Thesis for the Degree of Master of Science

# A state-space length-based assessment model for the Korea chub mackerel (*Scomber japonicus*) stock

by

DoYul Kim

Department of Marine Biology

The Graduate School

Pukyong National University

August 2021

# A state-space length-based assessment model for the Korea chub mackerel (*Scomber japonicus*) stock

# [한국 고등어(*Scomber japonicus*) 자원평가를 위한 상태공간 체장기반 모델]

Advisor: Prof. Saang-Yoon Hyun

by
DoYul Kim

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 2021

# A state-space length-based assessment model for the Korea chub mackerel (*Scomber japonicus*) stock

A thesis

by

DoYul Kim

Approved by:

(Chair) Dr. Hyun-Woo Kim

Seoyongil

(Committee) Dr. Young Il Seo

(Committee) Dr. Saang-Yoon Hyun

August 27, 2021

# Contents

# List of figures

# List of tables

A state-space length-based assessment model for the Korea chub mackerel (*Scomber japonicus*) stock

김도윤

부경대학교 대학원 해양생물학과

## 요 약

본 연구를 통해, 어류의 체장 자료를 이용하는 자원평가 모델인 상태공간 체장기반 모델을 개발하였다. 어류의 연령 자료가 이용가능한 경우에는 연령구조 모델을 이용하여 자원을 평가하는 것이 일반적이다. 하지만, 어류의 연령 자료를 수집하는 것은 체장 자료를 수집하는 것에 비해 시간 및 비용적 측면에서 훨씬 더 부담이 크기 때문에, 대부분의 수산자원에 대해 이용가능한 연령 자료가 부재한 경우가 일반적이다. 이러한 이유로, 우리나라에서는 어류의 연령 자료에 비해 체장 자료가 잘 수집이 되어 있는 실정이며, 이는 우리나라 자료 수집 현황에 적합한 체장기반 모델 개발 연구에 대한 직접적인 동기가 되었다. 본 연구의 또 다른 동기는, 자료를 수집하는 과정에서 일어날 수 있는 관측 오차(observation error)와 몇몇 시간 변이성을 갖는 모수에 의해 일어날 수 있는 과정 오차(process error)를 명확하게 구분하는 자원평가 모델을 개발하고자 하는 것이었다. 이를 위해, 시간 변이성을 갖는 모수를 random effects (state variable)로 취급하는 상태기반(state-space) 체계를 도입하였으며, 특히 해당 연구에서는 연도별 연령별 마릿수를 random effects 로 간주하였다. 해당 모델의 주요 특징으로는 개체군 내 연급군의 체장 빈도 분포를 연령 및 연도별로 동일하게 가정하는 것이 아니라, 가입연령의 체장 빈도 분포로부터 사망과 확률론적 von Bertalanffy 모델을 이용한 체장 계급별 체성장을 고려하여 가입연령 이후의 체장 빈도 분포를 직접적으로 유도한다는 것이다. 또한, 일반적인 자원평가 모델에서 추정이 잘 되지 않아 입력값으로써 선언되는 자연 사망률 값을 자유 모수로써 추정한다는 특징이 있다. 본 연구를 통해 모델을 개발하면서, 모델 내 몇몇 가정에 따른 모델 성능을 비교, 분석하였다. 이에, 모델 내 산란-가입 관계식 유무 및 과정 오차 고려 여부에 따라 총 6 개의 시나리오를 구성하였으며, 자료에 대한 적합도, retrospective 패턴, 모의 실험(simulation study)의 결과를 이용하여 모델을 검증하였다. 본 연구에서 개발한 자원평가 모델은 TMB 언어로

직접 코딩하였으며, 모수의 점추정치 뿐만 아니라 추정치의 표준 오차를 제시하였다. 자원평가를 위해 모델에 적용된 자료는 우리나라 연근해에서 대형선망 어업에 의해 주로 어획되는 고등어(*Scomber japonicus*)의 2000 년부터 2019 년 까지의 연도별 어획 체장 조성 자료, 단위노력당 어획량 자료(CPUE), 총 어획량 자료이다. 종합적인 결과로써, 산란-가입 관계식이 고려되지 않으면서, 모든 연령별 연도별 마릿수에 대한 과정오차를 고려한 모델(시나리오)이 가장 우수하였다. 해당 시나리오에서 자연 사망률 값은 약 0.1/년으로 추정되었으며, 추정된 고등어의 자원량과 산란 어미 생체량은 각각 약 5.0 x $10^5$ ~ 10.0 x $10^5$ MT, 1.0 x $10^5$ ~ 2.2 x $10^5$ MT 의 범위에서 나타났다.

# 1. Introduction

For most fisheries, gauging the age information of fish via fishery-independent surveys used in age-structured assessment models is considerably expensive and difficult to obtain (Rudd and Thorson 2018). By contrast, measuring and collecting body size information such as the lengths of fish caught by a fishing fleet is easier and more economical. In this context, a length-based assessment model that is more informative and has a higher resolution than a surplus production model could be a viable option to assess the exploited fish stock in most fisheries. With this in mind, I develop a length-based assessment model.

This model, which is extended from Cohen and Fishman (1980), Deriso and Parma (1988), and Quinn et al. (1998), has some differences from other prominent length-based assessment models in CASAL, MULTIFAN-CL, SCALE, Stock Synthesis, and others. Unlike the model in CASAL, a software package developed by NIWA for modeling the population dynamics of marine species (Bull et al. 2012), the proposed model utilizes the structure of imaginary age class, which enables it to segregate cohorts based on the length distribution of the population. MULTIFAN-CL, which uses identical assumptions in MULTIFAN (Fournier et al. 1998), predicts the observed size-composition data as the sum of normal distributions (Punt et al. 2013). By contrast, the proposed model assumes that each cohort of the population follows a length distribution. Furthermore, the proposed model assumes that length-based proportions of numbers follow a normal distribution with a given mean and variance only at the age of recruitment; the length distributions at the age after recruitment are calculated by considering the process of mortality and growth, which differs from the assumption made in SCALE. In SCALE, a statistical Catch-at-Length model included in the NOAA Fish and Fisheries Toolbox, all cohorts are assumed to follow given normal distributions for length at each age considered. As the length-based assessment module in Stock Synthesis, which has been applied in a wide variety of fish assessments globally, is based on an age-structured model, some functions associated with fish length, such as gear selectivity and fish body growth, are converted to the function for age (Methot and Wetzel 2013; Methot et al. 2020). In the proposed model, most of the functions are customized for fish lengths and not for age.

Another difference of the proposed model from other prominent length-based assessment models is that it explicitly estimates the natural mortality rate. As estimating the natural mortality rate in stock assessment models is difficult (Vetter 1988; Quinn and Deriso 1999), it is commonly assumed to be known and constant (Hilborn and Walters 2013). Although natural mortality in fish likely varies among ages (sizes) and years (Deroba and Schueller 2013), it is treated as a constant across fish of various lengths and years, reflecting that a constant natural mortality rate performed better than the allometric relationship of natural mortality to mass (Miller and Hyun 2018).

The other main features in this model include considering observation and process errors, leading to a state-space framework, where I treat some time-varying parameters as random effects (a.k.a. state variables). Traditional assessment models do not separate variance of the observations from those attributable to time-varying processes (Miller and Hyun 2018). However, the utility of formal state-space models wherein process errors in time-varying parameters are modeled and estimated is statistically significant (Miller and Hyun 2018). Accounting for random variation and observation error arising from the process of sampling from the population helps to better identify the true state of the fish population (Rudd and Thorson 2018). In this context, the proposed model treats the annual abundance at age as a random effect and the probability distribution related to fishing mortality as a penalty to the likelihood function.

I also explored the effects of the presence or absence of a stock–recruitment relationship and process errors in annual abundance at age in the model framework. The relationship between the biomass of reproductively mature individuals (spawning stock biomass) and the resulting offspring added to the population (recruitment), i.e. the stock–recruitment relationship, is a fundamental challenge in population biology (Mangel et al. 2010). In some cases, it is argued that recruitment is mainly determined by the environment and is independent of spawning stock biomass. The steepness, which was popularized by Mace and Doonan (1988), is set to 1 (Mangel et al. 2010). I try to determine the impact of stock–recruitment relationship on the result of estimation in this model. I also try to consider a length-based model without random effects as well as a state-space length-based model and compare each model's performance to determine whether accounting for variation arising from natural processes or measurement processes separately (Rudd and Thorson 2018) improves model performance.

The purpose of this study is to develop and demonstrate this state-space length-based assessment model. This model requires yields, catch-per-unit-effort (CPUE), and length composition data, all of which are readily available. For demonstrating this model, I used actual data on Korean chub mackerel (*Scomber japonicus*) stock. By conducting a simulation study and comparing the goodness-of-fit for data and Mohn's (1999) $\rho$, the measure of retrospective pattern, I compared the relative performance of the models chosen for comparison in this study.

# 2. Materials and Methods

## 2.1. A state-space length-based model

### 2.1.1. Fish length as discrete variable

I assumed that a length class *x* of individuals at recruitment, *r* in the beginning of a year follows a discrete normal (Gaussian) random variable, $X_r$ : i.e., $X_r \sim Normal_D(\mu_r,\ \sigma_r^2)$ where the subscript *D* indicates the random variable is 'discrete'.

Thus, the probability mass function (PMF) of $X_r$ can be written

$$f_r(x) = \exp\left(-\frac{1}{2\sigma_r^2}(x-\mu_r)^2\right) \Bigg/ \sum_x \exp\left(-\frac{1}{2\sigma_r^2}(x-\mu_r)^2\right) \tag{1}$$

where constants with respect to *x* are ignored. See Table 1 for notations. Then, the abundance of recruitment at each length class at the beginning of year *t* is

$$N_{t,r}(x) = N_{t,r} \cdot f_r(x). \tag{2}$$

### 2.1.2. Mortality

A distribution of population's length frequency and abundances at age after recruitment change over time with the processes of growth and mortality. I followed Quinn et al. (1998)'s assumption that the mortality process occurs first and then the growth process does. The total mortality rate is the sum of natural and fishing mortality rate:

$$Z_t(x) = M + F_t(x). \tag{3}$$

I assumed the natural mortality rate to be constant across length classes and years but have fishing mortality rate to differ by each length class and year. Fishing mortality rate of length class *x* in year *t* is separable into the selectivity for length class and annual fully-selected fishing mortality in year *t*:

$$F_t(x) = S(x) \cdot F_t. \tag{4}$$

I assumed a logistic selectivity at length class for the fishing fleet which is considered to be large purse-sein in the proposed model:

$$S(x) = \frac{1}{1 + e^{-\gamma(x - L_{50\%})}}. \tag{5}$$

I treated annual fully-selected fishing mortality as a random walk with the known variance, $\sigma^2_{\log F}$ :

$$\log(F_1) \sim Normal(\log(F_{init}),\ \sigma^2_{\log F}) \tag{6}$$

and

$$\log(F_{t+1}) \sim Normal(\log(F_t),\ \sigma^2_{\log F}) \tag{7}$$

where $t = 1, \ldots, T - 1$.

The relative length distribution at age $a$ at the end of year $t$ where the process of morality took place during year $t$ is derived from the PMF of the length class at age $a$ at the beginning of year $t$:

$$p_{t,a,Z}(x) = f_{t,a}(x) \cdot e^{-Z_t(x)}. \tag{8}$$

### 2.1.3. Body growth

To account for body growth after fish experienced mortality, I used Quinn et al. (1998)'s assumption that an individual of length class $x$ will growth to length class $l$ in one time step according to a stochastic growth model (Quinn et al. 1998). The deterministic von Bertalanffy growth model is

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

This equation can be transformed into the relationship of $L_{a+1}$ against $L_a$ with the stochastic error term:

$$L_{a+1} = L_\infty(1-\rho_G) + \rho_G L_a + \varepsilon_G \tag{9}$$

where $\varepsilon_G \sim Normal(0,\ \sigma_G^2)$ and the Brody coefficient $\rho_G$ is $e^{-\kappa}$. By modifying this equation, the expected length and variance at age $a$ + 1 can be expressed by equation for length class $x$ and age $a$, respectively, from Cohen and Fishman (1980):

$$\mu(x) = L_\infty(1-\rho_G) + \rho_G x \tag{10}$$

and

$$\sigma_{a+1}^2 = \sigma_G^2 \frac{1-\rho_G{}^{2(a+1-r)}}{1-\rho_G{}^2} + \rho_G{}^{2(a+1-r)}\sigma_r^2. \tag{11}$$

Thus, length $l$ of each individual at age $a$ + 1 is considered to be a discrete variable and has a distribution which is discrete normal with the mean, $\mu_G(x)$ and the variance, $\sigma_{a+1}^2$ : i.e., $L \mid x \sim Normal_D(\mu_G(x),\ \sigma_{a+1}^2)$. The PMF of length $L$ at age $a$ + 1 given length class $x$ can be written

$$f_{a+1,G}(l \mid x) = \exp\left(-\frac{1}{2\sigma_{a+1}^2}(l-\mu_G(x))^2\right) \Bigg/ \sum_l \exp\left(-\frac{1}{2\sigma_{a+1}^2}(l-\mu_G(x))^2\right). \tag{12}$$

### 2.1.4. Combining the process of mortality and body growth

The relative length distribution at age $a$ + 1 at the beginning of year $t$ + 1 is

$$p_{t+1,a+1}(l) = \sum_x p_{t,a,Z}(x) \cdot f_{a+1,G}(l \mid x). \tag{13}$$

The number of individuals of length class $l$ at age $a$ + 1 at the beginning of year $t$ + 1 is calculated using the number of individuals at age $a$ at the beginning of year $t$ and eq. (13):

$$N_{t+1,a+1}(l) = N_{t,a} \cdot p_{t+1,a+1}(l). \tag{14}$$

Then, the abundance at age $a$ + 1 at the beginning of year $t$ + 1 is

$$N_{t+1,a+1} = \sum_{l} N_{t+1,a+1}(l). \tag{15}$$

The PMF of a length class can be written

$$f_{t+1,a+1}(l) = \frac{p_{t+1,a+1}(l)}{\sum p_{t+1,a+1}(l)} \tag{16}$$

or

$$f_{t+1,a+1}(l) = \frac{N_{t+1,a+1}(l)}{N_{t+1,a+1}}. \tag{17}$$

### 2.1.5. Abundances at the first year

In the propsed model, the logarithm of abundances at all ages at the first year ($\log(N_{t=1,a})$) is assumed to be distributed normally with mean $\log(\mu_{N_{t=1}})$ and the known variance $\sigma^2_{\log N_{t=1}}$:

$$\log(N_{t=1,a}) \sim Normal(\log(\mu_{N_{t=1}}),\ \sigma^2_{\log N_{t=1}}) \tag{18}$$

where $a$ = r, …, A. The known variance provides the constraining penalty for the estimates of deviations of abundance at age at the first year (Methot et al.

2020).

As I assumed the distribution of length frequency at recruited age, I could derive length frequency distribution at age after recruitment for each cohort by considering both the mortality and the body growth process. However, as abundances at age at the first year are not from certain one common cohort, I needed to define the length frequency distribution by age. I assumed the equilibrium state for total mortality rate to resolve it. I considered the mean value, $\overline{Z}$ of the total mortality rates in all years by length class. This can be written as

$$p_{t=1,a,Z}(x) = f_{t=1,a}(x) \cdot e^{-\overline{Z}(x)}. \tag{19}$$

Then, the process of growth occurs:

$$p_{t=1,a+1}(l) = \sum_{x} p_{t=1,a,Z}(x) \cdot f_{a+1,G}(l \mid x). \tag{20}$$

The PMF of length class at age $a + 1$ at the first year can be written:

$$f_{t=1,a+1}(l) = \frac{p_{t=1,a+1}(l)}{\sum p_{t=1,a+1}(l)}. \tag{21}$$

By repeating this calculation until the maximum age, each PMF of length classes at all ages can be obtained. The number of individuals of length class $x$ at age $a$ at the beginning of the first year is calculated

$$N_{t=1,a}(l) = N_{t=1,a} \cdot f_{t=1,a}(l) \tag{22}$$

where $a = 2, \ldots, \text{A+}$.

### 2.1.6. Spawning stock biomass

I estimated the spawning stock biomass by estimating the number of spawners and considering maturation rates, the ratio of female fish to both sex fish, and body weights. I applied the length-maturation relationship of a logistic form (eq. (23)) and the ratio of the female fish, 0.5 following Kim et al. (2020):

$$Mat(x) = \frac{1}{1 + e^{(b_0 - b_1 \cdot x)}}. \tag{23}$$

Then, the number of spawners of length class $x$ at age $a$ at the beginning of year $t$ is expressed as

$$Spawners_{t,a}(x) = N_{t,a}(x) \cdot Mat(x) \cdot ratio_{female}. \tag{24}$$

Summing spawners from eq. (24) over age and length classes, I get the number of spawners at the beginning of year $t$:

$$Spawners_t = \sum_a \sum_x Spawners_{t,a}(x) \tag{25}$$

Also, multiplying the number of spawners of length class $x$ at age class $a$ at the beginning of year $t$ in the right side of eq. (25) by the mean weight at length class $(W(x))$, spawning stock biomass at the beginning of year $t$ can be acquired

$$SSB_t = \sum_a \sum_x Spawners_{t,a}(x) \cdot W(x) \tag{26}$$

where the form of allometric length-weight relationship $W(x)$ is as follows:

$$W(x) = \alpha \cdot x^{\beta}. \tag{27}$$

### 2.1.7. Recruitment

I considered two assumptions for estimating recruitment in all remaining years except the first year. The first assumption is that Beverton–Holt stock–recruitment model in which recruitment is related to spawning stock biomass is applied to estimate recruitments. The second is the assumption that recruitment is related independently spawning stock biomass. Two alternative assumptions are as follows:

$$N_{t+1,r} = \begin{cases} \dfrac{SSB_t}{\alpha_{SR} + \beta_{SR} \cdot SSB_t} \\ \\ \mu_R \end{cases} \tag{28}$$

where $t = 2, \ldots, T$. In the case that the second assumption is utilized in this model, logarithm of recruitment at the beginning of year $t$ $(\log(N_{t,r}))$ is assumed to be distributed normally with mean $\log(\mu_R)$ and fixed variance $\sigma^2_{\log R}$:

$$\log(N_{t,r}) \sim Normal(\log(\mu_R),\ \sigma^2_{\log R}) \tag{29}$$

where $t = 2, \ldots, T$. The fixed variance provides the constraining penalty for the estimates of recruitment deviations and it is not affected by data (Methot et al. 2020).

### 2.1.8. Abundance at length class x at each age

The following equation describes how the abundances at length class $x$ at each age are calculated. The eq. (22) presents those in the first year:

$$N_{t=1,a}(l) = N_{t=1,a} \cdot f_{t=1,a}(l). \tag{22}$$

And the eq. (30) presents those for year > 1:

$$N_{t,a}(x) = \begin{cases} N_{t,r} \cdot f_r(x) & \text{for } a = r \\ N_{t-1,a-1} \cdot p_{t,a}(x) & \text{for } 1 < a < A \\ N_{t-1,a-1} \cdot p_{t,a}(x) + N_{t-1,a} \cdot p_{t,A}(x) & \text{for } a = A \end{cases} \tag{30}$$

### 2.1.9. Random effects

I considered abundance at age and years as a random effects. Logarithm of abundance for age and year > 1 are normally distributed conditional on the calculated deterministic numbers at age (Miller and Hyun 2018):

$$\log(N^{*}_{t,a}) \sim Normal(\log(N_{t,a}),\ \sigma^2_{\log N}). \quad \text{for } t > 1 \tag{31}$$

In this model, I considered process error in abundance at age separated by recruitment and remaining abundance at older ages. Therefore, I can include process errors in annual deviations around Beverton–Holt model stock–recruitment curve or logarithm of recruitment $\mu$ and in interannual transitions of abundance at older ages. Although I considered random effects as two parts, I assume there is only one common variance for the abundance at age and estimate this parameter. The reason why I treat it is estimating a variance in stock assessment is difficult, so I tried to lessen the number of free parameters difficult to be estimated. Now, I can take into account several scenarios, depending on which assumption for estimating recruitment to select, and whether to consider both or either process error or not. How to set up a series of scenarios will be explained further later.

### 2.1.10. Measurement error

I assumed the logarithm of fishery yields have observation errors, a normal (Gaussian) random variable with the known variance $\sigma^2_{\log Y}$ :

$$\log(Y_t) \sim Normal(\log(Y_t),\ \sigma^2_{\log Y}). \tag{32}$$

The predicted catch at length class *x* in year *t* by the fishing fleet is

$$C_t(x) = \sum_a \frac{F_t(x)}{Z_t(x)} \cdot (1 - e^{-Z_t(x)}) \cdot N_{t,a}(x). \tag{33}$$

The predicted aggregate yield by the fishing fleet in year *t* is

$$Y_t = \sum_x C_t(x) \cdot W(x). \tag{34}$$

The observed logarithm of aggregate relative biomass indices for fishing fleet which are equivalent to catch-per-unit-effort (CPUE) in this model are assumed to follow a normal distribution with the known variance $\sigma^2_{\log I}$:

$$\log(I_t) \sim Normal(\log(\hat{I}_t),\ \sigma^2_{\log I}). \tag{35}$$

The predicted relative biomass index at the beginning of year *t* is

$$\hat{I}_t = q \cdot B_t. \tag{36}$$

The biomass at the beginning of year *t* is

$$B_t = \sum_a \sum_x N_{t,a}(x) \cdot W(x). \tag{37}$$

The catchability for the relative biomass index is assumed to be constant regardless of year and estimated on a logit-scale to avoid boundary problems during estimation (Miller and Hyun 2018):

$$q = l_q + \frac{u_q - l_q}{1 + e^{-\mathrm{logit}(q)}} \tag{38}$$

where $l_q$ and $u_q$ are lower and upper bounds of parameter $q$ and $\mathrm{logit}(q)$ is the actual free parameter estimated in the model (Miller and Hyun 2018).

The length frequencies by the fishing fleet catch are assumed to follow a multinomial distribution:

$$\underset{\sim}{m}_t \sim Multinomial(n_t, \underset{\sim}{\hat{p}}_t) \tag{39}$$

where $n_t$ is the sample size of the length frequencies by the fishing fleet catch in year *t*, and $\underset{\sim}{\hat{p}}_t$ is the vector of predicted catch proportions by length class in year *t*:

$$\underset{\sim}{\hat{p}}_t = \frac{C_t(x)}{\sum_x C_t(x)} \tag{40}$$

## 2.2. Model scenarios

I considered six model scenarios and summarize them in (Table 2). The criteria determining the model scenario are as follows: (i) model for estimating recruitment after the first year, and (ii) how the population abundances at age after the first year were treated. In the respective criterion, there are two and three options to choose from, respectively.

In the first criterion, the options are (see eq. (28)): (i-1) Beverton–Holt stock–recruitment model, (ii-2) mean-deviations which includes assumption that recruitment is related independently spawning stock biomass. In the second criterion, the options I can select are as follows: (ii-1) no random effects, (ii-2) treating only annual recruitment after the first year as random effects, and (ii-3) treating all abundances at age after the first year as random effects. Thus, a total of 6 scenarios can be constructed by combining (i) and (ii) (Table 2). The first three (M1-M3) are assumed to have Beverton–Holt stock–recruitment model, while the later three (M4-M6) are assumed that recruitment is related independently spawning stock biomass. The difference among model scenarios M1-M3 and among M4-M6 are in the assumptions for process errors. Model scenarios M1 and M4 are assumed to have no random effects. Model scenarios M2 and M5 are assumed to treat only recruitment after the first year as random effects, and model scenarios M3 and M6 are assumed to consider abundance at age after the first year as random effects. Therefore, model scenarios M1 and M4 are referred to as a length-based model and model scenarios M2, M3, M5, and M6 are referred to as a state-space length-based model.

## 2.3. Objective function

Free parameters in the proposed model are the mean value of abundance and all abundances at age at the first year $(\mu_{N_{t=1}}, N_{t=1,a})$, catchability for relative biomass index $(q)$, all selectivity parameters $(\gamma, L_{50\%})$, some growth parameters $(\kappa, \sigma_G)$, initial fully-selected fishing mortality rate and annual fully-selected fishing mortality rate at all years $(F_{init}, F_t)$, and constant natural mortality rate across length classes and years $(M)$. In model scenarios M1, M2, and M3, the two parameters in Beverton–Holt stock–recruitment model $(\alpha_{SR}, \beta_{SR})$ are commonly added as free parameter. However, in model scenarios M4, M5, and M6, mean value of recruitment $(\mu_R)$ is commonly added as free parameter. Further, annual recruits after the first year $(N_{t,r})$ are treated as free parameter in model scenario M4, whereas these are treated as random effects in model scenarios M5, and M6. State-space model scenarios M2, M3, M5, and M6 estimate additional free parameter for variance for random effects $(\sigma^2_{\log N})$. Therefore, it is different that what kind of and the number of free parameters estimated for each model scenario fitted to the Korea chub mackerel data (Table 3).

I estimated free parameters using the maximum likelihood approach (MLE), and the likelihood components are defined according to eq. (6), (7), (18), (29), (31), (32), (35), and (39). The likelihood component for eq. (6), and (7) for logarithm of fully-selected fishing mortality rate is as follows:

$$L_1 = Normal\left(\log(F_1) \mid \log(F_{init}), \sigma_{\log F}{}^2\right) \cdot \prod_{t=1}^{T-1} Normal\left(\log(F_{t+1}) \mid \log(F_t), \sigma_{\log F}{}^2\right).$$

The notation, *Normal(X|Y, Z)* indicates the normal pdf of X whose population mean and variance are Y and Z , respectively.

The likelihood component for eq. (18) for logarithm of abundances at all ages at the first year is as follows:

$$L_2 = \prod_{a=r}^{A} Normal\left(\log(N_{t=1,a}) \mid \log(\mu_{N_{t=1}}), \sigma^2_{\log N_{t=1}}\right).$$

The likelihood component for eq. (29) for logarithm of recruitment is as follows:

$$L_3 = \prod_{t=2}^{T} Normal\left(\log(N_{t,r}) \mid \log(\mu_R), \sigma^2_{\log R}\right).$$

The likelihood component for eq. (31) for logarithm of abundance at age can be divided into two parts depending on model scenario related to random effects.

$$L_4^{(1)} = \prod_{t=2}^{T} Normal\left(\log(N^*_{t,r}) \mid \log(N_{t,r}), \sigma^2_{\log N}\right)$$

where $L_4^{(1)}$ includes assumption corresponding (ii-2) option in model scenario.

$$L_4^{(2)} = \prod_{t=2}^{T} \prod_{a=r}^{A} Normal\left(\log(N^*_{t,a}) \mid \log(N_{t,a}), \sigma^2_{\log N}\right)$$

where $L_4^{(2)}$ includes assumption corresponding (ii-3) option in model scenario. The likelihood component for eq. (32) for observed logarithm of aggregate yields by the fishing fleet is as follows:

$$L_5 = \prod_{t=1}^{T} Normal\left(\log(Y_t) \,|\, \log(Y_t),\, \sigma_{\log Y}^2\right).$$

The likelihood component for eq. (35) for observed logarithm of aggregate relative biomass indices for fishing fleet is as follows:

$$L_6 = \prod_{t=1}^{T} Normal\left(\log(I_t) \,|\, \log(\hat{I}_t),\, \sigma_{\log I}^2\right).$$

The likelihood component for eq. (39) for the multinomial distribution for the length composition data for fishing fleet is as follows:

$$L_7 = \prod_{t=1}^{T}\left(n_t!\prod_{x}\frac{\hat{p}_t^{\,o_{t,x}}}{o_{t,x}!}\right).$$

The likelihood components $L_1$, $L_2$, $L_5$, $L_6$, and $L_7$ are commonly included in all of model scenarios I considered, whereas the likelihood components $L_3$, $L_4^{(1)}$, and $L_4^{(2)}$ are only involved in corresponding model scenarios (Table 2). Therefore, according to each model scenario, the joint likelihood function is the product of the corresponding likelihood components in the model scenario. I implemented the state-space length-based model in R package, TMB

(Kristensen et al. 2015; Developer Core Team 2019). TMB uses Laplace approximation to integrate joint log-likelihood over the random effects, then, the marginal log-likelihood is computed and obtained. The “nlminb” function for optimization is used to minimize the negative of the marginal log-likelihood function (objective function). I showed my TMB code (cpp file) in the appendix.

## 2.4. Application to Korea chub mackerel (*Scomber japonicus*) stock

I illustrated this model for the Korea chub mackerel stock. Chub mackerel are one of the most important stocks for fishery in Korea. Chub mackerel are mainly caught by the large purse-sein fishery, and the fishery catch accounts for about 92% of the total catch of Chub mackerel in Korea. For this reason, I assumed that catch of Chub mackerel by large purse-sein represents the whole of catch of that. This model requires three kind of data that are fishing fleet yields, relative biomass indices, and length frequencies by the fishing fleet catch. I used annual total yields from 2000–2019 provided by the Korean Statistical Information Service (KOSIS), and annual CPUE and length frequency data for the fishing fleet catch from large purse-sein from 2000–2019 provided by the National Institute of Fisheries Science (NIFS).

To apply this model to Korea chub mackerel, I made a few assumptions. I defined the age of recruitment as age 1 and the terminal age as age 6 for Korea chub mackerel. Also, length bin width is defined as 1 cm and the number of classes is 43 whose range is from 10.5–52.5 cm. In the length frequency data by the fishing fleet catch assumed to follow a multinomial distribution, the numbers of annual total sample size are assumed to be same and set as 200 individuals. For some parameters that are difficult to be estimated or have some information from previous studies or estimated value using related data outside this model, those are declared as an input (Table 4).

In terms of weights for likelihood, the sample size of the length frequency data for the fishing fleet catch is assumed to be same throughout all years to allocate the same weight to every year I consider. Other than that, I did not have any information for some weights for likelihoods such as fixed variances of observation error and penalized likelihood $\sigma^2_{\log Y}$, $\sigma^2_{\log I}$, $\sigma^2_{\log F}$, $\sigma^2_{\log R}$, and $\sigma^2_{\log N_{t=1}}$. For that reason, I performed a sensitivity analysis for these weights. I expressed a variance of logarithm of random variable X as a coefficient of variation of X, using following equation $\sigma_{\log X} = \sqrt{\log((CV_X)^2 + 1)}$ . In sensitivity analysis for weights for likelihood components, I first assumed a set of plausible candidates for each value of weights and then estimated parameters repeatedly conditioned on those. In the next step, I compared the results of sensitivity analysis using several criteria which include whether the numerical optimization was successful, maximum gradient value, etc. Consequently, the best set of weights was determined based on the results and used in estimation of parameters.

## 2.5. Model performance

To compare in performance among each model scenarios, I used measure of a retrospective pattern and did a simulation study. For measure of a retrospective pattern, I calculated Monh's $\rho$ and checked the consistency of model estimates. I peeled off total five years from the terminal year in sequence and simultaneously estimated parameters using data peeled off. Then, I compared the estimates which are from data peeled off based on the estimates which are from all of data.

$$\rho(\theta) = \frac{1}{m}\sum_{y=1}^{m}\frac{\theta_{T-y,T-y} - \theta_{T-y,T}}{\theta_{T-y,T}}$$

$$\rho_{t,y}(\theta) = \frac{\theta_{t,T-y} - \theta_{t,T}}{\theta_{t,T}}$$

I defined $T = 2019$, and $y = 1, \ldots, 5$ and calculated Mohn's $\rho$ for estimates of annual fully-selected fishing mortality rate and annual spawning stock biomass.

For simulation (Figure 1), I generated 1000 sets of pseudo data which consisted of annual yields by the fishing fleet, annual catch-per-unit-effort for relative biomass indices for fishing fleet, and annual length frequencies by the fishing fleet catch. I commonly considered observation errors in yield, CPUE, length frequency data by the fishing fleet in all model scenarios, whereas process errors in recruitment or abundance at age after the first year are included only in a state-space model scenario (Table 5). The estimates of parameters by each model scenario from the fits to the Korea chub mackerel data are used as the true parameter values for generating pseudo data. I estimated parameters using sets of pseudo data and checked the number of convergences of each model scenario fitted to corresponding sets of pseudo data.

I used four criteria of convergence: (i) convergence flag that the nlminb function returns, (ii) whether to generate Hessian matrix, (iii) whether estimates are within the bounds I set, and (iv) the maximum gradient value. I use relative difference to compare true parameter values I set and estimates of parameters from the fits to the set of pseudo data. The relative difference of a parameter estimate $\theta_i$ from the true value $\theta_i$ for pseudo data set *i* is as follows (Miller and Hyun 2018):

$$\mathrm{RD}_i(\theta) = \frac{\theta_i - \theta_i}{\theta_i}$$

**Table 1.** Description of notations.

| Notation | Description |
|---|---|
| **Indices** | |
| $a$ | Imaginary age class |
| A+ | Terminal imaginary age class and its above classes |
| $r$ | Recruitment age |
| $t$ | Year |
| $T$ | Terminal year in all year considered |
| $x$ | Length class before the process of growth occurs |
| $l$ | Length class after the process of growth occurs |
| **Data** | |
| $Y_t$ | Observed aggregate yield by the fishing fleet in year *t* |
| $I_t$ | Observed aggregate relative biomass indices for fishing fleet in year *t* |
| $\underset{\sim}{o}_t$ | Vector of observed catch composition frequencies by length class in year *t* |
| $n_t$ | Sample size of the length frequencies by the fishing fleet catch in year *t* |
| **Parameter(s)** | |
| $\mu_r$ | Mean of length distribution of recruitment |
| $\sigma_r^2$ | Variance of length distribution of recruitment |

| | |
|---|---|
| $M$ | Length- and time-invariant natural mortality rate |
| $F_{init}$ | Initial fully-selected fishing mortality rate |
| $F_t$ | Fully-selected fishing mortality rate in year $t$ |
| $\sigma^2_{\log F}$ | Variance of random walk for fully-selected fishing mortality rate |
| $\gamma,\ L_{50\%}$ | Logistic selectivity parameters |
| $L_\infty,\ \kappa,\ a_0$ | Parameters in von Bertalanffy growth model |
| $\sigma^2_G$ | Variance of stochastic error term in von Bertalanffy growth model |
| $\mu_{N_{t=1}}$ | Mean abundance at the first year |
| $N_{t=1,a}$ | Abundance at age $a$ at the beginning of the first year |
| $\sigma^2_{\log N_{t=1}}$ | Variance of distribution of abundance at age at the first year |
| $b_0,\ b_1$ | Logistic maturation parameters |
| $ratio_{female}$ | Ratio of female fish |
| $\alpha,\ \beta$ | Parameters in allometric length-weight relationship |
| $\alpha_{SR},\ \beta_{SR}$ | Parameters in Beverton–Holt stock–recruitment model |
| $\mu_R$ | Mean of recruitment |
| $\sigma^2_{\log R}$ | Variance of distribution of recruitment |
| $\sigma^2_{\log N}$ | Variance of process error in abundance at age |
| $\sigma^2_{\log Y}$ | Variance of observation errors for fishing fleet yield |
| $\sigma^2_{\log I}$ | Variance of observation errors for relative biomass indices for fishing fleet |
| $q$ | Catchability for relative biomass index for fishing fleet |

| | |
|---|---|
| $u_q$, $l_q$ | Lower and upper estimation bounds for catchability $q$ |
| $\text{logit}(q)$ | Actual parameter for estimating catchability $q$ |
| **Derived parameter(s)** | |
| $N_{t,a}$ | Abundance at age $a$ at the beginning of year $t$ |
| $N_{t,a}(x)$ | Abundance of length class x at age $a$ at the beginning of year $t$ |
| $X_r$ | Discrete random variable for length of recruitment |
| $f_r(x)$ | Probability mass function of $X_r$ |
| $F_t(x)$ | Fishing mortality rate at length class $x$ in year $t$ |
| $S(x)$ | Selectivity at length class $x$ |
| $Z_t(x)$ | Total mortality rate at length class $x$ in year $t$ |
| $p_{t,a,Z}(x)$ | Relative length distribution at age $a$ at the end of year $t$ where the process of morality in year $t$ took place |
| $L_a$ | Length at age $a$ |
| $\rho_G$ | Brody coefficient, $e^{-\kappa}$ |
| $\mu(x)$ | Expected length when individuals of length class $x$ grow up |
| $\sigma^2_{a+1}$ | Variance of length distribution at age $a + 1$ when individuals of length class $x$ at age a grows up |
| $f_{a+1,G}(l \mid x)$ | Conditional probability of individuals of length class $l$ of which length class is $x$ before growing up |
| $p_{t+1,a+1}(l)$ | Relative length distribution at age $a + 1$ at the beginning of year $t + 1$ |
| $f_{t,a}(x)$ | Probability of individuals of length class $x$ at the beginning of year $t$ |

| | |
|---|---|
| $\overline{Z}$ | Mean of total mortality rate in all years |
| $Mat(x)$ | Maturation at length class *x* |
| $Spawners_{t,a}(x)$ | Number of spawners of length class *x* at age *a* at the beginning of year *t* |
| $SSB_t$ | Spawning stock biomass at the beginning of year *t* |
| $W(x)$ | Average weight corresponding length class *x* |
| $B_t$ | Biomass at the beginning of year *t* |

**Predicted value(s)**

| | |
|---|---|
| $Y_t$ | Predicted aggregate yield by the fishing fleet in year *t* |
| $C_t(x)$ | Predicted catch of length class *x* by the fishing fleet in year *t* |
| $\hat{I}_t$ | Predicted aggregate relative biomass indices for fishing fleet in year *t* |
| $\underset{\sim}{\hat{p}}_t$ | Vector of predicted catch proportions by length class in year *t* |

**Random effects**

| | |
|---|---|
| $N^*_{t,a}$ | Abundance at age *a* at the beginning of year *t* treated as random effects |

**The others**

| | |
|---|---|
| $L_i$ | Likelihood component *i* |
| $CV_X$ | Coefficient of variation of *X* |

| | |
|---|---|
| $\rho_{t,y}(\theta)$ | Relative difference of estimate for parameter $\theta$ in year $t$ when using data of which terminal year is $T - y$ and estimate for parameter $\theta$ in year $t$ when using data of which terminal year is $T$ |
| $\mathrm{RD}_i(\theta)$ | Relative difference of estimate and true value for parameter $\theta$ of pseudo data set $i$ |

**Table 2.** Description of model scenarios.

| Model scenario (corresponding likelihood components) | | (i) Stock–recruitment relationship | |
|---|---|---|---|
| | | (i-1) Beverton–Holt model | (i-2) mean-deviations |
| (ii) Random effects | (ii-1) No random effects | M1<br>($L_1$, $L_2$, $L_5$, $L_6$, and $L_7$) | M4<br>($L_1$, $L_2$, $L_3$, $L_5$, $L_6$, and $L_7$) |
| | (ii-2) Only recruitment after the first year | M2<br>($L_1$, $L_2$, $L_4^{(1)}$, $L_5$, $L_6$, and $L_7$) | M5<br>($L_1$, $L_2$, $L_3$, $L_4^{(1)}$, $L_5$, $L_6$, and $L_7$) |
| | (ii-3) Abundance at age after the first year | M3<br>($L_1$, $L_2$, $L_4^{(2)}$, $L_5$, $L_6$, and $L_7$) | M6<br>($L_1$, $L_2$, $L_3$, $L_4^{(2)}$, $L_5$, $L_6$, and $L_7$) |

**Table 3.** Number of free parameters estimated for each model scenario fitted to Korean chub mackerel data.

| Parameter | M1 | M2 | M3 | M4 | M5 | M6 |
|---|---|---|---|---|---|---|
| $\log(\mu_{N_{t=1}})$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\log(N_{t=1,a})$ | 6 | 6 | 6 | 6 | 6 | 6 |
| $\log(\mu_R)$ | 0 | 0 | 0 | 1 | 1 | 1 |
| $\log(N_{t,r})$ | 0 | 0 | 0 | 19 | 0 | 0 |
| $\log(\alpha_{SR})$ | 1 | 1 | 1 | 0 | 0 | 0 |
| $\log(\beta_{SR})$ | 1 | 1 | 1 | 0 | 0 | 0 |
| $\log(F_{init})$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\log(F_t)$ | 20 | 20 | 20 | 20 | 20 | 20 |
| $\log(M)$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\log(\kappa)$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\log(\sigma_G)$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\log(L_{50\%})$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\log(\gamma)$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\mathrm{logit}(q)$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\sigma_{\log N}$ | 0 | 1 | 1 | 0 | 1 | 1 |
| Total | 36 | 37 | 37 | 54 | 36 | 36 |

**Table 4.** Input values used. The value of parameters $\mu_r$ , $L_\infty$ , and $ratio_{female}$ are taken from previous studies (Choi et al. 2000; Kim et al. 2020) and the value of parameter $\sigma_r^2$ is assumed to be CV of 10% about $\mu_r$. Using related data from other sources, parameters $b_0$ , and $b_1$ are externally estimated in the length-maturation relationship, and $\alpha$ and $\beta$ in the allometric length-weight relationship. The estimates are used as input values in this model.

| Parameter | Input value |
|---|---|
| $\mu_r$. | 18.0 (cm) |
| $\sigma_r^2$ | $(0.1\cdot18.0)^2$ ($cm^2$) |
| $L_\infty$ | 51.67 (cm) |
| $b_0$ | 20.11 |
| $b_1$ | 0.7 ($cm^2$) |
| $\alpha$ | 0.0028 ($g/cm^{3.43}$) |
| $\beta$ | 3.43 |
| $ratio_{female}$ | 0.6 |

**Table 5.** Level of observation and process errors used in generating pseudo data for each model scenario.

| | M1 | M2 | M3 | M4 | M5 | M6 |
|---|---|---|---|---|---|---|
| Sample size | 200 | 200 | - | 200 | 200 | 200 |
| $CV_Y$ | 0.1 | 0.1 | - | 0.1 | 0.1 | 0.1 |
| $CV_I$ | 0.1 | 0.1 | - | 0.1 | 0.1 | 0.1 |
| $CV_R$ | - | 0.1 | - | - | 0.1 | 0.1 |
| $CV_N$ | - | - | - | - | - | 0.1 |

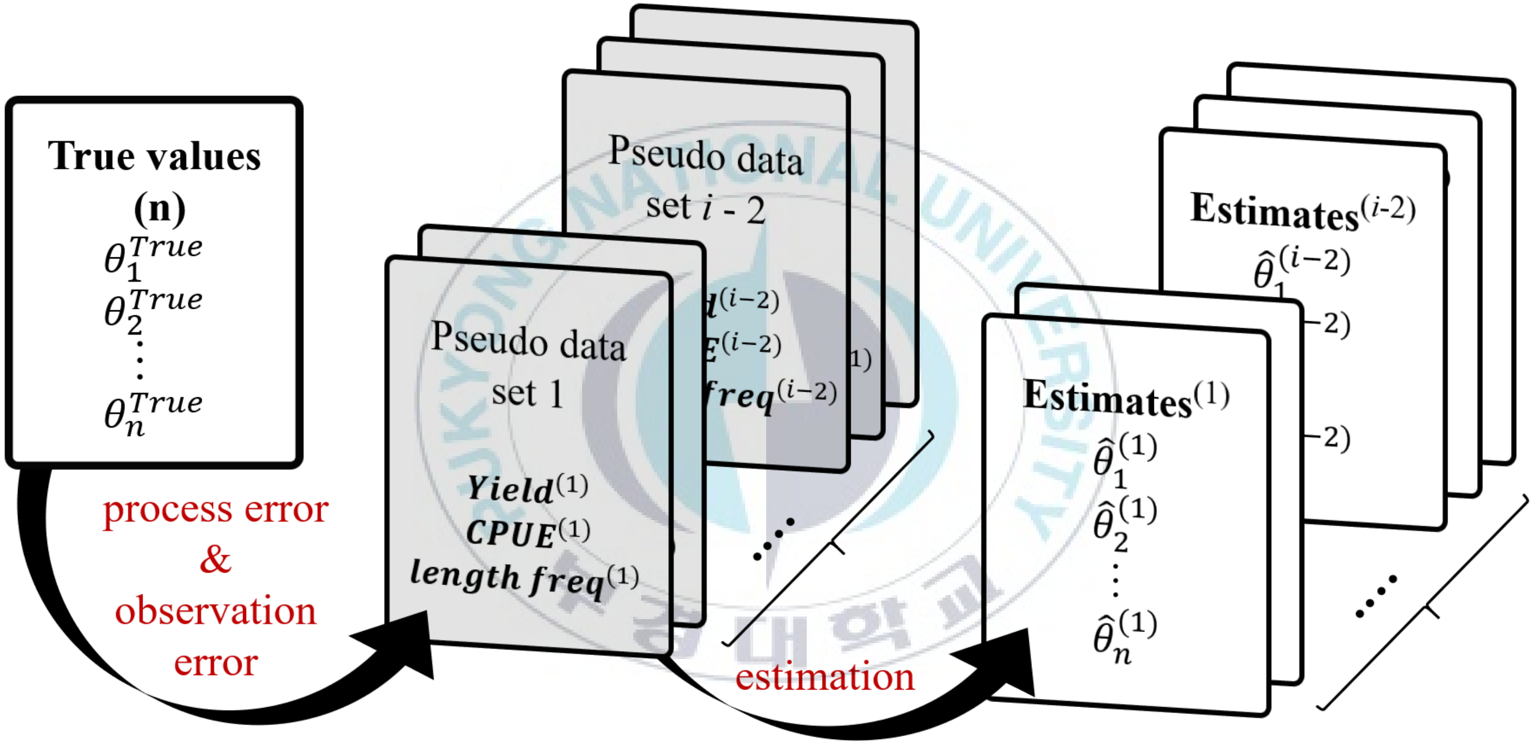

**Figure 1.** Overall structure of the simulation study. First, all true values and the number of iterations (*i*) are set. A total of *i* pseudo data sets considering process errors and observation errors are generated. A total of *i* estimations are performed using each pseudo data set.

# 3. Results

Among the six models I fit to the Korean chub mackerel data, five model scenarios converged. Model scenario 3 (M3), which allowed process errors in abundance at age and used the Beverton–Holt model for predicting annual recruits, did not. The converged five model scenarios had an invertible Hessian providing variance estimation of parameters estimated by ML (Miller and Hyun 2018).

## 3.1. Goodness-of-fit

The goodness-of-fit values of each model scenario (except M3) for annual yields, CPUEs, and length frequency data were generally good (Figure 2, Figure 3, and Figure 4.). In particular, M6, the state-space model in which abundances at age after the first year are treated as random effects and recruitment is related independent of spawning stock biomass showed the best outcomes with respect to goodness-of-fit for all data used in this length-based model. To compare results, all model scenarios were divided into two classes depending on the form of the stock–recruitment relationship, i.e., M1 and M2 (Beverton–Holt), and M4, M5, and M6 (mean-deviations).

Of the two model scenarios, M1 and M2, that use the Beverton–Holt model for annual recruits, M1 performed better than M2 in goodness-of-fit for yields

(Figure 2). Of the three model scenarios, M4, M5, and M6, where the logarithm of recruitment is normally distributed about a mean value, the result of M6 was best in goodness-of-fit for yields (Figure 2b). Although the results of M1 and M6 in each class were the best in their respective classes, all models showed good results in general.

M2 performed much better than M1 in goodness-of-fit for CPUE (Figure 3). Although the predicted line in M1 penetrated all CPUEs, it did not seem to explain the annual variation of CPUE satisfactorily. In goodness-of-fit for the CPUEs of M4, M5, and M6, all predicted lines of each model scenario seemed to explain the annual variation of CPUE well (Figure 3). Within each class of model scenarios (Beverton–Holt or mean-deviations), state-space models (M2, M5, and M6) outperformed non-state-space models (M1 and M4) with respect to goodness-of-fit for CPUE. Furthermore, comparing the results of goodness-of-fit for CPUE of M1 and M4, M4 (which included penalized likelihood for annual recruitment) performed better than M1.

In goodness-of-fit for the length frequency data by the fishing fleet catch assumed to follow a multinomial distribution, the difference between all model scenarios was negligible (Figure 4.). Their annual predicted lines were almost bell-shaped. For this reason, some annual length frequency data that do not have a unimodal shape or have a skewed distribution did not fit well with the predicted line.

## 3.2. Retrospective pattern

Initially, when I peeled off the data from the terminal year for checking the retrospective pattern and achieving Mohn's $\rho$, the state-space models (M2, M5, and M6) failed to converge. To mitigate this issue, the variance of process error in state-space models was treated as an input value, which was set by the estimate that is achieved when the state-space model converged for all Korean chub mackerel data. The variance of process error in the state-space model out of some free parameters was treated as the input value because it was the most difficult to estimate compared to the others.

The highest absolute value of Mohn's $\rho$ estimates for fully-selected fishing mortality rate and spawning stock biomass for all model scenarios was approximately 0.227 and the suggested retrospective patterns were relatively minor (Figure 5). M1 had the lowest absolute value of Mohn's $\rho$ estimates, followed by M6. With respect to consistency of model estimates when data are peeled off in sequence, M1 and M6 had better results than other model scenarios (Figure 6).

## 3.3. Simulation study

In the simulation study, state-space models (M2, M5, and M6) did not noticeably converge owing to the difficulty of estimating the variance of process errors. For this reason, the variance of process errors in state-space models was also treated as an input value. Consequently, the relative difference of the variance of process errors in state-space models could not be calculated.

The number of convergences among model scenarios in the simulation study varied considerably (Table 6). M6, which outperformed the other model scenarios with respect to the goodness-of-fit and retrospective pattern, converged most frequently. Within each class of model scenarios (Beverton–Holt or mean-deviations), state-space models (M2, M5, and M6) outperformed the other models (M1 and M4). The more random effects that were considered, the better the result of convergence (M5<M6). Among model scenarios including varying stock–recruitment relationships with the same assumption about random effects ((M1 and M4) and (M2 and M5)), the model scenarios that included the Beverton–Holt model converged less than others.

The relative differences of each parameter estimate were generally distributed around zero in the simulation study (Figure 7). However, the logarithm of the initial fully-selected fishing mortality rate $(\log F_{init})$ and the variance of the stochastic error term in the von Bertalanffy growth model $(\log \sigma_G)$ showed a wide distribution of relative differences. Furthermore, fits of some model scenarios in the simulation study produced a bias in the estimate of these $(\log F_{init}, \log \sigma_G)$. Whether the bias was positive or negative among model scenarios was not consistent. M1 and M4 performed better than the other model scenarios with respect to the degree of distribution of relative difference and bias of each parameter estimate. This result is expected, given that M1 and M4, which have no random effects, considered only observation error in generating pseudo data sets. Given that state-space models (M2, M5, and M6) included observation error as well as process error in abundance at age in generating pseudo data sets, the results of relative difference seemed to be good.

### 3.4. Estimates

The estimates of the natural mortality rate for all model scenarios except M3 were reliable but differed depending on the model scenario (Table 7). The estimate of natural mortality rate in M6 was 0.10 year$^{-1}$. This value was much lower than that in the other model scenarios. Instead of having a lower estimate of the natural mortality rate in M6 than in the others, however, estimates of annual fully-selected fishing mortality rate were generally higher than those of the others (Figure 8). Although the scale of estimates of annual fully-selected fishing mortality rate differed for all model scenarios, their annual trends were similar.

The estimates of parameters related to body growth caused length distributions by age to be different by each cohort. For example, the probability mass function of length distributions by age of the cohort recruited at the beginning of 2014 is described in Figure 9. As it was assumed that the length distribution for recruitment is fixed, all length distributions for age 1 were the same regardless of converging model scenarios. Although widening the variance of length distribution with aging was similar for all model scenarios, the mean values of length distribution by age were not. Remarkably, the gaps between the mean values of length distribution by age decreased with aging until the maximum age. This was consistent with the assumption that the process of body growth follows the von Bertalanffy growth model where the growth rate of the individual reduces as it grows. However, the gap between age 5 and maximum age, age 6+, increased again, unlike the behavior for previous ages. This was because age plus group was considered. In other words, the length distribution for the maximum age of a certain cohort was calculated considering not only the individuals of the cohort but also the remaining individuals from previous cohorts that survived without dying. Comparing various models, the gaps between the mean value of length distribution by age were the largest in M6. As a result, the mean value of length distribution for the maximum age, 6+, was the largest, i.e., approximately 31.5 cm.

There was no significant difference in the estimated selectivity curves for various converging model scenarios (Figure 10). The overall shapes were similar; however, there was a slight difference in the selection below 30 cm. The values of $L_{50\%}$ ranged between 29–31 cm and the values of selection of 1.0 were represented at approximately 40 cm for converging model scenarios.

The annual estimate of biomass and spawning stock biomass for all model scenarios showed a similar trend, but the scale and range of those differed considerably (Figure 11). The scale and range of estimated annual biomass and stock spawning biomass in M6 was the lowest.

Within the class of models that included the Beverton–Holt model for recruitment (M1 and M2), the shape and scale of estimated lines for the Beverton–Holt model were different (Figure 12). The shape of the estimated line for the Beverton–Holt model in M1 increased rapidly at the beginning of spawning stock biomass. Once the spawning stock biomass was above a certain level, there seemed to be no significant change in recruitment as the spawning stock biomass changed. As M1 was assumed to have deterministic annual recruitment depending on the estimated line for the Beverton–Holt model, all points representing estimated annual recruit were on the line. These results implied that recruitment in M1 was not significantly affected by spawning stock biomass, though recruitment is related to spawning stock biomass when using the Beverton–Holt model. By contrast, the shape of the estimated line for the Beverton–Holt model in M2 gently increased as compared to that of M1, which suggested that as the spawning stock biomass increases, recruitment also increases. As M2 treated recruitment as a random effect, however, all points representing predicted annual recruit were distributed above or below the line, unlike for M1.

**Table 6.** Number of convergences by model scenario. 1000 denotes the total number of iterations in the simulation study.

| Model | Convergence (1000) |
|---|---|
| M1 | 364 |
| M2 | 585 |
| M3 | - |
| M4 | 448 |
| M5 | 787 |
| M6 | 946 |

**Table 7.** Estimates of natural mortality rate by model scenario.

| Model | Natural mortality rate ($\text{year}^{-1}$) |
|---|---|
| M1 | 0.374 |
| M2 | 0.500 |
| M3 | - |
| M4 | 0.527 |
| M5 | 0.577 |
| M6 | 0.100 |

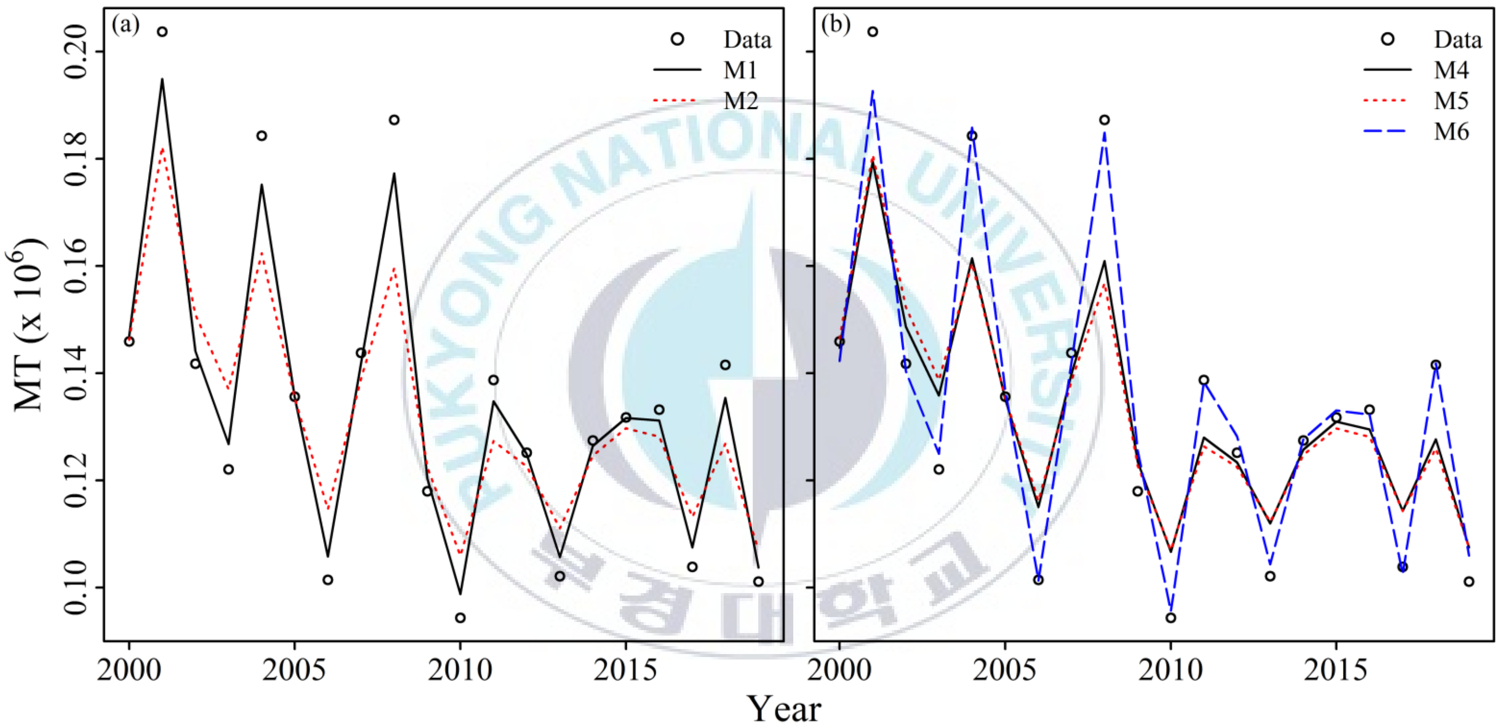

**Figure 2.** Goodness-of-fit for annual total yields of chub mackerel from 2000–2019. Panel (a) describes the results of M1 and M2, and panel (b) represents those of M4, M5, and M6, respectively. The points are observed yields and each line represents predicted values by model scenario.

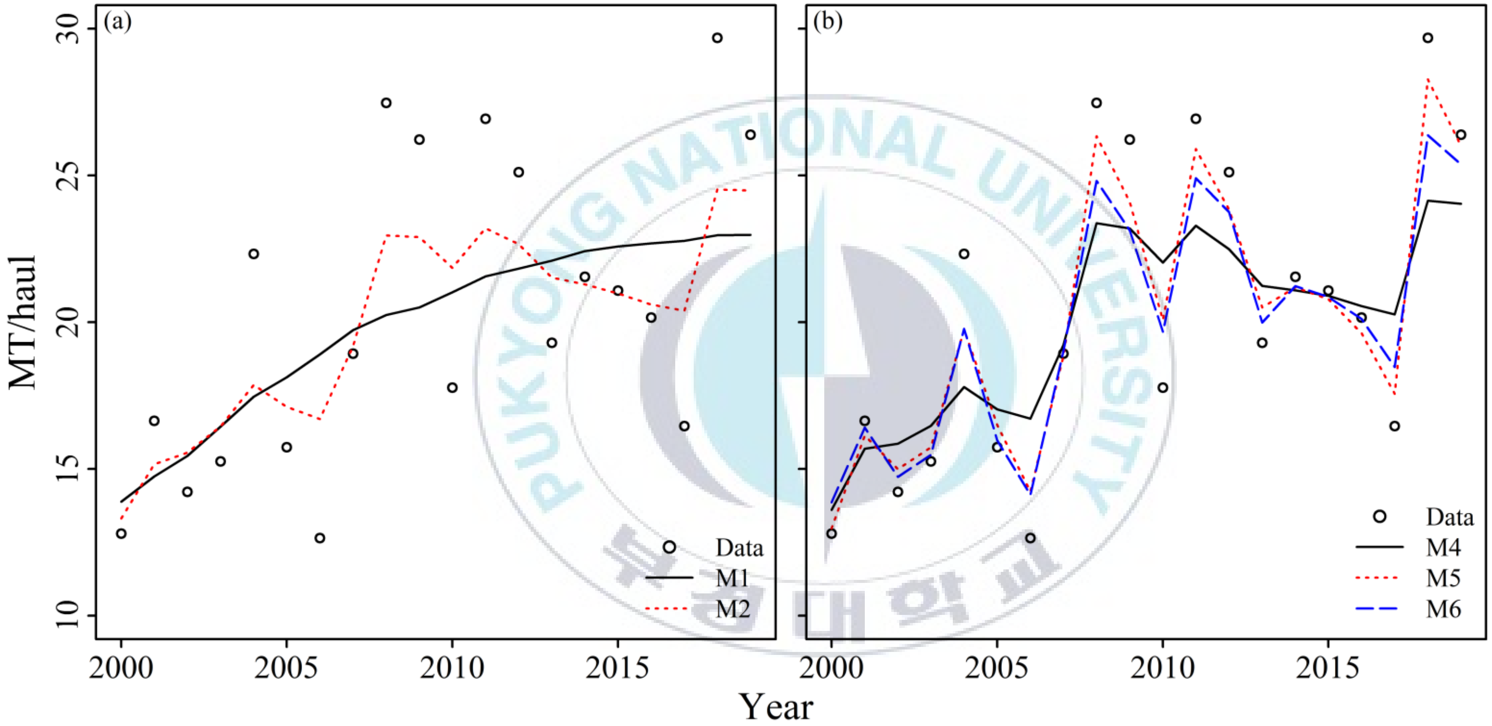

**Figure 3.** Goodness-of-fit for annual CPUE's of chub mackerel from 2000–2019. Panel (a) describes the results of M1 and M2, and panel (b) represents those of M4, M5, and M6, respectively. The points are observed CPUE's and each line represents predicted values by model scenario.

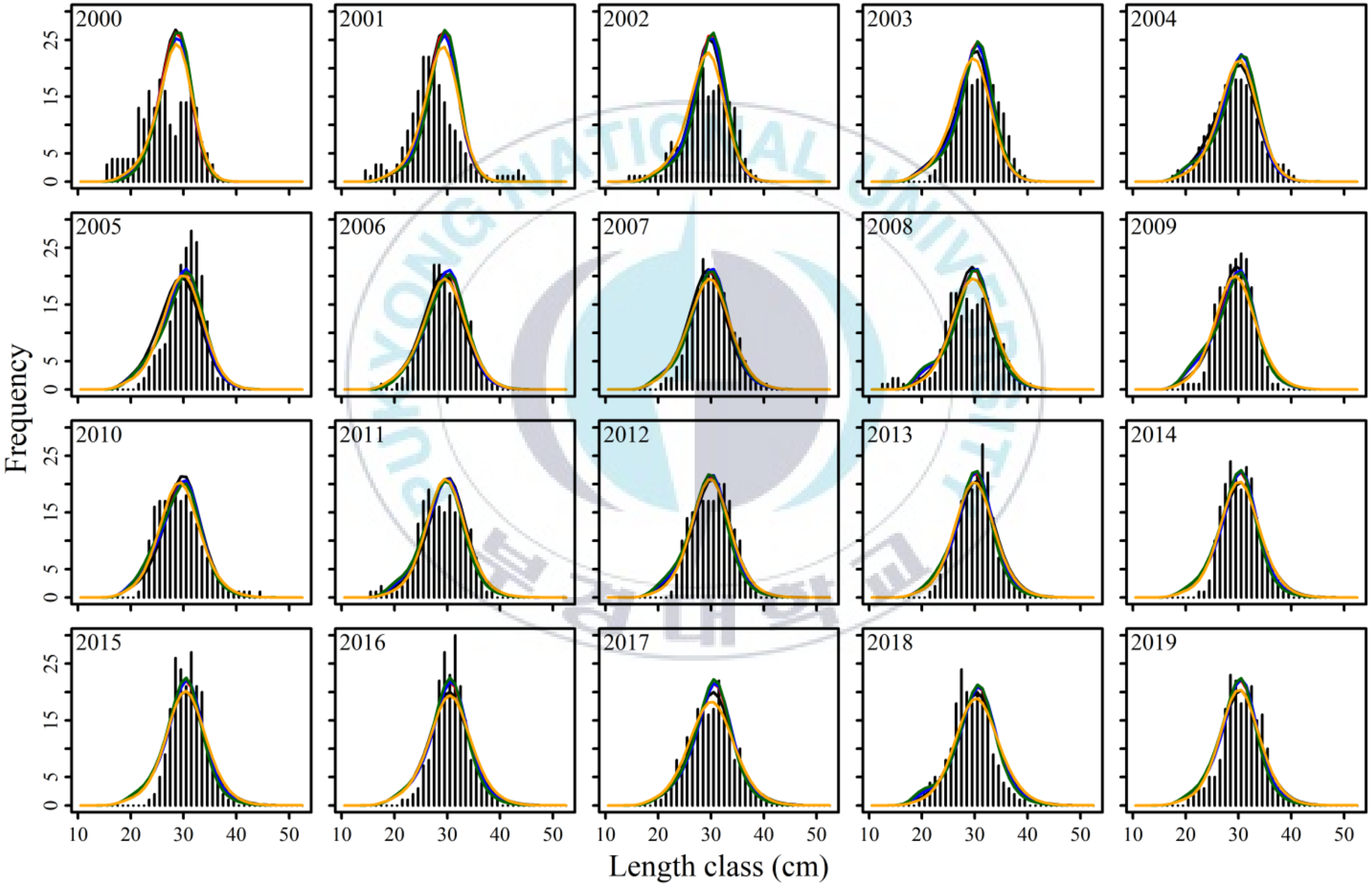
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
0
5
15
25
Frequency
10
20
30
40
50
Length class (cm)

**Figure 4.** Goodness-of-fit for length frequency data of chub mackerel by the fishing fleet catch from 2000–2019. The histograms are observed data and each line is predicted values by model scenario.

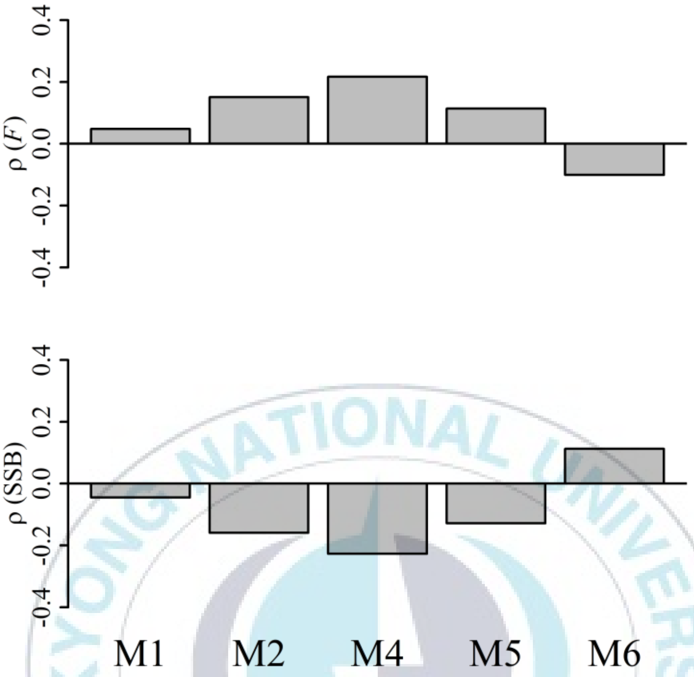

**Figure 5.** Mohn's $\rho$ for fully-selected fishing mortality and spawning stock biomass by model scenario.

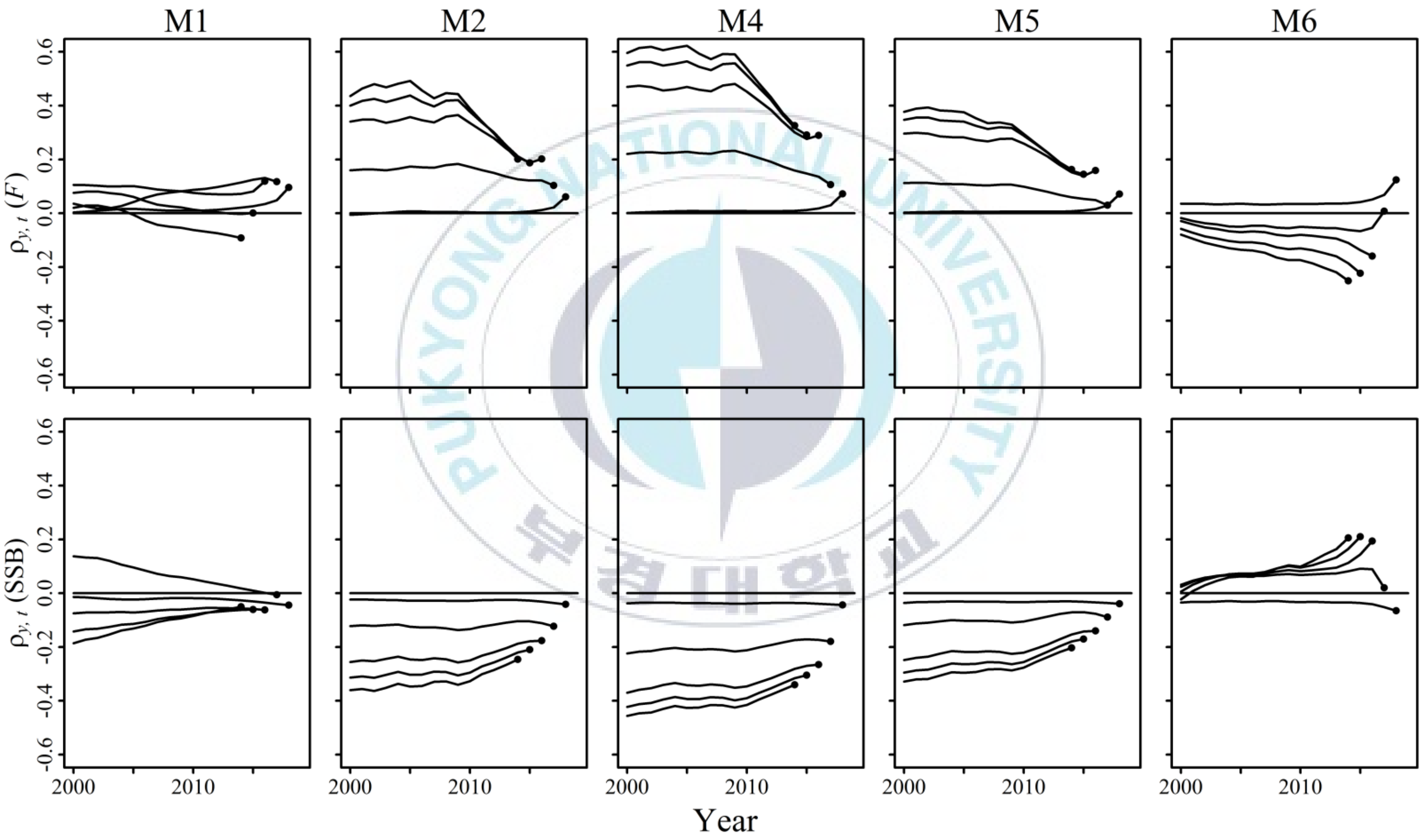
M1
M2
M4
M5
M6
$\rho_{y,t}$ ($F$)
$\rho_{y,t}$ (SSB)
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
2000
2010
Year

**Figure 6.** Relative difference plot of retrospective patterns for fully-selected fishing mortality and spawning stock biomass by model scenario. Five years from the terminal year in sequence are peeled off.

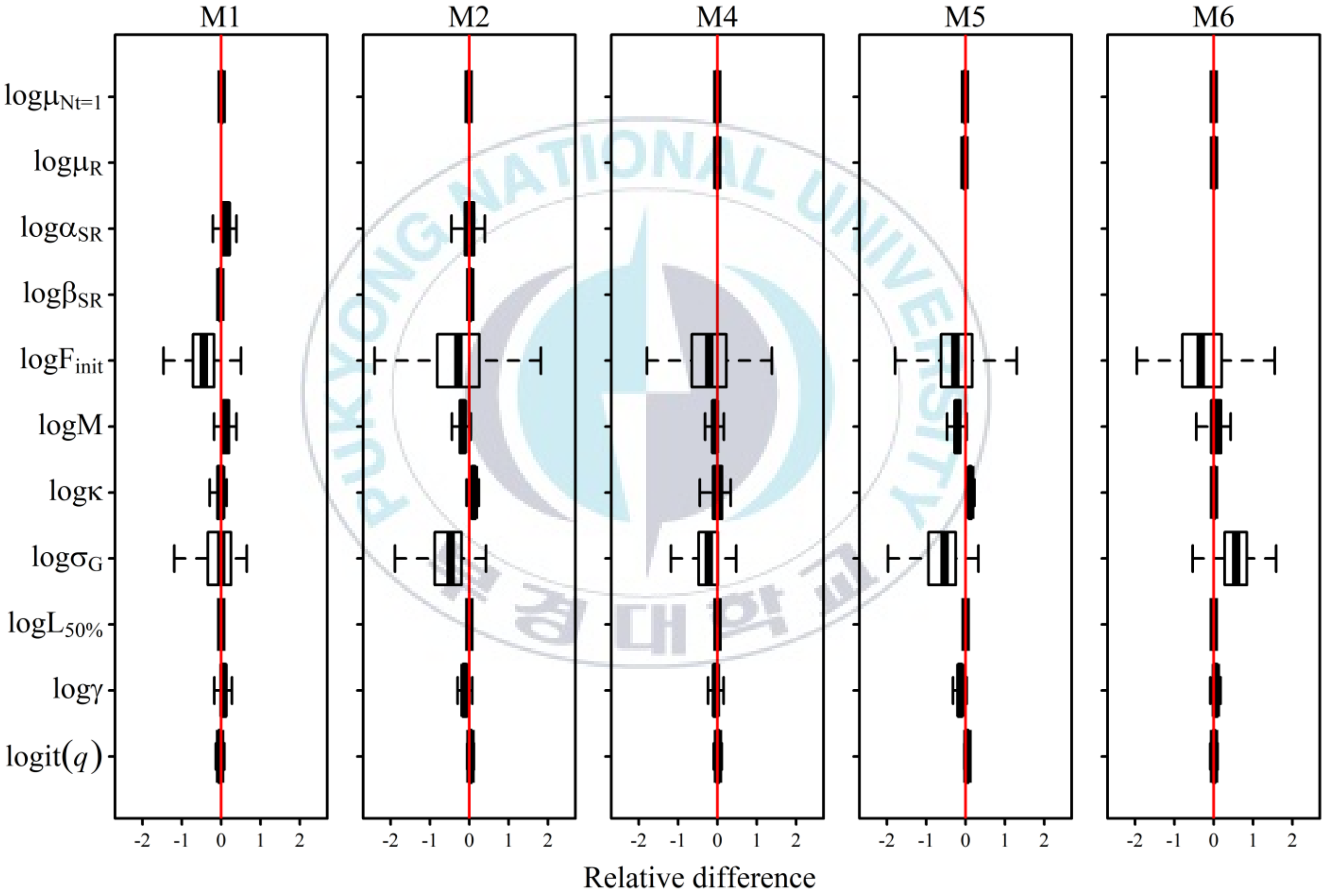
M1
M2
M4
M5
M6
$\log\mu_{Nt=1}$
$\log\mu_{R}$
$\log\alpha_{SR}$
$\log\beta_{SR}$
$\log F_{init}$
$\log M$
$\log\kappa$
$\log\sigma_{G}$
$\log L_{50\%}$
$\log\gamma$
$\mathrm{logit}(q)$
-2
-1
0
1
2
Relative difference

**Figure 7.** Box plot of relative differences of estimates for pseudo data generated in simulation study by model scenario. The red line indicates that the relative difference is zero.

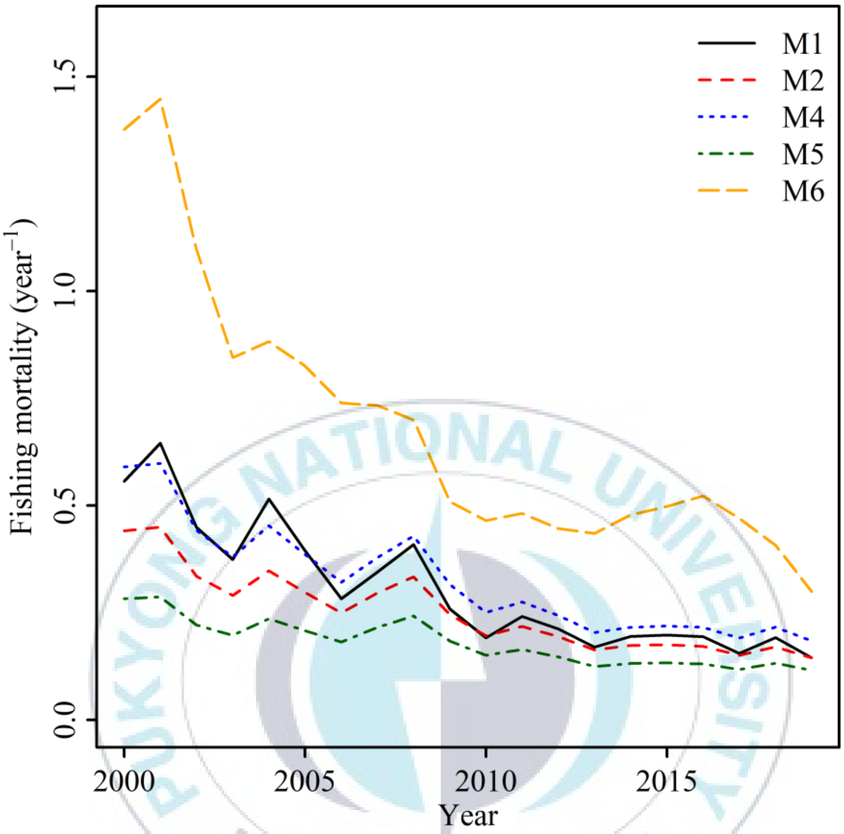

**Figure 8.** Estimates of fully-selected fishing mortality rate by model scenario.

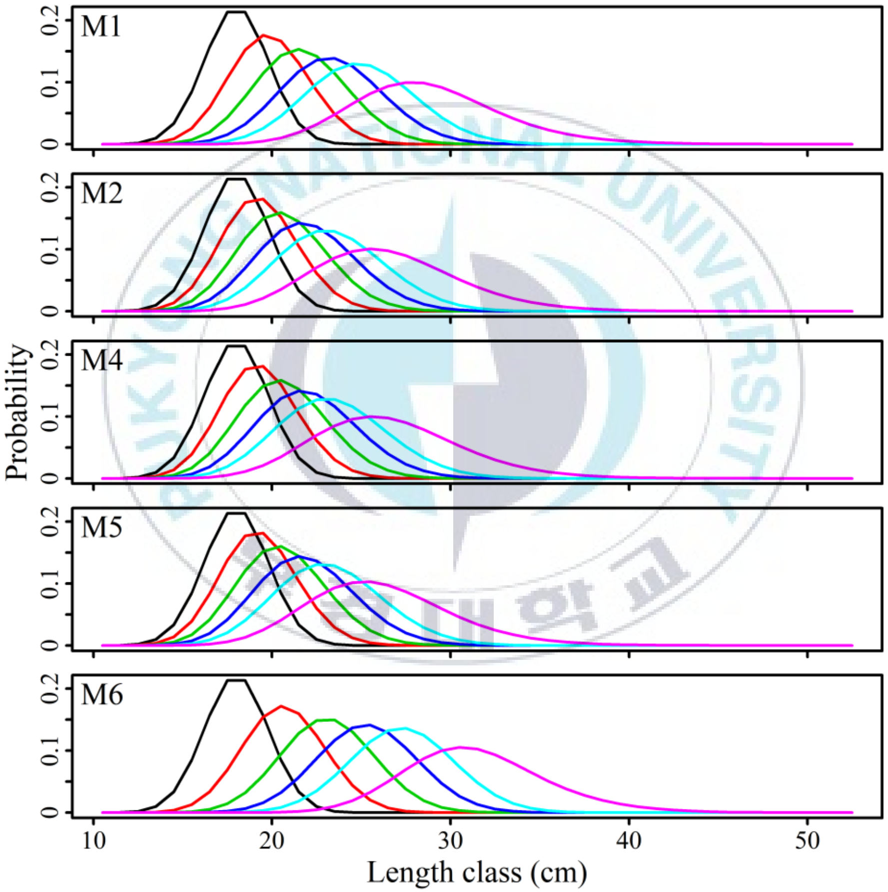
M1
M2
M4
M5
M6
0
0.1
0.2
Probability
10
20
30
40
50
Length class (cm)

**Figure 9.** Estimated probability mass function (PMF) of length class by age of the cohort recruited at the beginning of 2014 by model scenario. The black line represents estimated PMF of length class for age 1 in 2014. The red line represents estimated PMF of length class for age 2 in 2015. This process is repeated and the pink line on the far right represents estimated PMF of length class for age 6+ in 2019.

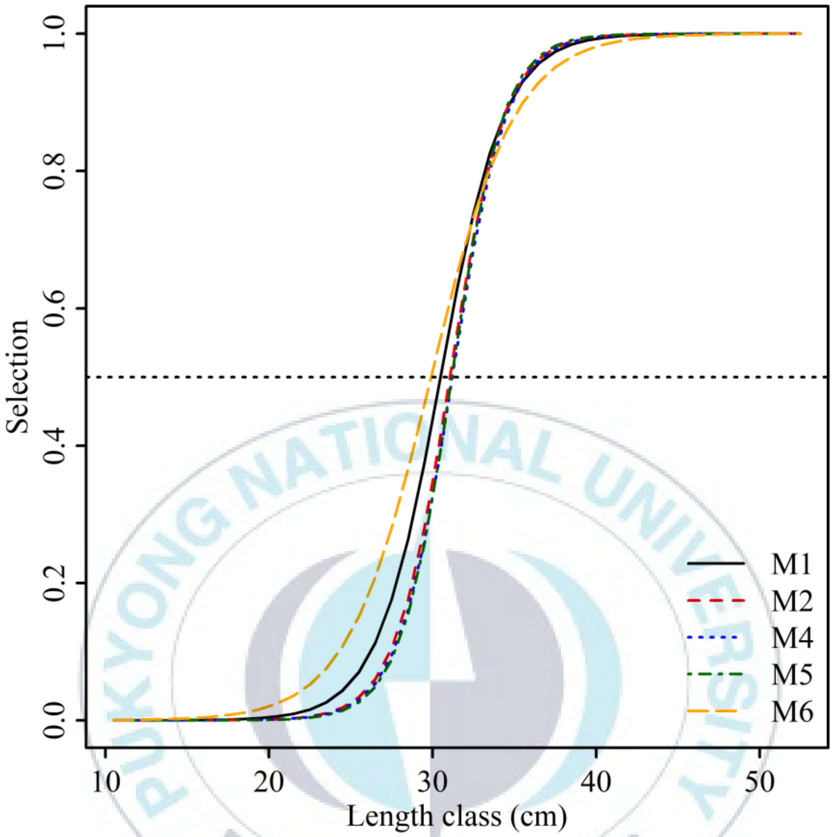

**Figure 10.** Estimated selectivity curve by model scenario. The horizontal dotted line indicates the value of selection is 0.5.

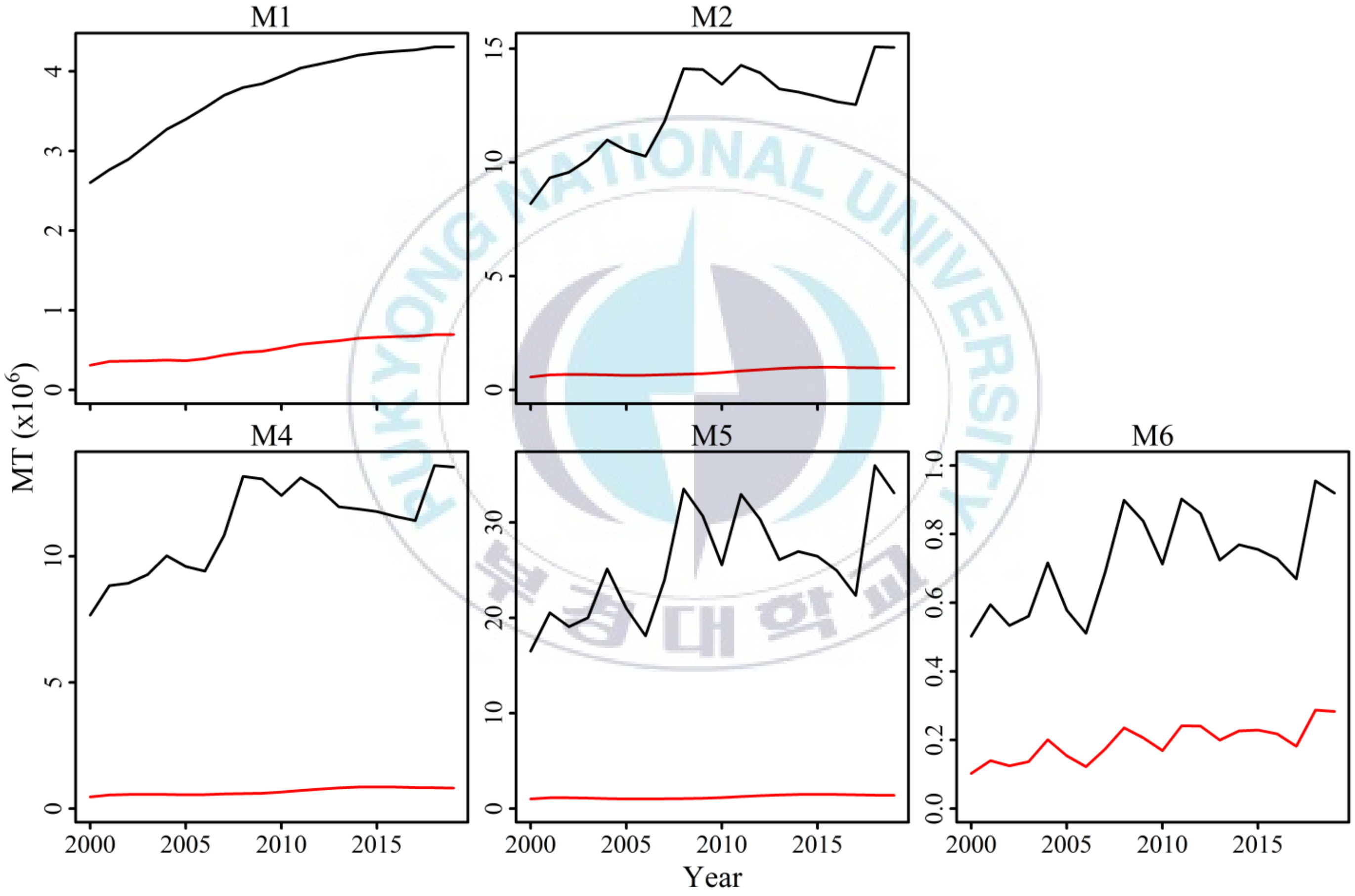
M1
M2
M4
M5
M6
MT (x10$^6$)
Year
0
1
2
3
4
5
10
15
20
30
0.0
0.2
0.4
0.6
0.8
1.0
2000
2005
2010
2015

**Figure 11.** Estimated (M1, M2, M4, and M5) or predicted (M6) annual biomass and spawning stock biomass by model scenario. The black line and the red line describe estimated (M1, M2, M4, and M5) or predicted (M6) annual biomass and spawning stock biomass, respectively. Note the change in the y-axis.

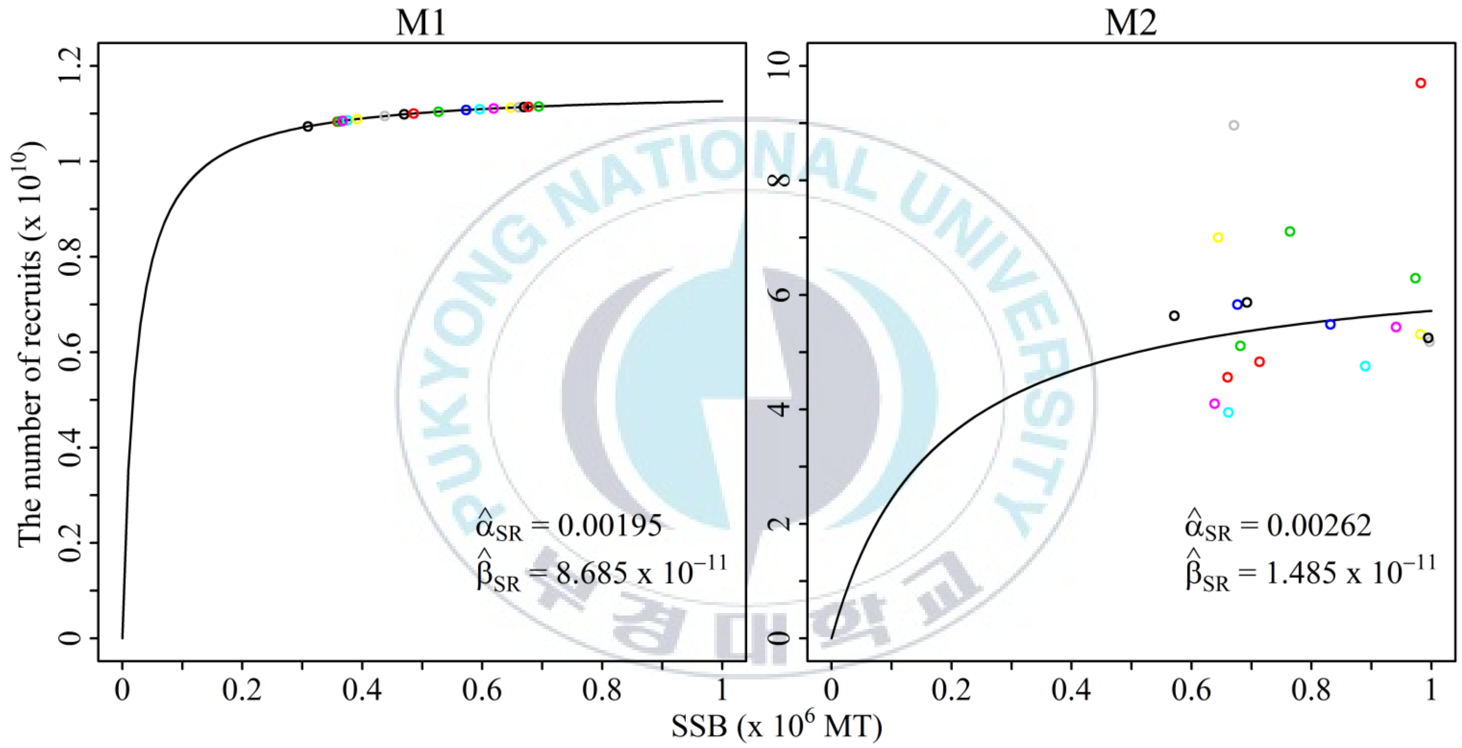

**Figure 12.** Beverton–Holt stock–recruitment relationship by model scenario assumed that recruitment is related to spawning stock biomass (M1 and M2). The solid line represents the estimated line for the Beverton–Holt model and the points are estimated (M1) or predicted (M2) annual recruits. Note the change in the y-axis.

# 4. Discussion

It was observed that the state-space models (M2, M3, M5, and M6) have difficulty in reaching convergence when the variance of process error in abundance at age and year is treated as a free parameter. Specifically, the variances of process error in recruitment as well as in abundance at age except recruitment in M3 and M6 were estimated, but convergence failed to occur. By assuming that there is only one common variance of process error in all abundance at age, M6 converged, but M3 did not. This seems to be because M3, which allowed process error in abundance at age and used the Beverton–Holt model for predicting annual recruitment, has more burdens with respect to having multiple random effects and some free parameters that are difficult to estimate, such as two parameters in the Beverton–Holt model $(\alpha_{SR}, \beta_{SR})$ and natural mortality rate $(M)$.

Although M3 did not converge, the converged state-space model scenarios (M2, M5, and M6) outperformed the models without random effects (M1 and M4) with respect to goodness-of-fit for all data and convergence rate in the simulation study. Furthermore, retrospective patterns in all model scenarios were negligible. Considering the comprehensive results, it was observed that M6, the state-space model in which the abundances at age after the first year are treated as random effects and recruitment is related independent of spawning stock biomass, performed best among all models. In terms of the number of free parameters, M4, which has the maximum number of free parameters (54) and no random effects, was expected to show good results. However, even though M5 and M6 have 36 free parameters, those that have random effects outperformed M4. This might suggest that considering observation errors in data and process errors in time-varying parameters is meaningful in stock assessment.

However, there were some disadvantages when allowing stochasticity in the interannual transition in abundance. As can be seen from the result of convergence in M3 and treating the variance of process error as an input value in a retrospective pattern and simulation study, there were difficulties in numerical optimization. In particular, the state-space models considered in this study are sensitive to the weighing terms for likelihood in the model. Furthermore, considering process error and including random effects in the model considerably increases the computation time. Even though TMB, known to be fast for implementing and fitting state-space models, was used, the process was still time-intensive.

Annual variation of recruits was also found to potentially affect the result of goodness-of-fit for CPUE. The M1's predicted line about CPUE could not adequately explain the annual variation of CPUE. The main difference between M1 and the other models is that recruitment in M1 was not considered a random variable. Consequently, there was no likelihood component for recruitment in M1. This might imply that considering recruitment as a random variable would result in a better outcome in goodness-of-fit for CPUE, regardless of the form of penalized likelihood (M4) or random effects (M2, M5, and M6).

In the simulation study, it was determined that fishing mortality seems to be most affected by variation of a set of pseudo data based on the result of the relative difference of each parameter estimate. For various sets of pseudo data generated by considering observation errors and process errors, most estimated parameters, except fishing mortality, did not show a significant difference depending on variation in pseudo data. However, the relative difference of logarithm of initial fully-selected fishing mortality rate showed a wide distribution, which might imply that the estimate of fishing mortality is more sensitive to data than any other estimates. In other words, data used in estimating parameters might be mainly explained by estimates of fishing mortality in this model. Bias in the variance of stochastic error in the von Bertalanffy growth model could also imply that the estimate was relatively imprecise. This implies that estimating a variance in error term is not easy. In future studies, if some biological information regarding the variance related to body growth could be obtained, it could be treated as an input value, which may yield better results.

All free parameters in each model were estimated, except in M3. Specifically, the length- and time-invariant natural mortality rate, which is conventionally fixed, was estimated. Crone and Hill (2015) assumed this natural mortality rate as 0.5 $\text{year}^{-1}$ in stock assessment for the Pacific mackerel (*Scomber japonicus*). Takahashi et al. (2019) provided the range of estimates of natural mortality explicitly as 0.3–0.5 $\text{year}^{-1}$ for the North Pacific Chub mackerel; Castro and Santana (2000) presented a slightly wider range. The estimates of natural mortality in M1, M2, M4, and M5 were somewhat consistent with the values in previous studies, but in M6 the value was quite different.

In respect to estimates of the parameter related to body growth of fish, there was a considerable difference in the growth coefficient $(\kappa)$ in the von Bertalanffy growth model and the value in the previous study referenced and estimated in this model. While estimating parameters in this study, the input value of $L_{50\%}$ was set to 51.67 cm with reference to Choi et al. (2000), whereas $\kappa$ was treated as a free parameter. Comparing to the value of $\kappa$, i.e. 0.299 provided in Choi et al. (2000), the value in M6, i.e. 0.078, was much smaller. Shiraishi et al. (2008), which studied the growth of Chub mackerel using body length and age data estimated $\kappa$ using the von Bertalanffy growth model, estimated this value as 0.372. This also suggests that estimate of $\kappa$ in the proposed model was very low. However, it might be significant that this study estimated $\kappa$ without age data. Although estimate of $\kappa$ in the proposed model was different from that in previous studies, length distributions by age in Figure 9 were reliable and seem to follow the assumption in the von Bertalanffy model where the growth rate of the individual reduces as it grows. Estimating $\kappa$ as well as $L_{50\%}$ in a future study could provide valuable insights.

In this model, applying the Beverton–Holt model for recruitment seems inadequate to fulfill the purpose of associating annual recruitment with spawning stock biomass. In M1, as the estimated line for Beverton–Holt model had a drastic slope at the beginning of spawning stock biomass, the annual recruitment appeared at a similar level, regardless of the level of spawning stock biomass. In other words, the line in M1 that implied recruitment is almost independent of spawning stock biomass. Although the estimated line for the Beverton–Holt model in M2 indicated that recruitment is related to spawning stock biomass, unlike that of M1, the annual predicted recruitment for which process error is considered is distributed very far from the line. This also suggested that recruitment is relatively independent of spawning stock biomass.

This study has some limitations such as the short period of time series data used and the lack of actual survey data. The unit of time in this model was defined as a year. The period of all annual data used was from 2000–2019, i.e., 20 annual data. Given some details in this study such as the structural complexity of this model, the presence of process errors, and the difficulty in estimating some parameters, this period is not enough. Although the fact that all parameters in this model could be estimated using short-term data is encouraging, long-term data must be applied to this model. In a future study, long-term pseudo data could be used. The model could also be applied to another stock for which actual long-term data is available. Another limitation in terms of data was the lack of actual survey data. All actual data used in this study were from fishing fleet catches. Owing to this lacuna, CPUEs were used for the biomass index. If reliable survey data can be obtained, the estimation can be more reliable.

# References

Bull, B., Francis, R.I.C.C., Dunn, A., McKenzie, A., Gilbert, D.J., Smith, M.H., Bian, R., and Fu, D. 2012. CASAL (C++ algorithmic stock assessment laboratory): CASAL User Manual v2.30-2012/03/21. NIWA Tech. Rep. **135**: 280 pp.

Castro, J.J., and Santana, A.T. 2000. Synopsis of Biological Data on The Chub Marckerel (*Scomber japonicus* Houttuyn, 1782). FAO Fish. Synopsis **39**(157): 1–77.

Choi, Y.M., Park, J.H., Cha, H.K., and Hwang, K.S. 2000. Age and growth of common mackerel, *Scomber japonicus* Houttuyn. Korean Waters. J Korean Soc Fish Res **3**: 1–8.

Cohen, M.-D., 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. NRC Research Press Ottawa, Canada.

Crone, P.R., and Hill, K.T. 2015. Pacific mackerel (*Scomber japonicus*) stock assessment for USA management in the 2015–16 fishing year. Pacific Fish. Manag. Counc. **7700**: 131.

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.

NRC Research Press Ottawa, Canada.

Deroba, J.J., and Schueller, A.M. 2013. Performance of stock assessments with misspecified age-and time-varying natural mortality. Fish. Res. **146**: 27–40. Elsevier.

Developer Core Team, R. 2019. R: A language and environment for statistical computing. Vienna, Austria R Found. Stat. Comput.

Fournier, D.A., Hampton, J., and Sibert, J.R. 1998. MULTIFAN-CL: A length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci. **55**(9): 2105–2116. doi:10.1139/f98-100.

Hilborn, R., and Walters, C.J. 2013. Quantitative fisheries stock assessment: choice, dynamics and uncertainty. Springer Science & Business Media.

Kim, S. ra, Kim, J.J., Park, H.W., Kang, S., Cha, H.K., and Baek, H.J. 2020. Maturity and Spawning of the Chub Mackerel *Scomber japonicus* in the Korean Waters. Korean J Fish Aquat Sci **53**(1): 9–18.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell, B. 2015. TMB: automatic differentiation and Laplace approximation. arXiv Prepr. arXiv1509.00660.

Mace, P.M., and Doonan, I.J. 1988. A generalised bioeconomic simulation model for fish population dynamics. MAFFish, NZ Ministry of

Agriculture and Fisheries.

Mangel, M., Brodziak, J., and DiNardo, G. 2010. Reproductive ecology and scientific inference of steepness: A fundamental metric of population dynamics and strategic fisheries management. Fish Fish. **11**(1): 89–104. doi:10.1111/j.1467-2979.2009.00345.x.

Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86–99. Elsevier B.V. doi:10.1016/j.fishres.2012.10.012.

Methot, R.D., Wetzel, C.R., Taylor, I.G., and Doering, K. 2020. Stock Synthesis User Manual: Version 3.30. 15.

Miller, T.J., and Hyun, S.-Y. 2018. Evaluating evidence for alternative natural mortality and process error assumptions using a state-space, age-structured assessment model. Can. J. Fish. Aquat. Sci. **75**(5): 691–703. NRC Research Press.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. **56**(4): 473–488. Oxford University Press.

Punt, A.E., Huang, T., and Maunder, M.N. 2013. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. **70**(1): 16–33. Oxford University

Press.

Quinn, T.J., and Deriso, R.B. 1999. Quantitative fish dynamics. oxford university Press.

Quinn, T.J., Turnbull, C.T., and Fu, C. 1998. A length-based population model for hard-to-age invertebrate populations.

Rudd, M.B., and Thorson, J.T. 2018. Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. Can. J. Fish. Aquat. Sci. **75**(7): 1019–1035. doi:10.1139/cjfas-2017-0143.

Shiraishi, T., Okamoto, K., Yoneda, M., Sakai, T., Ohshimo, S., Onoe, S., Yamaguchi, A., and Matsuyama, M. 2008. Age validation, growth and annual reproductive cycle of chub mackerel *Scomber japonicus* off the waters of northern Kyushu and in the East China Sea. Fish. Sci. **74**(5): 947–954. Springer.

Takahashi, N., Kamimura, Y., Ichinokawa, M., Nishijima, S., Yukami, R., Watanabe, C., and Okamura, H. 2019. A Range of Estimates of Natural Mortality Rate for Chub Mackerel in the North Pacific Ocean. Available from https://www.npfc.int/range-estimates-natural-mortality-rate-chub-mackerel-north-pacific-ocean.

Vetter, E.F. 1988. Estimation of natural mortality in fish stocks: a review.

Collect. Repr.

# Appendix. TMB code for the state-space length-based assessment model (M6)

```
//A size-based model for the Korean mackerel stock assessment;
//Author: Doyul Kim and Saang-Yoon Hyun

 #include <TMB.hpp>

 // pass missing values
 template<class Type>
 bool isNA(Type x){
     return R_IsNA(asDouble(x));
 }

 // square
 template<class Type>
 Type square(Type i) {
     return i*i;
 }

 //objective function
 template<class Type>
 Type objective_function<Type>::operator() () {

 //Data section
 DATA_INTEGER(nages);       //number of imaginary age classes;

 //Yield and CPUE data
 DATA_MATRIX(data_yieldCPUE);

 //length frequency data
 DATA_VECTOR(x);
 DATA_MATRIX(data_length_freq);
 DATA_SCALAR(neff);          //effective sample size in the multinomial
likelihood for the length data;
 //input values of model structure
 DATA_VECTOR(range_q);

 //length-weigth relationship
 DATA_MATRIX(data_LW);
 DATA_SCALAR(log_alpha_LW);
 DATA_SCALAR(log_beta_LW);

 //length-maturation relationship
```

```
 DATA_MATRIX(data_maturation);
 DATA_SCALAR(b0_mat);
 DATA_SCALAR(b1_mat);
 DATA_SCALAR(ratio_female);

 //likelihood weights
 DATA_SCALAR(lambda);     //weight for length freq data

 //Parameter section
 PARAMETER(mu_r);

 //abundances in initial time
 PARAMETER(logN_1st_time_mean);
 PARAMETER_VECTOR(logN_1st_time);

 //random effects in recruitments
 PARAMETER(logRecruit_mean);
 PARAMETER_VECTOR(logRecruits_re);     //random effects; the number of
parameters are 19 (2001 - 2019)

 //random effects in abundances
 PARAMETER_MATRIX(logN_re);            //random effects; the matrix
dimension is 19 x 5 (except first row and column)

 //fishing mortality
 PARAMETER(logFt_init);
 PARAMETER_VECTOR(logFt);     //annual fishing mortality over time;

 //natural mortality
 PARAMETER(log_M_constant);

 //biomass index
 PARAMETER(logit_q);   //q is from index = q*Bimoass;

 //growth
 PARAMETER(Linf);
 PARAMETER(log_kappa);             //von-Bertalanffy growth equation
 PARAMETER(log_sig_L);             //sig_L, where L is the L_{a+1} equation;

 //selectivity
 PARAMETER(log_L50);                //gear selectivity parameter;
 PARAMETER(log_gamma);              //gear selectivity parameter;

 //stock-recruitment relationship
 //PARAMETER(log_aSR);
 //PARAMETER(log_bSR);
```

```
 //sd of likelihood (weight)
 //random effect or penalized likelihood
 PARAMETER(logN_1st_time_sd);          //known value (outside CPP file)
 PARAMETER(logRE_sd);
 PARAMETER(logFt_sd);                  //known value (outside CPP file)

 //measurement
 PARAMETER(logYield_sd);               //known value (outside CPP file)
 PARAMETER(logCPUE_sd);                //known value (outside CPP file)


 //Derived quantities
 //Derived quantities of data
 vector<Type> yield=data_yieldCPUE.col(3);    //yield data in
data_yieldCPUE
 vector<Type> CPUE=data_yieldCPUE.col(4);   //cpue data in
data_yieldCPUE

 int nlengths=x.size();              //the number of length classes;
                                     //median length of maxFL = 53.5 cm;
 int nyrs=yield.size();              //the number of years
 int r=1;                            //recruitment is defined as the pop size at age
1;

 vector<Type> L=x;                   //length after growth;

 vector<Type> data_length_LW=data_LW.col(0);
 vector<Type> data_weight_LW=data_LW.col(1);
 Type alpha_LW=exp(log_alpha_LW);
 Type beta_LW=exp(log_beta_LW);
 int size_data_LW=data_length_LW.size();

 vector<Type> data_length_maturation=data_maturation.col(0);
 vector<Type> data_rate_maturation=data_maturation.col(1);
 int size_data_maturation=data_length_maturation.size();

 Type q_lower=range_q(0);
 Type q_upper=range_q(1);


 //Derived quantities of paremeter

 Type sig2_r=square(0.1*mu_r);   //variance in length distribution of
recruitment (CV: 10%)

 vector<Type> Recruits(nyrs-1);
 vector<Type> Ft=exp(logFt);
```

```
 Type kappa=exp(log_kappa);                //von Bertalanffy;
 Type Rho=exp(Type(-1.0)*kappa);          //Brody coefficient;
 Type L50=exp(log_L50);                   //gear selectivity;
 Type gamma=exp(log_gamma);                //gear selectivity;
 Type sig_L=exp(log_sig_L);

 //instantaneous natural mortality
 vector<Type> M(nlengths);
 Type M_constant=exp(log_M_constant);

 //constant
 M.fill(M_constant);

 vector<Type> pred_Wt(nlengths);          //body weight by length classes;
 pred_Wt.setZero();
 vector<Type> pred_Maturation(nlengths);   //maturation rate by length
classes;
 pred_Maturation.setZero();
 vector<Type> Sel_PS(nlengths);            //gear selectivity by length
classes;
 Sel_PS.setZero();
 matrix<Type> F_tx(nyrs,nlengths);         //instantaneous fishing mortality;
considering selectivity
 F_tx.setZero();
 matrix<Type> Z_tx(nyrs,nlengths);         //instantaneous fishing mortality +
instantaneous natural mortality
 Z_tx.setZero();
 matrix<Type> ExpZ_tx(nyrs,nlengths);      //survival rate
 ExpZ_tx.setZero();
 vector<Type> Mu(nlengths);               //differ by length
 Mu.setZero();
 vector<Type> SS(nages);                   //differ by age
 SS.setZero();
 vector<Type> p(nlengths);
 p.setZero();
 vector<Type> p_plus(nlengths);            //the last age class;
 p_plus.setZero();
 matrix<Type> f(nages,nlengths);           //length frequency as pmf
 f.setZero();
 array<Type> f_total(nages, nlengths, nyrs);
 f_total.setZero();
 array<Type> pp(nlengths,nlengths,nages);        //pp(L,x,a); //where nlengths
(row) x nlengths (columns) at each level;
 pp.setZero();
 matrix<Type> N(nages,nyrs);
 N.setZero();
```

```
  array<Type> NL(nyrs,nlengths,nages);               //at the level of each age
  NL.setZero();
  matrix<Type> N_save(nages,nyrs);
  N_save.setZero();
  array<Type> NL_re(nyrs,nlengths,nages);            //at the level of each age
  NL_re.setZero();
  matrix<Type> Spawners(nages,nyrs);
  Spawners.setZero();
  array<Type> SpawnersL(nyrs,nlengths,nages);       //at the level of each age
  SpawnersL.setZero();
  matrix<Type> SpawnerBiomass(nages,nyrs);
  SpawnerBiomass.setZero();
  array<Type> SpawnerBiomassL(nyrs,nlengths,nages); //at the level of each
age
  SpawnerBiomassL.setZero();

  Type CNum;
  CNum=Type(0.0);
  Type CWt;
  CWt=Type(0.0);
  vector<Type> TCatch(nyrs);
  TCatch.setZero();
  matrix<Type> Catch(nyrs,nlengths);
  Catch.setZero();
  vector<Type> Yieldhat(nyrs);
  Yieldhat.setZero();

  vector<Type> Pop(nyrs);
  Pop.setZero();
  array<Type> ENx(nyrs,nlengths,nages);
  ENx.setZero();
  vector<Type> EN(nyrs);
  EN.setZero();
  vector<Type> B(nyrs);
  B.setZero();
  vector<Type> EB(nyrs);
  EB.setZero();

  vector<Type> nll(7);          //elements of the objective funtion, which is the
negative loglikelihood;

//Weight, gear selectivity, and maturation rate by length
  for(int xind=0; xind<nlengths; xind++) {
    pred_Wt(xind)=alpha_LW*pow(x(xind),beta_LW)/Type(1000);
    //the division of 1000 is to convert gram to kg;
    Sel_PS(xind)=Type(1.0)/(Type(1.0)+exp(Type(-1.0)*gamma*(x(xind)-
L50)));
```

```
    pred_Maturation(xind)=Type(1.0)/(1+exp(b0_mat-b1_mat*x(xind)));
  }; //mortality and survival rate by time and length

  for(int t=0; t<nyrs; t++) { //t is year; //it is m, month in Quinns code;
    for(int xind=0; xind<nlengths; xind++){
        F_tx(t,xind)=Sel_PS(xind)*Ft(t);
        Z_tx(t,xind)=M(xind)+F_tx(t,xind);
        ExpZ_tx(t,xind)=exp(Type(-1.0)*Z_tx(t,xind));    //survival;
    };
  };

  vector<Type> ExpZ_tx_colsums(nlengths);
  ExpZ_tx_colsums=ExpZ_tx.colwise().sum();
  vector<Type> ExpZ_tx_mean(nlengths);
  ExpZ_tx_mean=ExpZ_tx_colsums/nyrs;

  //LVB body growth;
  SS(0)=sig2_r;                          //SS(0): Var{lengths at age 1}

  f.row(0)=dnorm(x, mu_r, sqrt(SS(0)))/sum(dnorm(x,mu_r,sqrt(SS(0))));

  for(int xind=0; xind<nlengths; xind++) {
      Mu(xind)=Linf-(Linf-x(xind))*Rho;
  };

  for(int a=1; a<nages; a++) {
      //this SS is from Cohen and Fishman (1980); //it was used for the shrimp
in the Quinns paper;
    SS(a)=square(sig_L)*(Type(1.0)-pow(Rho,(Type(2.0)*(a+1)-
Type(2.0)*r)))/(Type(1.0)-square(Rho))+pow(Rho,(Type(2.0)*(a+1)-
Type(2.0)*r))*sig2_r;
  };

  Type kkk;
  for(int a=0; a<nages; a++) {
      for(int xind=0; xind<nlengths; xind++) {
          kkk=Type(0.0);
          for(int Lind=0; Lind<nlengths; Lind++) {
              pp(Lind,xind,a)=Type(0.0);

              if(Lind>=xind){
                 pp(Lind,xind,a)=dnorm(L(Lind),Mu(xind),sqrt(SS(a)));
                 // f(L|x) in Quinn et al. (1998);
                 kkk=kkk+pp(Lind,xind,a);
              };
          };
          for(int Lind=0; Lind<nlengths; Lind++) {
```

```
            pp(Lind,xind,a)=pp(Lind,xind,a)/kkk;    //normalize f(L|x);
        };
    };
};

//Initial time length frequency
matrix<Type> f_1st_time(nages, nlengths);
f_1st_time.setZero();

f_1st_time.row(0)=f.row(0);

for(int a=1; a<nages; a++) {
    for(int Lind=0; Lind<nlengths; Lind++) {
        for(int xind=0; xind<nlengths; xind++) {
            f_1st_time(a,Lind)+=f_1st_time(a-
1,xind)*ExpZ_tx_mean(xind)*pp(Lind,xind,a);
        };
    };
};

for(int a=1; a<nages; a++) {
    f_1st_time.row(a)=f_1st_time.row(a)/(f_1st_time.row(a).sum());
}


//Start of cohort loop
int a;
for(int m=0; m<nyrs; m++) {
    if(m == 0) {
      for(int a=0; a<nages; a++) {
          for(int xind=0; xind<nlengths; xind++) {
              NL(m,xind,a)=Type(0.0);
              NL_re(m,xind,a)=Type(0.0);

              NL(m,xind,a)=exp(logN_1st_time(a))*f_1st_time(a,xind);
              N_save(a,m)+=NL(m,xind,a);
          };
          N(a,m)=N_save(a,m);
      };
    } else if(m>0) {
      a=0;

      for(int xind=0; xind<nlengths; xind++) {
          NL(m,xind,a)=Type(0.0);
          NL_re(m,xind,a)=Type(0.0);

          NL(m,xind,a)=exp(logRecruits_re(m-1))*f(a,xind);
```

```
         N_save(a,m)+=NL(m,xind,a);
      };
      N(a,m)=N_save(a,m);

      for(int a=1; a<nages; a++) {
         for(int Lind=0; Lind<nlengths; Lind++) {
            p(Lind)=Type(0.0);
            for(int xind=0;xind<nlengths;xind++) {
               p(Lind)+=f(a-1,xind)*ExpZ_tx(m-1,xind)*pp(Lind,xind,a);
            };
         };
         if(a!=nages-1) {
           for(int Lind=0;Lind<nlengths;Lind++) {
              NL(m,Lind,a)=Type(0.0);
              NL_re(m,Lind,a)=Type(0.0);
              NL(m,Lind,a)=N(a-1,m-1)*p(Lind);
              N_save(a,m)+=NL(m,Lind,a);
           };
           //N(a,m)=N_save(a,m);
           N(a,m)=exp(logN_re(a-1,m-1));

         } else if(a == nages-1) {
           for(int Lind=0;Lind<nlengths;Lind++) {
              p_plus(Lind)=Type(0.0);
              for(int xind=0;xind<nlengths;xind++) {
                 p_plus(Lind)+=f(a,xind)*ExpZ_tx(m-
1,xind)*pp(Lind,xind,a);
              };
           };

           for(int Lind=0;Lind<nlengths;Lind++) {
              NL(m,Lind,a)=Type(0.0);
              NL_re(m,Lind,a)=Type(0.0);
              NL(m,Lind,a)=N(a-1,m-1)*p(Lind)+N(a,m-1)*p_plus(Lind);
              N_save(a,m)+=NL(m,Lind,a);
           };
           //N(a,m)=N_save(a,m);
           N(a,m)=exp(logN_re(a-1,m-1));

         };
      };// a(2~6) ends here
    };

    for(int a=0;a<nages;a++) {
       for(int Lind=0;Lind<nlengths;Lind++) {
          f(a,Lind)=NL(m,Lind,a)/N_save(a,m);
       };
```

```
      };

      for(int a=0;a<nages;a++) {
           for(int Lind=0;Lind<nlengths;Lind++){
                f_total(a, Lind, m)=f(a, Lind);
           };
      };

      for(int a=0;a<nages;a++) {
           for(int Lind=0;Lind<nlengths;Lind++) {
                NL_re(m,Lind,a)=N(a,m)*f(a, Lind);
           };

           for(int Lind=0;Lind<nlengths;Lind++) {
                SpawnersL(m,Lind,a)=NL_re(m,Lind,a)*pred_Maturation(Lind)*rati
o_female;

SpawnerBiomassL(m,Lind,a)=SpawnersL(m,Lind,a)*pred_Wt(Lind);
                Spawners(a,m)+=SpawnersL(m,Lind,a);
                SpawnerBiomass(a,m)+=SpawnerBiomassL(m,Lind,a);
           };
      };
 }; // m ends here


 for(int m=0;m<nyrs;m++) {
      for(int a=0;a<nages;a++) {
           for(int xind=0;xind<nlengths;xind++) {
                CNum=NL_re(m,xind,a)*(F_tx(m,xind)/Z_tx(m,xind))*(Type(1.0)-
ExpZ_tx(m,xind));
                CWt=CNum*pred_Wt(xind);                //in kg
                Catch(m,xind)+=CNum;
                TCatch(m)+=CNum;
                Yieldhat(m)+=CWt;
                B(m)+=NL_re(m,xind,a)*pred_Wt(xind);
           };
      };
 };   //m ends here;


 //objective functions
 nll.setZero();
 //part 1 of the objective funcion: multinomial for length-frequency data
 matrix<Type> matrix_for_multinomial_prob(nyrs, nlengths);
 // effective sample size (yearly)
   for(int m=0;m<nyrs;m++) {
        vector<Type>
```

```
data_length_freq_eff_ss_row=data_length_freq_eff_ss.row(m);
        vector<Type> prob_length_freq=Catch.row(m)/Catch.row(m).sum();

        matrix_for_multinomial_prob.row(m)=prob_length_freq;

        nll(0)-=lambda*dmultinom(data_length_freq_eff_ss_row,
prob_length_freq, true);
   };


  //part 2 of the objective function: lognormal for yield data
  for(int m=0;m<nyrs;m++) {
      nll(1)-= dnorm(log(yield(m)), log(Yieldhat(m)/Type(1000)), logYield_sd,
true);
  };


  //part 3 of the objective function: lognormal for cpue data
  //log(CPUE) ~ normal(log(q*Bt), sig2_logCPUE);
  Type q=Type(0.0);
  q=q_lower+(q_upper-q_lower)/(1+exp(-logit_q));

  for(int m=0;m<nyrs;m++) {
      nll(2)-= dnorm(log(CPUE(m)), log(q)+log(B(m)), logCPUE_sd, true);
  };


  //part 4 of the objective function: fishing mortality
  //likelihood for logF1
  nll(3)-= dnorm(logFt(0), logFt_init, logFt_sd, true);

  //likelihood for logF2 to logFT
  for(int m=1; m<nyrs;m++) {
      nll(3)-= dnorm(logFt(m), logFt(m-1), logFt_sd, true);
  };


  //part 5 of the objective function: abundances in initial time
  for(int a=0; a<nages; a++) {
      nll(4)-= dnorm(logN_1st_time(a), logN_1st_time_mean,
logN_1st_time_sd, true);
  };


  //part 6 of the objective function: recruitments from the 19 years of 2001 -
2019;
  for(int m=0;m<nyrs-1;m++) {
```

```
        nll(5)-= dnorm(logRecruits_re(m), logRecruit_mean, logRE_sd, true);
  };


  //part 7 of the objective function: logN_re;
  for(int m=0; m<nyrs-1; m++){
        for(int a=0; a<nages-1; a++) {
               nll(6)-= dnorm(logN_re(a,m), log(N_save(a+1, m+1)), logRE_sd,
true);
        };
  };

    Type jnll=nll.sum();

    return jnll;
```

# Acknowledgements

본 연구는 재정적으로 국립수산과학원(연구과제번호: R2020022) 및 한국연구재단(연구과제번호: 2019R1I1A2A01052106)의 지원을 받았으며, 본 연구에 사용된 자료는 국립수산과학원, 통계청으로부터 제공받았습니다.

많은 분들의 도움이 있었기에 연구를 잘 마무리할 수 있었습니다. 특히, 연구의 방향성을 잃지 않을 수 있도록 언제나 올바른 방향을 제시해주시고, 학문적 지도뿐만 아니라 인생에 대한 아낌없는 조언으로 은혜를 베풀어 주신 현상윤 지도교수님께 심심한 감사의 말씀을 드립니다. 제 연구의 심사위원을 맡아 주시고, 연구의 개선을 위해 많은 도움을 주신 김현우 교수님과 서영일 박사님께 감사드립니다. 또한, 매일같이 연구실에서 함께 많은 시간을 보내며, 동고동락했던 수산자원학 연구실의 김진우 형, 이승준 군, 정동원 군, 박민규 군, 안동영 형을 비롯하여 타국에서도 많은 신경을 써주신 김규한 선배에게 감사드립니다. 더불어, 힘들 때 힘이 되어주고 좋은 일에는 무한한 축하를 보내준 고등학교 동창 김종환, 박희근, 이동현, 서규헌, 임건욱, 임형택, 조광재, 차민승, 하동훈, 하병문, 홍지용 그리고 대학교 동창 김유창, 유민호, 이효태에게 감사드립니다. 또한, 언제나 나의 편이 되어주고 가장 가까이에서 항상 큰 힘이 되어준 양수정에게 진심으로 감사의 말씀을 전합니다. 마지막으로, 저와 평생을 함께하며 항상 믿어주시고 조건없는 무한한 사랑을 베풀어 주신 나의 부모님, 부모님처럼 저를 사랑해주고 아껴준 누나, 삶의 활력소가 되어준 조카 소윤이, 성윤이 모두 사랑하고 감사합니다.