**Chapter 2**

**Evaluating estimation performance of a length-based multispecies stock assessment model**

## **Introduction**

Scientific advice for fisheries management has most often been based on single species population dynamics (Skern-Mauritzen et al., 2016; Karp et al., 2023). However, this approach assumes that changes in abundance are only due to fishing exploitation and natural mortality as general sources of mortality without explicitly represent the sources and components of mortality where natural mortality M is often a constant value between years and ages (Plagányi et al., 2022). Multispecies fisheries stock assessment models can improve accuracy in estimation of population dynamics by explicitly including the effect of predator-prey interactions (Wooton et al., 2021) and capture variations in natural mortality rates (Plagányi et al., 2022). Predation mortality has the largest impact in stock perception, and ignoring trophic interactions can lead to a bias in the estimated state of populations (Trijoulet et al., 2020). Multispecies models can help meet a variety of ecosystem management goals while providing improved insights for managing individual fisheries. Unfortunately, multispecies models have had limited use within the management process (Karp et al., 2023) and they are rarely used for stock assessment (Trijoulet et al., 2020).

Early attempts to apply ecosystem approaches in assessment models include incorporating environmental factors into single-species stock assessments such us oceanographic and physical variables and species interaction through predation mortality in age-based and length-based assessment models (Hunsicker et al., 2011; Skern-Mauritzen et al., 2016). Models of Intermediate Complexity for Ecosystem Assessment (MICE) account for ecological processes among a limited number of components or species (Plagányi et al., 2014), providing a link between full ecosystem models and single-species assessment models typically used to provide tactical fisheries management advice (Punt et al., 2016).

To ensure that models accurately reflect population dynamics, it’s crucial to test and validate them under controlled conditions before applying them to real data. Simulation testing has been widely utilized in fisheries science management to assess model performance, including the ability of stock assessment models to replicate population dynamics and stock conditions across various scenarios (Basson, 2002; Punt et al., 2002, Deroba et al., 2015), as well as to determine if an assessment model can provide reliable catch advice. Understanding the impacts of uncertainty in structural assumptions—such as the extent of trophic interactions—can enhance the practical application of multispecies models and facilitate the integration of ecosystem information into fisheries management. Consequently, running models under controlled scenarios or simulation testing, is an essential step for verifying model consistency and the robustness of our system's dynamics under known initial conditions.

Hydra is a length-structured, multispecies-multifleet MICE (Gaichas et al., 2017) designed to simulate population dynamics within the Georges Bank ecosystem for testing the performance of simpler multispecies assessment models and management strategies. Utilizing Hydra as an estimation model helps address long-term, multi-faceted management objectives in the framework of ecosystem-based fisheries management, offering assessments and catch advice for multiple species. To assess Hydra's potential as an estimation model, we evaluated its performance and robustness across various scenarios, examining how it responds to different parameterizations, input data, and conditions, as well as its capacity to accurately estimate key model parameters and important variables such as catch, biomass, and recruitment.

To understand how the strength of trophic interactions impacts model performance and the estimation of key management metrics, this chapter aimed to conduct simulation testing of a multispecies statistical size-structured integrated assessment model for a subset of stocks within the Georges Bank ecosystem in the Northwest Atlantic, while varying the availability of other food sources. This approach provides valuable insights into the model's performance, enabling fisheries managers to make better-informed decisions that support sustainability and ecosystem health.

## **Methods**

Hydra simulation model was already applied as a basis for testing EBFM procedures (pMSE reference) and later modified to be used as an estimation model by incorporating likelihood functions to fit the model to real data from the Georges Bank ecosystem, components of the objective function can be found in **Table 2**. Hydra incorporates length-structured data since predation and harvesting are length-based processes. Its goal is to provide accurate assessments of fish stocks within multispecies and multifleet contexts, including various data sources and parameterization; multiple forms of growth; recruitment and environmental covariates. Predation mortality is included to account for mortality on preys and to explore the impacts of trophic interactions. In this chapter we used Hydra as a simulation model (operating model) and as an estimation model.

We simplified our system with simple dynamics to two predators, Atlantic cod and Spiny dogfish, and two prey Atlantic mackerel and Atlantic herring, two fleets demersal and pelagic, and one survey (NEFSC spring survey), considering an equal number of size bins for each species in centimeters. Parameterizations for growth, recruitment, and fishery size selection were based on Georges Bank survey and fishery data; fishery selectivity, catchability and fishing effort were parameterized following guidance from stock assessment reports to reproduce plausible historical dynamics of the Georges Bank ecosystem. Input data series for each species are commercial catch in weight, survey index, length proportions for both commercial catches and survey and diet proportions (mention sources of all these data sets). The total biomass on Georges Bank or “other food” value was taken from Tsou & Collie (2001) and it was estimated in 15 million metric tons.

**Operating model description**

The operating model Hydra (Gaichas et al., 2017) describes population dynamics, including various forms of growth, reproduction, and mortality; biological interactions between species, such as predation and competition, and their effects on population structure; models different fishing fleets, considering effort, catch, and management. This allows for the evaluation of how fishing activity affects fish populations. The main components of the model include initial states, recruitment, predation M2, fishing effort, survey biomass, ecosystem indices, and assessment (model structure equations in **Table 1**).

The Operating Model OM was parametrized to reflect plausible dynamics in our system and to capture the trend of historical exploitation in the Northeast fishery. We mimic the trends of fishing effort time series for pelagic and groundfish species by guidance from stock assessment reports (**Figure 1**). (Describe average recruitment and recruitment deviations rec sigma average F for each fleet and deviations selectivity fishery and survey and residual mortality and other food value).

To estimate predation mortality in multispecies models, natural mortality is divided in two components: mortality due to predation (M2) and mortality due to all other natural causes or residual mortality (M1) (Helgason and Gislason, 1979). M1 was constant through length bins and years, and it was chosen so that M1+M2 is similar to M from stock assessments.

To understand the implications of the strength of trophic interactions we evaluated two scenarios with high and low interactions changing the amount of other food available in the ecosystem.

The multispecies portion of this model is dominated by five main equations and includes the total consumption by predators, the suitability of that prey to each predator, and the total suitable prey biomass available to each of its predators.

Annual fishing mortality deviations were chosen to force the trends from the stock assessment reports, for demersal fleet we assumed high fishing rates for the first 20 years and constant deviations for pelagic fleet. Different values of residual mortality M1 and other food OF were used until total natural mortality M was similar to those used in the stock assessments for each species.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Atlantic cod | Atlantic herring | Atlantic mackerel | Spiny dogfish |
| M1 | 0.28 | 0.32 | 0.2 | 0.102 |

Recruitment standard deviations were set in 0.5 for cod, herring and mackerel and 0.1 for spiny dogfish to reduce the variability in recruitment. Fishing mortality in the model is defined as fishing effort multiplied by catchability for each fleet and summed across the two fleets, also catchability q determines the proportion of secondary species on each fleet, for demersal fleet (Spiny dogfish) q was set in 0.25 and for pelagic fleet (Atlantic mackerel) in 0.1. Selectivity at length is assumed logistic for both fleets. Initial parametrization of the operating model detailed in **Table 3**.

To understand the consequences of uncertainty in structural assumptions and how the strength of trophic interactions affects model performance to improve the operational management use of multispecies models; two different assumptions about the magnitude of the interactions were included:

* Increase magnitude of trophic interactions to approximate the dynamic of a single species approach.
* Decrease the magnitude of trophic interactions between predators and prey varying the amount of other food and residual mortality.

**Simulated data**

Multispecies and size-structured data was simulated using the operating model to simulate 100 datasets including a 42-years period with constant values for the magnitude observation error variance e.g., CV and composition sample size (**Table 4**) of:

* catch (by species for each fleet).
* size composition of the catch (by species for each fleet.
* survey abundance index (by species.
* survey size composition (by species).
* survey diet proportions (stomach weight by predator size bin for each predator species).

Log based catches and indices for each species and survey/fleet were simulated assuming lognormal error, size compositions in numbers for each fleet and survey were modeled as multinomial given an input annual sample size for each year and species. Survey diet proportions by prey species were assumed multinomial distributed with an input sample size for each year and species (**Table 4**).

**Estimation model**

The estimation model is a multispecies statistical size-structured integrated assessment model including the same 4 stocks and the same population dynamics described in the Operating Model section. To evaluate the impact of the strength of trophic interactions in model performance assume different structural assumptions varying the parameter governing the amount of other food available to the modeled predators, considering high values of other food available in the ecosystem to reduce the strength of the modeled species interactions and smaller values to increase the magnitude of predation mortality, assuming there is less availability of non-modeled species in the ecosystem.

The estimation model will share the same population dynamics with the operating model and fits to stomach contents data in addition to survey and catch size compositions, catch, and survey abundance indices (components of the objective functions in **Table 2**).

Estimated parameters

* + - initial year numbers at size for each stock
    - average annual recruits for each stock
    - annual F for each fleet
    - fishery catchability for each stock
    - fishery logistic selectivity at length for each fleet
    - survey catchability for each stock
    - survey logistic selectivity at length

**Model performance**

The values for the estimated parameters and derived quantities from the fitted estimation models were compared to assess similarities, differences, and the coverage of estimated confidence intervals relative to the operating model quantities. Overall goodness-of-fit was calculated for each simulation, along with measures such as absolute error and mean squared error for each scenario, to determine if the model is sensitive to changes in the input data. Finally, a sensitivity analysis was conducted to quantify differences in the estimated parameters across models and scenarios, evaluating the sensitivity of the predictions.

Add equations for performance metrics (**Table 5**)

* + Distribution of relative errors
    - Bias and precision in the estimates of derived quantities of interest from the estimations model (relative error in estimates).
    - Relative error: compare the values of estimated parameters from the estimation model to the true values from the operating model.

## **Results**

OM description and plots with results:

**Figure 1** Operating model time series of biomass, fishing mortality, natural mortality, consumption and selectivity.

Scenarios results output plots

Compare results across scenarios and simulations

Summarize OM values and performance metrics (REE)

Show relative errors in plots for OM derived quantities

○ boxplots of the REE time series across simulations

○ plots of the relative errors for base scenario estimated parameters

Time series plots: SSB, recruitment, F, consumption and/or M2

○ boxplots of the REE for each estimated parameter

● Create plots of the relative errors for no interaction scenario derived quantities

○ boxplots of the REE time series across simulations, SSB, recruitment, F

● Create plots of the relative errors for no interaction scenario estimated parameters

○ boxplots of the REE for each estimated parameter

● Summarize (list form) the main results of the simulation experiments

write paragraph summarizing estimation performance for base scenario for derived quantities

write paragraph summarizing estimation performance for base scenario for estimated parameters

write paragraph comparing estimation performance for derived quantities for no interaction scenario

write paragraph comparing estimation performance for estimated parameters for no interaction scenario

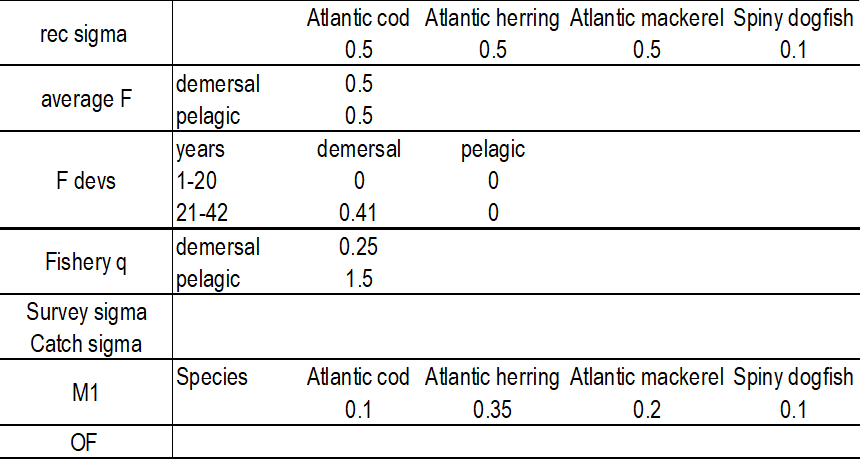
**Table 1**



**Table 2** Components of the objective function



**Table 3 incomplete**



**Table 4**