# Cara Rodgveller

## Skipped Spawning (Sablefish focus) and influence on management

Skipped spawning generally appears to be a plastic trait in many marine teleost fish, and is more common than we think. Variation in skipped spawning is likely a result of environmental conditions and has been attributed to poor feeding conditions, given that the probability of spawning is a trade-off between mortality, energetic costs associated with reproduction and migration, and maintenance.

With respect to Alaska sablefish, skipped spawning rates varied drastically (from about 21% to 6% in 2011 and 2015, respectively; Rodgveller *et al.*, 2016; Rodgveller, 2018a, 2018b), suggesting the plasticity of skipped spawning. In these papers, this was attributed to a potential shift from a cool PDO phase to a warmer PDO phase, which has been speculated to be beneficial for sablefish (i.e., when we have seen large recruitment events). Sablefish studied in the west coast are believed to be a separate population from Alaska sablefish; therein they found no evidence of skipped spawning, which might be attributed to genetics or latitudinal variation, but could also be due to the limited geographic scope of the study (Guzmán *et al.*, 2017). Given recent high recruitment events, it might be more likely that the younger large cohorts might be more susceptible to skipped spawning, considering that large recruitment events can lead to density-dependent effects, to which can result in reduced growth and the need to partition more energy towards somatic growth instead of reproduction. Further, considering the life-history of Alaska sablefish (long-lived, low natural mortality), it may be more beneficial to skip-spawn as there may be a higher probability of increasing your *lifetime* reproductive potential, if you decide to forego spawning in one year. This is somewhat evidenced by Rodgveller et al. (2018), where they found skipped-spawning decreases with age.

Unlike fish that undergo mass atresia (rapid reabsorption of vitellogenic oocytes; abortive maturation), sablefish display a “resting” type of skipped spawning. For Alaska sablefish, skipped spawning has primarily been identified using histological methods, where it is generally indicated by a thick ovarian wall, blood vessels in lamellae, and thick tunica walls. However, for some other species, skipped spawning can be identified using electronic tags (e.g., Pacific Halibut display spawning vertical migrations and distinct patterns), physiological measures (-estradiol, levels of water-soluble proteins, absence of spawning migrations at a given size, growth annuli, and isotopic signatures).

In terms of the significance to fisheries management and Alaska sablefish in general, the incorporation of skipped-spawning into maturity curves used for management has demonstrated decreases in spawning stock biomass (SSB) and less optimistic stock status. This is not surprising given that maturity is directly used to calculate reproductive potential. In Rodgveller et al. (2016), they classified skipped-spawning fish as immature, resulting in a shift in the age-at-50% maturity to about 9-years. However, I do not think this is an adequate approach to incorporate skipped-spawning into the calculations of reproductive potential, because we are reclassifying the maturity curve as functional maturity. Instead, I believe a potentially fruitful approach would be as follows:

where *SSB* is the spawning stock biomass, *PSS* is the proportion of skipped spawners, which is modelled as a function of age and potentially other covariates. The rest of the equation represents our standard calculations for *SSB*. I believe this approach might be better, because we partition out the females into two groups: 1) skipped spawners, and 2) spawners. Of the individuals that do not skip-spawning, they are either mature or immature, which then forms the rest of our *SSB* calculations. Additionally, we seldom have a large amount of skipped-spawning data and using this approach allows us to define an average skipped-spawning curve, which may be more representative of the population dynamics, rather than annually adjusting the *functional* maturity curve. Nonetheless, failing to incorporate skipped-spawning will result in more optimistic perceptions of stock status.

## Sablefish Reproductive Biology

Sablefish are demersal species that are typically found beyond the depths of 300m. They exhibit ontogenetic movements, wherein they begin to move offshore as they get larger. In coastal Washington, spawning takes place around December - February (late winter into early spring) and the reproductive cycle is initiated around March/April (Guzmán *et al.*, 2017). By contrast, in the Bering Sea, spawning is thought to take place around February although by some of the size of larvae, it can extend into late May (Mason *et al.*, 1983). Sablefish age-at-50% maturity tends to be around 5-6 years old. They are determinate spawners in that oocytes recruit as large groups for development, and some follicles maintain their perinucleolar stage; this stage happens around March, while for males, the recruitment of spermatozoa happens around April (Guzmán *et al.*, 2017).

Spawning tends to take place at depths of around 300m and deeper, while eggs tend to occupy depths of 400m, where the hydrographic conditions are fairly stable in this region. Sablefish larvae can be advected offshore or inshore in British Columbia, however, most larvae will attempt to exhibit shoreward movements into shallow waters, bays, and coasts (Mason *et al.*, 1983), as they grow, they will exhibit migration patterns back towards their spawning grounds (>300m depths). Sablefish eggs tend to be found around January – February suggesting that peak spawning can occur as early as late January, while their larvae are typically found in April – indicating a gestation period from the egg to larval stage. Fecundity is correlated with length/size where sablefish produce more eggs as they age, however, their relative fecundity is fairly constant (i.e., no exponential or allometric relationships with size or age), although some studies have actually found relative fecundity declines with age (although I attribute that to mostly noise, and the relationships tend to be fairly small; Rodgveller *et al.*, 2016; Rodgveller, 2018a).

Sablefish also exhibit skipped-spawning where they are classified as “resting” skipped-spawners (Rideout and Tomkiewicz, 2011; Rodgveller *et al.*, 2016) and do not recruit oocytes for development (i.e., no vitellogenesis) despite being mature. This is generally identified as having thick ovarian and tunica walls. Although there is some conflicting evidence, I believe that skipped-spawning rates tend to decrease with age. This is generally consistent with life-history theory where the energy allocated towards maintenance, somatic growth, and reproduction should maximize the lifetime reproductive potential. Thus, as you reach these older sizes, the prioritization of energy allocation should be towards reproduction given that there are only small gains in somatic growth as you reach your asymptotic sizes. Decreasing skipped spawning with age was identified in (Rodgveller, 2018a) and increasing skipped spawning with age was identified in (Rodgveller *et al.*, 2016). This was likely attributed to differences in sampling, where the 2018 study sampled slopes (shallow water with young fish) and shelves (deeper waters with old fish) fairly equally, while the 2016 sampled predominately slopes. As most young fish occupy the shelves and the high prevalence of skipped spawning in these regions, young fish are most likely to be the dominant skipped-spawners within a population. Additionally, skipped-spawning rates appeared to be quite variable; such characteristics were not detected in Washington (Guzmán *et al.*, 2017), but were detected at rates of about 6 – 20% in Alaska (Kodiak Island; Rodgveller *et al.*, 2016; Rodgveller, 2018a). This suggests that skipped-spawning is a plastic trait, which has been hypothesized to be driven primarily by environmental and feeding conditions (Rideout and Tomkiewicz, 2011). Lastly, the number of oocytes per gram produced by a young individual (<12 years) compared to an old individual is different (less for older individuals), suggesting that older individuals likely produce eggs that are of better condition (Rodgveller *et al.*, 2016) and the potential for maternal age effects in spawning, despite relative fecundity being age-agnostic.

With respect to sampling methods, maturity is poorly determined using at-sea macroscopic methods and can underestimate the age-at-50% maturity, which can eventually result in overestimating management quantities (SSB). Towards the later part of the survey, the proportion of fish that were in the later vitellogenic stages increased (stage 3), which will make it easier to determine maturity from an at-sea macroscopic perspective. Furthermore, sampling and determining maturity during the summer might not be representative because sablefish might initiate maturity right after that summer survey period, and could be classified as immature, despite actually maturing later into the year. Macroscopic methods that are not standardized (i.e., the picture method) will tend to underestimate maturity, however, these standardized macroscopic methods performed fairly well with respect to accurately determining maturity when compared to histological methods. Sampling during the last leg of the survey will allow for the most accurate determination of maturity status because it is easier to tell apart oocytes that are in vitellogenic stages and oocytes that are at a perinucleolar stage. Other methods could also be coupled with macroscopic methods to ensure the accuracy of maturity determination. In particular, the use of a gonadosomatic and hepatosomatic index has been found to be successful in determining sablefish with periovulatory follicles, which are indicative of mature fish, although this tends to be more indicative of maturity towards the spawning season (Guzmán *et al.*, 2017). In a similar vein, (Rodgveller, 2019) used a body condition index and a hepatosomatic index as covariates to predict sablefish maturity, and found that incorporating such covariates allowed for maturity estimates to correspond with histological samples. The same study did also find significant differences in relative body condition and the hepatosomatic index between immature and mature females, suggesting that these indices can be of value in determining maturity status. These indices have similarly been found to be of use in determining maturity status in other studies (Wuenschel *et al.*, 2019).

## Use of SSB in fisheries management

Big Picture Questions

* Why does maturity and fecundity vary over time?
* Why are reproductive dynamics important to understand and incorporate in the management process?
* What are some methods we can use to better predict skipped spawning and the probability of being mature? What sorts of data are needed in this context?
* How can reference points better consider the reproductive biology of a species?
* What are some ways of accounting for skipped spawning in an assessment context?
* What are some alternative metrics we can use in our assessments?
* Why is it important to account for age-diversity in the SSB for our metrics in stock assessments?
* Is SSB a good measure of egg production?
* What are some core assumptions of SSB? Consequences if SSB is not proportional to egg production?

# Pete Hulson

## Magnusson Stevens Act and Management Objectives

The MSA is the primary legislature governing fisheries and uses the precautionary approach where the Council incorporates forward looking conservation measures that address differing levels of uncertainty. Further, the Council seeks to make decisions based on sound scientific decision and provide management proactively rather than reactively. FMP plans have to conform to the 10 National Standards set forth by the MSA. These are:

1. Prevent overfishing and achieve OY on a continuing basis,
2. Conservation and management based on the best available science,
3. Manage individual stocks as a single unit and coordinate with neighboring stocks if necessary,
4. Conservation measures should not be discriminatory or benefit others inequitably,
5. Conservation and management measures should account for variability in catch and the resource,
6. Conservation and management measures should be as efficient as possible,
7. Conservation and management measures should minimize costs,
8. Conservation and management measures should avoid harm and hardship to communities (economic) and provide opportunities for participation in the fishery,
9. Bycatch should be minimized,
10. Safety of human life should be prioritized.

In terms of management objectives for the council, they are broadly grouped into the following categories:

1. To prevent overfishing (conservative measures of harvest, evaluation of F40 rule, 2 million mt OY cap, adaptive management for dynamic range of OY),
2. Promote sustainable fisheries and communities (conservation measures balance both harvest of socio-economic considerations, equitable allocation of resources, and safety at sea),
3. Preserve food web (development of ecosystem indices, adjust ABC based on ecosystem factors, limit forage fish harvest),
4. Reduce bycatch (formation of bycatch incentive programs, bycatch limits and research on population status of non-targeted species, reduce economic-related discards, seasonal allocations and gear restrictions, account for bycatch mortality in TAC, reduce waste and maximize retention and utilization),
5. Avoid impacts to sea birds and sea lions,
6. Reduce impacts to habitats (area closures, HAPC, EFH designation and mapping),
7. Promote equitable and efficient fishery resource use,
8. Increase Native Consultation,
9. Improve data quality, monitoring, and enforcement.

## Tier System and Control Rules for Stock Assessments

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Tiers | OFL | ABC | a | b | c | HCR (Threshold) | Stock Recruit |
| 1 | FOFL = arithmetic mean of Fmsy | FABC = harmonic mean of Fmsy | Threshold control with = 0.05 using Bmsy | Threshold control with = 0.05 using Bmsy | Threshold control with = 0.05 using Bmsy | Only uses Bmsy reference points | Yes (Age-structured) |
| 2 | FOFL = Fmsy | FABC = Fmsy \* (F40 / F35) | Threshold control with = 0.05 using Bmsy and Fx% | Threshold control with = 0.05 using Bmsy and Fx% | Threshold control with = 0.05 using Bmsy and Fx% | For ABC setting, uses Bmsy and Fx% | Yes (Age-structured) |
| 3 | FOFL = F35 | FABC = F40 | Threshold control with = 0.05 using Bx% | Threshold control with = 0.05 using Bx% | Threshold control with = 0.05 using Bx% | Uses both Bx% and Fx% | No (Age-structured) |
| 4 | FOFL = F35 | FABC = F40 | NA | NA | NA | NA | No (Generally Age-structured) |
| 5 | FOFL = M | FABC = 0.75 \* M | NA | NA | NA | NA | No (Generally index method) |
| 6 | FOFL = Avg Catch (1979 – 1995) | FABC = 0.75 \* Avg Catch (1979 – 1995) | NA | NA | NA | NA | No (Catch only methods) |

## Status Determination

**Overfished Status:** Occurs when stock falls below the minimum stock-size threshold, which is defined as ½ of Bmsy or B40% depending on the tier of the stock. For tiers 4 – 6, since there is not estimate of Bmsy, overfished status is undefined. Thus, to declare overfished status, if the current year SSB is less than ½ of Bmsy, then the stock is overfished. If it is above, it is not overfished. If it is in between ½ of Bmsy and Bmsy, then we need to do stochastic simulations for 10 years, while fishing at the OFL and look at the mean SSB over that 10 year span – if it is below ½ of Bmsy, then the stock is overfished.

Once declared overfished, a rebuilding plan must be put in place with levels of FOFL and Fmsy that will rebuild the stock in a reasonable time frame.

**Approaching Overfished Status**: MSA requires us to determine whether a stock is approaching overfishing. This is done by projecting the current year SSB forward by two years and fishing at maxFABC so see if we fall below ½ of Bmsy. If we do, we are approaching an overfished status, vice versa. If we are in between ½ of Bmsy and Bmsy, we need to do stochastic simulations where the first two years are fished at maxFABC and the last 10 are fished at FOFL. If we fall below ½ of Bmsy in year 12, then we are approaching an overfished condition, and amendments need to be made.

For stocks assessed under tiers 4 – 6, it is not possible to determine whether it is overfished or approaching an overfished condition. Such determinations requires an estimate of B20% and B17.5%. However, it is possible to conclude whether overfishing is occurring, if catches surpass the OFL.

## Harvest Specifications

Definitions:

1. Maximum Sustainable Yield – the largest long term catch that can be taken from a stock or complex under the current prevailing environmental and fishery characteristics,
2. Optimum Yield – the amount of fish that provides the greatest benefit to the nation, is prescribed according to MSY with adjustments due to uncertainty, and provides a rebuilding plan to which allows us to reattain MSY if the stock is overfished. For the BSAI, the system wide OY is equal to 1.4 million mt – 2 million mt, while in the GOA, the OY is 116,000 mt – 800,000 mt.

Generally, harvest specifications follow a hierarchy. This hierarchy is as follows: 1) OFL = Maximum Fishing Mortality Threshold, 2) ABC, 3) ACL, which is generally equal to ABC, 4) TAC, which can be adjusted down from the ABC due to uncertainty, and 5) MSST (minimum stock size threshold), below which we are not allowed to fish (i.e., below ½ of B35% for Tier 3 stocks).

With respect to setting TACs, the general process is usually as follows: 1) determine maximum permissible ABC from Plan Team and SSC, 2) determinate an acceptable level of TAC based on current scientific information and socio-ecological considerations, and 3) make sure the sum of TACs fall within the BSAI OY range.

## NPFMC Management Measures

1. OY set at 1.4 million mt to 2 million mt, which may not be exceeded,
2. Area closures to reduce bycatch of salmon, crab, herring, crab, etc,
3. Area closures and transit closures to protect stellar sea lion foraging areas and walrus transit areas,
4. Area closures to protect unique habitats (Bowers Ridge, Alaska Seamounts),
5. Prohibited Species Catch limits and caps, wherein exceeding these limits can lead to area and fishery closures (Chinook, Herring, Crab),
6. EFH designations to understand the life-stages at which certain species are most vulnerable (although not really extensively used in management),
7. Use of prevailing ecosystem information and uncertainty in the management process to reduce ABC to set TACs,
8. Input controls to limit vessel tonnage, horsepower, and size,
9. No directed commercial fishery allowed on forage species

## AI Pacific Cod

### Biology and Life-History

Pacific cod generally occur at depths from the shoreline to about 500m. Eggs are demersal, and the larval duration is about 90 days. The survival of eggs and hatching success have been found to be dependent on temperature – recruitment is also highly influenced by temperature. Eggs move quickly to the surface upon hatching and larval stages have fairly good swimming abilities. Larval drift can be great, where larval stages have been found to be transported by currents from the Kenai and Kodiak to Unimak. Larvae tend to exhibit shoreward movement (i.e., inshore) but this relationship can vary.

Juvenile Pacific cod tend to settle near the seafloor and can also be dependent on nursery habitats (as inferred from Atlantic cod). Habitat use of juveniles are generally in shallower waters from coastal-demersal to shelf-pelagic (0 – 80m). Habitat distribution of juveniles is hypothesized to be as a result of density-dependence, temperatures, and prevalence of demersal predators.

Although juveniles have some flexibility in being somewhat pelagic to demersal, adults are strongly associated with the seafloor and diel vertical migration has been observed. Adults tend to form large spawning aggregations and are annual spawners. Furthermore, they undertake spawning migrations as well as feeding migrations (has been observed to travel about 100nmi to 500nmi).

### Fishery Characteristics

The fishery is a multi-gear fishery that is composed predominately of trawl, pot, and hook-and-line gear. The catches among these gears fluctuate quite a bit and the selectivities are likely different, which may suggest the need for time-varying selectivity. Nonetheless, most the catch results from trawl gear, followed by longline gear, and pot gear (pot gear seems to be the least). The fishery during the feeding season tends to be longline gear, while trawl nets are typically used during the spawning season, which tend to select larger fish. Given the migratory patterns there and the different gears used, some thought needs to put into how selectivity should be structured in this respect.

### 2020 AI Pacific Cod

#### SSC and Plan Team Comments

1. SSC suggests exploring averaging multiple surveys (not sure if there is data for that) as well as using a VAST model for the purpose of apportionment, but not updates were made and no VAST models were explored due to a lack of a survey,
2. The SSC appreciates the efforts in the exploration of age-structured methods and recommends further explorations. They also recommend fitting the maturity curve inside the assessment as well as using *M­*-prior methods from Jason Cope to explore the estimate of natural mortality used in the assessment. Neither of these were done in this current assessment cycle. The age-structured model was not updated due to a lack of survey data, although I believe an effort should have been made to look at an alternative estimator for *M*.

#### Assessment Structure

The 2020 AI Pacific Cod stock assessment is defined as a Tier 5 stock ***(REMA model).*** As such, the FOFL is defined as *M* (0.34) while the FABC is defined as 0.75*M* (0.255), and the resulting catch advice would be the biomass estimated, multiplied by these quantities. Thus, the 2019 AI Pacific Cod assessment uses a simple random effects model, wherein the index of the NMFS Bottom Trawl Survey in the AI region is used (triennial early on and biennial in recent periods):

where is the observed index of abundance (trawl survey biomass), are lognormally distributed deviates constrained by the variance observed from the trawl survey, and are treated as random effects and represent the true unobserved trawl survey biomass, which follows a random walk process:

where are deviations from a random walk, which are constrained by a process error variance term (. In the assessment, only one process error variance term is estimated as a fixed-effect, and are estimated as latent unobserved random effects. Essentially, this model smooths over the observed survey biomass using a state-space random walk model.

Considering that this is a tier 5 stock, an estimate of *M* is needed to provide management advice, which is multiplied with to provide an estimate of catch advice (either *M* or 0.75*M*). This value was updated several times in accordance with natural mortality assessments from the EBS Pacific cod assessment, but this practice is no longer done given concerns with the EBS *M* estimate not necessarily being equal to that of the AI. Thus, I believe the authors here have continued their use of Jensen’s age-at-maturity estimator for *M*, which is currently specified at 0.34.

The assessment estimates fairly wide confidence intervals for 2020, which they attributed to a lack of survey data since 2018 (i.e., no survey data in 2019 and 2020 because of COVID).

#### Harvest Apportionment

For AI Pacific Cod, there are several issues that need to be addressed with respect to apportionment of harvest. These include: 1) apportionment of harvest from the state, and 2) apportionment of harvest given Stellar Sea Lion protection measures.

Prior to 2014, apportionment of BSAI Pacific Cod for the state was done by multiplying 3% of the TAC from the assessment. Following 2014, when the AI Pacific Cod stock was managed separately from EBS, apportionment to the state was done by multiplying about 27% - 39% of the AI Pacific Cod to the state. The percentage increases by 4% if the catch reaches 90% of the guideline harvest level from the state in the previous year, but may not exceed 39% (6804 t).

Prior to 2014, there was a regulation that prevented the harvest of Pacific Cod in Area 543 (in the AI region). However, there was an amendment in 2015, that now puts a harvest limit in Area 543, instead of fully restricting fishing. There, an apportionment of the TAC is made by:

where is the harvest limit in area 543 and is calculated by subtracting the TAC by the guideline harvest level from the state, and then multiplying the proportion of biomass in area 543 relative to the total biomass of the assessed stock (). This can be calculated several ways including: 1) using the average survey raw proportions, 2) the most recent survey raw proportions, 3) using the average survey estimate from Area 543 and dividing that by the average survey estimate from the entire area, and 4) using the most recent survey estimate from 543 and dividing that by the survey estimate in the most recent year. The last approach is what was used in this assessment.

### 2021 AI Pacific Cod

#### SSC and Plan Team Comments

1. SSC suggests exploring averaging multiple surveys as well as using a VAST model for the purpose of apportionment, but not updates were made and no VAST models again this year,
2. The SSC appreciates the efforts in the exploration of age-structured methods and recommends further explorations. They also recommend fitting the maturity curve inside the assessment but this was not done due to potential confounding with ageing error – which I do not think is a great excuse – fitting the maturity curve in the assessment would allow for the propagation of ageing error in the maturity curve,
3. Several growth models were investigated and it was eventually decided based on AIC and parsimony that the von Bertalanffy growth model would fit the best,
4. The SSC requested the author to look into the use of Jason Cope’s method for determining *M*. This was done in this assessment cycle and the estimate came out to 0.36, although a point estimate of 0.4 was used in the assessment model this cycle,
5. The Plan Team wanted the author to bring forward both model runs of maturity (observer and Stark 2007). The teams recommended the observer data because it has more samples and is more representative than the Stark (2007) estimates, but wanted histology verification,
6. The authors requested guidance on data-weighting from the SSC, but no data-weighting exercises were really attempted this assessment cycle (survey ages weighted by hauls, and fishery was set to standardize at a mean of 20 for ISS),
7. Results from 3 age-structured assessments were brought forward this cycle, and the SSC recommended not bringing forward one of the models that dropped fishery length data. These assessments differed in 1) the estimate of M (0.34 vs. 0.4) and 2) the maturity curves used.
8. Despite the *M­-*prior methods suggesting a point estimate of 0.36, the SSC and authors recommended a value of 0.4, to balance the tradeoff in the likelihood profile indicated by the fishery and survey (0.3 vs. 0.8). The estimate of 0.4 follows the general mode of *M* used in this assessment in previous years. However, there is no firm justification as to why the value of 0.4 was really chosen.

#### Assessment Structure

##### Data

The 2021 assessment brings forward 3 age-structured assessment models ***(old maturity curve + M = 0.34, new maturity curve + M = 0.34, new maturity curve + M = 0.4)*** and 1 tier 5 assessment model (rolled over from 2020). For the tier 5 model, it uses what was described in the 2020 AI Pacific Cod assessment model – BTS biomass estimates using a random-walk smoother.

The age-structured assessment model uses the following datasets:

1. Fishery catch (1991 – 2021; more during spawning season (winter), larger fish)
2. Fishery size compositions (1991 – 2021)
3. Biomass index from BTS survey (biennial and triennial, most recent = 2018; more during summer months, smaller fish),
4. Age composition from BTS survey (biennial and triennial, most recent = 2018)

In the age-structured assessment model, there are a decent number of age-composition samples available for use in modelling (500 – 1000 samples per survey year). Length-composition data are not used for the survey because age-data are available. The survey biomass index is a design-based index that is expanded to the strata and summed. The NMFS LL survey is also conducted in this same region, which I believe to be biennial. I believe that the author should make efforts to incorporate these data into the age-structured model and investigate the quality and quantity of length and age-composition data from the LL survey.

For fishery data, gears and statistical areas from AI are combined and modelled together. Catch data are simply summed. For length composition data, most of these data come from longline and trawl fisheries. The assessment document indicates that length compositions are combined by weighting by the relative catch in each statistical area. Depending on how well each gear, season, and area overlap, this might not be an appropriate approach, and the length frequency data should instead be weighted by relative catch across seasons, gears, and areas given that there appears to be differences in the sizes of fish caught depending on these characteristics (larger fish during winter months, which tend to be trawl gears). However, given that most of the fishing takes place in the winter months, season weighting might not be necessary. Nonetheless, I think gear \* area might be an appropriate relative weighting scheme for these data.

##### Model Structure

The model used is an age-structured model with a single-sex and a 1:1 sex-ratio (10 ages, 10+ is the plus group). The survey and fishery both assume logistic age-based selectivity. A growth function following von Bertalanffy dynamics is estimated outside the assessment, which is used to construct an age-length matrix as well as compute weight-at-age. An ageing error matrix is used in the assessment as well. Recruitment and fishing mortality parameters are time-varying, while all other parameters are constant. Survey catchability and selectivity are estimated within the assessment, while maturity is estimated outside of the assessment. Natural mortality is fixed inside the assessment (either as 0.34 or 0.4) and fishery length frequencies are weighted by the relative catch in each statistical area.

Data-weighting methods were attempted to weight survey age-composition and length frequencies but led to unreasonably high likelihood weights – this resulted in decreased survey catchability and biomass estimates, which do not seem to make a whole lot of sense. You would expect decreased survey catchability to increase biomass estimates. Nonetheless, they ended up weighting the samples using the number of hauls in each year for survey age compositions. Different weighting values resulted in expected changes in fits to data sources. Increasing weights to length-frequencies led to better fits to these data, but poorer fits to survey indices. However, this tended to result in poor convergence criteria. Thus, they ended up weighting the length frequency data as the number of lengths sampled to retain annual sampling variability, but weighted so that the mean of the weights were 20, so as not to overwhelm the model by forcing fits to length data.

A new maturity curve was also implemented in this model, because the old maturity curves was based off the EBS and had limited samples. The new maturity curve was similar but had a slightly lower age-at-50% maturity (4 years old). The maturity curve was estimated outside the assessment using length-based maturity, and was then converted to age-based.

Length-at-age and associated age-length keys were also constructed using a von Bertalanffy growth model, instead of more highly parameterized and flexible models as they all performed similarly. The age-length transition matrix was constructed by simulating values conditioned on the mean length-at-age and its associated CV/variance. Length-at-age was then converted to weight-at-age using an allometric relationship.

With respect to natural mortality, a value of 0.34 was used based on age-based maturity estimators of *M* initially. The estimate of natural mortality differ quite a bit from the GOA and the BSAI stocks (0.47 vs. 0.3ish). Likelihood profiles of *M* showed strong data conflicts – fishery length data indicated a value of 0.3 would be best, while survey ages, biomass, and the recruitment penalty suggested a value of 0.8 which is very unreasonable (but this could also be due to large fish moving out of the surveyed area – i.e., not seeing any large fish there which would result in *M* being estimated fairly high, while the fishery sees large fish, which is why you see that data conflict there; the conflict could also be due to the seasonal and spawning migrations as well as the fishery and survey operating in different times). Following this, a value of *M* = 0.4 was used instead to compromise between these data conflicts, which also matched some of the modes of *M* used in this assessment. 0.4 was considered a good starting point given estimates of 0.47 from GOA and 0.348 from the EBS assessment. The *M*-prior methods from Jason Cope suggested a value of 0.36. Given all of the above, *M* was set at 0.4 in this assessment.

Catchability estimates were estimated inside the model for the survey, but the fishery was set at 1 based on *a priori* knowledge and studies of cod availability to surveys (fairly high). Catchability estimates changed when *M* was fixed at different values (lower catchability for higher *M*), and it might be worthwhile to attempt to estimate this value as opposed to fixing it, given the large difference in catchability when *M* is fixed at a higher value. Selectivity for both the fishery and survey were estimated as logistic age-based. One could make an argument for a more flexible selectivity form for the fishery given the combined fleet structure, although I realize its informed by primarily length-composition data. However, it might be worthwhile to look at length-based selectivity given the reliance on length-data here, although it may not work as expected given that Pacific cod can demonstrate dramatic fluctuations in size-at-age throughout the season and annually. The justification of similar selectivities between the survey and fishery is that the length distributions match up fairly well at those initial ages. Furthermore, looking at the data, dome-shaped selectivity does not appear to be warranted – likely because they do not reside in untrawlable habitats and larger cod do not appear to leave the region entirely.

##### Model Results

Regarding the use of diagnostics, they examine goodness of fits statistics and residuals for the biomass index and composition data. They also use likelihood profiles for *M* to examine potential data conflicts and also use retrospective diagnostics. However, given the model instability to data-weighting, I think an additional jitter analysis should be conducted to see how much results change with incremental changes in initial values. Additionally, while a profile of *M* was conducted, I think a profile of *q* for the survey would also be appropriate to understand the values governing this parameter and whether composition data conflict with this estimate. A conflict in *q* could also be due to mis-specified selectivity (which does not seem to be an issue based on fits). In due to fixing *M* at 0.4, which may lead to some inconsistencies if 0.4 is not reasonable as determined by the model and could general, fits to these individual data sources appeared adequate. However, retrospective analysis indicated that SSB was consistently positive biased (Mohn’s rho = 0.15). They use Hurtado-Ferro’s paper to justify that the retrospective inconsistencies were not significant – constitutes best available science at that time.

To conclude, the authors recommend this stock to use tier 5 designations for harvest specifications, but for the SSC to consider upgrading the assessment to tier 3 designation.

##### Summary on Problems in Assessment

Below is a summary of what I think is going wrong with the age-structured assessment at this point:

1. The estimate of *M* used does not have good justification. Furthermore, there seems to be data conflicts of what *M* should be based on likelihood profiles between fishery data and survey data,
2. There seems to be some misfits and large shoulders in the fishery length composition data, which can be dealt with by separating out the fleets,
3. Given that the fleets operate at very distinct times, it may be worthwhile to disaggregate some of them (pot and longline = aggregated, trawl = separate),
4. Furthermore, while the authors suggest that there is no evidence for dome-shaped selectivity in the survey, this seems a bit dubious because of the life-history of Pacific cod (seasonal and spawning migrations), which can lead to dome-shaped selectivity (O’Boyle *et al.*, 2016). Looking at the age-and length-composition data for the survey and comparing them against the fishery, the survey selects more younger fish with a strong peak in those middle age-classes, while the fishery selects a more uniform spread of age-classes. This may suggest a potential for dome-shaped selectivity and should be explored,
5. There is no biomass index available for the new age-structured assessment, leading to increased uncertainty,
6. Additionally, the biomass-based index is not well fit in the assessment,
7. Despite the authors suggesting the retrospective bias is not significant, I tend to disagree. The retrospective pattern is potentially a result of aggregating all the fleets, but treating them as time-invariant – there seems to be shifts in the catch ratio among fleets,
8. The uncertainty in ageing from the maturity curve is not propagated into the assessment,
9. Data-weighting methods were unsuccessful – further methods (e.g., Francis instead of McAllister and Ianelli can be explored),
10. Lastly, the interpretation of fishery performance and the use of CPUE in this assessment is incorrect – you cannot just aggregate all those gears with different units of effort and interpret fishery performance from that.

##### Recommendations on Model Structure

Given that the survey and fishery operate at very distinct times, I think the estimate of *M* might be different in terms of the likelihood profiles between these different data sources because the survey isn’t seeing as many old fish, while the fishery explicitly targets old fish. Furthermore, the difference in survey and fishery timing might resulting in some confounding in the estimate of *M* because of the seasonal migrations of feeding and spawning (i.e., large Pacific Cod move out of the area to feed), suggesting the potential need to construct a spatial model to examine potential movement rates, and better reconcile the estimates of *M*. However, looking at the likelihood profiles, it seems like the survey data suggest that the profile for *M* is really flat beyond 0.3, and the likelihoods do seem to dip at 0.3. I think 0.3ish is a good estimate, but definitely would be interesting to look at some of the spatial structuring in this respect to see if the descending limb of the likelihood profiles for these datasets bound back up.

Spatial structure aside, the fishery also appears to operate at very distinct times. In particular, the trawl fishery seems to primarily operate during the months of Feb – Mar, while the longline and pot fisheries operate year-round. Considering these seasonal movements, I think that the pot gear and hook-and-line gear should be combined into its own fleet, while modelling the trawl fishery as its separate fleet and following *pulse-*fishing dynamics potentially. This would be a fruitful avenue to explore. This is potentially supported by some of the bimodal length frequencies that are observed in these data and some misfits to fishery data (long shoulders could be because of combining data from gears with distinct selectivities).

### 2022 AI Pacific Cod

#### SSC and Plan Team Comments

1. The SSB recommends changing the point estimate of *M* to 0.36 as in Jason Cope’s method. However, the assessment opted to estimate *M* inside the assessment and without a prior,
2. Maturity was requested to be fit within the assessment model, but was opted to be estimated outside the model in this cycle,
3. The SSC recommended using the LL survey in the age-structured models, where this was used in the assessment during this cycle,
4. The plan team recommended some data-weighting exercise and sigmaR being tuned, to which was employed in this assessment cycle (Francis and iterative tuning of sigmaR)

#### Assessment Structure

##### Data

In the 2022 AI Pacific Cod assessment, two age-structured assessments are brought forward ***(all data without LL survey and combined fishery, another with all data including LL survey but disaggregated fisheries)***. A tier 5 biomass-based index assessment is also brought forward. The age-structured assessment models uses the following datasets:

1. Fishery catch (1991 – 2022),
2. Fishery size compositions (1991 – 2021)
3. Age composition data from the fishery (2020 – 2021),
4. Biomass index from BTS survey (biennial and triennial, most recent = 2022; more during summer months, smaller fish),
5. Age composition from BTS survey (biennial and triennial, most recent = 2018),
6. Abundance index from LL Survey (biennial, most recent = 2022),
7. Size composition data from LL Survey (biennial, most recent = 2022)

##### Model Structure

As noted above, two age-structured models are presented. These are:

1. Model 22.0, where the fisheries are combined and the length frequency data are weighted by catch, season, gear, and area, and also uses the BTS Trawl Survey (164 parameters),
2. Model 22.1, where fisheries are disaggregated and uses both the BTS Trawl Survey and the LL Survey (274 parameters).

In these models, they are single sex, with growth estimated *within* the model, an ageing-error matrix is incorporated, fishing mortality, recruitment, selectivity, and catchability are estimated within the model, natural mortality is also estimated *within* the model (presumably with a prior), and the trawl survey has a prior of 1 for catchability. It appears that the standard deviation of the catchability prior is at 0.01, and thus is essentially fixed in the assessment. Additionally, the maturity curve is estimated outside the model using observer data. Additionally, these models were weighted using Francis-reweighting. However, no weights from Francis-reweighting are provided, which makes it impossible to know which data sources are being up or down-weighted.

For selectivity, all fisheries and surveys were fit using a double normal. For most fisheries and surveys, the double normal reached an upper bound so the descending limb was fixed to resemble logistic selectivity. However, for the LL survey, the descending limb was estimated and was dome-shaped due to a lower proportion of older fish caught.

With respect to natural mortality, the document says that it was estimated within the model, but provides no detail on whether a prior was used. In terms of the recruitment standard deviation, this was tuned iteratively, following methods of Thompson et al. (2008).

##### Model Results

The combined fleet model (22.0) fit the trawl survey biomass estimate much better than the disaggregated fleet model (22.1), but both models generally fell within the confidence intervals. I believe that there is model instability and confounding caused by the number of parameters estimated by the disaggregated fleet model (22.1), resulting in poor survey index fits as well as unreasonable estimates of catchability when estimated without a strong prior.

Likelihood profiles were constructed and profiled across R0. In general, for model 22.0, the likelihood profile for the fishery agreed with the MLE, while for the survey, it showed higher likelihoods at larger values. However, looking at the plots, its hard to figure out what component is attributed to what. But for model 22.0, the estimate of R0 was in between the fishery and survey data. In terms of model 22.1, the trawl fishery suggested a much higher R0 value relative the longline fishery and the trawl survey. The pot fishery and LL survey showed an increasing likelihood of R0 as it increased, and no local minima were detected. The fact that likelihood profiles show conflicting results for R0 is likely to suggest that there is some mis-specification for selectivity. This is likely/potentially the case potentially attributed to the time-invariant selectivity estimated, although the use of the R0 profile has sometimes been shown to be inconsistent (Lee *et al.*, 2014). Additionally, it could be due to selectivity for some of the fisheries being dome-shaped rather than asymptotic, which can result in the scaling parameter/absolute abundance being conflicted between those data sources.

Both models estimated somewhat reasonable levels of natural mortality (0.35 and 0.36), which are consistent with other methods and assessments that estimate *M* for Pacific cod. However, it remains unclear whether a prior was used, and whether use of a prior for natural mortality would help stabilize some of these results.

In terms of retrospective analysis, both models showed significant retrospective inconsistency when peeling back 10 years, which were about 0.25 to 0.31. While the authors attribute this to a lack of survey data in 2020, it is most likely that fishery selectivity is mis-specified and there are likely some time-varying components within that, that remain unmodeled in the assessment. Additionally, it remains unclear how exactly retrospective analysis was conducted – for example, were the same data-weights used from the terminal assessment, or was retrospective analysis conducted in conjunction with Francis-weighting. Additionally, a positive retrospective inconsistency could also be a result of *M* being higher during heatwave years and the use of a time-block as well as other *a priori* knowledge during this period to estimate a separate value would be valuable.

Overall, model 22.1 led to a poor fit to the trawl survey biomass relative to model 22.0. This misfit could be due to conflicting signals from the LL survey given that it may be unrepresentative of AI Pacific cod population trends. Additionally, likelihood profiles were particularly conflicted in this model, potentially due to certain data components suggesting larger stock sizes, which could be attributed to differences in fishery and survey timing. By contrast, model 22.0 performed reasonably with smooth likelihood profiles about R0, however, a fairly strong retrospective pattern was detected, likely attributed to a lack of a survey for multiple years.

In terms of the time-series trends, the stock has experienced a continued decrease from 1992, although in recent years there are several indications of increases in the population. Total biomass estimates from 22.0 predict smaller increases, whereas the scale of increase from model 22.1 are much larger (2-fold of bottom trawl survey). Furthermore, model 22.0 shows less optimistic stock trajectories as model 22.1 (22.0 is at about B20, while 22.1 is a decent amount above that). Recruitment estimates also seem to be higher for 22.1, although both models follow similar trends. Interestingly, despite all the surveys showing downward trends, the population trajectories from the age-structured assessment models trend up. However, the assessment notes that the LL survey may not be representative because: 1) it only targets half of the Aleutian Islands, and 2) variable sampling due to gear loss and Pacific cod are a non-target species. Looking at some of the trends between the surveys, they track each other fairly nicely, despite the difference in spatial scale – although the LL survey shows signs of increase in 2015 while the BTS survey shows stability – which could potentially be due to differences in selectivity (asymptotic vs. domed). However, it is unclear what these catchabilities for the LL survey are estimated at from the document. Knowing if the catchability was much lower for the survey would provide indication of what is driving the difference in scale between these two models.

Comparing model 22.0 and 22.1, the projected SSB is almost two-times more than for model 22.1 in the terminal period. The recommended harvest levels are also almost two-times more for model 22.1. It is apparent that there is a scaling issue present in these models, given difficulties in estimating survey catchability as well as conflicts in the R0 parameters. In particular, there is almost a 20% difference in R0 between the two models. However, it is unclear whether the scale issue is due to: 1) disaggregating the fleets, or 2) the use of the LL survey, and would be understood if the author used incremental changes instead of skipping certain developments of the model. The difference in scale also starts to begin in year 2009, prior to which, the LL survey still had data. This is similarly seen in recruitment estimates, while model 22.1 had slightly larger recruitment estimates prior to 2009, they significantly diverged from 22.0 after 2009. Note that during this period, data from the BTS survey and the pot fishery (size-composition and index) were lacking and the only available data that were available were from the LL survey. Furthermore, the LL survey index trends up and this tracks with when this divergence in biomass also begins trending up.

Given sparse data available for the pot fleet, I do not think there are sufficient data to do a fully fleet-disaggregated approach, which may be leading to model instability. It would have been helpful to provide the standard deviations of selectivity parameters for these fleets and the potential exploration of a combined hook-and-line and pot fleet model would be warranted, where the trawl fleet is separate. Furthermore, looking at some of the selectivity profiles, the longline and pot fleet have similar selectivities, suggesting that it is likely appropriate to combine these two fleets together.

##### Summary on Problems in Assessment

Below is a summary of what I think is going wrong with the age-structured assessment at this point and what was also pointed out by the authors:

1. The biomass index from the BTS survey is not well fit for both models,
2. In model 22.1, the LL survey is also not well fit,
3. A large retrospective pattern was detected in these models, which is potentially due to: 1) not modelling time-varying *M* during the heatwave period, and 2) selectivity for the fisheries are varying,
4. For trawl survey catchability, the model with multiple fleets (22.1, 274 parameters) led to a very low estimate of catchability when estimated freely, which is inconsistent with how trawls interact with cod. Thus, a strong prior with a mean of 1 and a sd of 0.01 was imposed. This might be happening due to the estimation of *M* within the model, which does not appear to be constrained by a prior,
5. The BTS survey size-composition data does not appear to be fit well for model 22.1. Furthermore, no plots of model fits to CAAL data are provided for either models, making it hard to determine model adequacy,
6. It is unclear how well estimated the selectivities for model 22.1 are. Given the lack of composition data for some of these fleets and the reliance on only length composition data, some of these parameters may not be well estimated, which might be confounding with other parameters (*M* and *q*). Associated uncertainty for these parameters should be provided,
7. For the point above, give the lack of data for the pot fleet and the similarities in estimated selectivities between the two gears, these should be combined into a fixed-gear fleet, while the trawl fleet should be separate given the season in which the gears operate in,
8. While the assessment authors suggest that dome-shaped selectivity is unreasonable in the BTS survey, the size-compositions suggest (and the lack of fit for model 22.1) that dome-shaped selectivity options should be explored, especially when considering the life-history of the species (seasonal migrations can lead to availability differences resulting in dome-shaped patterns),
9. The author notes that a double normal was used to estimate all selectivities for the fisheries. However, a double-normal consists of 6 parameters and it remains unclear how many parameters were fixed. If more than 2 parameters were estimated for these logistic selectivities, additional exploration looking at a 2 parameter logistic selectivity would be warranted to helps stabilize the model,
10. Lastly, the likelihood profiles for model 22.1 suggested substantial data conflicts, with the trawl fishery suggesting a much higher scale compared to all other data sources. This could be due to a variety of factors, including mis-specified selectivities, natural mortality (i.e., might be time-varying), growth mis-specification (given the reliance on length data), and just general model stability and sparse data for some fisheries.

##### Recommendations on Model Structure

Firstly, the assessment does a poor job of showing the influence of particular data sources. It would be incredibly beneficial to have an exercise that sequentially adds new data into the model, to understand the influence of particular data components, surveys and fisheries. Next, it somewhat worries me and seems dubious that *M* is so well estimated with low uncertainty even without a prior imposed on it. Given this point, I was surprised to see no attempt some form of a jitter analysis to see how much *M* changes as you change the starting values. It seems dubious that length-composition data are so influential in determining these natural mortality estimates and a likelihood profile on *M* should be done. My next point is in regard to the strong prior placed on survey catchability – which is essentially fixed at 1. It is possible that this strong prior placed on catchability might result in a fairly well estimated *M*, given these two factors are confounded, and might be interplaying with some of the lack of fit to the indices. In addition, it is unclear from the assessment document what the value of catchability for the LL survey is and how it is parameterized. If it still has a strong prior of 1, that seems unreasonable for a LL survey, especially given that the survey does not directly target Pacific cod. Clarification needs to be provided here. The assessment notes that dome-shaped selectivity is not reasonable for the BTS survey. However, given these seasonal movements, I would actually argue that it might be reasonable and would warrant exploration (O’Boyle *et al.*, 2016).

Another key issue with the use of the LL survey is that it only covers half of the area occupied by the stock, which may be a driver of the lack of fit in to the index. I think a potential solution could be to model a single aggregate survey fleet by standardizing the indices between the BTS and LL survey, while modelling and controlling for spatial-temporal effects. Although there is little literature on this specific topic, it may be a fruitful avenue to explore if we want to incorporate all possible data sources into the analysis. However, I recognize that there might be some difficulties in implementing such an approach given the large spatial imbalance, the need to aggregate composition data. The other issue I have is it remains unclear how catchability can be 1 for your survey, but the biomass estimates are almost 2-times larger.

With respect to fleet structure in model 22.1, given the lack of data for the pot fishery and similarities between pot and longline gears (selectivity and seasons), exploration should be conducted to combine these two gears, while modelling the trawl fishery as a separate fleet. Perhaps the longline and pot gear fleets should incorporate some form of dome-shaped selectivity as it appears that the larger age/size-classes are less selected relative to the trawl fishery. Furthermore, selectivtiies that are monotonically increasing should be modelled as an asymptotic logistic curve, rather than as a double-normal to reduce the number of parameters estimated and improve model stability, if more than 2 parameters are used to estimate the double-normal.

My next point is in regards to the use of conditional-age-at-length data, despite this method generally thought to be reserved for data-rich assessments, where this stock imo should be considered relatively data poor. Explorations should be conducted into fitting marginal age-compositions and length-compositions instead of conditional-age-at-length, given that the use of CAAL can lead to biases in growth and management quantities if these data are not representative of the age-structure of the population – no model fits nor justification was provided in terms of the use of these data (Lee *et al.*, 2019).

### 2023 AI Pacific Cod

#### SSC and Plan Team Comments

#### Assessment Structure

##### Data

In the 2023 AI Pacific Cod assessment, three age-structured assessments were brought forward. The general model structure for this assessment combines fisheries which uses fishery length composition data and catch data, while only one survey fleet is used (BTS survey with CAAL data). ***The three assessment models are: 1) three growth time-blocks to account for shifts in growth and warmer temperatures, 2) five selectivity time-blocks to accommodate shifts in fishery targeting practices, and 3) two time-blocks on growth and two time-blocks on natural mortality to accommodate warmer temperatures.*** A tier 5 biomass-based index assessment is also brought forward. The age-structured assessment models use the following datasets:

1. Catches from the combined fishery (1991 – 2023),
2. Abundance indices from the BTS survey (1991 – 2022),
3. Length composition data from the combined fishery (1991 – 2022),
4. Length composition data from the BTS survey (1991 – 2022),
5. CAAL data from the BTS survey (1991 – 2022)

Note that in the current assessment year, no trawl survey was conducted. To summarize, the model uses CAAL data for the survey, length composition from the fishery, an index from the survey, and catch data from the fishery. A fair number of ages are sampled from the survey, almost surpassing sablefish in some cases.

##### Model Structure

As noted above, three age-structured models are presented. These are:

1. Model 23.0, where three growth time-blocks are incorporated due to a documented change in growth using Maia’s method for detecting shifts in growth,
2. Model 23.1, five fishery selectivity time-blocks are incorporated on the ascending width of a double-normal,
3. Model 23.2, two time-blocks on natural mortality and two growth time-blocks due to changes in regimes (marine heatwave).

In these models, they are single sex, with growth estimated *within* the model, an ageing-error matrix is incorporated, fishing mortality, recruitment, selectivity, and catchability are estimated within the model, natural mortality is also estimated *within* the model (***without a prior***), and the trawl survey has a prior of 0 for catchability in log space. It appears that the standard deviation of the catchability prior is at 0.01, and thus is essentially fixed in the assessment. Additionally, the maturity curve is estimated outside the model using observer data. Additionally, these models were weighted using Francis-reweighting. Furthermore, input sample sizes were developed using a bootstrap method to reflect annual variability in sampling using Pete and Ben’s method.

For selectivity, all fisheries and surveys were fit using a double normal. A single combined fishery fleet was modeled and it its generally appropriate how the length composition were weighted and expanded and used in the model. Similar to the previous assessment, Francis-reweighting was used.

For growth, which was examined outside the assessment initially, a shift in growth in 2002 was detected. The author indicates that a shift in growth occurred in two time-blocks, but does not provide data or figures to support this (i.e., the first derivative plots in Maia’s paper). Furthermore, the authors indicate that there was a shift in growth by using length-frequency data pre and post 2004 (2004 because lack of survey data in 2003), and shows that there were differences in the length frequencies. However, this could have entirely been due to a change in the age-structure of the population and not necessarily attributed to growth. An additional time-block for growth was implemented at about year 2017, which was not supported by Maia’s method, but was justified due to a decline in survey biomass. The authors justify estimating a growth time block during this period to account for a decline in survey biomass, which seems inappropriate – declines in survey biomass could be due to below average recruitment, increased mortality, and a variety of other factors; not necessarily because of a change in growth. This was not supported by a data-driven method, and I don’t see why it was attempted here. Nonetheless, it appeared that incorporating a new growth time-block (3 to be clear) allowed for better fits to the survey biomass data in recent years (the declines).

Natural mortality was suggested to change in 2015 due to a thermal lag from marine heatwaves in 2013-2014, which appears appropriate and justified from other Pacific cod assessments. Incorporating a change in natural mortality appeared to allow for better fits to the survey biomass index, which seems appropriate.

For BTS survey catchability, two options are presented in the document. In general, catchability is estimated in log space with a strong prior of 0 and a small standard deviation of 0.01. The first approach used a prior to estimate catchability, while the second approach was a sensitivity test to examine the influence of the prior, wherein catchability was analytically calculated. The author posits that small differences were detected between these two options, however, the virgin SSB and natural mortality were fairly different. This might be due to the inherent correlation in *M* and *q*, and the removal of the prior led to this difference (prior is very informative).

While the author indicates that selectivity for model 23.1 was estimated with five time-blocks, no justification was given as to why these blocks were used. Lastly, natural mortality was freely estimated as two time-blocks in the model.

***Note that in this assessment, time-blocked parameters for growth and natural mortality have an extra estimated parameter. In particular, if two time-blocks were implemented, three parameters were estimated. The correct and more appropriate method would be to estimate two separate parameters, instead of three. It appears that the current parameterization uses a mean parameter and estimates additional deviations about this mean.***

##### Model Results

In general, all three models fit the survey biomass and length frequencies well. However, it appears that model 23.2 fits the biomass index the best. The author uses AIC to compare models and says that it is appropriate given that the same datasets are used. However, this is incorrect because Francis-reweighting is used to weight the likelihoods, and thus, your likelihood components are going to be different, and AIC is not a valid tool to use in this respect. However, if Francis-reweighting is not used (it doesn’t really seem like it and no weights are provided), then this approach is appropriate.

Likelihood profiles about R0 did not seem to indicate anything jarring and was a compromise between fishery and survey data. Retrospectives were best for model 23.2, however, there are several issues with the retrospective analysis, especially when time-blocks are used. In particular, the retrospective peels prior to the implementation of time-blocks are not comparable as they constitute entirely different models. Nonetheless, I still believe that it provides insight with respect to how the assessment would have performed if those parametrizations were used during that particular period. Furthermore, it remains unclear whether reweighting methods were used in tandem with retrospective analysis or if the likelihood weights were kept the same. For all 3 age-structured models, the retrospective patterns were within reasonable bounds.

At present, the stock does not appear to be in good condition, which could be due to adverse ecosystem conditions and impaired recruitment/survival. SSB has declined to about B20%, which would indicate a fishery closure is likely necessary. ***This is one of the detriments of using Tier 5 methods, is that it is not possible to tell whether a stock is in overfished condition. Perhaps an alternative would be to use the start of the time-series and assume that it is in unfished condition, take 20% of that value and use it as our limit reference point, while applying potentially a knife-edged or threshold control rule. Another issue I have with the Tier 5 control rule is that it doesn’t take into account the sex-ratio of the population, where we apply M to the total biomass observed. In an assessment that is age-structured, we often assume a 50:50 sex-ratio and your reference points (Bx% and Fx%) are conditioned upon that. Nonetheless, the Tier 5 control rule might not be as conservative as we actually think it is and could potentially lead to recruitment overfishing if we treat both the male and female population of the stock equally.***

Lastly, the assessment did their harvest projections using the most recent growth block but average natural mortality. I believe they presented more projections in their slides, but these are not present in the document. Nonetheless, none of the projections presented indicated that the stock was subject to overfishing or was in an overfished condition. However, note that there is potential for the stock to actually be below B20% based on the confidence intervals, and there is concern for fishery closures there. Based on ecosystem considerations, it appears that the stock may be experiencing better conditions in the recent period (lower summer temps (but still above average), higher winter temps, more Pacific cod in seabird diet data), but signals are variable based on other indicators (temps that are still above average indicate higher bioenergetic demand, etc, lower sea lion abundance).

##### Summary on Problems in Assessment

Several problems are identified in this assessment:

1. Unnecessary parameters are estimated for growth and natural mortality time blocks. Furthermore, there is not enough sufficient justification for some of the time-blocks for the fishery as well as the most recent time-block,
2. The growth time-block parameters are estimated very similarly. There is no reason why these should be treated as separate parameters. Additionally, a figure of length-frequencies is presented to justify a growth block. I do not see how that is appropriate given that this could be due to a shift in age-structure, not necessarily growth. However, it does potentially help justify a shift in *M*, given the abundance of younger fish observed, relative to older fish (i.e., individuals are dying more quickly and are unable to reach older ages),
3. Additionally, the use of such an informative prior on *q* is somewhat concerning. I think that the reason the analytical *q* method was so different in terms of *M* and SSB estimated was because of the strong prior imposed on *q*, and the strong correlation between the two parameters,
4. Next, the retrospective patterns are not necessarily comparable within a given model, considering the use of time-blocks changes the model structure, and does not result in a apples-to-apples comparison, although I do believe that it provides insights as to how the assessment would have performed given its retrospective assessment structure. The variation in retrospective trends in those initial peels for models 23.0 and 23.1,
5. Additionally, the use of AIC to compare models in this assessment is invalid. This is because the likelihood weights applied are different when Francis-reweighting is used. Although the document indicates that Francis-reweighting is used, no likelihood weights are provided and the likelihood components are very similar. Clarification is needed there,
6. There does appear to be a shift in selectivity within the fishery although figures of these blocks are not provided. Given the dynamic nature of fleet structure in this stock, time-varying selectivity really should be considered in conjunction with mortality. Furthermore, given some of the size-compositions relative to the trawl fishery for longline and pot, I think it would be prudent to investigate dome-shaped selectivity for the fixed-gear fleet***. I have the same sentiment for the bottom trawl survey, where dome-shaped selectivity might be considered, which may help explain the need for such a strong prior on q and previous unreasonable estimates (model 19.0b in 2021, M = 0.4, with observer maturity curve) of q when it was freely estimated. It has been documented that mis-specified selectivity forms can lead to unreasonable estimates of catchability, given that these two variables are related (Cadrin et al., 2016).***

##### Recommendations on Model Structure

Some recommendations I have on model structure are detailed below:

1. In general, I think model 23.2 is a good starting point for next years assessment, although I would make the following changes – remove the time-block on growth given the similarities in kappa. I think the *M* time-block is well justified in this case, and does appear to help model performance significantly,
2. I would consider aggregating the longline and pot fleets, while separating out the trawl fleet for fleet structure. Given the pot fleet has limited data, this approach would allow us to increase the information content available to use a semi-fleet disaggregate approach. This is also most likely appropriate because the longline and pot fleets operate at very similar times of the year and previous years assessments where a fleet-specific model was employed, has shown that longline and pot selectivity are estimated fairly similar (composition data also indicate this),
3. In a similar vein, investigations should be attempted to allow for dome-shaped selectivity in the fixed-gears because of differences in size-composition relative to the trawl fleet (see Figure 10 in the 2022 assessment),
4. Similar in the survey, there does seem to be potential for dome-shaped selectivity especially when we consider that seasonal movements can lead to dome-shaped selectivity patterns – domed selectivity can manifest because of availability differences (Cadrin *et al.*, 2016; O’Boyle *et al.*, 2016). The reason that such a strong catchability prior is needed may be due to a mis-specified survey selectivity form and may help explain why models brought forward in 2021 estimated lower catchabilities for the survey. However, I do note that the length-composition data are generally well-fit, although the average length compositions for the survey towards those older size-bins are descending much faster than the fishery. Fits to CAAL data are not provided and ghost fits to age-composition are not provided, making it difficult to evaluate the adequacy of logistic selectivity for the survey (best practice is to use OSA residuals, so I refrained from interpreting the Pearson residual plots).
5. Further expanding on point 4, while the authors indicate that there is not direct evidence for dome-shaped selectivity in the survey using Weinberg et al. 2016, this study investigate such dynamics by increasing vessel tow speed and escape from nets. However, this study did not look into the temporal aspect of availability, especially considering the seasonal migrations Pacific cod undertake (i.e., stock assessment selectivity is both contact and availability),
6. Next, I recommend imposing priors on estimating natural mortality. It seems dubious that it is so well estimated especially given that composition data are not directly available after a small amount of exploitation from the fishery – which has been hypothesized to allow for adequate estimation of natural mortality (Lee *et al.*, 2011). The reason that *M* is so well estimated could be due to the strong prior imposed on catchability, and might be supported by the fact that the estimates for *M* change a decent bit when *q* is analytically determined,
7. I further recommend likelihood profiles on *M* to understand how it changes and how different data sources inform this estimate – it could be that the likelihood profile is super shallow and that the estimate of *M* is actually stuck at a local-minima. A jitter analysis would be incredibly helpful.

# Dan Goethel (Stock Structure)

## Pacific Cod Stock Structure

* Wide-ranging all the way from Japan to California,
* Isolation-by-distance mechanism in stock structure (i.e., further populations are more genetically distinct),
* Strong natal homing instincts back into spawning areas, which could be a dominant driver of genetic differentiation,
* Local adaptation can also be driving differences in stock structure, as a result of different environments (genes related to vision as well as ZP3 – protein coating for embryo),
* Most likely potential for stock mixing and gene flow, although there is a migration-selection balance, where selection pressures are more influential than gene flow,
* However, there is also evidence of landscape influencing stock structure, wherein deep passes along the Aleutian Chain (Amitchka Pass, Samalga Pass) might prevent movement of adult cod since they prefer depths of around 260m. Additionally, currents going into the Bering Sea might be causing genetic differentiation in the Unimak and Pribilof Islands – could also be related to difference in spawn timing,
* In general, the stock-structure in the eastern Pacific ocean is thought to be: 1) Bering Sea, and within the Bering Sea, there is potential for certain subpopulations, 2) Aleutian Islands, with potential of subpopulation structure – although some of the genetic samples looking at the ZP3 haplotype suggest that some of the western GOA (Shumigan and Kodiak Islands – more ZP3 related genes for antifreeze, and likely have a narrower thermal egg range, potentially…) are more similar to some of the AI stocks, 3) GOA, which are thought to be separated by a western and eastern GOA stock (eastern GOA after heatwave was stable, but CGOA and WGOA were more impacted potentially related to differences in thermal tolerance in the ZP3 haplotype).

Pacific cod are wide ranging and are found from California to Norton Sound, as well as from the Gulf of Anadyr to the Yellow Sea. Their distribution is generally limited from 34N to 65N. In the BSAI region, they are widely distributed across the EBS, although in recent years, there has been evidence of a northward movement into the NBS. Stock structure within Pacific cod are generally thought to be a result of an isolation-by-distance mechanisms, where more genetic differentiation occurs the farther apart subpopulations occur. Tagging studies have found that they migrate both within and between the EBS, AI, and GOA, suggesting potential for stock mixing (from straying, despite natal homing). Mixing between the EBS and AI stocks likely occur because genetic differentiation between these stocks only occur in a few regions (islands of divergence). This is generally consistent with the framework that gene flow does occur between populations and primarily seeks to homogenize allele frequencies, except for regions that are strongly impacted by selection processes (Spies *et al.*, 2022).

Nonetheless, local adaptation between the EBS and AI stocks does appear to be driving stock structure, likely due to a variety of factors. Considering that most of the genetic differentiation appears to be related to genes regulating vision, this structure is likely due to the dynamics of the respective regions the stock primarily occupies. For example, EBS exhibits lower light penetration while the AI exhibits higher light penetration. Furthermore, the EBS is largely influenced by sea-ice and wind dynamics whereas the AI is more influenced by current dynamics (from the AK Coastal current and Alaska stream). Studies have found that differentiation in some loci are correlated with differences in velocity and salinity, providing support for some of these hypotheses (Spies *et al.*, 2022). Other factors that could be determining stock structure in face of local adaptation are the differences in species composition between these regions. For example, Pacific cod predominately feed on pollock in the EBS. By contrast, Pacific cod in the AI predominately feed on Atka mackerel. The genetic differentiation in vision-related genes could be a result of local adaptation to different prey species.

In general, Pacific cod are likely to perform natal homing during the spawning months (Jan – Apr) but migrate around to feed outside of that duration.

Genetics also indicate that the EBS and AI are discrete spawning stocks. Using SNPs, a high assignment success was found in assigning the origin of spawning populations (Gulf of Alaska, Hecate Strait, Kodiak Island, and Prince William Sound populations). Some studies have also found that the ZP3 locus have distinct sets of haplotypes (spawning populations from Kodiak vs. Prince William Sound). There seems to be some differentiation there between the Bering Sea group and other groups in the Gulf of Alaska, Hecate Strait, and Prince William Sound. There might be some selection processes on these northern (Bering Sea) populations wherein local adaptation is driving differences in haplotype frequencies. Building upon this, the population structure of Pacific cod appears to exhibit isolation-by-distance patterns, where samples in closer areas of spawning are more similar to those in distant areas. This pattern was observed between the western GOA through Unimak Pass to the eastern Al. There also appears to be a significant break between western and eastern GOA samples, and the AI samples appear to have some genomic regions that are highly divergent (local adaptation), suggesting the appropriateness of separating out the EBS and AI stock as distinct management units.

## Stock Identification Methods