



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

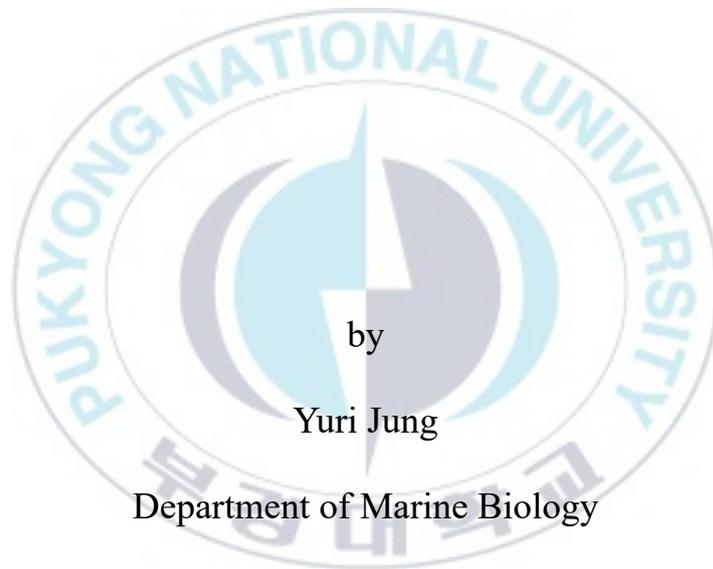
저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Thesis for the Degree of Master of Science

A Bayesian state-space production model
for Korean chub mackerel (*Scomber
japonicus*) stock



by

Yuri Jung

Department of Marine Biology

The Graduate School

Pukyong National University

August 2019

A Bayesian state-space production model
for Korean chub mackerel (*Scomber
japonicus*) stock

[고등어(*Scomber japonicus*) 개체군 자원평가를
위한 베이지안 상태공간 잉여생산량 모델]

Advisor: Prof. Saang-Yoon Hyun

by
Yuri Jung

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 Bayesian state-space production model for Korean chub mackerel
(*Scomber japonicus*) stock

A thesis
by
Yuri Jung

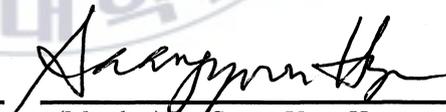
Approved by:



(Chairman) Dr. Seongbaek Yi



(Member) Dr. Young Il Seo



(Member) Dr. Saang-Yoon Hyun

August 23, 2019

Table of contents

List of figures	iii
List of tables	vi
요약	viii
Abstract.....	x
1. Introduction.....	1
2. Materials and Methods.....	5
2.1 Fishery data on mackerel.....	5
2.2 Surplus production model.....	8
2.3 A state-space production model.....	13
2.4 Likelihood function	15
2.5 Prior distributions	17
2.6 MCMC sampling.....	23
2.7 Various production models.....	25
3. Results.....	29
3.1 State-space production model results	29
3.2 Comparison with various production models	39
4. Discussion	45
4.1 Estimates from various production models	45
4.2 Implementation of a state-space production model	48
4.3 Stock assessment using data series from 1999 to 2017	50
4.4 Suggestions for future research.....	60
5. Conclusions	62
6. Acknowledgements	63
7. References	65

8. Appendices.....	72
Appendix A: Normal prior on $\log P_{1976}$ in ADMB-RE.....	72
Appendix B: ADMB-RE code for a Bayesian state-space production model.....	77



List of figures

Figure 1. Chub mackerel data from 1976 to 2017. Panel (a) shows the annual yield collected from entire fisheries on mackerel, and panel (b) shows the catch per unit effort (CPUE) data. Panel (c) presents the percentage of the yield achieved by large purse seine fisheries to total yield on chub mackerel. Units for the yield and CPUE are metric ton (MT) and metric ton per haul (MT/haul).7

Figure 2. Logistic growth (a) and surplus production curve (b) of the Schaefer model under the absence of fishing. Panel (a) shows the population size converges to carrying capacity, where panel (b) depicts the surplus production as the population size grows.12

Figure 3. Posterior densities obtained with the specified priors. Bold lines represent the priors and histograms show the Markov Chain Monte Carlo samples from the posterior distributions in each panel. Units for k , q and σ_o^2 are metric ton (MT), 1/haul and $\log[\text{MT/haul}]$, respectively. Note that a uniform prior for $\log q$ is considered while it is not shown in panel (c).35

Figure 4. Scatter plots of joint posteriors of two parameters. Units for k , q and σ_o^2 are metric ton (MT), 1/haul and $\log[\text{MT/haul}]$, respectively.37

Figure 5. Results from the state-space production model including

predicted CPUE, annual biomass, annual harvest rate and management references. CPUE data (filled squares) is compared with the predicted CPUE (solid line) in panel (a). Panel (b) shows the Maximum sustainable yield (MSY, dashed line) with annual yield (solid line with filled circles). Panel (c) depicts the predicted annual biomass (solid line) with carrying capacity (two-dashed line) and biomass that yields the MSY (B_{MSY} , dashed line). Panel (d) shows the annual harvest rate (solid line) with the harvest rate that corresponds to MSY (H_{MSY} , dashed line).....38

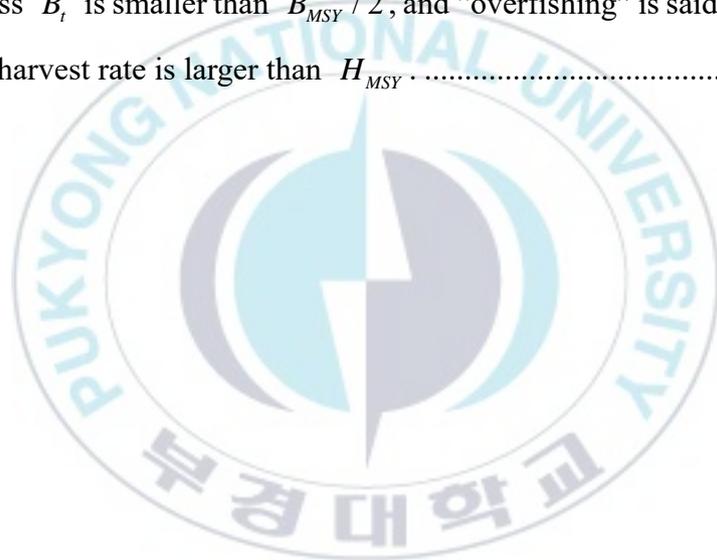
Figure 6. Predicted CPUE and biomass from the four different production models: Clarke-Yoshimoto production model, ASPIC program, Observation model and the state-space production model. Plots in the first row show the predicted CPUE whereas plots in the second row show the predicted biomass. Unit for the CPUE is metric ton per haul (MT/haul) and biomass for 10^6 metric tons (MT).....44

Figure 7. Posterior densities obtained with the specified priors. Dashed lines represent the priors and histograms depict the Markov Chain Monte Carlo samples from the posterior distribution in each panel. Units for k , q and σ_o^2 are metric ton (MT), 1/haul and $\log[\text{MT/haul}]$ respectively. Note that a uniform prior for $\log q$ is considered while it is not shown in panel (c).....57

Figure 8. Comparison of management references and predicted values. Panel (a) shows the CPUE data (filled squares) and the predicted CPUE

(thick solid line) with the 95% credible intervals (dotted lines). Annual yields are shown (solid line with filled circles) with MSY (dashed line) in panel (b). In panel (c), the predicted annual biomass (thick solid line) and 95% credible intervals (dotted lines) are presented with estimated B_{MSY} (dashed line) and carrying capacity (two-dashed line). Panel (d) contrasts the predicted harvest rate (thick solid line) and H_{MSY} (dashed line).....58

Figure 9. Kobe plot showing the predicted trajectories of B_t / B_{MSY} and H_t / H_{MSY} from 1999 to 2017. The stock is said to be “overfished” when the biomass B_t is smaller than $B_{MSY} / 2$, and “overfishing” is said to occur when the harvest rate is larger than H_{MSY} 59



List of tables

Table 1. Selected priors set which satisfies the numerical optimization of the likelihood function. Values in the parentheses are parameters for each prior distribution.....22

Table 2. Diagnostics for Markov Chain Monte Carlo samples for parameters $(r, k, q, \sigma_p^2, \sigma_o^2, P_{1976})$: Dependence factor of Raftery-Lewis statistics (DF), lag-1 autocorrelation, the ratio of the naïve standard error to the time series standard error, and the shape of posterior distributions were checked.33

Table 3. Posterior summary for each parameter $(r, k, q, \sigma_p^2, \sigma_o^2, P_{1976})$ and management references. Note that the units for k, MSY, B_{MSY} are metric ton (MT), q and σ_o^2 are 1/haul and $\log[MT/haul]$, respectively.34

Table 4. Parameter estimates obtained from various production models - Fox model (Fox), Yoshimoto-Clarke model (YC), Schnute model (Schnute), ASPIC program (ASPIC), Observation model (Observation) which the process error is not considered, and the state-space production model (SSP). NAs in the table indicate that the model does not estimate the corresponding parameter(s). Except for the ASPIC results, Units for k, q and σ_o^2 are metric ton (MT), 1/haul and $\log[MT/haul]$ respectively.

For ASPIC, the units for r and q are year^{-1} and $\text{year}^{-1}\text{haul}^{-1}$, respectively. The 95% confidence intervals and 95% credible intervals for estimates from the Observation and SSP are provided in parentheses, respectively. Only positive ranges were showed for the 95% confidence intervals for k and q of the Observation model results.42

Table 5. Selected priors set that satisfied the numerical optimization of the likelihood function. Values in the parentheses are parameters for each prior distribution.54

Table 6. Diagnostics for Markov Chain Monte Carlo samples for parameters $(r, k, q, \sigma_p^2, \sigma_o^2, P_{1999})$: Dependence factor of Raftery-Lewis statistics (DF), lag-1 autocorrelation, the ratio of the naïve standard error to the time series standard error and the shape of posterior distributions were checked.55

Table 7. Posterior summaries of each parameter $(r, k, q, \sigma_p^2, \sigma_o^2, P_{1999})$ and management reference points: maximum sustainable yield (MSY), biomass that yields MSY (B_{MSY}), and harvest rate which corresponds to MSY (H_{MSY}). Units for k , MSY, B_{MSY} are metric ton (MT), q and σ_o^2 are 1/haul and $\log[\text{MT}/\text{haul}]$, respectively.56

고등어 (*Scomber japonicus*) 개체군 자원평가를 위한 베이지안 상태공간
모델

정유리

부경대학교 대학원 해양생물학과

요 약

수산자원 관리를 위해서는 어업이나 과학조사를 통해 수집된 자료에 자원평가 모델을 적용하여 환경수용력이나 개체군의 고유성장률과 같은 개체군에 대한 정보(모수)를 추정하는 단계가 필요하다. 이 단계에서 얻은 모수 추정치들은 수산자원 관리자가 자원 관리방안을 설정하는 데 도움을 준다. 자료의 수집 여부에 따라 다양한 자원평가 모델을 적용할 수 있으나, 그 중에서도 잉여생산량 모델은 타 자원평가 모델에 비해 자료의 요구량이 적고(어획량 자료와 단위노력당 어획량 혹은 과학조사지수 자료), 추정해야 할 모수의 개수가 적다는 장점으로 현재까지도 연구가 되어 왔다. 그러나 잉여생산량 모델을 적용하여 모수를 추정하는 대부분의 연구에서는 단위노력당 어획량자료가 전체적으로 단순하게 증가하거나 감소하는 추세를 보인 반면, 1976년부터 2017년까지 한국에서 수집된 고등어의 단위노력당 어획량자료는 뚜렷한 변동(fluctuation)을 보였다. 본 연구는 뚜렷한 변동을 보이는 고등어 자료를 적합시키고, 모델의 모수들을 추정하고자 상태공간 잉여생산량 모델을 적용하였다. 잉여생산량 모델에 상태공간 구조를 적용함으로써, 모델이 설명하지 못하는 자원량의 자연적인 변동(natural variability)과 자료를 수집하는 과정에서 발생하는 오차를 동시에 고려하였다. 또한, 수치적 최적화(Numerical optimization)를 만족시키는 사전확률 분포를

고려하여 사후확률분포를 얻었고, 이를 통해 모수의 점추정치(point estimates) 및 추정치들에 대한 불확실성을 제시하였다. 모델이 자료를 잘 설명하는 지 비교하기 위해 다른 잉여생산량 모델을 비교하였으며, 본 논문에서 사용된 상태공간 잉여생산 모델이 주어진 고등어 자료를 가장 잘 설명하였다.



Abstract

Surplus production models are considered the simplest stock assessment models, with more parsimonious model parameters, compared to other stock assessment models. In addition to the simple model structure, the models require only two sets of time series data including annual yield and catch-per-unit-effort (CPUE) from commercial fishery or survey index from a scientific survey. While many researchers had applied production models for such advantages, most of the data fitting the models had only shown relatively stable trends of increase or decrease. Korean chub mackerel (*Scomber japonicus*) CPUE data, however, displays a *mixed trend*, showing fluctuations in quantity throughout multiple periods. Thus, leading researchers to get unreliable estimates on model parameters. While simpler model structures provide multiple benefits, production models have been criticized for sacrificing biological realism for mathematical simplicity. A state-space production model is one of the solutions to such objection by accounting for both unmodelled variability on biomass dynamics (process error) and measurement uncertainty (observation error). The main purpose of this

study is to fit the CPUE data with a *mixed trend* using a Bayesian state-space production model. To stabilize the numerical optimization, prior distributions were considered. Implementation is performed in script software ADMB-RE, because it reduces the computational cost on high-dimensional integration and provides Markov Chain Monte Carlo sampling which is required to a Bayesian approach. Applying the state-space production model to annual yield and CPUE data collected from a commercial fishery during 1976-2017, the model estimated key parameters and predicted annual CPUE and biomass. Comparison with results from various production models showed that the state-space production model explained the *mixed trend* in data best. The results suggested that the state-space production model should be preferred to the other production models in fitting CPUE data with a *mixed trend*.

1. Introduction

The main goal of stock assessment is to maximize profit from the catch while also effectively conserving a fish stock (Quinn and Deriso, 1999). Therefore, stock assessment involves quantitative predictions about fish populations using mathematical and statistical models. Surplus production models, or biomass dynamics models fall into a category of stock assessment models, which have been studied extensively for several advantages. Compared with other stock assessment models, production models have parsimonious number of parameters making the model structure simpler. Moreover, the models have lighter data requirement: historical data on yield, and relative abundance index, which can be catch-per-unit-effort (CPUE) from commercial fisheries or survey index from scientific surveys. It is not surprising the production models have long been primarily applied to data-limited situations where age or size data on a stock are not available. Many studies which used various versions of production models, however, have used CPUE data only showed increase or decrease in the overall trend (Carruthers et al., 2011; Chaloupka and Balazs, 2007; Millar and Meyer, 1999; Polacheck et al., 1993; Rankin and Lemos, 2015; Zeller et al., 2008). On

the other hand, the CPUE data on Korean chub mackerel showed conspicuous fluctuations or a *mixed trend*. Although some researchers attempted to estimate the model parameters of production models with the mackerel data, one of the study eliminated the data in 1996 regarding as an outlier which produces a significant peak (Choi et al., 2004). Cho et al. (2009) failed to estimate model parameters of various production models. While production models have many advantages mentioned above, they are criticized for sacrificing biological realism for model simplicity (Millar and Meyer, 1999). To be specific, the model describes population growth with one parameter, intrinsic growth rate, by aggregating natural factors including growth, recruitment, natural mortality. Pella and Tomlinson (1969) pointed out the model structures are too simple to explain the population dynamics with various sources of variability, such as interactions among species and abiotic conditions (Pedersen and Berg, 2017).

A state-space production model is one of the attempts that can alleviate the doubt on the simplicity of production models. It explicitly includes process errors and observation errors to account unmodelled factors and noise in data, respectively. Despite the ability to consider the two sources of

uncertainty, process errors demand the estimation of a large number of free parameters and require high-dimensional integration as well. Because of the computational costs, studies have been implemented linear state-space models, which is ecologically unrealistic (Rivot et al., 2004). Another action for fitting a state-space model is assuming the ratio between the variance of the process and observation errors to be known (Kimura et al., 1996; Ludwig et al., 1988). Fortunately, the development of statistical software such as script software ADMB-RE (Fournier et al., 2012; Skaug and Fournier, 2017) and R package TMB (Kristensen et al., 2016) enabled the implementation of state-space models with Gaussian error structures through the Laplace approximation. Another advantage of ADMB-RE is provision of Markov Chain Monte Carlo (MCMC) sampling without any revision of the computer code. In this study, prior distributions for model parameters are considered to stabilize the numerical optimization.

The main purpose of this study is to apply a state-space production model to Korean mackerel data for which CPUE showed a *mixed trend*. The state-space production model was implemented within ADMB-RE, and prior distributions were considered to aid numerical optimization. The present

study provided estimates of key parameters, including intrinsic growth rate, carrying capacity and annual biomass, as well as their uncertainties. The model also provided management references such as maximum sustainable yield (MSY), harvest rate that correspond to the MSY (H_{MSY}), and biomass that yields the MSY (B_{MSY}). Applying various production models to the same data, predicted CPUE and biomass trajectories were compared to show the model performance.



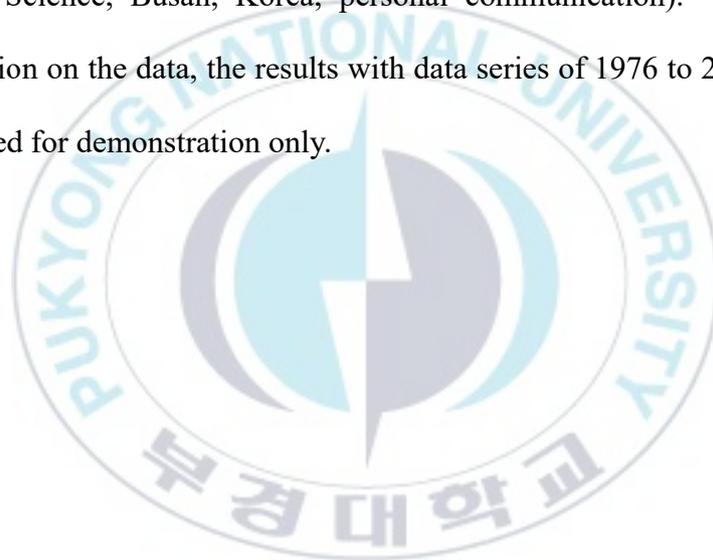
2. Materials and Methods

2.1 *Fishery data on mackerel*

Hilborn (2002) introduced a quote from John Shepherd, stating that counting fish is like counting invisible trees moving around. In other words, counting fish stock size underwater is impossible, unless fishing them all. To resolve such crucial problem, relative index of biomass (CPUE or survey index) is utilized in estimating stock biomass.

Two sets of historical data available on chub mackerel: annual yield and CPUE with units of metric ton (MT) and metric ton per haul (MT/haul). Figure 1a and 1b show the annual yield and CPUE of mackerel from 1976 to 2017. Every year, Statistics Korea have collected the yield from entire fisheries on mackerel. Korean National Institute of Fisheries Science (KNIFS) have gathered the yield and effort data from 70% of the entire large purse seine fisheries to calculate the CPUE data. When the fishing trips over, fishermen fall into the 70% provided the total yield and effort (number of hauls). Effort is calculated by multiplying a constant to the number of fishing days (Y. Seo, Korean National Institute of Fisheries Science, Busan, Korea, personal communication). In this study, I regarded the CPUE

(Figure 1b) to be representative of all chub mackerel fisheries, as large purse seine fisheries produced more than 90% of total yield on average (Figure 1c). As shown in Figure 1, trajectories of yield and CPUE have the peaks with highest value of 415,003MT and 32.44MT/haul in 1996. Several scientists in KNIFS argued the CPUE data lost consistency after the Korea-Japan fisheries agreement, which was signed in 1998, because of the reduction of the fishing ground (S. Kang and H. Cha, Korean National Institute of Fisheries Science, Busan, Korea, personal communication). Regarding their opinion on the data, the results with data series of 1976 to 2017 should be accepted for demonstration only.



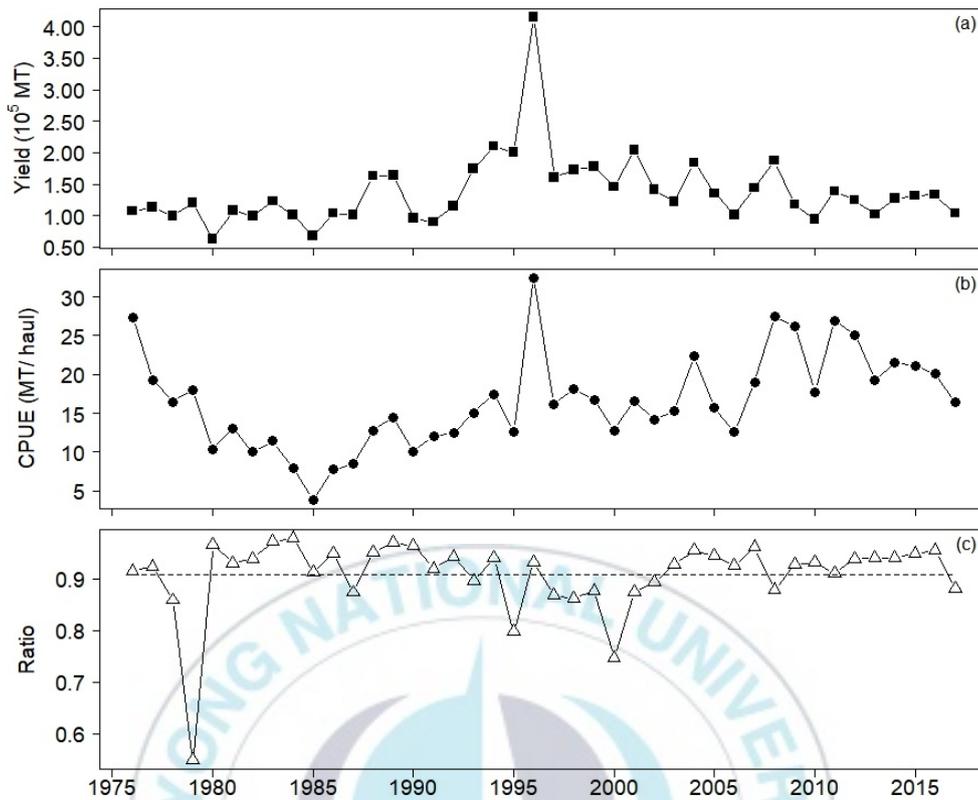


Figure 1. Chub mackerel data from 1976 to 2017. Panel (a) shows the annual yield collected from entire fisheries on mackerel, and panel (b) shows the catch per unit effort (CPUE) data. Panel (c) presents the percentage of the yield achieved by large purse seine fisheries to total yield on chub mackerel. Units for the yield and CPUE are metric ton (MT) and metric ton per haul (MT/haul).

2.2 Surplus production model

In stock assessment, mathematical models involve that describe the dynamic system and observation which linked to the population. A surplus production model depicts the biomass dynamics with three terms – biomass in the previous year, surplus production and catch from commercial fisheries.

$$B_{t+1} = B_t + f(B_t) - Y_t \quad (1)$$

In equation (1), B_t is biomass in year t , Y_t is fisheries yield in year t . $f(B_t)$ represents the ‘surplus production’ and contains the entire changes in natural increase and decrease, such as growth in weight, recruitment, and natural mortality (Hilborn and Walters, 1992). Equation (1) says that the stock size in the next year ($t+1$) is the sum of the biomass and surplus production, which are removed with yield in the previous year (t). Among the various versions of surplus production models, I chose the discrete version of the Schaefer model (Schaefer, 1954) proposed by Hilborn and Walters (1992):

$$B_{t+1} = B_t + rB_t \left(1 - B_t / k\right) - Y_t \quad (2)$$

where r is intrinsic growth rate of a population and k is carrying capacity. The biological assumption lying behind this model is that the stock increase by the growth rate until the stock size reaches to the carrying capacity. From the logistic relationship (Figure 2a) described by equation (2), surplus production is maximized at $B = k / 2$, and called B_{MSY} under absent of fishing (Figure 2b). MSY and harvest rate which correspond to MSY (H_{MSY}) are calculated as $rk / 4$ and $r / 2$ by plugging $B = k / 2$ into equation (2) under absence of fishing.

The Schaefer model has been criticized because the surplus production is always maximized when the biomass is 50% of the carrying capacity (i.e., $B = k / 2$). To improve flexibility, Pella and Tomlinson (1969) introduced a shape parameter that allows the production curve to be asymmetric (Quinn and Deriso, 1999).

$$f(B_t) = \frac{r}{\phi} B_t \left(1 - \left(\frac{B_t}{k}\right)^\phi\right) \quad (3)$$

Despite such advantage, the model should estimate one more parameter (the shape parameter ϕ) than the Schaefer model. In addition, Hilborn and Walters (1992) pointed out that few data sets would allow modelers to obtain reliable estimate of the shape parameter. Therefore, I chose the Schaefer production model in this study. Equation (2) is called deterministic process equation because it describes biomass dynamics system without statistical errors.

Likewise, deterministic observation equation depicts the measurement of biomass, which links the data to biomass as follows:

$$I_t = qB_t \quad (4)$$

where I_t represents the CPUE collected in year t , and q is catchability coefficient. If survey data is present, I_t can be survey index. Equation (4) demonstrates that the CPUE collected in year t is directly related to biomass with a coefficient q . Since the biomass in the model may be subjected to autocorrelation (Millar and Meyer, 1999, 2000), which reduces

the efficiency of parameter estimation, I used relative biomass P_t by scaling the biomass B_t with carrying capacity k ($P_t = B_t / k$). The resulting equations are

$$P_{t+1} = P_t + rP_t(1 - P_t) - Y_t / k \quad (5)$$

and

$$I_t = qkP_t \quad (6)$$



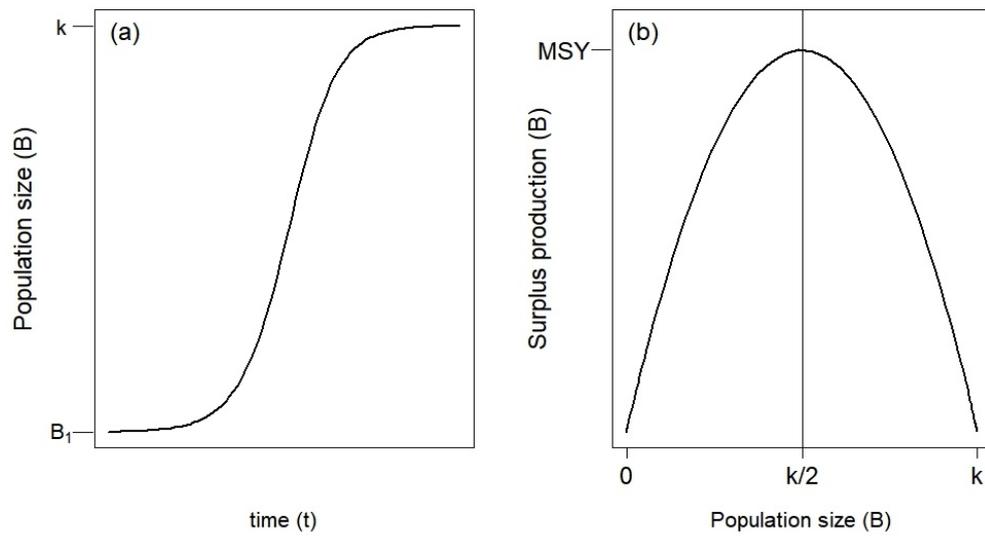
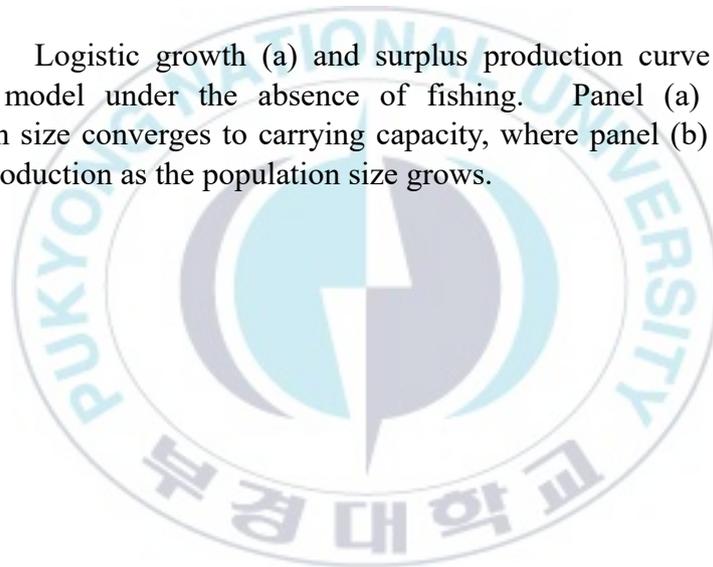


Figure 2. Logistic growth (a) and surplus production curve (b) of the Schaefer model under the absence of fishing. Panel (a) shows the population size converges to carrying capacity, where panel (b) depicts the surplus production as the population size grows.



2.3 A state-space production model

Assuming multiplicative errors on equations (5) and (6), a state-space production model is formulated as below.

$$P_{t+1} = [P_t + rP_t(1 - P_t) - Y_t / k] \exp(\varepsilon_t^p) \quad (7)$$

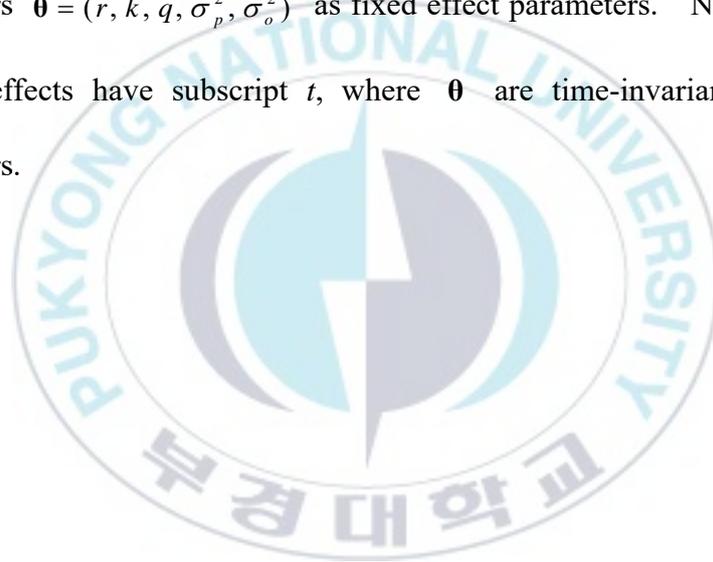
$$I_t = qkP_t \exp(\varepsilon_t^o) \quad (8)$$

In equations (7) and (8), ε_t^p is process error and ε_t^o is observation error. The errors are assumed to follow normal distributions with the mean of 0 and variance of σ_p^2 and σ_o^2 respectively, i.e.,

$$\begin{aligned} \varepsilon_t^p &\sim N(0, \sigma_p^2) \\ \varepsilon_t^o &\sim N(0, \sigma_o^2) \end{aligned} \quad (9)$$

With those errors included, equations (7) and (8) are called process equation and observation equation. Process error in equation (7) accounts the unmodelled natural variability which influences the biomass dynamics.

Punt (2003) mentioned that natural variabilities arise from the growth of the population, recruitment, and natural mortality for production models. The process error allows the biomass to be treated as a random variable. On the other hand, the observation error describes variability in CPUE data including measurement errors and reporting errors (Winker et al., 2018). To categorize the different types of parameters, I denoted the relative biomass $\mathbf{P} = (P_{1976}, P_{1977}, \dots, P_{2018})$ as random effect parameters or state variables, and parameters $\boldsymbol{\theta} = (r, k, q, \sigma_p^2, \sigma_o^2)$ as fixed effect parameters. Note that the random effects have subscript t , where $\boldsymbol{\theta}$ are time-invariant constant parameters.



2.4 Likelihood function

In statistics, a likelihood function describes the likelihood of parameters being true parameters for the given data. Parameters are estimated by maximizing the function.

The $\log P_{t+1}$ is normally distributed with the mean of $\log[P_t + rP_t(1 - P_t) - Y_t/k]$ and variance of σ_p^2 . That is,

$$\log P_{t+1} \sim N\left(\log[P_t + rP_t(1 - P_t) - Y_t/k], \sigma_p^2\right) \quad (10)$$

Likewise, $\log I_t$ is normally distributed with the mean of $\log[qkP_t]$ and variance of σ_o^2 .

$$\log I_t \sim N\left(\log[qkP_t], \sigma_o^2\right) \quad (11)$$

Assuming mutual independence, the likelihood functions for process and observation equations have the following forms.

$$L(r, k, \sigma_p^2, \mathbf{P} | \mathbf{Y}) = \prod_{1976}^{2017} \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left(-\frac{\left(\log P_{t+1} - \left[\log\{P_t + rP_t(1-P_t) - Y_t/k\}\right]\right)^2}{2\sigma_p^2}\right) \quad (12)$$

$$L(q, k, \sigma_o^2, \mathbf{P} | \mathbf{I}) = \prod_{1976}^{2017} \frac{1}{\sqrt{2\pi}\sigma_o} \exp\left(-\frac{\{\log I_t - \log[qkP_t]\}^2}{2\sigma_o^2}\right) \quad (13)$$

Note that $\mathbf{Y} = (Y_{1976}, Y_{1977}, \dots, Y_{2017})$ and $\mathbf{I} = (I_{1976}, I_{1977}, \dots, I_{2017})$.

With the likelihood functions (12) and (13), the joint likelihood function have the following form, given the data $\mathbf{D} = (\mathbf{Y}, \mathbf{I})$.

$$L(r, k, q, \sigma_p^2, \sigma_o^2, \mathbf{P} | \mathbf{D}) = L(r, k, \sigma_p^2, \mathbf{P} | \mathbf{Y}) \cdot L(q, k, \sigma_o^2, \mathbf{P} | \mathbf{I}) \quad (14)$$

2.5 Prior distributions

Within the script software ADMB-RE, the state-space production model estimates both fixed and random effect parameters using the empirical Bayes method (Skaug and Fournier, 2017). Under the approach, fixed effects are estimated by maximum likelihood estimation, and random effects are provided by modes of the posterior distribution (Vincenzi et al., 2014) using automatic differentiation. First, the marginal likelihood function for $\boldsymbol{\theta} = (r, k, q, \sigma_p^2, \sigma_o^2)$ is provided by the Laplace approximation.

$$\int \log L(\boldsymbol{\theta}, \mathbf{P} | \mathbf{D}) d\mathbf{P} = \log L(\boldsymbol{\theta} | \mathbf{D}) \quad (15)$$

In other words, random effects are separated from the joint likelihood through the approximation.

Then the fixed effects are obtained from the maximum likelihood estimation.

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \log L(\boldsymbol{\theta} | \mathbf{D}) \quad (16)$$

With the point estimates of fixed effects, estimates of random effects are provided by the mode of posterior distributions of \mathbf{P} .

$$\hat{\mathbf{P}} = \text{mode} \left[p(\mathbf{P} | \mathbf{D}, \hat{\boldsymbol{\theta}}) \right] \quad (17)$$

And the uncertainties on fixed effects $\boldsymbol{\theta}$ and random effects \mathbf{P} are given as below, respectively.

$$\text{cov}(\boldsymbol{\theta}) = \left[-\frac{\partial^2 \log L(\boldsymbol{\theta} | \mathbf{D})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]^{-1} \quad (18)$$

$$\text{cov}(\mathbf{P}) = \left[-\frac{\partial^2 \log L(\mathbf{P} | \hat{\boldsymbol{\theta}}, \mathbf{D})}{\partial \mathbf{P} \partial \mathbf{P}'} \right]^{-1} + \frac{\partial \mathbf{P}}{\partial \boldsymbol{\theta}} \text{cov}(\boldsymbol{\theta}) \left(\frac{\partial \mathbf{P}}{\partial \boldsymbol{\theta}} \right)' \quad (19)$$

While the classical statistics treat parameters as unknown constant values, Bayesian approach considers each parameter's probability distribution. The probabilities of parameters are updated as data obtained. Through the Bayes rule described by equation (20), the probabilities for

parameters are updated by the prior distribution and the likelihood built on data.

$$p(\mathbf{P}, \boldsymbol{\theta} | \mathbf{D}) \propto p(\mathbf{P}, \boldsymbol{\theta})p(\mathbf{D} | \mathbf{P}, \boldsymbol{\theta}) \quad (20)$$

Note that the joint posterior distribution $p(\mathbf{P}, \boldsymbol{\theta} | \mathbf{D})$ is proportional to the product of joint prior $p(\mathbf{P}, \boldsymbol{\theta})$ and joint likelihood $p(\mathbf{D} | \mathbf{P}, \boldsymbol{\theta})$.

While Bayesian approach does not involve the numerical optimization of the likelihood function, the numerical approach did not stabilize without priors. Therefore, I considered prior distributions which aid numerical optimization. Without previous studies on Korean mackerel, I specified prior distributions for parameters $\boldsymbol{\theta} = (r, k, q, \sigma_p^2, \sigma_o^2)$, and the relative biomass in 1976, P_{1976} . Specifically, I randomly generated values for mode and coefficient of variation (CV) for each parameter which determine the hyperparameters. Once the optimal prior set stabilizes the optimization, I switched to Bayesian approach, which involves Markov Chain Monte Carlo sampling as described below. In practice, I chose log-normal distributions for r and k , whose domains of the distributions are positive. I considered

inverse gamma distributions for the variances of errors, σ_p^2 and σ_o^2 , to assign prior distributions with positive support. These prior distributions are used in various papers which involved Bayesian stock assessment (Chaloupka and Balazs, 2007; Meyer and Millar, 1999; Millar and Meyer, 2000; Winker et al., 2018). Since the catchability coefficient is a scaling factor, which range should be positive and smaller than 1, I assigned a uniform prior for $\log q$ (McAllister et al., 1994; Millar and Meyer, 2000). I assigned a normal distribution for $\log P_{1976}$ which is equivalent to log-normal distribution with the same hyper parameters. I provided the details about the relationship between the log-normal distribution and normal distribution in appendix A, by re-expressing the mean and variance of a normally distributed random variable, $\log X$, using the mode and CV of a log-normally distributed random variable, X . I also provided the ADMB-RE code in appendix B. Table 1 lists the selected prior set with modes and CVs.

Assuming the mutual independence of priors, joint prior probability can be written as below.

$$\pi(r, k, q, \sigma_p^2, \sigma_o^2, P_{1976}) = \pi(r)\pi(k)\pi(q)\pi(\sigma_p^2)\pi(\sigma_o^2)\pi(P_{1976}) \quad (21)$$

With the joint likelihood and joint prior, the joint posterior is defined by

Bayes' rule:

$$p(r, k, q, \sigma_p^2, \sigma_o^2, \mathbf{P} | \mathbf{D}) \propto \pi(r, k, q, \sigma_p^2, \sigma_o^2, P_{1976}) L(r, k, q, \sigma_p^2, \sigma_o^2, \mathbf{P} | \mathbf{D}) \quad (22)$$



Table 1. Selected priors set which satisfies the numerical optimization of the likelihood function. Values in the parentheses are parameters for each prior distribution.

Parameter	Prior	Mode	CV
r	Log-normal (-1.04, 0.46)	0.28	0.49
k	Log-normal (14.92, 0.59)	2,137,000	0.64
q	Uniform (-90, -1) on $\log q$	Noninformative	
σ_p^2	Inverse gamma (3.09, 0.80)	0.20	0.96
σ_o^2	Inverse gamma (3.52, 0.43)	0.09	0.81
P_{1976}	Normal (-0.57, 0.93) on $\log P_{1976}$	0.22	1.24

2.6 MCMC sampling

Since full posterior distributions require high-dimensional integration, they cannot be obtained in closed form when the model contains numbers of parameters. Instead, Markov Chain Monte Carlo (MCMC) sample sets were generated to get (approximated) posterior distributions in the software ADMB-RE. To be specific, ADMB-RE samples within the Metropolis Hastings Algorithm. One in every 50,000 samples set was thinned out from a total of 0.2 billion iterations to reduce the autocorrelation between parameter samples. The first 100 sample sets (i.e., initial 5,000,000 sample sets) were removed as a burn-in period, ultimately resulting in 3,900 sample sets remained. Once the posterior samples obtained, diagnosis of convergence of MCMC samples is required. In this study, four criteria were applied to check the convergence of the MCMC samples for each parameter $(r, k, q, \sigma_p^2, \sigma_o^2, P_{1976})$: (i) the dependence factor of the Raftery-Lewis statistics, (ii) lag-1 correlation, (iii) the ratio between the naïve standard error and the time series standard error, which is corrected with autocorrelation, and (iv) unimodal shape of histogram of MCMC samples. I checked the first three criteria within R using package CODA (Plummer et al., 2006).

The posterior samples were said to be converged when the dependence factor of the Raftery-Lewis statistics is smaller than 5, lag-1 correlation is close to 0, the ratio of the naïve standard error to the time series standard error is around 1, as well as the shape of the posterior histogram has unimodal shape. Summaries on the posterior distributions gave modes as point estimates, and uncertainties as 95% credible intervals for each parameter.



2.7 Various production models

I chose several production models to compare and evaluate the performance of the state-space production model; the selected models included Fox production model (Fox, 1970), Yoshimoto-Clarke model (Clarke et al., 1992; Yoshimoto and Clarke, 1993), Schnute regression model (Schnute, 1977), ASPIC program (Prager, 2016, 1994) and Observation model (equation (5) and (8)). The Fox model, commonly referred to as the exponential (production) model, assumes an exponential relationship between fishing effort and population size (Fox, 1970). Since the model regards the surplus production as yield, the stock is in equilibrium (i.e., $B_{t+1} = B_t$). Fox (1970) provided the following regression model to estimate parameters r and q with annual CPUE and effort.

$$\log I_t = \log I_\infty - \frac{q}{r} E_t \quad (23)$$

where the I_∞ is catch per unit effort proportional to carrying capacity, k , and E_t is effort in year t .

However, Hilborn and Walters (1992) criticized that the equilibrium conditions lead to overestimate the stock size. Thus, they warned their readers never to use the equilibrium methods.

Clarke et al. (1992) and Yoshimoto and Clarke (1993) modified the Fox model under a non-equilibrium condition. I denote the model as the YC model which is given as below:

$$\log I_t = \frac{2r}{2+r} \log qk + \frac{2-r}{2+r} \log I_{t-1} - \frac{q}{2+r} (E_{t-1} + E_t) \quad (24)$$

Yoshimoto and Clarke (1993) showed that their model predicted CPUE data even with negative estimates of q and k .

Schnute (1977) transformed the Schaefer model into a linear regression model which allowed one to use explanatory variables with geometric means of CPUE and effort data. The Schnute model is given as follows.

$$\log \left(\frac{I_t}{I_{t-1}} \right) = r - q \left(\frac{E_t + E_{t-1}}{2} \right) + \frac{r}{qk} \left(\frac{I_t + I_{t-1}}{2} \right) \quad (25)$$

I chose the former three models to compare, as they have been used frequently in recent publications and technical reports (Cho et al., 2009; Jeong and Nam, 2017; Kim et al., 2018; Kwon et al., 2013).

The ASPIC program is developed by Prager (1994) and has been included in the NOAA toolbox. Among the various modes available for fitting the data, I chose the Schaefer model with least squares method to provide results without process and observation error. Whereas the former three models estimate r , k , and q as free parameters, the ASPIC program estimates four free parameters, r , k , q and B_{1976} .

$$\frac{dB_t}{dt} = (r - B_t / k)B_t - F_t B_t \quad (26)$$

Equation (26) says that the rate of change of the biomass is determined with the surplus production and fishing. Here, F_t is the fishing mortality rate. Since the ASPIC program fits the differential equation, units for the r and q are year⁻¹ and year⁻¹ haul⁻¹, respectively.

Without process error, the Observation model depicts the biomass dynamics within a deterministic system and variability in measurement. In

the Observation model, the relative biomass from 1977 onward ($P_{1977}, P_{1978}, \dots, P_{2017}$) treated as derived parameters, which do not require the Laplace approximation. Therefore, the estimation of the model parameters was performed within ADMB through maximum likelihood estimation instead of ADMB-RE.



3. Results

3.1 State-space production model results

I regarded the MCMC samples set built the posterior distributions with diagnostics which showed the MCMC samples for each parameter ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1976}$) converged (Table 2, Figure 3). The point estimates (posterior modes) of intrinsic growth rate r and carrying capacity k were 0.30 and 2,143,122MT, which the value was about five times larger than the largest yield achieved in 1996 (415,003MT, Figure 1). For the catchability coefficient, the point estimate was $5.10 \times 10^{-6} \text{ haul}^{-1}$. Estimates of the variance of process error and observation error were 0.09 and 0.05, respectively. I calculated the CVs for process error and observation error as follow:

$$CV_{\text{process}} = \frac{\sigma_p}{E(\log P_t)} = \frac{\sqrt{\sigma_p^2}}{\frac{1}{42} \sum_{t=1976}^{2017} \log P_t} \quad (27)$$

$$CV_{\text{observation}} = \frac{\sigma_o}{E(\log I_t)} = \frac{\sqrt{\sigma_o^2}}{\frac{1}{42} \sum_{t=1976}^{2017} \log I_t} \quad (28)$$

The CV for the process error was 11 times larger than that of observation error, calculated at 89% and 8%, respectively. The relative biomass in 1976 (P_{1976}) was 1.12, which shows the initial biomass was about 10% larger than the estimated carrying capacity. Management references MSY, B_{MSY} and H_{MSY} were 174,298MT, 1,071,152MT, and 0.15, respectively. Table 3 lists the 95% credible intervals for each parameter.

I also compared the prior and posterior distributions for ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1976}$) in Figure 3. All posterior distributions were skewed to the right, and the posterior distributions of r and k showed similar to the shape of their priors. The posterior distributions for variances of process and observation errors updated into narrower distributions compared to their priors (Figure 3d, 3e). While the modes of the variances of process error and observation error estimated smaller than their modes of priors, the relative biomass in 1976, P_{1976} , predicted a larger value than the mode of its prior (Figure 3f).

Figure 4 shows the joint posteriors of two parameters with scatter plots. The notable negative relationship between k and q is shown as a banana shape. When scaling the biomass in the model with carrying capacity, it results in the relative biomass P_t having no dimension. Because the CPUE in this study have the unit of MT/haul, carrying capacity and catchability coefficient have a negative relationship. For example, if the model predicts a same value of I_t , two possible values come from larger k and smaller q or smaller k and larger q .

The state-space production (hereafter SSP) model fitted the CPUE data with the *mixed trend* by following the upward and downward phase (Figure 5a). Since there is no rigid criterion on fitting CPUE data which can determine to be a ‘good fit’ or a ‘bad fit’, I only presented the distance (i.e., $\sqrt{(I_t - \hat{I}_t)^2}$) between the predicted values and the data. The model showed the largest distance (9.21MT/haul) with the CPUE data (22.23MT/haul) in 1996 (i.e., $\sqrt{(I_{1996} - \hat{I}_{1996})^2}$), while the others remained within 5MT/haul. Estimated MSY from the model suggests that the stock was overexploited in the years of 1993-1996, 1999, 2001, 2004 and 2008 (Figure 5b). Predicted annual biomass showed fluctuations in the range of 4×10^5 MT to 2.4×10^6

MT, and was 1.5×10^6 MT in average (Figure 5c). Because the stock did not reach a steady state during the entire period, the population size larger than the carrying capacity can be regarded as temporary variations in 1976, 1996, 2008, 2009, 2011 and 2012. The result argued that mackerel stock had a smaller size than B_{MSY} in 1982-1988 and 1990. Only four years achieved a larger harvest rate than H_{MSY} (i.e., 1986, 1988, 1995, 1996) and has declined since 1996 (Figure 5d).



Table 2. Diagnostics for Markov Chain Monte Carlo samples for parameters ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1976}$): Dependence factor of Raftery-Lewis statistics (DF), lag-1 autocorrelation, the ratio of the naïve standard error to the time series standard error, and the shape of posterior distributions were checked.

Parameters	DF	Lag-1 autocorrelation	Naïve / Time series	Posterior shape
r	1.04	-0.02	1.08	Unimodal
k	1.02	-0.03	1.04	Unimodal
q	1.00	0.02	1.00	Unimodal
σ_p^2	1.04	-0.05	1.05	Unimodal
σ_o^2	1.04	0.01	1.00	Unimodal
P_{1976}	1.15	0.01	1.00	Unimodal

Table 3. Posterior summary for each parameter ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1976}$) and management references. Note that the units for k, MSY, B_{MSY} are metric ton (MT), q and σ_o^2 are 1/haul and log[MT/haul], respectively.

Parameters	Summary			
	Mode	2.5%	50%	97.5%
r	0.30	0.12	0.32	0.57
k	2,143,122	950,266	2,770,000	6,408,487
q	5.05×10^{-6}	1.13×10^{-6}	7.60×10^{-6}	1.88×10^{-5}
σ_p^2	0.09	0.05	0.09	0.15
σ_o^2	0.05	0.03	0.05	0.09
P_{1976}	1.12	0.43	1.20	2.13
MSY	174,298	55,227	212,762	561,277
B_{MSY}	1,071,152	473,829	1,380,000	3,207,033
H_{MSY}	0.15	0.06	0.16	0.29

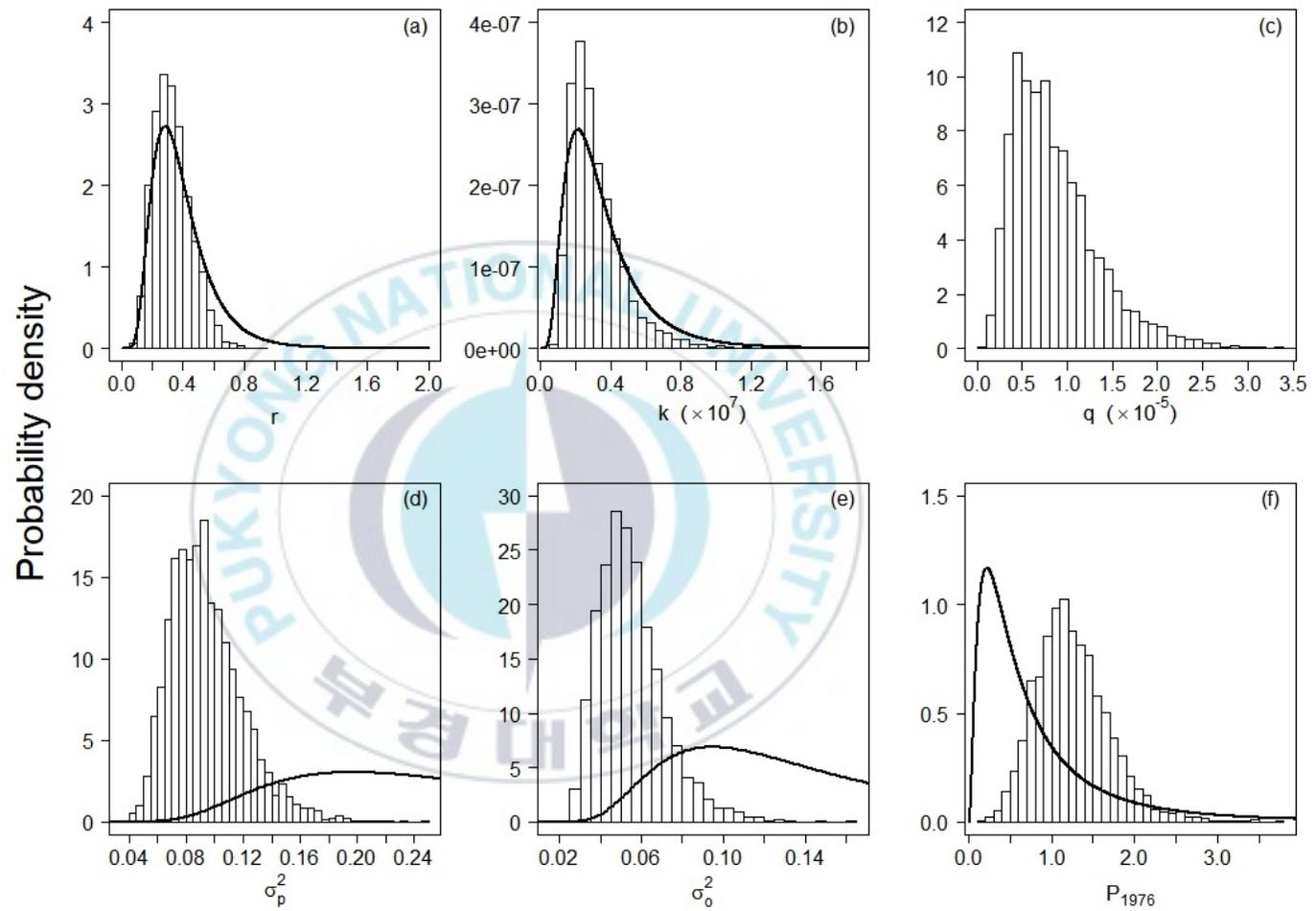


Figure 3. Posterior densities obtained with the specified priors. Bold lines represent the priors and histograms

show the Markov Chain Monte Carlo samples from the posterior distributions in each panel. Units for k , q and σ_o^2 are metric ton (MT), 1/haul and $\log[\text{MT/haul}]$, respectively. Note that a uniform prior for $\log q$ is considered while it is not shown in panel (c).



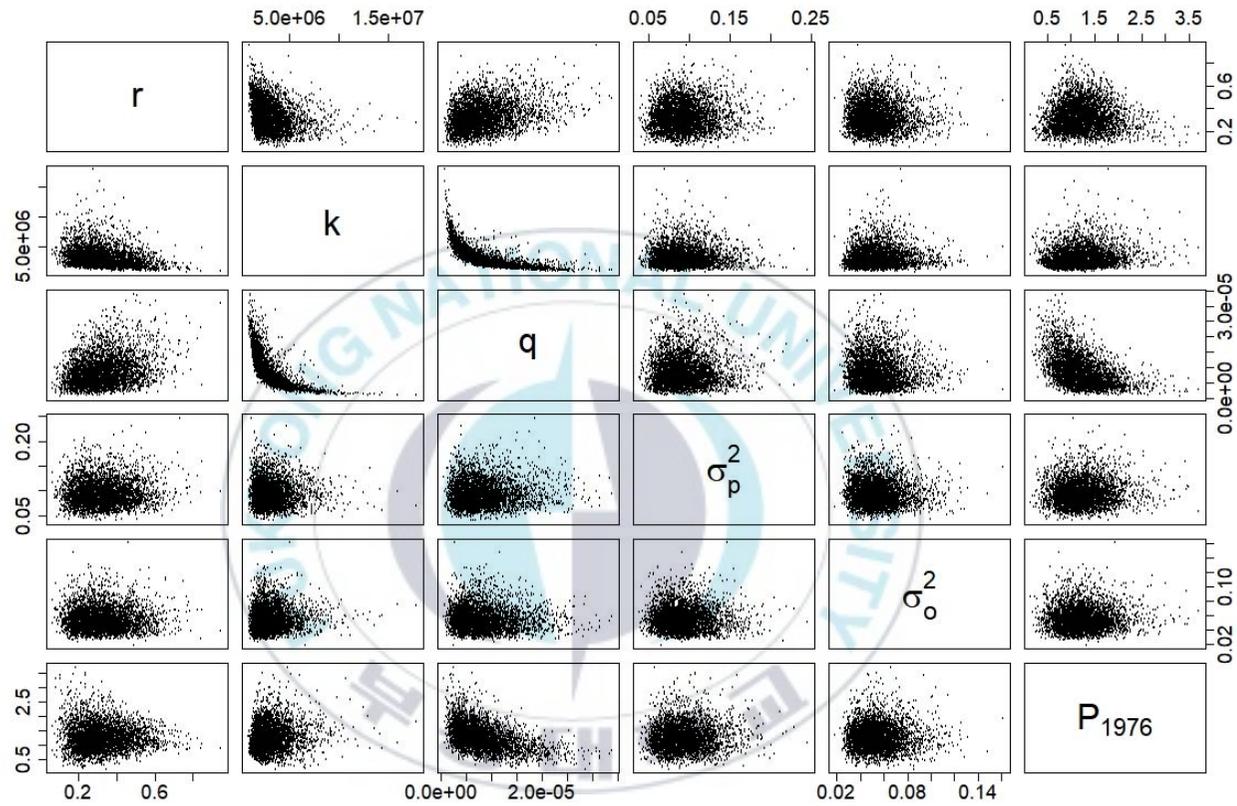


Figure 4. Scatter plots of joint posteriors of two parameters. Units for k , q and σ_o^2 are metric ton (MT), 1/haul and log[MT/haul], respectively.

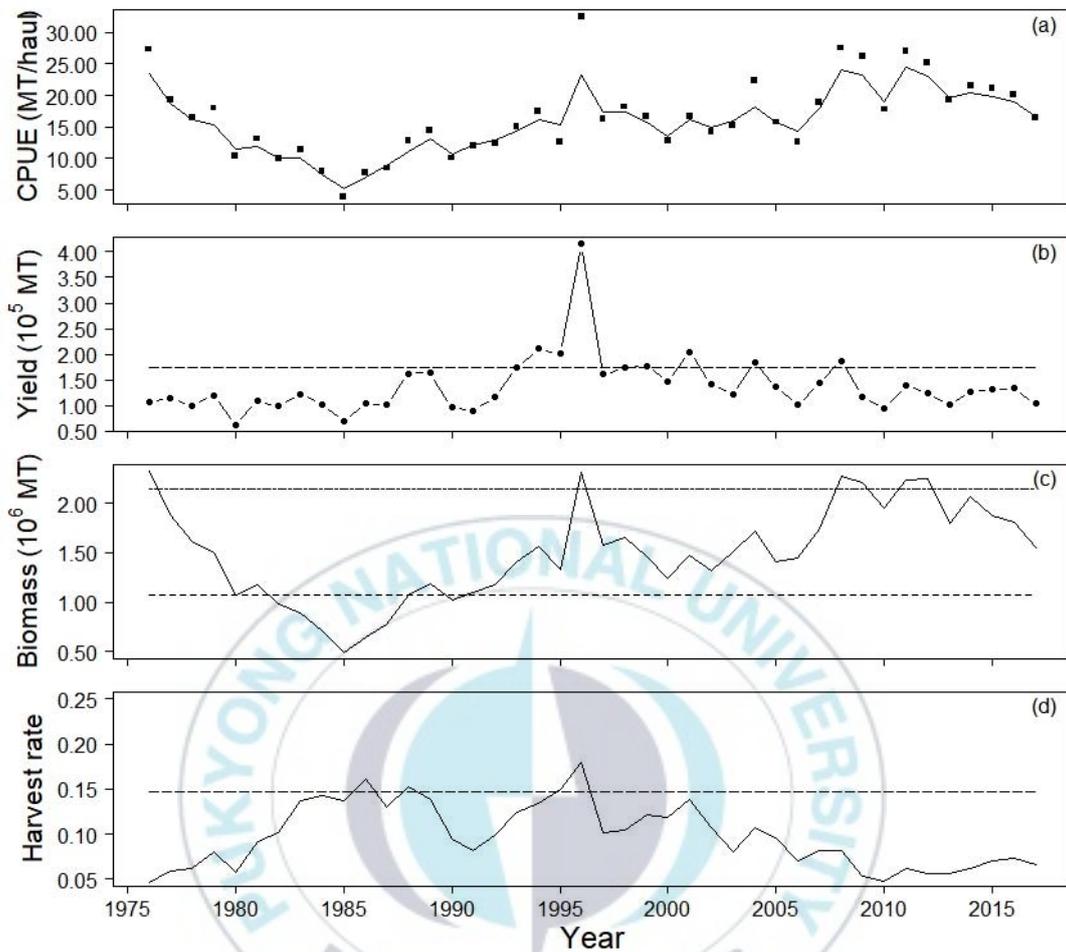


Figure 5. Results from the state-space production model including predicted CPUE, annual biomass, annual harvest rate and management references. CPUE data (filled squares) is compared with the predicted CPUE (solid line) in panel (a). Panel (b) shows the Maximum sustainable yield (MSY, dashed line) with annual yield (solid line with filled circles). Panel (c) depicts the predicted annual biomass (solid line) with carrying capacity (two-dashed line) and biomass that yields the MSY (B_{MSY} , dashed line). Panel (d) shows the annual harvest rate (solid line) with the harvest rate that corresponds to MSY (H_{MSY} , dashed line).

3.2 Comparison with various production models

Applying the various production models, each model provided parameter estimates. Table 4 lists the estimates. The Fox model, YC model, and Schnute model estimated three parameters r , k , q , whereas the ASPIC program estimated an additional parameter, the relative biomass in 1976, P_{1976} . Since the Observation model considers the observation error, the model estimates one more parameter, σ_o^2 , than the ASPIC program. The Observation model provided 95% confidence intervals for parameter estimates. The Fox model and YC model estimated the intrinsic growth rate about 0.7 which values are larger than that of the SSP model. However, the Schnute model provided negative value on growth rate (-0.37), which implies the stock decreasing. ASPIC program showed that the stock grew by 0.02 per year and the Observation model estimated the growth rate as 0.46. The Fox model and Schnute model estimated carrying capacity as -556,659MT and -1,913,152MT, respectively. The YC model estimated the smallest carrying capacity (585,284MT). The Observation model and ASPIC program estimated the maximum population size to be greater than that of SSP model with 3.18×10^{13} and 4.86×10^{12} in MT, respectively. Regarding the average yield was 1.40×10^5 MT in the entire duration (1976-2017), those two models suggest the stock was underexploited. The ASPIC program also estimated that the biomass

level was low (0.1%) in 1976 compared to the carrying capacity, while the observation model estimated the stock size in 1976 was greater than its carrying capacity (177%). The Fox model and Schnute model gave negative estimates for carrying capacity and catchability coefficient. The YC model estimated the catchability coefficient which value is eight times larger than the estimated catchability coefficient from SSP model. The observation model estimated the catchability coefficient with the smallest value, 4.73×10^{-13} in haul^{-1} , where ASPIC program estimated $2.09 \times 10^{-9} \text{ year}^{-1} \text{ haul}^{-1}$. The ASPIC program and Schnute model estimated the relative biomass in 1976 smaller than 1, whereas the Fox and Schnute model provided negative values. Observation error variance estimated from the Observation model was 0.15, three times larger than that of the SSP model. Since the Observation model provided the lower 95% confidence limits for k and q with negative values (i.e., -1.66×10^{17} MT for k and $-2.47 \times 10^{-9} \text{ haul}^{-1}$ for q), which should have positive domains, only positive ranges of the 95% confidence intervals for k and q shown in Table 4. I excluded the Fox and Schnute model in comparison of the predicted CPUE and biomass with other models, for producing unreliable estimates such as negative carrying capacity and catchability coefficient.

I showed the predicted CPUE and biomass from the four different models in

Figure 6. While the YC model moderately described the CPUE data (Figure 6a), it predicted the biomass in 1996 as 2.63×10^5 MT and failed to account the largest yield (Figure 6e). The predicted CPUE from the ASPIC program showed a monotonic increase in years of 1976-2017, resulting in the stock size increase during the period (Figure 6b, 6f). On the other hand, the Observation model predicted the declining CPUE which is similar to the data for initial years (i.e., 1976-1980), but remained the same from 1986 onward (15.8MT/haul) (Figure 6c). In consequence, the predicted biomass graph showed a flat line from 1982 (Figure 6g). In addition to the flat graphs, 95% confidence intervals did not include the CPUE data, where the 95% confidence intervals for annual biomass was too wide to be shown with the predicted biomass. The SSP model predicted the CPUE by picking up the fluctuation in the CPUE data (Figure 6d) and included the data within the 95% credible intervals. In contrast to the YC model, which failed to explain the yield in 1996, the SSP model predicted the annual stock size to be larger than the annual yield from 1976 to 2017. Not surprisingly, the goodness-of-fit statistics (i.e., $\sum (I_t - \hat{I}_t)^2$) suggested that the SSP model outperformed in fitting the CPUE data (YC: 1209, ASPIC: 1221, observation: 1459, SSP: 191).

Table 4. Parameter estimates obtained from various production models - Fox model (Fox), Yoshimoto-Clarke model (YC), Schnute model (Schnute), ASPIC program (ASPIC), Observation model (Observation) which the process error is not considered, and the state-space production model (SSP). NAs in the table indicate that the model does not estimate the corresponding parameter(s). Except for the ASPIC results, Units for k , q and σ_o^2 are metric ton (MT), 1/haul and log[MT/haul] respectively. For ASPIC, the units for r and q are year^{-1} and $\text{year}^{-1}\text{haul}^{-1}$, respectively. The 95% confidence intervals and 95% credible intervals for estimates from the Observation and SSP are provided in parentheses, respectively. Only positive ranges were showed for the 95% confidence intervals for k and q of the Observation model results.

Parameters	Fox	YC	Schnute	ASPIC	Observation	SSP
r	0.71	0.74	-0.37	0.02	0.46 (0.06, 0.87)	0.30 (0.12, 0.57)
k	-556,659	585,284	-1,913,152	4.86×10^{12}	3.18×10^{13} (0, 1.66×10^{17})	2,143,122 (950 266, 6 408 487)
q	-5.77×10^{-5}	4.39×10^{-5}	-2.55×10^{-5}	2.09×10^{-9}	4.73×10^{-13} (0, 2.47×10^{-9})	5.10×10^{-6} (1.13×10^{-6} , 1.88×10^{-5})

σ_o^2	NA	NA	NA	NA	0.15 (0.09, 0.22)	0.05 (0.03, 0.09)
σ_p^2	NA	NA	NA	NA	NA	0.09 (0.05, 0.15)
P_{1976}	1.37	1.06	0.55	0.001	1.77 (0.47, 3.10)	1.12 (0.40, 2.10)



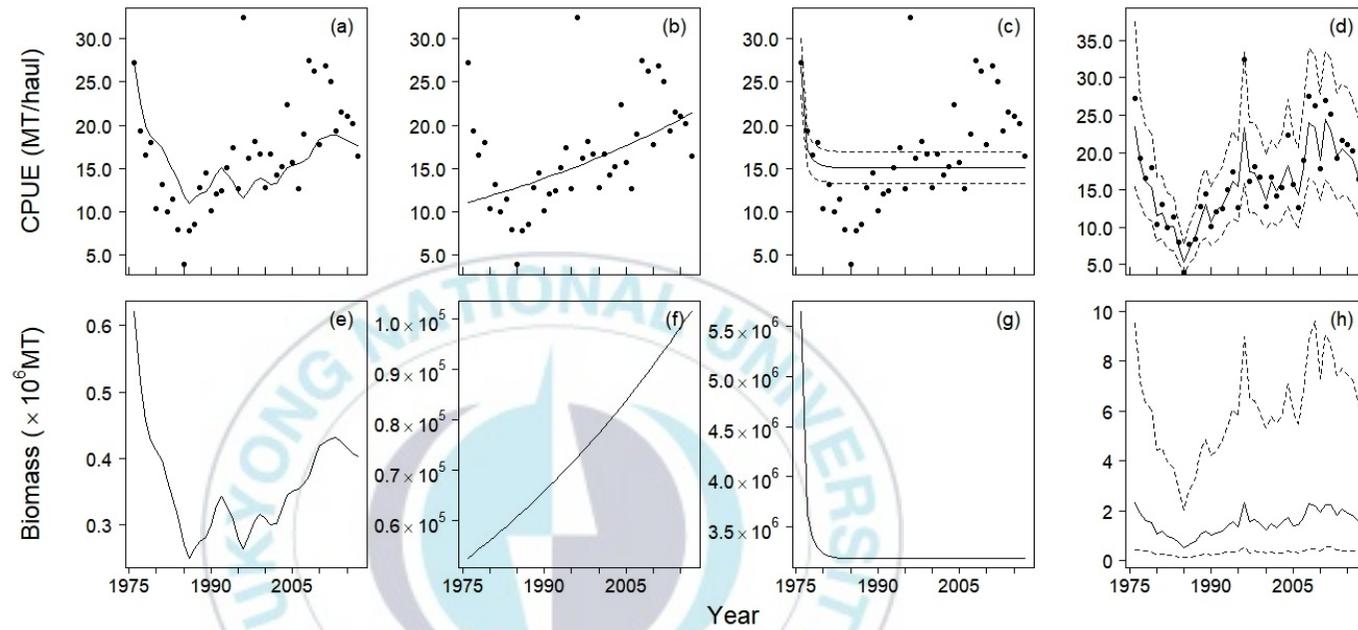


Figure 6. Predicted CPUE and biomass from the four different production models: Clarke-Yoshimoto production model, ASPIC program, Observation model and the state-space production model. Plots in the first row show the predicted CPUE whereas plots in the second row show the predicted biomass. Unit for the CPUE is metric ton per haul (MT/haul) and biomass for 10^6 metric tons (MT).

4. Discussion

4.1 Estimates from various production models

Because of the simplicity concerning data requirement and model structures, surplus production models have been applied until recently. However, the results presented in Table 4 suggest that the simplicity does not guarantee the reliable estimates of model parameters and biomass, especially when the relative biomass index data shows a *mixed trend*. For example, the Fox and Schnute model estimated negative values for carrying capacity and catchability coefficient which values should be positive. While the YC model moderately fitted the CPUE data, it failed to estimate stock size larger than the yield in 1996. The Observation model provided the uncertainties for the predicted CPUE as the SSP model had. However, the intervals of the model were too narrow to include the CPUE data (Figure 6c), which, in contrast, the SSP model could (Figure 6d). This may be the case the Observation model considers the observation errors only, whereas the SSP model incorporated both process and observation errors, resulting in the wider credible intervals for the predicted CPUE. The ASPIC program and Observation model estimated the growth rate and carrying capacity as positive values, those two models suggest the optimum yield level with 3.89×10^9 MT and 1.94×10^{10} MT respectively, which imply the stock was severely underexploited throughout the

whole duration. In other words, the ASPIC program and Observation model may have an optimistic view of mackerel stock, which allows fishermen to exploit the stock more than the present level. This is because the production models which were compared with the SSP model cannot account the fluctuations of CPUE data especially in the mid-90s. In short, a state-space production model can be a solution for CPUE data with a *mixed trend*.

Hilborn and Walters (1992) coined the term ‘data failures’ to explain that data containing insufficient information on model parameters, leading one to have unreliable estimates of carrying capacity and catchability coefficient (Table 4). Although the Observation model provided variances for point estimates, while the regression models and ASPIC program did not, the 95% confidence intervals of k and q (Table 4) included negative support. Since the curvature of a likelihood function is associated with the amount of information in the data (Pawitan, 2001), the high variance of the maximum likelihood estimates (i.e., \hat{k} and \hat{q}) imply the data is not very informative in estimating k and q . Moreover, posterior results from the SSP model support this point. While the posteriors of P_{1976} , q , σ_p^2 , and σ_o^2 updated from the prior and likelihood, the posteriors of r and k remained similar to their priors (Figure 3). Regarding the posterior distribution on a parameter updated with prior and likelihood, the mackerel data did not provide

sufficient information to estimate r and k , resulting in the posterior distributions to reflect their priors. Hilborn and Walters (1992) argued the data for production models are responsible for unreliable estimates. They explained that the biomass dynamics models require data with historical change on stock size and fishing effort to get reliable estimates on parameters including r , k , and q . For example, stock size consistently reduced from carrying capacity as the fishing effort increase at the beginning of fishing. Once the stock size reduced about half of the virgin state, fishing effort decrease, and the stock has a chance to recover. The recovery phases provide information on the intrinsic growth rate, where the depletion phases provide information on the catchability coefficient. In practice, however, collecting data which provide sufficient information for parameters rarely accomplished. Therefore, 'data failure' can be translated into 'model failure', which implies insufficient description of biomass dynamics, and many researchers have mentioned (Carruthers et al., 2011; Millar and Meyer, 1999; Musick and Bonfil, 2005; National Research Council, 1998; Pedersen and Berg, 2017).

4.2 *Implementation of a state-space production model*

While state-space models have been favored for addressing the variability of a system which the model describes and noise in data, the estimation of random effects is computationally demanding. As explained earlier, the state-space production model estimates both fixed effect parameters and random effect parameters. Prager (1994) suggested to provide auxiliary information or assumptions for state-space models. In the same context, prior distributions were considered as additional information for the state-space production model. Without previous studies of population parameters on chub mackerel, which aid in specifying prior distributions on model parameters, I borrowed prior distributions for model parameters from various studies. Since priors address the probability on parameters, biologically absurd prior values were excluded. For example, positive domains were chosen for modes of growth rate r , catchability coefficient q and P_{1976} . Mode for carrying capacity k was only selected when its value was larger than the largest yield (415,003MT). Since the priors were not supported by previous researches, minimum CV for each informative prior was set by 40% to avoid assigning highly informative priors. In addition to the priors, software ADMB-RE performed both numerical optimization and MCMC sampling. While the numerical optimization was an intermediate stage to find optimal priors for the

state-space production model, the software provides Laplace approximation without restrictive assumptions such as fixing the ratio between process and observation error variance or linear structure of models. Once the priors selected, posteriors also can be calculated through the sampling without editing codes. Therefore, I recommend taking advantages of the statistical tool for the implementation of state-space models under both presence or absence of priors.



4.3 Stock assessment using data series from 1999 to 2017

Yet, the state-space production model outperformed various production models in fitting a *mixed trend* data, the results should be regarded as a demonstration, not as a stock assessment for the stock. While the CPUE data have been collected quite prolonged period (about 40 years), scientists in KNIFS conjectured that CPUE data lost consistency for reduction of fishing ground for the Korea-Japan Fisheries Agreement which was signed in 1998. Therefore, I applied the state-space production model to shorter time series data (1999-2017) on mackerel and included the results in this section. I denote the data set as ‘short data’ to distinguish it from the data set mentioned in the method section. Since the numerical optimization was not achieved with the short data, I repeated the same method for implementation of the state-space production model, which described in section 2. Table 5 lists the priors set which stabilized the numerical optimization. Again, the length of 3900 posterior sample sets were obtained from the 0.2 billion iterations with a thinning interval of 50,000, and the initial 500,000 samples were removed as a burn-in period. Table 6 lists the diagnostics of samples, which showed the MCMC samples passed all four criteria. The posterior summaries were provided in Table 7 and the comparison of priors and posteriors for each parameter was provided in Figure 7. The posterior distributions of r , k ,

and P_{1999} reflected the priors, which indicate the amount of information contained in the data was not enough to estimate the three parameters (Figure 7a, 7b, 7f).

The estimates of intrinsic growth rate, carrying capacity, and catchability coefficient, were 0.16 and 1,832,623MT, $5.49 \times 10^{-5} \text{ haul}^{-1}$, respectively. Variances for the process error and the observation error were 0.13 and 0.07, respectively. The CVs of the process and observation error were 84% and 9% respectively, which indicate the relative variability of the process error was 9 times larger than the relative variability of the observation error. The relative biomass in 1999 (P_{1999}) was 0.51, which indicates the stock size in 1999 was roughly half of the carrying capacity. Calculated MSY, B_{MSY} and H_{MSY} were 152,677MT, 910,866MT and 0.08, respectively. 95% credible intervals for the estimates were provided in Table 7, as uncertainties of estimates.

With the short data, the SSP model fitted the CPUE data by following its trend, and the 95% credible intervals included the CPUE data (Figure 8a). Compared to the calculated MSY, annual yield achieved in 1999, 2001, 2004 and 2008 were larger than MSY, where the other years remained below the MSY (Figure 8b). Predicted annual stock size was in a range of 1.0×10^6 MT to 1.8×10^6 MT, and the average biomass during the recent decade was about 1.2×10^6 MT (Figure 8c). Mean biomass during 1999 to 2017 was 1,064,924MT, and the smallest and the

largest biomass were 805,645MT and 1,419,278MT, respectively. Harvest rate, the proportion of yield to biomass, showed decrease during the period (Figure 8d). Annual harvest rate was kept above the calculated H_{MSY} .

According to the National Oceanic and Atmospheric Administration (NOAA), ‘overfishing’ and ‘overfished’ applied when a stock having harvest rate larger than H_{MSY} (e.g., $H_t > H_{MSY}$) and the stock size is smaller than $0.5B_{MSY}$ (e.g., $0.5B_{MSY} > B_t$). Regarding these definitions, I presented the historical trajectories of the annual harvest rate and biomass in a Kobe plot (Figure 9). The plot suggests the stock was subject to overfishing from 1999 to 2017 and that the harvest rate exceeded H_{MSY} . After 1999, the annual harvest rate showed a decrease. On the other hand, the annual biomass was larger than $0.5B_{MSY}$, indicating that the stock was not overfished during 1999-2017. The graph shows a small increase in the ratio between the biomass and B_{MSY} from 1.00 in 1999 to 1.07 in 2017.

Quota management have been applied to mackerel since 1999 for large purse seine fisheries. In 2017, the total allowable catch (TAC) was set at 123,000MT (Korean Ministry of Ocean and Fisheries, 2017), which was 80.56% of the MSY calculated in this study. For fisheries managers, Hilborn (2010) introduced the concept of a ‘Pretty good yield’ defined as 80% of the MSY (Hilborn, 2010).

However, since the method of calculating TAC is not available to the public in Korea, the ‘pretty good yield’ from the result and the TAC cannot be directly compared. Because the minimum legal size for mackerel is 21cm, the predicted annual biomass in this study must be interpreted as the biomass of fish larger than 21cm. The quota in 2017 can therefore be said to be optimal only if the TAC targets stock biomass within the legal size.



Table 5. Selected priors set that satisfied the numerical optimization of the likelihood function. Values in the parentheses are parameters for each prior distribution.

Parameter	Prior	Mode	CV
r	Log-normal (-0.75, 0.86)	0.22	1.05
k	Log-normal (15.14, 0.96)	1,454,000	1.26
q	Uniform (-90, -1) on $\log q$	Noninformative	
σ_p^2	Inverse gamma (2.68, 1.06)	0.29	1.21
σ_o^2	Inverse gamma (4.78, 0.66)	0.11	0.60
P_{1999}	Normal (-1.37, 1.35) on $\log P_{1999}$	0.07	1.69

Table 6. Diagnostics for Markov Chain Monte Carlo samples for parameters ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1999}$): Dependence factor of Raftery-Lewis statistics (DF), lag-1 autocorrelation, the ratio of the naïve standard error to the time series standard error and the shape of posterior distributions were checked.

Parameters	DF	Lag-1 autocorrelation	Naïve / Time series	Posterior shape
r	1.04	0.00	1.03	Unimodal
k	1.02	-0.02	1.00	Unimodal
q	1.03	0.00	1.00	Unimodal
σ_p^2	0.99	-0.00	1.00	Unimodal
σ_o^2	0.98	-0.01	1.00	Unimodal
P_{1999}	1.06	0.02	0.94	Unimodal

Table 7. Posterior summaries of each parameter ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1999}$) and management reference points: maximum sustainable yield (MSY), biomass that yields MSY (B_{MSY}), and harvest rate which corresponds to MSY (H_{MSY}). Units for k, MSY, B_{MSY} are metric ton (MT), q and σ_o^2 are 1/haul and log[MT/haul], respectively.

Parameters	Summary			
	Mode	2.5%	50%	97.5%
r	0.16	0.05	0.28	1.03
k	1,832,623	851,710	2,890,000	12,900,000
q	5.49×10^{-5}	1.70×10^{-6}	1.06×10^{-5}	4.18×10^{-5}
σ_p^2	0.13	0.08	0.14	0.30
σ_o^2	0.07	0.04	0.08	0.15
P_{1999}	0.51	0.13	0.56	1.49
MSY	152,677	33,208	205,295	1,020,000
B_{MSY}	910,866	425,856	1,450,000	6,450,000
H_{MSY}	0.08	0.03	0.14	0.52

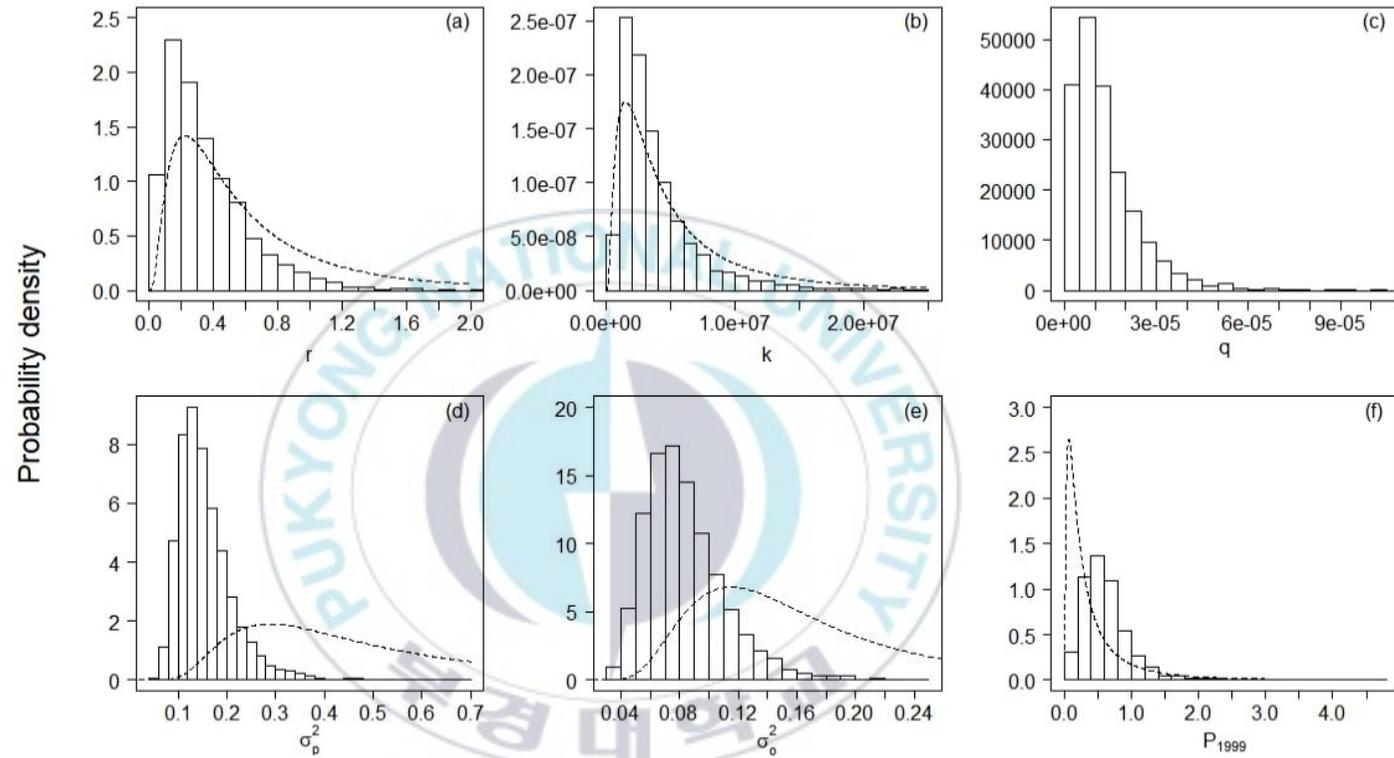


Figure 7. Posterior densities obtained with the specified priors. Dashed lines represent the priors and histograms depict the Markov Chain Monte Carlo samples from the posterior distribution in each panel. Units for k , q and σ_o^2 are metric ton (MT), 1/haul and $\log[\text{MT}/\text{haul}]$ respectively. Note that a uniform prior for $\log q$ is considered while it is not shown in panel (c).

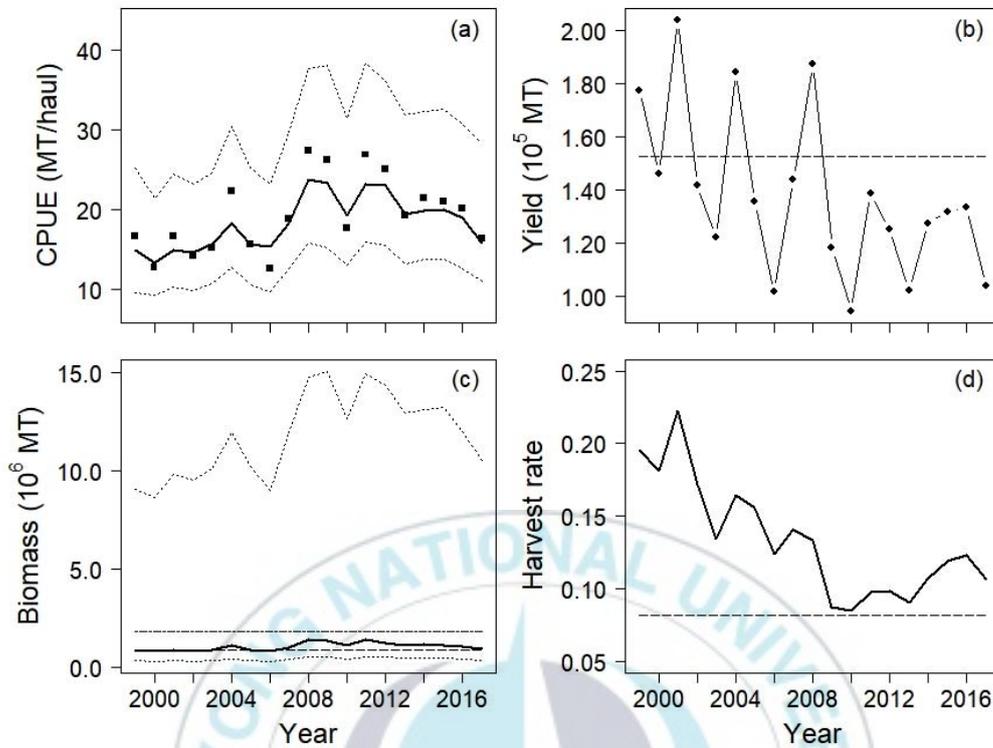


Figure 8. Comparison of management references and predicted values. Panel (a) shows the CPUE data (filled squares) and the predicted CPUE (thick solid line) with the 95% credible intervals (dotted lines). Annual yields are shown (solid line with filled circles) with MSY (dashed line) in panel (b). In panel (c), the predicted annual biomass (thick solid line) and 95% credible intervals (dotted lines) are presented with estimated B_{MSY} (dashed line) and carrying capacity (two-dashed line). Panel (d) contrasts the predicted harvest rate (thick solid line) and H_{MSY} (dashed line).

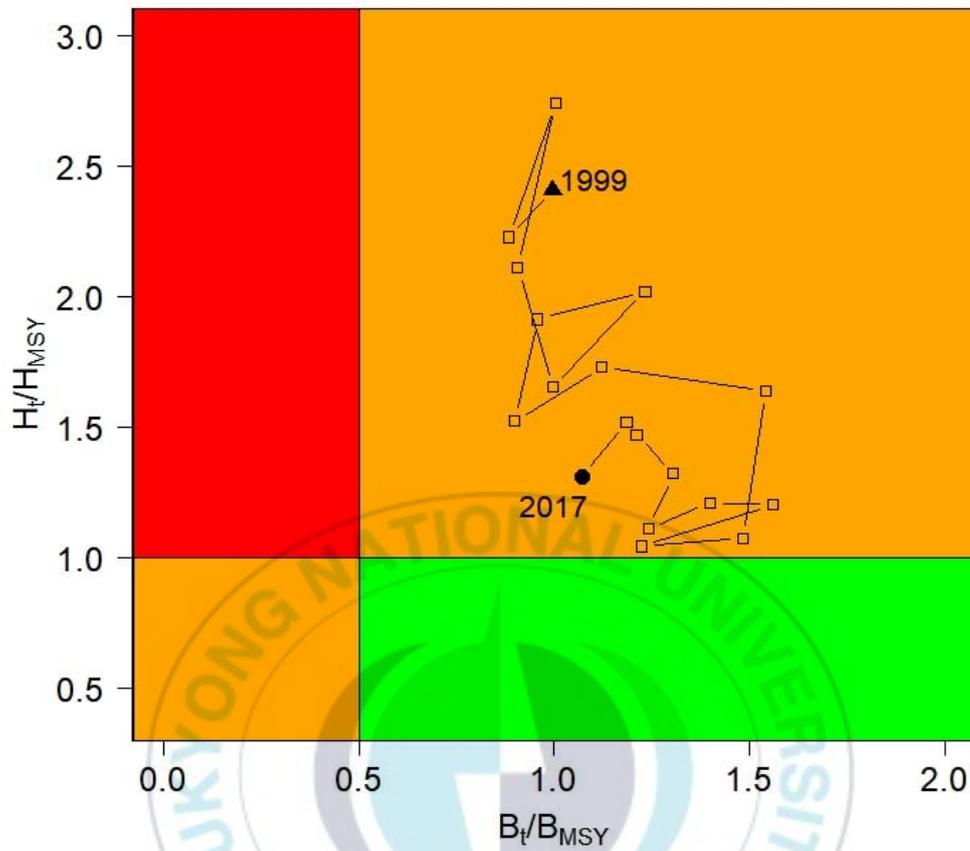


Figure 9. Kobe plot showing the predicted trajectories of B_t/B_{MSY} and H_t/H_{MSY} from 1999 to 2017. The stock is said to be “overfished” when the biomass B_t is smaller than $B_{MSY}/2$, and “overfishing” is said to occur when the harvest rate is larger than H_{MSY} .

4.4 Suggestions for future research

CPUE data are often influenced by various factors such as oceanic conditions or fishing gears (Hinton and Maunder, 2003; Hoyle et al., 2014). To be specific, the observation equation in this study addresses that the CPUE is proportionally related to biomass with scaling factor q , where fishing vessels usually have selectivity to the target stock. Therefore, Hinton and Maunder (2003) and Hoyle et al. (2014) suggested to present the observation equation as a generalized linear model to bring factors that may affect CPUE. In case of Korean mackerel, fishing area can be included as a predictor variable.

Another suggestion for assessing Korean mackerel is to combine the data from three countries, including Korea, China, and Japan to view as one population. While it is common to ignore movement (i.e., immigration or emigration) in production models, chub mackerel is known as a highly migratory species, distributed from the East China Sea to Kurile Island, including East/ Japan Sea (Castro Hernandez and Ortega Santana, 2000; Hiyama et al., 2002; Lee, 2018). In addition to the wide range of habitats, chub mackerel is exploited by Korea, China and Japan in adjacent fishing grounds. This weakness could be addressed by aggregating annual yield and CPUE data from all three countries and applying the state-space production model to assess the stock as a single population. Similarly,

Hiyama et al. (2002) used two data sets, one each collected from Korea and Japan, and estimated the stock size as a single population in the East China Sea and the Japan Sea based on chub mackerel migration patterns. This simple approach may provide better estimates of model parameters once rigorous studies have been conducted on the migratory pattern of the species.



5. Conclusions

In this study, I applied a state-space production model to Korean mackerel data collected from 1976 to 2017 with prior distributions for model parameters with the script software ADMB-RE. The model provided parameter estimates including intrinsic growth rate, carrying capacity, catchability coefficient, variances of process and observation error and annual biomass, as well as their uncertainties. The state-space production model outperformed various production models in fitting CPUE data with a *mixed trend*. Result of applying the model to data from 1999 to 2017 indicated that the stock suffered from overexploitation in terms of harvest rate. In conclusion, the state-space production model should be preferred when fitting the relative biomass index which trajectories show fluctuations in the overall period. Future studies regarding the migration pattern of the species and modification of the observation equation in the state-space production model would be worthwhile.

6. Acknowledgements

본 논문은 국립수산과학원의 연구과제 (C-D-2018-3000)의 재정적 지원을 받아 수행되었으며, 고등어 자료 중 어획량 자료는 통계청으로부터, 단위노력당 어획량 자료는 국립수산과학원으로부터 제공받았습니다.

수많은 조언과 아낌없는 격려를 주신 현상윤 교수님의 지도 덕분에 학문이라는 넓디넓은 바다에 한 걸음을 내딛는 계기가 되었습니다. 2년동안 해 주셨던 진심 어린 말씀들을, 앞으로 살아가는데 나침반처럼 활용할 수 있도록 마음 깊이 새겼습니다.

바쁜 와중에도 논문에 대한 조언과 검토를 아낌없이 해 주신 이성백 교수님께, 그리고 서영일 박사님께 감사드립니다. 교수님의 섬세한 강의로부터 통계학이라는 철학을 조금 더 이해할 수 있는 깊은 시간을 보냈습니다. 또한 고등어와 자료에 대해 많은 질문을 드렸음에도, 자세하게 답변해주신 서영일 박사님과 이재봉 박사님 덕분에 자료에 대한 이해가 한층 더 깊어질 수 있었습니다.

건물에서 마주칠 때마다 따뜻한 한마디로 격려해 주셨던 김수암 교수님, 남기완 교수님, 백혜자 교수님, 오철웅 교수님, 김현우 교수님, 박원규 교수님께 진심으로 감사드립니다. 교수님들의 지도 덕분에 해양생물이라는 분야의 여러 면모를 맛볼 수 있었습니다.

많은 질문에 일일이 시간 내주시고, 다방면으로 조언을 주신 윤민 교수님께 감사드립니다. 교수님 덕분에 제가 학문을 대하는 자세를 다시 돌아보고 고민해보는 시간을 가졌습니다.

졸업 이후에도 학문적으로 많은 도움을 주신 김규한, 이효태 선배님께 감사드립니다.

니다. 두 분의 피드백 덕분에 이 페이지까지 무사히 도착했습니다. 실험실에서 함께 의지하며 공부했던 김진우 동기님께, 그리고 안동영, 박민규, 이승준 후배님께 감사드립니다.

아울러 늘 저를 응원해준 무진언니, 늘 걱정해주고 챙겨주신 화현언니, 부족한 영어를 다듬는데 많은 시간을 할애해주신 김도균 선생님, 늘 도움의 손길을 내밀어주셨던 윤정훈 선생님, 제 작은 고민에도 주기적으로 시간 내주시던 가람언니, 시도때도 없는 디버깅을 도와주신 동원오빠, 항상 용기를 주었던 현정언니와 주란언니, 바쁜 와중에도 불구하고 디펜스 들으러 먼 길을 해주신 태우오빠, 공감률 100% 의철오빠, 서글서글한 서하 회장님, 어느덧 박사님이 되신 영선언니와 곧 박사님이 되실 승은언니, 미래가 기대되는 부용이, 열심히 살아가는 태용이, 꽃길만 걸게 될 대근오빠, 한 학기 먼저 졸업하고 힘내라는 말을 아낌없이 던져준 수진이, 끈기파 대학원생 소정이, 파워 긍정 소영이, 훌륭하게 학업을 마무리한 해인이, 사회인으로서 성실히 살아가고 있는 윤정, 민경, 수진, 네 학기 내내 고마웠던 온화한 선생님 경란이와 어엿한 직장인 예지, 그리고 이진호 과장님에게 진심으로 감사드립니다.

마지막으로, 제 성취도에 관계없이 따끔하고 따뜻한 말로 저를 보듬어준, 사랑하는 나의 가족에게 이 논문을 바칩니다.

7. References

- Carruthers, T.R., McAllister, M.K., Taylor, N.G., 2011. Spatial surplus production modeling of Atlantic tunas and billfish. *Ecol. Appl.* 21, 2734–2755.
<https://doi.org/10.1890/10-2026.1>
- Castro Hernandez, J.J., Ortega Santana, A.T., 2000. Synopsis of Biological Data on The Chub Marckerel (*Scomber japonicus* Houttuyn, 1782), FAO Fishery Synopsis.
- Chaloupka, M., Balazs, G., 2007. Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock. *Ecol. Modell.* 205, 93–109. <https://doi.org/10.1016/j.ecolmodel.2007.02.010>
- Cho, J., Lee, J.-S., Nam, J., 2009. A study on estimating the fishery optimal production by using a Bio-economic model. Korea Maritime Institute.
- Choi, Y.M., Zhang, C.I., Lee, J.B., Kim, J.Y., Cha, H.K., 2004. Stock Assessment and Management Implications of Chub Mackerel, *Scomber japonicus* in Korean Waters. *Korean Soc. Fish. Resour.* 7, 90–100.
- Clarke, R.P., Yoshimoto, S.S., Pooley, S.G., 1992. A Bioeconomic Analysis of the Northwestern Hawaiian Islands Lobster Fishery. *Mar. Resour. Econ.* 7, 115–140. <https://doi.org/10.1086/mre.7.3.42629029>
- Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder,

- M.N., Nielsen, A., Sibert, J., 2012. AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim. Methods Softw.* 27, 233–249.
<https://doi.org/10.1080/10556788.2011.597854>
- Fox, W.W., 1970. An Exponential Surplus-Yield Model for Optimizing Exploited Fish Populations. *Trans. Am. Fish. Soc.* 99, 80–88.
[https://doi.org/10.1577/1548-8659\(1970\)99<80:AESMFO>2.0.CO;2](https://doi.org/10.1577/1548-8659(1970)99<80:AESMFO>2.0.CO;2)
- Hilborn, R., 2010. Pretty Good Yield and exploited fishes. *Mar. Policy* 34, 193–196. <https://doi.org/10.1016/j.marpol.2009.04.013>
- Hilborn, R., 2002. The dark side of reference points. *Bull. Mar. Sci.* 70, 403–408.
- Hilborn, R., Walters, C.J., 1992. *Quantitative Fisheries Stock Assessment*. Springer US, Boston, MA. <https://doi.org/10.1007/978-1-4615-3598-0>
- Hinton, M.G., Maunder, M.N., 2003. Methods for standardizing CPUE and how to select among them.
- Hiyama, Y., Yoda, M., Ohshimo, S., 2002. Stock size fluctuations in chub mackerel (*Scomber japonicus*) in the East China Sea and the Japan/East Sea. *Fish. Oceanogr.* 11, 347–353. <https://doi.org/10.1046/j.1365-2419.2002.00217.x>
- Hoyle, S.D., Langley, A.D., Campbell, R.A., 2014. Recommended approaches for standardizing CPUE data from pelagic fisheries.

- Jeong, M., Nam, J., 2017. Estimation of Fishery Resource Rebuilding and Economic Effects on Coastal Gill-net Fishery as a Result of Korean Vessel Buy-back Program. *Ocean Polar Res.* 39, 221–232.
<https://doi.org/10.4217/OPR.2017.39.3.221>
- Kim, H.-A., Seo, Y.-I., Cha, H.K., Kang, H.-J., Zhang, C.-I., 2018. A study on the estimation of potential yield for Korean west coast fisheries using the holistic production method (HPM). *J. Korean Soc. Fish. Technol.* 54, 38–53.
<https://doi.org/10.3796/KSFOT.2018.54.1.038>
- Kimura, D.K., Balsiger, J.W., Ito, D.H., 1996. Kalman filtering the delay-difference equation: Practical approaches and simulations. *Fish. Bull.* 94, 678–691.
- Korean Ministry of Ocean and Fisheries, 2017. Management plan for total allowable catch for 11 species in 2017 [WWW Document]. URL <http://www.mof.go.kr/article/view.do?menuKey=376&boardKey=10&articleKey=14439> (accessed 3.28.19).
- Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M., 2016. TMB : Automatic Differentiation and Laplace Approximation. *J. Stat. Softw.* 70.
<https://doi.org/10.18637/jss.v070.i05>
- Kwon, Y., Zhang, C.I., Pyo, H.D., Seo, Y. Il, 2013. Comparison of models for estimating surplus productions and methods for estimating their parameters.

- J. Korean Soc. Fish. Technol. 49, 18–28.
<https://doi.org/10.3796/KSFT.2013.49.1.018>
- Lee, H., 2018. Developmental characteristics of the eggs and larval chub mackerel *Scomber japonicus* and its potential use for spawning biomass estimation. Pukyong National University.
- Ludwig, D., Walters, C.J., Cooke, J., 1988. Comparison of two models and two estimation methods for catch and effort data. *Nat. Resour. Model.* 2, 457–498. <https://doi.org/10.1111/j.1939-7445.1988.tb00068.x>
- McAllister, M.K., Pikitch, E.K., Punt, A.E., Hilborn, R., 1994. A Bayesian Approach to Stock Assessment and Harvest Decisions Using the Sampling/Importance Resampling Algorithm. *Can. J. Fish. Aquat. Sci.* 51, 2673–2687. <https://doi.org/10.1139/f94-267>
- Meyer, R., Millar, R.B., 1999. BUGS in Bayesian stock assessments. *Can. J. Fish. Aquat. Sci.* 56, 1078–1087. <https://doi.org/10.1139/f99-043>
- Millar, R., Meyer, R., 1999. Bayesian stock assessment using a nonlinear state-space model. *Can. J. Fish Aquat. Sci.* 56, 37–52.
- Millar, R.B., Meyer, R., 2000. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *J. R. Stat. Soc. Ser. C (Applied Stat.)* 49, 327–342.
<https://doi.org/10.1111/1467-9876.00195>

- Musick, J.A., Bonfil, R., 2005. Management techniques for elasmobranch fisheries, FAO Fisheries technical paper. Rome.
- National Research Council, 1998. Improving Fish Stock Assessments, The National Academies Press. National Academies Press, Washington, D.C.
<https://doi.org/10.17226/5951>
- Pawitan, Y., 2001. In All Likelihood: Statistical Modelling and Inference Using Likelihood. Oxford University Press.
- Pedersen, M.W., Berg, C.W., 2017. A stochastic surplus production model in continuous time. *Fish Fish.* 18, 226–243. <https://doi.org/10.1111/faf.12174>
- Pella, J.J., Tomlinson, P.K., 1969. A generalized stock production model.
- Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R news* 6, 7–11.
<https://doi.org/10.1159/000323281>
- Polacheck, T., Hilborn, R., Punt, A.E., 1993. Fitting Surplus Production Models: Comparing Methods and Measuring Uncertainty. *Can. J. Fish. Aquat. Sci.* 50, 2597–2607. <https://doi.org/10.1139/f93-284>
- Prager, M.H., 2016. User's guide for ASPIC suite, version 7: A Stock-Production Model Incorporating Covariates and auxiliary programs.
- Prager, M.H., 1994. A suite of extensions to a nonequilibrium surplus-production model. *Fish. Bull.* 92, 374–389.

- Punt, A., 2003. Extending production models to include process error in the population dynamics. *Can. J. Fish. Aquat. Sci.* 60, 1217–1228.
<https://doi.org/10.1139/f03-105>
- Quinn, T.J., Deriso, R.B., 1999. *Quantitative Fish Dynamics*. Oxford University Press.
- Rankin, P.S., Lemos, R.T., 2015. An alternative surplus production model. *Ecol. Modell.* 313, 109–126. <https://doi.org/10.1016/j.ecolmodel.2015.06.024>
- Rivot, E., Prévost, E., Parent, E., Baglinière, J.L., 2004. A Bayesian state-space modelling framework for fitting a salmon stage-structured population dynamic model to multiple time series of field data. *Ecol. Modell.* 179, 463–485. <https://doi.org/10.1016/j.ecolmodel.2004.05.011>
- Schaefer, M.B., 1954. Some Aspects of the Dynamics of Populations Important to the Management of the Commercial Marine Fisheries. *Inter-American Trop. Tuna Comm.* 1, 23–56.
- Schnute, J., 1977. Improved Estimates from the Schaefer Production Model: Theoretical Considerations. *J. Fish. Res. Board Canada* 34, 583–603.
<https://doi.org/10.1139/f77-094>
- Skaug, H., Fournier, D., 2017. *Random Effects in AD Model Builder, ADMB-RE user guide*. admb-project.org.
- Vincenzi, S., Mangel, M., Crivelli, A.J., Munch, S., Skaug, H.J., 2014.

- Determining Individual Variation in Growth and Its Implication for Life-History and Population Processes Using the Empirical Bayes Method. *PLoS Comput. Biol.* 10, e1003828. <https://doi.org/10.1371/journal.pcbi.1003828>
- Winker, H., Carvalho, F., Kapur, M., 2018. JABBA: Just Another Bayesian Biomass Assessment. *Fish. Res.* 204, 275–288. <https://doi.org/10.1016/j.fishres.2018.03.010>
- Yoshimoto, S.S., Clarke, R.P., 1993. Comparing Dynamic Versions of the Schaefer and Fox Production Models and Their Application to Lobster Fisheries. *Can. J. Fish. Aquat. Sci.* 50, 181–189. <https://doi.org/10.1139/f93-020>
- Zeller, D., Darcy, M., Booth, S., Lowe, M.K., Martell, S., 2008. What about recreational catch? Potential impact on stock assessment for Hawaii's bottomfish fisheries. *Fish. Res.* 91, 88–97. <https://doi.org/10.1016/j.fishres.2007.11.010>

8. Appendices

Appendix A: Normal prior on $\log P_{1976}$ in ADMB-RE

In practice, within the software ADMB-RE, prior distributions for model parameters $(r, k, q, \sigma_p^2, \sigma_o^2, P_{1976})$ were considered using mode and CV. A normal distribution was specified for $\log P_{1976}$ using the mode and CV for P_{1976} . A simple proof of the relationship between a normal distribution and a log-normal distribution with transformation of a random variable given as below.

Let X be a random variable which follows a log-normal distribution with a mean of μ_x and variance of σ_x^2 . That is, $X \sim \log N(\mu_x, \sigma_x^2)$. Thus, X have the probability density function (PDF) as follows.

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} \cdot \frac{1}{x} \exp\left(-\frac{(\log(x) - \mu_x)^2}{2\sigma_x^2}\right), \quad x > 0, \sigma_x^2 > 0$$

Let $y = g(x) \equiv \log(x)$. That is, $g^{-1}(y) = x = \exp(y)$. Note that the function $g^{-1}(y)$ is monotonic increasing function.

Let $\frac{dx}{dy} = \frac{dg^{-1}(y)}{dy}$. Then, the cumulative density function (CDF) for Y is

given that

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y))$$

Hence, the PDF for Y is given by

$$f_Y(y) = \frac{d}{dy} F_Y(y) = f_X(g^{-1}(y)) \frac{dx}{dy} = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

$$\because g^{-1}(y) = \exp(y); \quad \left| \frac{dg^{-1}(y)}{dy} \right| = \exp(y),$$

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{(\log(x) - \mu_X)^2}{2\sigma_X^2}\right) \frac{\exp(y)}{x} \quad \because y = \log(x) \\ &= \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left(-\frac{(y - \mu_X)^2}{2\sigma_X^2}\right), \quad y \in \mathbb{R}, \sigma_X^2 > 0 \end{aligned}$$

Since the random variable X is log-normally distributed, Y is said to follow a normal distribution with the same parameters of the distribution of X . Note that the parameters μ_X and σ_X^2 for the random variable X are not mean and variance,

but they are for $Y \equiv \log X$ which follows a normal distribution. Mean and variance of X are given as

$$E(X) = \exp\left(\mu_x + \frac{1}{2}\sigma_x^2\right) \quad (\text{A.1})$$

$$\text{var}(X) = \exp(2\mu_x + \sigma_x^2) \times (\exp(\sigma_x^2) - 1) \quad (\text{A.2})$$

And the mode and CV are given as

$$\text{mode}(X) = \exp(\mu_x - \sigma_x^2) \quad (\text{A.3})$$

$$\text{CV} = \sqrt{\exp(\sigma_x^2) - 1} \quad (\text{A.4})$$

ADMB-RE code shown in next section involves re-expression of μ_x and σ_x^2 with the mode and CV of X , and I provided how I derived them from (A.1) and (A.2).

First, switch the first term in variance of X with $[E(X)]^2 = \exp(2\mu_x + \sigma_x^2)$.

Then the (A.2) can be written as below.

$$\text{var}(X) = [E(X)]^2 \times \exp(\sigma_x^2) - [E(X)]^2$$

Then the variance of X can be expressed in terms of $\exp(\sigma_x^2)$. That is,

$$\exp(\sigma_x^2) = \frac{\text{var}(X) + [E(X)]^2}{[E(X)]^2} \quad (\text{A.5})$$

Taking the logarithm of (A.5), variance of $Y \equiv \log X$ is obtained as

$$\begin{aligned} \sigma_x^2 &= \log \left[\frac{\text{var}(X) + [E(X)]^2}{[E(X)]^2} \right] \\ &= \log [\text{var}(X) + [E(X)]^2] - \log [[E(X)]^2] \end{aligned} \quad (\text{A.6})$$

Taking the logarithm of both sides of (A.1), and re-express the σ_x^2 with the right-hand side of (A.6). The resulting equation is:

$$\log[E(X)] = \mu_x + \frac{1}{2}\sigma_x^2$$

That is,

$$\begin{aligned}\mu_x &= \log[E(X)] - \frac{1}{2}\sigma_x^2 \\ &= \log[E(X)] - \frac{1}{2}\log\left[\text{Var}(X) + [E(X)]^2\right] + \frac{1}{2}\log\left[[E(X)]^2\right]\end{aligned}$$

Mean of $Y \equiv \log X$ can then be expressed with mean and variance of X , $E(X)$ and $\text{var}(X)$.

$$\mu_x = 2\log[E(X)] - \frac{1}{2}\log\left[\text{Var}(X) + [E(X)]^2\right] \quad (\text{A.7})$$

Given the mode and CV in (A.3) and (A.4), μ_x and σ_x^2 are re-expressed as below.

$$\mu_x = \log\left[\text{mode}(X) \cdot (\text{CV}(X)^2 + 1)\right] \quad (\text{A.8})$$

$$\sigma_x^2 = \log\left([\text{CV}(X)]^2 + 1\right) \quad (\text{A.9})$$

Appendix B: ADMB-RE code for a Bayesian state-space production model

TPL file

```
//a state-space production model
//Target data: Chub Mackerel (KNIFS)
//Author: Saang-Yoon Hyun and Yuri Jung
TOP_OF_MAIN_SECTION
  arrmblsize=50000000;
  gradient_structure::set_MAX_NVAR_OFFSET(50000);
  gradient_structure::set_NUM_DEPENDENT_VARIABLES(100000);

DATA_SECTION
  init_ivector ddim(1,3);
  int nyrs;
  int ncol; //data column
  !!nyrs=ddim(1);
  !!ncol=ddim(2);

  init_matrix catchcpue(1,nyrs,1,ncol);
  vector yrs(1,nyrs);
  vector Ct(1,nyrs); //Yield
  vector It(1,nyrs); //cpue data
  vector log_It(1,nyrs);
  !!yrs=column(catchcpue,1);
  !!Ct=column(catchcpue,2);
  !!It=column(catchcpue,3);
  !!log_It=log(It);

  number min_logK;
  !!min_logK=log(max(Ct));

  init_vector mode_P1(1,3);
  init_vector CV_P1(1,3);
```

```

number mu_logP1; // mean of the log_P1
number var_logP1; // variance of the log_P1
number mean_P1;
number var_P1;
number check_mean_logP1;
number check_var_logP1;

//prior for sig2_obs ~ inverse gamma (alpha, beta);
init_vector mode_sig2_obs(1,3);
init_vector CV_sig2_obs(1,3);
number alpha_sig2_obs;
number beta_sig2_obs;

//prior for sig2_proc ~ inverse gamma (alpha, beta);
init_vector mode_sig2_pr(1,3);
init_vector CV_sig2_pr(1,3);
number alpha_sig2_pr;
number beta_sig2_pr;

//prior for K ~ lognormal (mu, sigma2); - lnorm(mu,sigma2)
init_vector mode_K(1,3);
init_vector CV_K(1,3);
number sig_K;
number mu_K;

//prior for r ~ lognormal (mu, sigma2);
init_vector mode_r(1,3);
init_vector CV_r(1,3);
number sig_r;
number mu_r;

PARAMETER_SECTION
init_number log_sig2o(4);
init_number log_sig2p(5);

```

```

init_bounded_number log_K(min_logK,17.0,2);
init_bounded_number log_q(-90.0,-1.0,3);
init_number log_r(1);
random_effects_vector log_Pt(1,nyrs+1,6); //Pt= Bt/K;

number sig2o; //observation error
number sdo;
number sig2p; //process error
number sdp; //sd of the process error
number K;
number q;
number r;
number P1;
number aic;

sdreport_vector Pt(1,nyrs+1);

sdreport_number MSY;
sdreport_number Bmsy;
sdreport_number Hmsy;
sdreport_number Fmsy;

sdreport_vector Bt(1,nyrs+1); //Pt= Bt/K;
sdreport_vector fCPUE(1,nyrs+1); //fitted CPUE

objective_function_value jnll; // joint negative log-likelihood

```

PRELIMINARY_CALCS_SECTION

```

mu_logP1=log( mode_P1(1)*( square(CV_P1(1))+1) );
var_logP1=log( square(CV_P1(1))+1 );
mean_P1=mfexp( mu_logP1+0.5*var_logP1 );
var_P1=mfexp( 2.0*mu_logP1+var_logP1 )*( mfexp(var_logP1)-1 );
check_mean_logP1=2.0*log(mean_P1)-
0.5*log( var_P1+square(mean_P1) );

```

```
check_var_logP1=-2.0*log(mean_P1)+log( var_P1+square(mean_P1) );
```

```
alpha_sig2_obs=(1.0/square(CV_sig2_obs(1)))+2.0;  
beta_sig2_obs=mode_sig2_obs(1)*(alpha_sig2_obs+1);  
alpha_sig2_pr=(1.0/square(CV_sig2_pr(1)))+2.0;  
beta_sig2_pr=mode_sig2_pr(1)*(alpha_sig2_pr+1);  
mu_K=log(mode_K(1)*(square(CV_K(1))+1));  
sig_K=sqrt( log(square(CV_K(1))+1) );  
mu_r=log(mode_r(1)*(square(CV_r(1))+1));  
sig_r=sqrt( log(square(CV_r(1))+1));
```

PROCEDURE_SECTION

```
jnl1=0.0;  
  
sig2o=mfexp(log_sig2o); //exp(log(a))=a  
sdo=sqrt(sig2o);  
sig2p=mfexp(log_sig2p);  
sdp=sqrt(sig2p);  
  
//sdp=sdo*scale; //scale //sdp=mfexp(log_sdp);  
K=mfexp(log_K);  
q=mfexp(log_q);  
r=mfexp(log_r);  
P1=mfexp(log_Pt(1));  
Pt=mfexp(log_Pt);  
  
fCPUE=mfexp(log_Pt)*q*K; //fitted CPUE  
Bt=mfexp(log_Pt)*K;  
MSY=r*K/4;  
Bmsy=K/2;  
Hmsy=r/2;  
Fmsy=MSY/Bmsy; //Compare the results from Hmsy, Fmsy!  
  
//prior for log_Pt(1);
```

```

    jnll+=0.5*log(2.0*M_PI)+0.5*log(var_logP1)+square(log_Pt(1)-
mu_logP1)/(2.0*var_logP1); //log(P1)~normal(mu_logP1, var_logP1);

    for(int i=2;i<=(nyrs+1);++i) {
        step(log_Pt(i-1),log_Pt(i),sdp,K,r,i-1);
    };
    for(int i=1;i<=nyrs;++i) {
        obs(q,K,log_Pt(i),sdo,i);
    };

    //prior for sig2_obs ~ inverse gamma //as the negative logarithm;
    jnll+=-
1.0*alpha_sig2_obs*log(beta_sig2_obs)+gammln(alpha_sig2_obs)+1.0*(alp
ha_sig2_obs+1.0)*log(sig2o)+beta_sig2_obs/(sig2o);

    //prior for sig2_process ~ inverse gamma; //as the negative logarithm;
    jnll+=-
1.0*alpha_sig2_pr*log(beta_sig2_pr)+gammln(alpha_sig2_pr)+1.0*(alpha_s
ig2_pr+1.0)*log(sig2p)+beta_sig2_pr/(sig2p);

    jnll+=0.5*log(2.0*M_PI)+log(sig_K)+log_K+square(log(K)-
mu_K)/(2.0*sig_K*sig_K);

    //prior for r ~lognormal distribution (-1.38, 0.51^2) // as the negative
logarithm;
    jnll+=0.5*log(2.0*M_PI)+log(sig_r)+log_r+square(log(r)-
mu_r)/(2.0*sig_r*sig_r);

    aic=2.0*jnll+2.0*(nyrs+6); // number of free parameters: 5 free parameters
and the Pt(1, nyrs+1)

    if(mceval_phase()) Get_outputs();

FUNCTION Get_outputs

```

```

out1<<sig2o<<" "<<sig2p<<" "<<K<<" "<<q<<" "<<r<<" "<<MSY<<"
"<<Bmsy<<" "<<Hmsy<<" "<<Pt<<" "<<Bt<<endl; //see Global_section

```

```

SEPARABLE_FUNCTION void step(const dvariable& log_P1, const
dvariable& log_P2, const dvariable& sdp, const dvariable& K, const
dvariable& r, int i)

```

```

    dvariable varp=square(sdp);
    dvariable P1=mfexp(log_P1);
    dvariable predP=(P1+r*P1*(1-P1)-Ct(i)/K);
    jnl1+=0.5*log(2.0*M_PI)+0.5*log(varp)+square(log_P2-
log(predP))/(2.0*varp);

```

```

SEPARABLE_FUNCTION void obs(const dvariable& q, const dvariable& K,
const dvariable& log_Pt, const dvariable& sdo, int i)

```

```

    dvariable varo=square(sdo);
    dvariable log_predIt=(log(q)+log(K)+log_Pt);
    jnl1+=0.5*log(2.0*M_PI*varo)+square(log_Pt(i)-log_predIt)/(2.0*varo);

```

```

REPORT_SECTION

```

```

    report<<"#prior for sig2_obs ~ inverse gamma(alpha, and beta)"<<endl;
    report<<"#alpha.o, beta.o, mode.o, CV.o: "<<endl;
    report<<alpha_sig2_obs<<" "<<beta_sig2_obs<<"
"<<mode_sig2_obs(1)<<" "<<CV_sig2_obs(1)<<endl;
    report<<"#alpha.p, beta.p, mode.p, CV.p: "<<endl;
    report<<alpha_sig2_pr<<" "<<beta_sig2_pr<<" "<<mode_sig2_pr(1)<<"
"<<CV_sig2_pr(1)<<endl;
    report<<"#sdp, sdo: "<<endl;
    report<<sdp<<" "<<sdo<<" "<<endl;
    report<<"#mu_r, sig_r: "<<endl;
    report<<mu_r<<" "<<sig_r<<endl;
    report<<"#mu_K, sig_K: "<<endl;
    report<<mu_K<<" "<<sig_K<<endl;
    report<<"#min_logK: "<<endl;

```

```

report<<min_logK<<endl;
report<<"#mu_logP1, check_mean_logP1, var_logP1, check_var_logP1,
mean_P1, var_P1"<<endl;
report<<mu_logP1<<" "<<check_mean_logP1<<" "<<var_logP1<<"
"<<check_var_logP1<<" "<<mean_P1<<" "<<var_P1<<endl;
report<<"#MSY, jnll, AIC: "<<endl;
report<<MSY<<" "<<jnll<<" "<<aic<<endl;
report<<"#yr Ct It predIt Pt Bt max.grad: "<<endl;
for(int i=1;i<=nyrs;i++) {
    report<<yrs(i)<<" "<<Ct(i)<<" "<<It(i)<<"
"<<q*K*mfexp(log_Pt(i))<<" "<<mfexp(log_Pt(i))<<"
"<<K*mfexp(log_Pt(i))<<" ";
    report<<objective_function_value::gmax<<endl;
};

```

GLOBALS_SECTION

```

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

ofstream out1("mcmc.out");

```

DAT file with mackerel data (1976-2017)

```

# dimension
42 3 -99
# year, # yield (Metric ton) # CPUE (Metric ton per haul)
1976 107382 27.3
1977 113051 19.3
1978 99519 16.5
1979 120283 18
1980 62690 10.4
1981 108082 13.1

```

1982 99447 10
1983 122883 11.42
1984 101714 7.95
1985 68479 3.87
1986 103511 7.75
1987 101337 8.49
1988 162828 12.79
1989 163617 14.47
1990 96297 10.1
1991 89738 12.03
1992 115619 12.47
1993 174684 15.07
1994 210442 17.45
1995 200481 12.66
1996 415003 32.44
1997 160448 16.19
1998 172925 18.17
1999 177540 16.73
2000 145908 12.8
2001 203717 16.64
2002 141751 14.22
2003 122044 15.26
2004 184274 22.33
2005 135596 15.74
2006 101427 12.65
2007 143776 18.93
2008 187240 27.47
2009 117960 26.22
2010 94331 17.77
2011 138729 26.93
2012 125143 25.11
2013 102114 19.3
2014 127452 21.55
2015 131735 21.08



2016 133200 20.16
2017 103870 16.46
mode for P1
0.22244 -99 -99
CV for P1
1.24 -99 -99
mode for variance of observation error
0.09451 -99 -99
CV for variance of observation error
0.81 -99 -99
mode for variance of process error
0.19654 -99 -99
CV for variance of process error
0.96 -99 -99
mode for K
2137000 -99 -99
CV for K
0.64 -99 -99
mode for r
0.28336 -99 -99
CV for r
0.49 -99 -99

DAT file with mackerel data (1999-2017)

dimension
19 3 -99
year, # yield (Metric ton) # CPUE (Metric ton per haul)
1999 177540 16.73
2000 145908 12.8
2001 203717 16.64
2002 141751 14.22
2003 122044 15.26
2004 184274 22.33
2005 135596 15.74

2006 101427 12.65
2007 143776 18.93
2008 187240 27.47
2009 117960 26.22
2010 94331 17.77
2011 138729 26.93
2012 125143 25.11
2013 102114 19.3
2014 127452 21.55
2015 131735 21.08
2016 133200 20.16
2017 103870 16.46
mode for P1
0.06603 -99 -99
CV for P1
1.69 -99 -99
mode for variance of observation error
0.11479 -99 -99
CV for variance of observation error
0.6 -99 -99
mode for variance of process error
0.28697 -99 -99
#CV for variance of process error
1.21 -99 -99
mode for k
1454000 -99 -99
CV for k
1.26 -99 -99
mode for r
0.22498 -99 -99
CV for r
1.05 -99 -99

