



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

Yuri Jung¹, Young Il Seo², Saang-Yoon Hyun^{1,*}

¹ College of Fisheries Sciences, Pukyong National University, Busan 48513, Korea

² East Sea Fisheries Research Institute, National Institute of Fisheries Science, Gangneung 25435, Korea

Abstract

The main purpose of this study is to fit catch-per-unit-effort (CPUE) data about Korea chub mackerel (*Scomber japonicus*) stock with a state-space production (SSP) model, and to provide stock assessment results. We chose a surplus production model for the chub mackerel data, namely annual yield and CPUE. Then we employed a state-space layer for a production model to consider two sources of variability arising from unmodelled factors (process error) and noise in the data (observation error). We implemented the model via script software ADMB-RE because it reduces the computational cost of high-dimensional integration and provides Markov Chain Monte Carlo sampling, which is required for Bayesian approaches. To stabilize the numerical optimization, we considered prior distributions for model parameters. Applying the SSP model to data collected from commercial fisheries from 1999 to 2017, we estimated model parameters and management references, as well as uncertainties for the estimates. We also applied various production models and showed parameter estimates and goodness of fit statistics to compare the model performance. This study presents two significant findings. First, we concluded that the stock has been overexploited in terms of harvest rate from 1999 to 2017. Second, we suggest a SSP model for the smallest goodness of fit statistics among several production models, especially for fitting CPUE data with fluctuations.

Keywords: ADMB-RE, Process error, State-space, Stock assessment, Surplus production model

Introduction

This study is the first attempt to apply a Bayesian state-space production (SSP) model to Korea chub mackerel (*Scomber japonicus*) population, which is a coastal-pelagic fish species distributed in a wide range of depths from the surface to continental slopes, down to around 300 m in warm and temperate waters (Castro Hernández & Ortega Santana, 2000). Given the

data about annual yield from all fisheries, and annual catch-per-unit-effort (CPUE) collected from the large purse seine fishery, we chose a state-space surplus production model. Previous assessments for the population (e.g., Cho et al., 2009; Choi et al., 2004b; Choi et al., 2004a) did not consider process errors in their analysis.

Surplus production models or biomass dynamics models have been studied because of their advantages over other

Received: Feb 3, 2021 Revised: Mar 22, 2021 Accepted: Mar 24, 2021

*Corresponding author: Saang-Yoon Hyun

College of Fisheries Sciences, Pukyong National University, Busan 48513, Korea

Tel: +82-51-629-5929, Fax: +82-51-629-5931, E-mail: shyun@pknu.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2021 The Korean Society of Fisheries and Aquatic Science

assessment models. These advantages include simpler model structures and fewer model parameters to estimate (Carruthers et al., 2011; Chaloupka & Balazs, 2007; McAllister et al., 2001; Millar & Meyer, 2000b; Pedersen & Berg, 2017). Management reference points, such as the maximum sustainable yield (MSY) and the biomass that yields MSY are estimated (Prager, 1994; Punt, 2003). Despite these benefits, biomass dynamics models have been criticized for aggregating growth, recruitment, and natural mortality into a surplus production term, at the cost of biological realism (Millar & Meyer, 1999).

de Valpine (2012) proposed that a more realistic model provides a better estimation. Since a stock assessment model is a simplified version of a dynamic system, and fits fisheries data that are generally noisy, the variability should be considered for both the model and the data (Aeberhard et al., 2018). Adopting a state-space framework for production models can mitigate the simplicity of production models by incorporating process errors and observation errors. Process errors in a biomass dynamics model describe uncertainties of random effect parameters such as biomass, and observation errors account for the noise in the data. Despite the ability to account for these two sources of uncertainty, the inclusion of process errors requires the estimation of a number of free parameters, and high-dimensional integration. Over the past few decades, researchers have implemented state-space models under assumptions such as model linearity, which is ecologically unrealistic, within the frequentist statistical approaches. The development of statistical software such as ADMB-RE (Fournier et al., 2012; Skaug & Fournier, 2017) and TMB (Kristensen et al., 2016) has enabled the implementation of state-space models with Gaussian error structure using the Laplace approximation. Another advantage of ADMB-RE is that the program provides Markov Chain Monte Carlo (MCMC) sampling without any revision of the computer code. Because the numerical optimization of parameters could not be achieved without priors, we specified prior distributions to stabilize the numerical optimization.

Using the ADMB-RE software for implementing Bayesian SSP models, we estimated the annual biomass, intrinsic growth rate carrying capacity, and associated uncertainties. We further provided management reference points, including MSY, biomass that yields MSY (B_{MSY}), and harvest rate that corresponds to MSY (H_{MSY}). We also compared the performance of our model with those of others.

Methods

Data

Two sets of historical data were used in this study: annual yield and catch-per-unit-effort (CPUE) for chub mackerel collected from Korean waters from 1999 to 2017 (Fig. 1). The Korean National Institute of Fisheries Science provided standardized CPUE with yield and effort data for 70% of the total large purse seine fisheries in Korea. We considered the chub mackerel CPUE (Fig. 1b) to be representative of all chub mackerel fisheries, as large purse seine fisheries accounted for more than 90% of the total on average (Fig. 1c).

Surplus production model

We adopted a quadratic form of the Schaefer model formulated by Walters & Hilborn (1976) based on the original Schaefer model (Schaefer, 1954):

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

where B_t is the biomass in year t , Y_t is the yield achieved by fishermen in year t , r is the intrinsic growth rate of the population, and k is the carrying capacity, or maximum population size. The second term on the right side of equation above contains the entire change in biomass, including growth, recruitment, and natural mortality in year t (Prager, 1994; Ricker, 1975), and this term is called 'surplus production' (Hilborn & Walters, 1992). Therefore, the Schaefer production model describes how the biomass at year $t + 1$ is related to the biomass in the previous year, with the model parameters r , k , and yield. Because the model assumes logistic growth or a symmetric curve of surplus production, the maximum surplus production occurs at $k/2$, which equals B_{MSY} . MSY and H_{MSY} can also be directly calculated based on $rk/4$ and $r/2$, respectively.

The model also depicts the measurement of the biomass using the following relationship:

$$I_t = qB_t$$

where I_t is the CPUE collected in year t , and q is the catchability coefficient. The measurement shows that the CPUE obtained in year t is proportional to the biomass in the same year.

We then re-parameterized the model by dividing each term in the Schaefer model and the measurement equation above to downscale the biomass. The P_t terms in the following equations

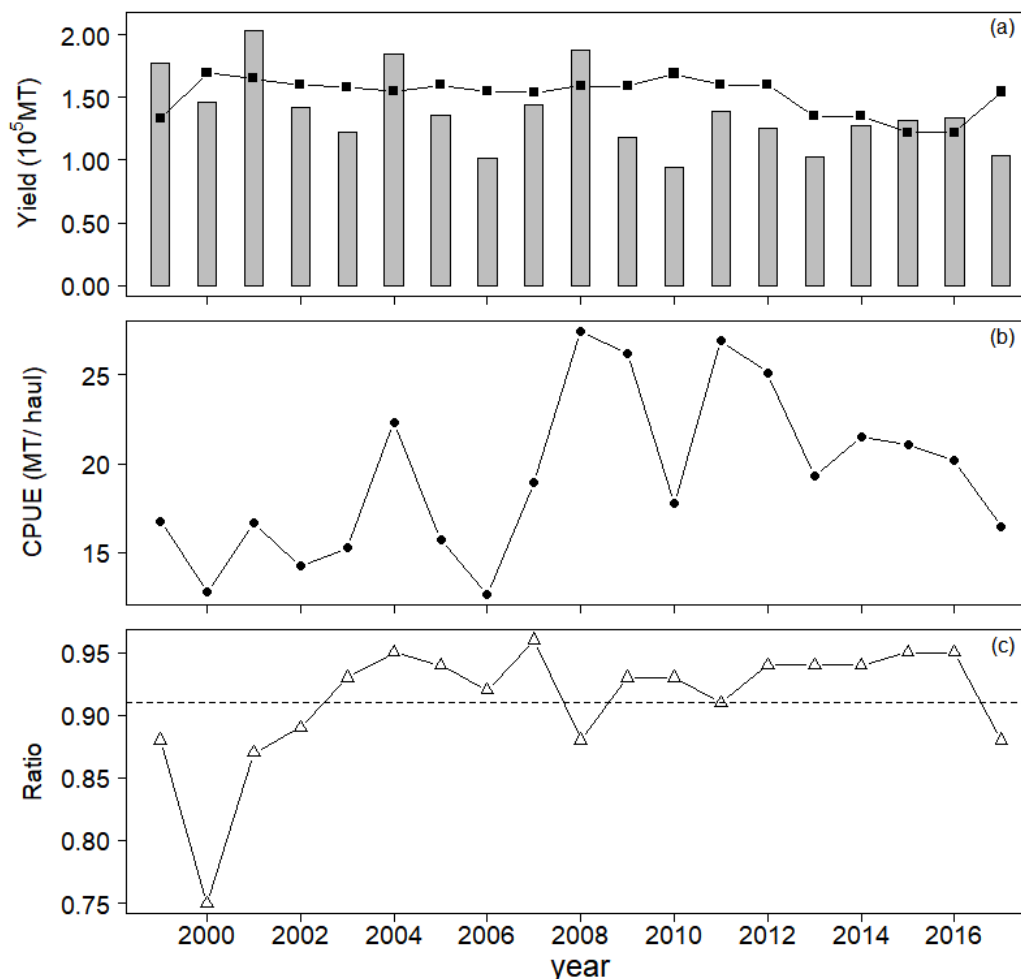


Fig. 1. Historical data on Korean mackerel. Bars in panel (a) shows the yield of chub mackerel caught by the large purse seine fisheries, and line on squares indicates the total allowable catch. Panel (b) shows the annual CPUE collected from the large purse seine fisheries. Panel (c) shows the ratio of yield from the large purse seine fisheries to the total yield from all chub mackerel fisheries. In panel (c), the broken horizontal line indicates the mean ratio of yield from the large purse seine fisheries to total yield from 1999 to 2017. CPUE, catch-per-unit-effort.

are the ratio of biomass to carrying capacity, or relative biomass (i.e., $P_t = B_t / k$). As Millar & Meyer (2000b) pointed out, using the relative biomass P_t rather than B_t in the model can reduce the extent of autocorrelation. Scaling B_t with k can improve the efficiency of numerical optimization, as described below.

The state-space production model

We constructed a SSP model using the deterministic surplus production model described by Millar & Meyer (2000a). We assumed multiplicative errors that follow normal distributions, with means of 0 and variances of σ_p^2 and σ_o^2 for the biomass dynamics and observation, respectively.

$$P_{t+1} = (P_t + rP_t(1 - P_t) - Y_t / k) \exp(\epsilon_t^p),$$

where $\epsilon_t^p \sim N(0, \sigma_p^2)$ (1)

$$I_t = qkP_t \exp(\epsilon_t^o), \text{ where } \epsilon_t^o \sim N(0, \sigma_o^2)$$
 (2)

Equation (1) is called the process equation, because it describes the dynamics of a system with process error. Equation (2) is called the observation equation, and specifies the relationship between the measurement and the state. Compared to the deterministic production model, the SSP model assumes that errors can arise from either of two distinct sources: model

and data. The process error in equation (1) can be interpreted as the uncertainty of the model structure in describing system dynamics. Thus, P_t is treated as a random variable for process errors. The observation error in equation (2) accounts for errors in observation (i.e., CPUE), which can arise from measurement error or reporting error (Winker et al., 2018). Therefore, the parameters to be estimated in our model include relative biomass $\mathbf{P}=(P_{1999}, P_{2000}, \dots, P_{2018})$ and $\Theta = (r, k, q, \sigma_p^2, \sigma_o^2)$. To distinguish the different types of parameters in our model, we denote the \mathbf{P} as random-effect parameters or states, and the time invariant constant parameters, Θ , as fixed effect parameters.

Likelihood functions

We built likelihood functions for the process equation and observation equation described above. The P_{t+1} follows a log-normal distribution, with $\log P_{t+1}$ having the mean of $\log\{P_t + rP_t(1 - P_t) - Y_t / k\}$ and variance of σ_p^2 , written as follows:

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

Likewise, $\log I_t$ follows a normal distribution with the mean of $\log\{qkP_t\}$ and variance of σ_o^2 :

$$\log I_t \sim N[\log\{qkP_t\}, \sigma_o^2]$$

Therefore, the likelihoods built using the process and observation equations have the following form, with assumptions of independence:

$$L(r, k, \sigma_p^2, \mathbf{P} | \mathbf{Y}) = \prod_{1999}^{2017} \frac{1}{\sqrt{2\pi} \sigma_p} \exp\left(-\frac{(\log P_{t+1} - [\log\{P_t + rP_t(1 - P_t) - Y_t / k\}])^2}{2\sigma_p^2}\right) \tag{3}$$

$$L(q, k, \sigma_o^2, \mathbf{P} | \mathbf{I}) = \prod_{1999}^{2017} \frac{1}{\sqrt{2\pi} \sigma_o} \exp\left(-\frac{(\log I_{t+1} - [\log\{qkP_t\}])^2}{2\sigma_o^2}\right) \tag{4}$$

With likelihood functions (3) and (4), the joint likelihood function is as follows.

$$L(\Theta, \mathbf{P} | \mathbf{D}) = L(r, k, \sigma_p^2, \mathbf{P} | \mathbf{Y}) \cdot L(q, k, \sigma_o^2, \mathbf{P} | \mathbf{I}) \tag{5}$$

Note that $\mathbf{D} = (\mathbf{I}, \mathbf{Y})$, where $\mathbf{Y} = (Y_{1999}, Y_{2000}, \dots, Y_{2017})$, and

$$\mathbf{I} = (I_{1999}, I_{2000}, \dots, I_{2017}).$$

Prior and posterior distributions

Haddon (2011) argued that data availability governs the complexity of stock assessment models. While the SSP model takes into account the variability in biomass dynamics, which improves realism, it has a large number of free parameters that must be estimated with the given data. Prager (1994) suggested that additional information is required for parameter estimation in state-space models.

As mentioned earlier, the script software ADMB-RE implements SSP models using the Laplace approximation, to estimate fixed and random effects separately (Skaug & Fournier, 2017). That is, random effects are separated from the joint likelihood function (i.e., $\int \log L(\theta, \mathbf{P} | \mathbf{D}) d\mathbf{P} = \log L(\theta | \mathbf{D})$) in numerical optimization. Because numerical optimization was not achieved, we introduced prior distributions for parameters, $r, k, q, \sigma_p^2, \sigma_o^2$, and the relative biomass in the first year, P_{1999} . Although informative priors should be considered carefully, the population parameters of chub mackerel in Korean waters have not previously been studied. Therefore, we tested various prior sets and selected the one best suited for numerical optimization. In practice, we chose log-normal distributions for r and k , whose distribution domains are positive. We chose 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 have been used in previous studies focusing on Bayesian stock assessments (Chaloupka & Balazs, 2007; Meyer & Millar, 1999; Millar & Meyer, 2000b; Winker et al., 2018). Because the catchability coefficient is a scaling factor, we assigned a uniform prior for $\log q$ (McAllister et al., 1994; Millar & Meyer, 2000b). We considered a log-normal prior for relative biomass in 1999, P_{1999} . To address uncertainties on priors, we chose priors with the coefficient of variations (CVs) larger than 50%. The selected parameter set is shown in Table 1.

Taking a Bayesian approach, the probability that a parameter is updated is based on the prior distribution and likelihood. Assuming mutual independence of the priors, the joint prior has the following form:

$$\pi(r, k, q, \sigma_p^2, \sigma_o^2, P_{1999}) = \pi(r) \pi(k) \pi(q) \pi(\sigma_p^2) \pi(\sigma_o^2) \pi(P_{1999}) \tag{6}$$

With the joint likelihood and the joint prior, the posterior is defined by Bayes' rule:

Table 1. Prior probability distributions with mode and coefficient of variation for parameters specified in the state-space production model

Parameter	Prior	Mode	CV
r	Lognormal (-0.75, 0.86)	0.22	1.05
k	Lognormal (15.14, 0.96)	1,454,000	1.26
q	Uniform on $\log q$ (-90, -1)	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 on $\log p_{1999}$ (-1.37, 1.35)	0.07	1.69

CV, coefficient of variation.

$$p(\Theta, \mathbf{P} \mid \mathbf{D}) \propto \pi(r, k, q, \sigma_p^2, \sigma_o^2, P_{1999}) L(\Theta, \mathbf{P} \mid \mathbf{D}) \quad (7)$$

Inference of parameters

Because high-dimensional integration is required to calculate the full posterior distributions, which cannot be achieved in closed form, we generated MCMC sample sets using the ADMB-RE software. We thinned out every 50,000th MCMC sample sets over 0.2 billion iterations to reduce the autocorrelation for each parameter. The first 100 sample sets (i.e., the initial 5,000,000 sample sets) were then removed as a burn-in period. We assessed four criteria to check the convergence of the MCMC samples for parameters r , k , q , σ_p^2 , σ_o^2 , and P_{1999} : (i) the dependence factor of the Raftery-Lewis statistics, (ii) the lag-1 correlation, (iii) the ratio between the naïve standard error and the time series standard error, which is corrected for autocorrelation, and (iv) the unimodal histogram of MCMC samples for each parameter. The first three were carried out with R package, “coda (Plummer et al., 2006), where we used default values (e.g., error margin (= 0.005) and probability of obtaining an estimate in the interval (= 0.95) for Raftery-Lewis statistics). We determined the convergence of the samples when the dependence factor of the Raftery-Lewis statistics was less than 5, the lag-1 correlation was close to 0, and the ratio of the naïve standard error to the time series standard error was around 1, indicating that no autocorrelation was present. The summaries of the posterior distributions give modes as point estimates and uncertainties as 95% credible intervals.

Other methods

To compare our results, we applied various other production models to the same data. The selected models included the Fox production model (Fox, 1970), Yoshimoto-Clarke model (Clarke et al., 1992; Yoshimoto & Clarke, 1993), the Schnute

regression model (Schnute, 1977), the ASPIC program (Prager, 2016, 1994), and the Observation model. The Fox model, commonly referred to as the exponential (production) model, assumes an exponential relationship between fishing effort and population size (Fox, 1970):

$$\log I_t = \log I_\infty - \frac{q}{r} E_t$$

where I_∞ is catch per unit effort, which is proportional to the carrying capacity, k , and E_t is the effort in year t . Since the model regards surplus production as yield, it assumes that the stock is at equilibrium. Hilborn & Walters (1992) criticized the model on the grounds that equilibrium conditions rarely occur, and stock size is usually overestimated. Clarke et al. (1992) and Yoshimoto & Clarke (1993) modified the Fox model, implementing non-equilibrium conditions, as a linear regression model. Using this model, Yoshimoto & Clarke (1993) showed that their model predicted CPUE data well, even with negative estimates of q and k . We denote this model the Yoshimoto-Clarke (YC) model:

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

Schnute (1977) transformed the Schaefer model into a linear regression model which allowed the use of explanatory variables with geometric means of CPUE and effort data.

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

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

The ASPIC program was developed by Prager (1994), and has been included in the NOAA toolbox (<https://nmfs-fish-tools.github.io/>). Among the various modes available for fitting the data, we fitted the Schaefer model through the least squares method, to provide results without process and observation errors. The other former models estimate r , k , and q as free parameters, and the ASPIC program estimates four free parameters: r , k , q , and B_{1999} .

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

F_t in the model is fishing mortality rate. The unit for the

estimates of r and q are year⁻¹ and year⁻¹ haul⁻¹, respectively.

When the process error is removed from equation (1), the Observation model depicts deterministic biomass dynamics and variability in measurement. Without process errors, the relative biomasses from 2000 onward (i.e., $P_{2000}, P_{2001}, \dots, P_{2018}$) are calculated as derived parameters, which do not require the Laplacian approximation. Therefore, the estimation of the model parameters was performed within ADMB through maximum likelihood estimation, instead of using ADMB-RE. Because the Observation model considers the observation error, the model estimates one more parameter, σ_o^2 , than the ASPIC program.

Results

With the specified priors and data about yield and CPUE, we obtained posterior distributions for each parameter ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1999}$) using MCMC samples. We determined our MCMC samples built converged posterior distributions for each parameter ($r, k, q, \sigma_p^2, \sigma_o^2, P_{1999}$), while accounting for the diagnostic results provided in Table 2. For each sample set of parameters, the dependence factor of the Raftery-Lewis statistics was around 1, the lag-1 correlation was around 0, the ratio between naïve and time series standard errors was around 1, and the posterior distribution was unimodal (Table 2 and Fig. 2).

The estimates (posterior modes) of intrinsic growth rate r , carrying capacity k , and catchability q were 0.16, 1,832,623 metric tons (MT), and 5.49×10^{-5} haul⁻¹, respectively. The CV of the process error was nine times larger than that of the observation error, calculated as 84% and 9%, respectively. Based on the estimates of r and k , the calculated management reference points MSY, B_{MSY} , and H_{MSY} were 152,677 MT, 910,866 MT, and 0.08,

respectively. Table 3 lists the 95% credible intervals for each parameter. Based on the scatter plots of joint posterior samples for both parameters, r and q exhibited negative relationships with k (Fig. 3). The former relationship can arise when the stock is large with low productivity, or small with high productivity (Millar & Meyer, 2000b; Walters & Martell, 2004). Millar & Meyer (2000b) also mentioned that the carrying capacity k in equation (2) scales the biomass B_t down and is highly correlated with q .

Our model result followed the CPUE data with increasing or decreasing trends (Fig. 4a). The annual yields achieved in 1999, 2001, 2004, and 2008 were larger than MSY , whereas those in other years were below MSY (Fig. 4b). The average yield of chub mackerel during 1999–2017 was 90% of MSY , but it was 78% (119,393 MT) during 2009–2017.

The annual biomass was below the carrying capacity, and the biomass was smaller than B_{MSY} in 1999, 2000, 2002, 2003, 2005, and 2006 (Fig. 4c). From 2007 onward, the predicted stock size was greater than B_{MSY} . The mean biomass from 1999 to 2017 was 1,064,924 MT, and the smallest and largest biomass values were 805,645 MT and 1,419,278 MT, respectively. The harvest rate, the proportion of yield to biomass, decreased during this period (Fig. 4d). The annual harvest rate remained above the calculated H_{MSY} .

Applying the various production models, we obtained parameter estimates for each model (Table 4). The Fox and Schnute regression models gave negative estimates for carrying capacity and catchability coefficient, which are biologically absurd values. The population growth rate estimated with the Fox, YC, and Observation models was larger by about 10 to 17 times than that with the SSP model. The Schnute model interpreted the stock size as decreasing (-0.35), whereas the ASPIC program showed that the stock grew by 0.18 per year. The YC model estimated the smallest carrying capacity among the production models (160,198 MT), whereas the ASPIC program estimated the largest value (85,693,999 MT). The Observation model estimated the carrying capacity as 596,002 MT, which is about 33% of that of the SSP model. The YC model estimated the largest value for catchability coefficient (2.05×10^{-5} haul⁻¹ year⁻¹), and the ASPIC program estimated the smallest value (2.59×10^{-7} haul⁻¹ year⁻¹) among the production models. The Observation model estimated a catchability coefficient roughly 10 times larger than that estimated by the SSP model (4.88×10^{-5} haul⁻¹). The Observation model estimated the variance of the observation error as 0.03, roughly 42% of the corresponding SSP model estimate. All pro-

Table 2. Diagnostics for the Markov Chain Monte Carlo posterior samples for parameters $r, k, q, \sigma_p^2, \sigma_o^2, P_{1999}$

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

The dependence factor of Raftery-Lewis statistics (DF), lag-1 autocorrelation, ratio of naïve standard error to time series standard error, and the shape of posterior distributions were checked.

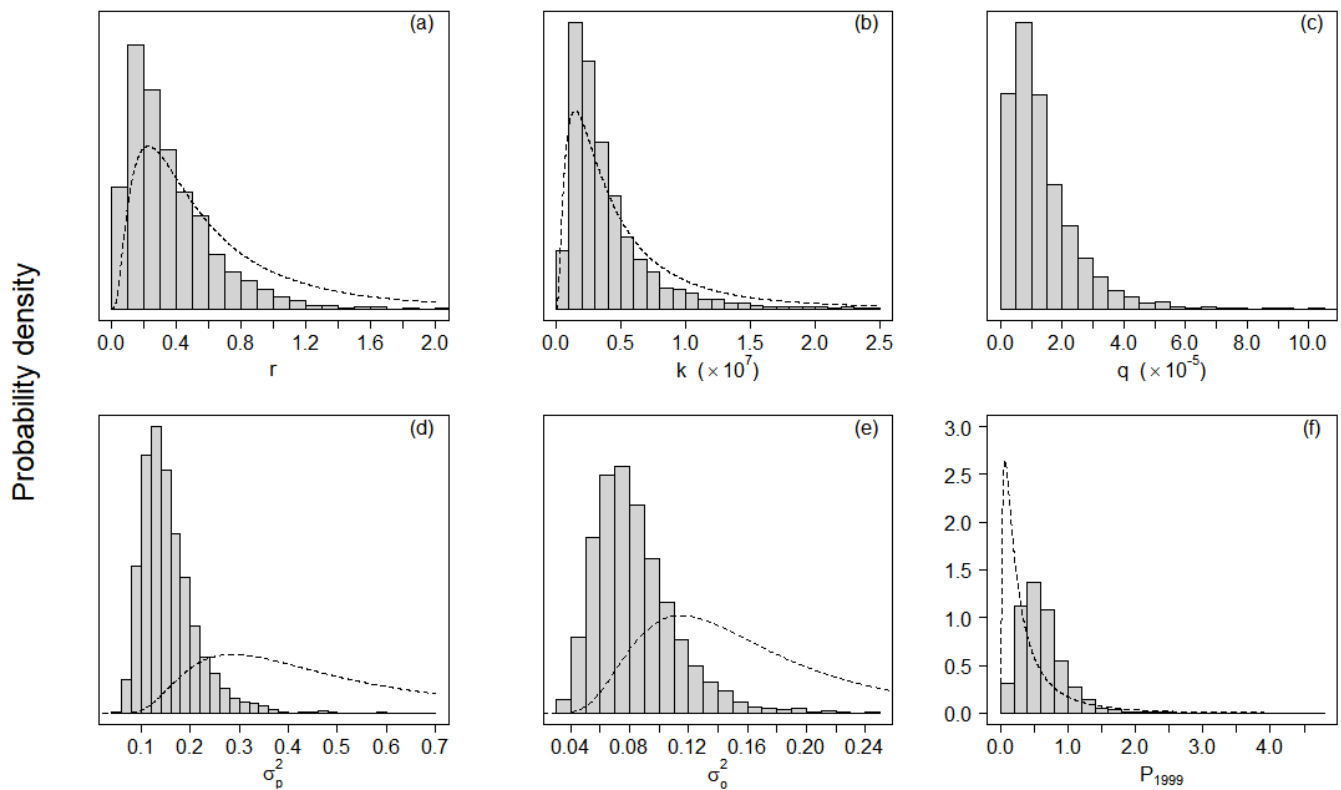


Fig. 2. Prior probability distribution (dotted line) for parameters r , k , q , σ_p^2 , σ_o^2 , P_{1999} and posterior probability distribution (histogram) obtained from Markov Chain Monte Carlo sample sets. We specified log-normal priors for r and k and P_{1999} inverse gamma priors for σ_p^2 and σ_o^2 , and uniform priors for $\log q$. Units of k and q are metric tons (MT), and 1/haul, respectively.

Table 3. Posterior summaries of parameters r , k , q , σ_p^2 , σ_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})

Parameters	Summaries			
	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

duction models estimated the relative biomass in the first year, P_{1999} , within the range 0.48 to 0.60. We excluded the Fox and Schnute models in comparisons of predicted CPUE and biomass

trajectories because of their unreliable estimates, as evidenced by their negative values for k and q .

The predicted CPUE and biomass from the four models are listed in Figs. 5 and 6. Although all four models moderately explained the CPUE data, the residual sum of squares (RSS, equation (8)) showed that the SSP outperformed the other three models in predictive accuracy (RSS_{YC}: 227.44, RSS_{ASPIC}: 280.78, RSS_{Observation}: 248.29, RSS_{SSP}: 79.91):

$$\sum_{1999}^{2017} (I_t - \hat{I}_t)^2 \tag{8}$$

The 95% confidence intervals for the Observation model did not include the CPUE data, whereas the 95% credible intervals calculated from the SSP model did (Fig. 5c, 5d).

Although the YC model had second smallest value of RSS, it only explained the yield in years 2010 and 2013 (Fig. 6a). In the ASPIC program, the stock biomass showed monotonic increases from 1999 to 2017 (Fig. 6b), and its stock size was the largest among the four production models. The ASPIC program

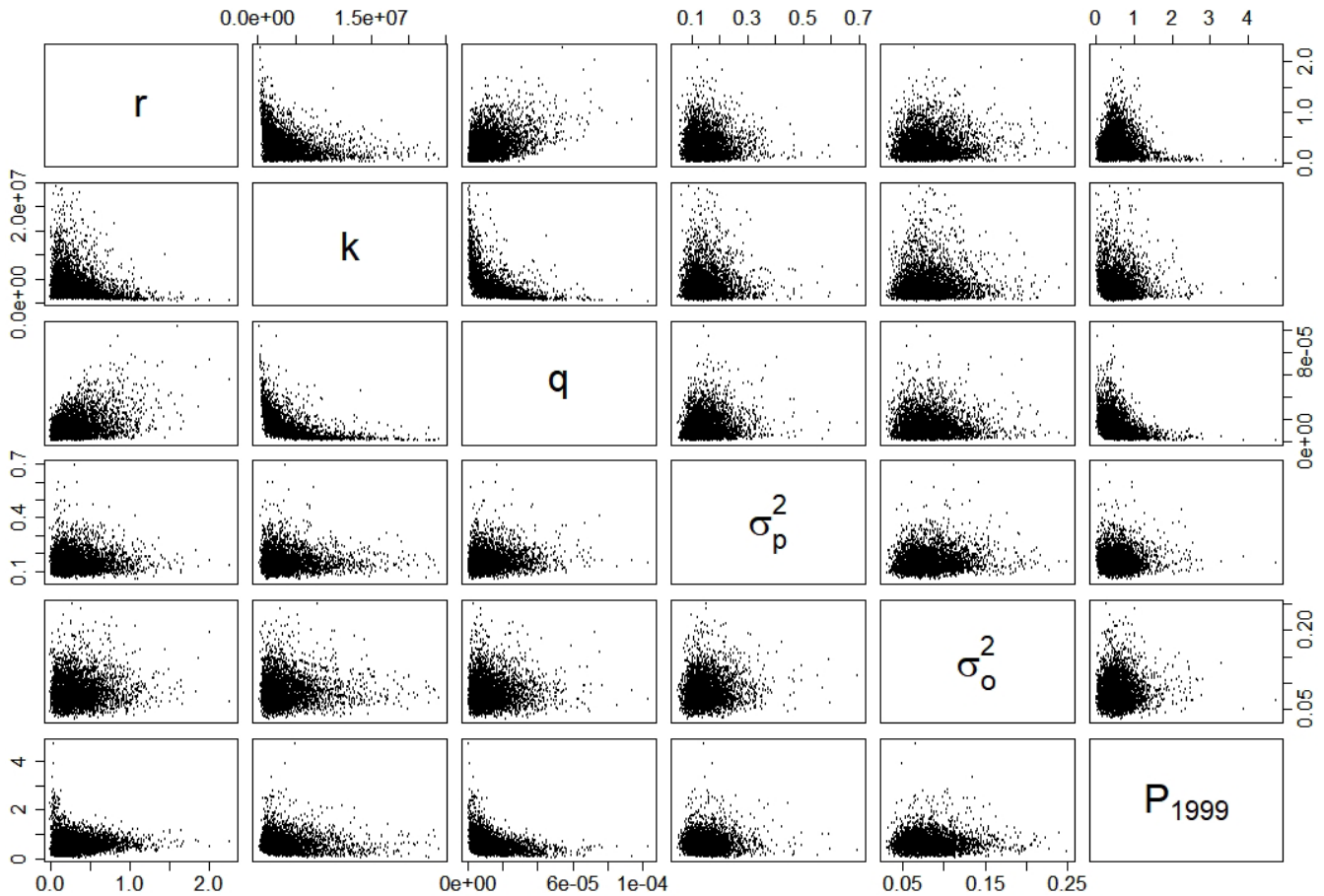


Fig. 3. Scatter plots of joint posteriors for two parameters obtained using Markov Chain Monte Carlo sampling.

estimated that the average yield from 1999 to 2017 was 137,779 MT, suggesting that the stock was underexploited. The Observation model and the SSP model predicted the annual biomass to be larger than the annual yield.

According to the National Oceanic and Atmospheric Administration (NOAA), the terms ‘overfishing’ and ‘overfished’ apply when stocks have a harvest rate larger than H_{MSY} ($H_t > H_{MSY}$) and the stock size is smaller than $0.5B_{MSY}$ (e.g., $0.5B_{MSY} > B_t$). Using these definitions, we depicted the historical trajectories of the annual harvest rate and biomass in a Kobe plot (Fig. 7). The plot suggests that 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. However, the annual biomass was larger than $0.5B_{MSY}$, indicating that the stock was not overfished during the 1999–2017 period. The graph shows a small increase in the ratio of annual biomass to B_{MSY} , from 1.00 in 1999 to 1.07 in 2017.

Discussion

The goal of fisheries management is to maximize the profit from the catch, while also effectively conserving fish stocks (Quinn & Deriso, 1999). Estimates of growth rate and carrying capacity are directly used to calculate management policies, so estimating model parameters is crucial not only to understanding the biological properties of a population, but also to determining optimal management practices for the species.

Chub mackerel have been fished by large purse seine fisheries under quota management since 1999. In 2017, the total allowable catch (TAC) was set at 123,000 MT (Korean Ministry of Ocean and Fisheries, 2017), which was 80.56% of the MSY calculated in the present study. Hilborn (2010) introduced the concept of a ‘pretty good yield’ for managers, defined as 80% of the MSY. However, since the method of calculating TAC is not available to the public in Korea, these two values cannot be

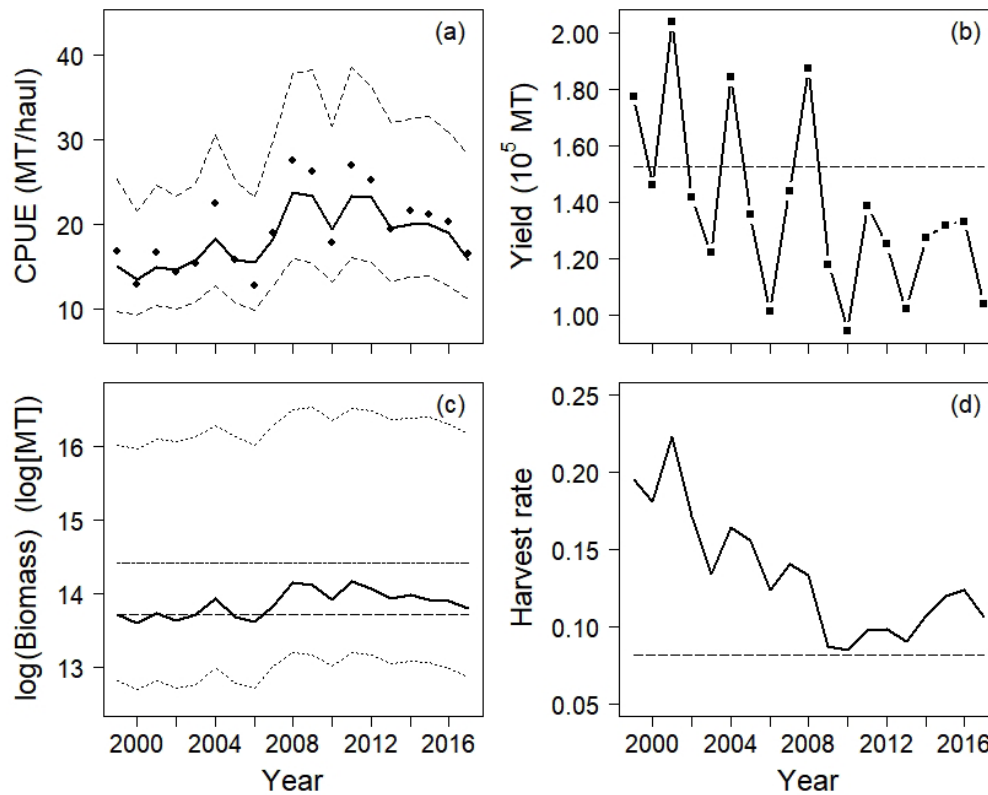


Fig. 4. Results of the state-space production model. Predicted CPUE (thick solid line) and observed CPUE data (black circles) were compared in panel (a) with 95% credible intervals (dotted lines). Annual yield data are indicated by cross marks with estimated MSY (long dashed line) in panel (b). In panel (c), the predicted annual biomass (thick solid line) with 95% credible intervals (dotted lines) is presented, with estimated B_{MSY} (long-dashed line) and carrying capacity (double-dash line) in log scales. Panel (d) contrasts the harvest rate and H_{MSY} (long-dashed line). CPUE, catch-per-unit-effort.

Table 4. Parameter estimates obtained from the Fox model (Fox), the Yoshimoto-Clarke model (YC), the Schnute regression model (Schnute), the ASPIC program, the Observation model (Observation), and the state-space production model (SSP)

Parameters	Fox	YC	Schnute	ASPIC	Observation	SSP
r	0.94	2.76	-0.35	0.18	1.13 (0.69, 1.57)	0.16 (0.05, 1.03)
k	-483,829	160,198	-16,88,001	85,693,999	596,002 (338,613, 853,392)	1,832,623 (851,710, 12,900,000)
q	-6.56×10^{-5}	2.05×10^{-4}	-2.06×10^{-5}	2.59×10^{-7}	4.88×10^{-5} (2.59×10^{-5} , 7.17×10^{-5})	5.49×10^{-6} (1.70×10^{-6} , 4.18×10^{-5})
σ_p^2	NA	NA	NA	NA	NA	0.13 (0.08, 0.30)
σ_o^2	NA	NA	NA	NA	0.03 (0.02, 0.04)	0.07 (0.04, 0.15)
P_{1999}	0.53	0.51	0.48	0.60	0.53 (0.47, 0.58)	0.51 (0.13, 1.49)

NAs in the table indicate that the model does not estimate the corresponding parameter(s).

The 95% confidence intervals for parameter estimates for the Observation model and 95% credible intervals for the SSP model are provided in parentheses.

Except for the ASPIC program results, the units of k and q are metric tons (MT), and 1/haul, respectively.

For the ASPIC program, units for r and q are year^{-1} and $\text{year}^{-1}\text{haul}^{-1}$, respectively.

YC, Yoshimoto-Clarke; SSP, state-space production.

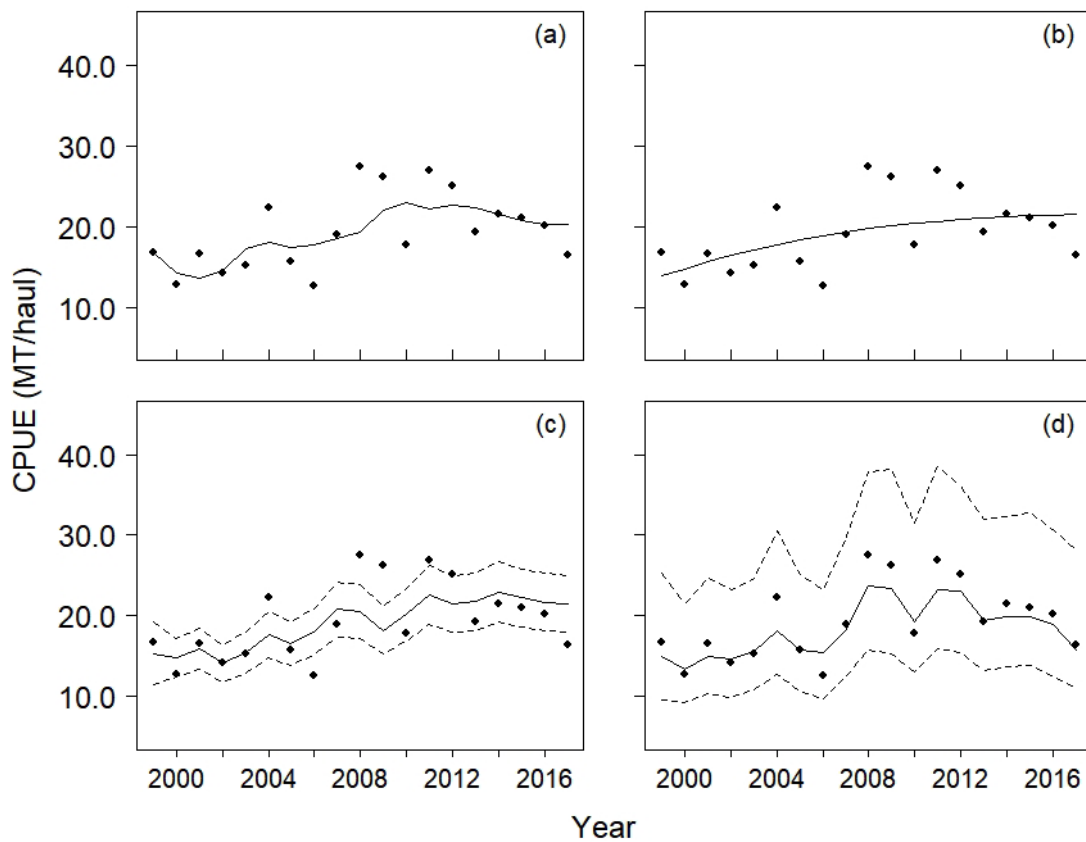


Fig. 5. Plots of predicted CPUE from the four production models. Black circles in all panels illustrate the annual CPUE data, and solid lines show the predicted CPUE from the four models. Panels (a–d) show the results from the Yoshimoto-Clarke model, the ASPIC program, the Observation model, and the state-space production model, respectively. Dotted lines along with the solid lines in panels (c) and (d) are 95% confidence intervals and 95% credible intervals, respectively, for predicted CPUE. CPUE, catch-per-unit-effort.

directly compared. Because the minimum legal size for chub mackerel is a total length of 21 cm, the predicted annual biomass in the present study must be interpreted as the biomass of fish larger than 21 cm. The quota in 2017 can therefore be said to be optimal only if the TAC targets the stock biomass within the legal size.

Although this study provides management principles for Korean chub mackerel, it can be dangerous to base management strategies for a species on estimates produced by a single model. We therefore strongly suggest evaluating various alternative models before setting management goals based on our results.

Compared with age-structured models, surplus production models have fewer requirements for historical data, fewer model parameters to be estimated, and simpler model structure. Setting aside these advantages, the main weakness of the present study is related to the model structure. First, because the

surplus production term in the model is combined with recruitment, natural mortality, and growth, our model cannot generate separate estimates of annual recruitment and natural mortality. Therefore, age data for the stock should be collected and applied within an age-structured model, such as a statistical catch-age model, not only to assess the stock year-class strength, but also for cross-validation of the estimates. Second, while chub mackerel is a highly migratory species, distributed from the East China Sea to Kurile Island (Castro Hernández and Ortega Santana, 2000; Lee & Kim, 2011; Wang et al., 2014), the model does not consider immigration or emigration. In addition to its wide range of habitats, chub mackerel is exploited by Korea, China, and Japan, in adjacent fishing grounds, which are located in the East China Sea and adjoining waters. This weakness could be addressed by aggregating annual yield and CPUE data from all three countries and applying the SSP model to assess the stock

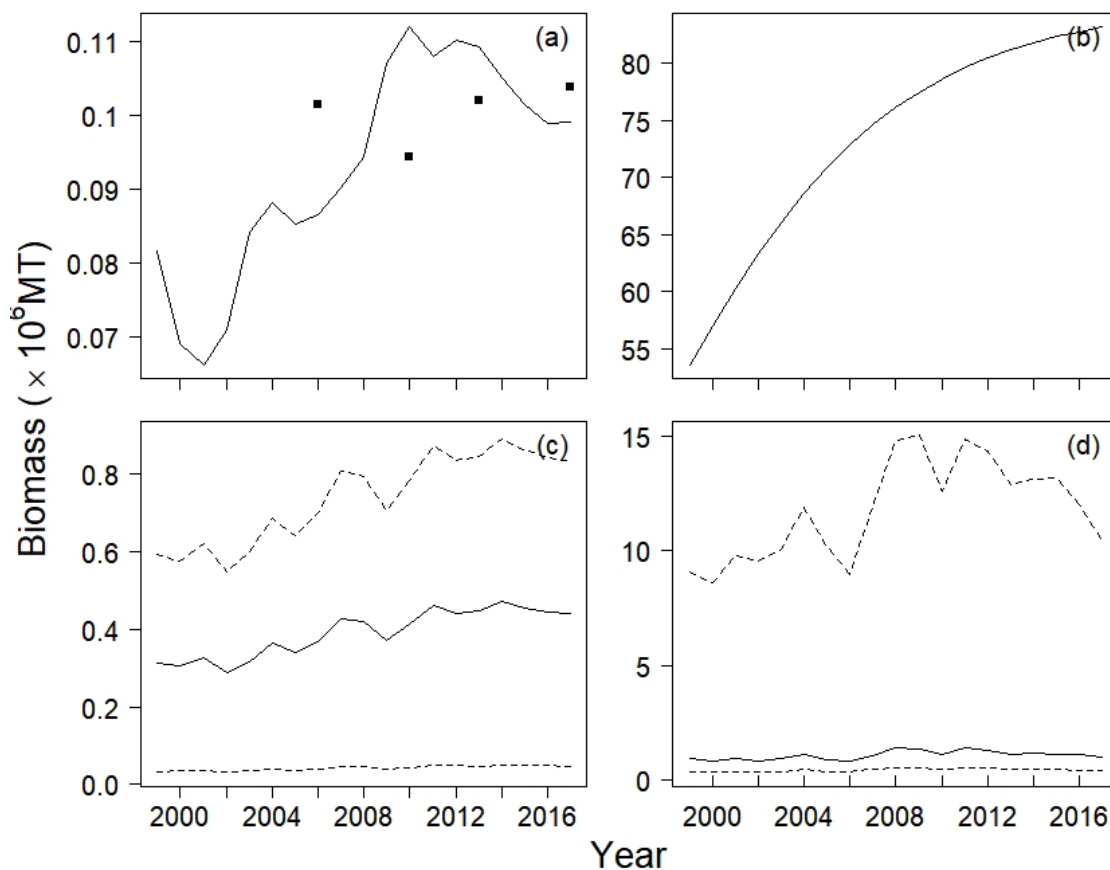


Fig. 6. Plots of predicted biomass from the four production models. The black squares in panel (a) show the yield data, and solid lines in all panels show the predicted biomass. Panels (a–d) show the results from the Yoshimoto-Clarke model, the ASPIC program, the Observation model, and the state-space production model, respectively. Dotted lines along with the solid lines in panel (c) and (d) are 95% confidence intervals and 95% credible intervals, respectively, for predicted biomass.

as a single population. Hiyama et al. (2002) used two data sets, one collected from Korea and one from Japan, and estimated the stock size as a single population in the East China Sea and 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.

State-space models, comprising process and observation equations, explicitly quantify uncertainties arising from model structure and measurement; process errors and observation errors. As mentioned above, the inclusion of process errors in our model increases the number of free parameters, such as annual biomass, and therefore necessitates additional information for parameter estimation (Prager, 1994). Carruthers et al. (2011) recommended the inclusion of prior distributions on r and k for production models, because fisheries data often contain

insufficient information to obtain reliable estimates. Hilborn & Walters (1992) coined the term ‘data failure’ for instances in which the fisheries data include insufficient information about model parameters, leading to unreliable parameter estimates. For example, the negative values for carrying capacity and catchability coefficient in the Fox and Schnute model resulted from insufficient information in the data. In addition to ‘data failure’, the posteriors obtained from SSP also support this point. The posterior distributions of r and k reflected the shape of prior distributions (Fig. 2a, 2b). Posterior distributions of parameters are updated based on priors and likelihood, and the yield and CPUE data on chub mackerel did not provide sufficient information about r and k ; as a result, the posteriors mirrored the corresponding priors. In practice, however, collecting data which provide sufficient information for parameter estimation is rarely possible. Therefore, the term ‘data failure’ can be

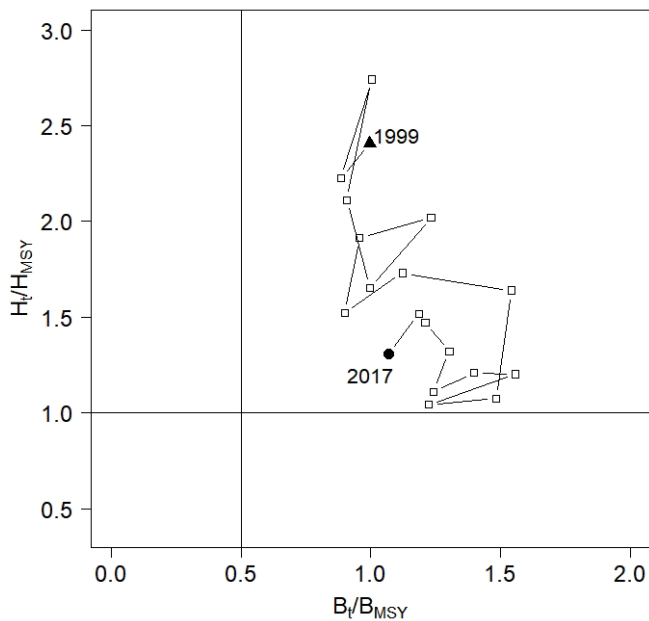


Fig. 7. Kobe plot showing the predicted trajectories of B_t/B_{MSY} and H_t/H_{MSY} from 1999 to 2017. The stock is said to be “overfished” when the biomass, B_t , is smaller than $B_{MSY}/2$, and “overfishing” is said to occur when the harvest rate is larger than H_{MSY} .

translated to ‘model failure’, as models do not accurately depict biomass dynamics (Carruthers et al., 2011; Millar & Meyer, 1999; Musick & Bonfil, 2005; National Research Council, 1998; Pedersen & Berg, 2017).

While the four production models (YC, ASPIC, Observation, and SSP) fitted the CPUE data moderately well, the results showed that the SSP production model should be preferred over the others for explaining CPUE data with fluctuations. In particular, the RSSs showed that the SSP model performed best in explaining the CPUE data. While the RSS suggested that the YC model explained the CPUE data second best among the four models, the YC model failed to account for the yield in 2010 and 2013. The ASPIC program indicated that the stock was severely underexploited throughout the whole duration of the study. To be specific, the mean ratio of the B_t to B_{MSY} was 1.6, and the mean ratio of F_t to F_{MSY} (fishing mortality rate that corresponds to MSY) was 0.02. The ASPIC program may therefore provide an optimistic view of chub mackerel stock, which allows fishermen to exploit the stock more extensively than does the present level. The Observation model provided uncertainties for the predicted CPUE, as did the SSP model. However, the intervals of the model were too narrow to include the CPUE data

(Fig. 5c), which could be addressed by the SSP model (Fig. 5d). The Observation model therefore may consider the observation errors only, whereas the SSP model incorporated both process and observation errors, resulting in wider credible intervals for the predicted CPUE. In short, a SSP model can be a solution for CPUE data with fluctuations.

Priors were a key factor in the implementation of the SSP model. In practice, we used the mode and CV to determine hyperparameters for each model parameter, and then obtained prior distributions through trial and error, to stabilize the numerical optimization. As the priors contain information on parameters, biologically absurd priors were excluded. For example, positive values were chosen for the modes of growth rate r , catchability coefficient q , and P_{1999} . The mode for carrying capacity, k , was only selected when its value was larger than the largest yield. To address our uncertainty regarding priors, we set the CV to be larger than 50% for each parameter. The posterior distributions obtained with the priors and likelihood function in the present study can be used as priors in future studies on Korean chub mackerel. The script software ADMB-RE performed both numerical optimization and MCMC sampling. Although we selected priors that assisted in the numerical optimization of the likelihood function, it is also possible to obtain posterior distributions through the script software without editing its code. Therefore, we recommend taking advantage of this powerful statistical tool for the implementation of Bayesian state-space models.

Conclusions

This study aimed to fit CPUE data with fluctuations, and to assess Korean chub mackerel stock. We applied a SSP model to the annual yield and CPUE data collected from 1999 to 2017, with prior distributions for the model parameters (r , k , q , σ_p^2 , σ_o^2 , P_{1999}), using the software ADMB-RE. We estimated model parameters including intrinsic growth rate r , carrying capacity k , and annual biomass, and provided management references (MSY , B_{MSY} , and H_{MSY}). We concluded that Korean chub mackerel stock have been subjected to overfishing, but were not overfished during 1999–2017, based on the trajectories of H_t/H_{MSY} , and B_t/B_{MSY} . The findings of this study suggest that practitioners should choose a SSP model which considers two sources of variability, and to fit the CPUE data with fluctuations. Further studies into the migration patterns of chub mackerel must be carried out in order to assess the stock as a single popu-

lation fished in the East China Sea.

Competing interests

No potential conflict of interest relevant to this article was reported.

Funding sources

This study was financially supported by the National Institute of Fisheries Science (grant no: R2021032), and National Research Foundation of Korea (grant no: NRF-2019R1I1A2A01052106).

Acknowledgements

The National Institute of Fisheries Science provided the CPUE data, and Statistics Korea offered fishery yield data. Handling editor, Dr. Yong-Woo Lee's advice improved our manuscript.

Availability of data and materials

Upon reasonable request, the datasets of this study can be available from the corresponding author.

Ethics approval and consent to participate

This article does not require IRB/IACUC approval because there are no human and animal participants.

ORCID

Yuri Jung <https://orcid.org/0000-0003-2636-4999>
 Young Il Seo <https://orcid.org/0000-0002-1329-7697>
 Saang-Yoon Hyun <https://orcid.org/0000-0003-2517-5619>

References

- Aeberhard WH, Flemming JM, Nielsen A. Review of state-space models for fisheries science. *Annu Rev Stat Appl.* 2018;5:215-35.
- Carruthers TR, McAllister MK, Taylor NG. Spatial surplus production modeling of Atlantic tunas and billfish. *Ecol Appl.* 2011;21:2734-55.
- Castro Hernández JJ, Ortega Santana AT. Synopsis of biological data on the chub mackerel (*Scomber japonicus* Houttuyn, 1782). Rome: Food and Agriculture Organization of the United Nations; 2000.
- Chaloupka M, Balazs G. Using Bayesian state-space modelling to assess the recovery and harvest potential of the Hawaiian green sea turtle stock. *Ecol Model.* 2007;205:93-109.
- Cho J, Lee JS, Nam J. A study on estimating the fishery optimal production by using a bio-economic model. Busan: Korea Maritime Institute; 2009.
- Choi YM, Zhang CI, Kim YS, Baik CI, Park YC. Ecological characteristics and biomass of chub mackerel, *Scomber japonicus* Houttuyn in Korean waters. *Korean Soc Fish Resour.* 2004a ;7:79-89.
- Choi YM, Zhang CI, Lee JB, Kim JY, Cha HK. Stock assessment and management implications of chub mackerel, *Scomber japonicus* in Korean waters. *Korean Soc Fish Resour.* 2004b;6:90-100.
- Clarke RP, Yoshimoto SS, Pooley SG. A bioeconomic analysis of the northwestern Hawaiian Islands lobster fishery. *Mar Resour Econ.* 1992;7:115-40.
- de Valpine P. Frequentist analysis of hierarchical models for population dynamics and demographic data. *J Ornithol.* 2012;152:393-408.
- Fournier DA, Skaug HJ, Ancheta J, Ianelli J, Magnusson A, Maunder MN, et al. AD model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optim Methods Softw.* 2012;27:233-49.
- Fox WW. An exponential surplus-yield model for optimizing exploited fish populations. *Trans Am Fish Soc.* 1970;99:80-8.
- Haddon M. Modelling and quantitative methods in fisheries. 2nd ed. Boca Raton, FL: CRC Press; 2011.
- Hilborn R. Pretty good yield and exploited fishes. *Mar Policy.* 2010;34:193-6.
- Hilborn R, Walters CJ. Quantitative fisheries stock assessment. Boston, MA: Springer; 1992.
- Hiyama Y, Yoda M, Ohshimo S. Stock size fluctuations in chub mackerel (*Scomber japonicus*) in the East China sea and the Japan/East sea. *Fish Oceanogr.* 2002;11:347-53.
- Jeong M, Nam J. 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.* 2017;39:221-32.
- Kim HA, Seo YI, Cha HK, Kang HJ, Zhang CI. A study on the estimation of potential yield for Korean west coast fisheries using the holistic production method (HPM). *J Korean Soc Fish Technol.* 2018;54:38-53.
- Korean Ministry of Ocean and Fisheries. Management plan for total allowable catch for 11 species in 2017 [Internet]. 2017 [cited 2019 Mar 28]. URL <http://www.mof.go.kr/article/view.do?menuKey=376&boardKey=10&articleKey=14439>
- Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM. TMB:

- automatic differentiation and Laplace approximation. *J Stat Softw.* 2016;70:1-21.
- Kwon Y, Zhang CI, Pyo HD, Seo YI. Comparison of models for estimating surplus productions and methods for estimating their parameters. *J Korean Soc Fish Technol.* 2013;49:18-28.
- Lee HN, Kim HS. Variation of fisheries conditions of mackerel (*Scomber japonicus*) fishing ground for large purse seine fisheries. *Bull Korean Soc Fish Technol.* 2011;47:108-17.
- McAllister MK, Pikitch EK, Babcock EA. Using demographic methods to construct Bayesian priors for the intrinsic rate of increase in the Schaefer model and implications for stock rebuilding. *Can J Fish Aquat Sci.* 2001;58:1871-90.
- McAllister MK, Pikitch EK, Punt AE, Hilborn R. A Bayesian approach to stock assessment and harvest decisions using the sampling/importance resampling algorithm. *Can J Fish Aquat Sci.* 1994;51:2673-87.
- Meyer R, Millar RB. BUGS in Bayesian stock assessments. *Can J Fish Aquat Sci.* 1999;56:1078-87.
- Millar RB, Meyer R. Bayesian state-space modeling of age-structured data: fitting a model is just the beginning. *Can J Fish Aquat Sci.* 2000a;57:43-50.
- Millar RB, Meyer R. Bayesian stock assessment using a nonlinear state-space model. *Can J Fish Aquat Sci.* 1999;56:37-52.
- Millar RB, Meyer R. Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *J R Stat Soc Ser C Appl Stat.* 2000b;49:327-42.
- Musick JA, Bonfil R. Management techniques for elasmobranch fisheries. Rome: Food and Agriculture Organization of the United Nations; 2005.
- National Research Council. Improving fish stock assessments. Washington, DC: National Academies Press; 1998.
- Pedersen MW, Berg CW. A stochastic surplus production model in continuous time. *Fish Fish.* 2017;18:226-43.
- Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News.* 2006;6:7-11.
- Prager MH. A suite of extensions to a nonequilibrium surplus-production model. *Fish Bull.* 1994;92:374-89.
- Prager MH. User's guide for ASPIC suite, version 7: a stock-production model incorporating covariates and auxiliary programs. Portland, OR: Prager Consulting; 2016.
- Punt A. Extending production models to include process error in the population dynamics. *Can J Fish Aquat Sci.* 2003;60:1217-28.
- Quinn TJ, Deriso RB. Quantitative fish dynamics. Oxford: Oxford University Press; 1999.
- Ricker WE. Computation and interpretation of biological statistics of fish populations. *Bull Fish Res Board Can.* 1975;1-382.
- Schaefer MB. Some aspects of the dynamics of populations important to the management of the commercial marine fisheries. *Inter-Am Trop Tuna Comm.* 1954;1:23-56.
- Schnute J. Improved estimates from the Schaefer production model: theoretical considerations. *J Fish Res Board Can.* 1977;34:583-603.
- Skaug H, Fournier D. Random effects in AD model builder, ADMB-RE user guide 11.2. admb-project.org; 2017.
- Walters CJ, Hilborn R. Adaptive control of fishing systems. *J Fish Res Board Can.* 1976;33:145-59.
- Walters CJ, Martell SJD. Fisheries ecology and management. Princeton, NJ: Princeton University Press; 2004.
- Wang Y, Zheng J, Yu C. Stock assessment of chub mackerel (*Scomber japonicus*) in the central East China Sea based on length data. *J Mar Biol Assoc UK.* 2014;94:211-17.
- Winker H, Carvalho F, Kapur M. JABBA: just another Bayesian biomass assessment. *Fish Res.* 2018;204:275-88.
- Yoshimoto SS, Clarke RP. Comparing dynamic versions of the Schaefer and Fox production models and their application to lobster fisheries. *Can J Fish Aquat Sci.* 1993;50:181-89.