) denotes the growth parameter. $L_{50\%}$ (cm) and γ (cm^{-1}) denote the parameters in the selectivity function. M (year^{-1}) denotes the natural mortality and σ_L (cm) denotes the deviation of stochastic error term in the growth equation.

Parameters	Estimates	SE	Parameters	Estimates	SE
$\log N_{1996,r}$	21.37	0.42	$\log N_{2010,r}$	21.62	0.47
$\log N_{1997,r}$	20.63	0.46	$\log N_{2011,r}$	21.62	0.46
$\log N_{1998,r}$	21.70	0.43	$\log N_{2012,r}$	21.55	0.46
$\log N_{1999,r}$	21.53	0.44	$\log N_{2013,r}$	20.80	0.64
$\log N_{2000,r}$	21.49	0.44	$\log N_{2014,r}$	18.92	2.40
$\log N_{2001,r}$	21.30	0.46	$\log N_{2015,r}$	21.28	0.47
$\log N_{2002,r}$	21.32	0.45	$\log N_{2016,r}$	21.17	0.51
$\log N_{2003,r}$	21.47	0.45	$\log N_{2017,r}$	18.58	2.36
$\log N_{2004,r}$	21.48	0.45	$\log q$	-10.62	0.35
$\log N_{2005,r}$	21.29	0.49	$\log \kappa$	-2.25	0.03
$\log N_{2006,r}$	21.50	0.48	$\log L_{50\%}$	3.33	0.01
$\log N_{2007,r}$	21.64	0.47	γ	0.45	0.01
$\log N_{2008,r}$	21.68	0.48	$\log M$	-2.22	0.57
$\log N_{2009,r}$	21.65	0.48	σ_L	0.99	0.04

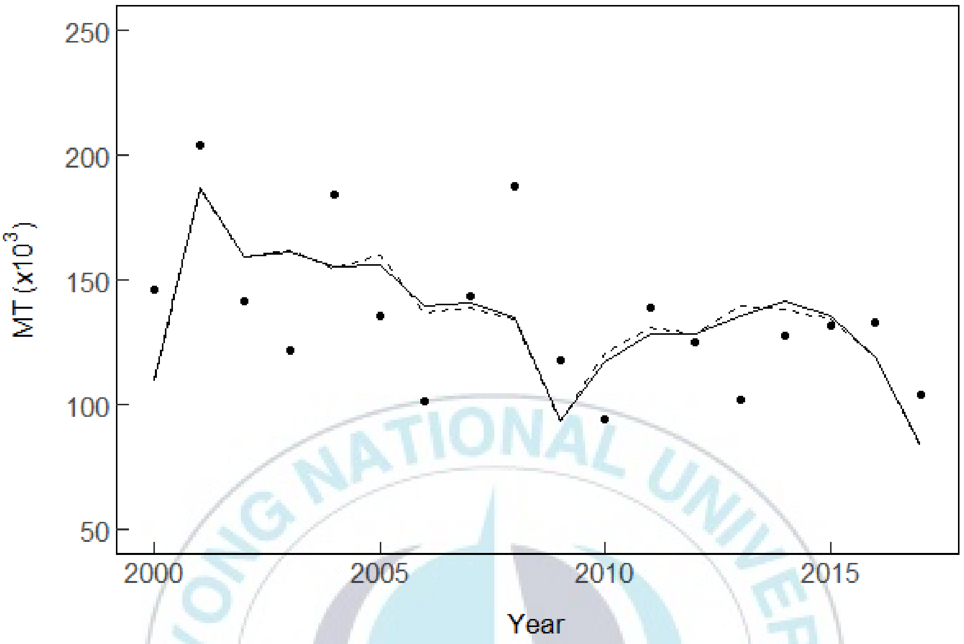
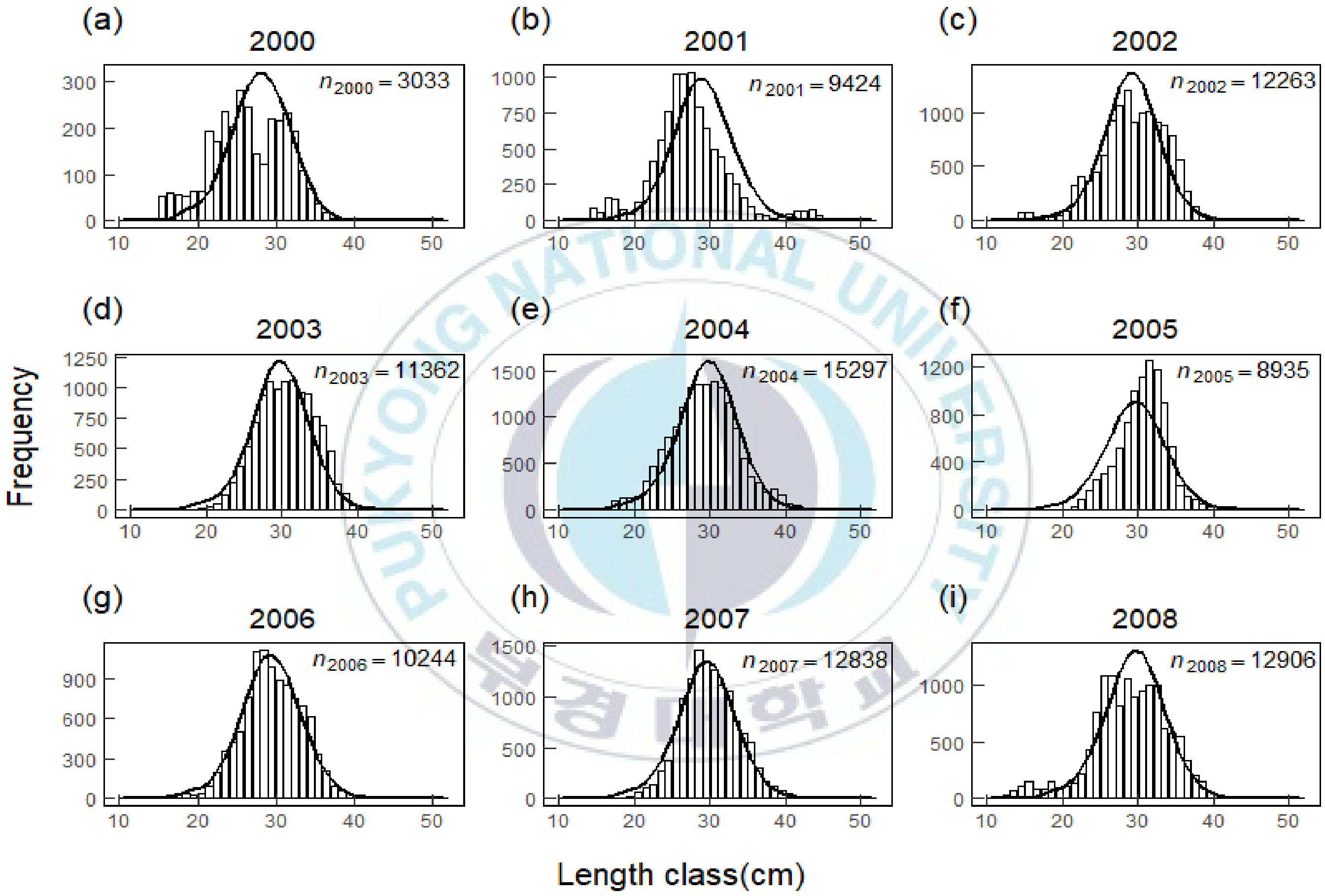


Fig. 3. Annual observed yield of Korean chub mackerel from in South Korea and the predicted yield from the length-based model with the best fit (Case I and Case II). The solid line and broken line denote the predicted yield from Case I (using aggregated length-weight data) and Case II (using annual length-weight data) respectively. The dots denote the observed yield. The vertical axis denotes yield in MT ($x10^3$) and the horizontal axis denotes time in years.


(a)
2000
$n_{2000} = 3033$
(b)
2001
$n_{2001} = 9424$
(c)
2002
$n_{2002} = 12263$
(d)
2003
$n_{2003} = 11362$
(e)
2004
$n_{2004} = 15297$
(f)
2005
$n_{2005} = 8935$
(g)
2006
$n_{2006} = 10244$
(h)
2007
$n_{2007} = 12838$
(i)
2008
$n_{2008} = 12906$
Frequency
Length class(cm)
10
20
30
40
50
0
100
200
300
250
500
750
1000
1250
1500
400
800
1200
600
900

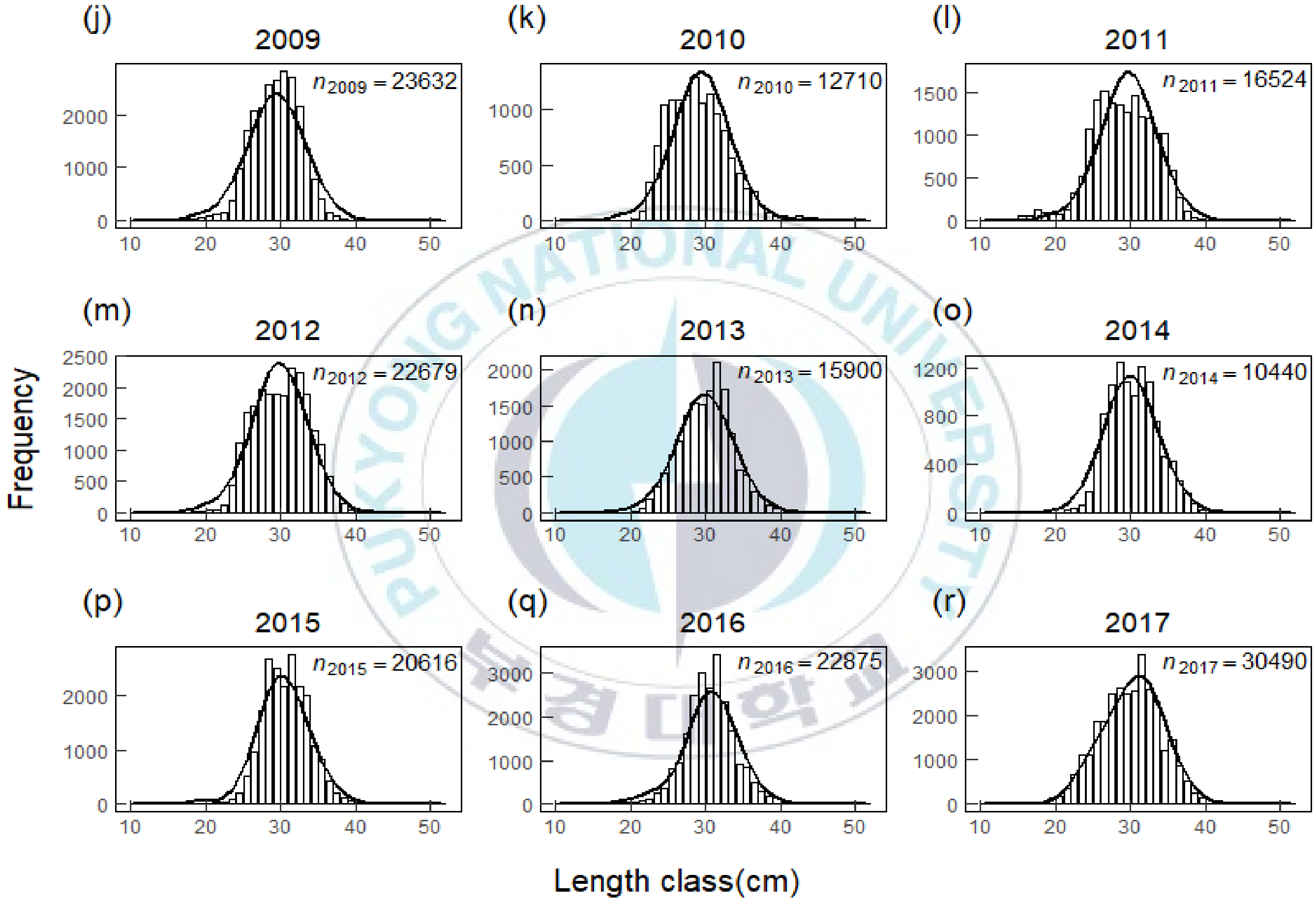
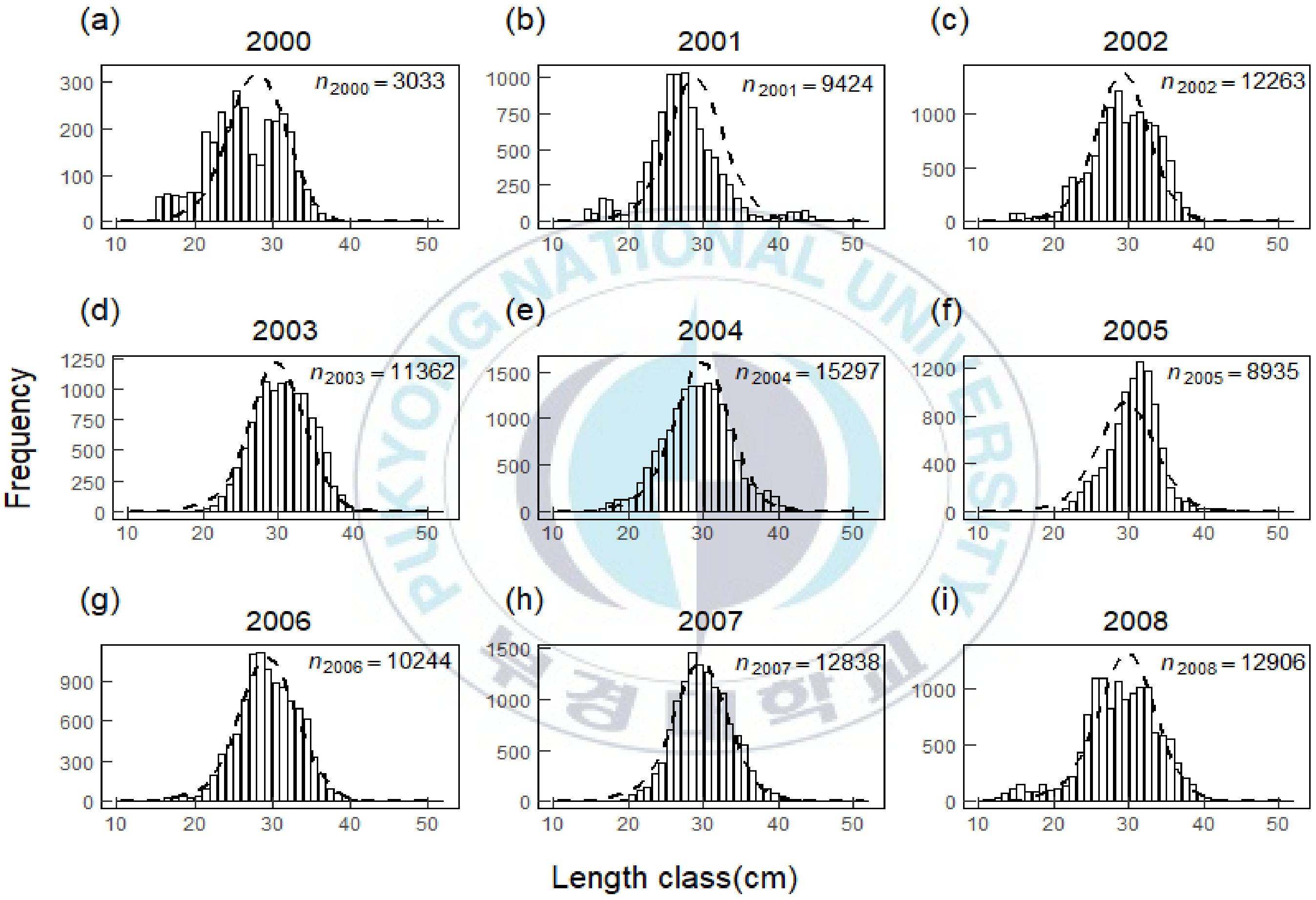
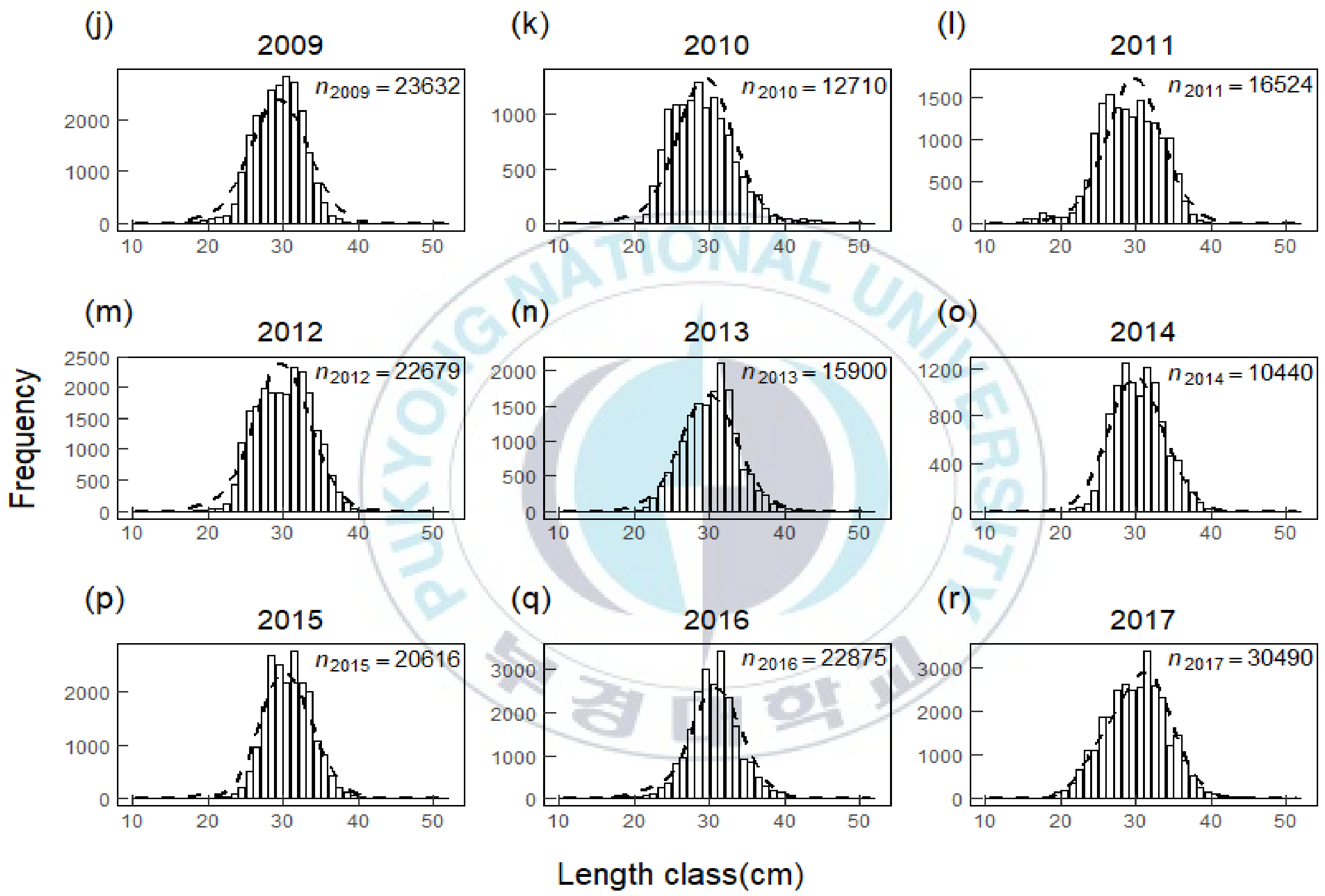

(j)
2009
$n_{2009} = 23632$
(k)
2010
$n_{2010} = 12710$
(l)
2011
$n_{2011} = 16524$
(m)
2012
$n_{2012} = 22679$
(n)
2013
$n_{2013} = 15900$
(o)
2014
$n_{2014} = 10440$
(p)
2015
$n_{2015} = 20616$
(q)
2016
$n_{2016} = 22875$
(r)
2017
$n_{2017} = 30490$
Frequency
Length class(cm)
0
500
1000
1500
2000
2500
3000
400
800
1200
10
20
30
40
50

Fig. 4. Observed length frequency of Korean chub mackerel from the large purse-seine fisheries in South Korea (2000-2017) and the predicted length frequency from the length-based model with the best fit in Case I (using aggregated length-weight data). The horizontal axis denotes length-classes with an interval of 1cm and the vertical axis denotes frequencies. The blank histograms denote the data and the solid lines denote predicted values from the model. n_j on each panel denotes the sample size of the annual dataset ($j = 2000, 2001, \cdots, 2017$).


(a)
2000
$n_{2000} = 3033$
(b)
2001
$n_{2001} = 9424$
(c)
2002
$n_{2002} = 12263$
(d)
2003
$n_{2003} = 11362$
(e)
2004
$n_{2004} = 15297$
(f)
2005
$n_{2005} = 8935$
(g)
2006
$n_{2006} = 10244$
(h)
2007
$n_{2007} = 12838$
(i)
2008
$n_{2008} = 12906$
0
100
200
300
250
500
750
1000
1250
1500
400
800
1200
600
900
10
20
30
40
50
Frequency
Length class(cm)



(j)
2009
$n_{2009} = 23632$
(k)
2010
$n_{2010} = 12710$
(l)
2011
$n_{2011} = 16524$
(m)
2012
$n_{2012} = 22679$
(n)
2013
$n_{2013} = 15900$
(o)
2014
$n_{2014} = 10440$
(p)
2015
$n_{2015} = 20616$
(q)
2016
$n_{2016} = 22875$
(r)
2017
$n_{2017} = 30490$
Frequency
Length class(cm)
0
500
1000
1500
2000
2500
3000
400
800
1200
10
20
30
40
50

Fig. 5. Observed length frequency data of Korean chub mackerel from the large purse-seine fisheries in South Korea (2000-2017) and the predicted length frequency from the length-based model with the best fit in Case II (using annual length-weight data). The horizontal axis denotes length-classes with an interval of 1cm and the vertical axis denotes frequencies. The blank histograms denote the data and the broken lines denote predicted values from the model. n_j on each panel denotes the sample size of the annual dataset ($j = 2000, 2001, \cdots, 2017$).

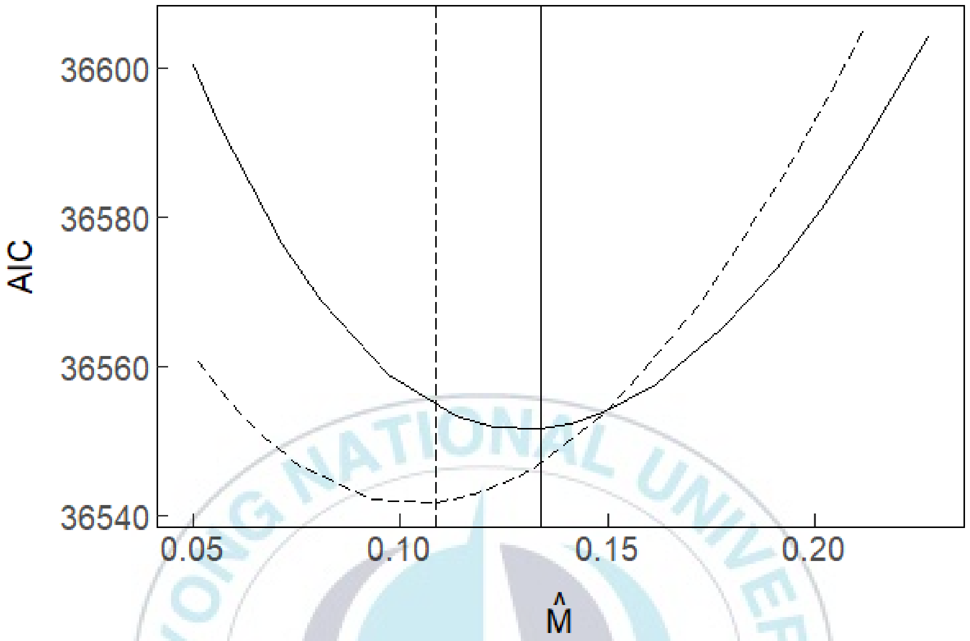


Fig. 6. AIC values for the estimated M ($\hat{M}$) using aggregated length-weight data (Case I, solid convex line) and annual length-weight data (Case II, broken convex line). The maximum age, A was set to be six years in both cases. $\hat{M}$ were estimated by adjusting the mode and coefficient of variation of the prior distribution for M. The solid vertical line denotes $\hat{M}$ (=0.13 year^{-1}) for the minimum AIC (=36551.6) from Case I. The broken vertical line denotes $\hat{M}$ (=0.11 year^{-1}) for the minimum AIC (=36541.6). For the estimates of the parameters other than M, see Tables 1 and 2.

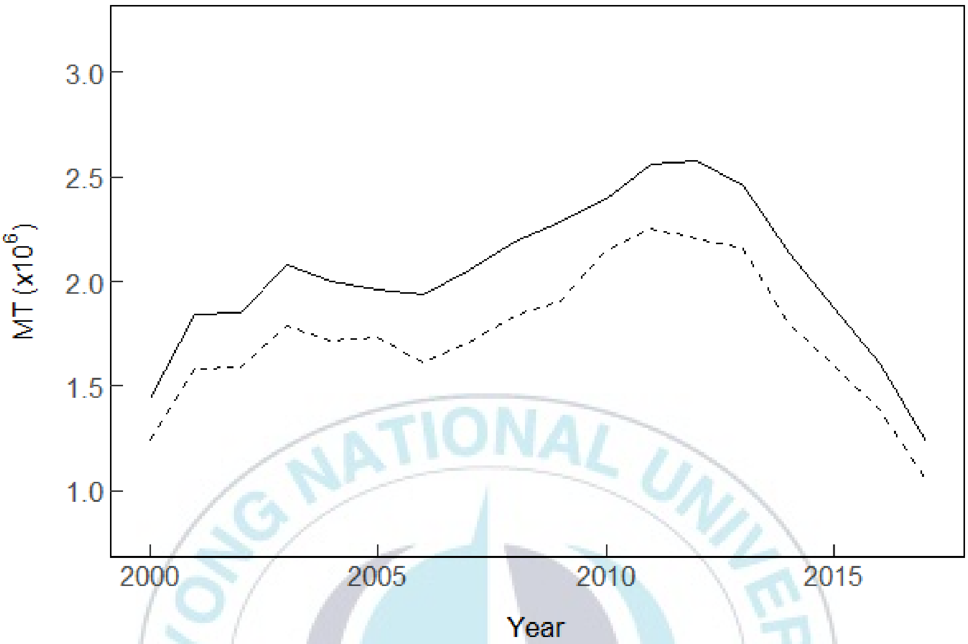


Fig. 7. Predicted annual biomass of Korean chub mackerel stock from the model with the best fit in Cases I and II. The solid line and broken line denote the predicted annual biomass from Case I (using aggregated length-weight data) and Case II (using annual length-weight data) respectively. The vertical axis denotes biomass in MT ($\times 10^6$) and the horizontal axis denotes time in years.

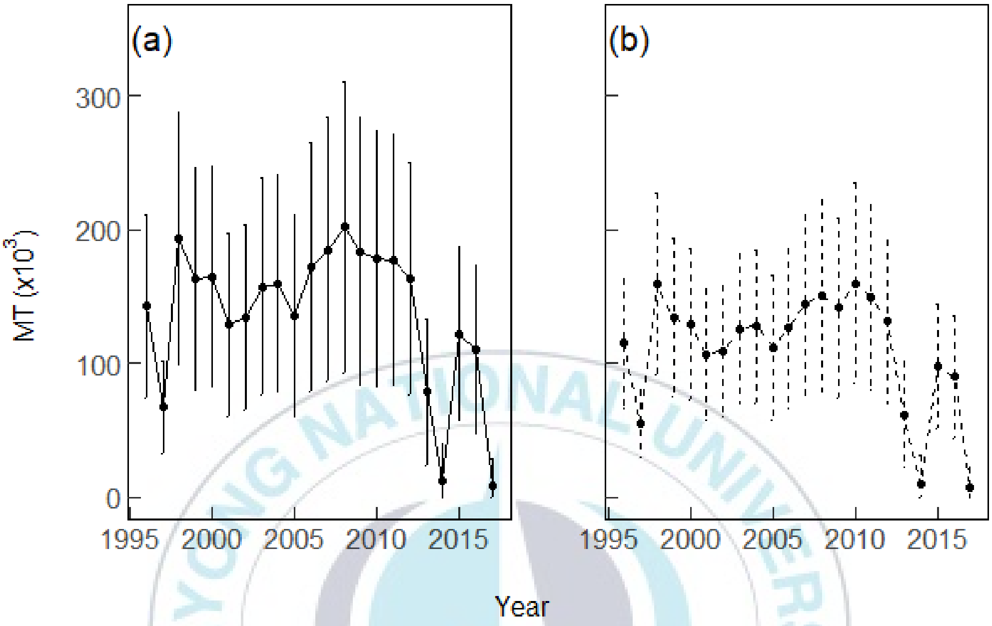


Fig. 8. Estimated the annual recruitment in weight (1996-2017) of Korean chub mackerel stock from the model with the best fit in Case I (panel (a)) and Case II (panel (b)). The points in panel (a) and (b) denote the estimated recruitments in weight from Case I (using aggregated length-weight data) and Case II (using annual length-weight data). The vertical lines around the points denote the standard errors of the estimates. The vertical axis denotes biomass in MT ($\times 10^3$) and the horizontal axis denotes time in years.

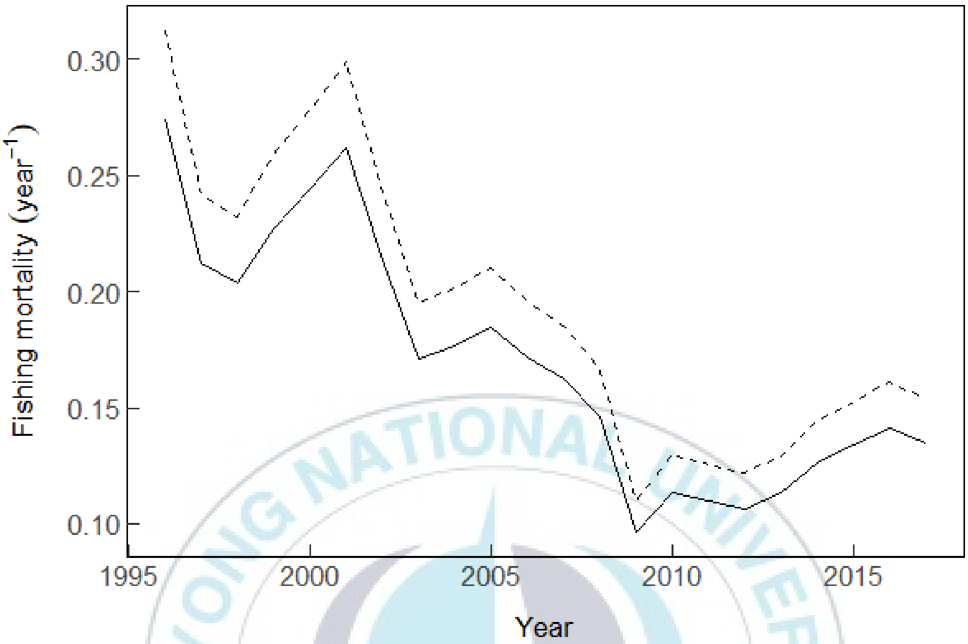


Fig. 9. Predicted annual instantaneous fishing mortality (1996-2017) of Korean chub mackerel stock from the model with the best fit in Case I and Case II. The solid and broken lines denote the predicted annual instantaneous fishing mortality from Case I (using aggregated length-weight data) and Case II (using annual length-weight data) respectively. The vertical axis denotes instantaneous fishing mortality per year and the horizontal axis denotes time in years.

4. Discussion

4.1. Recruitment parameters

In this model, the recruitment parameters accounted for most of the free parameters, as in Quinn's model. However, Quinn et al. (1998) were not able to estimate all the recruits, so they used values for eight of the 31 recruits as derived parameters. Unlike Quinn’s assumption, I tried to set all recruits as free parameters. Since the LF data begin from 2000, recruits start from 1995, under the assumption that A is six. However, in this case, the results were unacceptable because the predicted F_j or estimated M was extremely low. Hence, the recruits in 1995 were ignored, and recruits from 1996 were estimated (Fig. 10). The recruits in 1995 could be ignored because of the characteristics of the LF data from 2000. The recruits in 1995 were only supported by the LF data from 2000, and the length frequency distribution of the catch should be fitted as the last age-class with the LF data in 2000. However, the LF data in 2000 were unusual among the annual LF datasets (2000-2017) because the sample size is the smallest (n_{2000} =3033), and it is the only data set that does not have samples over 40cm in length (Fig. 4 (a)).

Consequently, there are insufficient data for LF in 2000, hindering estimation of recruits in 1995.

Year	Age1	Age2	Age3	Age4	Age5	Age6
1995	~~A~~					
1996	B	~~A~~				
1997	C	B	~~A~~			
1998	D	C	B	~~A~~		
1999	E	D	C	B	~~A~~	
2000	F	E	D	C	B	~~A~~
2001	G	F	E	D	C	B
2002	H	G	F	E	D	C
2003	I	H	G	F	E	D
⋮	⋮	⋮	⋮	⋮	⋮	⋮
2014	T	S	R	Q	P	O
2015	U	T	S	R	Q	P
2016	V	U	T	S	R	Q
2017	W	V	U	T	S	R

Fig. 10. Age-time structure in this model. The maximum age was assumed to be six years. The Age 1 class was assumed to be recruits (Column ‘Age1’). ‘A’ to ‘W’ denote different cohorts. Since the observed annual length frequency were fitted to the model from 2000 to 2017, the column ‘Age1’ begins from 1995. However, the best fit was obtained when cohort A was removed. The length frequency data from 2000 were fitted with five age classes, and the others (2001 to 2017) were fitted with six age classes.

4.2. Maximum age

The maximum age A refers to the number of age classes fitted to the annual LF data. In this model, there are no age classes over A. Choi et al. (2000) reported that the maximum age of Korean chub mackerel was six years, and Hernández and Ortega (2000) reported that 11-year-old mackerel had been caught in the Pacific Northwest. Based on this information, several values of A were tried. The best results were found at A=6.

Since estimations were failed when A was seven years or older, the longevity of chub mackerel of 11 years reported by Hernández and Ortega (2000) can be regarded as an extraordinary value of A under this model. Hernández and Ortega (2000) reported that the age composition of mackerel was dominated by the two- to four-year-old individuals, and individuals older than six years were very uncommon. This report is consistent with the results of our model, in which the best fit was shown at A=6. Consequently, A in our model can be understood as the number of major age classes in the catch, rather than as the longevity of a species.

4.3. Growth

This model shows the change in the length frequency distribution of a cohort over time. In practice, the variance of length at a specific age tends to increase with increasing age (Schnute and Founier 1980; Jones 1987; Parma and Dersio 1990), and the model results were consistent with this observation (Fig. 11). Since the length distribution of recruits of all cohorts in the model shared the parameters (μ_r, σ_r^2), and their growth also shared the parameters of the LVB growth equation (L_∞, κ, σ), the mean lengths of all cohorts at the same age were calculated as being similar. From Case II, the average of the mean lengths of all cohorts from ages one to six were 18.00cm, 21.38cm, 24.40cm, 27.08cm, 29.46cm, and 31.59cm. However, according to the LVB growth equation reported by Choi et al. (2000), the calculated lengths from ages one to six were 17.97cm, 26.67cm, 33.13cm, 37.92cm, 41,47cm, and 44.11cm (Fig. 11). Choi et al. (2000) used samples of Korean chub mackerel from May 1996 to Dec 1998 and I used LF data from Jan 2000 to Dec 2017. Therefore, it appears that the length of Korean chub mackerel stock decreased in 2000-2017 compared to 1996-1998.

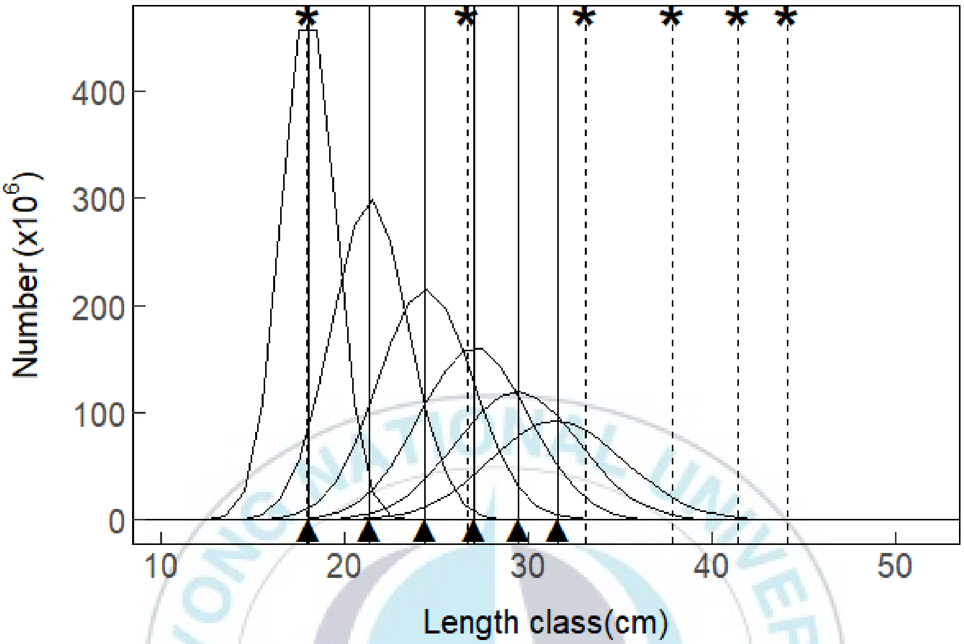


Fig. 11. Year-to-year changes in the length frequency distribution of the cohort recruited in 2005 from the model (Case II). The six bell-shaped curves represent the length frequency distribution of abundance from 2005 to 2010, from left to right. The class interval of the horizontal axis is 1cm. The vertical axis denotes the number of individuals. The vertical solid lines with lower triangles represent the mean length of the cohort by year. The broken lines with upper asterisks represent the lengths by age from one to six according to the LVB growth equation published by Choi et al. (2000).

4.4. Sample size in LF

The three years with the largest normalized difference between the LF data and the fitted values were 2000, 2001 and 2005, and only these datasets have less than 10,000 samples (n_{2000} =3,033, n_{2001} =9,424, n_{2005} =8,935, $n_{average}$ =15,120). This observation implies the existence of a relationship between the sample size and the model fit.

It is obvious that larger sample sizes are more representative of a population, and insufficient sample sizes may yield false representations. For example, in a length frequency analysis, a histogram of frequencies of length often shows distinct modes that hypothetically represent distinct age-classes (Quinn and Deriso 1999), and the number of modes can be affected by the sample size (see Figure 1 in Carlile 2005). However, taking either too few or too many samples can be wasted effort (Miranda 2007). It is not straightforward to determine an optimal sample size, because optimal sample size varies with bin width, the range of lengths, the number of age classes, and life history characteristics (Erzini 1990). According to Miranda (2007), smaller species, smaller populations, and populations with higher mortality required fewer samples.

Despite those difficulties, previous studies have suggested sample sizes for length data. Erzini (1990) suggested that sample sizes greater than 1000 are required in order to identify more than half the modes in a typical distribution. Carlile (2005) suggested that the use of 1000-2000 samples stabilizes the distribution. Miranda (2007) found that 1cm length-frequency histograms required 375-1200 samples. Grant et al. (1987) suggested that very large samples sizes (>1000) were necessary for computer-based methods if modes were obscured. Overall, the authors suggested sample sizes of more than 1000.

Therefore, accepting the above criteria and assuming that the optimum sample size for the monthly LF data is 1000, there will be 12,000 samples per year. This value lends weight to the hypothesis that the differences between the fitted values and the data shown in the 2000, 2001 and 2005 were due to small sample sizes. However, this suggestion is only based on the references and was not examined by simulation or experiment. Therefore, further study is needed to determine optimal sample sizes for this model.

4.5. Length-weight data

Comparing Case I and Case II, the lowest AIC was found in Case II; better results were obtained with more detailed length-weight information.

The α_j and β_j used in Case II are more effective because they did not lose the annual information in the length-weight data, as happened with the aggregated α and β used in Case I. However, the approach in Case II does not always guarantee better results. When data are divided into smaller time units, outliers can be a problem. The effect of outliers can be more significant when dividing the data by year (Fig. 2 (f), (k), (l)). Also, if the sample size is small, outliers can distort estimates (Fig. 2, (b)). However, in this study, the improved results of Case II were obtained without removing outliers, suggesting that attempts to include more detailed data are valuable.

4.6. Parameter estimation

Quinn et al. (1998) found the best fit in their model by reducing the number of free parameters. This reduction involved setting some of the recruits as the derived parameters, and defining M or q as an input value. For simplicity, Quinn et al. (1998) assumed three time-invariant factors: a discrete length frequency distribution of recruits (μ_r, ${\sigma_r}^2$); an LVB growth equation (L_∞, κ, σ_L); and instantaneous natural mortality (M). These factors were in our model.

The major difference between the model described here and that of Quinn and colleagues is the definition of the objective function as the negative log-likelihood. Considering the results, the distributions assumed for LF and yield data in the objective function seem reasonable. However, after estimating parameters, even if the model values fitted well to the data, the model assumptions were reexamined when the estimates or calculated values were unreasonable or extreme.

5. Conclusions

The model described here was based upon Quinn's length-based model and provides age-structured information on chub mackerel stock by applying time-series length data. Several factors are important for successful estimation, including the assumptions made about the data and parameters, the biological characteristics of the fish species, and the appropriate sample size for the length data. Therefore, if this method was to be applied to other fish species in the future, these factors should be considered.

This study was part of an effort to conduct a stock assessment in a situation in which data, particularly age data, are limited. A length-based age-structured model was used to overcome the lack of age data, but in practice, the model still suffers from a problem of lack of length data due to factors such as insufficient sample sizes and time series with missing data periods. Further study is needed to overcome the problems of availability and use of data when applying the length-based model.

Acknowledgments

There have been a number of people without whom this thesis could not have been completed. My deepest gratitude goes out to Prof. Saang-Yoon Hyun. His dedicated guidance and teachings regenerated me. Without him, this thesis would not have even begun. I would like to thank my committee members: Prof. Chul Woong Oh for his teachings and guidance; and Dr. Jae Bong Lee for his perceptive comments that has improved this thesis. Also, I wish to thank Dr. Young Il Seo for his answers to my inquiry. I would like to express sincere gratitude to Kyuhan Kim and Hyotae Lee for their careful advice. Also, I thank my colleagues, Min Gyou Park, Yuri Jung and Dongyoung An for their support.

This study was funded by the Korean National Institute of Fisheries Science (KNIFS) under project No. CD20180300. The data used in this thesis were provided from KNIFS and Statistics Korea.

References

Carlile, D. 2005. An assessment of age determination needs and samples sizes for groundfish fisheries managed by the state of alaska. Alaska Department of Fish and Game, Special Publication 05-12.

Choi, Y.M., Park, J.H., Cha, H.K., and Hwang, K.S. 2000. Age and growth of common mackerel, *Scomber japonicus* houttuyn , in korean waters. Journal of the Korean Society of Fisheries Resources. 3: 1-8.

Choi, Y.M., Zhang, C.I., Kim, Y.S., Baik, C.I., and Park, Y.C. 2004. Ecological characteristics and biomass of chub mackerel, *Scomber japonicus* houttuyn in korean waters. Journal of the Korean Society of Fisheries Resources. 7: 79-89.

Cohen, M., and Fishman, G.S. 1980. Modeling Growth–Time and Weight–Length relationships in a single year-class fishery with examples for north carolina pink and brown shrimp. Can. J. Fish. Aquat. Sci. 37(6): 1000-1011. doi: 10.1139/f80-128.

Deriso, R.B., and Parma, A.M. 1988. Dynamics of age and size for a stochastic population model. Can. J. Fish. Aquat. Sci. 45(6): 1054-1068.

doi: 10.1139/f88-129.

Erzini, K. 1990. Sample size and grouping of data for length-frequency analysis. Fisheries Research. 9(4): 355-366. doi: 10.1016/0165-7836(90)90053-X.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J. 2012. AD model builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Software. 27(2): 233-249. doi: 10.1080/10556788.2011.597854.

Grant, A., Morgan, P.J., and Olive, P.J.W. 1987. Use made in marine ecology of methods for estimating demographic parameters from size/frequency data. Mar. Biol. 95(2): 201-208. doi: 10.1007/BF00409006.

Hernández, J.J.C., and Ortega, A.T.S. 2000. Synopsis of biological data on the chub mackerel (Scomber japonicus houttuyn, 1782). Food and Agriculture Organization of the United Nations, FAO Fisheries Synopsis No. 157.

Hiyama, Y., Yoda, M., and Ohshimo, S. 2002. Stock size fluctuations in chub mackerel (Scomber japonicus) in the east china sea and the Japan/East sea. Fish. Oceanogr. 11(6): 347-353. doi: 10.1046/j.1365-

2419.2002.00217.x.

Jones, R. 1987. An Investigation of Length Composition Analysis Using Simulated Length Compositions. D. Pauly, G.R.s. Morgan (Eds.), Length-Based Methods in Fisheries Research, ICLARM & Kowait Institute for Scientific Research, Sofat, Kowait (1987), pp. 217-238.

Kim, K., Hyun, S., and Seo, Y.I. 2018. Inference of age compositions in a sample of fish from fish length data. Korean Journal of Fisheries and Aquatic Sciences. 51(1): 79-90.

Miranda, L.E. 2007. Approximate sample sizes required to estimate length distributions. Trans. Am. Fish. Soc. 136(2): 409-415. doi: 10.1577/T06-151.1.

NPFC. 2019. NPFC-2019-AR-Annual Summary Footprint – Chub&Spotted mackerels (Rev 1). https://www.npfc.int/summary-footprint-chub-mackerel-fisheries.

Parma, A.M., and Deriso, R.B. 1990. Dynamics of age and size composition in a population subject to size-selective mortality: Effects of phenotypic variability in growth. Can. J. Fish. Aquat. Sci. 47(2): 274-289. doi: 10.1139/f90-030.

Punt, A.E., Huang, T., and Maunder, M.N. 2013. Review of integrated sizestructured models for stock assessment of hard-to-age crustacean and mollusc species. Icesjms. 70(1): 16-33. doi: 10.1093/icesjms/fss185.

Quinn, T.J.,II, and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press.

Quinn, T.J.,II, Turnbull, C.T., and Fu, C. 1998. A length-based population model for hard-to-age invertebrate populations. In Fishery stock assessment models. Edited by F. Funk, T.J. Quinn II, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Rep. No. 98-01 of the Alaska Sea Grant College Program, University of Alaska–Fairbanks, Juneau. pp. 531–556.

Schnute, J., and Fournier, D. 1980. A new approach to Length–Frequency analysis: Growth structure. Can. J. Fish. Aquat. Sci. 37(9): 1337-1351. doi: 10.1139/f80-172.

Wang, Y., Zheng, J., and Yu, C. 2014. Stock assessment of chub mackerel (Scomber japonicus) in the central east china sea based on length data. J. Mar. Biol. Assoc. U. K. 94(1): 211-217. doi: 10.1017/S0025315413001434.

Appendix 1. Detailed descriptions of the length-based model

It was assumed that the length frequency of Korean chub mackerel at recruitment follows a discrete normal distribution ($Normal_D$) with mean μ_r and variance $\sigma_r^{\,2}$ ($X \sim Normal_D(\mu_r, \sigma^2_{\ r})$). The probability mass function (PMF) for the i-th fork length class x_i in the start of year j and at recruiting age r is:

$$f_{j,r}(x_i) = \exp\left(-\frac{1}{2\sigma^2_{\ r}} \left(x_i - \mu_r \right)^2 \right) \Big/ \xi_{j,r} \ , \qquad \text{(A.1)}$$

$$\text{where } \ \xi_{j,r} = \sum_x \exp\left(-\frac{1}{2\sigma_r^{\ 2}} \left(x_i - \mu_r \right)^2 \right).$$

The constant $\xi_{j,r}$ is a normalizing constant so that the sum of equation (A.1) over X equals 1. The number of individuals of the newly recruited mackerel in the i-th length class x_i is $N_{j,r}\left(x_i\right) = N_{j,r} \cdot f_{j,r}\left(x_i\right)$. To obtain the abundance $N_{j+1,a+1}\left(x_i\right)$ at the start of year $j+1$ at age $a+1$ from abundance $N_{j,a}\left(x_i\right)$ (starting at age r), the process of mortality was accounted for first, followed by growth.

The total mortality $Z_j\left(x_i\right)$ is composed of time-independent natural mortality M and fishing mortality $S\left(x_i\right)F_j$, where F_j is full fishing mortality:

$$Z_j\left(x_i\right) = M + S\left(x_i\right)\cdot F_j, \tag{A.2}$$

where $S\left(x_i\right)$ is the gear selectivity function that follows the logistic curve:

$$S\left(x_i\right) = \frac{1}{1+\exp\left(-\gamma(x_i - L_{50\%})\right)}, \tag{A.3}$$

where $L_{50\%}$ is the size where 50% of the fish are vulnerable and γ is the shape parameter.

Natural mortality M is assumed to be a constant over lengths and years, and full fishing mortality F_j may be approximately related to fishing effort as

$F_j = q \cdot Effort_j$, where the catchability q is assumed to be a constant, $Effort_j$ is the fishing effort in year j: $Effort_j = Y_j / CPUE_j$

For a given PMF $f_{j,a}(x_i)$ in year j at the start of age a, the relative distribution of lengths in the population after mortality occurs in year j is:

$$p_{j,a,Z}(x_i) = f_{j,a}(x_i) \cdot \exp\left(-M - S(x_i) F_j\right) = f_{j,a}(x_i) \cdot \exp(-Z_j(x_i)) \,. \qquad \text{(A.4)}$$

To account for growth, it is assumed that an individual of the i-th length x_i will grow to length l_i in one time step according to a stochastic growth model. One useful model is the von Bertalanffy (LVB) model with stochastic error, which was first derived in Cohen and Fishman (1980) and utilized in the length-based model of Deriso and Parma (1988). The deterministic LVB model is:

$$L_a = L_\infty (1 - e^{-\kappa(a - a_0)})$$

where L_∞ is asymptotic length, κ is a growth parameter, and a_0 is the age corresponding to length 0. An equivalent formulation for size L_{a+1} at age $a+1$ as a function of previous size L_a with the inclusion of a stochastic term is:

$$L_{a+1} = L_\infty (1-\rho) + \rho L_a + \varepsilon \tag{A.5}$$

where ρ is $\exp(-\kappa)$, ε is an independent, normally distributed random variable with mean zero and variance ${\sigma_L}^2$. From Cohen and Fishman (1980), the expected length and variance at age $a+1$ for an individual of the i-th length x_i at age a, considering that it was recruited at age r, are:

$$\mu_{a+1}(x_i) = L_\infty (1-\rho) + \rho x_i \tag{A.6}$$

and

$$\sigma^2{}_{a+1} = {\sigma_L}^2 \frac{1-\rho^{2(a+1-r)}}{1-\rho^2} + \rho^{2(a+1-r)} {\sigma_r}^2 . \tag{A.7}$$

The normal PMF for the length (l_i) distribution after one growth increment for an individual originally of the i-th length (x_i), $L \sim Normal_D(\mu_{a+1}(x_i), \sigma^2_{a+1})$, is given by:

$$f_{j+1,a+1,G}(l_i|x_i) = \exp\left(-\frac{1}{2\sigma^2_{a+1}}\left(l_i - \mu_{a+1}(x_i)\right)^2\right)\Big/\xi_{j+1,a+1,x_i}, \qquad \text{(A.8)}$$

$$\text{where} \quad \xi_{j+1,a+1,x_i} = \sum_{l} \exp\left(-\frac{1}{2\sigma^2_{a+1}}\left(l_i - \mu_{a+1}(x_i)\right)^2\right).$$

The relative distribution of lengths (l_i) at the start of age $a+1$ is then obtained from the relative distribution of lengths after mortality equation (Equation A.4) and the PMF for growth equation (Equation A.8), which results in:

$$p_{j+1,a+1}(l_i) = \sum_{x} f_{j+1,a+1,G}(l_i|x_i) \cdot p_{j,a,Z}(x_i) \qquad \text{(A.9)}$$

The number of individuals at the i-th length class l_i, $N_{j+1,a+1}(l_i)$, year $j+1$

at age $a+1$ is then:

$$N_{j+1,a+1}\left(l_i\right)=N_{j,a}\cdot p_{j+1,a+1}\left(l_i\right) \tag{A.10}$$

and

$$f_{j+1,a+1}\left(l_i\right)=p_{j+1,a+1}\left(l_i\right)\Big/\sum_l p_{j+1,a+1}\left(l_i\right). \tag{A.11}$$

The total number of individuals in year $j+1$ at age $a+1$ is then:

$$N_{j+1,a+1}=\sum_l N_{j+1,a+1}(l_i). \tag{A.12}$$

From Deriso and Parma (1988), the relative distribution of lengths in the catch followed the Baranov catch equation $C=N\left(F/Z\right)\left[1-\exp\left(-Z\right)\right]$ and is given by:

$$p_{j,a,C}(x_i) = f_{j,a}(x_i)\frac{S(x_i)F_j}{Z_j(x_i)}\left(1-\exp\left(Z_j(x_i)\right)\right). \tag{A.13}$$

The number of caught fish of the i-th length x_i in year j at age a follows:

$$C_{j,a}(x_i) = N_{j,a} \cdot p_{j,a,C}(x_i). \tag{A.14}$$

The total catch in year j at age a is:

$$C_{j,a} = \sum_x C_{j,a}(x_i). \tag{A.15}$$

Weight at length was modeled as the allometric relationship,

$$W_j(x_i) = \alpha_j \cdot x_i^{\beta_j}. \tag{A.16}$$

And then length-specific values were multiplied by the corresponding weight. Thus biomass (abundance in weight) and yield (catch in weight) were calculated by multiplying length-specific abundance and catch, respectively, by weight at that length and summing over length, or

$$B_{j,a} = \sum_x N_{j,a}\left(x_i\right) \cdot W\left(x_i\right) \tag{A.17}$$

and

$$Y_{j,a} = \sum_x C_{j,a}\left(x_i\right) \cdot W\left(x_i\right). \tag{A.18}$$

This formulation covered the progress of a single year class through its lifespan. Identical equations could be constructed for all year classes and referenced to the passage of time. Thus, the length distributions of the population at a given time from this model could be assembled by summing over all age classes present. The number of individuals at the i-th length x_i from a cohort in year j at age a follows:

$$N_{j,a}\left(x_i\right) = N_{j,a} \cdot f_{j,a}\left(x_i\right). \tag{A.19}$$

It is assumed that the maximum age of mackerel is six years, thus, the population would consist of six cohorts:

$$N_j\left(x_i\right)=\sum_{a=1}^{A} N_{j,a}\left(x_i\right). \tag{A.20}$$

In the same way, the number of caught fish of the i-th length in year j follows:

$$C_j\left(x_i\right)=\sum_{a=1}^{A} C_{j,a}\left(x_i\right). \tag{A.21}$$

Consequently, the total biomass of the population and total yield in year j follows:

$$B_j=\sum_{x} N_j\left(x_i\right)\cdot W\left(x_i\right) \tag{A.23}$$

and

$$Y_j = \sum_{i=1}^{42} C_j\left(x_i\right) \cdot W\left(x_i\right). \tag{A.24}$$

Appendix 2. The notations used in the model

Indices

a	Age.
r	Recruitment age.
A	Maximum age.
i	Index for length: $1, 2, \cdots, 42$.
j	Index for year.
x_i	Mid-point of i-th length class.
l_i	Mid-point of i-th length class after one growth increment.
L_a	Length at age a.
ε	Stochastic error term in the growth equation, $\varepsilon \sim Normal(0, {\sigma_L}^2)$.
ε_j	Stochastic error term in the multiplicative error model, $\varepsilon_j \sim Normal\left(0, {\sigma_Y}^2\right)$.
a_0	Age corresponding to length 0.

J	The objective function.
λ_1, λ_2	Weighting terms.
k	Number of free parameters (= 28).
^	In this paper, the hat notation denotes a predicted or fitted quantity.
~	In this paper, the tilde mark denotes a vector.

Data

m_{ij}	The number of samples of the i-th length class in the length frequency data in year j.
n_j	The sample size of the length frequency data in year j.
Y_j	The observed yield in year j.
d	The time length of the observed yield.
$CPUE_j$	The catch-per-unit-effort data from the large purse seine fisheries in year j.
$Effort_j$	The fishing effort in year j: $Effort_j = Y_j / CPUE_j$.
L_∞	The asymptotic length.
μ_r	The mean length at recruitment.

σ_r	The standard deviation of the length frequency normal distribution at recruitment.
α_j, β_j	Parameters in the allometric length-weight relationship in year j.
μ_M	Mean of a prior distribution for M.
$\sigma_M{}^2$	Variance of a prior distribution for M.
Mode_M	Mode of a prior distribution for M: $\mathrm{Mode}_M = \exp\left(\mu_M - \sigma_M{}^2\right)$.
CV_M	Coefficient of variance of a prior distribution for M: $\mathrm{CV}_M = \sigma_M / \mu_M$.

Free parameters

$N_{j,r}$	Recruiting individuals in year j. Also, $\underset{\sim}{N}_r = (N_{1996,r},\ N_{1997,r},\ \cdots,\ N_{2017,r})$.
M	Instantaneous natural mortality rate.
q	Catchability coefficient.
$L_{50\%}$	Length at which 50% of the fish entering the gear is retained.

γ	Shape parameter for the gear selectivity.
κ	Brody growth coefficient.
σ_L	The standard deviation of stochastic error term in the growth equation: $L_{a+1}=L_{\infty}(1-e^{-\kappa})+e^{-\kappa}L_a+\varepsilon,\ \ \varepsilon \sim N\left(0,\sigma_L{}^2\right).$

Derived parameters

$f_{j,a}\left(x_i\right)$	The probability of individuals at the i-th length class x_j of the cohort in the start of year j for age a.
$f_{j+1,a+1,G}\left(l_i \middle\| x_i\right)$	The conditional probability of individuals at the i-th length class l_i after one growth increment for individuals originally of the i-th length class x_i.
$\xi_{j,a}$, ξ_{j,a,x_i}	Normalizing constant.
$p_{j,a,C}\left(x_i\right)$	The relative distribution of lengths in the catch for the cohort at age a in year j.
$p_{j,a,Z}\left(x_i\right)$	The relative distribution of lengths in the cohort at age a after mortality occurs in year j.

$p_{j+1,a+1}\left(l_i\right)$	The relative distribution of lengths after one year of mortality and growth.
$N_{j,a}$	The number of individuals of the cohort at age a in year j.
$N_{j,a}\left(x_i\right)$	The number of individuals at the i-th length class x_i in the cohort at age a in year j.
$N_{j,\bullet}\left(x_i\right)$	The number of individuals at the i-th length class x_i in year j. Also, $N_{j,\bullet}\left(x_i\right)=\sum_{a=1}^{A}N_{j,a}\left(x_i\right)$.
$C_{j,a}$	The number of individuals of caught fish at age a in year j.
$C_{j,a}\left(x_i\right)$	The number of individuals of caught fish in the i-th length class x_i at age a in year j.
$C_{j,\bullet}\left(x_i\right)$	The number of individuals of caught fish at the i-th length class in year j. Also, $C_{j,\bullet}\left(x_i\right)=\sum_{a=1}^{A}C_{j,a}\left(x_i\right)$.
$W\left(x_i\right)$	Allometric length-weight relationship: $W\left(x_i\right)=\alpha\cdot x_i^{\beta}$

$B_{j,a}$	The biomass of the stock at age a in year j.
$B_{j,\bullet}$	Total biomass of the stock in year j. Also, $B_{j,\bullet} = \sum_{a=1}^{A} B_{j,a}$.
$\hat{Y}_{j,\bullet}$	The predicted yield in year j. Also, $\hat{Y}_{j,\bullet} = \sum_{a=1}^{A} \hat{Y}_{j,a}$.
${\sigma_{\log Y}}^2$	Time-constant variance, where $\log Y_j \sim Normal\left(\log \hat{Y}_j, {\sigma_{\log Y}}^2\right)$.
o_{ij}	The probability of how many caught fish are assigned in the i-th length class in year j.
F_j	Instantaneous fishing mortality rate in year j.
$S(x_i)$	Gear selectivity function; A probability that fish in the i-th length class is caught by the fishing gear: $S(x_i) = 1 \big/ \left[1 + \exp\left(-\gamma\left(x_i - L_{50\%}\right)\right)\right]$.
$Z_j(x_i)$	Total mortality for a length class x_i in year j.

$\mu_{a+1}\left(x_i\right)$	The expected length at age $a+1$ after growth for individuals in a length class x_i at age a from Cohen and Fishman (1980): $\mu_{a+1}\left(x_i\right)=L_\infty\left(1-\rho\right)+\rho\cdot x_i$.
$\sigma^2{}_{a+1}$	The variance at age $a+1$ after growth for individuals from Cohen and Fishman (1980): $\sigma^2{}_{a+1}=\sigma_L{}^2\dfrac{1-\rho^{2(a+1-r)}}{1-\rho^2}+\rho^{2(a+1-r)}\sigma_r{}^2$.
ρ	$\exp(-\kappa)$.

Appendix 3. ADMB code for the length-based model (Case I)

TPL file

```
//lengthSA: a length-based stock assessment for the Korean mackerel
population;
//data file name: mackerel_aggre.DAT
//Author: Saang-Yoon Hyun and Jinwoo Gim as of Aug 23, 2019
//
DATA_SECTION
   init_int nages;                    // 6 years
   init_int nlengths;                 // 42 classes
   init_int nyrs;                     // 22 years (1996-2017)
   init_matrix yieldcpue(1,nyrs,1,3);
   vector years(1,nyrs);
   !!years=column(yieldcpue,1);

   init_int indexMinyrLD;    //2000 ==> index of 5;
   int nyrsLD;           //nyears for the length data
   !!nyrsLD=nyrs-4;   //i.e., 2000-2017;

   init_vector x(1,nlengths);     //discrete lengths;
   vector L(1,nlengths);          //discrete lengths;
   !!L=x;

   init_matrix lengthfrq(1,nyrsLD,1,nlengths);

   vector yield(1,nyrs);              // in MT
   !!yield=column(yieldcpue,2);
   vector CPUE(1,nyrs);           // in MT/haul
   !!CPUE=column(yieldcpue,3);
   vector Effort(1,nyrs);
   !!Effort=elem_div(yield,CPUE); // vector(1,23) in hauls
```

```
   init_vector musig2_r(1,2);
   number mu_r;
   number sig2_r;
   !!mu_r=musig2_r(1);
   !!sig2_r=musig2_r(2);
   init_vector abWL(1,2);
   number aWL;
   number bWL;
   !!aWL=abWL(1);
   !!bWL=abWL(2);

   vector Wt(1,nlengths);

   //prior for M ~ lognormal(mu, sigma2)
   init_number mode_M;   //init_number M;   // Natural mortaility as a salar;
   init_number CV_M;
   number mu_M;
   number sig_M;

   init_number lambda1; // lambda in the objective function
   init_number lambda2; // lambda in the objective function

   init_number logLinf;
   number Linf;
   !!Linf=mfexp(logLinf);
   int r;
   !!r=1; //recruitment is defined as the pop size at age 1;
   int ncohorts;
   !!ncohorts=nyrs;

   ivector SamSize(1,nyrs);     //annual sample size for the length data;

PARAMETER_SECTION
   init_number logq(1);              //log of the catchability in F_yr = q*Effort_yr;
```

```
   init_bounded_vector logRec(1,nyrs,11.5,25.0,2);
   init_bounded_number logkappa(-3.0,-0.80,3);
   init_bounded_number logL50(3.0,4.0,4);              //selectivity
   init_bounded_number gamma(0.05,1.5,5);             //selectivity
   init_bounded_number log_M(-2.996,-0.799,6);       //natural mortality;
   init_bounded_number sigmaL(0.0,1.5,6);              //the    uncertainty
in the L_{a+1} equation

   number M;
   number q;
   vector Recruits(1,nyrs);
   number kappa;
   number L50;

   vector Sel(1,nlengths);
   number Rho;
   number kkk; //for the cumulative purpose
   matrix f(1,nages,1,nlengths);   //length frequency as pmf
   3darray pp(1,nages,1,nlengths,1,nlengths); //pp(1 To Ages,1 To x,1
To L);
                                                 // x --> (growth) --> L
   vector Mu(1,nlengths); //differ by length class
   vector SS(1,nages);     //assumed to be constant over length classes

   vector F_yr(1,nyrs);              // fishing mortality;
   matrix F_tx(1,nyrs,1,nlengths);     //F_{t,x};
   matrix Z(1,nyrs,1,nlengths);
   matrix ExpZ(1,nyrs,1,nlengths);      //Exp(-Z);

   3darray NL(1,nages,1,ncohorts,1,nlengths);
   matrix N(1,nyrs,1,nages);

   number SumP;
   vector p(1,nlengths);
   number CNum;
   number CWt;
   vector TCatch(1,nyrs);
```

```
   matrix Catch(1,nyrs,1,nlengths);
   vector Yieldhat(1,nyrs);
   vector Pop(1,nyrs);
   3darray ENx(1,nages,1,ncohorts,1,nlengths);
   vector EN(1,ncohorts);
   vector B(1,ncohorts);
   vector EB(1,ncohorts);
   matrix LF(1,nlengths,1,nyrs);       //expected length-frequency by

   number logmult;                  //log(multinomial);
   vector elem_obj2(indexMinyrLD,nyrs);   //elements in part 2 of the
objective function;
   number sig2_yield;
   number lognormal;                //log(normal);
   number maxloglike;               // for calculation of AIC
   number aic;

   objective_function_value obj;

PRELIMINARY_CALCS_SECTION
   //re-arrangement of data for calculation purposes;
   //Calculate the weight relationship
   Wt=aWL*pow(x,bWL)/1000; //the division of 1000 is to convert gram
to kg

   //annual sample size for the length data
   int i;
   i=0;
   SamSize=0;
   for(int m=indexMinyrLD;m<=nyrs;m++) {
        i=m-4;
        SamSize(m)=sum(lengthfrq(i));
   };

   //prior for M
   mu_M=log(mode_M*(square(CV_M)+1));
   sig_M=sqrt( log(square(CV_M)+1) );
```

```
PROCEDURE_SECTION
   q=mfexp(logq);
   kappa=mfexp(logkappa);
   L50=mfexp(logL50);
   for(int i=1;i<=nyrs;i++)
         Recruits(i)=mfexp(logRec(i));
   M=mfexp(log_M);

   obj=0.0;

   //Calculate the selectivity for each length class;
   Sel=1.0/(1+mfexp(-1.0*gamma*(x-L50)));

   //Calculate fishing mortality for each length using F = q*Effort*Sel;
   //Sum_over x_F_{yr,x}    =/= F_{yr}; "=/=" "is not equal to"
   F_yr=q*Effort;         //1996 - 2017 (i.e., catch data)
   for(int t=1;t<=nyrs;t++)
      for(int xind=1;xind<=nlengths;xind++)    {
           F_tx(t,xind)=F_yr(t)*Sel(xind); //F_tx;
           Z(t,xind)=M+F_tx(t,xind);
           ExpZ(t,xind)=mfexp(-1.0*Z(t,xind));
      };

   //Calculating the length frequency dsn of the recruits;
   //recruitment is at one year of age;
   //f(1,x) is the same for all cohorts so this can be in the initial
calculations;
   Rho=mfexp(-1.0*kappa);
   SS(1)=sig2_r;
   kkk=0.0;
   for(int xind=1;xind<=nlengths;xind++) {
        f(1,xind)=mfexp(-1.0*square(x(xind)-mu_r)/(2.0*SS(1)));
        kkk=kkk+f(1,xind);
   };
```

```
   for(int xind=1;xind<=nlengths;xind++) {
        f(1,xind)=f(1,xind)/kkk;    //normalize;
        Mu(xind)=Linf-(Linf-x(xind))*Rho;
   };

   for(int a=2;a<=nages;a++)    {
          SS(a)=square(sigmaL)*(1.0-pow(Rho,(2.0*a-2.0*r)))/(1.0-
square(Rho))+(pow(Rho,(2.0*a-2.0*r)))*sig2_r;
                     //this SS is for Shrimp; //see Cohen and Fishman (1980)
   };

   for(int a=1;a<=nages;a++) {
        for(int xind=1;xind<=nlengths;xind++)    {
             kkk=0.0;
             for(int Lind=1;Lind<=nlengths;Lind++)    {
                    pp(a,Lind,xind)=mfexp(-1.0/(2.0*SS(a))*square(L(Lind)-
Mu(xind)));
                    kkk=kkk+pp(a,Lind,xind);
             };
             for(int Lind=1;Lind<=nlengths;Lind++)
                    pp(a,Lind,xind)=pp(a,Lind,xind)/kkk;
        };
   };

   //Start of cohort loop
   int m;
   for(int c=1;c<=ncohorts;c++)    {
        int a=1;
        m=c;
        N(m,a)=0.0;
        for(int xind=1;xind<=nlengths;xind++)    {
             NL(a,m,xind)=Recruits(m)*f(a,xind);
             N(m,a)=N(m,a)+NL(a,m,xind);       //Be careful that N has yrs x
ages;
        };

        for(int a=2;a<=nages;a++)      { //note that a starts at 2, not 1;
```

```
            m=a+c-1;
            if(m>nyrs)
                SumP=0.0;
            else     {
                    SumP=0.0;
                for(int Lind=1;Lind<=nlengths;Lind++)   {
                      p(Lind)=0;
                      for(int xind=1;xind<=nlengths;xind++)
                            p(Lind)=p(Lind)+f(a-1,xind)*ExpZ(m-
1,xind)*pp(a,Lind,xind);

                      SumP=SumP+p(Lind);
               };
               N(m,a)=0.0;
               for(int Lind=1;Lind<=nlengths;Lind++)   {
                      f(a,Lind)=p(Lind)/SumP;   //normalize;
                      NL(a,m,Lind)=N(m-1,a-1)*p(Lind);
                      N(m,a)=N(m,a)+NL(a,m,Lind);
               };
          };
       };
   };

   //================   //note m starts at indexMinyrLD;
   //Sum through the population matrix to calculate;
   //expected catch distribution;
   //summing for each time by length class for all ages;

   for(int m=indexMinyrLD;m<=nyrs;m++) {     //note m starts at
indexMinyrLD;
             TCatch(m)=0.0;
             Catch(m)=0.0;
             Yieldhat(m)=0.0;
             Pop(m)=0.0;
             EN(m)=0.0;
             B(m)=0.0;
             EB(m)=0.0;
```

```
             for(int a=1;a<=nages;a++)     {
                 for(int xind=1;xind<=nlengths;xind++) {
                             CNum=NL(a,m,xind)*(F_tx(m,xind)/Z(m,xind))*(1-
ExpZ(m,xind));
                             CWt=CNum*Wt(xind);   //in kg
                             Catch(m,xind)=Catch(m,xind)+CNum;
                             TCatch(m)=TCatch(m)+CNum;
                             Yieldhat(m)=Yieldhat(m)+CWt;

                             Pop(m)=Pop(m)+NL(a,m,xind);     //population;
                             ENx(a,m,xind)=NL(a,m,xind)*Sel(xind);
//Exploitable population;
                             EN(m)=EN(m)+ENx(a,m,xind);
                             B(m)=B(m)+NL(a,m,xind)*Wt(xind);      //Biomass;
//in kg
                             EB(m)=EB(m)+ENx(a,m,xind)*Wt(xind);
//exploitable biomass;
                 };
         };
   };

   //The expected length-frequency
   for(int m=indexMinyrLD;m<=nyrs;m++)     { //note m starts at
indexMinyrLD;
       for(int xind=1;xind<=nlengths;xind++)   {
              LF(xind,m)=(Catch(m,xind)/sum(Catch(m)))*SamSize(m);
           };
   };

   //part 1 of the objective function: multinomial
   logmult=0.0;
   for(int i=indexMinyrLD;i<=nyrs;i++) {
       logmult+=gammln(SamSize(i)+1);
       for(int xind=1;xind<=nlengths;xind++) {
           logmult+=-1.0*gammln((lengthfrq(i-4,xind)+1))+lengthfrq((i-
4),xind)*log(Catch(i,xind)/sum(Catch(i)));
       };
```

```
   };
   obj+=lambda1*(-1.0*logmult); //lamda1*(the negative multinomial
likelihood);

   maxloglike=0.0;
   maxloglike+=logmult;

   //part 2 of the objective function: normal
   sig2_yield=0.0;
   for(int m=indexMinyrLD;m<=nyrs;m++)
       elem_obj2(m)=square(log(yield(m))-log(Yieldhat(m)/1000)); //in
MT

   sig2_yield=sum(elem_obj2)/(nyrs-indexMinyrLD+1); //MLE of
sigma2
   obj+=lambda2*(0.5*(nyrs-
indexMinyrLD+1)*log(sig2_yield)+sum(elem_obj2)/(2.0*sig2_yield));
//lamda2*(the negative normal loglikelihood);

   lognormal=0.0;
   lognormal=(-0.5*(nyrs-indexMinyrLD+1)*log(2*M_PI)-0.5*(nyrs-
indexMinyrLD+1)*log(sig2_yield)-sum(elem_obj2)/(2.0*sig2_yield));

   maxloglike+=lognormal;

   //part 3: prior for M
   //prior for M ~lognormal(mu,sig2); // as the negative logarithm;
   obj+=0.5*log(2.0*M_PI)+log(sig_M)+log_M+square(log(M)-
mu_M)/(2.0*sig_M*sig_M);

   aic=-2.0*maxloglike+2.0*(nyrs+6); //check the number of free
parameters.


REPORT_SECTION
   report<<"Yr Recruits Fyr Yield Yieldhat"<<endl;
   for(int i=1;i<=nyrs;i++)
```

```
      report<<years(i)<<" "<<Recruits(i)<<" "<<F_yr(i)<<" "<<yield(i)<<" "<<Yieldhat(i)/1000<<endl;

   report<<"yrs Pop B.in.kg ExploitB"<<endl;
   for(int i=1;i<=nyrs;i++)
      report<<years(i)<<" "<<Pop(i)<<" "<<B(i)<<" "<<EB(i)<<endl;

   report<<"yrs N"<<endl;
   for(int i=1;i<=nyrs;i++)
      report<<years(i)<<" "<<N(i)<<endl;

   report<<"N:"<<N<<endl;
   report<<"lengthfrq: "<<lengthfrq<<endl;
   report<<"Catchhat: "<<Catch<<endl;
   report<<"max.grad: "<<objective_function_value::gmax<<endl;
   report<<"Sample size in the length frequency: "<<SamSize<<endl;

   report<<"x(1,nlengths): "<<x<<endl;
   report<<"Mu(1,nlengths): "<<Mu<<endl;

   report<<"SS(1,nages): "<<SS<<endl;
   report<<"M, sigmaL, and aic: "<<M<<" "<<sigmaL<<" "<<aic<<endl;


RUNTIME_SECTION
   maximum_function_evaluations 100,150,300,10000
   convergence_criteria .01,.0001,1e-7


TOP_OF_MAIN_SECTION
   gradient_structure::set_MAX_NVAR_OFFSET(1000);      //maximum number of depdendent variables of 400 exceeded
   gradient_structure::set_NUM_DEPENDENT_VARIABLES(1000);
   gradient_structure::set_GRADSTACK_BUFFER_SIZE(100000);
   gradient_structure::set_CMPDIF_BUFFER_SIZE(1000000);
   arrmblsize=900000;
```

```
GLOBALS_SECTION
   #include <admodel.h>
   #include <math.h>
   #include <stdio.h>
   #include <stddef.h>
   #include <stdlib.h>
```

DAT file (mackerel_aggre.DAT)

```
#nages: number of age classes
6
#number of length classes; #median length of maxFL = 51.5 cm;
42
#number of years for the fishery catch data: nyrs of 1996-2017;
22
#fishery (large-purse seine) catch
#       Yield(MT)       CPUE(MT/haul)
1996    415003.0
1997    160448.0          CPUE data are not allowed by the National
1998    172925.0          Institute of Fisheries Sciences (NIFS) to
1999    177540.0          be revealed.
2000    145908.0          If someone wants those data, get the
2001    203717.0          permission of NIFS.
2002    141751.0
2003    122044.0
2004    184274.0
2005    135596.0
2006    101427.0
2007    143776.0
2008    187240.0
2009    117960.0
2010    94331.0
2011    138729.0
2012    125143.0
2013    102114.0
2014    127452.0
2015    131735.0
2016    133200.0
2017    103870.0
#
#
#indexMinyrLD: index of the minimum year for the length data; //2000
5
#
```

```
#midpoints (cm) of length classes;
10.5    11.5    12.5    13.5    14.5    15.5    16.5    17.5    18.5
        19.5    20.5    21.5    22.5    23.5    24.5    25.5    26.5
        27.5    28.5    29.5    30.5    31.5    32.5    33.5    34.5
        35.5    36.5    37.5    38.5    39.5    40.5    41.5    42.5
        43.5    44.5    45.5    46.5    47.5    48.5    49.5    50.5
        51.5
#length frequency data (18 x 42);
#18: year 2000 - 2017
#42: midpoints (cm) of length classes: 10.5, 11.5, 12.5, ... , 51.5
```

Length data are not allowed by NIFS to be revealed.
If someone wants those data, get the permission of NIFS.

Length data are not allowed by NIFS to be revealed.
If someone wants those data, get the permission of NIFS.

Length data are not allowed by NIFS to be revealed.
If someone wants those data, get the permission of NIFS.

```
#
#Prior info
#Korean mackerel
#the mean (cm) and the variance (cm^2) of the lengths at the recruit stage (age 1);
#18.00 was calculated from the Linf below, Linf=exp(3.95);
18.00   2.10
#
#parameters in W =a*(L^b); #W in gram; # L in cm;
#a and b
```

```
#We estimated these using actual data;
0.003   3.425
#
#M: Natural mortality (instantaneous rate) as a prior dsn;
#mode of M
5.0
#CV of M
2.0
#
#lambda1 for the length data in the objective function
0.05
#lambda2 for the yield data in the objective function;
10.0
#
#logLinf
3.95
```

PIN file

```
#logq
-9.5
#logRec(1,nyrs)
20      20      20      20      20      20      20      20      20      20
        20      20      20      20      20      20      20      20      20
        20      20      20
#logkappa
-1.21
#logL50 in the selectivity
3.38
#gamma in the selectivity
0.6
#log_M
-1.386
#sigmaL
0.66
```

Appendix 4. ADMB code for the length-based model (Case II)

TPL file

```
//lengthSA: a length-based stock assessment for the Korean mackerel population;
//data file name: mackerel_annual.DAT
//Author: Saang-Yoon Hyun and Jinwoo Gim as of Aug 23, 2019
//
DATA_SECTION
   init_int nages;                  // 6 years
   init_int nlengths;               // 42 classes
   init_int nyrs;                   // 22 years (1996-2017)

   init_matrix yieldcpue(1,nyrs,1,3);
   vector years(1,nyrs);
   !!years=column(yieldcpue,1);

   init_int indexMinyrLD;   //2000 ==> index of 5;

   int nyrsLD;             //nyears for the length data
   !!nyrsLD=nyrs-4;   //i.e., 2000-2017;

   init_vector x(1,nlengths); //discrete lengths;
   vector L(1,nlengths);          //discrete lengths;
   !!L=x;

   init_matrix lengthfrq(1,nyrsLD,1,nlengths);

   vector yield(1,nyrs); // in MT
   !!yield=column(yieldcpue,2);
   vector CPUE(1,nyrs);   // in MT/haul
   !!CPUE=column(yieldcpue,3);
```

```
   vector Effort(1,nyrs);
   !!Effort=elem_div(yield,CPUE); // vector(1,23) in hauls

   init_vector musig2_r(1,2);
   number mu_r;
   number sig2_r;
   !!mu_r=musig2_r(1);
   !!sig2_r=musig2_r(2);

   init_matrix abWL(1,nyrs,1,2);
   vector aWL;
   vector bWL;
   !!aWL=column(abWL,1);
   !!bWL=column(abWL,2);

   matrix Wt(1,nyrs,1,nlengths);

   //prior for M ~ lognormal(mu, sigma2)
   init_number mode_M;   //init_number M;   // Natural mortaility as a
salar;
   init_number CV_M;
   number mu_M;
   number sig_M;
   init_number lambda1; // lambda in the objective function
   init_number lambda2; // lambda in the objective function
   init_number logLinf;
   number Linf;
   !!Linf=mfexp(logLinf);

   int r;
   !!r=1; //recruitment is defined as the pop size at age 1;
   int ncohorts;
   !!ncohorts=nyrs;

   ivector SamSize(1,nyrs);     //annual sample size for the length data;

PARAMETER_SECTION
```

```
   init_number  logq(1);                  //log  of  the  catchability  in  F_yr  =
q*Effort_yr;
   init_bounded_vector logRec(1,nyrs,11.5,25.0,2);
   init_bounded_number logkappa(-3.0,-0.80,3);
   init_bounded_number logL50(3.0,4.0,4);            //selectivity
   init_bounded_number gamma(0.05,1.5,5);             //selectivity
   init_bounded_number log_M(-2.996,-0.799,6);    //natural mortality;
   init_bounded_number sigmaL(0.0,1.5,6);         //the uncertainty in the
L_{a+1} equation

   number M;
   number q;
   vector Recruits(1,nyrs);
   number Rbar;
   number kappa;
   number L50;
   vector Sel(1,nlengths);
   number Rho;
   number kkk; //for the cumulative purpose
   matrix f(1,nages,1,nlengths);    //length frequency as pmf
   3darray pp(1,nages,1,nlengths,1,nlengths); //pp(1 To Ages,1 To x,1
To L);
                                                          // x --> (growth) --> L
   vector Mu(1,nlengths); //differ by length class
   vector  SS(1,nages);         //assumed  to  be  constant  over  length
classes
   vector F_yr(1,nyrs); // fishing mortality;
   matrix F_tx(1,nyrs,1,nlengths);    //F_{t,x}; //FM in the VB code;
   matrix Z(1,nyrs,1,nlengths);
   matrix ExpZ(1,nyrs,1,nlengths);    //Exp(-Z);
   3darray NL(1,nages,1,ncohorts,1,nlengths);
   matrix N(1,nyrs,1,nages);    //dimension index is different from the VB
code;
   number SumP;
   vector p(1,nlengths);
   number CNum;
   number CWt;
```

```
   vector TCatch(1,nyrs);
   matrix Catch(1,nyrs,1,nlengths);
   vector Yieldhat(1,nyrs);
   vector Pop(1,nyrs);
   3darray ENx(1,nages,1,ncohorts,1,nlengths);
   vector EN(1,ncohorts);
   vector B(1,ncohorts);
   vector EB(1,ncohorts);
   matrix LF(1,nlengths,1,nyrs);      //expected length-frequency by
   number logmult; //log(multinomial);
   vector elem_obj2(indexMinyrLD,nyrs);   //elements in part 2 of the
objective function;
   number sig2_yield;
   number lognormal;   //log(normal);
   number maxloglike; // for calculation of AIC
   number aic;

   objective_function_value obj;

PRELIMINARY_CALCS_SECTION
   //re-arrangement of data for calculation purposes;
   //Calculate the weight relationship
   for(int w=1;w<=nyrs;w++)
   Wt(w)=aWL(w)*pow(x,bWL(w))/1000; //the division of 1000 is to
convert gram to kg

   //annual sample size for the length data
   int i;
   i=0;
   SamSize=0;
   for(int m=indexMinyrLD;m<=nyrs;m++) {
        i=m-4;
        SamSize(m)=sum(lengthfrq(i));
   };

   //prior for M
   mu_M=log(mode_M*(square(CV_M)+1));
```

```
   sig_M=sqrt( log(square(CV_M)+1) );


PROCEDURE_SECTION
   q=mfexp(logq);
   kappa=mfexp(logkappa);
   L50=mfexp(logL50);
   for(int i=1;i<=nyrs;i++)
         Recruits(i)=mfexp(logRec(i));
   M=mfexp(log_M);

   obj=0.0;

   //Calculate the selectivity for each length class;
   Sel=1.0/(1+mfexp(-1.0*gamma*(x-L50)));

   //Calculate fishing mortality for each length using F = q*Effort*Sel;
   //Sum_over x_F_{yr,x}    =/= F_{yr}; "=/=" "is not equal to"
   F_yr=q*Effort;         //1996 - 2016 (i.e., catch data)
   for(int t=1;t<=nyrs;t++)
      for(int xind=1;xind<=nlengths;xind++)    {
          F_tx(t,xind)=F_yr(t)*Sel(xind); //F_tx; //FM in the VB code;
          Z(t,xind)=M+F_tx(t,xind);
          ExpZ(t,xind)=mfexp(-1.0*Z(t,xind));
      };

   //Calculating the length frequency dsn of the recruits;
   //recruitment is at one year of age;
   //f(1,x) is the same for all cohorts so this can be in the initial calculations;
   Rho=mfexp(-1.0*kappa);
   SS(1)=sig2_r;
   kkk=0.0;
   for(int xind=1;xind<=nlengths;xind++) {
        f(1,xind)=mfexp(-1.0*square(x(xind)-mu_r)/(2.0*SS(1)));
        kkk=kkk+f(1,xind);
   };
```

```
   for(int xind=1;xind<=nlengths;xind++) {
       f(1,xind)=f(1,xind)/kkk;    //normalize;
       Mu(xind)=Linf-(Linf-x(xind))*Rho;
   };

   for(int a=2;a<=nages;a++)    {
         SS(a)=square(sigmaL)*(1.0-pow(Rho,(2.0*a-2.0*r)))/(1.0-
square(Rho))+(pow(Rho,(2.0*a-2.0*r)))*sig2_r;
                 //this SS is for Shrimp; //see Cohen and Fishman (1980)
   };

   for(int a=1;a<=nages;a++) {
       for(int xind=1;xind<=nlengths;xind++)    {
           kkk=0.0;
           for(int Lind=1;Lind<=nlengths;Lind++)    {
               pp(a,Lind,xind)=mfexp(-1.0/(2.0*SS(a))*square(L(Lind)-
Mu(xind)));
               kkk=kkk+pp(a,Lind,xind);
           };
           for(int Lind=1;Lind<=nlengths;Lind++)
               pp(a,Lind,xind)=pp(a,Lind,xind)/kkk;
       };
   };

   //Start of cohort loop
   int m;
   for(int c=1;c<=ncohorts;c++)    {
       int a=1;
       m=c;
       N(m,a)=0.0;
       for(int xind=1;xind<=nlengths;xind++)    {
           NL(a,m,xind)=Recruits(m)*f(a,xind);
           N(m,a)=N(m,a)+NL(a,m,xind);        //Be careful that N has yrs x
ages; //different from the VB code
       };
```

```
        for(int a=2;a<=nages;a++)     { //note that a starts at 2, not 1;
            m=a+c-1;
            if(m>nyrs)
                  SumP=0.0;
            else     {
                      SumP=0.0;
                  for(int Lind=1;Lind<=nlengths;Lind++)    {
                        p(Lind)=0;
                        for(int xind=1;xind<=nlengths;xind++)
                              p(Lind)=p(Lind)+f(a-1,xind)*ExpZ(m-
1,xind)*pp(a,Lind,xind);

                        SumP=SumP+p(Lind);
                }; 
                N(m,a)=0.0;
                for(int Lind=1;Lind<=nlengths;Lind++)    {
                        f(a,Lind)=p(Lind)/SumP;    //normalize;
                        NL(a,m,Lind)=N(m-1,a-1)*p(Lind);
                        N(m,a)=N(m,a)+NL(a,m,Lind);
                };
           };
        };
    };

    //================   //note m starts at indexMinyrLD;
    //Sum through the population matrix to calculate;
    //expected catch distribution;
    //summing for each time by length class for all ages;
    for(int m=indexMinyrLD;m<=nyrs;m++)  {      //note  m  starts  at
indexMinyrLD;
              TCatch(m)=0.0;
              Catch(m)=0.0;
              Yieldhat(m)=0.0;
              Pop(m)=0.0;
              EN(m)=0.0;
              B(m)=0.0;
              EB(m)=0.0;
```

```
            for(int a=1;a<=nages;a++)     {
                for(int xind=1;xind<=nlengths;xind++) {
                        CNum=NL(a,m,xind)*(F_tx(m,xind)/Z(m,xind))*(1-
ExpZ(m,xind));
                        CWt=CNum*Wt(m,xind);   //in kg
                        Catch(m,xind)=Catch(m,xind)+CNum;
                        TCatch(m)=TCatch(m)+CNum;
                        Yieldhat(m)=Yieldhat(m)+CWt;

                        Pop(m)=Pop(m)+NL(a,m,xind);     //population;
                        ENx(a,m,xind)=NL(a,m,xind)*Sel(xind);
//Exploitable population;
                        EN(m)=EN(m)+ENx(a,m,xind);
                        B(m)=B(m)+NL(a,m,xind)*Wt(m,xind);
//Biomass;   //in kg
                        EB(m)=EB(m)+ENx(a,m,xind)*Wt(m,xind);
//expoitable biomass;
                };
        };
  };

  //The expected length-frequency
  for(int  m=indexMinyrLD;m<=nyrs;m++)     {  //note  m  starts  at
indexMinyrLD;
      for(int xind=1;xind<=nlengths;xind++)   {
             LF(xind,m)=(Catch(m,xind)/sum(Catch(m)))*SamSize(m);
        };
  };

  //part 1 of the objective funcion: multinomial
  logmult=0.0;
  for(int i=indexMinyrLD;i<=nyrs;i++) {
      logmult+=gammln(SamSize(i)+1);

      for(int xind=1;xind<=nlengths;xind++) {
          logmult+=-1.0*gammln((lengthfrq(i-4,xind)+1))+lengthfrq((i-
4),xind)*log(Catch(i,xind)/sum(Catch(i)));
```

```
       };
   };
   obj+=lambda1*(-1.0*logmult); //lamda1*(the negative multinomial
likelihood);

   maxloglike=0.0;
   maxloglike+=logmult;

   //part 2 of the objective function: normal
   sig2_yield=0.0;
   for(int m=indexMinyrLD;m<=nyrs;m++)
       elem_obj2(m)=square(log(yield(m))-log(Yieldhat(m)/1000)); //in
MT

   sig2_yield=sum(elem_obj2)/(nyrs-indexMinyrLD+1); //MLE of
sigma2
   obj+=lambda2*(0.5*(nyrs-
indexMinyrLD+1)*log(sig2_yield)+sum(elem_obj2)/(2.0*sig2_yield));
//lamda2*(the negative normal loglikelihood);

   lognormal=0.0;
   lognormal=(-0.5*(nyrs-indexMinyrLD+1)*log(2*M_PI)-0.5*(nyrs-
indexMinyrLD+1)*log(sig2_yield)-sum(elem_obj2)/(2.0*sig2_yield));

   maxloglike+=lognormal;

   //part 3: prior for M
   //prior for M ~lognormal(mu,sig2); // as the negative logarithm;
   obj+=0.5*log(2.0*M_PI)+log(sig_M)+log_M+square(log(M)-
mu_M)/(2.0*sig_M*sig_M);

   aic=-2.0*maxloglike+2.0*(nyrs+6); //check the number of free
parameters.


REPORT_SECTION
   report<<"Yr Recruits Fyr Yield Yieldhat"<<endl;
```

```
   for(int i=1;i<=nyrs;i++)
        report<<years(i)<<" "<<Recruits(i)<<" "<<F_yr(i)<<" "<<yield(i)<<" "<<Yieldhat(i)/1000<<endl;

   report<<"yrs Pop B.in.kg ExploitB"<<endl;
   for(int i=1;i<=nyrs;i++)
        report<<years(i)<<" "<<Pop(i)<<" "<<B(i)<<" "<<EB(i)<<endl;

   report<<"yrs N"<<endl;
   for(int i=1;i<=nyrs;i++)
        report<<years(i)<<" "<<N(i)<<endl;

   report<<"N:"<<N<<endl;
   report<<"lengthfrq: "<<lengthfrq<<endl;
   report<<"Catchhat: "<<Catch<<endl;
   report<<"max.grad: "<<objective_function_value::gmax<<endl;
   report<<"Sample size in the length frequency: "<<SamSize<<endl;

   report<<"x(1,nlengths): "<<x<<endl;
   report<<"Mu(1,nlengths): "<<Mu<<endl;

   report<<"SS(1,nages): "<<SS<<endl;
   report<<"M, sigmaL, and aic: "<<M<<" "<<sigmaL<<" "<<aic<<endl;


RUNTIME_SECTION
   maximum_function_evaluations 100,150,300,10000
   convergence_criteria .01,.0001,1e-7


TOP_OF_MAIN_SECTION
   gradient_structure::set_MAX_NVAR_OFFSET(1000);      //maximum number of depdendent variables of 400 exceeded
   gradient_structure::set_NUM_DEPENDENT_VARIABLES(1000);
   gradient_structure::set_GRADSTACK_BUFFER_SIZE(100000);
   gradient_structure::set_CMPDIF_BUFFER_SIZE(1000000);
   arrmblsize=900000;
```

```
GLOBALS_SECTION
   #include <admodel.h>
   #include <math.h>
   #include <stdio.h>
   #include <stddef.h>
   #include <stdlib.h>
```

DAT file (mackerel_annual.DAT)

```
#nages: number of age classes
6
#number of length classes; #median length of maxFL = 51.5 cm;
42
#number of years for the fishery catch data: nyrs of 1996-2017;
22
#fishery (large-purse-seine) catch
#       Yield(MT)       CPUE(MT/haul)
1996    415003.0
1997    160448.0
1998    172925.0
1999    177540.0
2000    145908.0
2001    203717.0
2002    141751.0
2003    122044.0
2004    184274.0
2005    135596.0
2006    101427.0
2007    143776.0
2008    187240.0
2009    117960.0
2010    94331.0
2011    138729.0
2012    125143.0
2013    102114.0
2014    127452.0
2015    131735.0
2016    133200.0
2017    103870.0
#
#
#indexMinyrLD: index of the minimum year for the length data; //2000
5
#
```

CPUE data are not allowed by the National Institute of Fisheries Sciences (NIFS) to be revealed.
If someone wants those data, get the permission of NIFS.

```
#midpoints (cm) of length classes;
10.5    11.5    12.5    13.5    14.5    15.5    16.5    17.5    18.5
        19.5    20.5    21.5    22.5    23.5    24.5    25.5    26.5
        27.5    28.5    29.5    30.5    31.5    32.5    33.5    34.5
        35.5    36.5    37.5    38.5    39.5    40.5    41.5    42.5
        43.5    44.5    45.5    46.5    47.5    48.5    49.5    50.5
        51.5
#length frequency data (18 x 42);
#18: year 2000 - 2017
#42: midpoints (cm) of length classes: 10.5, 11.5, 12.5, ... , 51.5
```

Length data are not allowed by NIFS to be revealed.
If someone wants those data, get the permission of NIFS.

Length data are not allowed by NIFS to be revealed.
If someone wants those data, get the permission of NIFS.

Length data are not allowed by NIFS to be revealed.
If someone wants those data, get the permission of NIFS.

```
#
#Prior info
#Korean mackerel
#the mean (cm) and the variance (cm^2) of the lengths at the recruit
stage (age 1);
#18.00 was calculated from the Linf below, Linf=exp(3.95);
18.00   2.10
#
#parameters in W =a*(L^b); #W in gram; # L in cm;
#a and b
```

```
0.003   3.425 #1996 #1
0.003   3.425 #1997 #2
0.003   3.425 #1998 #3
0.003   3.425 #1999 #4
0.003   3.425 #2000 #5 $1$
0.003   3.425 #2001 #6
0.003   3.425 #2002 #7
0.003   3.425 #2003 #8
0.003   3.425 #2004 #9
0.0036 3.3813 #2005 #10
0.0028 3.4388 #2006 #11
0.0025 3.4768 #2007 #12
0.0023 3.5062 #2008 #13
0.0018 3.5805 #2009 #14
0.0038 3.373 #2010 #15
0.0027 3.4681 #2011 #16
0.0024 3.488 #2012 #17
0.002   3.5491 #2013 #18
0.0023 3.4858 #2014 #19
0.0026 3.4521 #2015 #20
0.0029 3.4222 #2016 #21
0.003   3.4081 #2017 #22

#
#M: Natural mortality (instantaneous rate) as a prior dsn;
#mode of M
2.0
#CV of M
2.0
#
#lambda1 for the length data in the objective function
0.05
#lambda2 for the yield data in the objective function;
10.0
#
#logLinf
```

3.95
#

PIN file

```
#logq
-9.5
#logRec(1,nyrs)
20      20      20      20      20      20      20      20      20      20
        20      20      20      20      20      20      20      20      20
        20      20      20
#logkappa
-1.21
#logL50 in the selectivity
3.38
#gamma in the selectivity
0.6
#log_M
-1.386
#sigmaL
0.66
```