Thesis for the Degree of Master of Science

A Bayesian state-space production assessment model for 15 fish stocks in Java Sea, Indonesia

by

Dwi Ramadya Risqiana Putri

Department of Marine Biology

The Graduate School

Pukyong National University

February 2024

A Bayesian state-space production assessment model for 15 fish stocks in Java Sea, Indonesia

[인도네시아 자바해 어류 15 종에 대한 베이지안 상태공간 자원평가 모델]

Advisor: Prof. Saang-Yoon Hyun

By

Dwi Ramadya Risqiana Putri

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

February 2024

A Bayesian state-space production assessment model for 15 fish stocks in Java Sea, Indonesia

A thesis

by

Dwi Ramadya Risqiana Putri

Approved by:

_____

(Chairman) Prof. Hyun-Woo Kim

_____ _____

(Member) Dr. Jinwoo Gim (Member) Prof. Saang-Yoon Hyun

February 16, 2024

Contents

List of figures

List of tables

A Bayesian state-space production assessment model for 15

fish stocks in Java Sea, Indonesia

Dwi Ramadya Risqiana Putri

부경대학교 대학원 해양생물학과

요 약

베이지안 상태 공간 잉여 생산 모델은 일반적으로 어류의 연령과 신체 크기에 대한 정보를 사용할 수 없는 데이터 제한 상황에서 개체군 크기와 어획 수준을 평가하는 데 사용되고 있다. 이 모델에는 어업생산량과 단위 노력당 어획량 (CPUE) 또는 과학 조사 지수 (survey indices)가 필요로 하다. 인도네시아는 어류의 연령과 신체 크기와 같은 정보가 부족하여, 베이지안 상태 공간 잉여 생산 모델을 이용하여, 자원 평가하고자 하였다. 2010 년부터 2021 년까지의 인도네시아 자바해의 15 개 어류 자원을 평가 모델에 적용했다. R 프로그래밍 환경 내의 TMB 패키지를 사용하여 모수를 추정하였다. 본 연구에서는 자바해의 다양한 어족(예: 15 개 어족)의 잉여 생산량을 구현하는 데 중점을 두었다. 자원 평가 결과는 자바해의 15 개 어류 모두에 대한 수확률 (harvest rate)이 MSY(F_{MSY})의 수확률 지점보다 낮고 각 개체군의 바이오매스가 $B_{MSY}/2$ 이상임을 나타낸다. 이는 어류 자원 중 어느 것도 개체군 크기 관점과 수확률 관점에서 남획의 징후를 보이지 않았음을 나타낸다. 이는 수익을 극대화하기 위해서는 해당 개체군의 어획량을 어느 정도 늘릴 수 있도록 허용해야 한다는 의미이다. 인도네시아의 어업 관리가 지나치게 보수적이었을 수 있으며 어부들이 더 많이 어획할 수 있도록 허용함으로써 더 큰 경제적 이익을 얻을 수 있다고 주장한다.

A Bayesian state-space production assessment model for 15 fish stocks in Java Sea, Indonesia

Dwi Ramadya Risqiana Putri

Department of Marine Biology. The Graduate School.
Pukyong National University

Abstract

A Bayesian state-space surplus production assessment model has commonly been used to assess a fish population size, and its associated fishery exploitation level in data-limited situations where information about ages and body sizes of fish is unavailable. The model simply requires two types of time series data: (i) fishery yields; and (ii) catch per unit effort, or survey indices from a scientific survey. I applied the assessment model for 15 fish stocks from Java Sea, Indonesia, whose data ranged from 2010 to 2021. The estimation process was executed using the TMB package within the R programming environment. In this study, I focused on implementing the surplus production to the multiple fish stocks (i.e. 15 fish stocks) from the same ecosystem, namely Java Sea. For future reference, this will be very beneficial to fisheries management. The results indicated that the harvest rate for all 15 fish stocks from the Java Sea was below the point of harvest rate at MSY (F_{MSY}), and their biomass was above $B_{MSY}/2$. This indicates that none of the fish stocks showed signs of being overfished or undergoing overfishing. To maximize

earnings, the implication would be to permit the fisheries for those stocks to somewhat raise their catches. I would argue that Indonesia's fisheries management might have been overly conservative, and greater economic benefits might be achieved by allowing fishermen to harvest larger quantities of fish.

1. Introduction

Seated between the Pacific and Indian Oceans, Indonesia is an archipelagic country in Southeast Asia with a bigger sea area than land area. Indonesia faces significant challenges in managing its fishing resources. Fisheries management has numerous challenges due to the size of the nation and its tens of thousands of islands, including illegal fishing, unreported fishing, weak regulatory frameworks, and many more.

There are many fishing grounds in Indonesia, one of the most active fishing grounds is the Java Sea. Java Sea has long been exploited, both by small-scale fisheries and large-scale fisheries. Java Sea resources make an essential contribution to Indonesian fisheries, being the main supplier of fish protein for the population of Java Island, where most of Indonesian population lives. The main objective of this study is to assess the current status of the 15 fish stocks from Java Sea using hierarchically structured Bayesian state-space models because only two sets of annual CPUE and catch data were available. I selected 15 commercially important fish in Java Sea, (i.e. 7 pelagic fish, 6 demersal fish, 1 crustacean, and 1 invertebrate). I looked at 15 significant commercial fish to see how the Java Sea fisheries are doing in terms of stock management, as this study, which involved various fish species in Indonesia, has never been done before.

Surplus production models are simple but robust non-linear models for stock assessment that are widely used to model biomass dynamics in state-space models for exploited fish populations. A useful feature of surplus production models is that they do not require analytic detailing about specific biological characteristics of target stocks under survey. This is important because detailed information about population dynamics may not be available for analysis of stock sustainability.

State-space models are a class of hierarchical models where the quantity of interest cannot be observed directly, but must be inferred from noisy, transformed observations. Since the state-space model is structured as a hierarchical Bayesian model, posterior distributions can be obtained using Markov Chain Monte Carlo (MCMC) stochastic simulation (Sant'Ana et al., 2017). Therefore, in this study, a Bayesian state-space modeling approach was employed to fit a stochastic population dynamics model for 15 fish stocks in the Java Sea, using available data on commercial harvest history and relative biomass indices.

2. Materials and Methods

2.1. Data on 15 Fishery stocks in Java Sea, Indonesia

The total catch data (Yield) on 15 fish stocks from 2010 to 2021 and the catch per unit effort (CPUE) for each stock during the same period, expressed in metric tons (MT) and metric tons per number of fishing vessels (MT/number of fishing vessels), respectively, were employed in this study. Population abundance was estimated using CPUE data, which is a relative biomass index (Hilborn & Walters, 2013). Due to the unavailability of other data such as fishing hours or haul totals, the number of fishing vessels was utilized as a proxy for fishing effort.

The data were obtained from the Directorate of Fish Resource Management, Indonesian Ministry of Marine Affairs and Fisheries. The yearly yield from 2010 to 2021 is displayed in Figure 1, and the CPUE of 15 commercial fish stocks is displayed in Figure 2 for the same period.

The following table presents a list of the 15 fish stocks assessed in this study, along with their scientific names and groups. Information regarding the scientific names and groups was sourced from FishBase (for fish) (https://www.fishbase.se/search.php) and SeaLifeBase (for crustacean and cephalopod) (https://www.sealifebase.ca/). These species are the main

commodities of Java Sea fisheries, primarily targeted for both export and domestic consumption.

Table 1. List of 15 fish stocks.

Fish Stock	Scientific Name	Group
Indian anchovy	*Stolephorus indicus*	Pelagic
Short mackerel	*Rastrelliger brachysoma*	Pelagic
Mackerel scad	Decapterus macarellus	Pelagic
Yellowtail scad	*Atule mate*	Pelagic
Goldstripe sardinella	*Sardinella gibbosa*	Pelagic
Longtail tuna	*Thunnus tonggol*	Pelagic
Narrow-barred Spanish mackerel	*Scomberomorus commerson*	Pelagic
Giant sea catfish	*Ariidae*	Demersal
Black pomfret	*Parastromateus niger*	Demersal
Pony fishes	*Leiognathus equulus*	Demersal
Red snappers	*Lutjanus bitaeniatus*	Demersal
Ornate threadfin bream	*Nemipterus hexodon*	Demersal
Hairtails	*Trichiurus lepturus*	Demersal
White shrimp	*Penaeus indicus*	Crustaceans
Common squid	*Todarodes pacificus*	Invertebrates

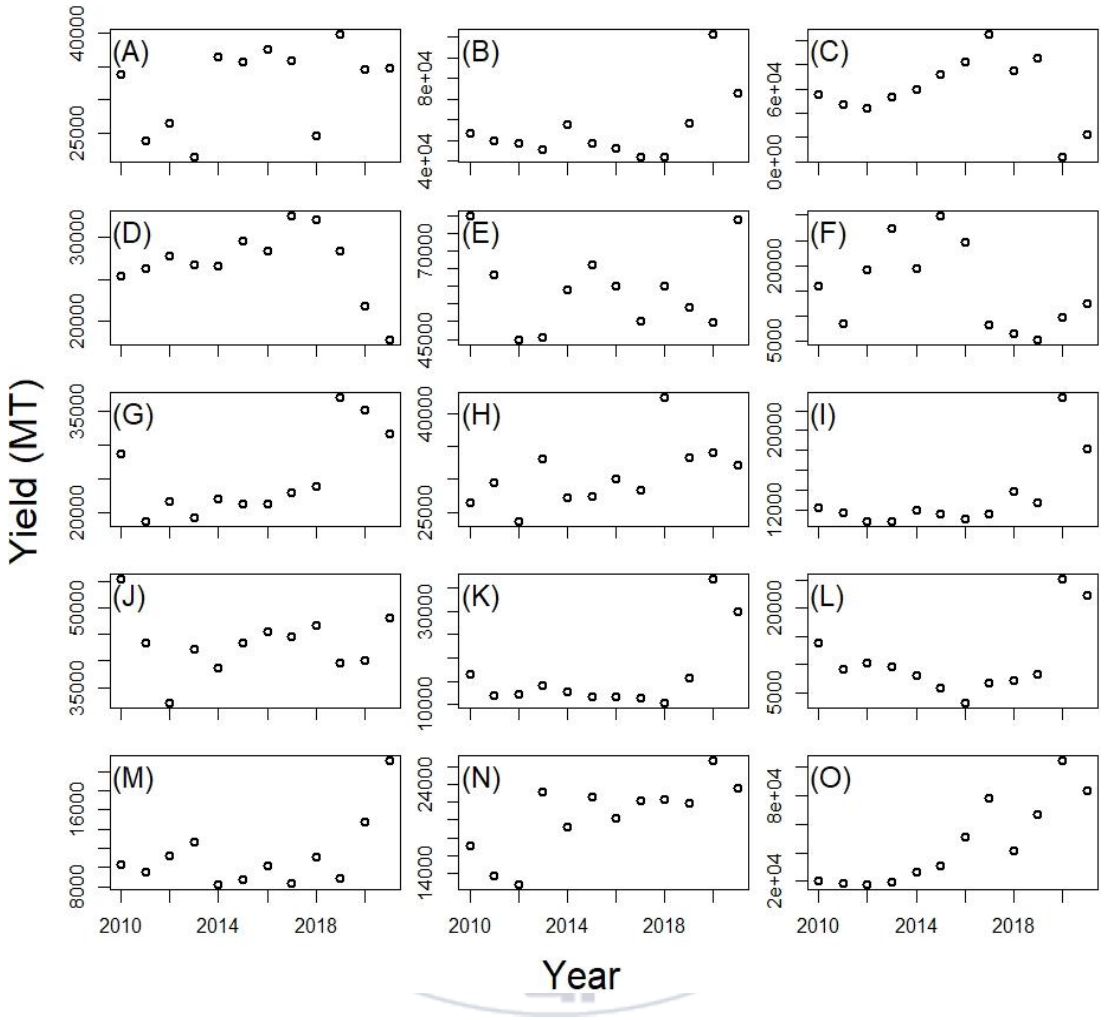

Figure 1. The annual yield in metric tons (MT) of 15 fish stocks from 2010 to 2021. Panel (A) show the yield of Indian anchovy, panel (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna, (G) Narrow-bared Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret, (J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

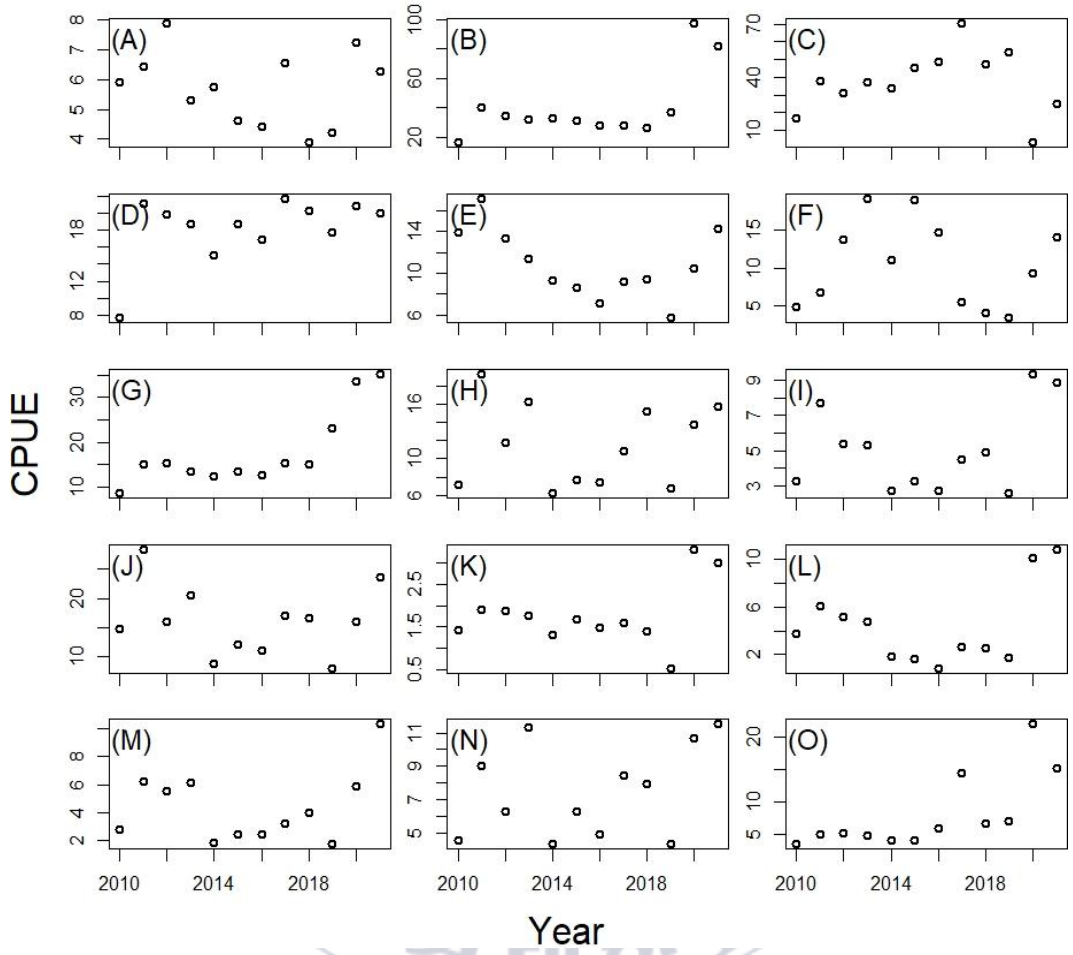

Figure 2. The annual catch per unit effort (CPUE) data in MT per the number of fishing vessels about 15 fish stocks from 2010 to 2021. Panel (A) show CPUE data of Indian anchovy, panel (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna, (G) Narrow-bared Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret, (J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

2.2. Surplus production model

Surplus production models are the only method for stock assessment in situations in which the only data available are time series of catches (i.e. yield) and some index of abundance (i.e. CPUE) (Punt, 2003). Due to their fewer parameters compared to other models, surplus production models are popular and widely used in practical stock assessments (Graham, 1935; Meyer & Millar, 1999; Millar & Meyer, 2000; Polacheck et al., 1993; Prager, 1994, 2002; Schaefer, 1954).

Surplus production models are employed to assess the biomass and exploitation levels of marine populations in situations with limited data, particularly where age and size information is unavailable (Jiao et al., 2011; Marandel et al., 2016; Punt, 2003). Generally, these models utilize catches per unit effort (CPUE) data and catch data on target species to estimate population abundances. Determining the stock size for the upcoming year is calculated by adding the current year's biomass to the surplus production and then subtracted by the catch amount in the current year.

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

In equation (1), B_t denotes the biomass quantity at time *t*, $f(B_t)$ is representing the surplus production at time *t*, and Y_t represents the catch at time

t. This study employed the discrete form of the Schaefer surplus production model.

$$B_{t+1} = B_t + r \cdot B_t \left(1 - \frac{B_t}{K}\right) - Y_t \quad (2)$$

$$I_t = q \cdot B_t \quad (3)$$

The Schaefer model is applied to the model in equation (2), where *K* is the carrying capacity and *r* is the intrinsic growth rate. Equation (3) is derived from the assumption that CPUE and biomass have a direct proportional relationship. The variables I_t and q represent the amount of catch per unit effort (CPUE) collected at time *t* and catchability coefficient, respectively. These notations are detailed in Table 2.

2.3. A state-space production model

A state-space model describes the temporal dynamics of two linked processes that may include error in either process: (1) a state process that describes the unobserved population dynamics in terms of biomass or abundance and (2) an observation process based on population-specific survey data that are a function of the unobserved state process (Buckland et al., 2004).

The state-space production model has both measurement sampling (observation error) and population dynamics (process error).

$$B_1 = b \cdot K \cdot exp(\varepsilon_t^p), \qquad \text{where } \varepsilon_t^p \sim N(0, \sigma_p^2) \tag{4}$$

$$B_{t+1} = \left[B_t + r \cdot B_t \left(1 - \frac{B_t}{K}\right) - Y_t\right] \cdot exp(\varepsilon_t^p), \text{ where } \varepsilon_t^p \sim N(0, \sigma_p^2) \tag{5}$$

$$I_t = q \cdot B_t \cdot exp(\varepsilon_t^o), \qquad \text{where } \varepsilon_t^o \sim N(0, \sigma_o^2) \tag{6}$$

The process equations are (4) and (5), whereas the observation equation is equation (6). The terms "process error" and "observation error," respectively, are represented by the variables ε_t^p and ε_t^o. It is presumed that they are independent and identically distributed random variables having a normal (Gaussian) probability density with variances of σ_p^2 and σ_o^2, respectively, and a mean of 0. Equation (4) defines the initial biomass (B_1) as a value that takes into account the carrying capacity (K), the constant (b), and the process error.

The unknown population size (B_t) can be regarded as a state variable or as a random effects when taking the process error into account. Assuming B_t to be a state variable, both point estimates and the associated uncertainties of B_t were calculated. On the other hand, the observation errors or measurement errors in the catch per unit effort data were considered.

However, in this study, to ensure the stability of numerical optimization, it was assumed that the variance of the process error (σ_p^2) and the variance of the observation error (σ_o^2) were equal ($\sigma_p^2 = \sigma_o^2$), and this common value was subsequently denoted as (σ^2).

$$B_1 = b \cdot K \cdot exp(\varepsilon_t^p), \qquad \text{where } \varepsilon_t^p \sim N(0, \sigma^2) \tag{7}$$

$$B_{t+1} = \left[B_t + r \cdot B_t \left(1 - \frac{B_t}{K}\right) - Y_t\right] \cdot exp(\varepsilon_t^p), \text{ where } \varepsilon_t^p \sim N(0, \sigma^2) \tag{8}$$

$$I_t = q \cdot B_t \cdot exp(\varepsilon_t^o), \qquad \text{where } \varepsilon_t^o \sim N(0, \sigma^2) \tag{9}$$

Therefore, the assumption was made that the variances of both the process error and the observation error are equal (σ^2), as demonstrated in the equations above. Since the auto-correlation may exist in the state variable B_t, the model was restructured by dividing B_t the state variable by the carrying capacity ($P_t \equiv B_t/K$).

$$P_1 = b \cdot exp(\varepsilon_t^p), \qquad \text{where } \varepsilon_t^p \sim N(0, \sigma^2) \tag{10}$$

$$P_{t+1} = \left[P_t + r \cdot P_t(1 - P_t) - \frac{Y_t}{K}\right] \cdot exp(\varepsilon_t^p), \text{ where } \varepsilon_t^p \sim N(0, \sigma^2) \tag{11}$$

$$I_t = q \cdot K \cdot P_t \cdot exp(\varepsilon_t^o), \qquad \text{where } \varepsilon_t^o \sim N(0, \sigma^2) \tag{12}$$

Table 2. Notations. Note that NA (not applicable/no dimension), MT (metric ton).

Notation	Description	Dimension
t	Time measured in discrete units	Year
Y_t	The number of fishery yield (catch) at time t	MT
I_t	Catch per unit effort data at time t	MT/number of fishing vessel
b	Coefficient for estimating the initial biomass: B_1 (*i.e.*, $B_1 = b\,K$)	NA
K	Carrying capacity	MT
r	Intrinsic growth rate	NA
q	Catchability coefficient	1/number of fishing vessel
σ_p^2	Variance of the process error	NA
σ_o^2	Variance of the observation error	NA
B_t	Biomass at time t	MT
P_t	Rescaled population size $\left(P_t \equiv \frac{B_t}{K}\right)$	NA

2.4. Likelihood functions

I constructed likelihood functions for the process equation and observation equation described in equations (10), (11), and (12) above. By taking log to both sides, the term $logP_1, logP_{t+1},$ and $logI_t$ follow a normally distribution, as shown in the following equations.

$$logP_1 \sim N(log(b), \sigma^2) \tag{13}$$

$$logP_{t+1} \sim N\left(log\left[P_t + r \cdot P_t(1 - P_t) - \frac{Y_t}{K}\right], \sigma^2\right) \tag{14}$$

$$logI_t \sim N(log[q \cdot K \cdot P_t], \sigma^2) \tag{15}$$

Equations (13), (14), and (15) together constitute $L(\boldsymbol{\theta}, \boldsymbol{P}|\boldsymbol{I}, \boldsymbol{Y})$, which is the joint log-likelihood encompassing both random effects and parameters. Within this framework, the parameters are denoted as $\boldsymbol{\theta} = (r, K, \sigma^2, b, q)$, while the random effects are presented as $\mathbf{P} = (P_1, P_2, \cdots, P_{n+1})$. The data are symbolized by **Y** and **I**, where **Y** consists of $(Y_{2010}, Y_{2011}, \cdots, Y_{2021})$ and **I** comprises $(I_{2010}, I_{2011}, \cdots, I_{2021})$.

$$logL(\boldsymbol{\theta}, \boldsymbol{P}|\boldsymbol{I}, \boldsymbol{Y}) = log[N(logP_1|logb, \sigma^2)]$$

$$+\sum_{t=1}^{n} log\left[N\left(logP_{t+1}\middle|log\left[P_t + r \cdot P_t(1 - P_t) - \frac{Y_t}{K}\right], \sigma^2\right)\right]$$

$$+\sum_{t=1}^{n} log[N(logI_t|log(q \cdot K \cdot P_t), \sigma^2)] \quad (16)$$

2.5. Prior distributions

The Bayesian methodology views the parameters to be estimated as probability distributions rather than constants, in contrast to classical statistics, which treats the parameters to be estimated as unknown constants. Prior distributions contribute vital information, assist in model estimation, and help reduce the uncertainty of estimates.

In this study, intrinsic growth rate (r) and carrying capacity (K) were assigned informative prior distributions. Other parameters were assigned non-informative prior distributions, as detailed in Table 3. Under the assumption that the prior distributions in this study are mutually independent, the expression for the joint prior distribution is formulated as the following equation.

$$p(r, K) = p(r) \cdot p(K)$$

The process of numerical optimization was conducted through the TMB package in the R software environment, leading to the estimation of both random and fixed effects. The marginal likelihood function for $\theta = (r, K, \sigma^2, b, q)$ is provided by the Laplace approximation. Following the principles of Bayes' theorem, the joint posterior distribution can be expressed as a function of the product of the joint likelihood function and joint prior distribution.

$$p(\boldsymbol{\theta}, \boldsymbol{P}|\boldsymbol{I}, \boldsymbol{Y}) \propto L(\boldsymbol{I}, \boldsymbol{Y}|\boldsymbol{\theta}, \boldsymbol{P}) \cdot p(r, K) \quad (17)$$

The process of MCMC sampling was carried out using the "tmbstan" R package, applied to the model object from "TMB" using "Stan". A total of 50,000 MCMC iterations were executed across four chains, with the initial 5,000 samples from each chain being discarded as a burn-in phase. To diminish autocorrelation among the parameter samples, thinning was applied, selecting every 100th sample.

After establishing the posterior distribution of the parameters, a diagnostic process is required to check whether the resultant MCMC draws were satisfactory or not. This diagnostic process was carried out using the R package "tmbstan" and "rstan". In this study, two criteria were employed to judge the satisfactory: e.g., the number of divergent transitions and the potential-scale reduction statistic $(\hat{R})$ which are standard diagnostic measures. MCMC posterior samples that failed to meet the diagnostic criteria were eliminated from consideration. Discarded set of MCMC draws which did not lead to no divergent transition which is 0, and the potential-scale reduction statistic $(\hat{R})$ must be around 1 (Best & Punt, 2020; Hyun & Kim, 2022; Lang, 1999). An $\hat{R}$ value around 1 indicate the variance within each chain about same as the variance across the chains, which means the chains have converged to build posterior distribution through MCMC sampling.

Table 3. Prior distributions of parameter

Parameter	Prior Distribution
K	Log-normal (13.19 , 0.47)
r	Log-normal (0.27 , 0.47)
q	Non-informative
b	Non-informative
σ^2	Non-informative

2.6. Determining informative prior *r* and prior *K*

In the Bayesian framework, the use of informative priors for parameters is essential. This study implemented such informative priors for the intrinsic growth rate (*r*) and the carrying capacity (*K*). The construction of prior distributions in this study was based on the fundamental concepts presented in (Froese et al., 2017). For establishing the prior on the intrinsic growth rate (*r*) of the species being evaluated, I used resilience information of the species as provided in FishBase. This information was then converted into *r*-ranges, as detailed in Table 4.

Based on prior ranges of the intrinsic growth rate (*r*) presented in Table 4, a prior density function of the corresponding parameter *r* was derived. First, the intrinsic growth rate (r) was assumed to follow a log-normal distribution, where the mean of its prior range was treated as the mode of this prior distribution. The coefficient of variation (CV) for the prior r was set at 0.5, which corresponds to 50%.

Subsequently, the prior range for the carrying capacity parameter (*K*) was established, based on the *r*-range previously outlined in Table 4.

$$K_{low} = \frac{\max(C)}{r_{high}}, K_{high} = \frac{4\max(C)}{r_{low}} \quad (18)$$

$$K_{low} = \frac{2\max(C)}{r_{high}}, K_{high} = \frac{12\max(C)}{r_{low}} \quad (19)$$

To establish the prior range for *K*, Equations (18) and (19) were applied. According to (Froese et al., 2017) it assumed that, if the last catch data in the time series is low relative to the maximum catch in the time series, use equation (18) to construct k-range. On the other hand, if the last catch in the time series is high relative to the maximum catch in the time series, use equation (19) to determine prior *K*-ranges. In this context, k_{low} and k_{high} represent the lower and upper limits, respectively, of the prior ranges for *K*. The term $\max(C)$ refers to the maximum catch in the time series, while r_{low} and r_{high} denote the lower and upper boundaries of the *r*-ranges in the Table 4.

The prior for *K* was presumed to have a log-normal distribution, wherein the mean of the prior range was treated as the mode of this distribution. Furthermore, the coefficient of variation (CV) for this prior *K* was set at 0.5, corresponding to 50%.

Table 4. Prior *r*-range based on classification of resilience in FishBase (Froese et al., 2017).

Resilience	Prior *r*-range
High	0.6 – 1.5
Medium	0.2 – 0.8
Low	0.05 – 0.5
Very low	0.015 – 0.1

3. Results

3.1. State space production model for 15 fish stocks

In term of conservation, the results of this study revealed that the 15 evaluated fish stocks are currently in a safe condition. The Maximum Sustainable Yield (MSY) points for these stocks were far above the annual yield. Detailed point estimates for the parameters concerning the 15 fish stocks are systematically displayed in Table 5. Successful estimation using Bayesian methods with the informative priors *r* and *K*, most of the resultant MCMC draws were satisfactory. The satisfactory percentage exceeded 87% for all evaluated stocks, with four fish stocks achieving perfect satisfaction, six fish stocks demonstrated approximately 99% satisfaction, three fish stocks were about 98% satisfied, one stock reached roughly 93% and on stock recorded 88% satisfaction, as detailed in Table 6. MCMC draw sets that did not pass the diagnostic test, specifically those with divergent transitions, were excluded. The potential-scale reduction statistic $(\hat{R})$ values for all 15 fish stocks consistently hovered around 1.0 (the range limits, 0.9999 – 1.006 rounded to 1.0). This indicates that all 15 fish stocks met the diagnostic criteria, signifying that the MCMC process for constructing the posterior distribution was successfully executed.

The observed annual CPUE were juxtaposed with the predicted CPUE indices derived from the model fit, as a measure of goodness of fit, depicted in Figure 3. This comparison serves to demonstrate the model's accuracy in fitting the observed data, with a closer alignment between predicted and observed data indicating a more precise numerical optimization process. The predicted CPUE for the 15 fish stocks closely approximated the observed CPUE time series, indicating effective model performance. Management reference points MSY, B_{MSY}, and F_{MSY} are presented in Figures 4, 5, and 6, respectively. Furthermore, the prior and posterior distributions for the parameters (r, K, q, b, σ^2) are illustrated in Figure 7 (7a, 7b, 7c).

Overall, the Maximum Sustainable Yield (MSY) levels for the 15 evaluated fish stocks were deemed satisfactory. However, exceptions as shown in Figure 4 include: the Short mackerel (panel B) and the Goldstripe sardinella (panel E), where the yields in 2020, 2010, and 2021 respectively exceeded the MSY levels; and the Common squid (panel O) in 2020, where the yield trajectory initially surpassed MSY before declining the next year. Additionally, Figure 6 indicates that in 2010, the annual harvest rates (F_{MSY}) for the Short mackerel (panel B) and the Narrow-barred Spanish mackerel (panel G), both part of the Scombridae family, also exceeded their respective MSY levels.

Figures 7a, 7b, and 7c display the prior distributions (depicted as curve lines) and the posterior distributions (represented by histograms) of parameters.

We could see that all posterior distributions assumed a unimodal form, indicating successful convergence of the MCMC sampling and effective construction of the posterior distribution. Additionally, the posterior distributions for the parameters r and K closely resembled the shape of their respective priors. Figure 8 illustrates the range of $\hat{R}$ values for the MCMC draws that met the satisfactory criteria, clearly indicating their adequacy, as the $\hat{R}$ values were predominantly around 1.

Table 5. Estimates of parameters in state space production models for 15 fish stocks (r, K, q, b, σ^2). Note that the units for K is metric ton (MT), r, b, σ^2 are dimensionless, and q is 1/the number of fishing vessels.

Fish Stock	Parameters				
	r	K	q	b	σ^2
Indian anchovy	1.36	1.77×10^5	3.83×10^{-5}	0.87	0.15
Short mackerel	0.88	3.79×10^5	2.36×10^{-4}	0.25	0.26
Mackerel scad	1.66	4.29×10^5	1.15×10^{-4}	0.37	0.32
Yellowtail scad	1.94	1.52×10^5	1.42×10^{-4}	0.41	0.1
Goldstripe sardinella	0.92	2.98×10^5	5.24×10^{-5}	0.94	0.18
Longtail tuna	0.29	1.24×10^6	8.68×10^{-6}	0.53	0.35
Narrow-barred Spanish mackerel	0.45	4.21×10^5	1.03×10^{-4}	0.24	0.16
Giant sea catfish	0.6	4.7×10^5	2.67×10^{-5}	0.72	0.28
Black pomfret	0.53	2.56×10^5	2.25×10^{-5}	0.7	0.31
Pony fishes	1.38	2.22×10^5	9.04×10^{-5}	0.78	0.26
Red snappers	0.54	4.07×10^5	4.5×10^{-6}	0.84	0.34
Ornate threadfin bream	0.49	2.7×10^5	1.83×10^{-5}	0.87	0.46
Hairtails	0.54	2.29×10^5	2.09×10^{-5}	0.72	0.36
White shrimp	1.26	1.08×10^5	8.92×10^{-5}	0.53	0.24
Common squid	0.83	4.03×10^5	3.88×10^{-5}	0.2	0.31

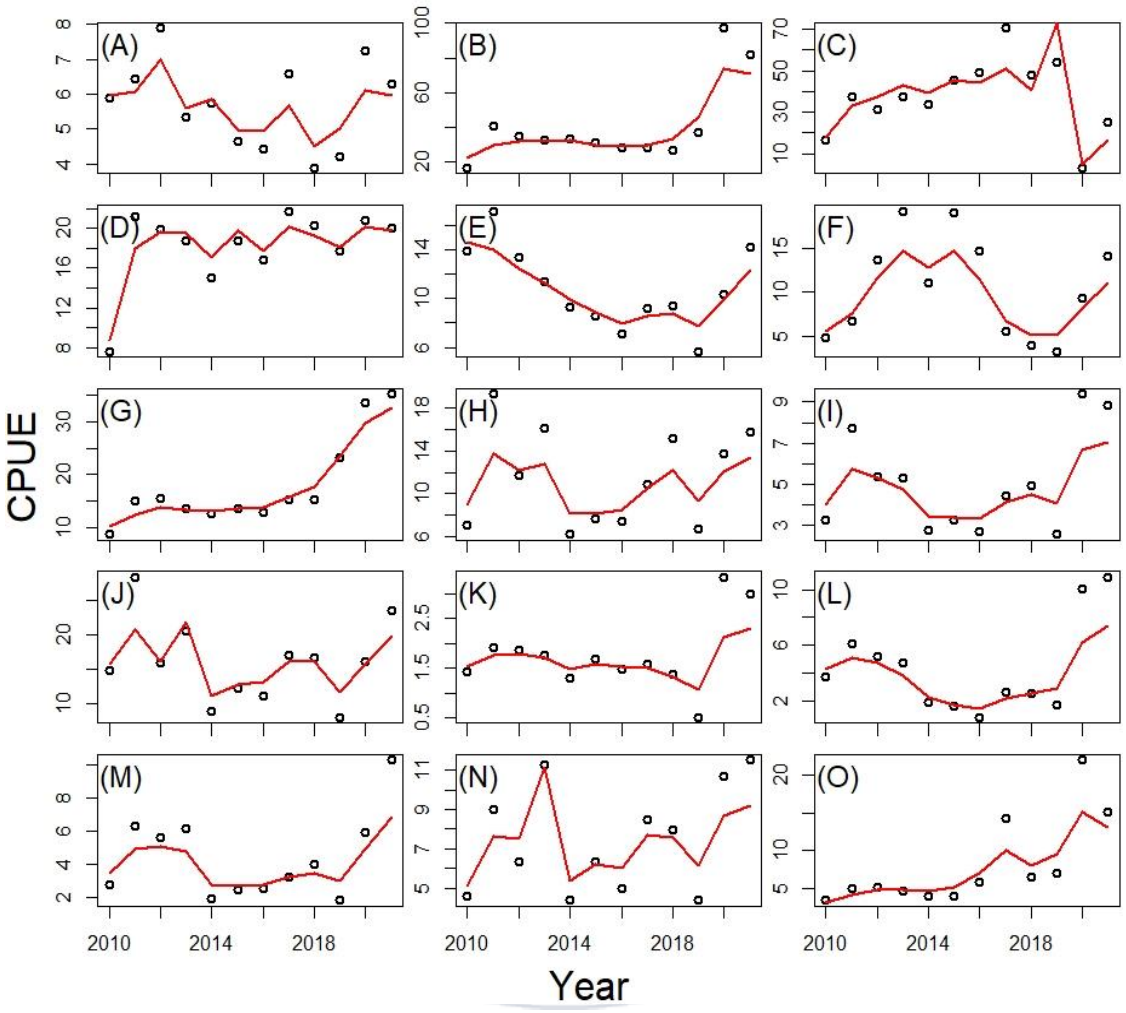

Figure 3. Goodness of fit graphs of 15 fish stocks. Comparison between observed CPUE and predicted CPUE shows how fit the model is to the observed data. The black circles are the observed CPUE and the red lines are the predicted CPUE by the model. Panel (A) Indian anchovy, (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna, (G) Narrow-bared Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret,

(J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

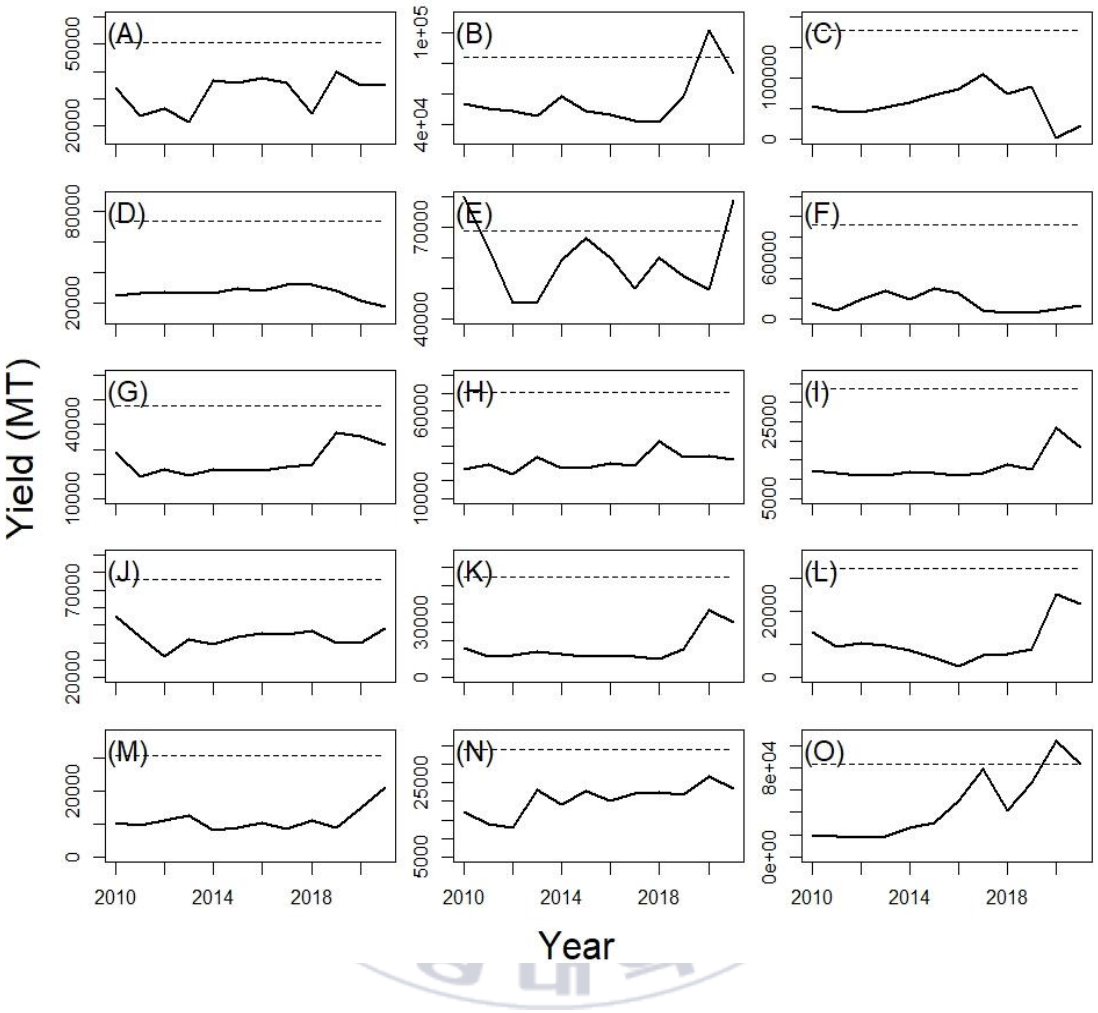

Figure 4. Maximum sustainable yield points of 15 fish stocks. (MSY, dashed line) with the annual yield (solid black line). Panel (A) Indian anchovy, (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna, (G) Narrow-barred Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret, (J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

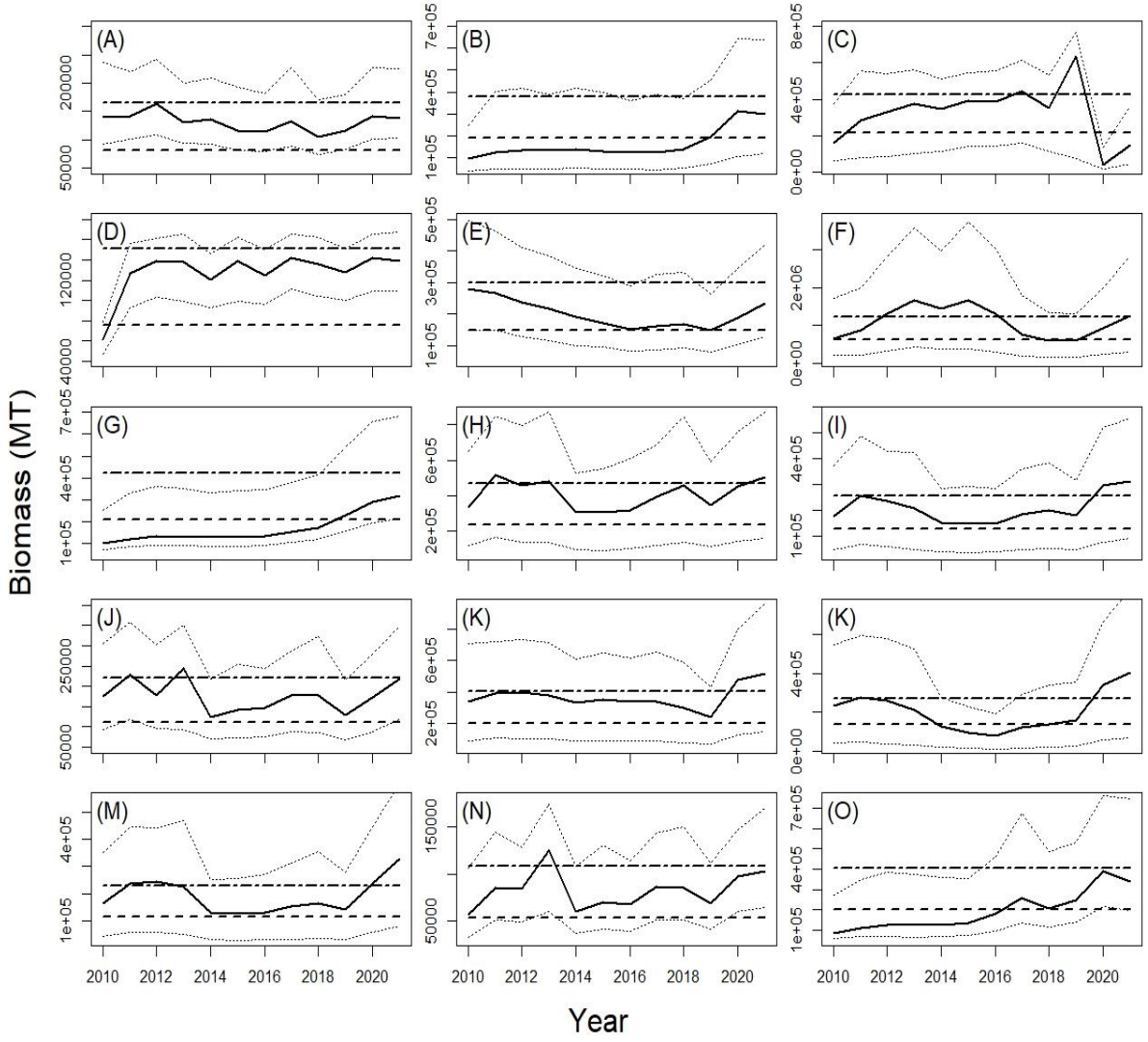

Figure 5. Estimated annual biomass of 15 fish stocks. The graphs above depict the predicted annual biomass (solid black line) with carrying capacity (*K*, two-dashed line), biomass at MSY (B_{MSY}, dashed line) and 95% credible intervals (dotted lines). Panel (A) Indian anchovy, (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna, (G)

Narrow-bared Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret, (J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

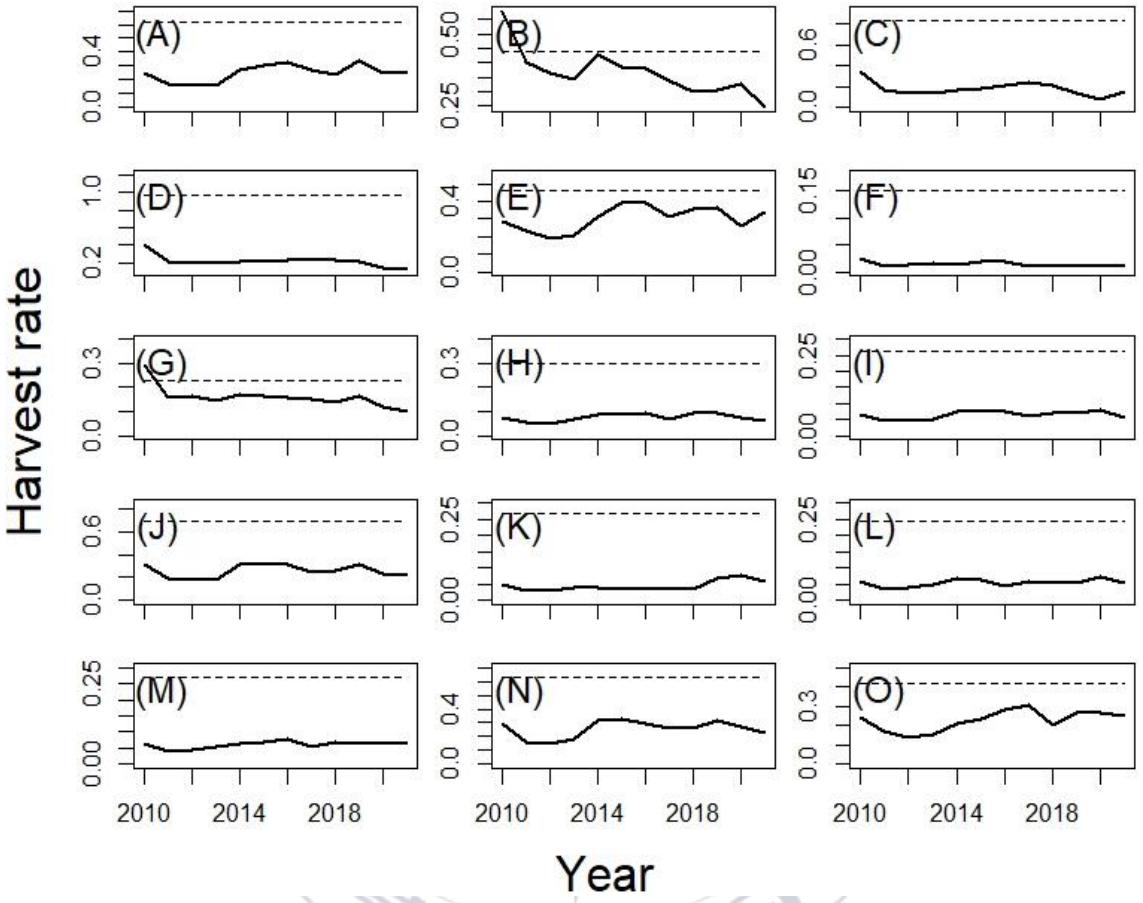

Figure 6. Estimated annual harvest rate of 15 fish stocks. The annual harvest rate (solid black line) and the harvest rate that corresponds to MSY (F_{MSY}, dashed line). Panel (A) Indian anchovy, (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna, (G) Narrow-bared Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret, (J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

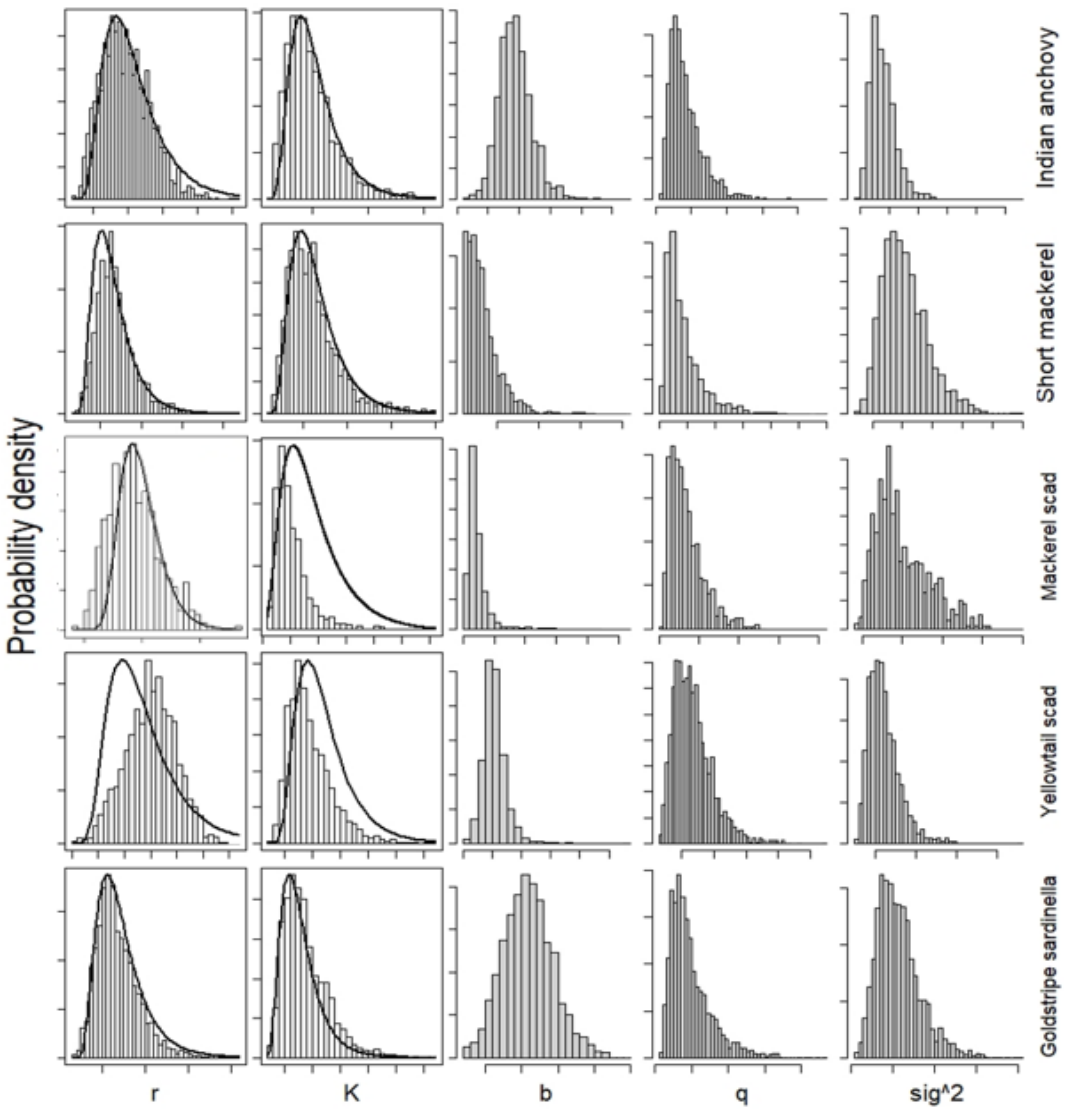

Figure 7a. Prior distributions (curve lines, only r and K) and posterior distributions (histogram) of parameters (r, K, b, q, σ^2) of Indian anchovy (first row), Short mackerel (second row), Mackerel scad (third row), Yellowtail scad (fourth row), and Goldstripe sardinella (fifth row).

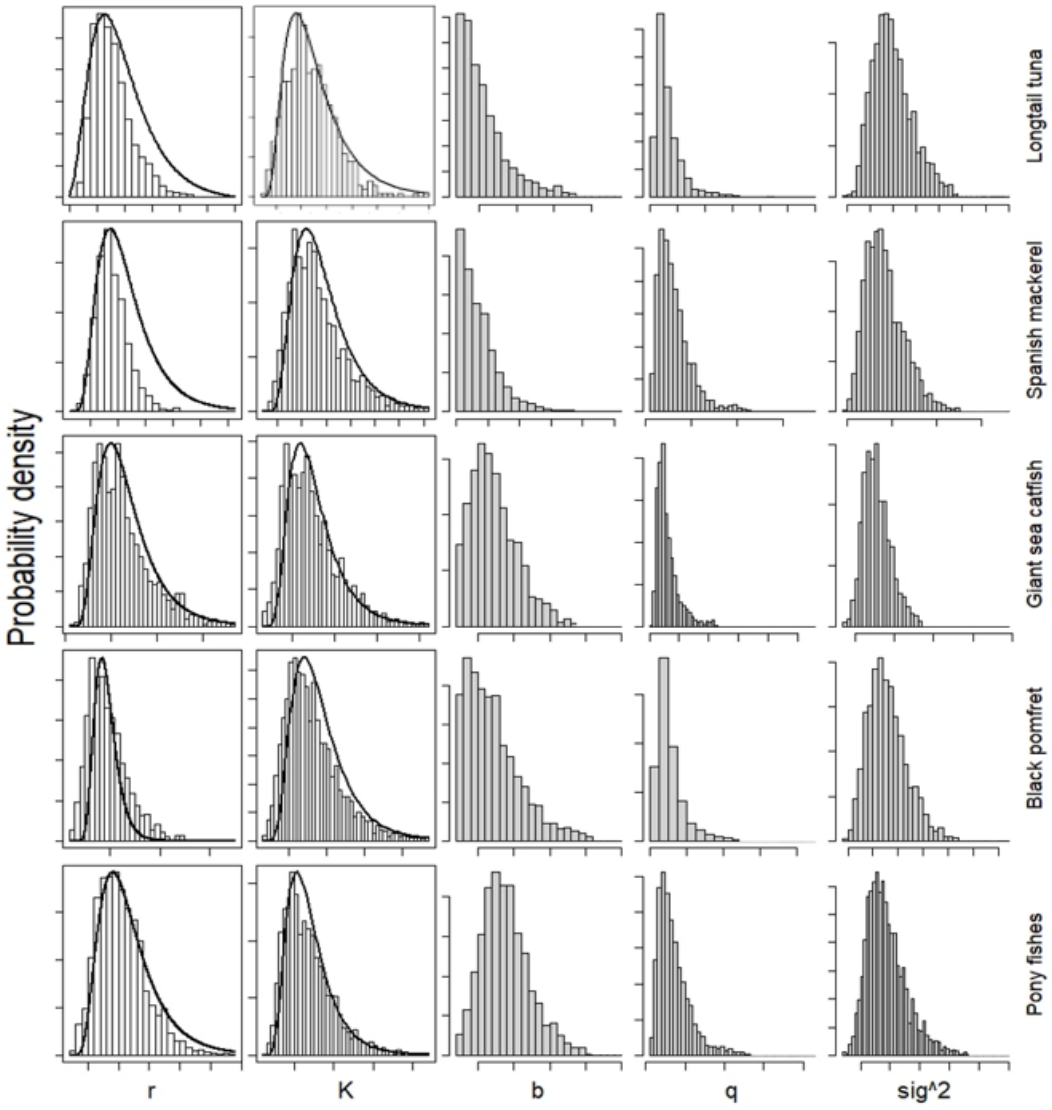

Figure 7b. Prior distributions (curve lines, only r and K) and posterior distributions (histogram) of parameters (r, K, b, q, σ^2) of Longtail tuna (first row), Narrow-barred Spanish mackerel (second row), Giant sea catfish (third row), Black pomfret (fourth row), and Pony fishes (fifth row).

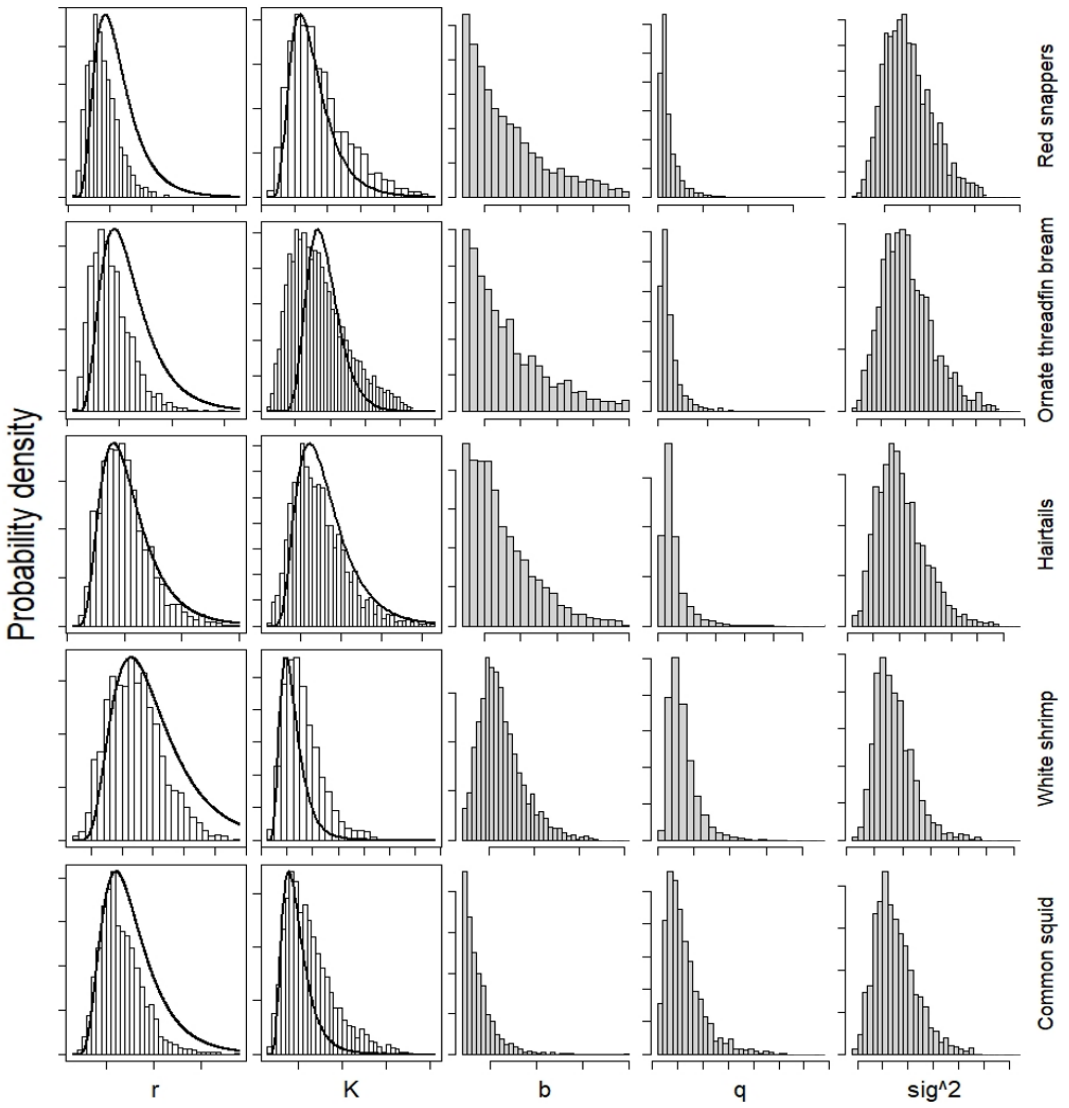

Figure 7c. Prior distributions (curve lines, only r and K) and posterior distributions (histogram) of parameters (r, K, b, q, σ^2) of Red snappers (first row), Ornate threadfin bream (second row), Hairtails (third row), White shrimp (fourth row), and Common squid. (fifth row).

Table 6. Successful convergence rate of MCMC sample.

Indian anchovy	Short mackerel	Mackerel scad	Yellowtail scad	Goldstripe sadrdinella	Longtail tuna	Narrow-bared Spanish mackerel
100%	99.3%	88%	99.1%	100%	99.8%	100%

Giant sea catfish	Black pomfret	Pony fishes	Red snappers	Ornate threadbream	Hairtails	White shrimp	Common squid
100%	99.8%	99.1%	98.2%	93.8%	98%	99.8%	98%

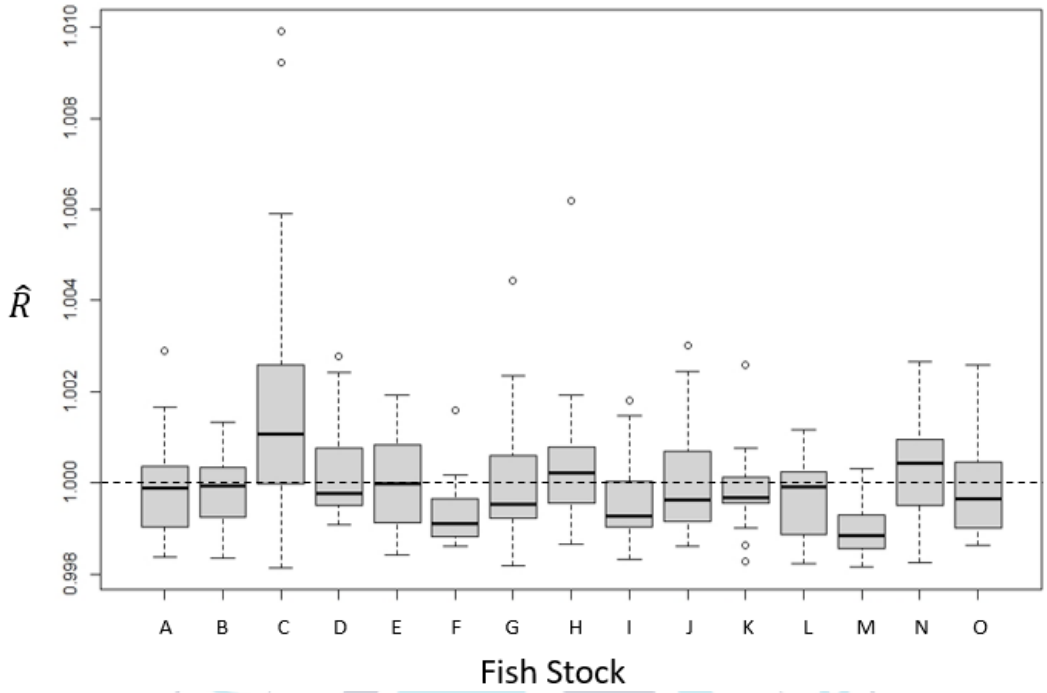

Figure 8. Potential-scale-reduction statistics ($\hat{R}$) of the satisfactory MCMC draws. The limit of the range $\hat{R}$ (0.998 to 1.01 are rounded to 1.0). The numbers on the x-axis represent the fish stock under assessment (i.e. A. Indian anchovy, B. Short mackerel, C. Mackerel scad, D. Yellowtail scad, E. Goldstripe sardinella, F. Longtail tuna, G. Narrow-barred Spanish mackerel, H. Giant sea catfish, I. Black pomfret, J. Pony fishes, K. Red snappers, L. Ornate threadfin bream, M. Hairtails, N. White shrimp, O. Common squid).

4. Discussion

4.1. Resilience level

Resilience denotes the ability of a species to endure and maintain its population despite disruptions or challenges arising from environmental factors. Information on resilience can be sourced from FishBase, where each species is classified according to its level of resilience. This classification encompasses four distinct resilience categories: high, medium, low, and very low. The resilience information is very helpful for the management perspective (i.e. as information which is then transformed into a prior distribution based on Froese's theory). Species classified as endangered often exhibit very low resilience levels, possibly indicating their reduced capacity to cope with surrounding environmental disturbances. In this research, I used the resilience levels, converting them into prior information, as elaborated in detail in the methodology section.

4.2. Management perspective

As per NOAA Fisheries (National Oceanic and Atmospheric Administration), stock status can be categorized into two distinct concepts: "overfishing" and "overfished." Overfishing refers to a situation in which the harvest rate (*F*) of a stock exceeds the Maximum Sustainable Yield's harvest rate (F_{MSY}). In contrast, a stock is considered “overfished” when its biomass (B) falls below half the biomass level that generates the Maximum Sustainable Yield (B_{MSY}): $B_t < {}^{B_{MSY}}/_{2}$. The trajectories of the harvest rate, as depicted in Figure 6, distinctly indicate that for the current year, all the stocks were below the Maximum Sustainable Yield's (MSY) harvest rate point. Regarding the assessment of being overfished, as shown in Figure 9, it is observed that the biomass trajectories remain above ${}^{B_{MSY}}/_{2}$.

Considering the previously discussed concepts, in the current year (the terminal year of time series) there are no signs of either overfishing or stocks being overfished among the 15 fish stocks in the Java Sea. Consequently, a more advantageous management approach might be to moderately increase catch levels to optimize profits. Each specific area possesses unique characteristics, necessitating tailored solutions for effective management. Factors such as regulations, local conditions, country-specific nuances, and others, should be taken into account by managers when determining the most suitable strategy for each area.

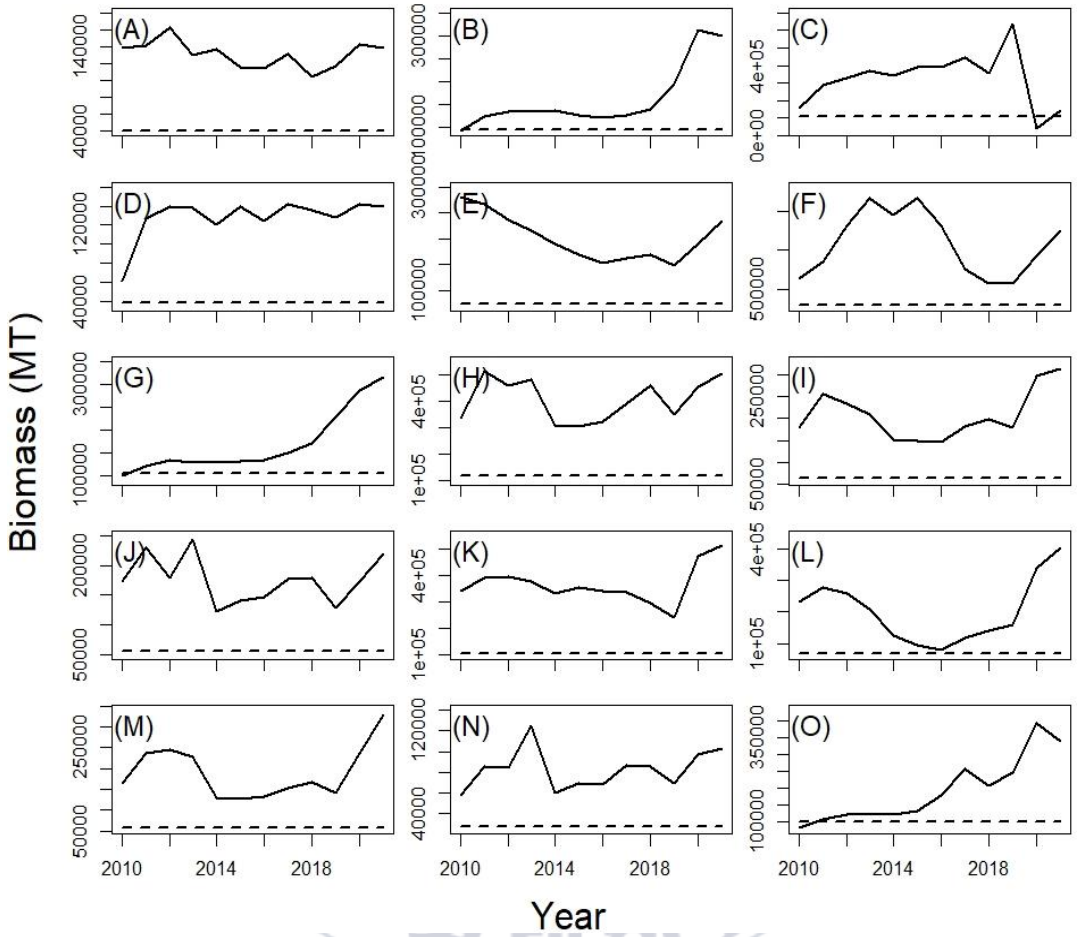

Figure 9. Predicted annual biomass (solid line) and biomass that produces MSY divided by two ($B_{MSY}/2$, dashed line). Panel (A) Indian anchovy, (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna, (G) Narrow-bared Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret, (J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

4.3. Kobe plot on 15 fish stocks

The Kobe plot, also known as the phase plot, is a tool for assessing stock status by examining the harvest rate (F) and biomass (B) concerning the Maximum Sustainable Yield (MSY; i.e. F_{MSY} and B_{MSY}) (Maunder & Aires-da-Silva, 2011). This approach is predicated on keeping the harvest rate below F_{MSY} and ensuring that stock biomass remains above B_{MSY}. Such a framework offers methodologies for encapsulating results from stock assessments and evaluations of management strategies in the Kobe format.

The Kobe plot comprises a four-quadrant display, each quadrant represented by different colors (red, orange, yellow, and green), facilitating the classification of stock status into four distinct categories: (1) Green, indicating stocks that are neither overfished ($B > {B_{MSY}}/{2}$) nor undergoing overfishing ($F < F_{MSY}$); (2) Orange (or upper-right yellow), for stocks that are not overfished ($B > {B_{MSY}}/{2}$) but are experiencing overfishing ($F > F_{MSY}$); (3) Yellow (lower-left), where stocks are overfished ($B < {B_{MSY}}/{2}$) but not undergoing overfishing ($F < F_{MSY}$); and (4) Red, signifying stocks that are both overfished ($B < {B_{MSY}}/{2}$) and undergoing overfishing ($F > F_{MSY}$) (Merino et al., 2020).

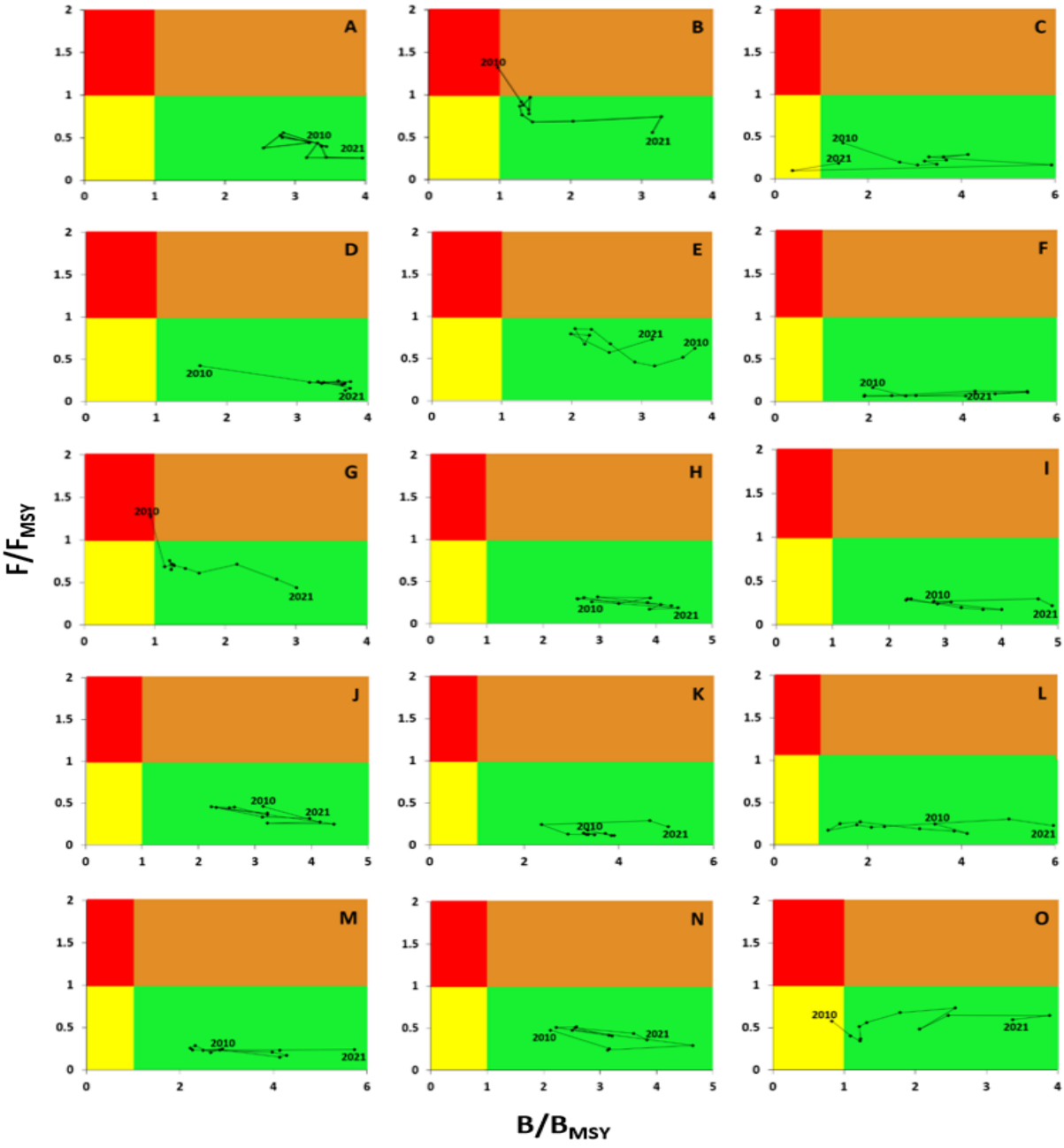

Figure 10. Kobe plot of 15 fish stocks. *F* represents the harvest rate, F_{MSY} is the harvest rate at the MSY point, B represents of Biomass, and B_{MSY} is biomass at the MSY point. Panel (A) Indian anchovy, panel (B) Short mackerel, (C) Mackerel scad, (D) Yellowtail scad, (E) Goldstripe sardinella, (F) Longtail tuna,

(G) Narrow-bared Spanish mackerel, (H) Giant sea catfish, (I) Black pomfret, (J) Pony fishes, (K) Red snappers, (L) Ornate threadfin bream, (M) Hairtails, (N) White shrimp, (O) Common squid.

Analyzing the Kobe plots depicted in Figure 10, it's evident that all 15 fish stocks were located within the green quadrant, indicating that they were neither overfished nor undergoing overfishing. However, there were instances where both the Short mackerel and the Narrow-barred Spanish mackerel temporarily fell into the red quadrant, signifying periods of being overfished and undergoing overfishing, although they predominantly remained in the green quadrant. Additionally, the Mackerel scad and Common squid were once positioned in the yellow quadrant, reflecting a state of being overfished but not actively undergoing overfishing. Crucially, in the most recent years of the time series, there was no evidence of either overfishing or being overfished among the 15 fish stocks from the Java Sea. Overall, these 15 fish stocks can be summarized as maintaining a satisfactory condition.

5. Conclusions

The implementation of a state-space production model effectively estimated the model parameters (r, K, q, b, σ^2) and calculated the annual biomass of 15 fish stocks in the Java Sea from 2010 to 2021. This model applied informative prior distributions for the intrinsic growth rate (r) and carrying capacity (K), utilizing the TMB package in R software. The outcomes of this study have facilitated a deeper understanding of the current status of the 15 fish stocks in the Java Sea, Indonesia.

As per the theories of NOAA Fisheries, there were no signs of either overfishing or stocks being overfished among the evaluated stocks. Considering all factors, all 15 fish stocks under analysis were not found to be in an unfavorable condition. To achieve the stock assessment goal, which is maximizing the catch is still necessary to maximize profits from the catch while effectively conserving fish stocks. I would argue that the fisheries management in Indonesia might have been excessively cautious. Allowing fishermen to capture a larger quantity of fish could result in economic benefits.

6. Acknowledgements

The author extends profound gratitude to Prof. Saang-Yoon Hyun for his invaluable guidance and numerous pieces of advice throughout this research; his support was fundamental to the initiation of this thesis. A special thanks is also owed to my senior, Jinwoo Gim, for his considerable assistance in this work. Heartfelt appreciation is given to my parents for their relentless support. Lastly, but importantly, sincere thanks are extended to all my lab colleagues for their consistent support and kindness. The study was funded by the ICFO Scholarship program. Directorate of Fish Resource Management, Indonesian Ministry of Marine Affairs and Fisheries provided the data.

7. References

Best, J. K., & Punt, A. E. (2020). Parameterizations for Bayesian state-space surplus production models. *Fisheries Research*, *222*, 105411.

Buckland, S., Newman, K., Thomas, L., & Koesters, N. (2004). State-space models for the dynamics of wild animal populations. *Ecological modelling*, *171*(1-2), 157-175.

Froese, R., Demirel, N., Coro, G., Kleisner, K. M., & Winker, H. (2017). Estimating fisheries reference points from catch and resilience. *Fish and Fisheries*, *18*(3), 506-526.

Graham, M. (1935). Modern theory of exploiting a fishery, and application to North Sea trawling. *ICES Journal of Marine Science*, *10*(3), 264-274.

Hilborn, R., & Walters, C. J. (2013). Quantitative fisheries stock assessment: choice, dynamics and uncertainty.

Hyun, S.-Y., & Kim, K. (2022). An evaluation of estimability of parameters in the state-space non-linear logistic production model. *Fisheries Research*, *245*, 106135.

Jiao, Y., Cortés, E., Andrews, K., & Guo, F. (2011). Poor-data and data-poor species stock assessment using a Bayesian hierarchical approach. *Ecological Applications*, *21*(7), 2691-2708.

Lang, J. B. (1999). Bayesian ordinal and binary regression models with a parametric family of mixture links. *Computational statistics & data analysis*, *31*(1), 59-87.

Marandel, F., Lorance, P., & Trenkel, V. M. (2016). A Bayesian state-space model to estimate population biomass with catch and limited survey data: application to the thornback ray (Raja clavata) in the Bay of Biscay. *Aquatic Living Resources*, *29*(2), 209.

Maunder, M. N., & Aires-da-Silva, A. (2011). Evaluation of the Kobe plot and strategy matrix and their application to tuna in the EPO. *Unpublished IATTC Scientific Advisory Committee document SAC-02-11 La Jolla, USA*, 14.

Merino, G., Murua, H., Santiago, J., Arrizabalaga, H., & Restrepo, V. (2020). Characterization, communication, and management of uncertainty in tuna fisheries. *Sustainability*, *12*(19), 8245.

Meyer, R., & Millar, R. B. (1999). Bayesian stock assessment using a state–space implementation of the delay difference model. *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(1), 37-52.

Millar, R. B., & Meyer, R. (2000). Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *49*(3), 327-342.

Polacheck, T., Hilborn, R., & Punt, A. E. (1993). Fitting surplus production models: comparing methods and measuring uncertainty. *Canadian Journal of Fisheries and Aquatic Sciences*, *50*(12), 2597-2607.

Prager, M. H. (1994). A suite of extensions to a nonequilibrium surplus-production model. *Fish. Bull.*, *92*, 374-389.

Prager, M. H. (2002). Comparison of logistic and generalized surplus-production models applied to swordfish, Xiphias gladius, in the north Atlantic Ocean. *Fisheries Research*, *58*(1), 41-57.

Punt, A. E. (2003). Extending production models to include process error in the population dynamics. *Canadian Journal of Fisheries and Aquatic Sciences*, *60*(10), 1217-1228.

Sant'Ana, R., Kinas, P. G., de Miranda, L. V., Schwingel, P. R., Castello, J. P., & Vieira, J. P. (2017). Bayesian state-space models with multiple CPUE data: the case of a mullet fishery. *Scientia Marina*, *81*(3), 361-370.

Schaefer, M. B. (1954). Some aspects of the dynamics of populations important to the management of the commercial marine fisheries.

Appendix: TMB code for the state-space production model with informative priors

CPP.file

```
sspmWithPrior <- "

//a state-space production model for Indonesia 15 fish stocks
//applying informative priors for parameter r and K

#include <TMB.hpp>

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

// square
template<class Type>
Type square(Type x){
return pow(x,2.0);
}

// dlnorm
template<class Type>
Type dlnorm(Type x, Type meanlog, Type sdlog, int give_log=0) {
   //return 1/(sqrt(2*M_PI)*sd)*exp(-.5*pow((x-mean)/sd,2));
   Type logres = dnorm( log(x), meanlog, sdlog, true) - log(x);
    if(give_log)
          return logres;
    else
          return exp(logres);
  }

//objective function
```

```
template<class Type>
Type objective_function<Type>::operator() () {
//data
DATA_VECTOR(Yt);          //Yt: yield (in weight) at time t
DATA_VECTOR(It);          //It: index or cpue at time t
DATA_VECTOR(modeCV_r);    //mode and CV of r, which are used for
informative prior for r;
DATA_VECTOR(modeCV_K); //mode and CV of K, which are used for
informative prior for K;

//parameters
PARAMETER(logsig_o);      //previously a=sig_o/sig_p;
PARAMETER(logb);       //P1 = b*lognormal(0, sigma_p^2) (note that
B1=b*K*lognormal(0, sigma_p^2), therefore E(logP1)=log(b), where B1 =
the initial biomass)
PARAMETER(logK);          //K: carrying capacity
PARAMETER(logb);          //q: scaling role, where It = q*Bt
PARAMETER(logr);          //r: intrinsic growth rate;
PARAMETER_VECTOR(logPt);  //log(Pt), where Pt=Bt/K; //Latent random
variable (i.e., random effect parameters)

//Derived quantities
int nyrs=It.size();       //nyrs: the number of years
Type sig_o=exp(logsig_o);
Type b=exp(logb);
Type sig_p=sig_o;      //variance of process error = variance of observation
error
Type K=exp(logK);
Type q=exp(logq);
Type r=exp(logr);
Type logMSY=logr+logK-log(4);
Type MSY=exp(logMSY);

vector<Type> Pt=exp(logPt);
vector<Type> E_logPt(nyrs+1);   //Expected values of log(Pt)'s
//vector<Type> E_logPt(nyrs);
vector<Type> logBt=logPt+logK;
```

```
vector<Type> logpredIt=logq+logPt+logK;
vector<Type> Bt=exp(logPt+logK);
vector<Type> predIt=exp(logq+logPt+logK);            //predicted It;
vector<Type> nll(5);   //negative log likelihoods that have three components

//objective function
nll.setZero();

//process equation
for(int i=1;i<=nyrs;i++) {
if(!isNA( Yt(i-1) )) {                                  //a zero-based indexing scheme
  E_logPt(i)=log( Pt(i-1)+r*(1.0-Pt(i-1))*Pt(i-1)-Yt(i-1)/K );  //because
Yt(nyrs) does not exist;
                                                  //Yt(0), Yt(1), ..., Yt(nyrs-1).
  nll(0) -= dnorm(logPt(i), E_logPt(i), sig_p, true);
  }
 }

//observation equation
for(int i=0;i<nyrs;i++) {   //a zero-based indexing scheme
if(!isNA( It(i) )) {
 nll(1) -= dnorm(log(It(i)), (logq+logPt(i)+logK), sig_o, true);
 }
}

//Only for the initial biomass, which is treated as the fixed effect parameter
//logPt(0) ~ normal(log(b), its variance), where its variance is assumed to be
sig_p;
nll(2)-= dnorm(logPt(0), log(b), sig_p, true);

//Informative prior for r
//r ~ lognormal;
Type prior_mode_r=modeCV_r(0);
Type prior_CV_r=modeCV_r(1);
Type sig2_r=log(square(prior_CV_r)+1);
Type mu_r=log(prior_mode_r)+sig2_r;
//=log(prior_mode_r*(square(prior_CV_r)+1));
```

```
nll(3)-=dlnorm(r,mu_r,sqrt(sig2_r), true);

//Informative prior for K;
//K ~ lognormal;
Type prior_mode_K=modeCV_K(0);
Type prior_CV_K=modeCV_K(1);
Type sig2_K=log(square(prior_CV_K)+1);
Type mu_K=log(prior_mode_K)+sig2_K;
nll(4)-=dlnorm(K,mu_K,sqrt(sig2_K),true);

Type jnll = nll.sum();              //joint negative loglikelihoods

//Reporting
REPORT(jnll);
REPORT(Bt);
REPORT(predIt);
REPORT(sig_o);
REPORT(b);
REPORT(sig_p);
REPORT(K);
REPORT(q);
REPORT(r);
REPORT(MSY);

ADREPORT(b);
ADREPORT(sig_o);
ADREPORT(sig_p);
ADREPORT(K);
ADREPORT(q);
ADREPORT(r);

ADREPORT(logb);
ADREPORT(logsig_o);
ADREPORT(logK);
ADREPORT(logq);
ADREPORT(logr);
```

```
ADREPORT(logBt);
ADREPORT(logpredIt);
ADREPORT(Bt);
ADREPORT(predIt);
ADREPORT(logMSY);
ADREPORT(MSY);

return jnll;

} "

#compile
write(sspmWithPrior, file="sspmWithPrior.cpp");
compile("sspmWithPrior.cpp");
dyn.load(dynlib("sspmWithPrior"))
```