How good is it? Using a 10-years hindcast to conclude the PBF assessment and projections provide a very good basis for management advice

Huihua Lee, Kevin Piner

Southwest Fisheries Science Center, NOAA Fisheries

December 2021

Working document submitted to the ISC Pacific bluefin tuna Working Group, International Scientific Committee for Tuna and Tuna-Like Species in the North Pacific Ocean (ISC), from 14 to 17 and 21 December 2021, Webinar.

Summary

A good stock assessment should be determined based on: 1) how accurate is the assessment model at recreating the past population dynamics, and 2) how good is the assessment model at predicting the future? This paper will focus on the second question that is the goal of fishery management. To assess the prediction quality of stock assessment for Pacific bluefin tuna, we used a hindcast to make a 10-years past prediction on the age-structured production model. As if we conducted the assessment 10 years ago using data only up to that year and forecast forward with the catches by fleets as actually did occur in the next 10 years, could we have predicted what happened to the stock? The result showed that the PBF assessment is a very reliable assessment for prediction because our production function accurately describes on average the effects of removing catches at age. Although we cannot prove that the model is accurate in recreating the past dynamics, it is very unlikely that a model with this predictive ability is not capturing a reasonable approximation of the dynamics. We strongly recommend that the following assessments provide the prediction tests, as they are the best gauge of an assessment's reliability for management.

Introduction

The fishery management relies on the stock assessment model to recreate the past population dynamics and project the future stock size. An accurate stock assessment model and/or a model with a good predictive skill earns credibility so that the managers and stakeholders are willing to base decisions on the information derived from the model. A good stock assessment should be judged on: 1) how accurate is the assessment model at recreating the past population dynamics, and 2) how good is the assessment model at predicting the future? The first criterion is hard to be proved in many assessments because the stock size and fishing mortality are not directly observed (unless the representative fishery-independent survey was available). Effort trying to prove how good the assessment is using a suite of diagnostics (e.g., convergence, goodness of fits, and model consistency analyses) is misinterpreted. These common diagnostics cannot tell us how accurate the model is. The second criterion, prediction quality, is the most crucial for managers and the ultimate proof of assessment reliability. Yet, this can be much easier to examine using own model (or ensemble of models) on historical data based on retrodiction. This test seeks to estimate the performance of an assessment model if it had been conducted during a past period and then used this back-tested model to predict/forecast the known future. This is a common practice in financial analysis and oceanography and is only now becoming more popular in fisheries (referred to as hindcasting; Kell et al. 2016).

The purpose of this paper is to show how good the Pacific bluefin tuna (PBF) assessment is at predicting the past using the age-structured production model (ASPM). The underlying assumption is that the assessment model that worked well in the past and predicted the past well indicates that the assessment model has a good prediction skill.

Methods

We first simplified the 2020 assessment model using the age-structured production model (ASPM) with the same fixed life history values as the assessment (Maunder and Piner 2015). The ASPM allows us to use the fundamental production function to make a hindcast given the known catch. The ASPM model was fitted to catch and indices only for 1952-2008 (eliminating the most recent 10 years (2009-2018) of catch and indices data). Then the dynamic was projected forward using this partial ASPM for 2009-2018 with the catches by fleets as actually did occur in these years. We compared the CPUE predictions for 2009-2018 to the CPUE observations not used in the fitting process to assess the prediction skill. Maybe the best way to think of this procedure is to imagine that we conducted the assessment 10 years ago using data only up to that year and forecast forward with the catches by fleets as actually did occur in the next 10 years; could we have predicted what happened to the stock?

We made forecasts using two options: 1) we knew future catches in weight or number as measured by that fleet and 2) we knew perfect future catch-at-age. The difference between these two is that option 1 does not consider how selectivity by fleet would change in the future nor how future recruitment variability would affect the catch-at-age. So you can think of option 1 as to how forecasting is actually done in the real world with all the real world problems and option 2 as a better measure of the model's ability to predict by controlling for the error in the catch-at-age (as if management had perfect control of catches).

Results and Discussion

Both option 1 (typical forecast) and option 2 (test of model prediction) were remarkably good at predicting future changes in abundance (Figure 1). This projection covered a period with large changes in management and resulting changes in adult stock trends (decreasing and then increasing). Obviously knowing the exact future catch-at-age improved the forecast, but both would have accurately described how trends in the abundance would have changed in that period.

The reason we can make such good predictions is that the PBF assessment uses a production function (made up of growth, natural mortality, and the spawner-recruitment function) that can accurately describe the average effects on abundance of catches at age over a range of stock sizes (Figure 2). When a model can accurately represent changes in abundance, considering only catches and the production function means that the model has a good measure of the effects of fishing. This model reasonably describes the trends in stock abundance over the last 50 years. In other words, once we have a good understanding of the average fishing effect, we do not need to know precisely how recruitment changes from year to year to predict on average how the stock changes in response to fishing. Remember that the actual assessment includes the estimates of annual recruitments which further improves the model performance without sacrificing our understanding of the fishing effect.

This paper is not meant to argue that assessment could not be improved, in fairness all assessments can be improved. The PBF working group has already recommended exploring alternative life history values (think production function) that could potentially produce similar reliable predictions. Those models could then be ensembled to provide a more complete (but still not full) inclusion of uncertainty.

We argue that models that make good predictions are better models than those that do not. Unquestionably the PBF assessment makes very good predictions because the production function can accurately assess the average effects of catches-at-age on stock abundance. We strongly recommend that the following assessments (whether a single model or ensemble of multiple models) provide these kinds of prediction tests, as they are the best gauge of an assessment's reliability for management. Given this view of models in general, we can only conclude that the assessment provides more than a reasonable basis for the management of the stock.

Reference

Kell, L.T., Kimoto, A., and Kitakado, T. 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fisheries Research. 183: 119-127. http://dx.doi.org/10.1016/j.fishres.2016.05.017

Maunder M.N. and Piner, K.R. 2015. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science. 72(1): 7-18. https://doi.org/10.1093/icesjms/fsu015

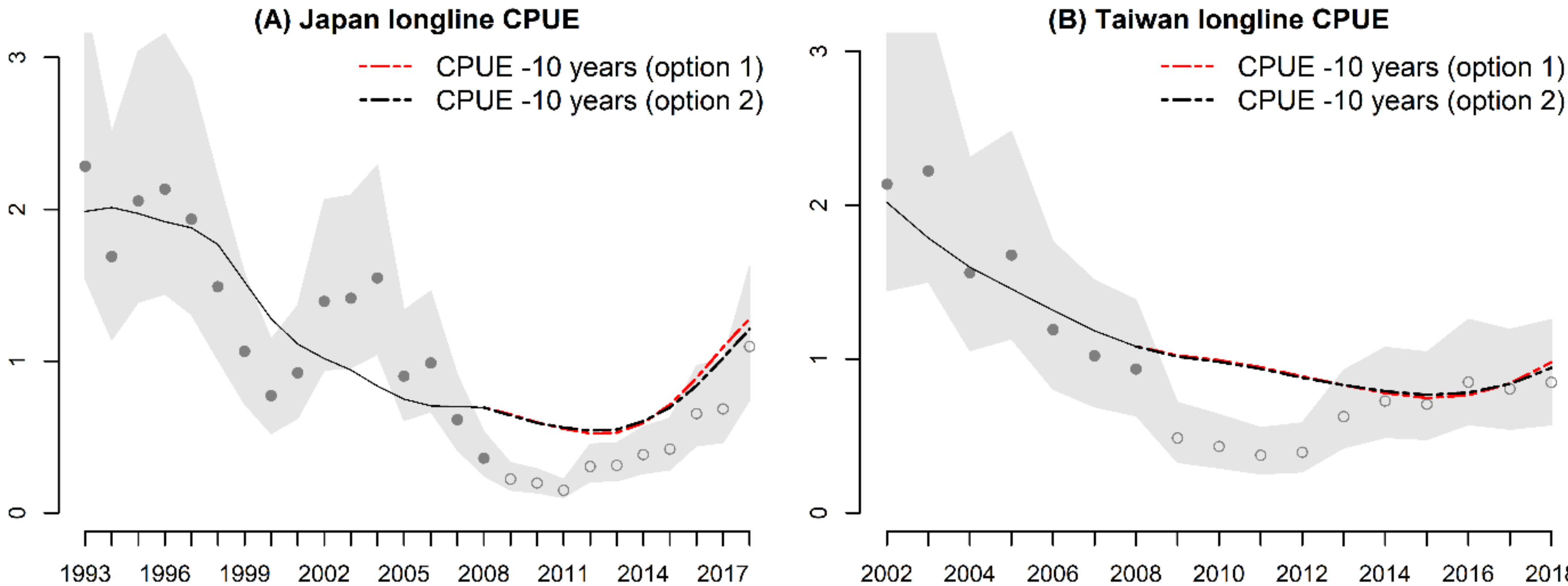

Figure 1. The expected (overlapped solid red and black lines) and predicted (dashed red and black lines) (A) Japan longline CPUE and (B) Taiwan longline CPUE from the age-structured production models, where only CPUE observations were removed for the recent 10 years (catch-at-age remained). The difference between the dashed red line and the dashed black line is that option 1 does not consider how selectivity by a fleet change in the future (typical forecast) and option 2 knows perfect future catch-at-age (test of model prediction). The solid circles represent the observations used in the models, open circles represent the missing values, and gray areas represent associated 95% confidence intervals.

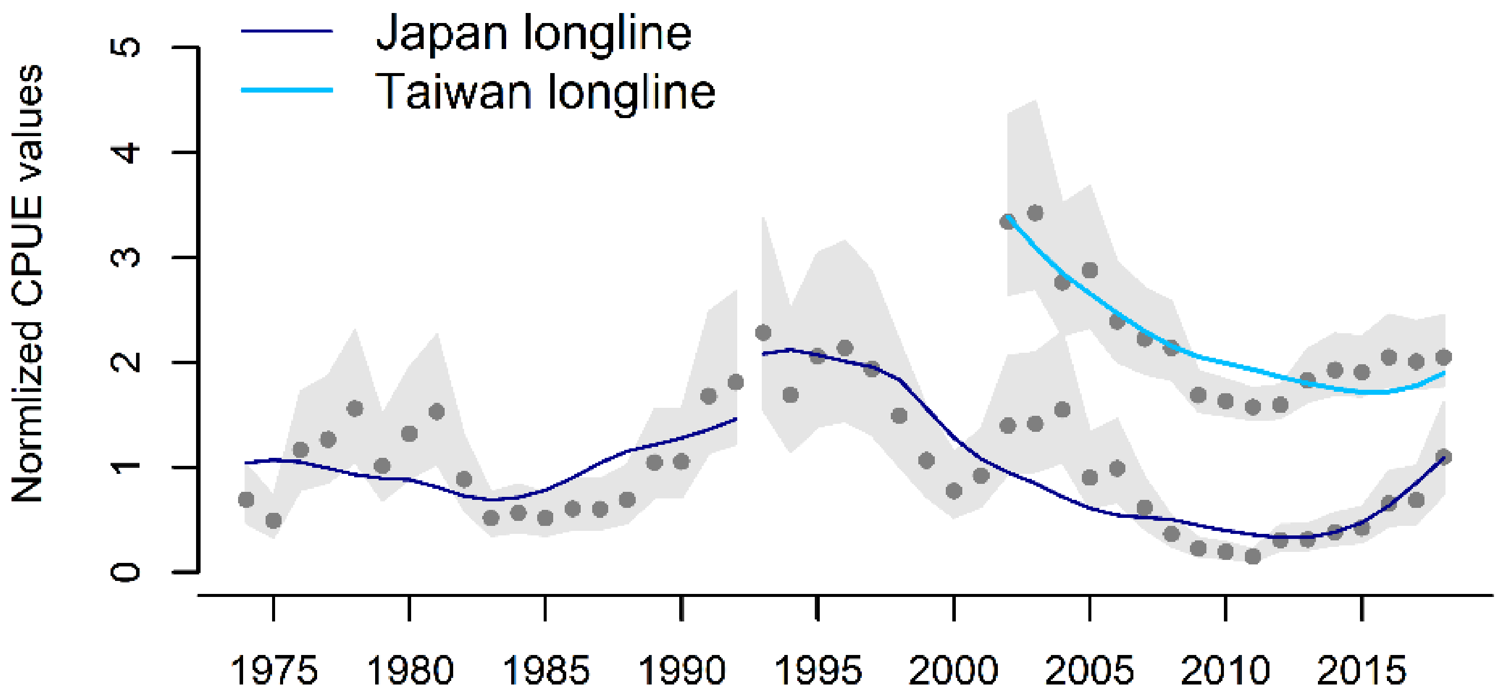

Figure 2. The expected adult indices (Japan longline in darker blue lines and Taiwan longline in lighter blue line) from the age-structured production model (with full CPUE observations) based on the 2020 stock assessment, where the solid circles represent the observations and gray areas represent associated 95% confidence intervals.