SEDAR

Southeast Data, Assessment, and Review

SEDAR 87

Gulf Pink Shrimp

SECTION III: Assessment Process Report

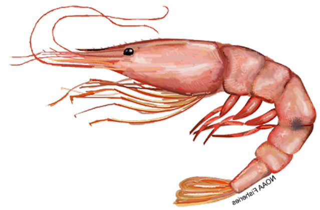
Updated: 5 June 2025
Original Release: 28 May 2025

SEDAR
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405

Document History

May 28, 2025 Original release

June 5, 2025 First letter of all species, regions, and size classes capitalized for consistency. Model landings and biomass units updated in table 11.



SEDAR87 Gulf Pink Shrimp Benchmark Assessment

Gulf Branch
Sustainable Fisheries Division
NOAA Fisheries - Southeast Fisheries Science Center

June 2025

Table of contents

1. Assessment Process Proceedings

On January 20, 2025, President Trump issued Executive Order 14172 to rename the Gulf of Mexico as the Gulf of America. Any reference to Gulf of America Pink Shrimp in SEDAR reports and other documents refers to the same species and fishery listed in *50 CFR part 622, Subpart C* (Shrimp Fishery of the Gulf of Mexico). As of the publication of this report, all efforts were made to use "Gulf of America" per Executive Order 14172. However, previous NOAA reports (cited herein) may have referred to this water body as the "Gulf of Mexico".

1.1 Introduction

1.1.1 Workshop Time and Place

The SEDAR 87 Assessment Process (AP) for Gulf Pink Shrimp was conducted via a series of webinars held between October 2024 and February 2025.

1.1.2 Terms of Reference

1. Review any changes in data or analyses following the Data Workshop. Summarize data as used in each assessment model. Provide justification for any deviations from Data Workshop recommendations.

2. Develop a management advice framework. Consider data availability (e.g., landings and catch-per-unit-effort [CPUE]) and management needs (e.g., harvest controls, stock status), and particular needs of the fishery and the biology of the resource.

3. Examine the impacts of social science factors on biological reference points as informed by stakeholders through industry input.

4. Recommend biological reference points for use in management.

    - Consider how reference points could be affected by management, ecosystem, climate, species interactions, habitat considerations, social or economic drivers, and/or episodic events.

5. Provide estimates of stock population parameters, including: Fishing mortality, biomass, selectivity, and/or other parameters as necessary to describe the population.

6. Characterize uncertainty in the assessment and estimated values.

- Consider uncertainty in input data, modeling approach, and model configuration.
- Provide appropriate measures of model performance, reliability, and 'goodness of fit'.
- Provide measures of uncertainty for estimated parameters and derived quantities such as biological reference points and stock status if feasible.

7. Provide recommendations for future research and data collection. Emphasize items that will improve future assessment capabilities and reliability. Consider data, monitoring, and assessment needs.

8. Complete an Assessment Workshop Report in accordance with project schedule deadlines.

1.1.3 List of Participants

Assessment Process Participants

Lisa Ailloud, Co-Lead Analyst .... NMFS Miami
Molly Stevens, Co-Lead Analyst .... NMFS Miami

Don Behringer .... NMFS SEFSC
Jie Cao .... NCSU/GMFMC SSC
Steve Munch .... NOAA NMFS
Jim Nance .... GMFMC SSC
Jason Saucier .... MS DMR
Katie Siegfried .... NMFS SEFSC
Brendan Turley .... NMFS

Appointed Observers

Leann Bosarge .... Industry Rep
Glenn Delany ....

Staff

Julie Neer .... SEDAR
Emily Ott .... SEDAR
Matt Freeman .... GMFMC Staff
Dominique Lazarre .... SERO
Michelle Masi .... SERO
Ryan Rindone .... GMFMC Staff
Carrie Simmons .... GMFMC Staff

Assessment Process Webinar Observers

Sarina Atkinson .... NMFS Miami
Peyton Cagle .... LWFD
Judd Curtis .... SAFMC Staff
Kyle Detloff .... NMFS SEFSC
Traci Floyd .... MS DMR
Carissa Gervasi .... NOAA
Bob Gill .... GMFMC

David Hanisko .... NMFS SEFSC
Kimberly Johnson .... NMFS SEFSC
Tricia Kimball ....
Christopher Liese .... NMFS SEFSC
Alan Lowther .... NMFS SEFSC
Richard Malinowski .... NOAA NMFS
Jessica Marchant .... AL DCNR
Fernando Martinez-Andrade .... TPWD
Akbar Marvasti .... NOAA
Cassidy Peterson .... NMFS SEFSC
Cheston Peterson .... NOAA
Adam Pollack .... NMFS SEFSC
David Records .... NOAA
Sarah Roberts .... NOAA
Skyler Sagarese .... NMFS SEFSC
Andrew Scalisi .... LDWF
Chris Schieble .... LDWF
Rebecca Smith .... NOAA
Jim Tolan ....
Michael Travis .... NOAA
Jo Williams .... NMFS SEFSC

1.1.4 List of Assessment Process Working Papers and Reference Documents

Document #	Title	Authors	Date Submitted
Documents Prepared for the Assessment Process			
SEDAR87-AP-01	Development of estuarine environmental indices for SEDAR 87 Gulf of Mexico White, Pink, and brown shrimp stock assessment	Brendan Turley, Lisa Ailloud, and Molly Stevens	25 July 2024
SEDAR87-AP-02	Price Indices for Shrimp Imports and Gulf of Mexico Shrimp Landings by Size and Season	Christopher Liese	18 December 2024
SEDAR87-AP-03	Developing a fishery-independent index of relative abundance for Gulf of Mexico Brown Shrimp using VAST	Lisa Ailloud, Molly Stevens, Brendan Turley, Adam Pollack, and David Hanisko	31 January 2025

SEDAR87-AP-04	Developing a fishery-independent index of relative abundance for Gulf of Mexico Pink Shrimp using VAST	Lisa Ailloud, Molly Stevens, Brendan Turley, Adam Pollack, and David Hanisko	31 January 2025
SEDAR87-AP-05	Developing a fishery-independent index of relative abundance for Gulf of Mexico White Shrimp using VAST	Lisa Ailloud, Molly Stevens, Brendan Turley, Adam Pollack, and David Hanisko	31 January 2025
Reference Documents			
SEDAR87-RD12	JABBA: Just Another Bayesian Biomass Assessment	Henning Winker, Felipe Carvalho, Maia Kapur	
SEDAR87-RD13	Empirical dynamic modeling for sustainable benchmarks of short-lived species	Cheng-Han Tsai, Stephan B. Munch, Michelle D. Masi, and Molly H. Stevens	
SEDAR87-RD14	Recent developments in empirical dynamic modelling	Stephan B. Munch, Tanya L. Rogers, George Sugihara	
SEDAR87-RD15	Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat	James T. Thorson and Lewis A. K. Barnett	

2. Data Review and Update

The following list summarizes the data inputs (and units) used in the assessment modeling process along with their corresponding available temporal scale based upon recommendations from the Data Workshop process. Two assessment modeling platforms were considered: a Bayesian surplus production model, JABBA (Just Another Bayesian Biomass Assessment), and an Empirical Dynamic Modeling (EDM) platform (see *Section 3*). Data for JABBA were on an annual time scale and included commercial landings (in million pounds of tails) and an index of abundance built with SEAMAP survey data using Vector Auto-Regressive Spatio-Temporal (VAST) modeling (Ailloud et al. 2025). EDM explored all the datasets listed below using various levels of stratification. JABBA allowed for different start years of data inputs, while EDM was limited by the start year of the survey data. For EDM, data were stratified by fishing area [A (*Figure 1*) : 1-10, 11-17, 18-21], size [S: >67 (Small), 67-31 (Medium), <=30 (Large) tails per pound], and quadrimester of the year [Q: January-April (Winter), May-August (Summer), September-December (Fall)] where possible, and are indicated as such in the data list. Stratifications were defined based on existing definitions of the ecological distribution of shrimp and the shrimping industry.

1. Commercial landings (million pounds of tails): 1960-2022 [A, S, Q]
2. SEAMAP survey data (number of shrimp per trawl hour): 2010-2022 [A, S]
3. Ex-Vessel price indices (2022 dollars): 1960-2023 [S, Q]
4. Imports (product volume in 100 million pounds): 1972-2022 [Q]
5. Salinity (practical salinity unit): 1980-2022 [A]
6. Bottom temperature (degrees Celsius): 1980-2022 [A]

Pink Shrimp are distributed primarily in the eastern Gulf (*Figure 1* : 1-10). SEAMAP survey data from this region were used to generate a VAST index (Ailloud et al. 2025). Possible data stratifications for Pink Shrimp EDM were defined as:

A) Aggregated: ANNUAL ; SIZE BINS AGG ; AREA AGG (01:10)
B) [Stratum N/A because single area]
C) Size: ANNUAL ; SIZE BINS (>67, 67-31, <=30) ; AREA AGG (01:10)
Csm) Size: ANNUAL ; SIZE BINS (>30, <=30) ; AREA AGG (01:10)

2.1 Stock Structure and Management Unit

The SEDAR 87 Gulf Pink Shrimp Benchmark Assessment stock boundary extends from the United States–Mexico border in the west through the northern Gulf of America waters (hereafter referred to as the Gulf) to the Dry Tortugas and Florida Keys. This includes all waters within the Gulf of Mexico Fishery Management Council (hereafter referred to as the Gulf Council) boundaries and extends to include fishing areas split by the eastern boundary off the Florida Keys (*Figure 1*: Areas 002, 001) in their entirety due to complications with reporting over time (Atkinson et al. 2024). This stock boundary distinction is most important for Pink Shrimp due to its distribution being centered in the eastern Gulf, but it was applied to all Gulf shrimp species.

2.2 Fishery-Independent Survey Data

2.2.1 Southeast Area Monitoring and Assessment Program (SEAMAP)

The Southeast Area Monitoring and Assessment Program (SEAMAP) survey is a collaborative effort between federal, state, and university programs, designed to collect, manage, and distribute fishery-independent data throughout the region. The SEAMAP survey design was improved and expanded in 2008. This design took a few years for reliable funding to fully sample the eastern Gulf and was deemed representative for Pink Shrimp since 2010 in the Summer and 2014 in the Fall for all provided size classes (Pollack and Hanisko 2024). Pink Shrimp abundance was explored by size class (*Figure 2*). Raw annual indices for Pink Shrimp are shown by size class in *Table 1* and overall in *Table 2*.

Due to the global COVID-19 pandemic, the 2020 Summer SEAMAP survey did not operate. There are variable time step methods available for dealing with missing data in EDM, but these approaches considerably increase model complexity and are typically only used when variable-sized gaps in data are present throughout a long time series. Using a variable time step method was determined to be inappropriate for addressing a single missing time step of data; however, it was also not desired to stop the model in 2019 when data were available through 2022. Therefore, the 2020 missing data point was replaced by an average of 2019 and 2021 on the finest resolution possible (e.g. annual aggregations averaged 2019/21, but size-stratified aggregations averaged Small 2019/21 for the Small 2020 estimate, and so forth).

2.2.2 Vector Autoregressive Spatio-Temporal (VAST) Index

VAST is a spatio-temporal modeling platform that can be used for standardizing indices of relative abundance. Data from one or more surveys are combined to predict population density based on both habitat covariates (that impact abundance) and spatial and spatio-temporal random effects, while controlling for catchability covariates (that impact sampling efficacy). A VAST index was developed for Pink Shrimp based on data from SEAMAP survey for input into JABBA. Details of the VAST index are documented in Ailloud et al. (2025).

2.3 Fishery-Dependent Data

2.3.1 Commercial Landings

Commercial landings of Pink Shrimp were constructed using data from the Gulf Shrimp System (GSS) and state trip ticket programs. Species-specific Gulf shrimp landings have been collected since the late 1950s, and their complex history within the federal and state databases, including justifications for the relative coefficients of variance (CVs) through time, is documented in great detail in Atkinson et al. (2024). Shrimp landings have been sold and recorded in eight market categories which were aggregated into three general size classes: Large, Medium, and Small (*Figure 3*). This dataset suffers from unreliable or missing size data for Pink Shrimp landings from Florida, with up to 30% of landings missing size information over the years overlapping with the SEAMAP survey data (*Table 3*). Landings were converted to tail weight for input to the assessment model (*Table 4*).

Pink Shrimp landings peaked in 1964 with landings totaling 20.99 million pounds of tails. For the years where there is an index of abundance, landings peaked in 2018 with landings totaling 12.99 million pounds of tails. The proportion of landings by season is shown in *Figure 4*.

2.4 Economics and Social Sciences

2.4.1 Imports and Ex-vessel Price Indices

Imported shrimp have exceeded the volume of domestically caught shrimp since the 1980's (Lowther 2023; Atkinson et al. 2024). In the mid-2000s, the volume of imported shrimp increased dramatically, particularly for Large shrimp which has a higher market value, causing domestic ex-vessel prices to plummet (*Figure 5*). Time series of imports and ex-vessel prices were both considered during EDM development (Liese 2024).

2.4.2 Industry Impacts

The globalization of the shrimp market with a focus on cheap aquaculture has resulted in dire economic operating conditions for the domestic fleet (Griffith et al. 2023). Increasing fuel costs and plummeting ex-vessel prices have created a situation in which most vessels struggle to remain profitable. Further, many vessels have exited the fleet, and those that remain may oscillate between narrowing profit margins and losses (*SEDAR87 data workshop report* 2023 pp. 84–94). With fewer vessels operating, the shrimping effort and associated landings have decreased overall, and the shrimp population size has increased.

Industry impacts were documented during a stakeholder listening session at the Data Workshop, with the intention of holding additional listening sessions throughout coastal Gulf shrimping communities. During this session, resource users stated that the troubles of the Shrimp Fishery cannot be improved by domestic fishery management solutions. The bulk of the problems are globally influenced, and this fishery was recommended to the National Seafood Strategy to address these problems if possible, informed by additional information gathered through the newly formed *Shrimp Futures Project*.

2.5 Environmental Indices

Annual shrimp recruitment has been tied to environmental drivers in the past (Browder et al. 2002; Zink et al. 2018; Schlenker et al. 2023). Within an assessment modeling framework, it is important to include drivers of abundance at the most meaningful spatio-temporal scale. At the SEDAR 87 Data Workshop, the Environment and Industry Working Group recommended that salinity and temperature in the nursery grounds during the months that the shrimp were in their respective nursery grounds were likely the primary environmental drivers for shrimp abundance. These two variables were hypothesized to best explain the magnitude of recruits into the population each year. The methodology used to derive Pink Shrimp temperature and salinity indices was outlined in Turley et al. (2023). These indices were considered in the construction of the VAST index and development of EDM.

Pink Shrimp is in its inshore nursery grounds September through October every year throughout its coastal range. It was hypothesized that the environment would affect the overall population abundance more directly through its impact on the young of the year in this volatile habitat (*Figure 6*). Everglades National Park buoy data were used to estimate temperature and salinity because Florida Bay has been documented to be the primary nursery ground for shrimp recruiting to the Tortugas fishing grounds, where the bulk of the landings originate (Sheridan 1996; Turley et al. 2023).

2.5.1 Temperature

Temperature in the Everglades from the buoy data is relatively warm with a median value of 29.1 degrees Celsius.

2.5.2 Salinity

Salinity in the Everglades is very high compared to other shrimp nursery grounds in the Gulf, with a median value of 31.12 PSU. Everglades watershed policy has a long history in Florida and numerous changes have occurred throughout the duration of the shrimp fleet’s operations. There have been freshwater flow diversions to supplement sugar crops as well as attempts to restore historical flow, with the Comprehensive Everglades Restoration Plan (CERP) approved by Congress in 2000.

3. Stock Assessment Model Configurations and Methods

Two modeling frameworks were evaluated for the Gulf Pink Shrimp SEDAR 87 Benchmark Assessment: Just Another Bayesian Biomass Assessment (JABBA) Model and Empirical Dynamic Modelling (EDM). These are described below.

3.1 Just Another Bayesian Biomass Assessment (JABBA) Model

JABBA is a Bayesian state-space surplus production model (SPM) framework that is documented in Winker et al. (2018) and is available as an R package on *GitHub*. SPMs pool the overall effects of recruitment, somatic growth, natural mortality, and associated density-dependent processes into a single production function dealing with undifferentiated biomass (Haddon 2021). The state-space formulation allows for the estimation of observation and process error, and the Bayesian formulation allows the user to define prior distributions for each parameter in the model to represent the initial beliefs about the parameter before observing any data. Primary data inputs into JABBA are indices of abundance proportional to the exploitable part of the stock biomass and a time series of fishery removals. The time series of removals can begin prior to the indices of abundance, and contrast in the data is required to appropriately map the stock dynamics.

The generalized surplus production function (Pella and Tomlinson 1969) used by JABBA is defined as

$$SPM_t = \frac{r}{m-1}\left(1 - \frac{{B_t}^{m-1}}{K}\right)$$

where r is the intrinsic rate of population increase at time t, K is the carrying capacity, B is the stock biomass at time t, and m is the shape parameter that determines at which B/K ratio maximum surplus production is attained. The Pella-Tomlinson function above is a generalized production function with Schaefer ($m = 2$) and Fox ($m = 1$) as special cases. The Schaefer may be the most well-known, with a symmetrical production curve and Maximum Sustainable Yield (MSY) attained at half the carrying capacity, $B = K/2$.

JABBA has several features including the ability to a) fit multiple CPUE time series and associated standard errors, b) estimate or fix the process variance, c) estimate additional observation variance on individual or grouped CPUE series, and d) specify either a Fox, Schaefer or Pella-Tomlinson production function. A full JABBA model description, including formulation and state-space implementation, prior specification options, and diagnostic tools is available in Winker et al. (2018).

3.1.1 Estimated Parameters

JABBA model parameters are defined in greater detail below.

K: Carrying capacity (million lb tail weight)

m: Shape parameter of the Pella-Tomlinson that determines at which B/K ratio maximum surplus production is attained. If $m = 2$, the model reduces to the Schaefer form, with the surplus production (SP) attaining MSY at exactly $K/2$. If $0 < m < 2$, SP attains MSY at biomass levels smaller than $K/2$; the converse applies for values of m greater than 2.

ψ: Ratio of the spawning biomass in the first year to K.

q: Catchability coefficient.

r: Intrinsic rate of population increase.

σ^2: Process variance.

τ^2 : Additional observation variance for the survey index.

3.1.2 Model Configurations and Prior Assumptions

The final VAST index built on SEAMAP survey data presented in Ailloud et al. (2025) was used as input to JABBA alongside an annual time series of commercial catches spanning 1960-2022 (*Section 2.3.1*). The following CVs were recommended by the WG and input into JABBA to reflect uncertainty in landings based on changes in the sampling programs through time. 1960-1983: CV = 0.2, 1984-2015: CV = 0.1, 2016-2022: CV = 0.05. The time series and associated confidence intervals are shown in *Figure 7* and *8*. Model configurations and prior distributions were defined as follows:

Carrying capacity (K): uninformative prior. Lognormal distribution specified using the "range" option in JABBA with lower and upper values ranging from maximum catch to 10x maximum catch (*Figure 9*)

Production function: Pella-Tomlinson (MSY at $B_{MSY}/K = 0.4$; $CV = 0.3$) where B_{MSY} is the biomass at MSY (*Figure 10*)

Process error variance (σ^2): Default $1/\gamma(4{,}0.01)$ (*Figure 11*). This matches the level of process error where state-space SPMs are most likely to adequately perform.

Observation error variance (if estimated) (τ^2): Default ~ $1/\gamma(0.001{,}0.001)$ (*Figure 12*)

r prior: informative priors were developed based on the Medium (0.2-0.8) and High (0.6-1.5) resilience categories in FishBase (Froese et al. 2019). Given that FishBase does not include any crustaceans and that shrimp are likely on the higher range of r compared to most fishes, an additional Very High (1.2-3) prior was tested (*Figure 13*)

Initial biomass depletion ratio (ψ): two alternative priors were tested to reflect Low initial depletion $lognormal(0.9{,}0.25)$ and and High initial depletion $lognormal(0.25{,}0.5)$ at the beginning of the catch time series ($\psi = B_{1960} / K$) (*Figure 14*)

A factorial design was used to test a suite of models with alternative prior assumptions about r, ψ and τ^2. The naming convention for candidate model is as follows:

SpeciesCode_ModelRun_ProductionCurve_rPrior (H:High,M:Medium,V:Very High)_PsiPrior(High:0.2,Low:0.9)_ObservationError(T=TRUE,F=False)_StartYearCatches

For example,

PSH_13_P_rH_psil0.9_sigF_60 : Pink Shrimp (PSH_) run number 13 (_13) using a Pella-Tomlinson surplus production curve (_P), high r prior (_rH), low initial depletion (_psil0.9) with no additional observation error being estimated (_sigF) and a catch time series starting in 1960 (_60)

3.1.3 Model Diagnostics

Candidate models were assessed based on the following four criteria (Carvalho et al. 2021):

3.1.3.1 Model Convergence

The Geweke convergence diagnostic (CONV_gw) compares the mean of the first and last part of Markov chain to see if they are significantly different. Z scores near 0 (between -1.96 and 1.96) are considered acceptable (Geweke 1992).

Heidelberger and Welch stationarity diagnostic (CONV_hs) shows the iteration number from which the chain is considered to have converged and an associated p value, where the null hypothesis is that the sampled values come from a stationary distribution (Heidelberger and Welch 1983). 'Failure' of the stationarity test indicates that a longer MCMC run is needed. The Heidelberger and Welch half-width test (CONV_hw) checks whether the Markov chain sample size is adequate to estimate the mean values accurately (Heidelberger and Welch 1983).

3.1.3.2 Model Fit

Catch-per-unit-effort (CPUE) residuals runs test: CPUE indices pass the runs test (CPUE_rt_rand) if there is no evidence of a non-random residual pattern ($p > 0.05$). Any year where the residuals are larger than the threshold limit [3 standard deviations (sd) away from the mean (Anhøj and Olesen 2014)] fail the outlier test (CPUE_rt_outl).

3.1.3.3 Model Consistency

Retrospective analysis: This test checks for systematic bias in the stock status estimates. The procedure involves sequentially removing all data from the most recent period (i.e. peeling), refitting the model, and then comparing terminal year estimates of stock status [e.g. spawning stock biomass (SSB), fishing mortality (F)] to the full model. A guiding practice proposed by Hurtado-Ferro et al. (2015), suggests values of Mohn's rho (RETRO_) that fall outside a set range (-0.22 to 0.30) for shorter-lived species indicates an undesirable retrospective pattern. In addition, the direction of the retrospective bias has implications for characterizing risk associated with management advice.

Process error: The annual process error deviations should exhibit a stochastic pattern with a constant average centered around the zero (ProcB_mu) and 95% credibility intervals covering the zero value (ProcB_CI).

3.1.3.4 Prediction Skill

Hindcast cross-validation (Kell et al. 2016, 2021): this test is to check that the model has prediction skill of future states under alternative management scenarios. The procedure involves sequentially removing CPUE data from the most recent period, refitting the model with the remaining data, and then comparing known CPUE values (observations) to model estimates.

Mean Absolute Scaled Error (HX_MASE): The MASE score scales the mean absolute error of the prediction residuals to the mean absolute error of a naive in-sample prediction (i.e. equal to the last observed value). A score of 0.5 indicates that the model forecasts of CPUE values are twice as accurate as a naive in-sample prediction, indicating that the model has prediction skill. A score higher than 1 indicated that the model forecasts are no better than a random walk. If MASE < 1, the model has some level of prediction skill and passes the test.

3.1.4 Goodness of Fit

Deviance Information Criteria (DIC) was used for model selection purposes, where a lower value generally indicates a better model fit. Root-Mean-Squared-Error (RMSE) was used to quantitatively evaluate the randomness of model residuals. These criteria were used to determine the best model of those that passed the model diagnostic tests described in the previous section.

3.2 Empirical Dynamic Modelling (EDM)

Empirical Dynamic Modelling (EDM) uses lags of time series data to reconstruct the state-space of a system (Sugihara 1994; Sugihara et al. 2012; Munch et al. 2017, 2022). This form of modeling is particularly useful for short-lived species with chaotic population dynamics where drivers are often not observed directly, yet the information is embedded within the time series of abundance. Lags of abundance indices are used to reconstruct the full dynamics of the system without needing data on variables impacting abundance or specifying model form. Gaussian-Process EDM (GP-EDM) version 0.0.0.9010 on *GitHub* was used to fit the SEAMAP survey data aggregated at levels defined in *Section 2*. We also tested the inclusion of economic and environmental variables as covariates since they are hypothesized drivers of shrimp abundance where measurements do exist.

3.2.1 Model Configurations

3.2.1.1 Formulation with Fishery Removals

Gaussian Process regression was used to approximate the Pink Shrimp population delay-embedding map f

$$P\left[y_t \,\middle|\, f, (X_{t-m} - q * C_{t-m}), z, V_e\right] \sim Normal(f(X_{t-m} - q * C_{t-m}, z), V_e)$$

where the probability of observing abundance y at time t is dependent on the function approximation f, vector of abundance indices X with m lags ($X_{t-m} = x_{t-1}, \dots, x_{t-m}$), optional covariates z, and process variance V_e. The delay embedding map defined above was expanded to include removals (C, catch or landings) scaled by a catchability parameter q which can be fit within or among populations. Here, catchability is a scalar used to translate units of landings into survey units. Covariates (z) can be included as direct drivers of abundance where measurements exist. Fitting to ‘escapement’, the composite variable $X_{t-m} - q * C_{t-m}$ is the number of individuals remaining after harvesting. GP-EDM with a single lag $m = 1$ can be thought of as a nonparametric production model (Thorson et al. 2014). f is dependent on the inverse length scales $\Phi = \phi_1, \ldots, \phi_{i=m+z}$ and pointwise prior variance τ and follows a Gaussian Process prior with mean zero and covariance function Σ, which assumes no relationship on the shape function.

$$P\left[f \,\middle|\, \Phi, \tau\right] \sim GP(0, \Sigma)$$

The covariance function Σ is defined for abundance y

$$\Sigma(y_t, y_s) = \tau * exp[-\Sigma_{i=1}^{m+z} \phi_i (X_{it} - X_{is})^2]$$

at times t and $s \in T$ where T is the time series length (Munch et al. 2022). The inverse length scale parameters ϕ and abundance observations X are provided for each $i = m + z$ where m is the lags of abundance and z is the covariates. This function is scaled by τ, and a prior is applied here that constrains the total variance of the predicted population size (y_{T+1}) to be less than twice the observed variance in $y_1, \ldots, y_T$. This prior specification for process and observed variances and length scale parameters are represented by

$$P[V_e, \tau, \Phi]$$

The covariance function and inverse length scales jointly control the degree of nonlinearity of the shape function f, where $\phi = 0$ indicates a flat relationship and a large estimate for ϕ indicates a higher degree of nonlinearity. The covariance function Σ can either tighten the relationship around the observed data, favoring a smaller length scale (i.e. a larger inverse length scale parameter) or relax the relationship, facilitating a smoother function with a larger length scale (smaller ϕ). Detailed GP prior specification for EDM variance and length scale parameters can be found in Munch et al. (2017).

An optional feature of GP-EDM is to assign a linear prior on f which can aid in grounding the population to 0 as the harvest rate, U, approaches 1 (i.e. the entire population is harvested). The linear prior option assumes that the mean function for the GP is linear with respect to the first input and fits the model on the residuals of

$$y_t = \beta_0 + \beta_1[x_{t-1} - q * c_{t-1}] + f(X_{t-m} - q * C_{t-m}, z)$$

where $[x_{t-1} - q * c_{t-1}]$ is first lag of escapement and f is the GP function approximation. If $y_t = log(x_{t+1}/x_t)$ and is backtransformed, this is equivalent to a Ricker model excluding f (Ricker 1954). In this case, we're working on deviations from growth under the assumed Ricker model. The model fits similarly to the previous configuration, but the primary difference can be observed outside of the range of observed data. This configuration helps linearly ground the fishery model abundance to zero as simulated removals approach the total population size. Without this prior, it's possible that outside of the observed range of the data, the abundance levels out to the flat prior where the population may never reach zero (and can result in extraordinarily high landings under simulated high harvest rates).

3.2.1.2 Embedding Dimension

EDM embedding dimension E is limited by the length T of the time series. An approximate maximum embedding dimension is $E \leq \sqrt{T}$. In the case of continuous seasonal data, the maximum embedding dimension is larger since the time series T is longer. Models were configured using Summer and Fall seasons as continuous time steps throughout a year and as a population-specific level within a hierarchical EDM, which will be explained in further detail below. The embedding dimension is defined as the number of population lags m (and covariates z if included) plus one, $E = m + z + 1$. For Pink Shrimp, the first year of the SEAMAP survey was 2010, resulting in 13 years of data, and a maximum embedding dimension of approximately 3 on an annual scale.

3.2.1.3 Hierarchical Model Scaling

Prior to fitting EDM models, all input data are standardized to a mean of 0 and standard deviation of 1. In the context of EDM, the term 'populations' is used to define data aggregations where information is expected to be informative. For Pink Shrimp, data aggregations and resulting populations that could be used to delineate levels of EDM are defined at the start of *Section 2*. For systems with multiple populations, these could be fit within a hierarchical EDM or independently.

In hierarchical models, the data must be scaled globally or locally across populations. For global scaling, the data across populations are expected to have the same mean. For Pink Shrimp, global scaling is likely inappropriate for most strata defined here. For example, we never expect the abundance of Large shrimp to equal the abundance of Small shrimp as would be implied by global scaling. Local scaling allows us to scale the data within the defined population time series of available data for each respective lag of population abundance or covariate. Both global and local scaling are applied within each predictor, not across all data. For example, each predictor is scaled to a mean of 0 and standard deviation of 1 for each lag and covariate. For global scaling, all data from all populations are used to scale the data; for local scaling, this is done within populations.

In independent models, definition of global or local scaling is obsolete because all data are scaled to a mean of 0 and standard deviation of 1. Independent models were tested for all data aggregations to ensure information was gained through the increased complexity and shared information from hierarchical models and with dynamic correlation.

3.2.1.4 Dynamic Correlation

Dynamic correlation ρ is defined as the degree to which the EDM population dynamics are correlated. This quantifies the similarity of population responses across predictor space and ranges from 0 to 1. Populations in hierarchical models will share the same embedding parameters and inverse length scale parameters (this includes models with $\rho = 0$, or independent dynamics). A dynamic correlation $\rho = 1$ means the dynamics of each population are identical. In other words, we assume that all delay vectors come from the same attractor. If fitting a single population or independent model, ρ reverts back to the mode of the prior, 0.5.

In hierarchical models, the dynamic correlation can be fixed or estimated. In cases where dynamic correlation is set to 0 within a hierarchical model, this will still yield different results when compared to independently fit models. This is because the hierarchical model shares information among the estimated length scale parameters ϕ for each embedding parameter.

3.2.1.5 Length Scale

Length scale parameters ϕ and the number of model inputs ($i = m + z$) define the complexity of the function represented by the GP. Each model input i incorporates an additional dimension of space, and their associated length scale parameter ϕ_i defines the wiggliness in that dimension. Low values of ϕ indicate stiff and mostly linear relationships, and large values of ϕ indicate more nonlinear relationships. A model with a single input and large ϕ_1 would have many degrees of freedom, while a model with many inputs but all ϕ_i close to 0 would have relatively few degrees of freedom (Tsai et al. 2024).

3.2.1.6 Data Transformations

Possible data transformations on the population are defined below. This is referred to as 'ytrans' in the GP-EDM R Package, but was defined as X_t above. This is the transformation that is applied before fitting the model.

- none: no transformation
- log: log transformation ($log(X_t)$)
- gr1: log difference transformation ($log(X_t/X_{t-1})$)
- gr2: log difference transformation on escapement ($log\left(X_t/(X_{t-1} - q * C_{t-1})\right)$)

3.2.1.7 Covariates

The underlying theory of EDM is that lags of the population have information on population drivers embedded within them (Munch et al. 2020). It is possible to include some covariates directly in EDM that are believed to influence population abundance. In the case of Gulf penaeid shrimp, economic conditions have had a massive impact on the domestic fishery, which in turn directly influences the amount of shrimp left in the water. Additionally, it has been hypothesized that environmental drivers such as salinity and temperature in the shrimp nursery grounds may have a direct impact on recruitment to the population the following year (Turley et al. 2023).

While covariates have the potential to improve model fits and short-term predictive accuracy, relying on lags of the population alone for estimating the biological MSY is simpler from an operational standpoint. Including covariates in the model requires making some assumption about the future states of that covariate in projections, which cannot be done with high confidence in this context. In addition, some of the variables considered may contain some level of covariation which the model is not set up to account for in its present form.

3.2.1.8 Cross Validation

Two different cross validation approaches were explored to evaluate prediction accuracy: "leave time out" and "sequential". Prediction method "leave time out" leaves out all data points (i.e., survey data, catch, covariates) taken at the same time across all populations where population is specified within hierarchical models. The "sequential" prediction method leaves out all future time points across all populations where population is specified. In both of these methods, training data are iteratively omitted for the predictions, but the inverse length scales and variances used are those obtained using all of the training data under the originally fit model. We anticipate that "sequential" would perform worse when compared to "leave time out". Both cross validation approaches were applied to all model configurations, but ultimately the "sequential" method was preferred for model selection because our ultimate objective is to project landings and harvest rates into the future in order to accurately estimate the system's maximum sustainable yield for fishery management.

3.2.2 Goodness of Fit

Goodness of fit was measured through the estimation of R^2.

In sample fit statistics for each prediction method:

- R^2 - proportion of variance explained by model (independent or hierarchical)
- R^2_{pop} - proportion of variance explained for each population within a hierarchical model
- R^2_{scaled} - proportion of variance explained by a hierarchical model, centered and scaled by population means
- $rmse$ - root mean square error
- df - degrees of freedom, trace of the smoother matrix

Out-of-sample fit statistics for each prediction method:

- R^2_{out} - out-of-sample R^2
- R^2_{outpop} - out-of-sample R^2_{pop}
- $R^2_{outscaled}$ - out-of-sample R^2_{scaled}
- $rmse_{out}$ - out-of-sample $rmse$

These fit statistics measure the models' overall performance and ability to perform outside of the training data. Within hierarchical models, population-specific R^2_{pop} metrics measure the model's ability to track the individual populations. For example, a model may be able to track one population well, but may fit another poorly. These population-specific R^2_{pop} metrics were centered and scaled around their respective model means in the R^2_{scaled} fit statistics to more appropriately measure the overall model performance. Population-specific R^2_{pop} and R^2_{scaled} statistics were compared to R^2 statistics obtained from independent model fits of each population to ensure that the complexity of the hierarchical model was warranted (i.e. improved overall prediction skill).

3.2.3 Estimated Parameters

Parameters estimated and priors specified in GP-EDM are defined below.

- $\phi_1 : \phi_i$ - length scale parameters for $1: i$ where i is the total m lags and z covariates ($i = m + z$); priors are set such that the expected number of local extrema for each ϕ_i is 1 (Munch et al. 2017)
- V_e - process variance
- τ - pointwise prior variance in f

- ρ - dynamic correlation between populations where values range from 0 to 1, with 0- independent no correlation and 1- identical dynamics
- q- catchability scalar that translates the units of landings into units of survey CPUE

The relative magnitude of the pointwise prior variance τ and process noise V_e gives information on how important the function is relative to the noise. Process variance is represented as a percentage of the total variance, whereas the pointwise prior variance cannot be directly translated to variance percentage because it interacts with the length scale parameters. If the model is purely deterministic, $V_e = 0$ and $\tau \approx 1$. If the model is not fitting the data well, τ is small and the process variance is close to 1.

Catchability could be estimated jointly (q=shared) or separately for each population in each model configuration. In some instances, the model obtained very good fits, but estimated catchability $q = 0$ and ignored the landings altogether. For the purposes of our work here, the link to landings is critical. To select a representative model for estimating MSY, the models were filtered to exclude any model where catchability < 0.001 (where the observed catchability in the data were typically above 0.01).

3.2.4 Estimating Maximum Sustainable Yield (MSY) with EDM

Maximum Sustainble Yield (MSY) estimates were generated following the methodology outlined in Tsai et al. (2024). Harvest rates ranging from 0:1 were projected into the future and an average of the long-term dynamics were taken for each population, then added up to obtain estimates of long-term landings. These averages were used to identify the harvest rate that maximizes landings. Models that were configured seasonally required landings and associated harvest rates to be translated to annual scales. Translating catch from a seasonal to an annual time scale was fairly simple

$$C_t = 2 * C_{t/2}$$

where t is defined as one year here, and $t/2$ represents 2 seasonal steps per year. Annual harvest rate U_t was estimated from a seasonal harvest rate $U_{t/2}$ as

$$U_t = 1 - \left(1 - U_{t/2}\right)^2$$

where the new estimated harvest rate U_t captures the portion of the population (0:1) removed via landings over the course of a year. Here, the estimated long-term biomass associated with the rate of removals does not need to be changed. The annual harvest rate U_t was further translated to an annual fishing mortality rate $F_t = -ln(1 - U_t)$. This allows for the calculation of the more familiar benchmark F_t/F_{MSY}, which is a measure of overfishing (estimated to be occurring if $F_t/F_{MSY} > 1$).

3.2.5 Model Diagnostics

Models were diagnosed and deemed reliable based on a set of criteria defined below. This methodology worked well for all Gulf shrimp species assessed within SEDAR 87. These decisions were applied to 'no covariate' models, since assumptions on the cyclical nature of environmental variables and the relationship between harvest rate and economic variables would be required for projections. It was determined that these assumptions should be avoided for the purposes of defining biological maxima if possible. The projection period was initially set to 50 timesteps then extended to 80 to ensure the reference points had stabilized before taking an average. The duration over which to average was determined by the length of a cycle, which was typically driven by the seasonal time steps in the model if present. The estimate of MSY is sensitive to setting an appropriate projection period that ensures the population has stabilized and an appropriate save interval that ensures only complete cycles are clipped, the latter ensures the estimate is not biased high or low (as would be observed if the time step just outside of a completed cycle is increasing or decreasing, respectively).

3.2.5.1 Model Fitting Performance

Model performance was determined by considering the suite of Goodness of Fit parameters defined above. The top 30 models from the hierarchical overall R^2_{out} and top 30 models from the $R^2_{outscaled}$ were pulled, and any overlapping models were considered. The top 5 from each of these criteria and the top 5 aggregated Gulf-wide models were considered to evaluate what was gained from added complexity.

3.2.5.2 Model Projection Performance

Projection performance was evaluated to ensure models extrapolate to MSY in a reasonable way. Model selection was already performed with this goal in mind when relying on predictmethod=sequential to obtain fit statistics. Additional diagnostics were developed to cull out unreasonable models. This included removing models that maximized catch at $U = 1$, which generally happened when models would predict that the population returns to the flat prior outside of the observed range of the data. These models were often paired with unrealistically high catch estimates due to the coupling of extreme harvest rates with populations that did not always ground to zero. It is intuitively not sustainable to remove the entire population, so these models were removed. Unrealistically high estimates of MSY were defined as greater than ten times the highest historic landings.

3.2.5.3 Model Robustness

From the remaining set of models that (1) had good fits, (2) did not solve on a bound ($U = 1$), and (3) did not estimate MSY at greater than 10x historical landings records, a retrospective pattern analysis was carried out where 1 to 5 time steps were peeled back and MSY was re-estimated. The Model Projection Performance selection criteria defined above were applied to each of these iterations. If any iteration failed, it was dropped from further consideration. This resulted in a final selection that balances model complexity and relative stability.

4. Stock Assessment Model Results

4.1 JABBA Results

4.1.1 Model Fit and Diagnostics

Diagnostic results for the top performing JABBA model runs are presented in *Figure 15*. Though these models were selected for being the best performing models of all the models tested, none of them performed well. There was no clear pattern in the CPUE residuals (*Figure 16*) but the models had no prediction power – the hindcast cross validation results were poor (MASE >1) with the model systematically underestimating the index value in years 2018 and beyond (*Figure 17*). All models exhibited strong patterns in the process error deviates, indicating that changes in biomass diverged from the model expectations. Models were very sensitive to prior assumptions made about r and initial depletion and a retrospective analysis of derived benchmarks showed sensitivity to final estimates as years of the model were peeled back (*Figure 18*). Detailed figures showing the results of each diagnostic test are reported for run 82.

4.1.2 Estimated Parameters and Derived Quantities

Estimated parameters for these models are provided in *Table 5* and *Figure 19*. Model posteriors did not deviate significantly from the priors (except for process error) as there was not much contrast in the data to inform the underlying surplus production model (*Figure 20*). Furthermore, some information could be seen as conflicting, in this developed fishery, both the landings and the index are seen as increasing between 2010 and 2022. Standard biomass dynamics might expect the removals to have a negative effect on the population size and could lead to errant results here.

For all biomass trajectories, the error remains large throughout the duration of the time frame (*Figure 21*). For the models being considered here, MSY was estimated between 14-30 million pounds of tails, a more than twofold difference in estimates (*Table 6*). There is a wide range of unfished biomass estimates shown in the surplus production models considered here (*Figure 22*), and temporal fits to B/B_{MSY} are relatively unaffected by the initial depletion prior (*Figure 23*). Most models dip below B/B_{MSY} periodically throughout the duration of the data until recent years, when the fishery began targeting almost exclusively Large shrimp (where the shift in size structure is not accounted for in biomass dynamic models).

4.2 EDM Results

Over 1,200 model configurations were evaluated for Pink Shrimp to explore assumptions and ensure that results from the various iterations made sense. Up to the maximum embedding dimension was considered, with preference given to simpler models where possible. Estimated parameters, model fits, and projection capabilities are discussed below.

4.2.1 Model Configurations

Model configurations were examined to test assumptions and ensure results were as expected. The impact of using the ‘sequential’ method when defining the training dataset for prediction accuracy is shown in *Figure 24*, where ‘leave-time-out’ generally yielded a higher out-of-sample R^2_{out} fit. In hierarchical models, large differences in population means could artificially inflate the R^2_{out} metric. Therefore, metrics were calculated to estimate goodness-of-fit that were centered and scaled around the population mean, $R^2_{scaledout}$. Hierarchical out-of-sample R^2_{out} generally yielded a higher value than the out-of-sample scaled by population-specific fits, $R^2_{scaledout}$. With these models, the goal is to fit and project each population within the model well, and $R^2_{scaledout}$ was the primary metric used to gauge model fits going forward.

Assumptions of global and local scaling were shown across all model configurations (embedding dimension, population, y transformation) using out-of-sample scaled $R^2_{scaledout}$ fit statistic. Local scaling generally yielded better fits compared to global scaling, which agrees with what we understand about these populations and their relative biomass (*Figure 25*). At this stage, only models with local scaling were considered for all configurations (except for models with one "population" where global scaling is inherent).

Some of the reported R^2 metrics were associated with models that ignored landings (i.e. $q \approx 0$). *Figure 26* shows the distribution of R^2 after these models were removed. From the set of models that account for landings, additional models were excluded due to the fact that they included covariates. Because of the short data series available for Pink Shrimp, removing the models with covariates resulted in relatively poorly fitting models remaining (*Figure 27*). This figure shows the model configurations that were analyzed for fit and eventual MSY estimation.

4.2.2 Model Fit and Residual Analysis

From the set of models described in the previous section, the procedures outlined in *Section 3.2.5.1* were applied, i.e. ranking the models by out-of-sample prediction accuracy for the model as a whole (R^2_{out}), scaled by population ($R^2_{outscaled}$), or both (*Section 3.2.2*). This resulted in 30 models going through MSY estimation and further model diagnostics (*Table 7*).

4.2.3 Model Diagnostics

The subset of models with the best fits were analyzed for consistent and robust projections. These models were investigated considering projection ability and model robustness outlined in *Section 3.2.5* down, but no models were acceptable. At this stage, no models were ran through retrospective analyses. The maximum landings for Pink Shrimp were observed in 1964 with landings totaling 20.99 million pounds of tails, but the maximum landings input into the model were 12.99 million pounds of tails, observed in 2018. While the true maximum value of landings was used as a threshold for gauging reasonable outputs, from a diagnostics perspective, it is important to also consider what information the model has predicted based on the inputs themselves. Models with estimates of MSY less than 5x the historical maximum are provided in *Table 8*.

Variable harvest rate projections of CPUE by size class for the model with the highest R^2 are shown in *Figure 28* - *30* and associated landings are shown in *Figure 31* - *33*. These data series were used to generate average biomass and landings under all harvest rates 0:1 in *Figure 34* and *Figure 35*. The population metrics for B_{MSY} and MSY are shown in *Figure 36* and *Figure 37* where dashed lines represent the annual rate at MSY.

The horizontal dotted line on the MSY figure (*Figure 37*) shows the maximum landings incorporated into the model. The maximum landings ever caught by the fishery was 20.99 million pounds of tails in 1964, which is less than half the projected MSY, 52.06 million pounds of tails. If any size class models were to be used for management, they would need to be scaled up by a factor using the proportion of landings with missing size information to account for total removals (*Table 3*).

4.2.4 Estimated Parameters and Derived Quantities

Estimated parameters from the top-performing model with failed diagnostics is shown in *Table 9*. The estimated length scale parameters ϕ_i are all less than 1, with the first two completely flat (=0). These are linear and smooth for all populations in the model (*Figure 38* - *41*). The pointwise prior variance was estimated to be 0.615, and the function process variance was estimated to be 0.544. The dynamic correlation of the model was 0.756, indicating a high correlation between these data, which can be observed visually in *Figure 2* and *Figure 3*. These size populations had distinct catchabilities that translates fishery removals to the units of the SEAMAP survey (number of shrimp per trawl hour divided by million pounds of tails): Large $q = 2.142$, Medium $q = 0.334$, and Small $q = 0.896$.

The R^2 statistics for the overall and population-scaled metrics both exceeded 0.5, but had no predictive power (*Table 10*). Out-of-sample $R^2_{out} = 0.297$ performed better than the $R^2_{outscaled} = 0$ scaled by population means, indicating the former was likely inflated by the magnitude of differences in population means.

4.2.5 Fishing Mortality

Fishing mortality rates were estimated based on model outputs are provided in *Table 11*. These are uninformative are not recommended for use.

4.2.6 Biomass and Abundance Trajectories

The Pink Shrimp stock has a relatively short time series of abundance available. It's difficult to tease out any increasing or decreasing trend over this interval, and model fits for the data from the model with no predictive capability are shown in *Figure 42* and *Figure 43*.

5. Discussion

Neither EDM nor JABBA models were recommended for providing Pink Shrimp stock status determination criteria. The JABBA models performed poorly across the board and were limited by the general constraints of surplus production models which aggregate dynamics, not accounting for size or spatial differences. This poor model performance was likely driven by the limited contrast present in the data, with landings exhibiting a one-way trip throughout the time period covered by the index and the index not appearing to respond to this increase in catches in a logical way. In addition, it is likely that Pink Shrimp stock dynamics are driven more by environmental and economic factors than by the catches, which could lead the stock to appear to respond to the fishery in unexpected ways (e.g., an increase in population alongside the increase in landings). The Gulf Pink Shrimp fishery landings size compositions have changed through time (*Figure 44*) due to global market conditions and increasing demand and prices yielded for Large shrimp. This likely causes a mismatch between the time series of CPUE and catches in terms of what each is indexing, further confusing the size aggregated model. Additionally, as the domestic fleet consolidated, larger and more efficient vessels remained and could be trawling in a different habitat compared to the historic distribution of the fleet. Lastly, it is well documented that physical changes to the Everglades has altered freshwater flows into South Florida estuaries throughout the last half century and that resulting changes in salinity have a direct impact on Pink Shrimp productivity in the area. The potential change in carrying capacity of the system brought about by these drastic changes in the extent of available habitat cannot be adequately captured in JABBA. These factors are all justifications against using surplus production models that assume catch levels reflect only changes in stock abundance and that patterns of exploitation are primarily driven by shrimp availability rather than environmental or economic considerations.

EDM models presented here did not have sufficient predictive capability to estimate MSY. Stacking the time series in a hierarchical model can improve the model's ability to fit to the data, sharing information among populations through length scales and dynamic correlation parameters. Hierarchical size models allowed for poor fits in one size class to be improved by informing it with trends from another size class but were still insufficient at predicting into the future, likely due to the relatively short time series available (T=13).

EDM was able to capture the cyclical nature of shrimp population abundance but was unable to predict outside of the data or project this information into the future. Lags of the population retain information on sometimes immeasurable drivers, including abundance of predators and important environmental influences. Direct inclusion of environmental and economic covariates improved model fits and allowed for better prediction for Pink Shrimp in some cases. They were not used to estimate MSY because additional assumptions would be required to forecast the state of the industry and environment into the future. Furthermore, relationships between the simulated harvest rates and these covariates responded in unexpected ways, requiring additional research on covariate impacts on escapement. Given the goal of providing a biological MSY estimate for this fishery, models with only lags of abundance were presented here. The models with covariates may serve other purposes and could be used to predict year-ahead abundance and landings more accurately than the model with lags alone, particularly for those tied to economic drivers.

If size structured EDM models are able to be used in the future, the landings with missing size data must be corrected. It is strongly recommended that this information is collected more reliably in the future since it provides such valuable information to this model through shared parameters across size classes. Additionally, feedback loops across size classes should be investigated if these models are pursued for management in the future. A harvest rate that maximizes Small shrimp may cause a negative feedback loop on Large shrimp (or vice versa) that is not accounted for in the current structure. Accounting for this type of negative feedback loop through mixed-age configurations is possible (Dolan et al. 2023), but it is complicated by mixing landings across calendar years to fit to population escapement, which could markedly increase management complexity, perhaps unnecessarily.

For Pink Shrimp, no models presented here are recommended for providing status determination criteria or fishing mortality metrics. Alternate strategies could be used including the third highest catch as a reference point and/or monitoring the trend of abundance using the VAST index.

6. Research Recommendations

Additional research into covariates should be investigated. Direct inclusion of covariates resulted in improved model fits and could likely improve forecasting efficiency for trends of abundance. To forecast MSY, covariates would need to be projected into the future. For environmental covariates, the cyclical nature of these trends would need to be captured. For economic covariates, the relationship with projected harvest rates would need to be explicitly defined.

Potential improvements to the modeling framework include accounting for removal of shrimp as it pertains to harvest rates that are optimized at varying size classes. Creating a feedback loop that appropriately represents the removal of larger shrimp that may not contribute to future generations as well as the removal of smaller shrimp that may not grow into Large shrimp should be accounted for. Sensitivities of these potential feedback loops and their impact on estimating optimal harvest rates should be investigated in both directions (i.e. Large to Small and Small to Large impacts).

As funding for scientific surveys is becoming increasingly sparse, implications of using an average of 2019/21 for missing summer 2020 SEAMAP data and resulting effects on model diagnostics should be investigated. EDM performs best on continuous, long time series of data, and quantifying implications of future gaps in survey data would be valuable.

7. Acknowledgements

The SEDAR 87 Gulf Pink Shrimp Benchmark Assessment would not have been possible without the efforts of the numerous NMFS, SEFSC, SERO, and GMFMC staff along with the many academic, research partners, and stakeholders involved throughout the Gulf listed in *Section 1.1.3*.

8. References

Ailloud, L., M. Stevens, B. Turley, A. Pollack, and D. Hanisko. 2025. Developing a fishery-independent index of relative abundance for Gulf of Mexico pink shrimp using VAST. Page 25pp. SEDAR, SEDAR87-AP-04, North Charleston, SC.

Anhøj, J., and A. V. Olesen. 2014. *Run charts revisited: A simulation study of run chart rules for detection of non-random variation in health care processes*. PLoS One 9(11):Article e113825.

Atkinson, S., A. Lowther, K. Dettloff, and S. Smith. 2024. Gulf of Mexico commercial brown, pink and white shrimp landings. Page 38pp. SEDAR, SEDAR87-DW-06, North Charleston, SC.

Browder, J. A., Z. Zein-Eldin, M. M. Criales, M. B. Robblee, S. Wong, T. L. Jackson, and D. Johnson. 2002. *Dynamics of pink shrimp (Farfantepenaeus duorarum) recruitment potential in relation to salinity and temperature in Florida Bay*. Estuaries 25(6):1355–1371.

Carvalho, F., H. Winker, D. Courtney, M. Kapur, L. Kell, M. Cardinale, M. Schirripa, T. Kitakado, D. Yemane, K. R. Piner, M. N. Maunder, I. Taylor, C. R. Wetzel, K. Doering, K. F. Johnson, and R. D. Methot. 2021. *A cookbook for using model diagnostics in integrated stock assessments*. Fisheries Research 240:105959.

Dolan, T. E., E. P. Palkovacs, T. L. Rogers, and S. B. Munch. 2023. *Age structure augments the predictive power of time series for fisheries and conservation*. Canadian Journal of Fisheries and Aquatic Sciences:cjfas-2022-0219.

Froese, R., H. Winker, G. Coro, N. Demirel, A. C. Tsikliras, D. Dimarchopoulou, G. Scarcella, M. L. D. Palomares, M. Dureuil, and D. Pauly. 2019. *Estimating stock status from relative abundance and resilience*. ICES Journal of Marine Science 77(2):527–538.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Pages 169–193 *in* A. P. D. J. M. Bernardo J. O. Berger and A. F. M. Smith, editors. Bayesian statistics. Clarendon Press, Oxford.

Griffith, D., C. Liese, M. Travis, M. Freeman, and D. Records. 2023. Social dimensions of Gulf of Mexico shrimping. Page 12pp. SEDAR, SEDAR87-DW-15, North Charleston, SC.

Haddon, M. 2021. Surplus production models. Pages Chapter 7 Using r for modelling and quantitative methods in fisheries. Chapman; Hall, London, UK.

Heidelberger, P., and P. D. Welch. 1983. Simulation run length control in the presence of an initial transient. Operations Research 31:1109–1144.

Hurtado-Ferro, F., C. S. Szuwalski, J. L. Valero, S. C. Anderson, C. J. Cunningham, K. F. Johnson, R. Licandeo, C. R. McGilliard, C. C. Monnahan, M. L. Muradian, K. Ono, K. A. Vert-Pre, A. R. Whitten, and A. E. Punt. 2015. *Looking in the rear-view mirror: Bias and retrospective patterns in integrated, age-structured stock assessment models*. ICES Journal of Marine Science 72(1):99–110.

Kell, L. T., A. Kimoto, and T. Kitakado. 2016. *Evaluation of the prediction skill of stock assessment using hindcasting*. Fisheries Research 183:119–127.

Kell, L. T., R. Sharma, T. Kitakado, H. Winker, I. Mosqueira, M. Cardinale, and D. Fu. 2021. *Validation of stock assessment methods: Is it me or my model talking?* ICES Journal of Marine Science 78(6):2244–2255.

Liese, C. 2024. Price indices for shrimp imports and Gulf of Mexico shrimp landings by size and season. Page 10pp. SEDAR, SEDAR87-AP-2, North Charleston, SC.

Lowther, A. 2023. Shrimp import data. Page 2pp. SEDAR, SEDAR87-DW-10, North Charleston, SC.

Munch, S. B., A. Brias, G. Sugihara, and T. L. Rogers. 2020. *Frequently asked questions about nonlinear dynamics and empirical dynamic modelling*. ICES Journal of Marine Science 77(4):1463–1479.

Munch, S. B., V. Poynor, and J. L. Arriaza. 2017. *Circumventing structural uncertainty: A Bayesian perspective on nonlinear forecasting for ecology*. Ecological Complexity 32:134–143.

Munch, S. B., T. L. Rogers, and G. Sugihara. 2022. *Recent developments in empirical dynamic modelling*. Methods in Ecology and Evolution:1–14.

Pella, J. J., and P. K. Tomlinson. 1969. *A generalized stock production model*. Inter-American Tropical Tuna Commission Bulletin 13:419–496.

Pollack, A. G., and D. S. Hanisko. 2024. Brown, white and pink shrimp abundance indices from SEAMAP groundfish surveys in the northern Gulf of Mexico. Page 113pp. SEDAR, SEDAR87-DW-13, North Charleston, SC.

Ricker, W. 1954. Stock and recruitment. J. Fish. Res. Bd. Canada 11(5):559--623.

Schlenker, L. S., C. Stewart, J. Rock, N. Heck, and J. W. Morley. 2023. *Environmental and climate variability drive population size of annual penaeid shrimp in a large lagoonal estuary*. PLOS ONE 18(5):e0285498.

SEDAR87 Data Workshop Report. 2023. Page 94pp. SEDAR, SEDAR87, North Charleston, SC.

Sheridan, P. 1996. Forecasting the fishery for pink shrimp, penaeus duorarum, on the tortugas grounds, florida. Fishery Bulletin 94:743–755.

Sugihara, G. 1994. Nonlinear forecasting for the classification of natural time series.

Sugihara, G., R. May, H. Ye, C. Hsieh, E. Deyle, M. Fogarty, and S. Munch. 2012. *Detecting Causality in Complex Ecosystems*. Science 338(6106):496–500.

Thorson, J. T., K. Ono, and S. B. Munch. 2014. *A Bayesian approach to identifying and compensating for model misspecification in population models*. Ecology 95(2):329–341.

Tsai, C.-H., S. B. Munch, M. D. Masi, and M. H. Stevens. 2024. *Empirical dynamic modeling for sustainable benchmarks of short-lived species*. ICES Journal of Marine Science:1–12.

Turley, B., L. Ailloud, and M. Stevens. 2023. Development of estuarine environmental indices for SEDAR 87 Gulf of Mexico white, pink, and brown shrimp stock assessment. Page 15pp. SEDAR, SEDAR87-AP-01, North Charleston, SC.

Winker, H., F. Carvalho, and M. Kapur. 2018. *JABBA: Just Another Bayesian Biomass Assessment*. Fisheries Research 204:275–288.

Zink, I. C., J. A. Browder, D. Lirman, and J. E. Serafy. 2018. *Pink shrimp Farfantepenaeus duorarum spatiotemporal abundance trends along an urban, subtropical shoreline slated for restoration*. PLOS ONE 13(11):e0198539.

9. Tables

Table 1: SEAMAP CPUE in number of shrimp per trawl hour for Pink Shrimp by size.

Year	Large	Medium	Small
2010	9.01	23.13	7.65
2011	5.24	13.01	27.45
2012	3.76	8.71	5.90
2013	7.74	7.41	2.13
2014	7.85	18.60	9.65
2015	9.13	10.85	3.15
2016	6.65	16.48	12.02
2017	6.11	16.11	6.04
2018	12.67	15.03	4.49
2019	6.94	5.68	1.96
2020	5.79	5.18	1.51
2021	4.64	4.67	1.06
2022	16.98	20.76	2.16

Table 2: SEAMAP CPUE in number of shrimp per trawl hour for Pink Shrimp.

Year	CPUE
2010	39.79
2011	45.71
2012	18.37
2013	17.28
2014	36.10
2015	23.13
2016	35.14
2017	28.26
2018	32.19
2019	14.58
2020	12.48
2021	10.37
2022	39.90

Table 3: Landings of Pink Shrimp in million pounds of tails, landings of Pink Shrimp with size category recorded in million pounds of tails, and proportion of landings with size category missing. This proportion has changed through time as data collection has changed (Atkinson et al. 2024).

Year	Landings	Landings_sizes	Proportion_missing
1960	20.66		
1961	9.46		
1962	15.33		
1963	18.00		
1964	20.99		
1965	14.11		
1966	12.99		
1967	8.97		
1968	10.17		
1969	9.89		
1970	11.93		
1971	10.12		
1972	10.81		
1973	13.99	13.99	0.00
1974	14.37	14.37	0.00
1975	13.75	13.75	0.00
1976	13.02	13.00	0.00
1977	16.20	16.20	0.00
1978	16.01	16.01	0.00
1979	13.85	13.85	0.00
1980	12.88	12.88	0.00
1981	18.77	18.77	0.00
1982	11.64	11.64	0.00
1983	12.63	12.63	0.00
1984	14.70	14.62	0.01
1985	15.93	15.74	0.01
1986	11.72	10.79	0.08
1987	10.49	10.32	0.02
1988	9.14	9.00	0.01
1989	8.62	8.48	0.02
1990	7.45	7.40	0.01
1991	6.79	6.77	0.00

Table 3 Continued: Landings of Pink Shrimp in million pounds of tails, landings of Pink Shrimp with size category recorded in million pounds of tails, and proportion of landings with size category missing. This proportion has changed through time as data collection has changed (Atkinson et al. 2024).

Year	Landings	Landings_sizes	Proportion_missing
1992	6.34	6.23	0.02
1993	9.49	9.41	0.01
1994	10.09	10.04	0.01
1995	14.06	14.01	0.00
1996	19.34	19.20	0.01
1997	12.69	12.52	0.01
1998	17.16	17.00	0.01
1999	8.03	7.95	0.01
2000	7.45	7.39	0.01
2001	9.70	9.57	0.01
2002	8.06	2.50	0.69
2003	8.07	3.58	0.56
2004	8.61	4.48	0.48
2005	7.27	3.55	0.51
2006	6.47	3.84	0.41
2007	3.46	1.89	0.45
2008	4.87	2.69	0.45
2009	4.03	2.47	0.39
2010	5.43	4.22	0.22
2011	4.55	3.34	0.27
2012	3.83	2.68	0.30
2013	4.03	2.64	0.34
2014	6.40	4.86	0.24
2015	5.54	4.00	0.28
2016	5.24	3.67	0.30
2017	11.39	8.41	0.26
2018	12.99	10.51	0.19
2019	7.76	6.42	0.17
2020	7.73	5.83	0.25
2021	7.93	6.40	0.19
2022	9.98	8.45	0.15

Table 4: Landings of Pink Shrimp in million pounds of tails by size class, where size is recorded.

Year	Large	Medium	Small
2010	0.87	2.87	0.48
2011	0.84	2.35	0.16
2012	0.67	1.63	0.38
2013	0.56	1.85	0.23
2014	0.89	3.26	0.72
2015	1.28	2.36	0.36
2016	1.22	2.15	0.30
2017	1.72	5.69	1.00
2018	3.26	6.51	0.73
2019	2.49	3.72	0.21
2020	1.80	3.62	0.40
2021	2.17	3.79	0.45
2022	2.70	5.12	0.63

Table 5: Pink Shrimp parameter estimates from JABBA where Runs are described using unique indentifiers (1:90), P indicates a Pella-Tomlinson surplus production curve with estimated shape parameter m, r is the relative level of the intrinsic rate of growth prior (M-Medium, H- High), psi is the initial depletion prior (0.9 low, 0.2 high), sig indicates whether additional observation error tau2 is estimated (T/F), and the last two numbers are the start year of the landings (1960). Median parameter estimates are provided with lower and upper credible intervals. K is reported in million lb tail weight.

Run	Parameter	Estimate	LCI.95	UCI.95
PSH_13_P_rH_psil0.9_sigF_60	K	116.96	57.04	400.45
PSH_13_P_rH_psil0.9_sigF_60	r	0.74	0.48	1.21
PSH_13_P_rH_psil0.9_sigF_60	q	0.01	0.00	0.02
PSH_13_P_rH_psil0.9_sigF_60	psi	0.87	0.55	1.29
PSH_13_P_rH_psil0.9_sigF_60	sigma2	0.03	0.01	0.04
PSH_13_P_rH_psil0.9_sigF_60	tau2	0.00	0.00	0.01
PSH_13_P_rH_psil0.9_sigF_60	m	1.29	0.69	2.62
PSH_16_P_rM_psil0.9_sigF_60	K	157.41	95.91	695.68
PSH_16_P_rM_psil0.9_sigF_60	r	0.29	0.17	0.53
PSH_16_P_rM_psil0.9_sigF_60	q	0.01	0.00	0.01
PSH_16_P_rM_psil0.9_sigF_60	psi	0.87	0.55	1.30
PSH_16_P_rM_psil0.9_sigF_60	sigma2	0.03	0.01	0.04
PSH_16_P_rM_psil0.9_sigF_60	tau2	0.00	0.00	0.01
PSH_16_P_rM_psil0.9_sigF_60	m	1.26	0.71	2.29
PSH_82_P_rM_psil0.2_sigF_60	K	197.65	98.78	451.24
PSH_82_P_rM_psil0.2_sigF_60	r	0.32	0.17	0.62
PSH_82_P_rM_psil0.2_sigF_60	q	0.00	0.00	0.01
PSH_82_P_rM_psil0.2_sigF_60	psi	0.32	0.14	0.72
PSH_82_P_rM_psil0.2_sigF_60	sigma2	0.03	0.01	0.04
PSH_82_P_rM_psil0.2_sigF_60	tau2	0.00	0.00	0.01
PSH_82_P_rM_psil0.2_sigF_60	m	1.12	0.63	2.03

Table 6: Pink Shrimp reference points from selected JABBA models in Table 5. K, Bmsy and MSY are reported in million lb tail weight.

Run	K	Bmsy	Fmsy	MSY
PSH_13_P_rH_psil0.9_sigF_60	116.96	50.35	0.57	27.50
PSH_16_P_rM_psil0.9_sigF_60	157.41	65.07	0.23	14.48
PSH_82_P_rM_psil0.2_sigF_60	197.65	77.45	0.29	20.74

Table 7: Pink Shrimp fit statistics for top performing models where run names are described by strata A:C, species, start year, landings units, shared catchability b (T/F), population, time step (YEAR2 is seasonal), embedding dimension E, scaling (global vs. local), and y transformations (log, gr1, gr2, none).

Run	R2_out	R2_outscaled
A20051_PSH2010_CPUEtailmp_bshareF_GULFYEAR_E4_global_ytransnone_ricker	0.128	0.128
C1035_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E3_local_ytranslog	0.327	0.016
C20039_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E4_local_ytransnone	0.297	0.000
C859_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E3_local_ytransnone	0.279	0.000
C10155_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E3_local_ytransnone	0.235	0.000
C309_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E2_local_ytranslog	0.221	0.000
C1013_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E2_local_ytranslog	0.221	0.000
C10133_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E2_local_ytransnone	0.160	0.000
C21069_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E4_global_ytransnone_ricker	0.154	0.000
C1519_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E3_global_ytransnone_ricker	0.150	0.000
C89_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E2_global_ytransnone	0.132	0.000
C793_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E2_global_ytransnone	0.132	0.000
C1695_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E3_global_ytranslog_ricker	0.120	0.000
C21037_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E4_global_ytransnone	0.120	0.000
C1497_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E2_global_ytransnone_ricker	0.115	0.000
C10815_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E3_global_ytransnone	0.114	0.000
C265_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E2_global_ytranslog	0.114	0.000
C10969_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E2_global_ytranslog	0.097	0.000
C10265_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E2_global_ytranslog	0.097	0.000
C10111_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E3_global_ytransnone	0.094	0.000
C10793_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E2_global_ytransnone	0.093	0.000
C10089_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E2_global_ytransnone	0.093	0.000
C12223_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E3_global_ytransnone_ricker	0.089	0.000
C20069_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E4_global_ytransnone_ricker	0.084	0.000
C11695_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E3_global_ytranslog_ricker	0.077	0.000
C11519_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E3_global_ytransnone_ricker	0.077	0.000
C12399_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E3_global_ytranslog_ricker	0.074	0.000
C21045_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E4_global_ytranslog	0.066	0.000
C287_PSH2010_CPUEtailmp_bshareT_SIZEYEAR_E3_global_ytranslog	0.062	0.000
C21109_PSH2010_CPUEtailmp_bshareF_SIZEYEAR_E4_global_ytranslog_ricker	0.052	0.000

Table 8: Pink Shrimp MSY estimates for models with predictive capability that are within 5x historical landings observed in 1964. The maximum landings input into the model was 12.99mp tails observed in 2018, and all models here converged on the maximum harvest rate = 1, which would advise to harvest the entire population. These models are not stable nor recommended, but outputs are investigated further.

Run	MSY	BMSY_mp	MSY_factor	F_drop
PSH_C20039	52.06	52.06	2.48	1
PSH_C859	78.01	78.01	3.72	1
PSH_C309	24.48	24.48	1.17	1
PSH_C1013	24.12	24.12	1.15	1
PSH_C1519	39.98	39.98	1.91	1
PSH_C89	33.91	33.91	1.62	1
PSH_C793	27.23	27.23	1.30	1
PSH_C1695	21.07	21.07	1.00	1
PSH_C1497	17.84	17.84	0.85	1
PSH_C265	34.94	34.94	1.66	1
PSH_C10969	31.31	31.31	1.49	1
PSH_C10265	20.67	20.67	0.99	1
PSH_C10111	66.13	66.13	3.15	1
PSH_C10793	30.83	30.83	1.47	1
PSH_C10089	47.22	47.22	2.25	1
PSH_C12223	49.07	49.07	2.34	1
PSH_C20069	24.57	24.57	1.17	1
PSH_C11695	41.21	41.21	1.96	1
PSH_C11519	47.88	47.88	2.28	1
PSH_C12399	44.36	44.36	2.11	1
PSH_C21109	36.08	36.08	1.72	1

Table 9: Pink Shrimp parameter estimates for the top performing model, which did not pass all diagnostic tests but is included here for discussion.

Parameter	PSH_C20039
CatchabilityLarge	2.142
CatchabilityMedium	0.334
CatchabilitySmall	0.896
DynamicCorrelation	0.756
LengthScale1	0.000
LengthScale2	0.000
LengthScale3	0.654
LengthScale4	0.224
PointwisePriorVariance	0.615
ProcessVariance	0.544

Table 10: Pink Shrimp MSY estimates for the top performing model, which did not pass all diagnostic tests but is included here for discussion.

Statistic	PSH_C20039
MSY_mptails	52.056
Fmsy	Inf
Umsy	1.000
Bmsy_mp	52.056
df	7.531
R2	0.682
R2Scaled	0.540
R2_outsample	0.297
R2Scaled_outsample	0.000

Table 11: Pink Shrimp status through time based on benchmarks from the top performing model, which did not pass all diagnostic tests but is included here for discussion. Landmp- landings in millions of pound of tails, Frate- fishing mortality rate, Best_mp- estimate of population size in millions of pound of tails, FFmsy- Frate relative to Fmsy, BBmsy- Best relative to Bmsy.

Year	landmp	Frate	Best_mp	FFmsy	BBmsy
2010	4.22	0.09	81.98	0	1.38
2011	3.34	0.06	72.03	0	1.59
2012	2.68	0.13	34.41	0	0.64
2013	2.64	0.12	28.16	0	0.60
2014	4.86	0.11	70.12	0	1.25
2015	4.00	0.18	40.25	0	0.80
2016	3.67	0.11	65.84	0	1.22
2017	8.41	0.26	57.82	0	0.98
2018	10.51	0.36	55.92	0	1.12
2019	6.42	0.62	22.44	0	0.51
2020	5.83	0.57	19.88	0	0.43
2021	6.40	0.94	17.33	0	0.36
2022	8.45	0.23	72.48	0	1.38

10. Figures

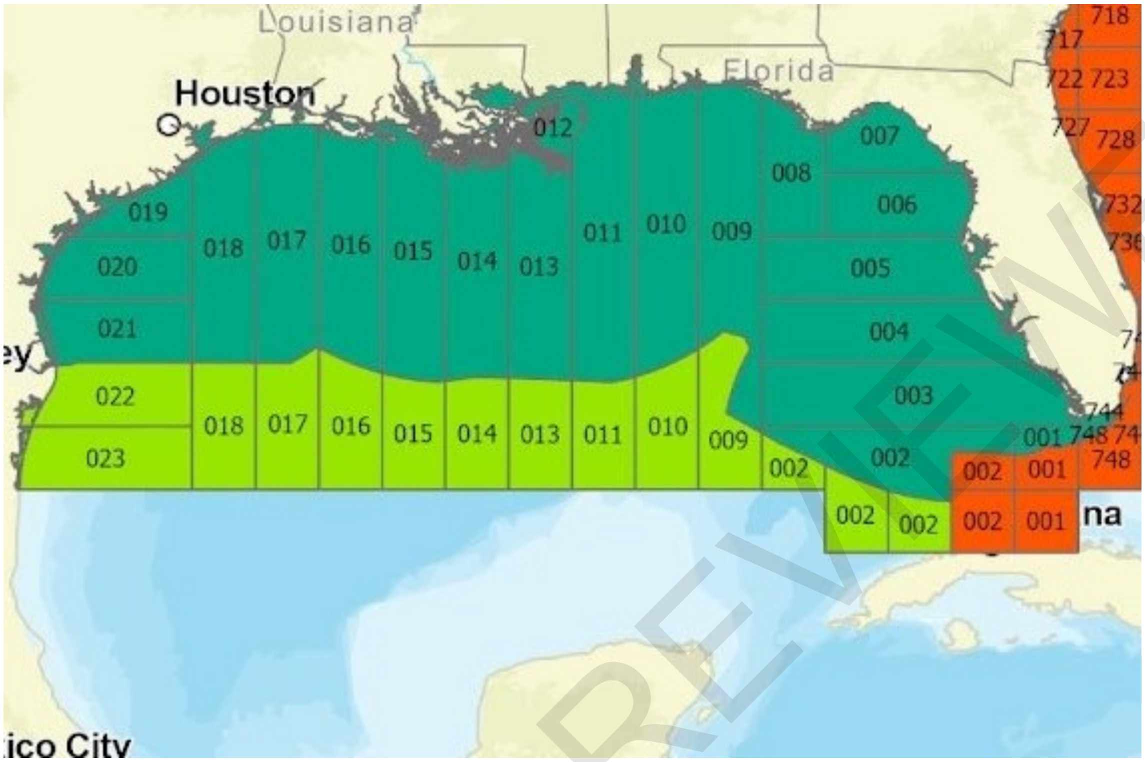


Figure 1: Map of the Gulf of America, where dark green is the Gulf defined by Gulf of Mexico Fishery Management Council boundaries, light green is Gulf international waters, and red is typically managed by the South Atlantic Fishery Management Council. Fishing areas 001 and 002 in their entirety were included in the analyses here per the recommendation of WP-06.

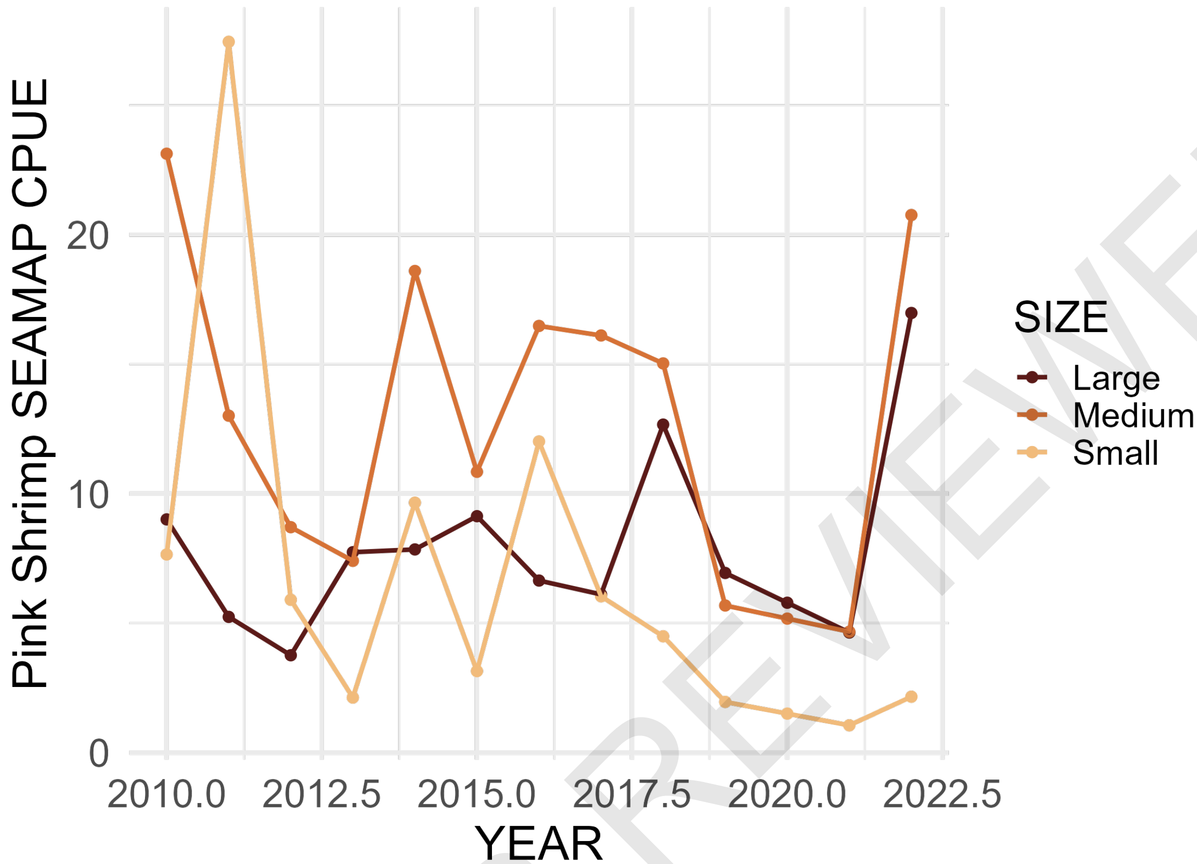


Figure 2: Pink Shrimp CPUE separated by size.

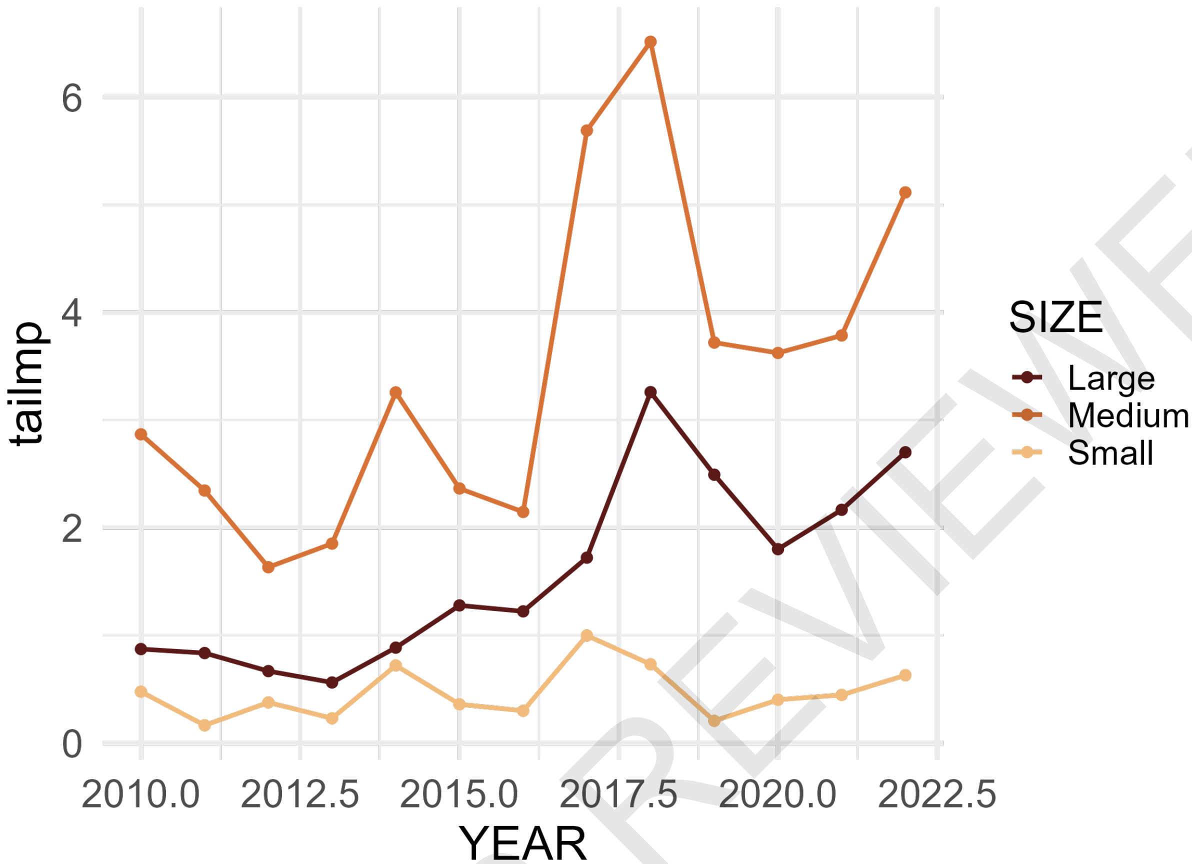


Figure 3: Pink Shrimp landings in millions of pounds of tails separated by size.

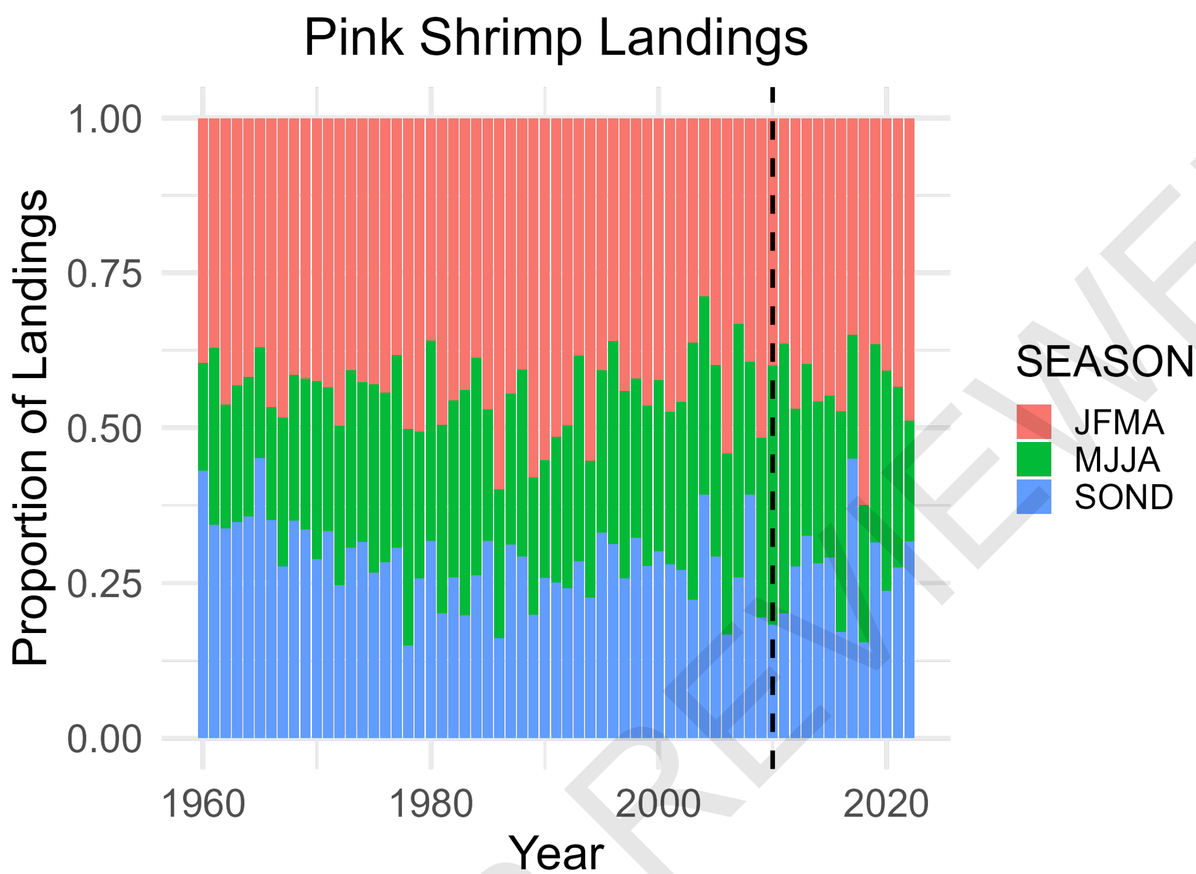


Figure 4: Seasonal distribution of Pink Shrimp landings.

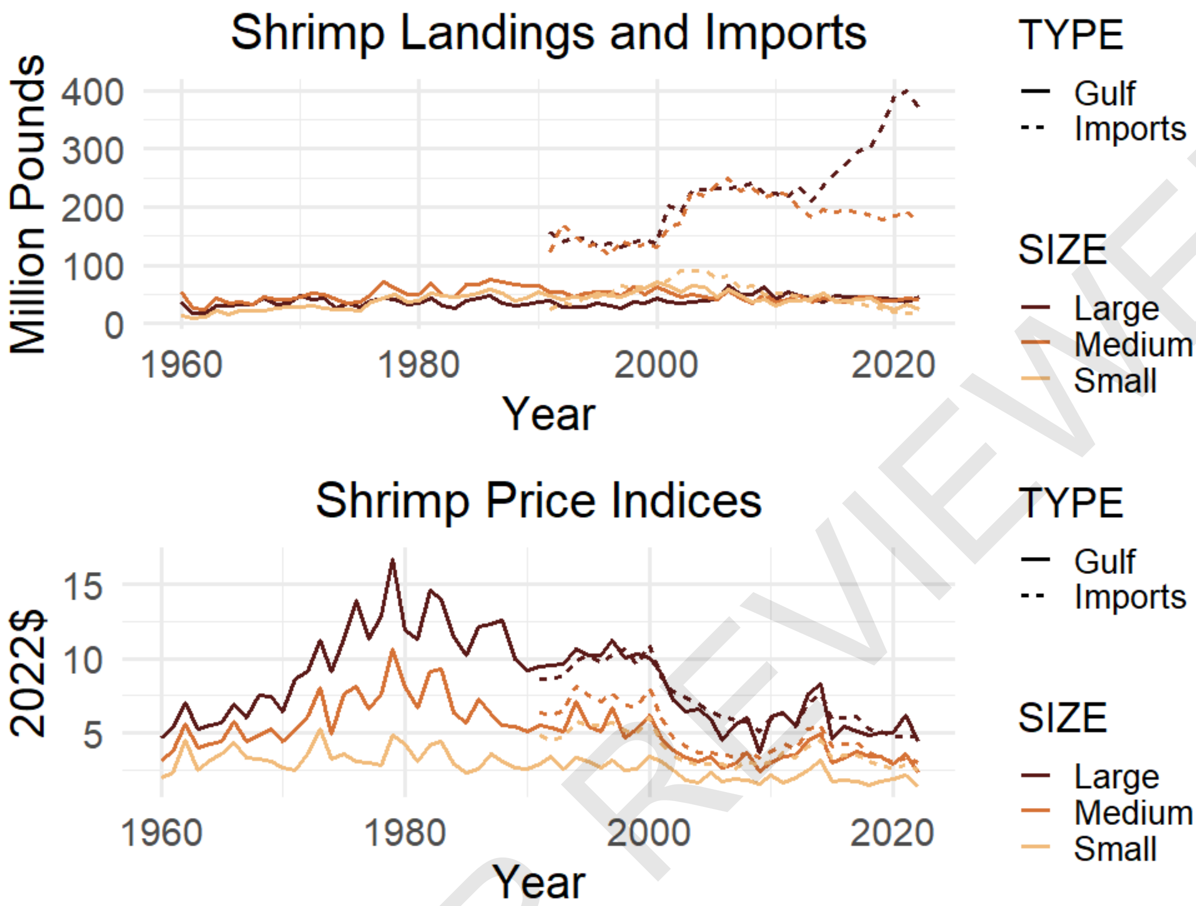


Figure 5: Domestic Gulf shrimp landings compared to global imports into the US by size category (top panel). This increase in supply has resulted in a crash of the ex-vessel price and domestic price index by size category, with all sizes decreasing, but Large yielding the highest amount (bottom panel).

Figure 6: Pink Shrimp environmental indices from Everglades National Park buoy data.

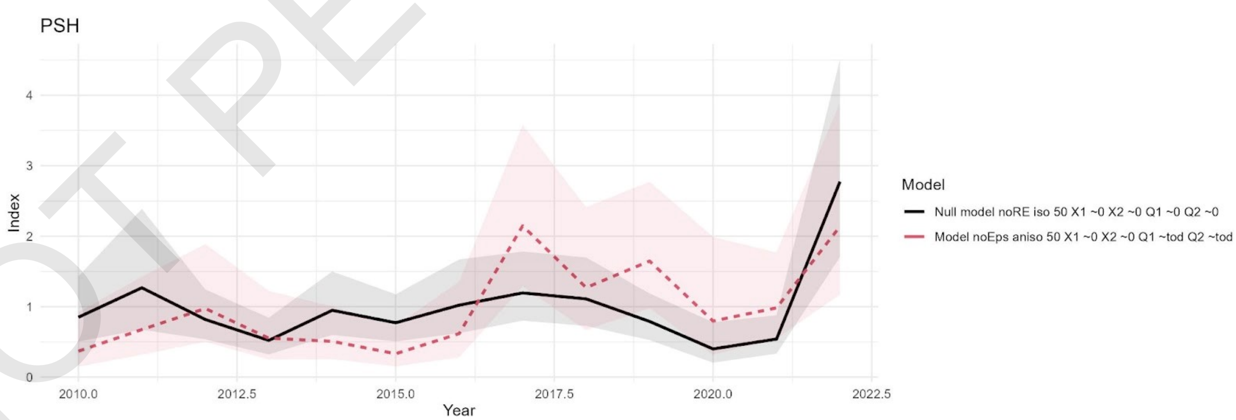


Figure 7: Final VAST index (red dashed line) and associated 95% confidence interval (red shading) incorporated into the JABBA model.

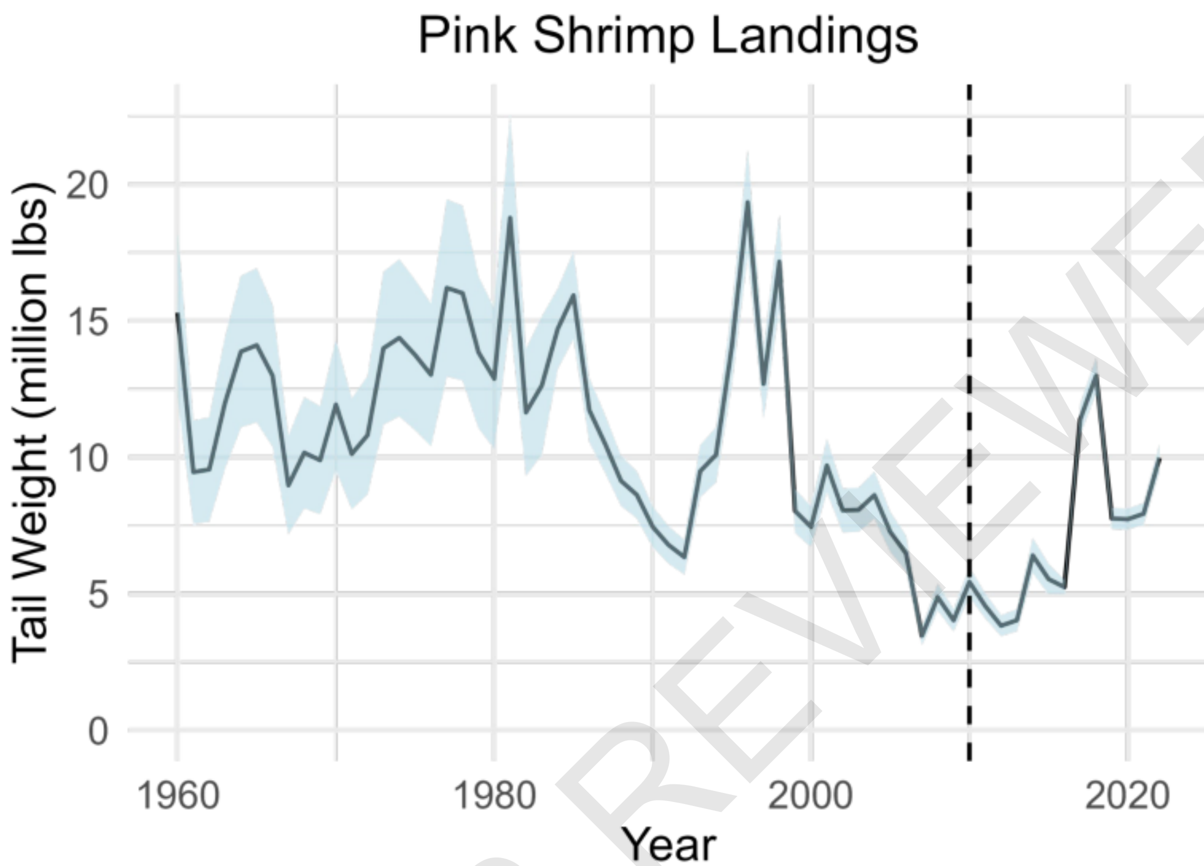


Figure 8: Final landings (blue line) and associated error (blue shading) input into JABBA. The dashed line indicates the start year of the index of relative abundance.

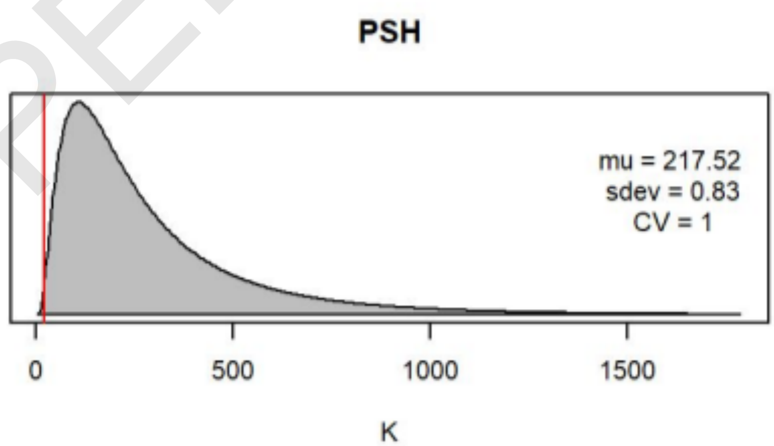


Figure 9: JABBA prior for carrying capacity, K, for all model configurations.

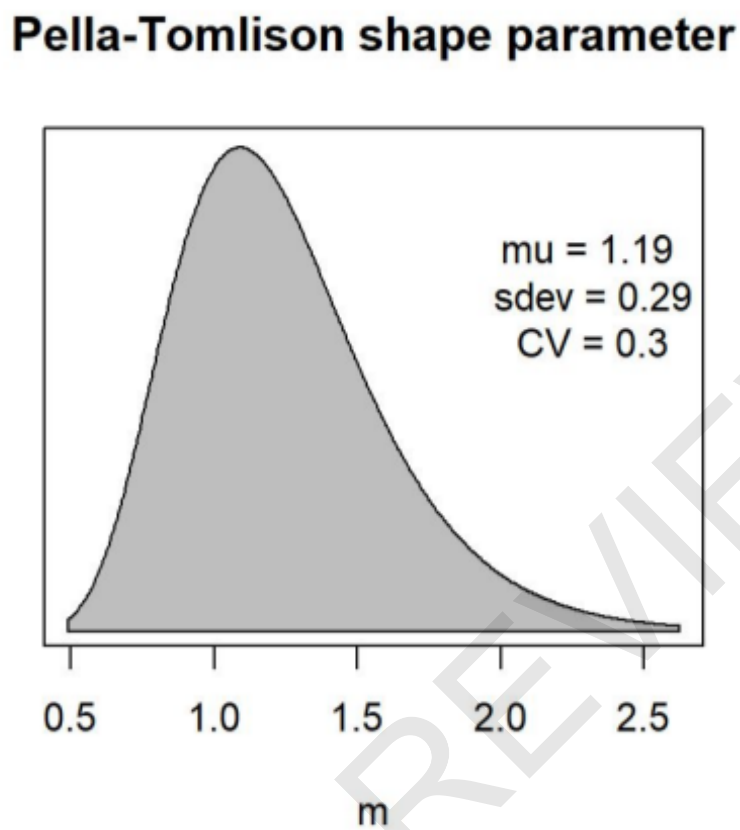


Figure 10: JABBA prior for Pella Tomlinson production function shape parameter, m, for all model configurations.

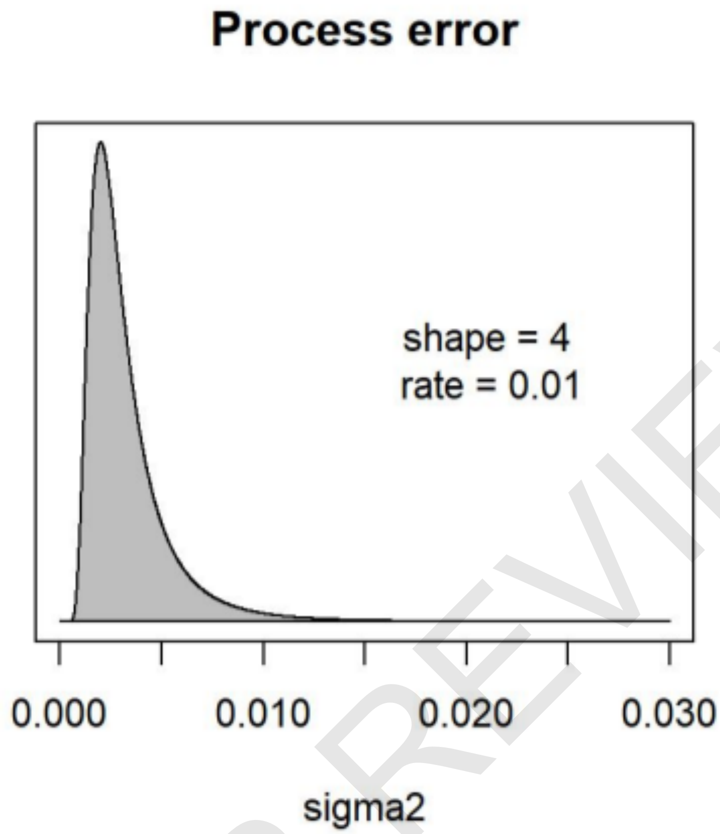


Figure 11: JABBA prior for process error for all model configurations.

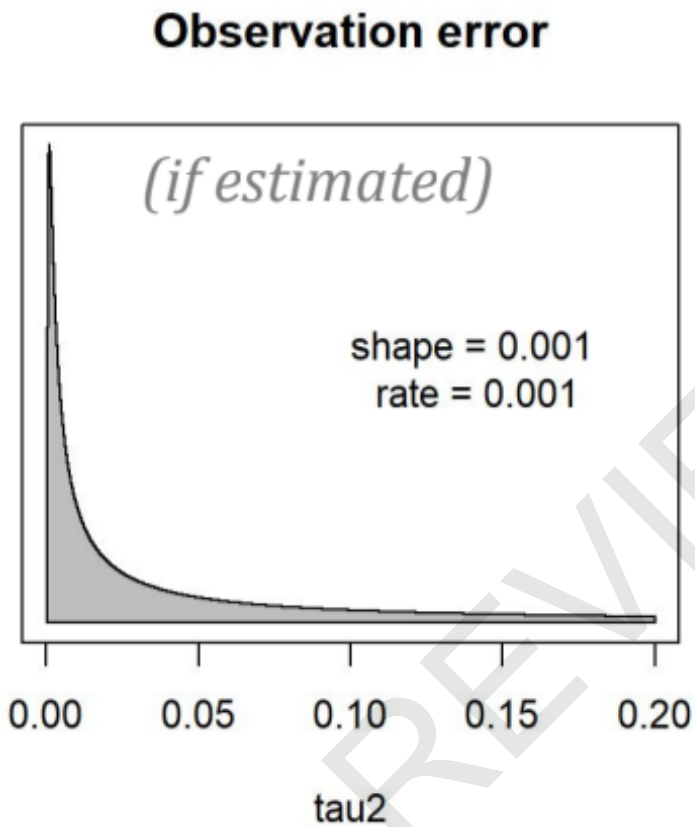


Figure 12: JABBA prior for observation error for all model configurations where estimated.

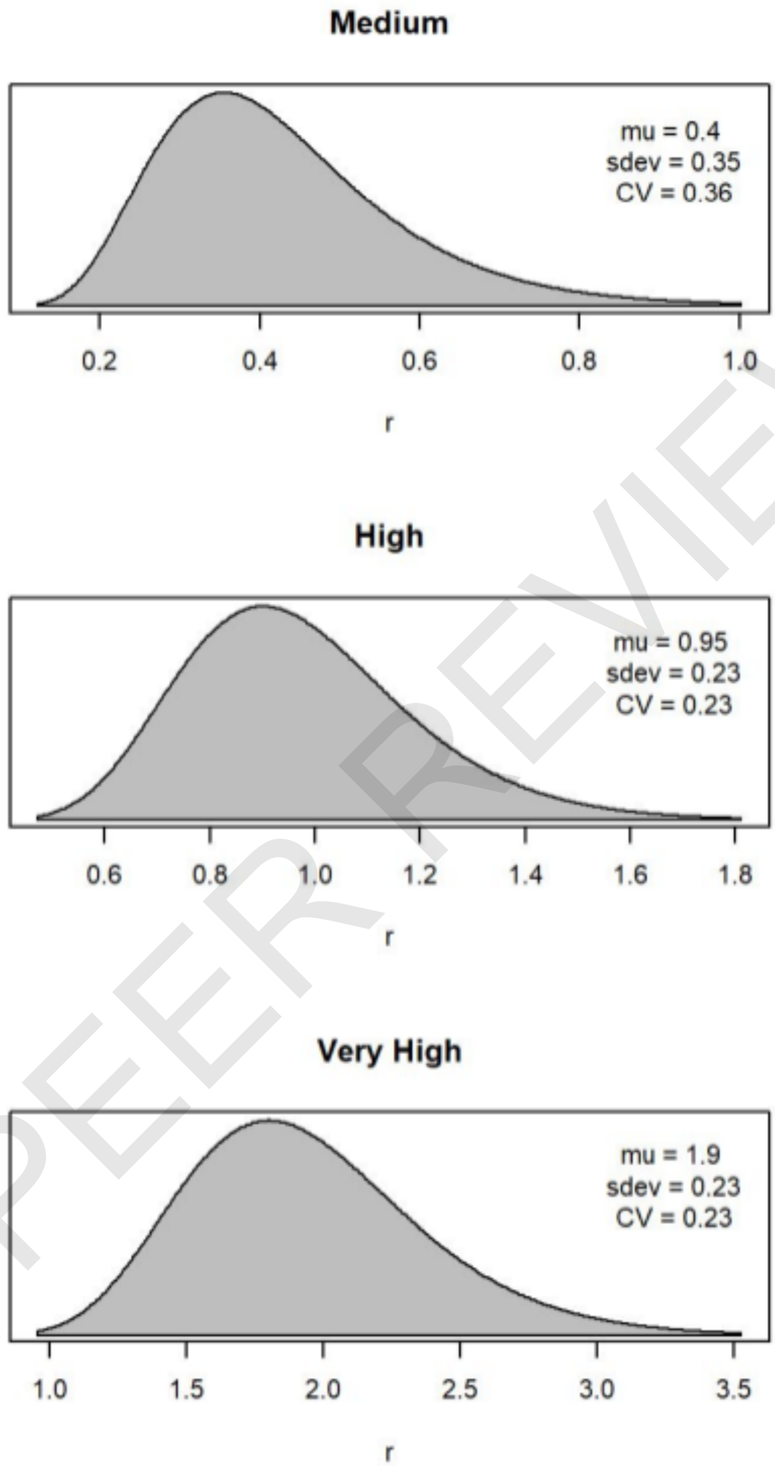


Figure 13: JABBA alternative prior assumptions for the intrinsic growth rate r.

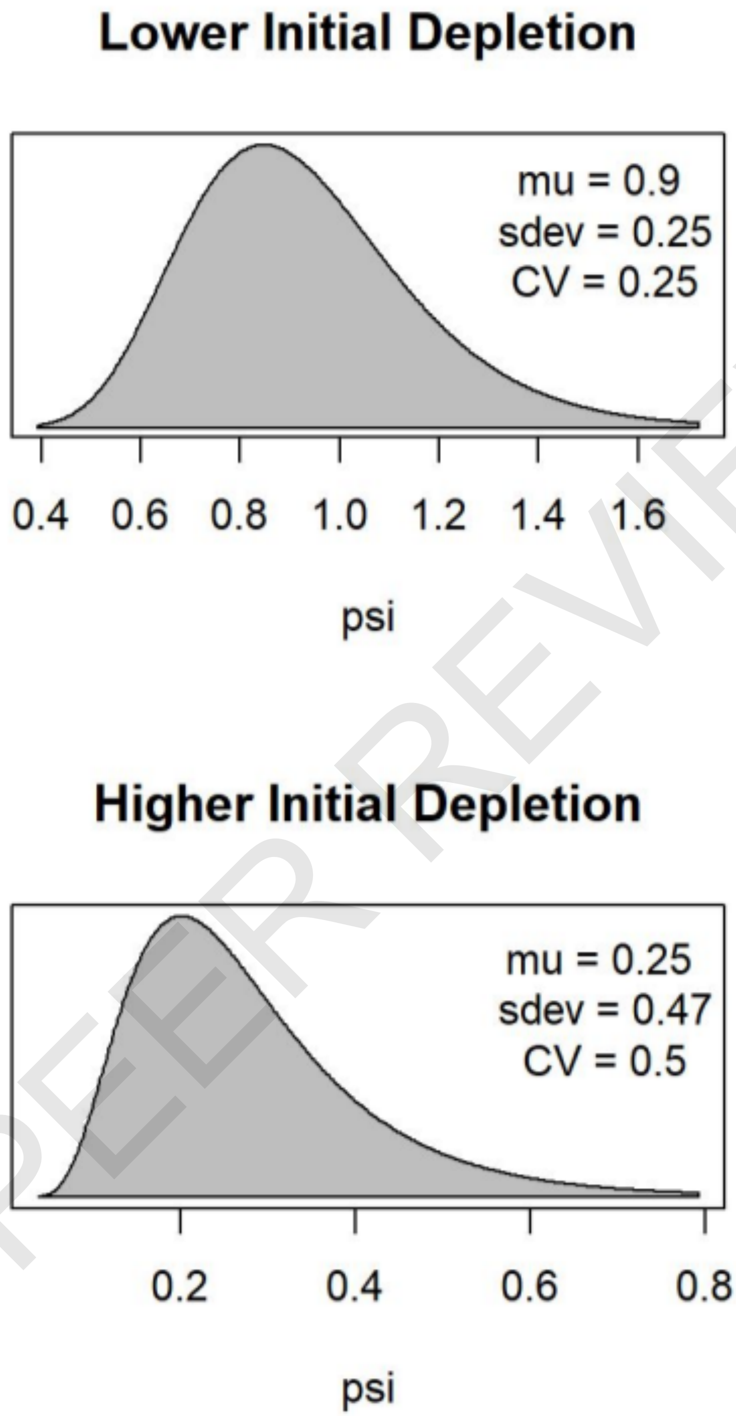


Figure 14: JABBA alternative prior assumptions for the initial biomass depletion ratio psi.

	Model Convergence			Model Fit		Model Consistency				Process Error	Prediction Skill	
run	CONV_gw	CONV_hw	CONV_hs	CPUE_rt_rand	CPUE_rt_outl	RETRO_B	RETRO_F	RETRO_B.Bmsy	RETRO_F.Fmsy	ProcB_CI	HX_MASE	DIC
PSH_13_P_rH_psil0.9_sigF_60	FAIL	PASS	PASS	FAIL	PASS	0.11	-0.06	0.02	-0.30	FAIL	1.76	-456.30
PSH_16_P_rM_psil0.9_sigF_60	PASS	PASS	PASS	PASS	PASS	0.29	-0.21	0.16	-0.41	FAIL	1.49	-466.80
PSH_82_P_rM_psil0.2_sigF_60	PASS	PASS	PASS	PASS	PASS	0.00	-0.00	0.01	-0.17	FAIL	1.49	-472.80

Figure 15: Diagnostic tests for top performing JABBA models, where Run 82 was the "best model" that passed the most diagnostic tests.

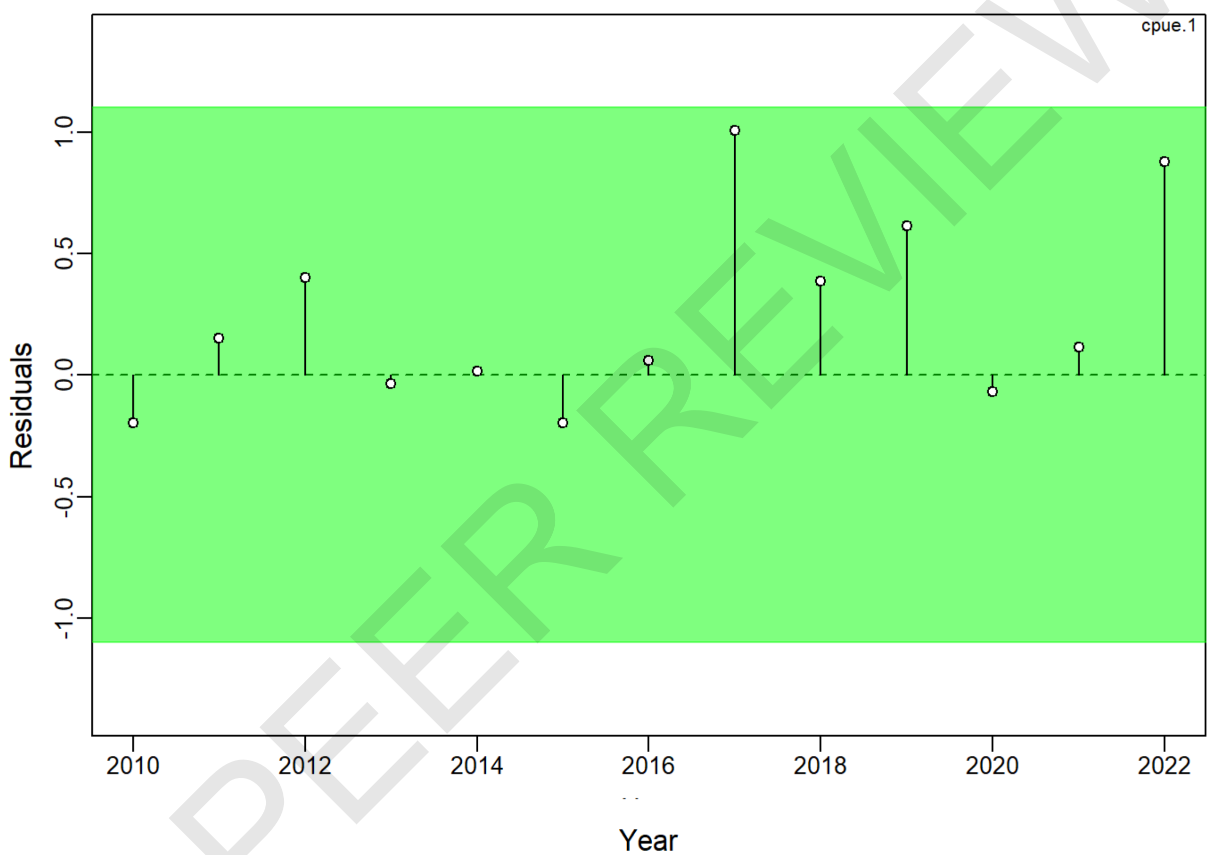


Figure 16: Residual runs test for top performing JABBA model. Green shading indicates no evidence (p = 0.05) and red shading evidence (p < 0.05) to reject the hypothesis of a randomly distributed time-series of residuals, respectively. The shaded (green/red) area spans three residual standard deviations to either side from zero. Here, all points are within the 'three-sigma limit' for that series.

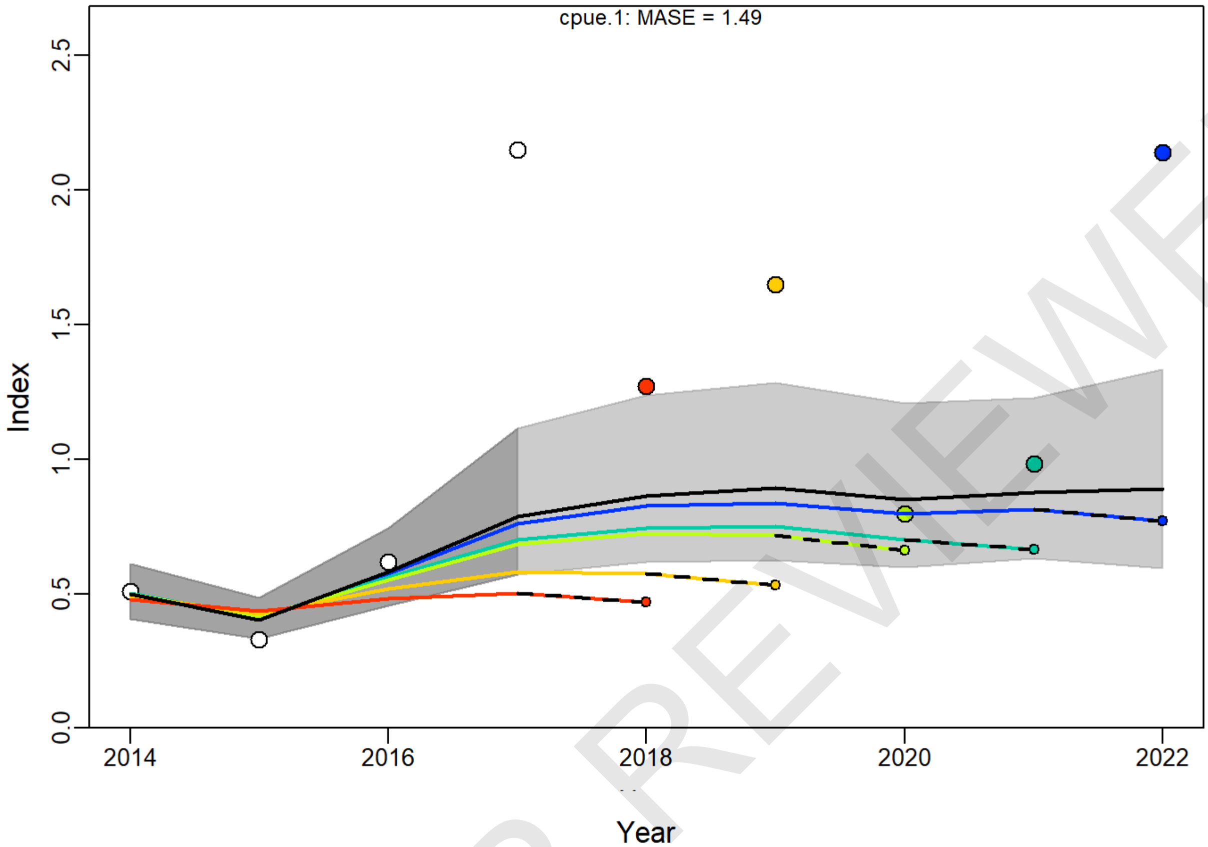


Figure 17: Hindcasting cross-validation (HCxval) results from CPUE fits, showing observed (large points), fitted (solid lines) and one-year ahead forecast values (small terminal points). HCxval was performed using one reference model (black line) and five hindcast model runs (colored lines with terminal years 2018 to 2022) relative to the expected CPUE. The mean absolute scaled error (MASE) score scales the mean absolute error (MAE) of forecasts (i.e., prediction residuals) to MAE of a naïve in-sample prediction (CPUE value this year = CPUE value from last year).

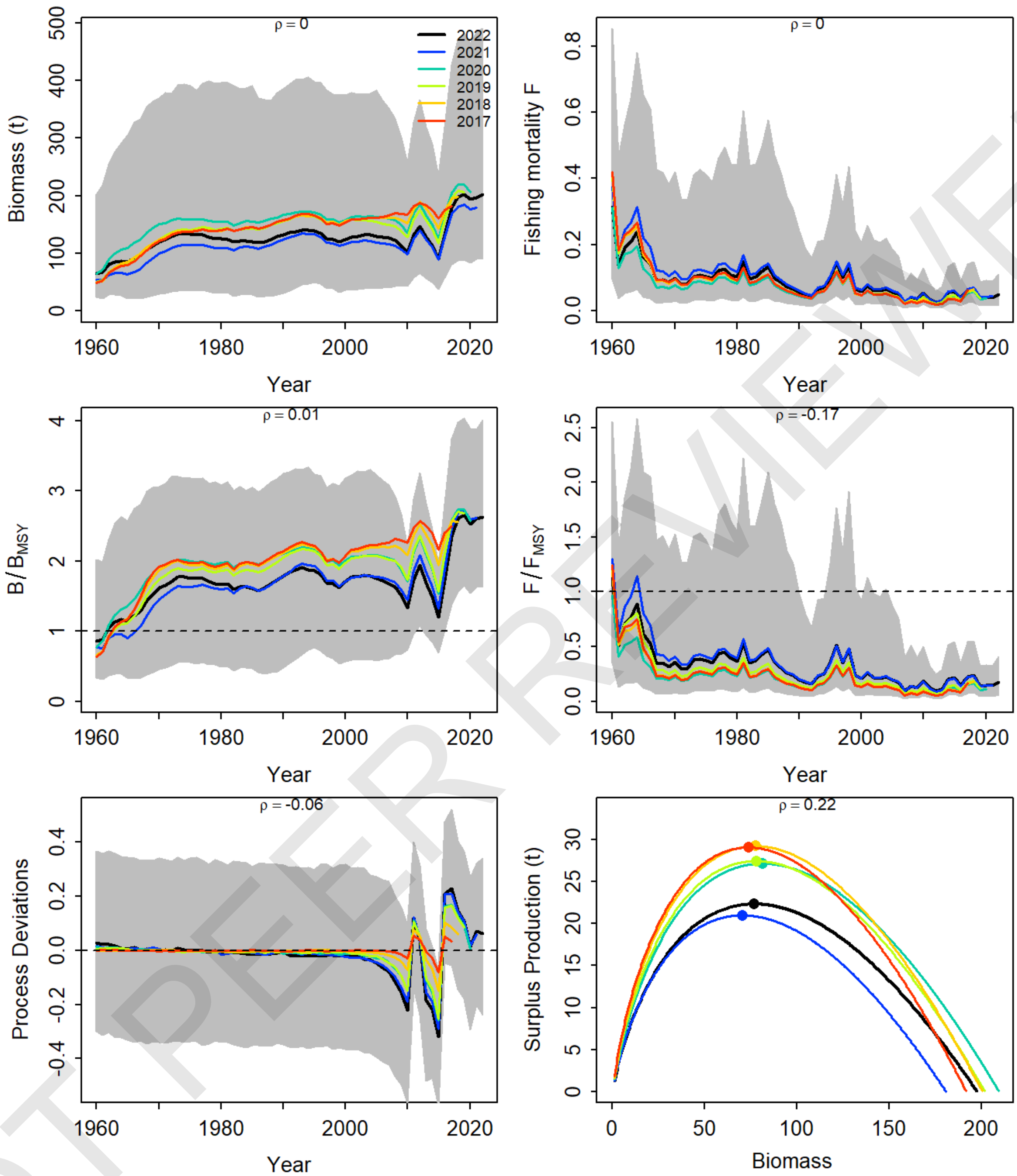


Figure 18: Retrospective analysis of key parameters and management quantities for top performing JABBA model run, with the line color corresponding to terminal years of data ranging from 2017:2022. Mohn's rho statistic (ρ) are denoted on top of the panels. Grey shaded areas are the 95% credible intervals from the reference model. Biomass and surplus production are reported in million lb tail weight.

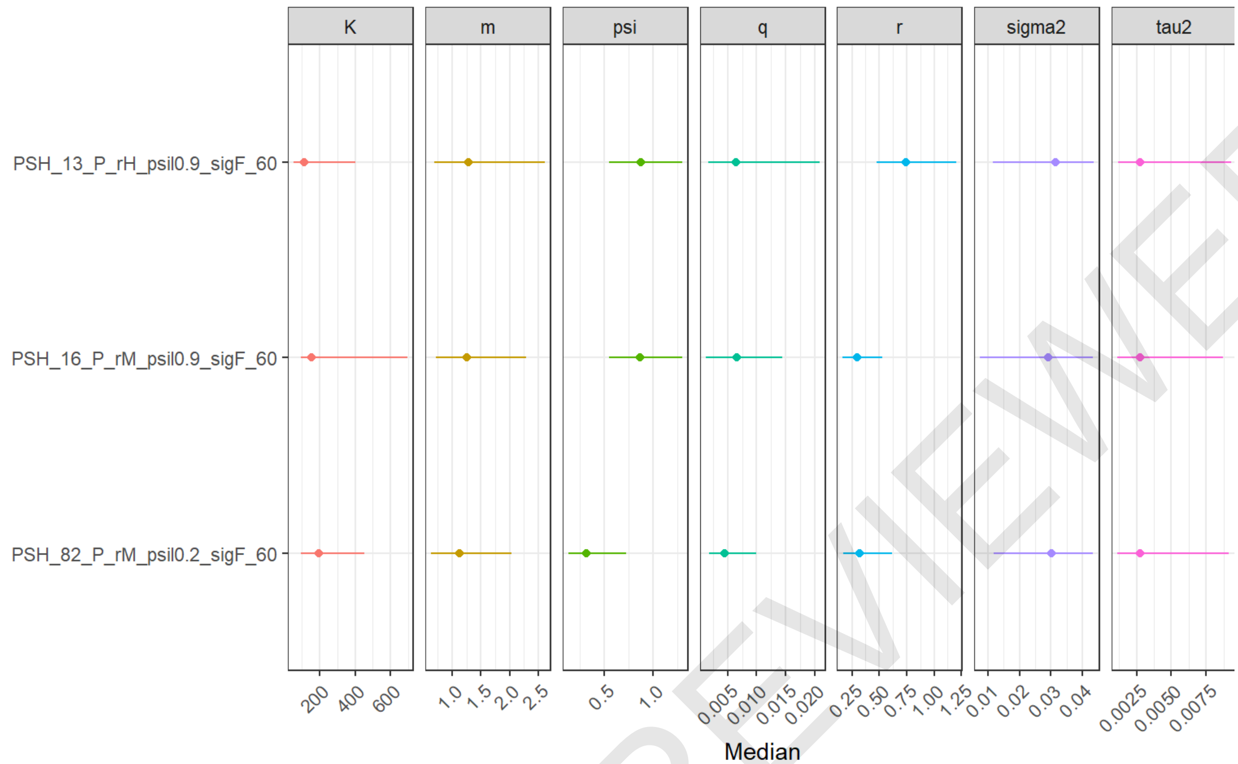


Figure 19: Parameter estimates and error for top performing JABBA models.

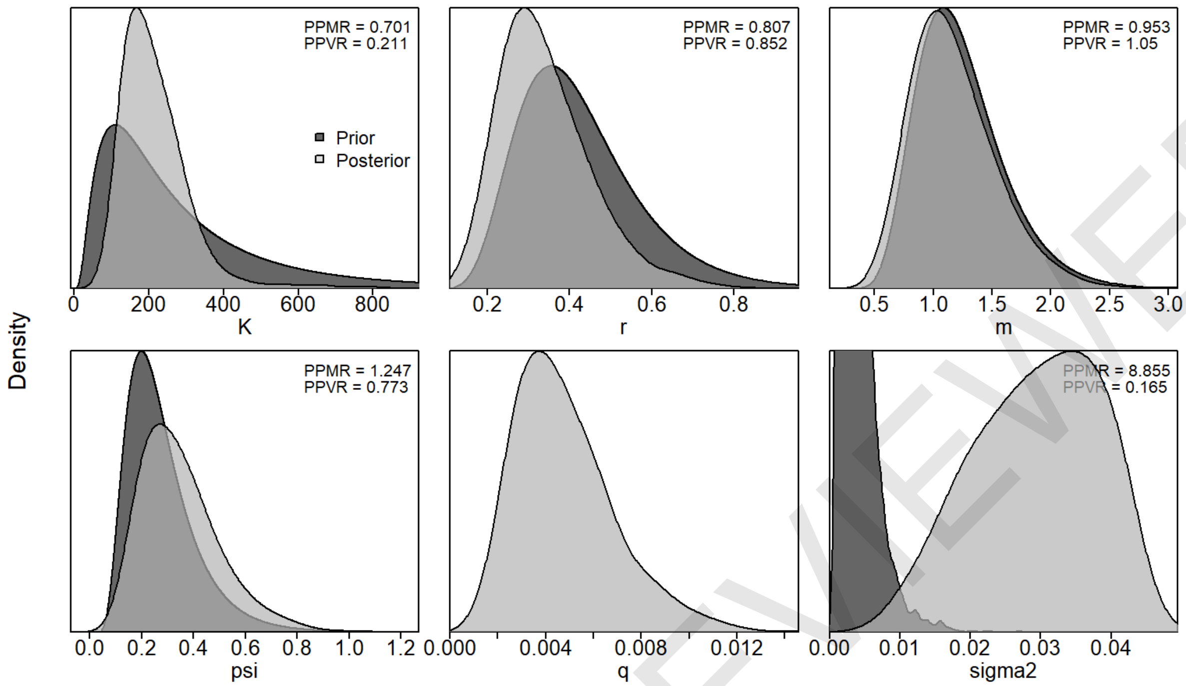


Figure 20: Posteriors for top JABBA model (did not pass diagnostic tests).

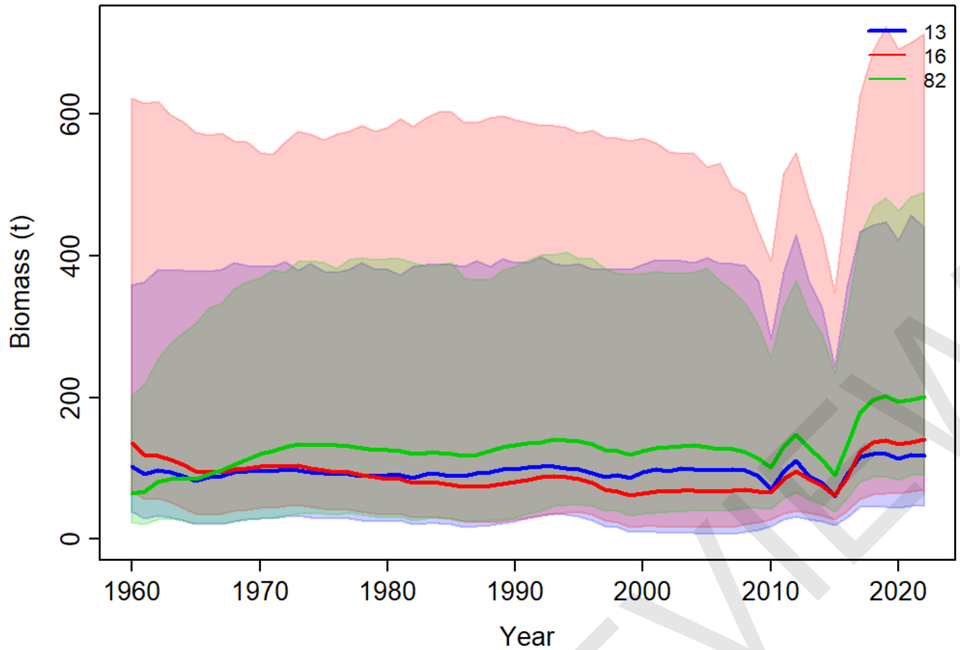


Figure 21: Biomass trajectories (in million lb tail weight) for top performing JABBA models, where Run 82 (green) was the "best model" but did not pass diagnostic tests.

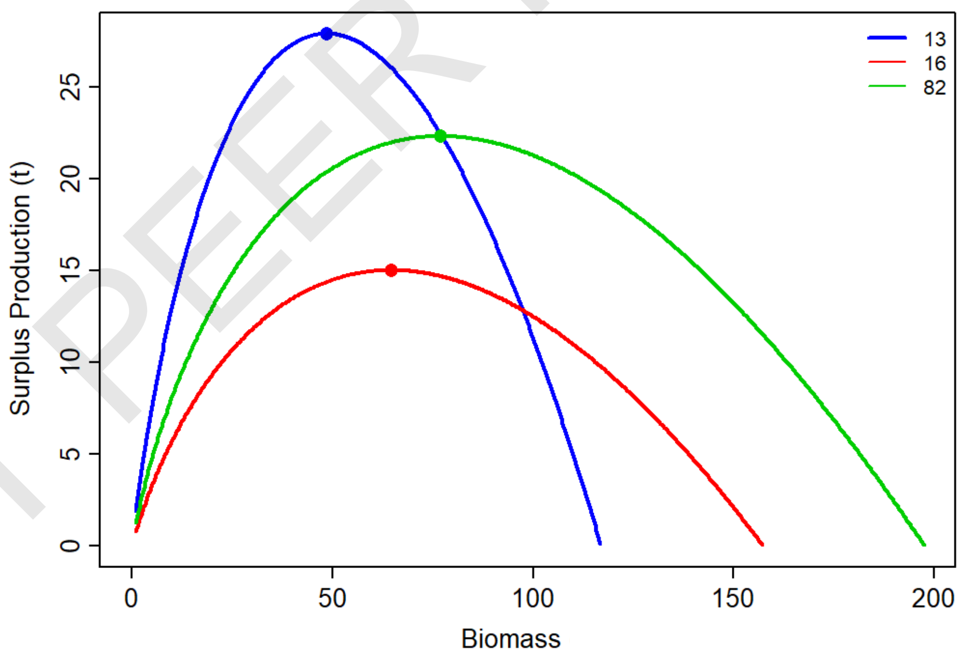


Figure 22: Surplus production and associated biomass estimated for all top performing models (in million lb tail weight), where Run 82 (green) was the "best model" but did not pass diagnostic tests.

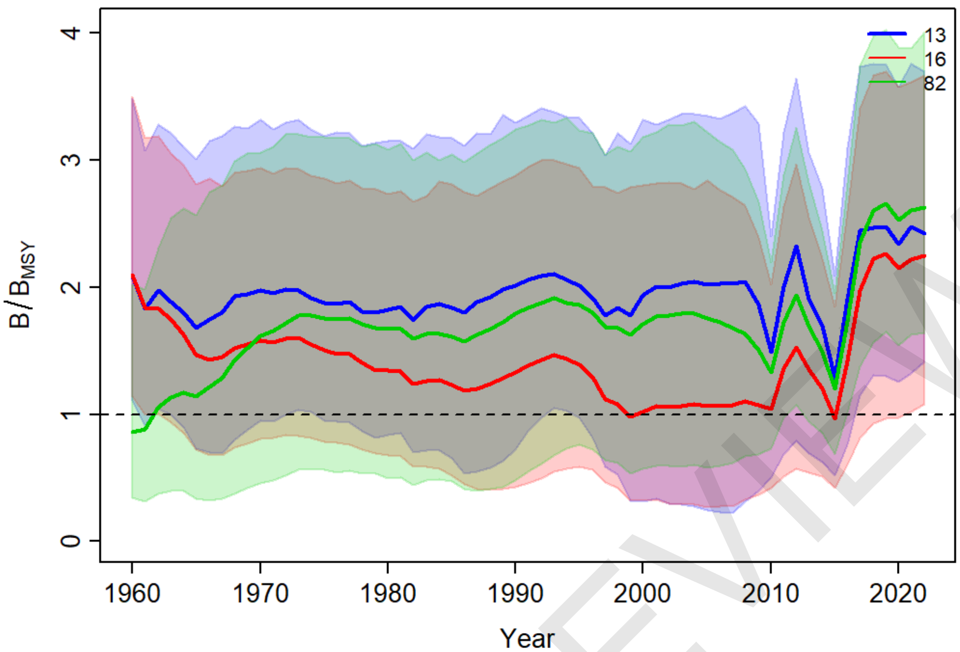


Figure 23: B/Bmsy trajectories for top performing JABBA models, where Run 82 (green) was the "best model" but did not pass diagnostic tests.

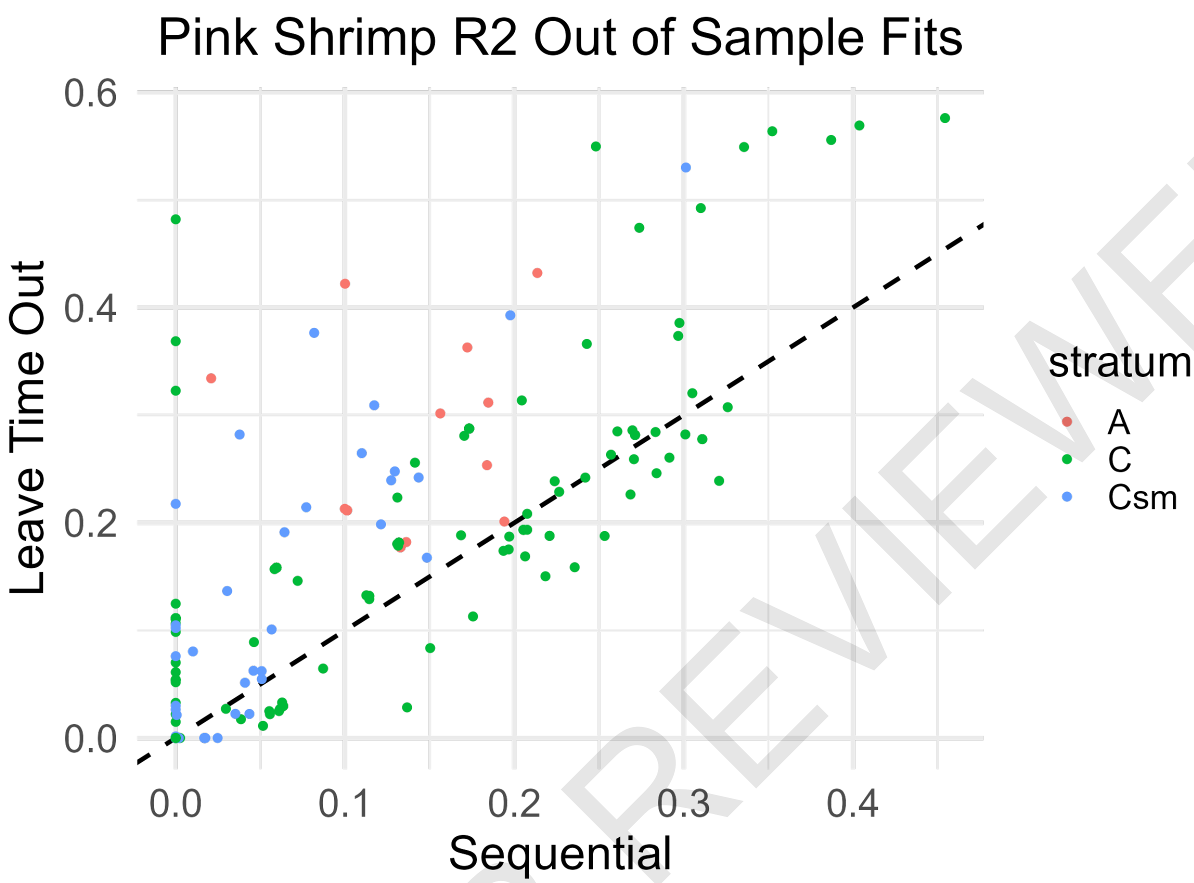


Figure 24: Out-of-sample R2 statistics for each model configuration using the 'leave time out' vs. the 'sequential' cross validation approach. While 'leave-time-out' obtains better model fits, the purpose here is to be able to project well into the future, which is better captured by the 'sequential' approach. Models are filtered based on the R2 statistics from the 'sequential' prediction method going forward.

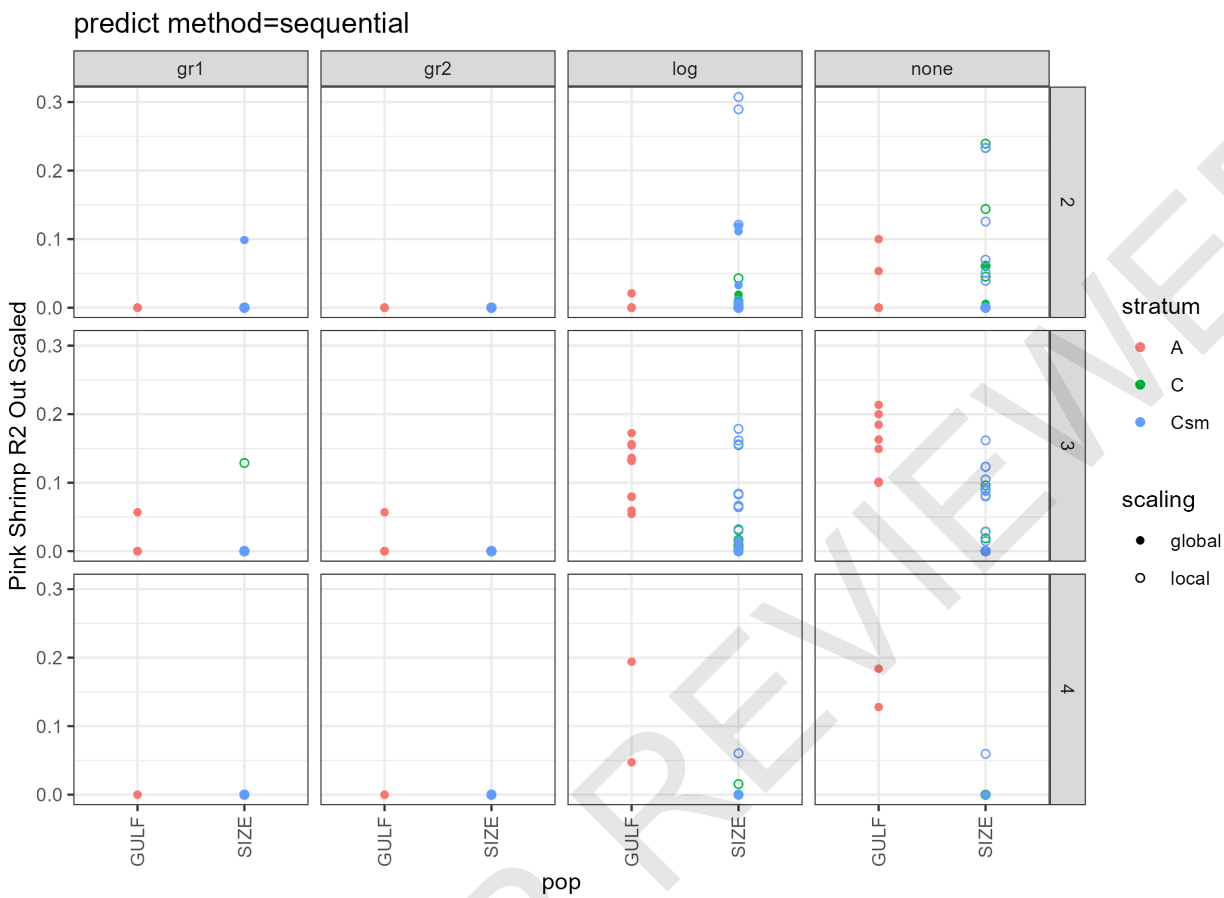


Figure 25: "Sequential" out-of-sample R2 fit statistic resulting from each model run. Facet columns show results based on different data transformations. Facet rows show results based on the embedding dimension. Within each facet, the x axis groups the models by the type of aggregation (spatial, size, season, or a combination). Models that locally scaled the data performed better than models that scaled globally for most transformations, which aligns with how we understand the data (e.g. populations are not directly comparable as defined here).

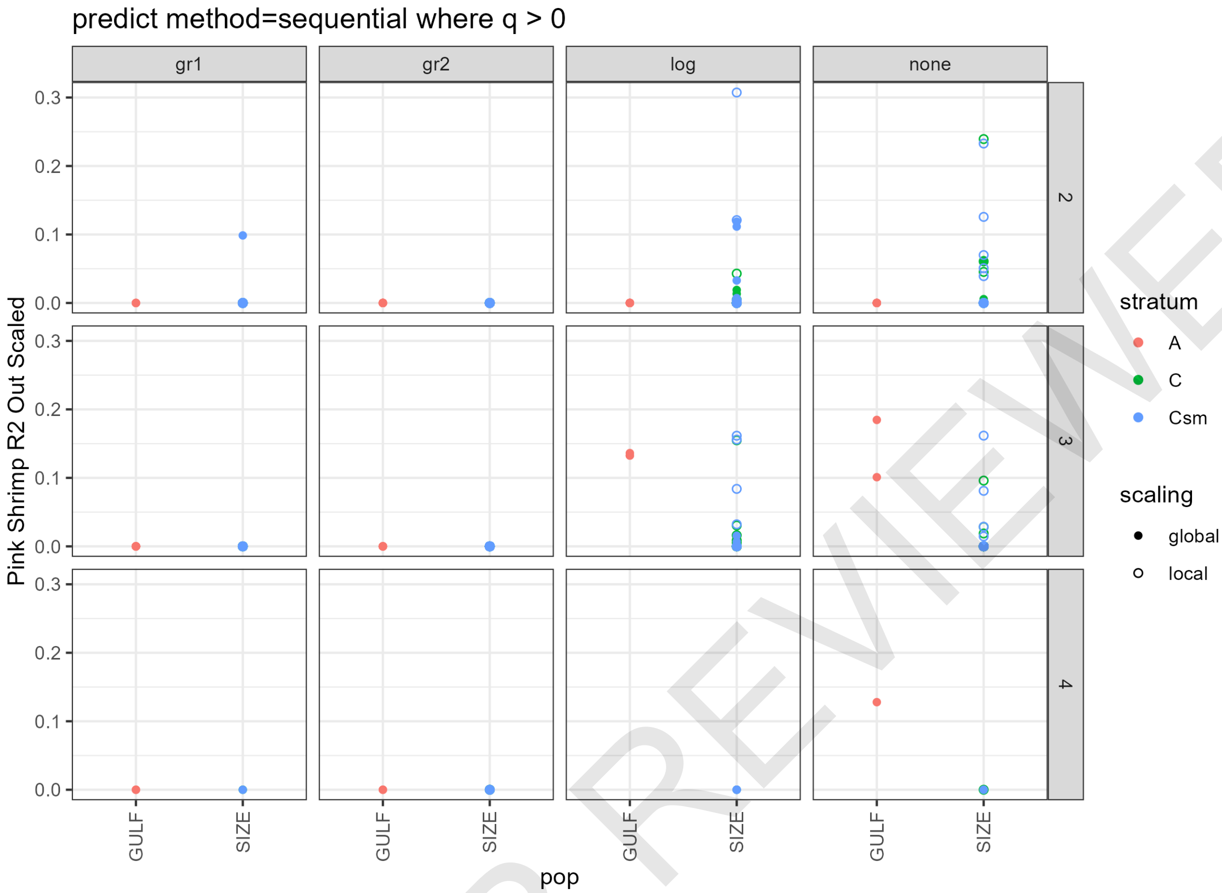


Figure 26: "Sequential" out-of-sample R2 fit statistic resulting from each model run with "local" scaling. Facet columns show results based on different data transformations. Facet rows show results based on the embedding dimension. Within each facet, the x axis groups the models by the type of aggregation (spatial, size, season, or a combination). Models that fit to the survey data and ignored landings (e.g. q=0) were removed from further consideration.

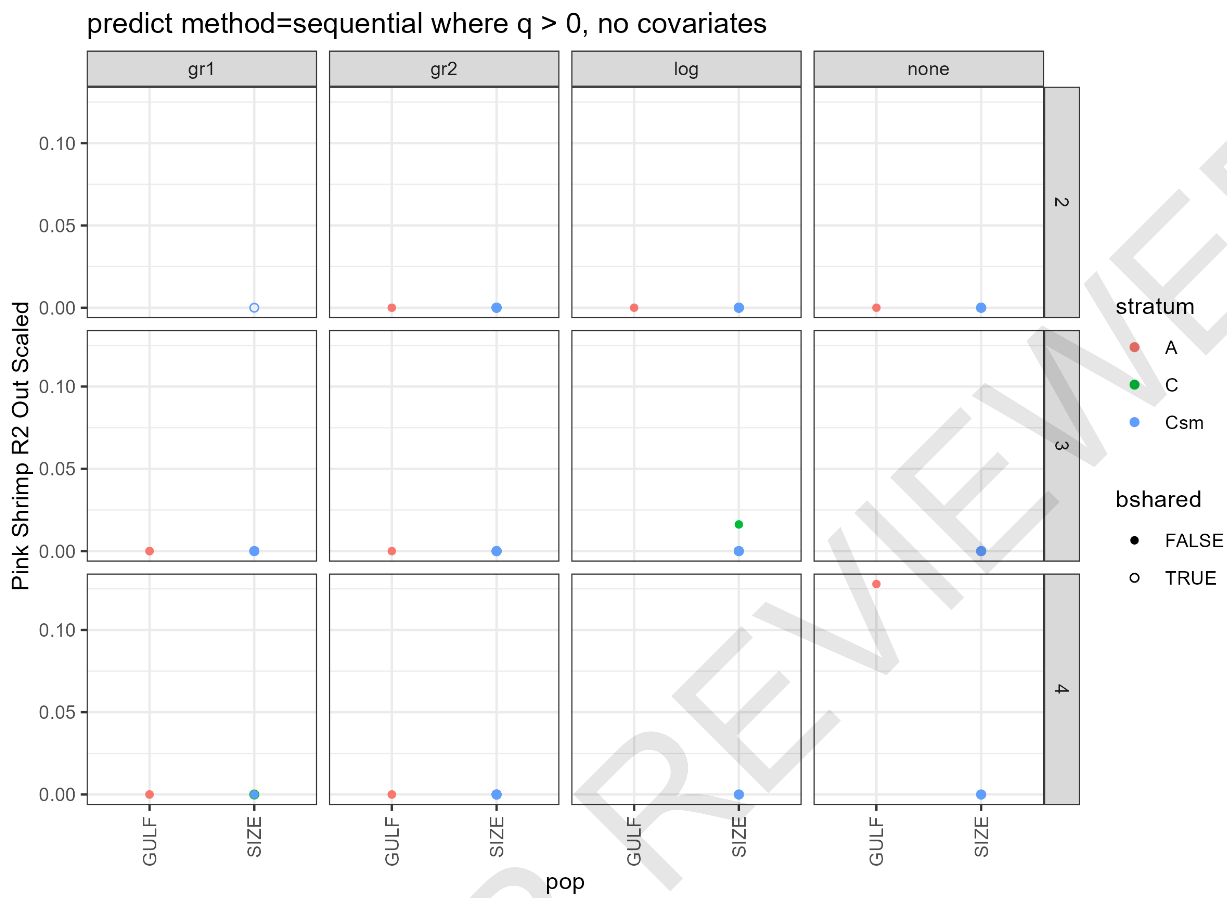


Figure 27: "Sequential" out-of-sample R2 fit statistic resulting from each model run with "local" scaling, q>0.001, and no covariates. Facet columns show results based on different data transformations. Facet rows show results based on the embedding dimension. Within each facet, the x axis groups the models by the type of aggregation (spatial, size, season, or a combination). In this figure, the shape fill was determined by whether or not the catchability parameter was shared among populations in the model (bshared = True / False, respectively). For the shorter time series under Pink Shrimp, removing covariates results in a sharp drop in predictive capability for the model.

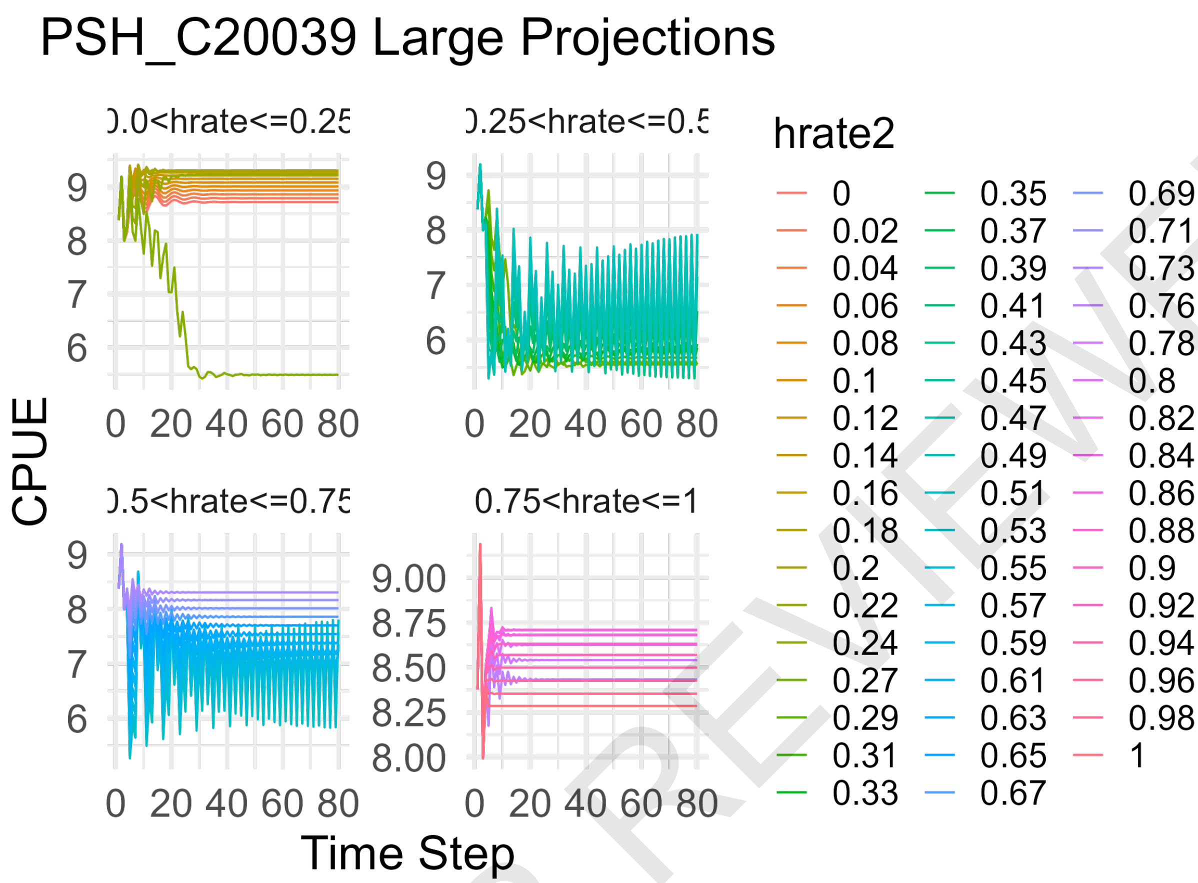


Figure 28: Variable harvest rate projections of CPUE from the best performing run for the Large shrimp population.

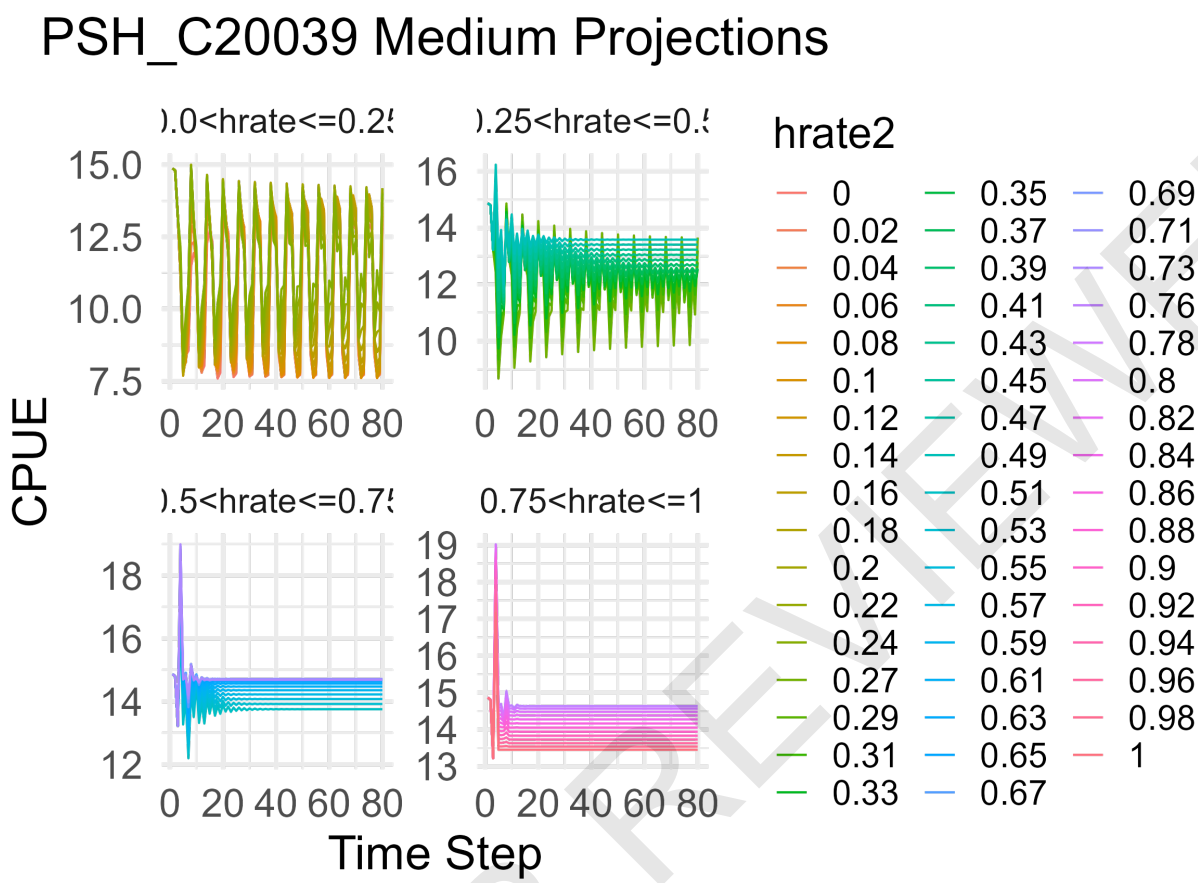


Figure 29: Variable harvest rate projections of CPUE from the best performing run for the Medium shrimp population.

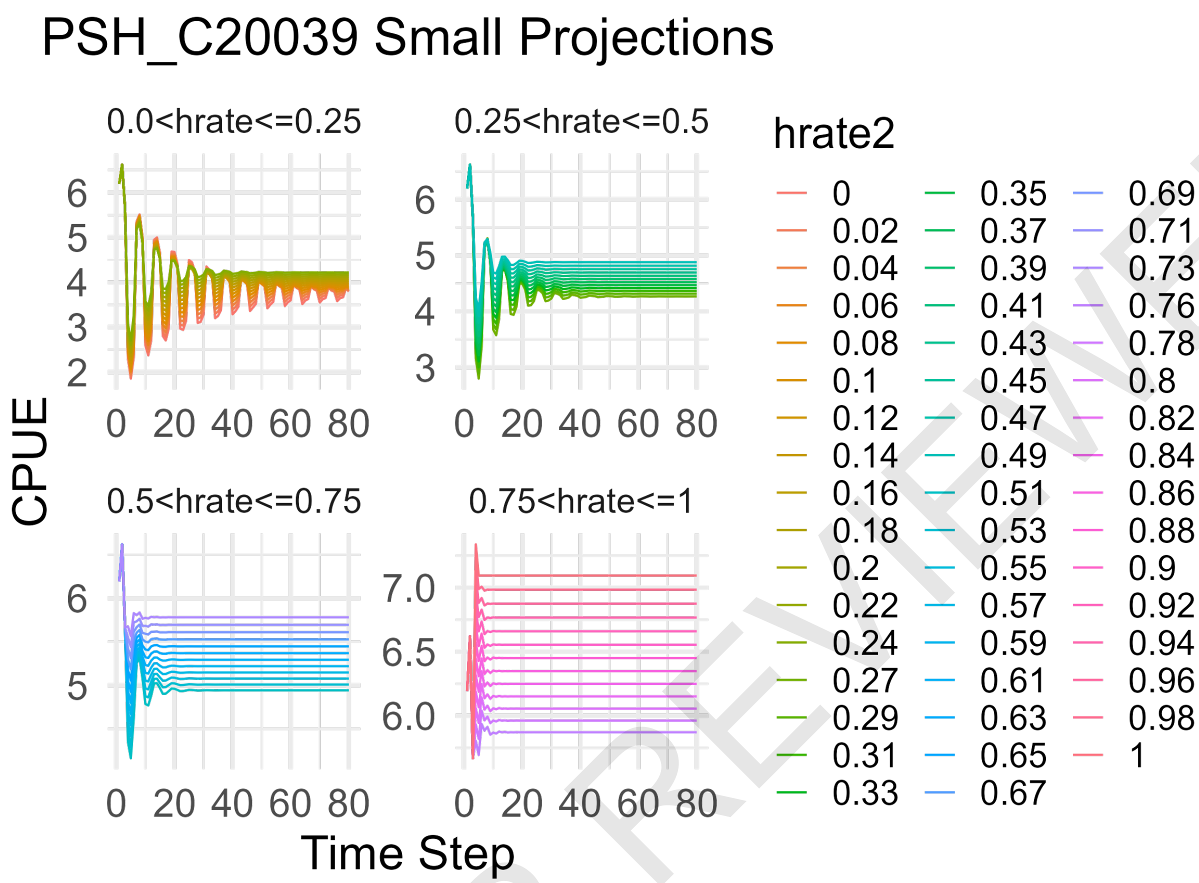


Figure 30: Variable harvest rate projections of CPUE from the best performing run for the Small shrimp population.

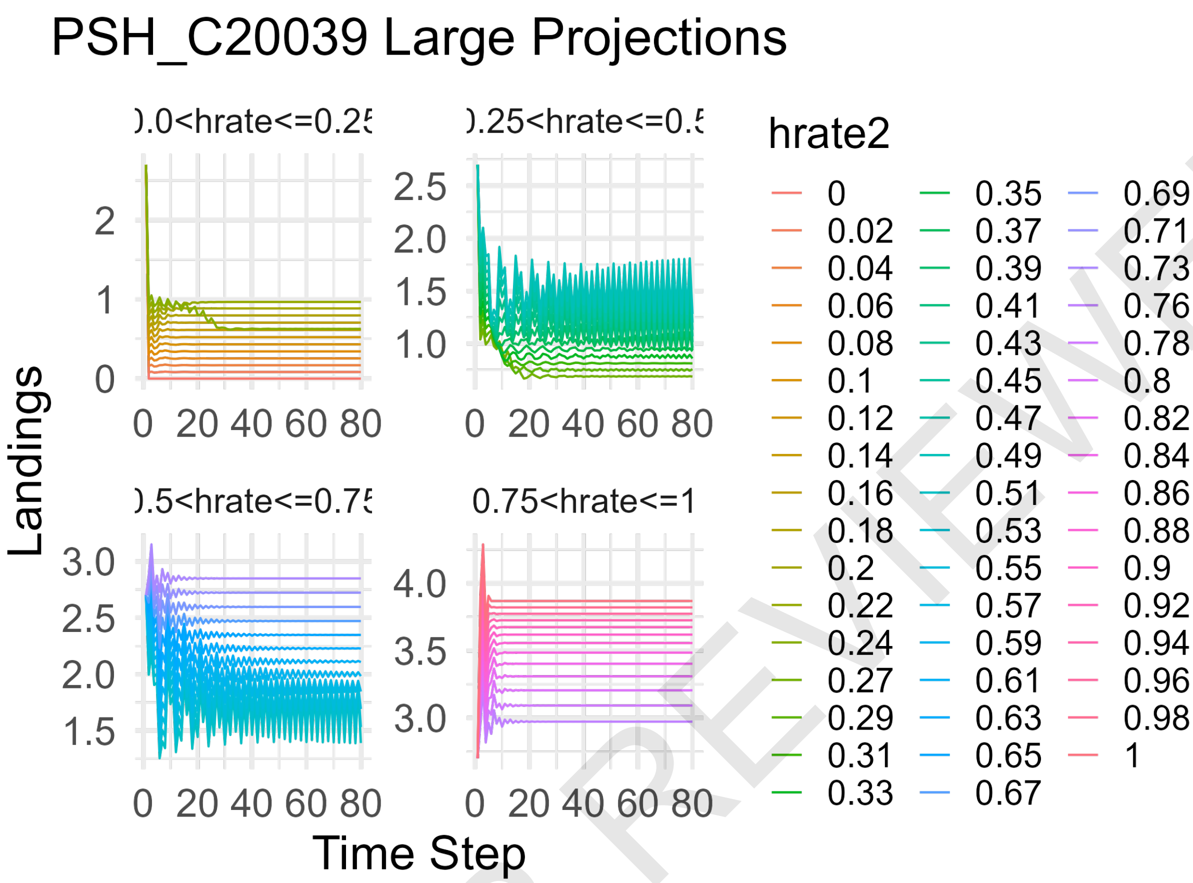


Figure 31: Variable harvest rate projections of landings from the best performing run for the Large shrimp population.

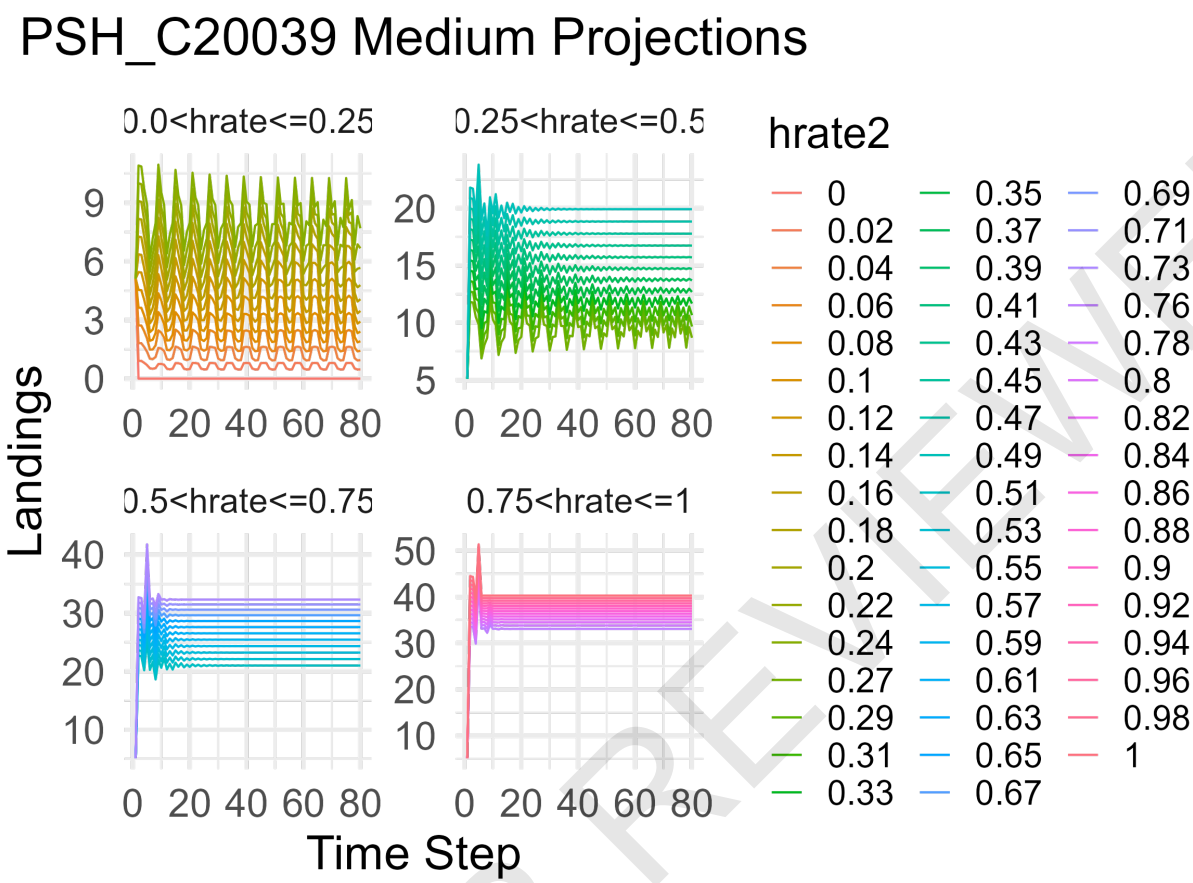


Figure 32: Variable harvest rate projections of landings from the best performing run for the Medium shrimp population.

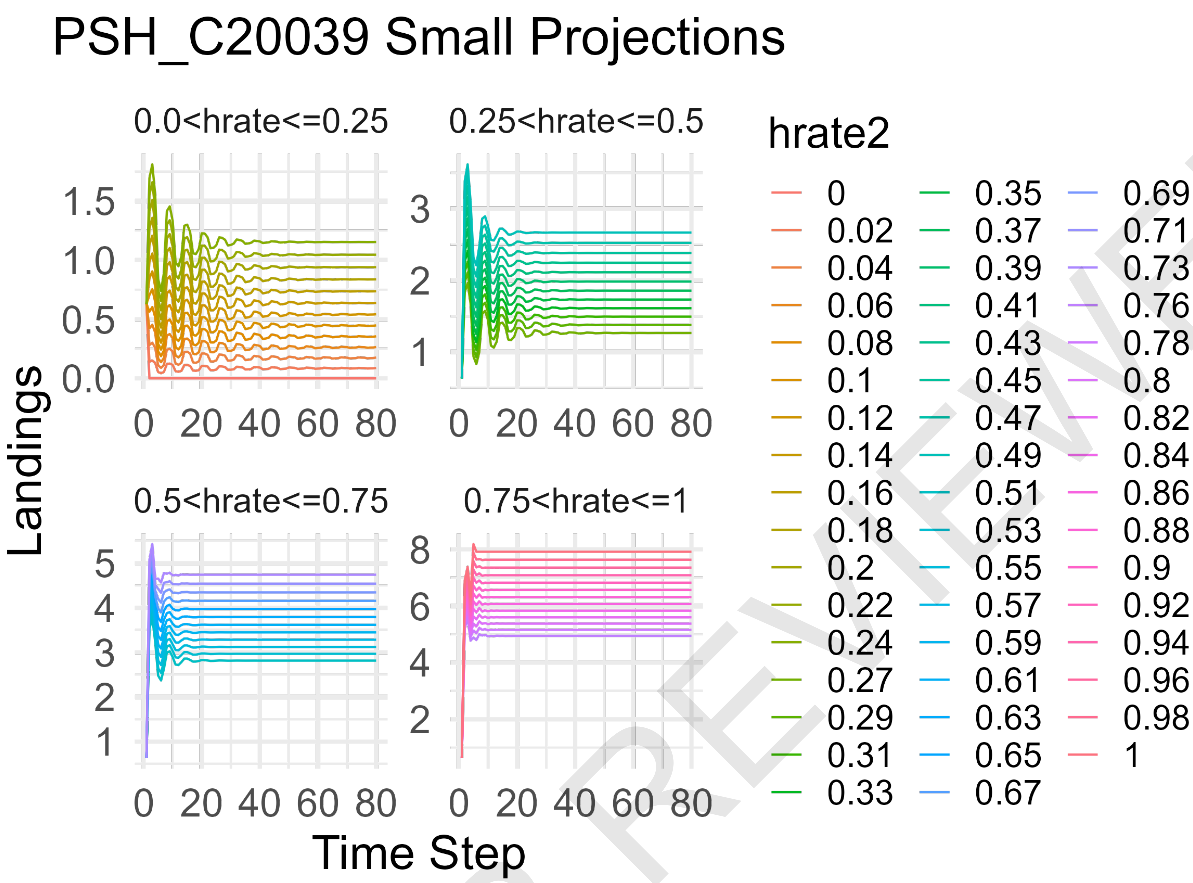


Figure 33: Variable harvest rate projections of landings from the best performing run for the Small shrimp population.

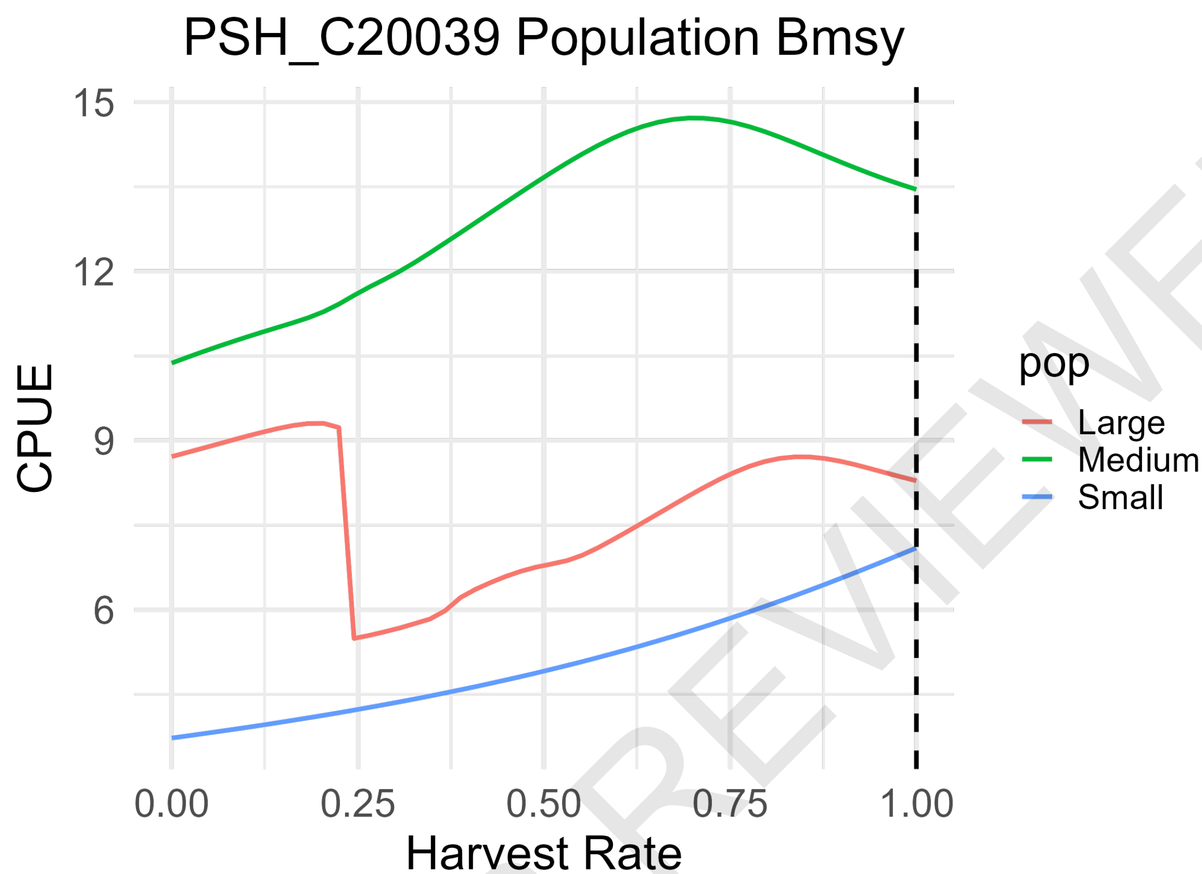


Figure 34: Average CPUE by harvest rate for individual populations for the best performing run. The dashed line indicates the seasonal harvest rate where MSY occurs, indicating population-wide Bmsy in units of CPUE for each population. This model is not recommended.

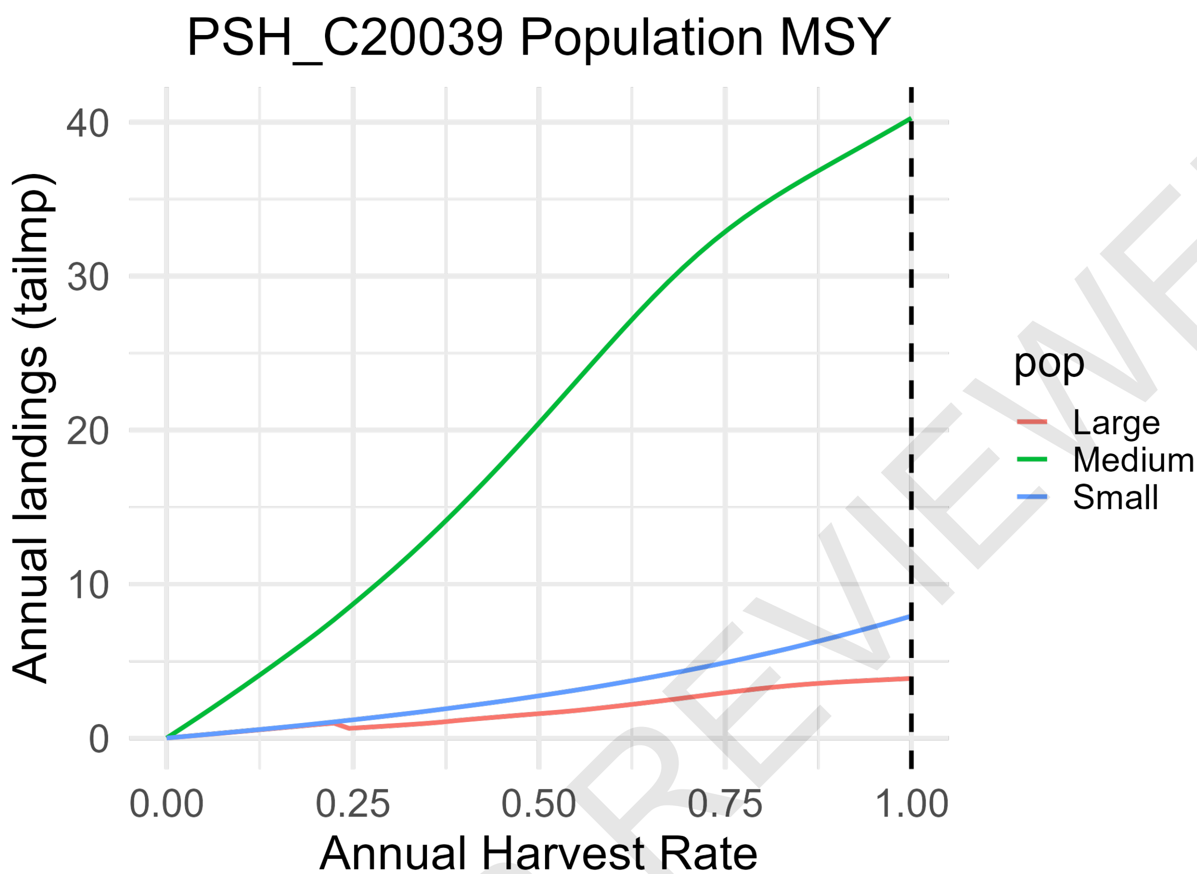


Figure 35: Average landings estimated under a range of annual harvest rates for the best performing run. The optimal annual harvest rate for all populations combined is shown in the dashed line. Individual populations see their landings maximized at slightly different harvest rates. This model is not recommended.

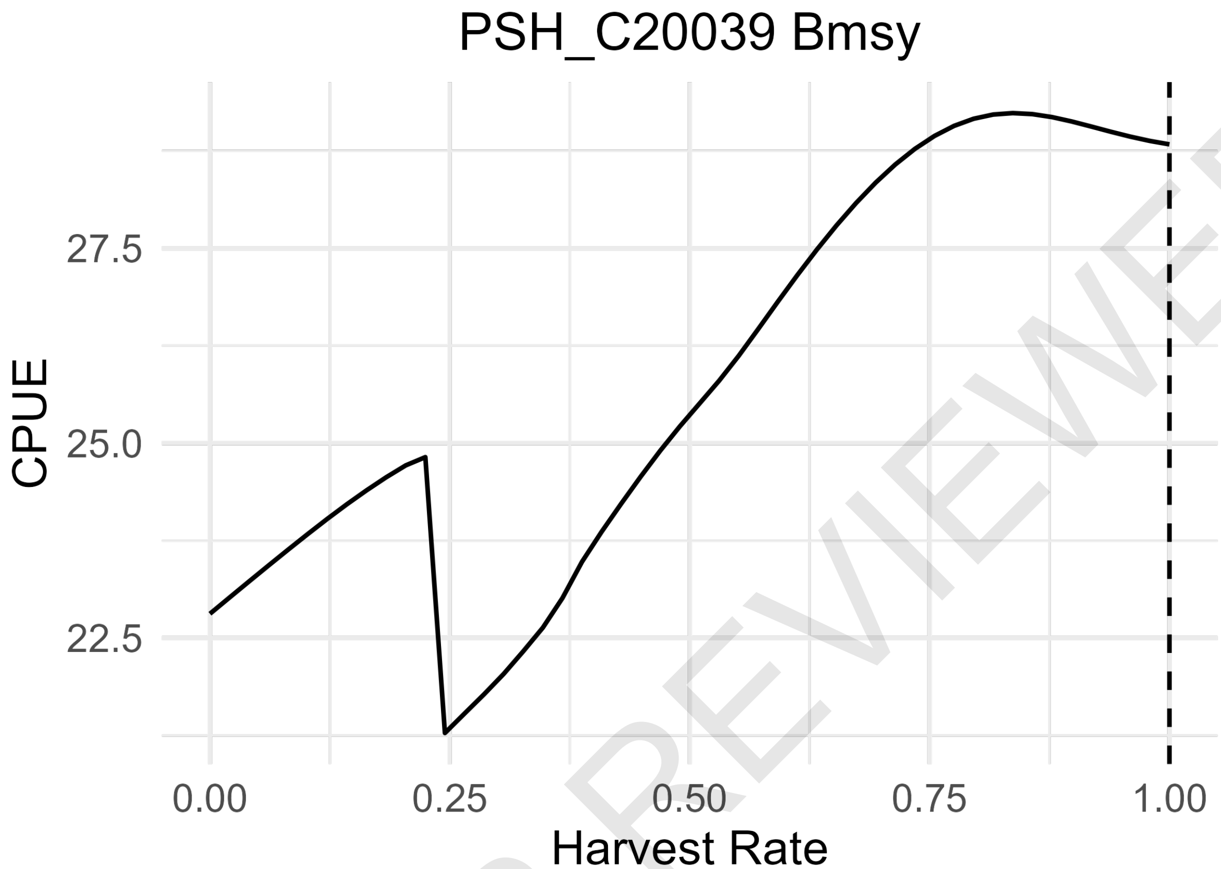


Figure 36: Average CPUE for all populations combined for the best performing run. The dashed line indicates the seasonal harvest rate where MSY occurs, indicating population-wide Bmsy in units of CPUE. This model is not recommended.

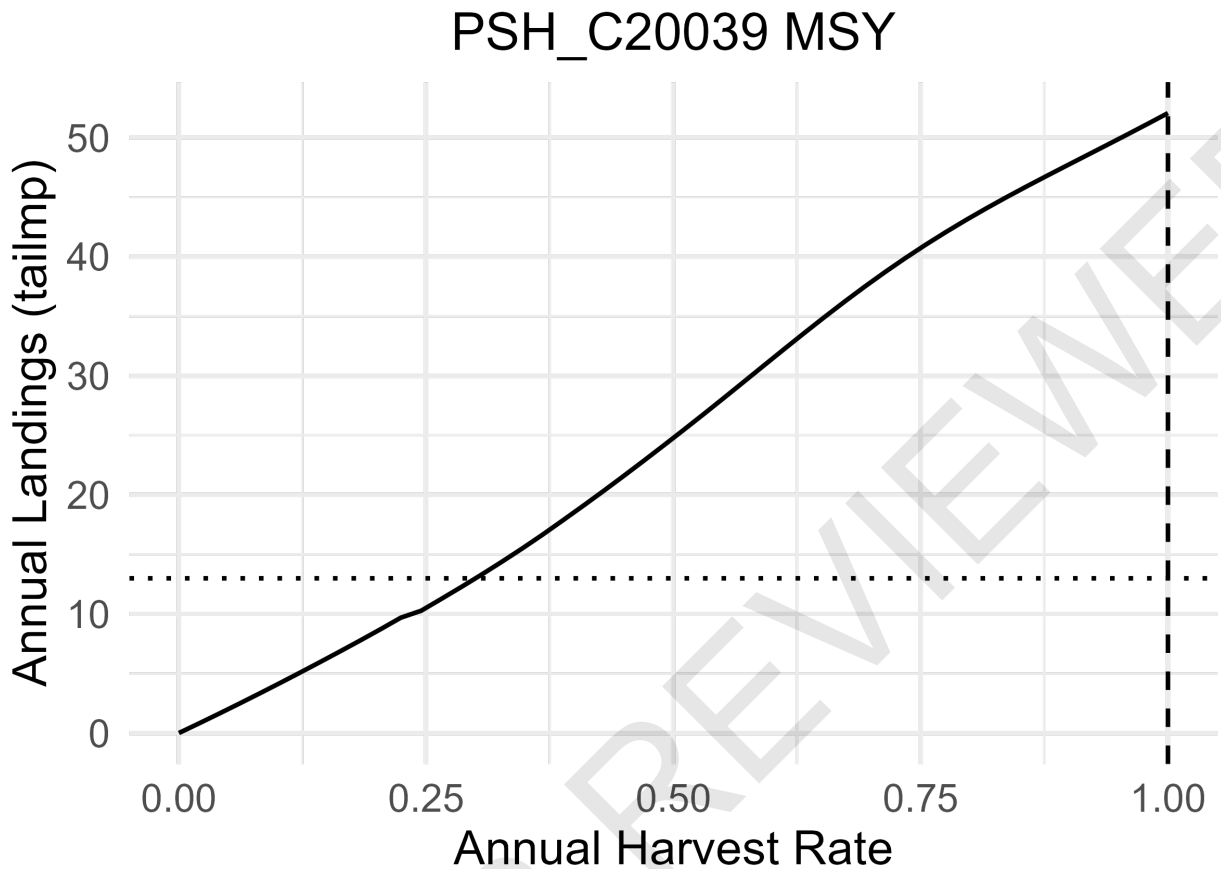


Figure 37: Average landings estimated under a range of annual harvest rates for the best performing run. The optimal annual harvest rate for all populations combined (MSY) is marked with a vertical dashed line. The maximum landings incorporated in the model are marked with a horizontal dotted line. This model is not recommended.

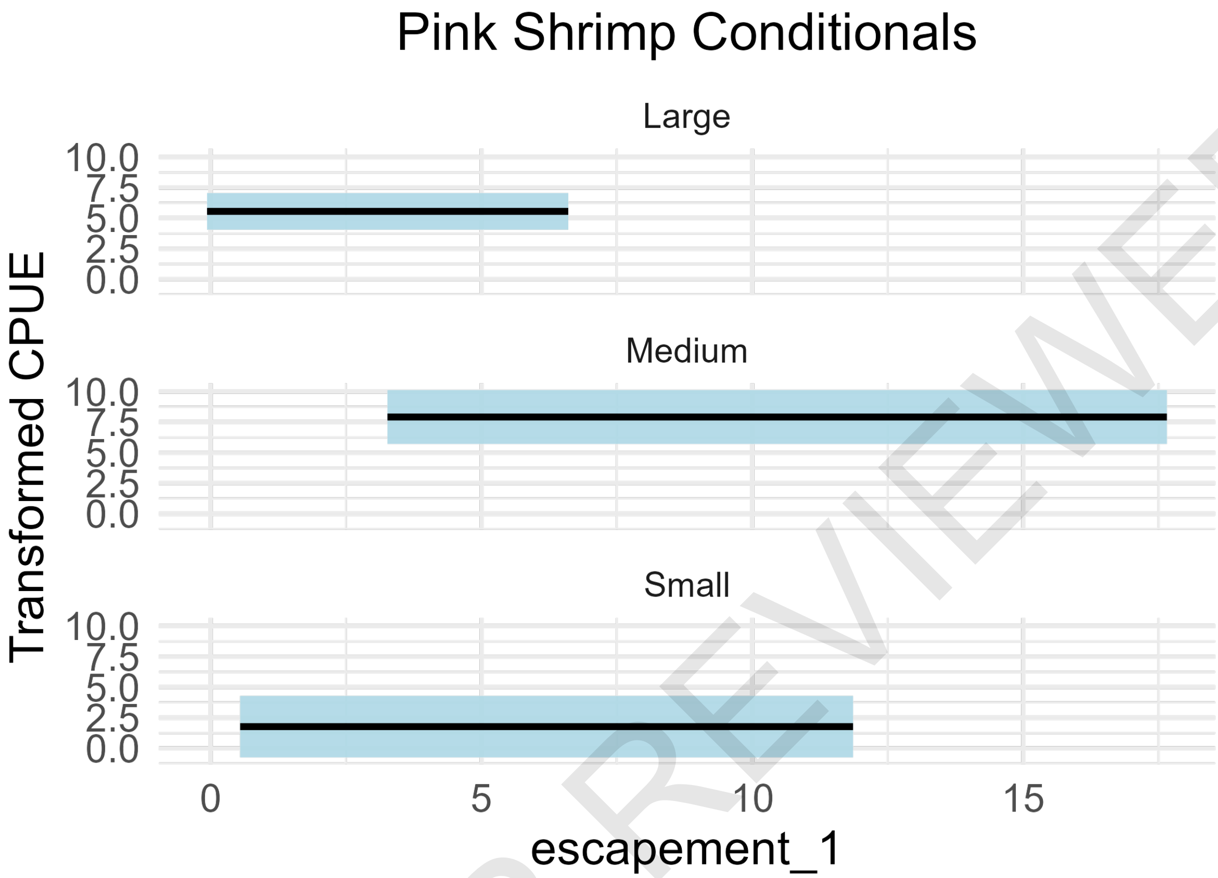


Figure 38: Length scale parameters for the 1st lag of abundance from the best performing model.

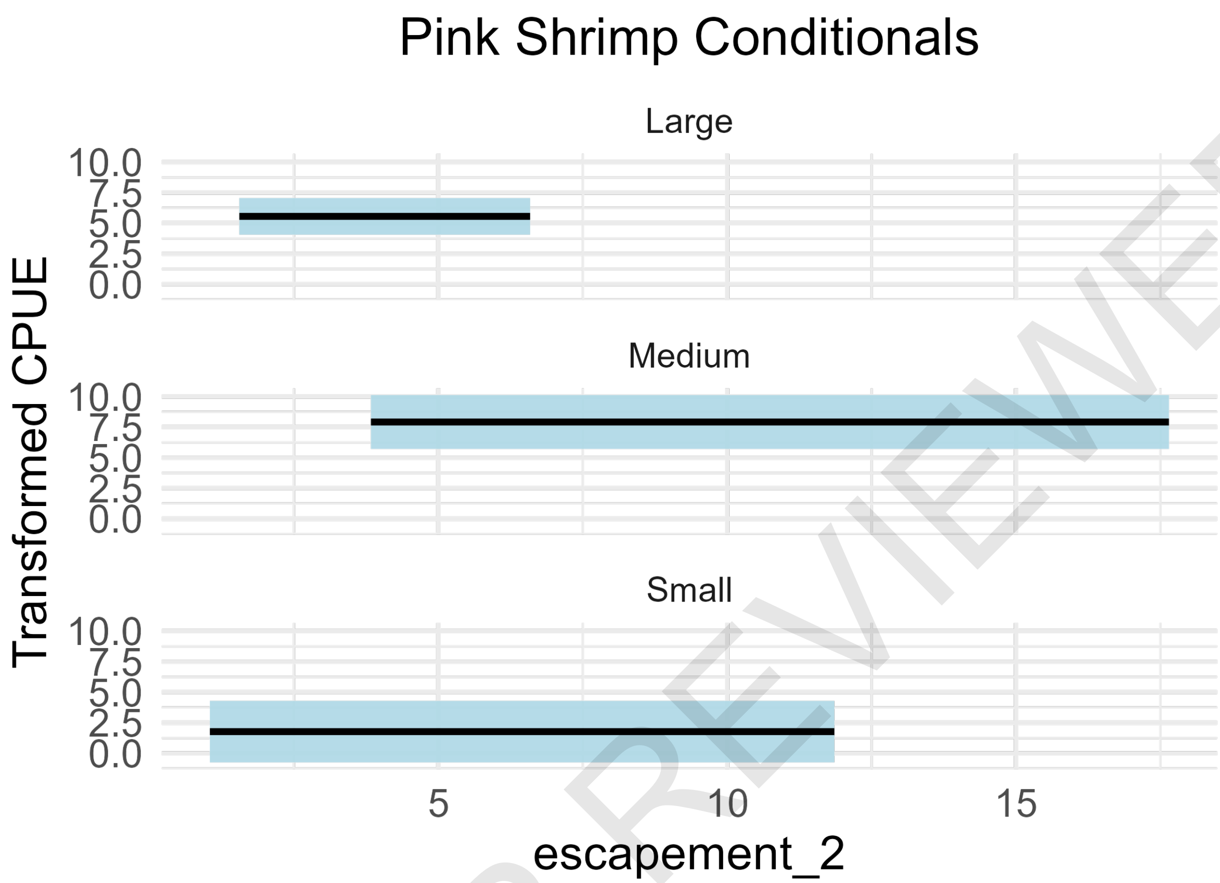


Figure 39: Length scale parameters for the 2nd lag of abundance from the best performing model.

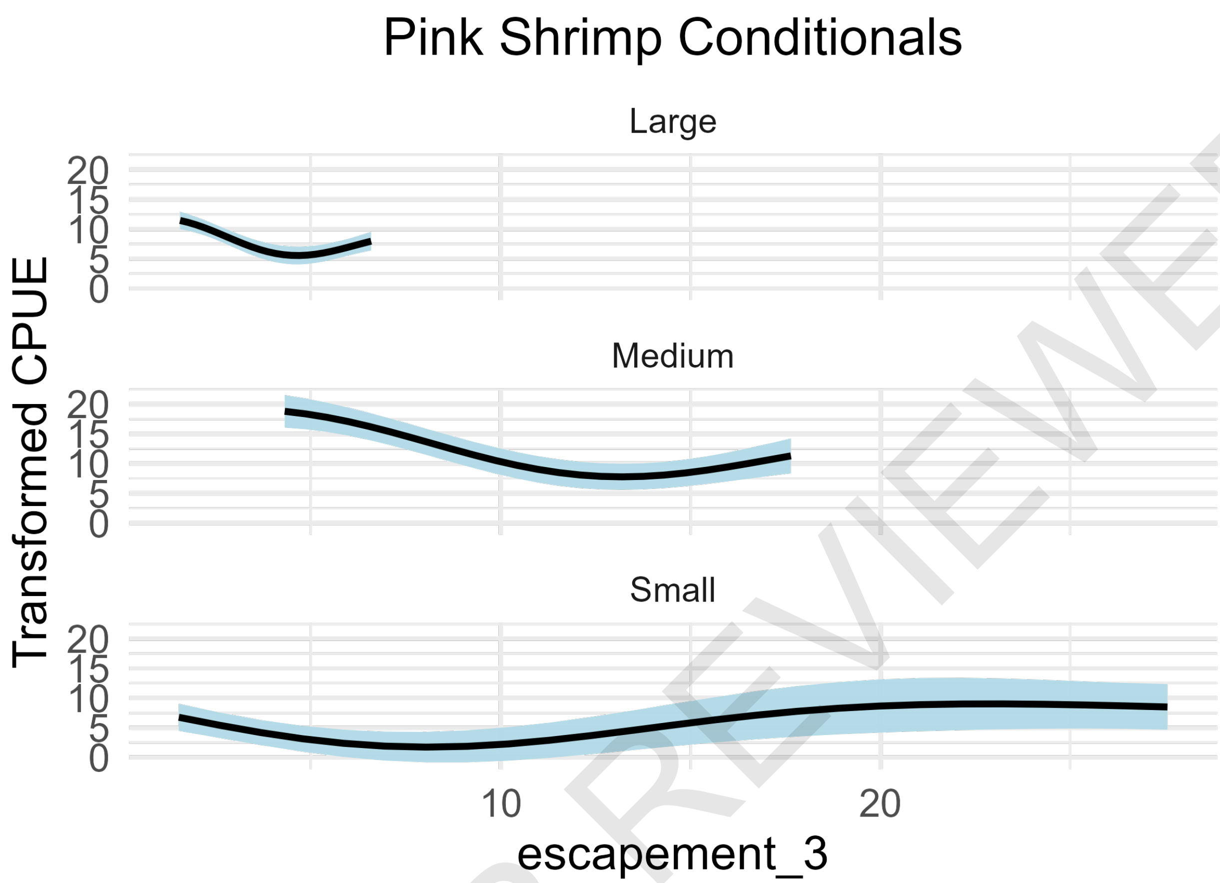


Figure 40: Length scale parameters for the 3rd lag of abundance from the best performing model.

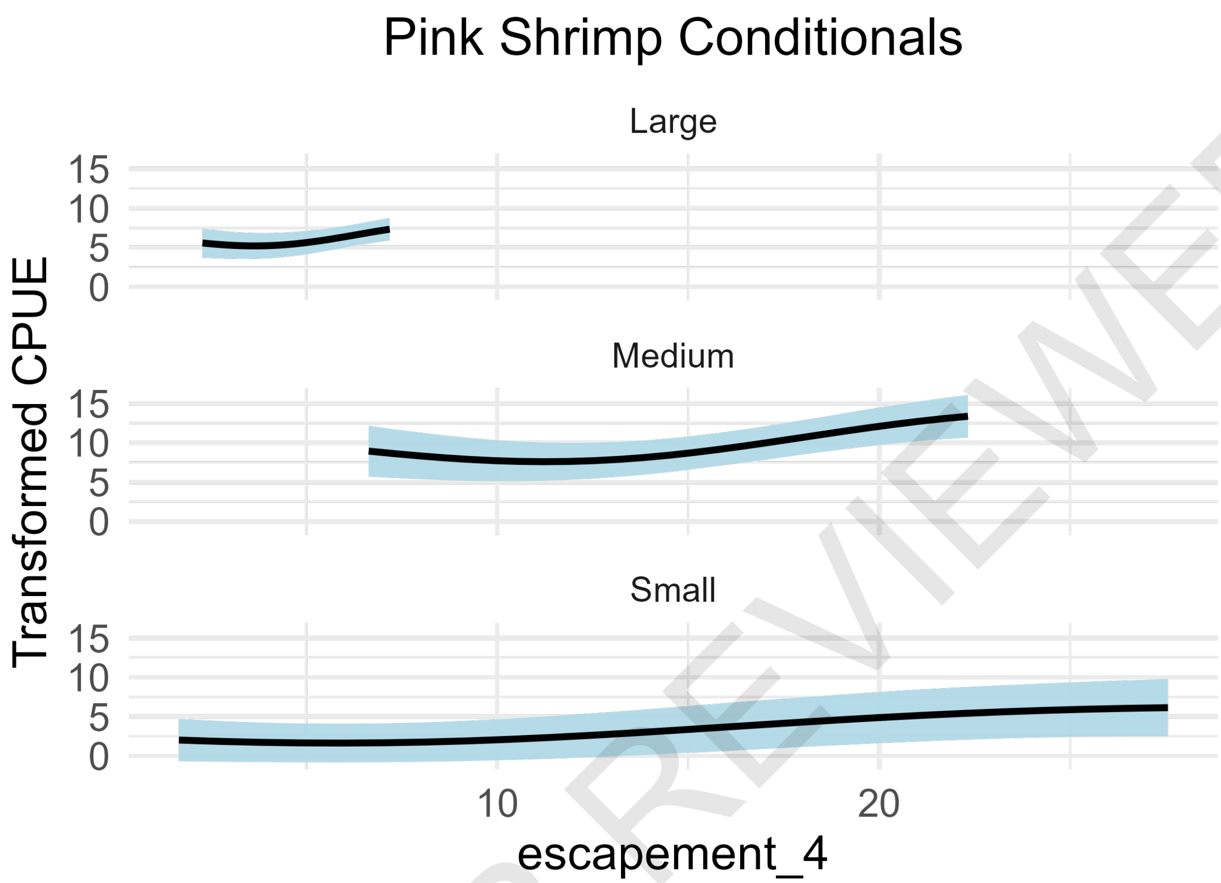


Figure 41: Length scale parameters for the 4th lag of abundance from the best performing model.

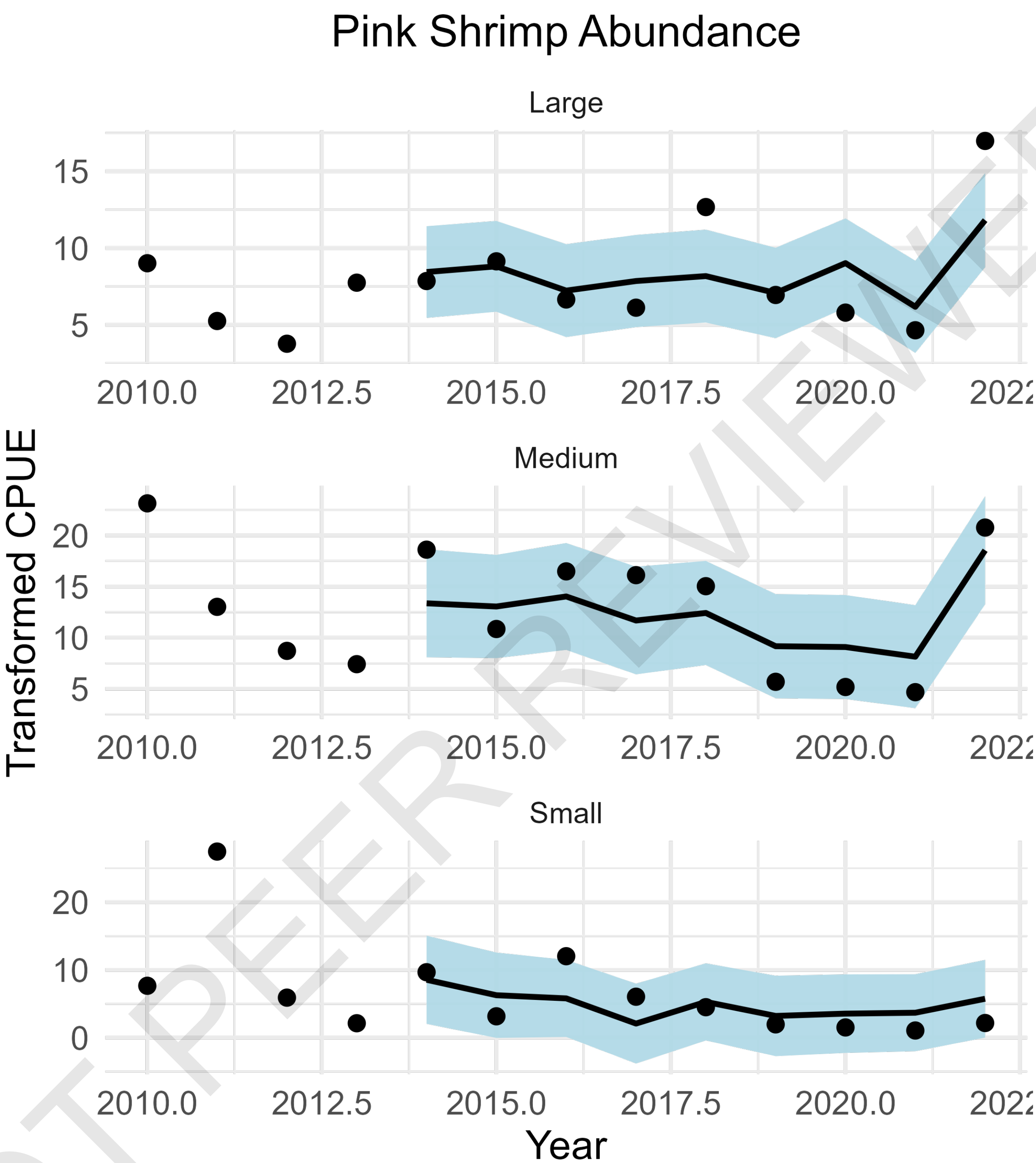


Figure 42: EDM model fits for the best performing run, transformed with error bars.

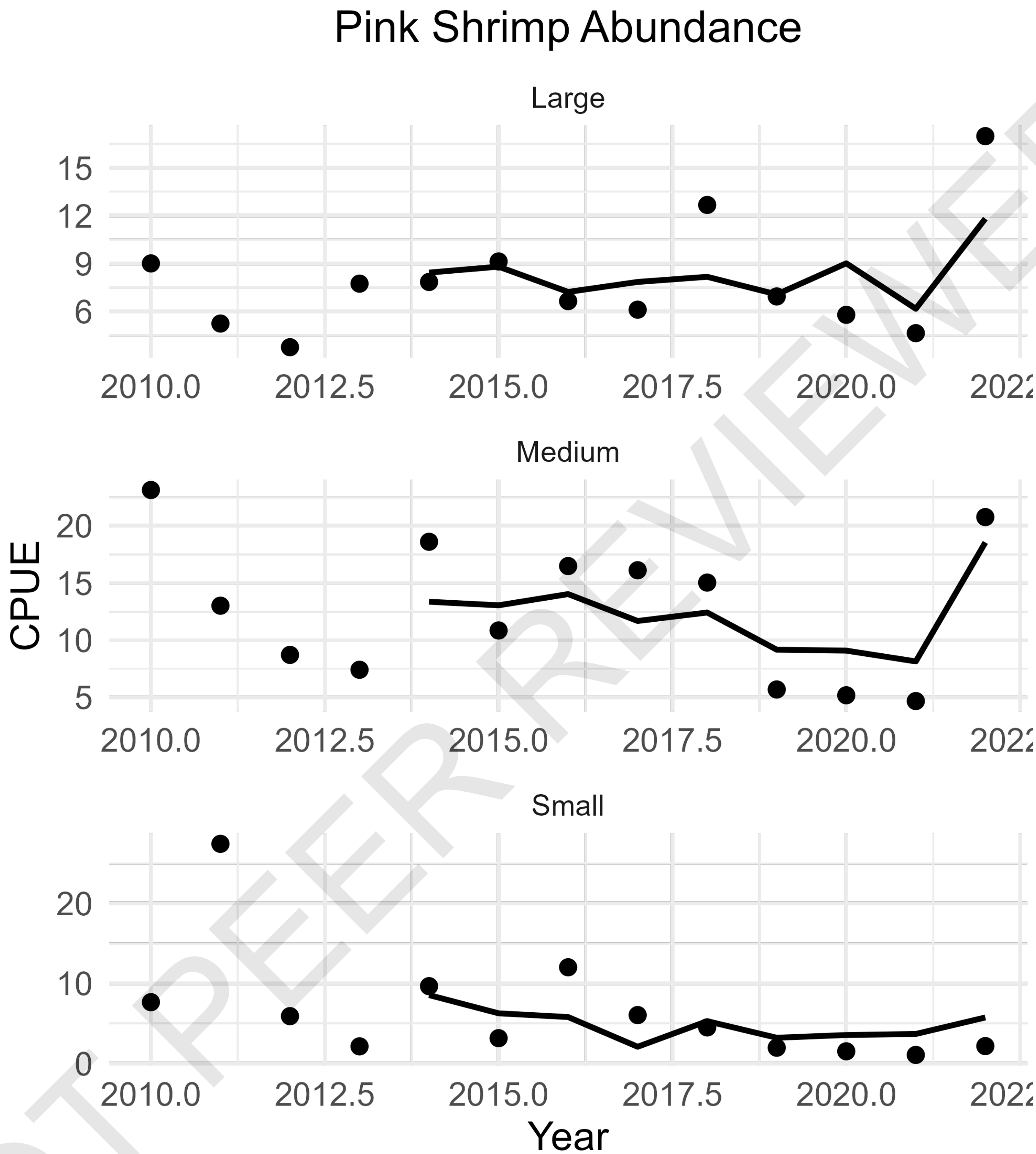


Figure 43: EDM model fits for the best performing run in raw units of SEAMAP CPUE.

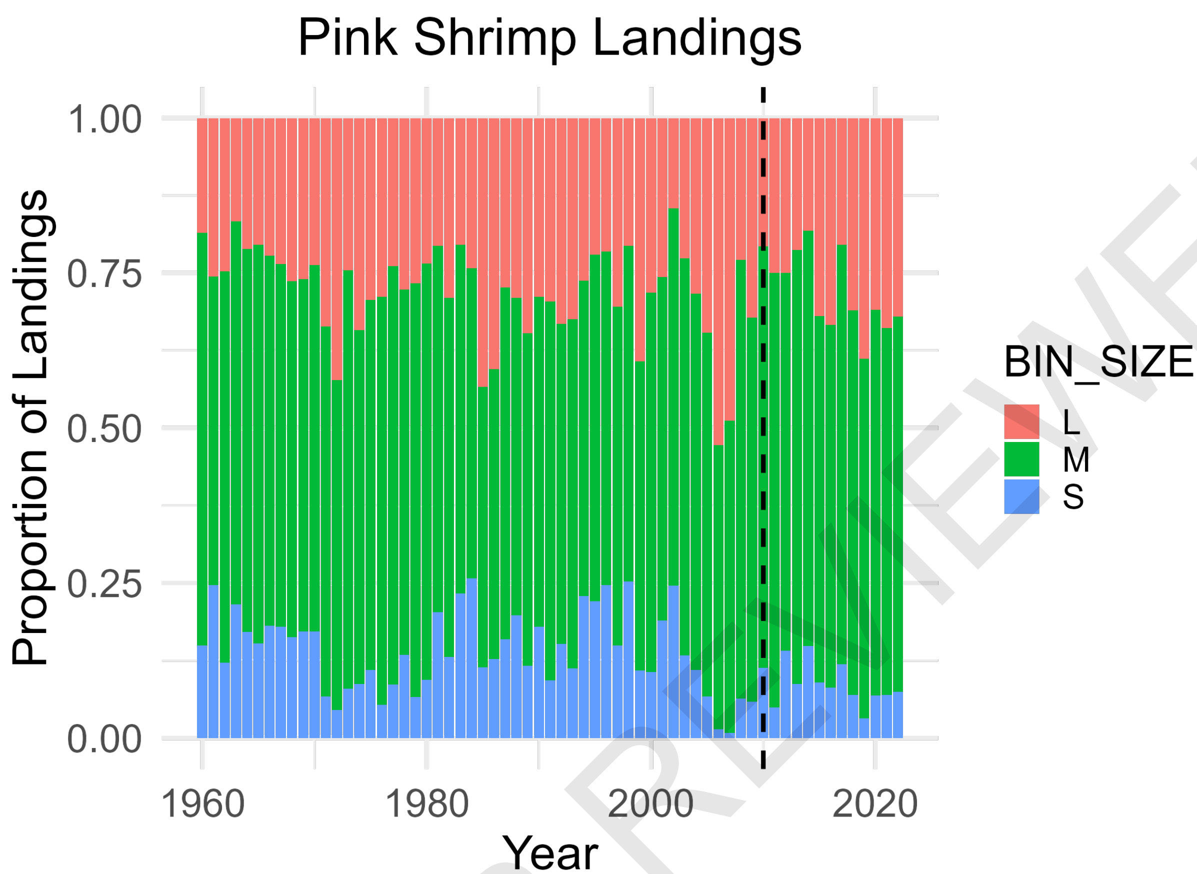


Figure 44: Proportion of landings by size class. The dashed line indicates the first year of the VAST index.