



Model selection and multi-model inference for Bayesian surplus production models: A case study for Pacific blue and striped marlin



Yi-Jay Chang^{a,*}, Jon Brodziak^b, Joseph O'Malley^b, Hui-Hua Lee^a, Gerard DiNardo^b, Chi-Lu Sun^c

^a Joint Institute for Marine and Atmospheric Research, University of Hawaii—NOAA Fisheries, Honolulu, HI 96818, USA

^b NOAA Fisheries, Pacific Islands Fisheries Science Center, Honolulu, HI 96818, USA

^c Institute of Oceanography, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan

ARTICLE INFO

Article history:

Received 30 March 2014
Received in revised form 29 August 2014
Accepted 30 August 2014
Handling Editor A.E. Punt
Available online 18 October 2014

Keywords:

Bayesian hierarchical surplus production model
Model selection uncertainty
Deviance information criterion
Multi-model inference
Billfish

ABSTRACT

Stock assessment typically involves developing a set of alternative models, fitting each to the available data, and then selecting the one that gives the most accurate estimates of management quantities of interest. In this context, it is important to consider model selection uncertainty because ignoring it can lead to unreliable estimates and overconfident inferences. For this study, four Bayesian surplus production models with symmetric or asymmetric production functions and either a constant or hierarchical time-varying intrinsic growth rate (r) were developed using data for Pacific blue marlin (*Makaira nigricans*) and Western and Central North Pacific striped marlin (*Kajikia audax*). The uncertainty resulting from model selection was evaluated using Monte Carlo simulation techniques to examine the consistency of model estimates within (self-tests) and among (cross-tests) the alternative models. Specifically, these tests evaluated the performance of the deviance information criterion (DIC) and Bayesian model averaging (BMA). The results of the simulation tests suggested that mis-specification of time-varying r can lead to large estimation errors for biomass and management quantities and that DIC may not reliably identify the true data-generating model. Although BMA did not provide more accurate point estimates than just selecting the data-generating model, it did provide a more accurate characterization of uncertainty in model results. Our study shows the value of using simulations to evaluate model performance and to account for model selection uncertainty.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

To develop a stock assessment, fishery scientists usually build several alternative hypotheses or models, fit the models to the available data, and decide which alternative (“best”) model is likely to give the most accurate estimates of the current status of the stock relative to management targets (NRC, 1998; Hilborn and Walters, 1992). For example, alternative hypotheses could involve the choice of production curve for biomass dynamics models (Prager, 2002) or the choice of stock–recruitment relationships and selectivity functions for statistical catch-at-age (SCA) models (McAllister and Kirchner, 2002; Punt et al., 2014). The single “best” model is then usually selected using statistical measures of goodness-of-fit or model selection criteria.

Model selection is an important aspect of stock assessment because estimated quantities of interest can be sensitive to model structure. For example, the ratio of the reference biomass to the carrying capacity derived from biomass dynamics models is pre-determined by the choice of production curve and intrinsic growth rate (Hilborn and Walters, 1992). Consequently, ignoring model selection uncertainty can lead to overconfident inferences if the choices are artificially limited (Hoeting et al., 1999; Burnham and Anderson, 2002).

Model goodness-of-fit tests (e.g., F -ratio tests) or model selection criteria (Akaike Information Criterion, AIC, Akaike, 1973; Bayesian Information Criterion, BIC, Schwarz, 1978) have been widely used to select among fisheries assessment models. For example, Prager (2002) used an F -ratio test to choose between a logistic surplus production and a generalized surplus production model. Butterworth et al. (2003) used AIC to compare assessment model configurations that assumed selectivity was constant or changing over time. Helu et al. (2000) evaluated the performance of AIC and BIC to assess model selection in Stock Synthesis

* Correspondence to: NOAA Daniel K. Inouye Regional Center, 1845 Wasp Boulevard, Building 176, Honolulu, Hawaii 96818, USA. Tel.: +1 808 7255705.

E-mail address: yi-jay.chang@noaa.gov (Y.-J. Chang).

models. For Bayesian methods, model selection has traditionally been based on Bayes factors (Aitkin, 1991) and more recently on the deviance information criterion (DIC, Spiegelhalter et al., 2002). Wilberg and Bence (2008) explored DIC to select the best SCA model and suggested that DIC performed well at choosing the structurally appropriate model. Simulation testing has been also used to evaluate the performance of stock assessment models, through self-tests, i.e., when the structure of the assessment model matches that of the data-generating model, and cross validation analyses (NRC, 1998; Deroba et al., 2014). Although statistical model-fitting criteria, such as root mean square error, AIC, BIC, and DIC have been applied in many studies, relatively few studies have directly evaluated the performance of these criteria (Helu et al., 2000).

Although most assessments identify a single “best” assessment configuration, it may be advisable to apply multi-model inference using model averaging based on weighted support from a set of competing models (Hilborn and Mangel, 1997; Hoeting et al., 1999; Burnham and Anderson, 2002). Parameter estimates or predictions obtained from model averaging can be robust, in the sense that model selection bias can be reduced and model selection uncertainty can be taken into account. Multi-model inference using Bayesian model averaging (BMA) has been proposed as a means for assessing population growth for the endangered dromedary pearly mussel (*Dromus dromas*) (Jiao et al., 2008), estimating the stock–recruitment relationship for the walleye (*Sander vitreus*) fishery (Jiao et al., 2009a), estimating the probable status of a fishery resource under assessment scenario uncertainty (Brodziak and Piner, 2010), and standardizing CPUE of bycatch species (Brodziak and Walsh, 2013).

Use of Bayesian hierarchical frameworks for stock assessment modeling is also becoming more common. For example, Bayesian state-space surplus production models including multilevel priors (hierarchical models) were developed to simulate variability in population growth rates of hammerhead sharks (*Sphyrna* spp.) along the Atlantic and Gulf of Mexico coasts of the United States (Jiao et al., 2009b, 2011); similar models were also applied to the Deep 7 bottomfish complex of the Main Hawaiian Islands (Brodziak et al., 2011). Hierarchical model structures have also been used to simulate time-varying catchabilities (Wilberg et al., 2010).

The objectives of this study were to examine how model estimates are related to model selection uncertainty, to test the performance of DIC in model selection, and to evaluate the BMA approach. We applied Monte Carlo simulation techniques to address these issues using assessment data for blue marlin (*Makaira nigricans*) in the Pacific Ocean and striped marlin (*Kajikia audax*) in the Western and Central North Pacific Ocean (WCNPO). A set of simulation tests was conducted for each species based on four Bayesian state-space surplus production models using pseudo-datasets generated from the same structural model as the assessment model (self-tests) and when the data-generating and assessment models differed (cross-tests). These models had different structural assumptions related to production curves and/or population intrinsic growth rates (hierarchically or non-hierarchically structured). The results of simulation tests were compared within and between species. The appropriateness of applying these statistical approaches for stock assessment model selection and the practical implications for management are discussed.

2. Methods

This study consisted of two major parts: (1) fitting four Bayesian surplus production models to the Pacific blue marlin and WCNPO striped marlin stock assessment data; and (2) conducting a

simulation study to evaluate consequences of model misspecification and to test the performance of DIC and BMA.

2.1. Data used

Fishery catch data used were gathered from the most recent assessments of Pacific blue marlin and WCNPO striped marlin conducted by the Billfish Working Group of the International Scientific Committee for Tuna and Tuna-like Species in the North Pacific Ocean (ISC) (ISC, 2012, 2013). The catch data were used to model the effects of fishery removals from the Pacific blue marlin and WCNPO striped marlin stocks during 1952–2011 and 1975–2009, respectively.

Standardized catch-per-unit-effort (CPUE) indices were also gathered from the most recent stock assessments (ISC, 2012, 2013). The Pacific blue marlin CPUE indices consisted of two Japanese distant water longline CPUE series (1975–1993 and 1994–2011) and two Taiwanese distant water longline CPUE series (1979–1999 and 2000–2011). The two time blocks of CPUE for each fishery corresponded to the changes in fishing practices. The WCNPO striped marlin CPUE indices consisted of three Japanese distant water longline CPUE series (1975–1986, 1987–1999, and 2000–2009). These CPUE indices were selected on the basis of examination of residuals and total likelihoods (ISC, 2012, 2013).

2.2. Bayesian surplus production models

Surplus production models were fitted using a Bayesian state-space approach with explicit observation and process error terms (e.g., Meyer and Millar, 1999; Punt, 2003). We employed a time-varying hierarchical surplus production model with process error, similar to Brodziak and Ishimura (2011), to represent temporal fluctuations in biomass based on density-dependent processes and fishery harvests:

$$B_t = \left(B_{t-1} + r_t B_{t-1} \left(1 - \left(\frac{B_{t-1}}{K} \right)^M \right) - C_{t-1} \right) \eta_t \quad (1)$$

where B_{t-1} and C_{t-1} denote exploitable biomass and catch, respectively, for year $t-1$; K is the carrying capacity of the population; r_t is the intrinsic population growth rate for year t (i.e., the population growth rate was allowed to vary annually in some models); and M is the production shape parameter, which determines where surplus production peaks as biomass varies relative to carrying capacity. The process errors (η_t) were assumed to be independent and lognormally distributed random variables, $\eta_t = e^{U_t}$, where U_t were random normal variables with mean 0 and variance σ^2 . A reparameterization which expressed the annual biomass as a proportion of carrying capacity ($P_t = B_t/K$) was used to improve the efficiency of the Markov Chain Monte Carlo (MCMC) algorithm (Millar and Meyer, 2000). Therefore, the state equations for the initial time period ($t=1$) and subsequent periods ($t>1$) were

$$P_1 = \eta_1 \quad \text{and} \quad P_t = \left(P_{t-1} + r_t P_{t-1} \left(1 - P_{t-1}^M \right) - \frac{C_{t-1}}{K} \right) \eta_t \quad (2)$$

The observation model related the unobserved states, B_t , to the relative abundance indices, $I_{i,t}$, which were observed with error. We incorporated the selected CPUE indices into a single estimation model by specifying a separate observation equation for each index. Each observation equation assumed that the CPUE index was proportional to exploitable biomass

$$I_{i,t} = q_i K P_t \varepsilon_{i,t} \quad (3)$$

where $I_{i,t}$ is the relative abundance for index i at time t ; q_i is the catchability coefficient for index i , which describes the effectiveness of each unit of fishing effort; and $\varepsilon_{i,t}$ is a lognormal random

Download English Version:

<https://daneshyari.com/en/article/4542834>

Download Persian Version:

<https://daneshyari.com/article/4542834>

[Daneshyari.com](https://daneshyari.com)