# Abstract

# Introduction

* Lack of documented case-studies extending panmictic assessment models to spatially explicit
* Why Sablefish – mobile species, great data, a lot of tagging data, previous studies to lean on but none have integrated tagging data with the assessment data
* Overview of the approach which is to describe our process/decision points, including exploratory analysis, model comparison techniques, goodness of fit methods for developing spatially explicit model for Sablefish

Continued emphasis on developing spatial stock assessment approaches has culminated in the recent literature including best practice review papers (Punt 2019a, 2019b, Goethel et al 2023), international workshops (space oddity and CAPAM) and simulation studies (Goethel xx) , which highlight the benefits and pitfalls of spatially explicit stock assessment models. However, in practice, development of spatially explicit stock assessment models is complex, requiring identification of primary spatial dynamics (Cadrin et al. 2023) and appropriate assumptions. Given the dearth of spatial assessment applications (Punt 2019), this paper aims to document and disseminate the model development process in the hope of highlighting lessons learned to aid future spatial applications.

We documented the development of a spatially explicit tag integrated age structured assessment model for Alaskan Sablefish (*Anoplopoma fimbria*), covering the Gulf of Alaska, Bering Sea, and Aleutian Islands regions (Goethel et al. 2023). This stock provides a unique case study to illustrate the spatial modeling process, due to its vast mobility, economic importance, and a long-term dedicated longline survey and over 40 years of tag releases and recaptures. Currently a single, panmictic population is modeled with quotas apportioned to management areas based on area-specific survey biomass (SAFE reference). However, significant spatial heterogeneity exists in the population distribution resulting from multiple hypothesized factors such as age-based habitat preferences and migration patterns.

Many spatial tagging models have been developed and or included the Alaskan Sablefish stock (Hanselman et al 2015, Heifetz and Fujioka 1991, Bracken 1982, Kimura & Shavy 1998) but none have attempted to integrate the tagging data with the assessment data and the assessment population structural assumptions in a full integrated assessment model. Computational power limitations previously hindered fully integrated analyses in past studies, but recent advancements in computers and efficient statistical software like TMB and Stan (references) have alleviated these barriers.

It is common for spatially explicit population model papers to skip the detail on how the spatial resolution was chosen or lean on default management boundaries and or previous research (**help with references, nervous on shitting on other peoples work**). However, as more and more spatially granular data becomes available, there is a need for continued scrutiny of spatial boundaries and assumptions. We discuss the exploratory approaches employed in a ground up approach to build our spatially explicit model.

**Probably need to link in Kari’s and Dana’s work in a little more detail somewhere…**

The purpose of this paper is to highlight decision points and model considerations when transitioning a panmictic model towards a spatially explicit model, using Alaskan Sablefish as the case study.

# Methods

## Literature review

Expanding panmictic stock assessments to include spatial regions will have existing assessment models and supporting material present in the grey literature. We found this to be the best initial resource for understanding what has been explored in addition to providing insight into the most important process dynamics for the stock.

The primary literature was the next step for providing *a priori* evidence for spatial delineations and spatial structure. Northwest pacific sablefish is fortunate to be the focus of genetics studies (Jasonowicz et al. 2017, **One more?**), morphometric studies (M. Kapur et al. 2020) and tagging studies (Hanselman et al 2015, Tripp-Valdez et al. 2012, Kimura and Shavy 1998, Heifetz and Fujioka 1991, Bracken 1982). However, we recognize these resources are not available for many stocks globally. In these cases, environmental or ecosystem meta studies could be used as proxies for spatial distribution factors and delineation (Planque et al. 2011, Landa et al. 2011).

## Data

Sablefish is also fortunate to have a dedicated fishery independent survey that provides a long-time series of highly spatial length, age and abundance data (Kimura and Zenger Jr 1997). Since 1978, the Alaskan sablefish survey has been conducted during the summer using a systematic design, offering increased flexibility for generating model inputs at different spatial resolutions (e.g., three-region vs. five-region models). This flexibility arises from the uniform sampling intensity along the coast, distinguishing it from commonly used stratified survey designs. Spatially stratified surveys typically optimize variance reduction across space and can serve as initial spatial boundaries for a spatial assessment model. However, most surveys have multiple objectives and consider ecosystem factors rather than being solely optimized for a single species.

We found flexible estimators, such as model-based estimators (VAST Thorson 2015 and sdmTMB Anderson et al. 2022)) were advantageous over design-based estimators due to their ability to generate assessment inputs such as length frequencies and abundance indices for various spatial resolution assumptions. This flexibility enables modelers to easily configure and test alternative spatial structures, which is a large source of uncertainty in spatially explicit age-structured models. However, applying model-based estimators to Alaskan sablefish was challenging, due to their preference to occupy narrow area on the slope but have long north-south spatial extent along the slope from Canada to the Bering Sea. This spatial extent complicated the use of spatial model-based estimators, as they required fine spatial resolution to capture abundance changes across the slope resulting in sluggish and impractical use.

Table : Input data sets and recorded spatial resolution.

|  |  |
| --- | --- |
| Data Set | Spatial Resolution |
| Catch pre-1979 | Available at stock resolution by gear |
| Catch post-1978 | Available at Fishing Management Plan (FMP) boundaries by gear (Figure 1) |
| Observer data (age, length and catch) | Latitude and Longitude positions |
| Survey data (age, length and catch) | Latitude and Longitude positions |
| Tagging data | Latitude and Longitude for releases, Latitude and longitude for approximately (get %) recoveries |

Reporting all data sets and the finest spatial resolution available is critical when considering how granular the assessment model can be. The dataset with the coarsest spatial resolution can set an upper limit for the number of regions considered during model development (Cadrin 2020). The limiting data set for sablefish was early catch history prior to 1979 (Table 1), which was only available at the stock level (Fenske 2023).

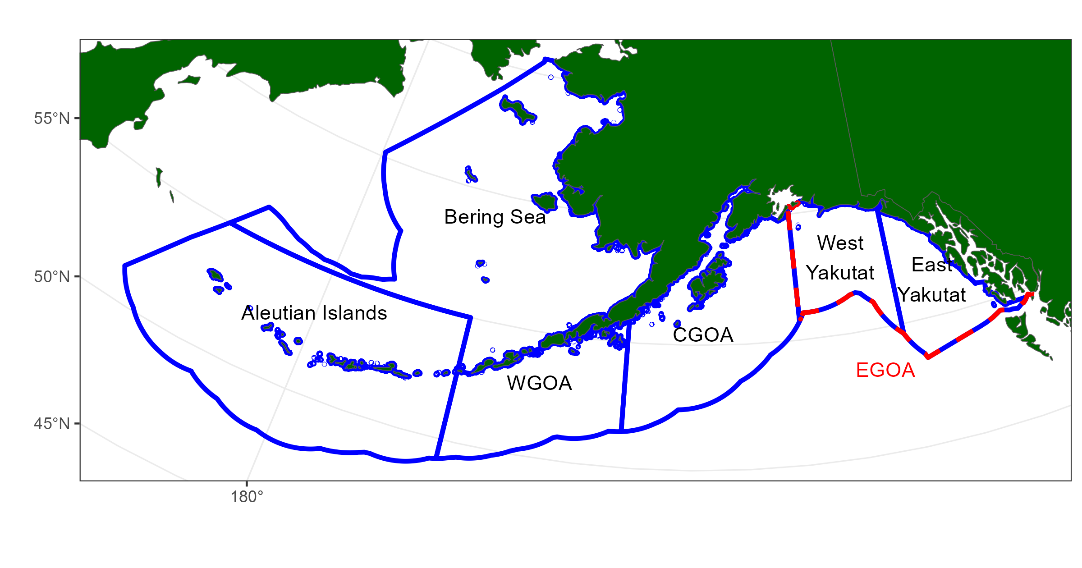


Figure 1: Fishery management plan (FMP) boundaries. The Eastern Gulf is often reported at subareas split up be East/West Yakutat and sometimes Southeast.

## Exploratory methods used for spatial boundary delineation

An approach utilized in this study to inform spatial boundaries was that described by Lennert-Cody et al. (2010, 2013). This analyzed length frequencies from both fishery and survey data sets at the one-degree latitude and longitude spatial resolution. This method uses a regression tree to identify clusters based on seasonal, latitudinal and longitudinal breaks as covariates. This approach identified three longitudinal breaks, two of which fell in the middle of an FMP boundary (Figure 2). The split in the Aleutian Island FMP was ignored due to the lack of catch and data available to the west of that split point. However, the longitudinal splits in the central Gulf region, did correspond with an ecosystem-based break identified by Kapur et al. 2020. Due to the spatial resolution of catch, an apportionment assumption would be required to allocate the Central Gulf catch to either side of this spatial boundary. A sensitivity simulation was thought necessary to identify if splitting the Central Gulf region at this break point would impact management advice, as of yet this sensitivity has not been explored.

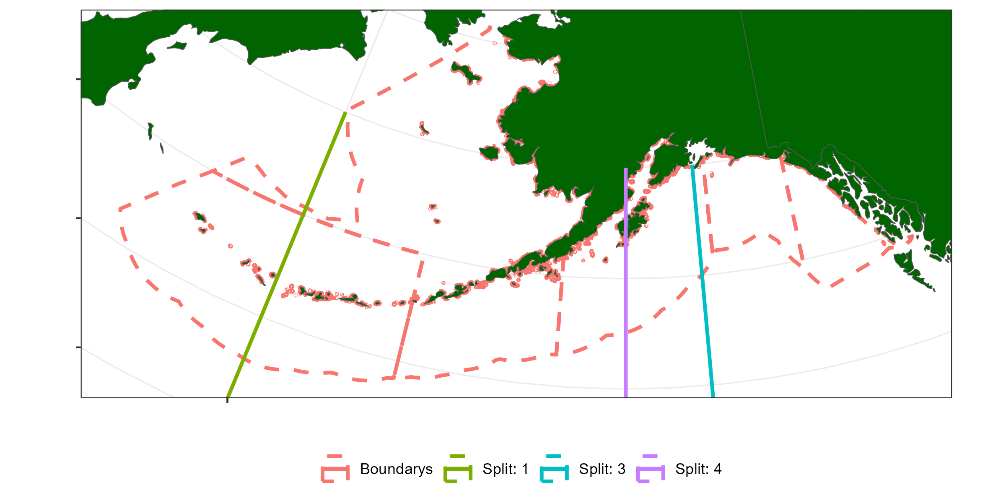


Figure 2: Longitude splits based on regression tree analysis.

Age-length pairs were visually plotted by FMP regions for each sex to inspect growth rate variation among regions. A more objective approach at identifying growth delineation is described by (Kapur et al 2020). Incorporating spatially specific growth in an age-based model would require modelling numbers by age and length (or length group), thus adding another dimension to the partition (numbers at age array). This would have a significant computational cost in a tag-integrated spatially explicit model, from the large number of dimensions when tagged fish are considered. Due to this consideration, we decided that very clear spatial growth patterns would be required to justify the inclusion of spatially varying growth which is why we chose to use visual plots (Figure 3). **Probably need to re-word this**

|  |  |
| --- | --- |
|  |  |

Figure : Age-length pairs from the survey for males (left panel) and females (right panel) amongst each FMP region.

Fishing effort distribution plots were used to provide insight into the possibility of changing fishing selectivities over space and time and help inform fleet assumptions in the base model. The fishing effort variables available for our case study were depth, duration, latitude and longitude (See appendix Figure A 1 – Figure A 4). These show a shift from hook and line (HAL) to pot fishing (POT), but no systematic shifts that would highlight the need to incorporate spatially varying selectivities.

In addition to the above explorations, we explored tag-recovery matrices to understand the mixing of fish between FMP regions (Figure 3). This analysis aggregates all recoveries by release region and size group and ignores spatial-temporal recovery rates as well as time-at-liberty. The aim was to get a general understanding of source and sink dynamics.

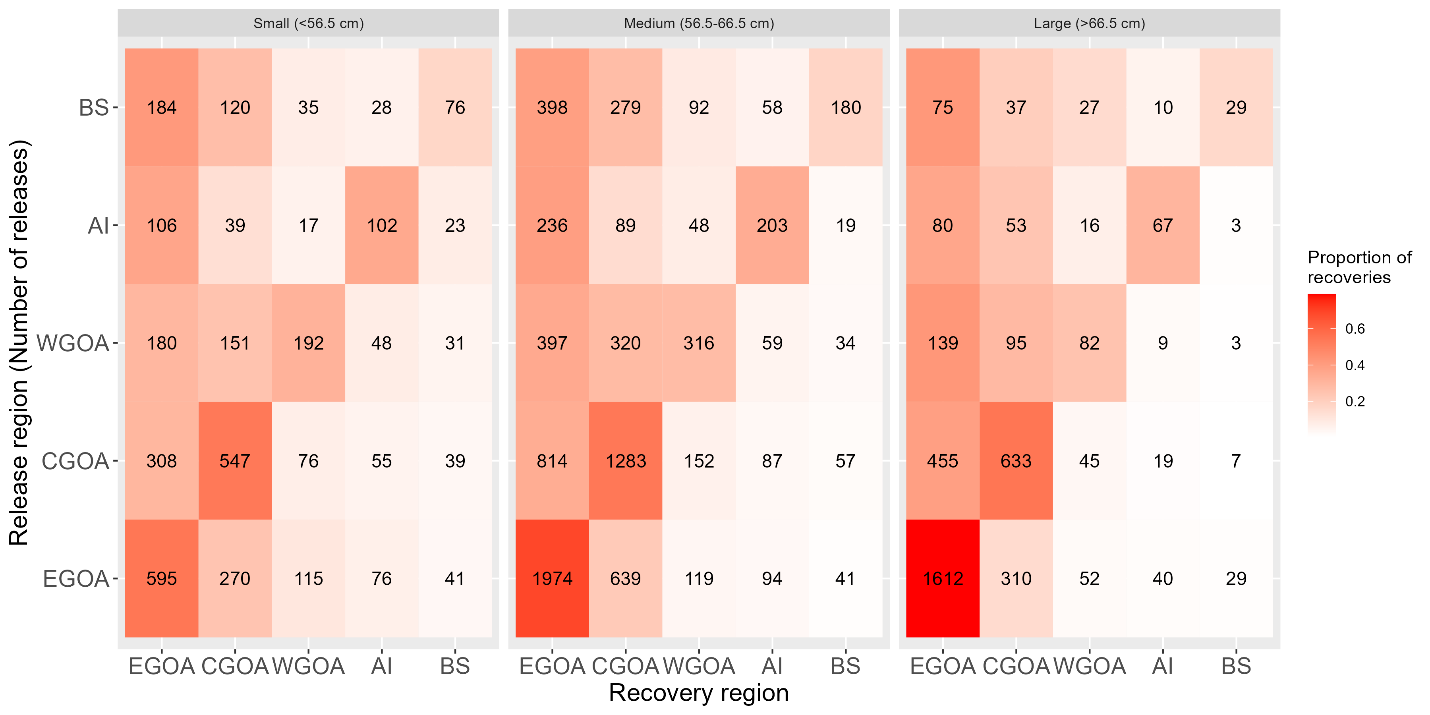


Figure : Tag-recoveries by FMP region and size group for all tags recovered.

Additional exploratory analysis available but not used in this study for informing spatial boundaries include; hierarchical clustering analyses from spatiotemporal models (Gruss et al. 2023), otolith chemistry (Tanner et al. 2016), and morphometric analysis (Kapur et al 2020).

Given the recorded spatial resolution of historical catch was at the FMP region and the length frequency analysis only had one split in the middle of an FMP region, we decided to develop a five-area model with the Eastern Gulf region consisting of both East/West Yakutat (Figure 1), due to low data availability. In the Western regions We also developed a three-area model which aggregated Bering Sea, Aleutian Islands and Western Gulf region similar to that developed by Fenske (2023) as well as a single panmictic model.

There was an *a priori* preference in the observation sub-model for distributions that had estimable dispersion parameters such as the negative binomial vs the Poisson and Dirichlet-Multinomial vs Multinomial (Thorson 2018). These distributions would remove the need to do any iterative data-weighting commonly conducted in stock assessments.

## Initial model structure

A generalized spatial model was developed in TMB for this project, a full description of the model is given in Appendix (**Currently in a standalone word document to reduce clutter**). The model allowed for a general number of spatial regions, input observations and movement/tagging assumptions, and is hereby referred to throughout this paper as “the model”.

An initial model structure was used for each spatial model configuration (1-area, 3-area & 5-area). From the initial model, we then explored a range of model assumptions and used model selection criteria to identify a suite of candidate models.

An overview of the core process dynamics are given in Table 2. Observations available to each spatial model are given in Appendix B, the initial model assumed composition data were multinomially distributed with bootstrapped standard errors converted to an equivalent input sample size and abundance and catch data were lognormally distributed with design based standard errors.

Table 2: Initial model process dynamic assumptions.

|  |  |
| --- | --- |
| Process dynamic | Assumptions |
| Recruitment | No stock recruitment relationship,  Regional specific and annual recruitment deviations |
| Fishing Mortality | Fishery gear specific selectivity constant across regions. Annual regional fishing mortalities solved using a Newton Raphson algorithm |
| Movement | Markovian movement, which is age & time-invariant |
| Natural mortality | Not estimated, age and time invariant |
| Tag release | See next section |

### Tag releases and recoveries

Tagged fish are released by the summer survey, which records the length of each fish released. There are two large benefits of having the survey release tags, the first being the presence of an age-length transition matrix. This is used to convert unsexed lengths frequencies of tagged fish at release to sex and age disaggregated releases which are compatible with the sex and age structured estimation model. The second, is the spatial coverage of releases, which results in a balanced spatial design.

Tagged fish are in the partition are assumed to have the same ageing, mortality and growth assumptions as the untagged members of the partition. When tagged fish are released into the partition, they are indexed by the release event index denoted by which is region and year specific (). Tag induced mortality and initial tag-loss is applied as an initial mortality rate denoted by ,

Ongoing annual tag-shedding denoted by is also applied as a mortality process whereby tagged fish are deleted from the partition as

Values for and were taken from the tagging study by Beamish and McFarlane (1988).

The final model consideration for tag-releases was for how many years to track the tagged cohorts within the model partition, also known as time-at-liberty denoted by . Figure 5 shows the skewed distribution of time-at-liberty for all tag-recoveries, which highlights sablefish’s ability to remain at liberty for long periods of time. We chose to use for the three-area model and for the five-area model, which was a compromise between computation time and potential movement information. These values were larger than the 9 years at liberty considered by Beamish and McFarlane (1988). The spatial model used a tag-release conditioned likelihood (Vincent, McGarvey) this means the model tracks numbers at age for each region and tag-release cohort. For a five-area model () without tagging and max age of 30 the partition has 150 elements (5 30). When tagging is included and , the partition has 7650 elements (150 + 30 5 ).

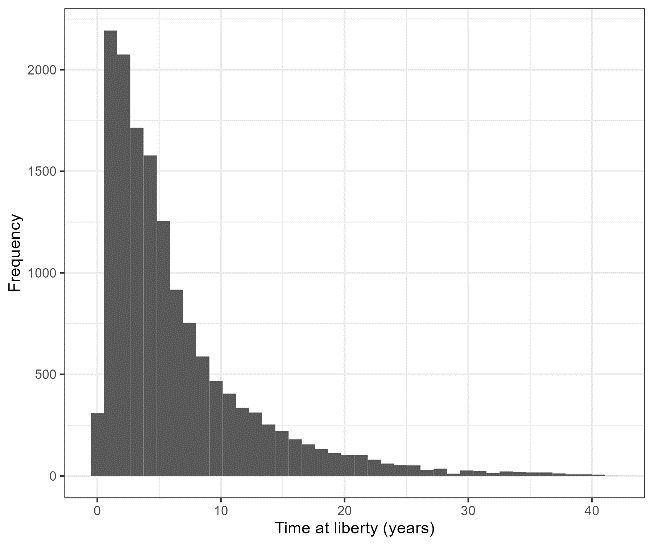


Figure 5: Time at liberty from all tag-recoveries.

Two assumptions regarding tag-recoveries were investigated which included the mixing periods and tag-recovery method.

* mixing period - Figure 6 displays the relative distribution of releases and recoveries by release year and years-at-liberty (normalized to have a max value of 1 for each year at liberty). This shows recovered fish that have been at liberty between 0 to 1 years have quite a different distribution to fish that were at liberty for 2+ years. We interpreted this as mixing signal, where by fish at liberty for one year were more likely to be captured than fish from 2+ years. For this reason, we chose to have a mixing period of two years before considering tag-recovery observations. This approach differed from those discussed in Kolody and Hoyle (2015), firstly it was qualitative in nature and secondly, we focused on the temporal distribution of tag recoveries by release events. The fishery dependent catch per unit effort data was not a the spatially granularity to test for gradients in recoveries from release events as proposed by Kolody and Hoyle (2015).
* Tag recovery was fishery dependent, choice of gear type – Most of the returns are from the fixed gear fishery (Longline & Trap/Pot). One trend in recent years was a dramatic drop off in tag-recoveries. Due to this trend, we did not consider recoveries after 2016 (Figure 7).

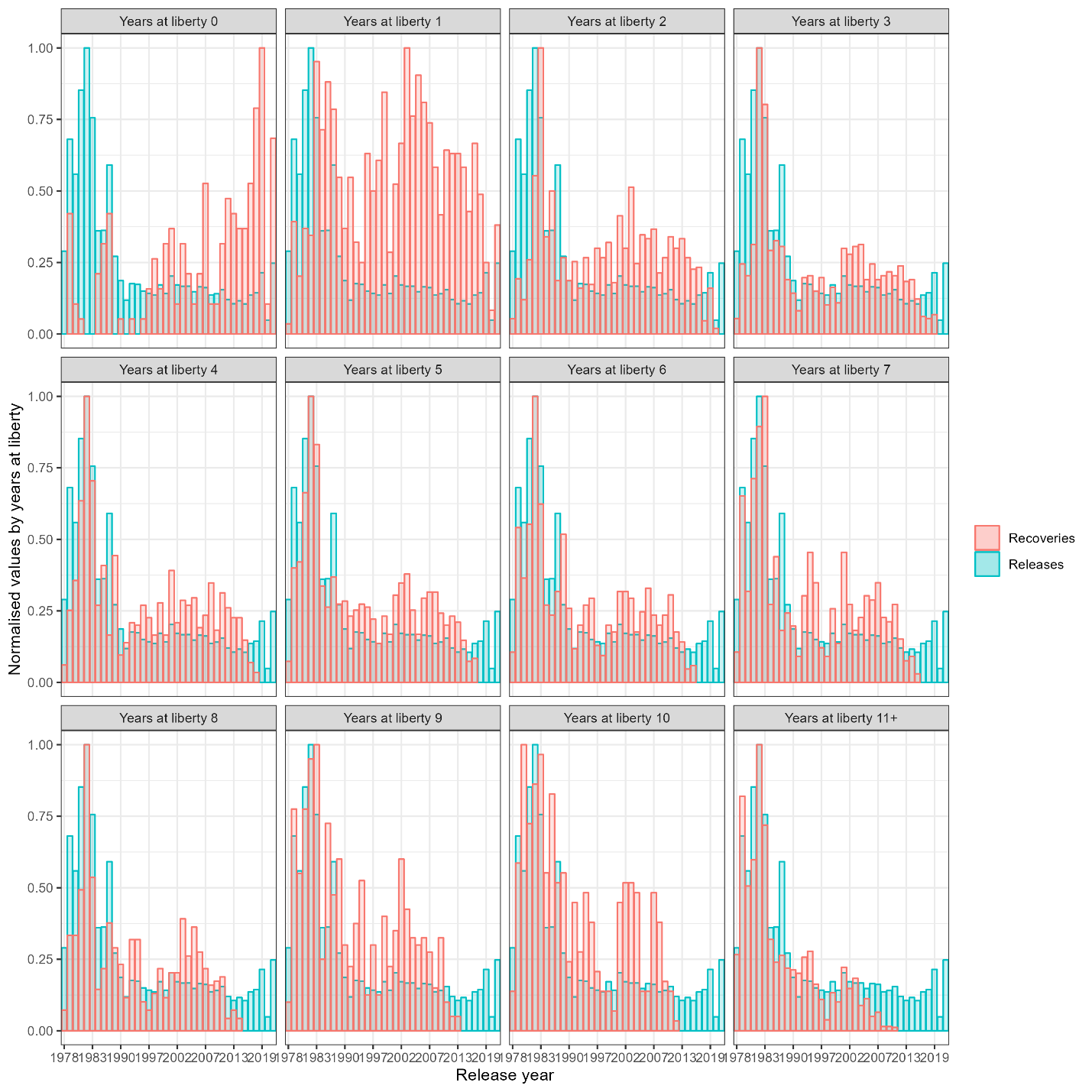


Figure 6: Relative distribution of tag recoveries and tag releases, by release year (x-axis) and time-at liberty (panels).

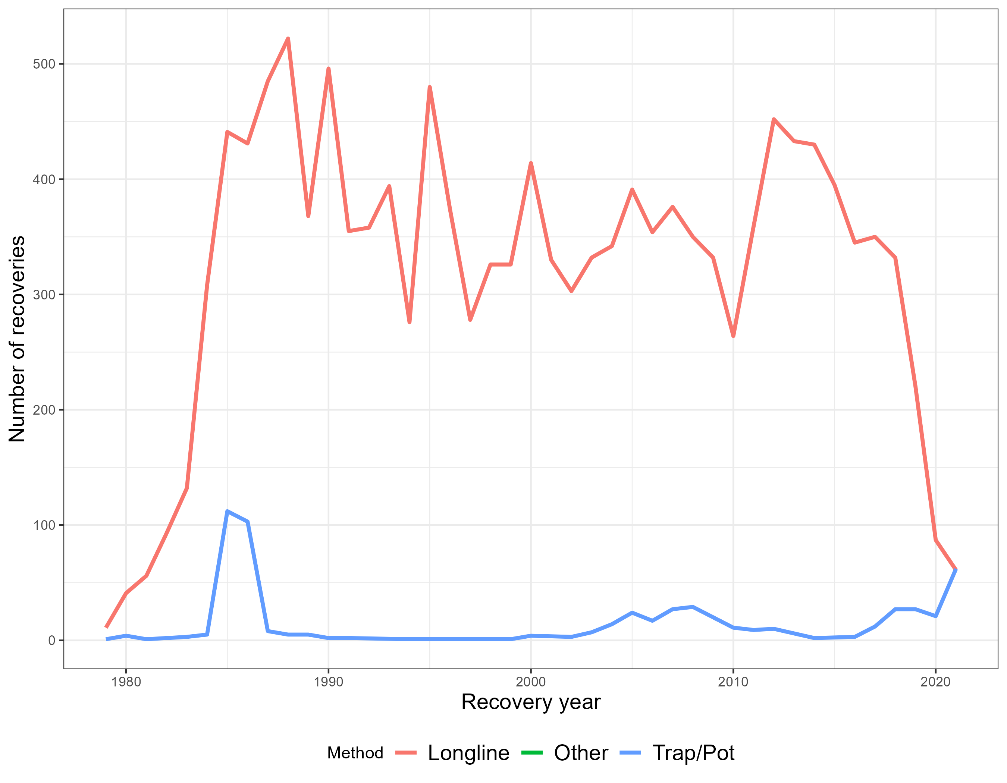


Figure 7: Tag recoveries by gear type. The fixed gear type includes both Longline and Trap/Pot.

## Model assumptions explored from initial model & model selection criteria

A range of alternative model assumptions were explored from the initial model structure for each spatial structure. These included,

* Model start year (equilibrium vs non-equilibrium conditions)
* Age-varying movement
* Time-varying selectivities
* Selectivity parameterizations
* Spatially-varying catchabilities
* Time-varying reporting rate
* Inclusion of the tag-recovery data
* Tag-likelihood (Poisson vs Negative Binomial)
* Compositional likelihood (Multinomial vs Dirichlet-Multinomial)

Model convergence was used to omit assumptions that resulted in unsatisfactory inference. A model was deemed not-converged if it exhibited any of the following conditions

1. Parameters estimated at or near boundaries
2. Non positive definite hessian
3. Maximum absolute gradient of any estimated parameter > 0.001

All models that converged were subjected to a “self-test” (Deroba et al, 2015). The self-test assumes the operating model (OM) and estimation model (EM) have identical model assumptions and are employed as an integrity check, to ensure model robustness and help identify coding errors. Any model that could not produce unbiased management quantities was classified as non-converged.

For the models that passed the above convergence criteria had residuals assessed for all data sets. Given the large number of observations we frequently found that alternative model assumptions, would result in a similar total objective score, but different fits across the observations, making model comparisons difficult. For this reason, we prioritized survey abundance index fits assuming the fits to composition and tagging data were satisfactory. This was firstly for convenience in model selection, but also follows the view of Francis 2011 & Francis 2014, where fits to indices of abundance should be prioritized over other data fits when providing catch-based management advice.

# Results

## Key model uncertainties

An additional sensitivity analysis focusing on the observation sub model of each converged model was also conducted. This iterated over each data set and dropped it during estimation whilst keeping all other datasets (except catch which was needed to estimate annual fishing mortalities). Spatial models that estimated movement would not converge when tagging data was omitted. This highlighted the necessity of tagging data to account for diffusion among regions (**Are there papers which show you don’t need tagging data? That either disagree or agree with this finding**).

Perhaps the most sensitive model assumption was the start year of the model. Across all three spatial resolution models we would get similar absolute SSB estimates over the data period but varying estimates of B0 and thus percent B0 depending on if we started the model in 1960, 1977 or 1990. This highlighted a key model sensitivity that was further explored using simulation. The issue with starting the model far away from B0 was estimability of the non-equilibrium parameters. The self-test frequently showed theses parameters were biased. There is a parallel project which is investigating optimal parameterization for spatial models that start in non-equilibrium states.

## Simulations

The following shows the results from the self-test Figures ?? show SSB trajectories from the 1A, 3A and 5A models (move these to appendix, not really that important). The self-test was found to be the criteria which rejected the most model formulations. In particular there were issues with early and later recruitment deviations which lacked data were fixed due to the self-test in the spatial models. Even though the more traditional convergence criteria wanted more sophisticated model the self-test would frequently expose that they could not produce unbiased management estimates.

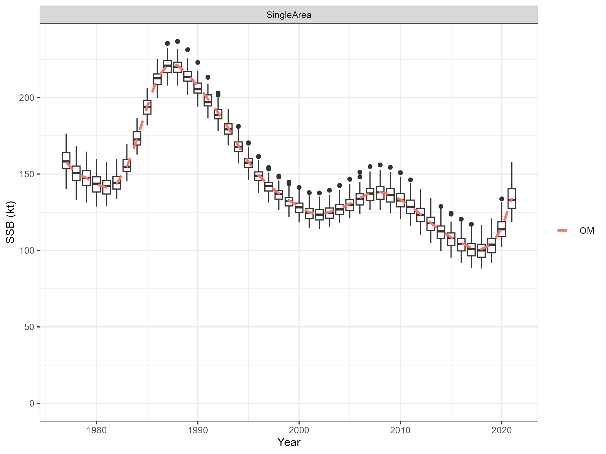


Figure 8: Estimated SSBs from the self-test for the 1-Area model.

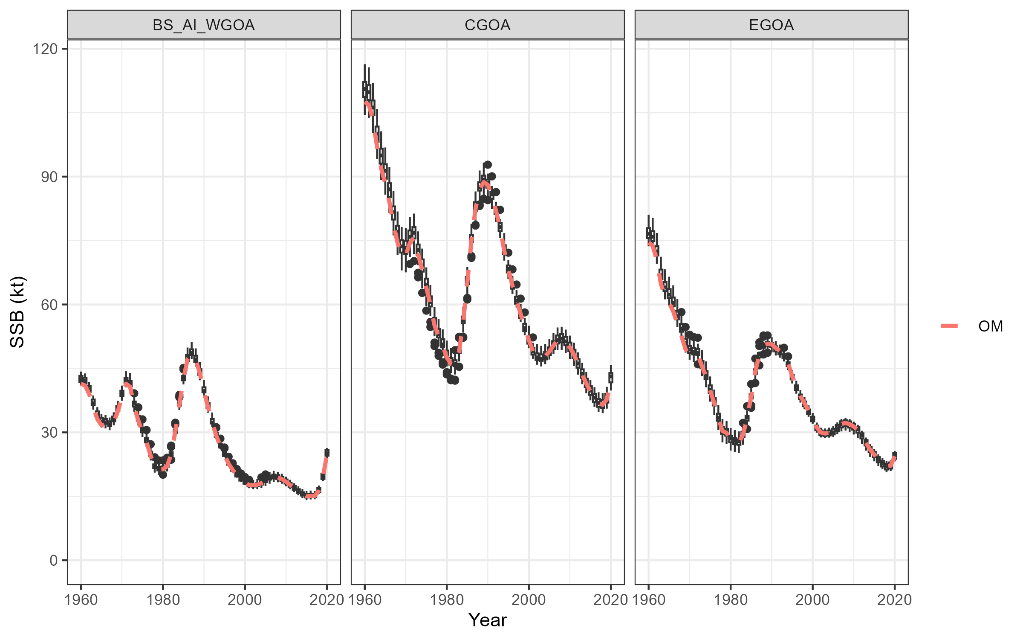


Figure 9:Estimated SSBs from the self-test for the 3-Area model

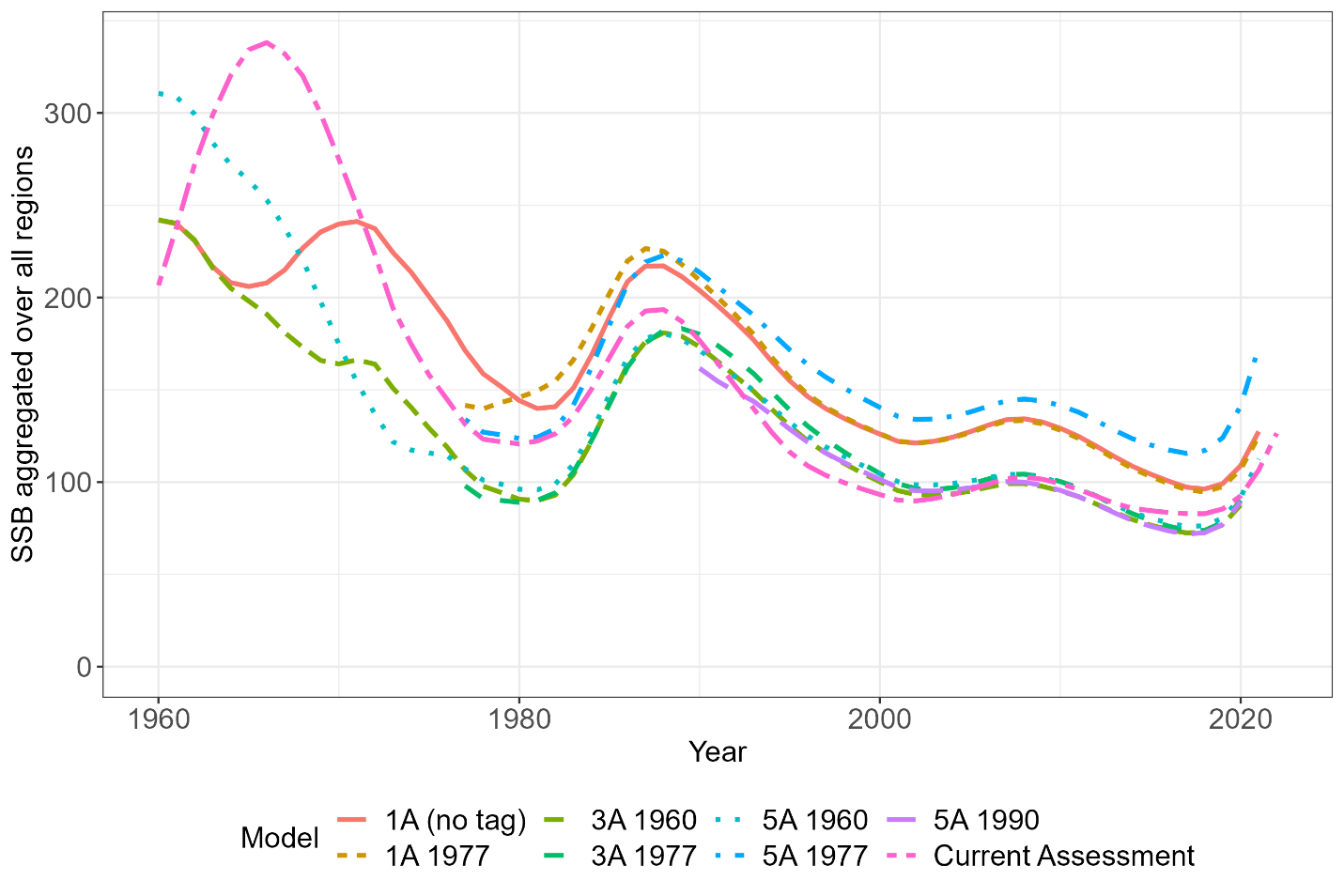


Figure 10: Spatially aggregated SSB comparison of all final models considered.

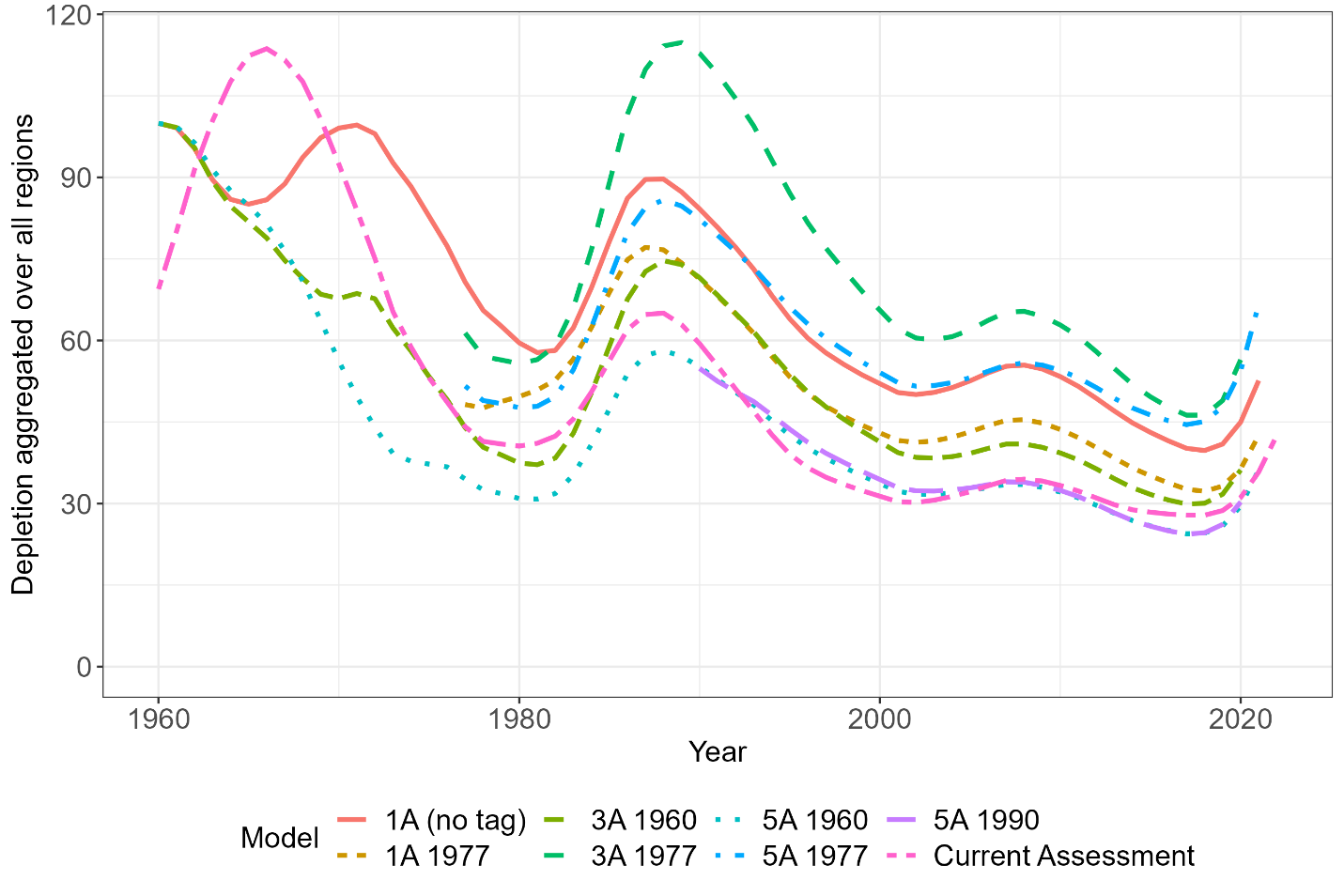


Figure 11: Depletion aggregated across all spatial regions.

Relationship between and , interestingly with the estimated movement from the 3-Area model, it estimated a large for the Aleutian Islan/Bering Sea/Western gulf complex (AI/BS/WGOA), but a small , see Table 3. This highlights a source sink dynamic from AI/BS/WGOA down to the Eastern Gulf of Alaska (EGOA).

Table 3: Key model quantities from the 1977 three area model

|  |  |  |  |
| --- | --- | --- | --- |
|  | (millions) | (kilo tonnes) | Total biomass |
| AI/BS/WGOA | 12.11 | 45.97 |  |
| CGOA | 3.60 | 107.45 |  |
| EGOA | 1.52 | 79.0 |  |

Table 4: Movement matrix from 3-Area model.

|  |  |  |  |
| --- | --- | --- | --- |
|  | AI/BS/WGOA | CGOA | EGOA |
| AI/BS/WGOA | 0.826 | 0.117 | 0.0567 |
| CGOA | 0.0262 | 0.885 | 0.0885 |
| EGOA | 0.0415 | 0.0952 |  |

Model assumptions that had convergence problems included age-varying movement, certain time-varying selectivity models, and specific selectivity parameterizations. In the case of age-based movement the self-test simulation showed that all parameters were identifiable for a conditioned OM. However, understanding why the model did not converge when estimated with actual data was not ascertained during this project, due to the number of possible reasons and time constraints.

# Discussion and next steps

* General summary - This project applied and documented methods and key decision points for developing a spatially explicit age-structured stock assessment model to the Gulf of Alaska Sablefish stock. This resulted in a suite of spatial candidate models that resulted in “similar” estimates of global SSB during the data period (Figure 5). When examining global depletion (Figure 6), there was a much more variability among candidate models due to the uncertainty in estimates of .
* How can you use these results in management - One approach often used when stock assessments produce a suite of models is to use a model ensemble approach to inform management decisions (Ducharme-Barth and Vincent 2022). Model ensembles utilize all models in a weighted framework, which can account for model-structural uncertainty, which was shown in the sablefish case with spatial assumptions resulting in variable estimates of depletion.
* MSE next steps - The next step of this research would be to use a full management strategy evaluation (MSE) to identify the “best” performance spatial resolution model. We used simulations to identify model pathologies and biases using the “self-test”. However, to truly compare management performance a full MSE is required which is the next step.

MSE may also identify alternative stock status reference points other than B0 which we found to be sensitivity to model assumptions such as the start year and conditions i.e., use where is the first year with reliable data (Punt 2003).

* Discussion topics - Data-weighting
* Things I would have changed and suggest others do (other than not doing a postdoc of course) - Change parameterization of (or ) to estimate a total value and apportionment parameters, that way you can profile the total to get an understanding of which data sets were influencing the absolute scale of the model which was sensitive to starting condition assumptions. The model developed, estimated a separate for each region, which made using log-likelihood profiles difficult to diagnose the source of this uncertainty.
* Discuss how this work builds on Kari’s work, movement assumptions. Are results consistent with her findings, is there anything that differs. Spatial EM chapter 2, movement is fixed. Estimate consistent movement.
* Exploratory analysis found possible boundary splits in the middle of Central Gulf that were not practical to implement in the model due to data reporting restrictions, which are often pinned to management regions (circularity).
* How to use this tool?
  + Run it in parallel with the assessment, use it as a “conditioned OM” to test assessment model
  + Use it as a sanity check on regional abundance and catch advise.
* A model assumption that should be explored in future research is the uncertainty from converting unsexed lengths of tag-releases to sex and age disaggregated tag releases which are inputs in the model and assumed known without error. One approach that we wanted to explore consisted bootstrapping the age-length transition matrix and re-running the model with different tag-releases at age and sex from each bootstrapped age-length transition matrix. This is quite computer intensive and should only be conducted on final models.
* Caveats
  + Time-varying movement

# Appendix A

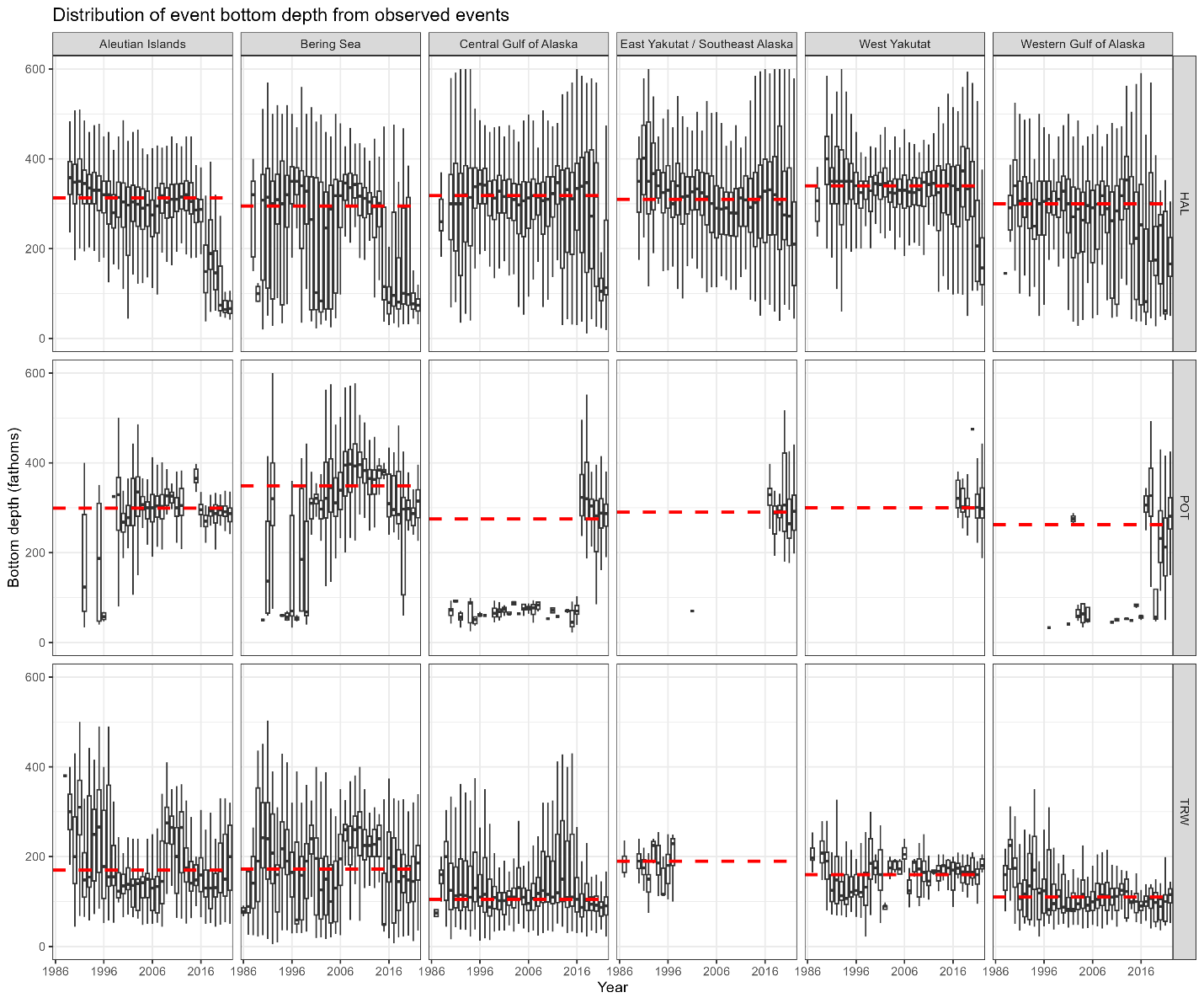


Figure A : bottom depth distributed by year, gear type, and FMP region.

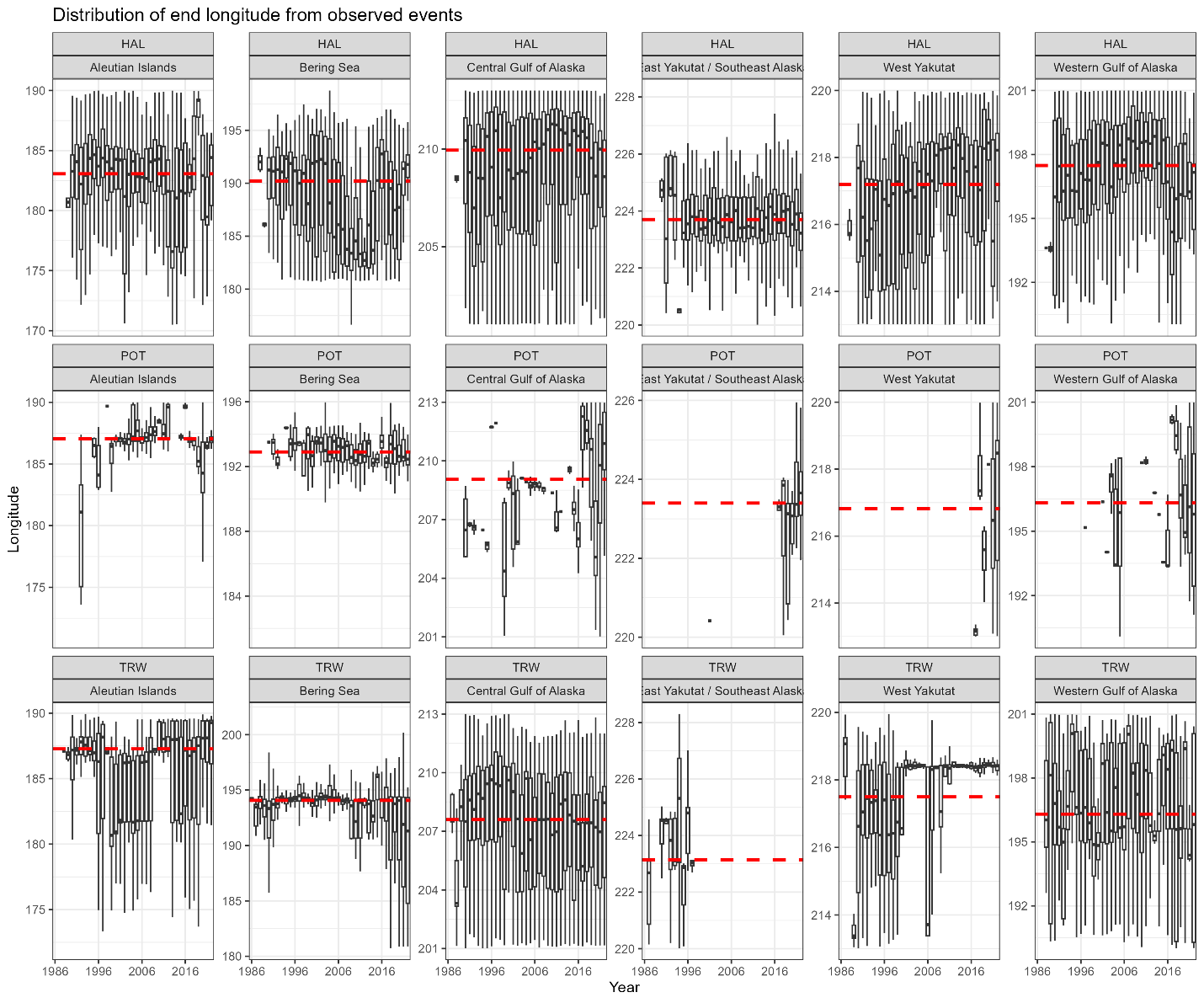


Figure A : Longitude distributed by year, gear type, and FMP region.

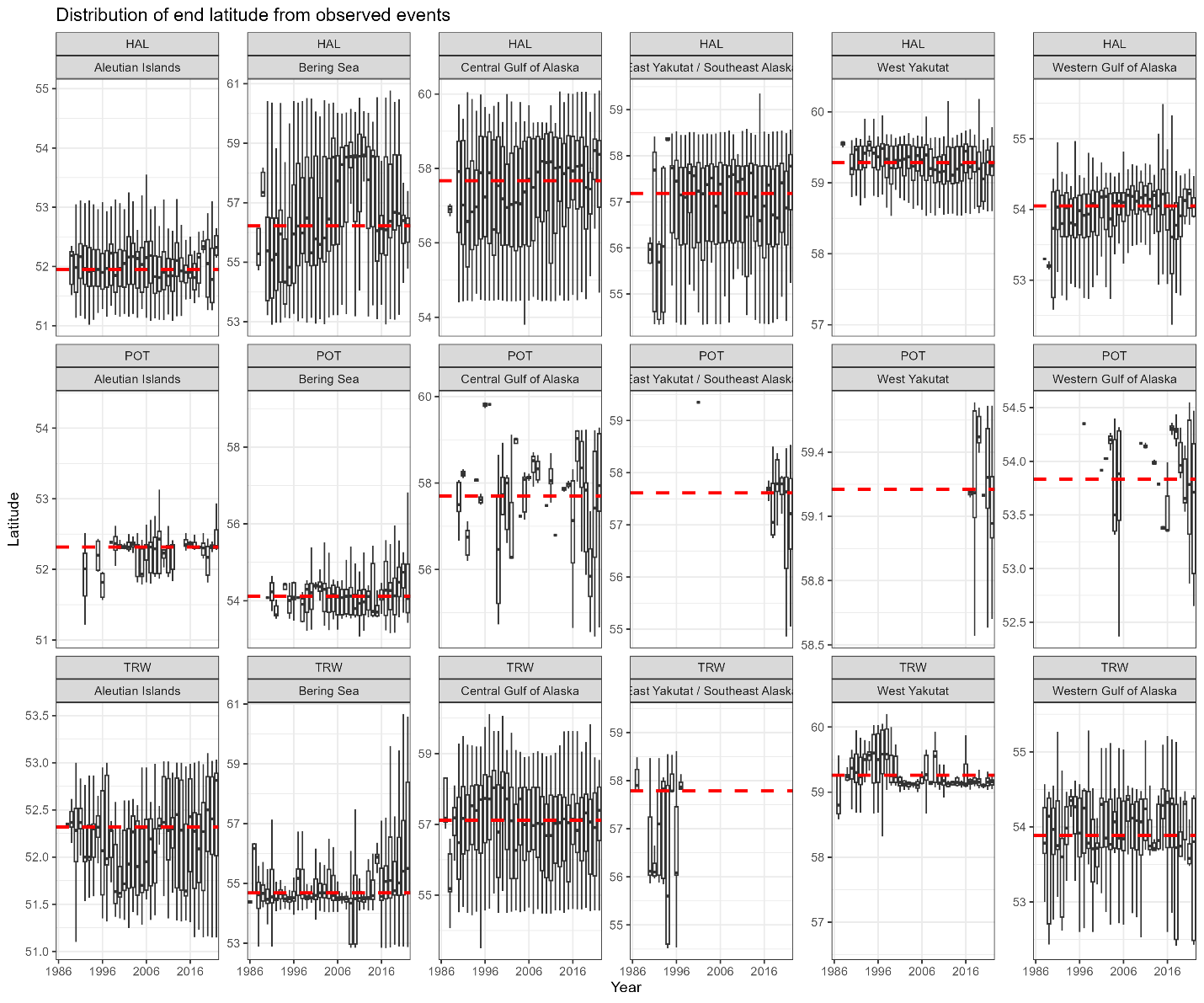


Figure A : Latitude distributed by year, gear type, and FMP region.

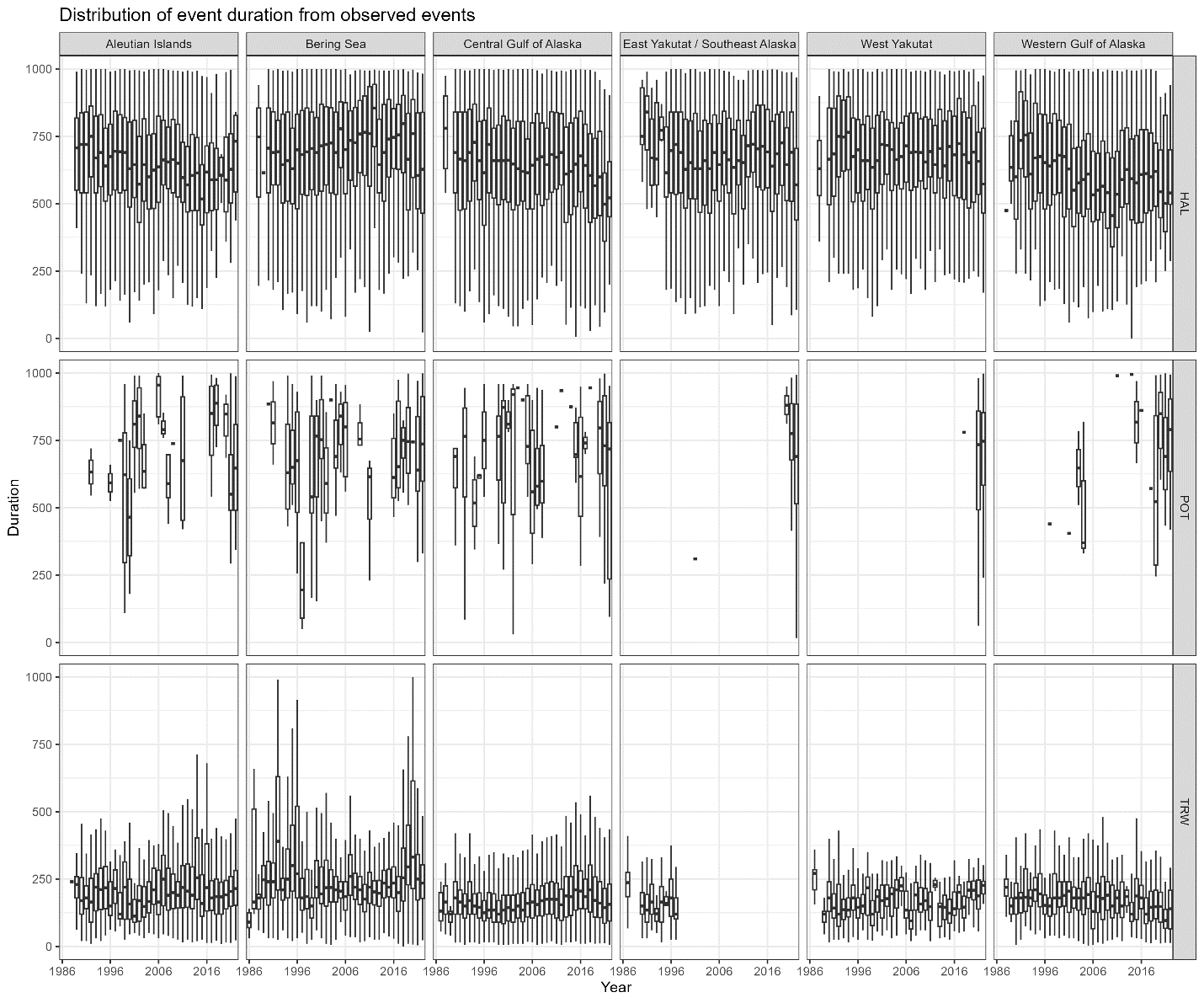


Figure A : Fishing duration distributed by year, gear type, and FMP region.

# Appendix B

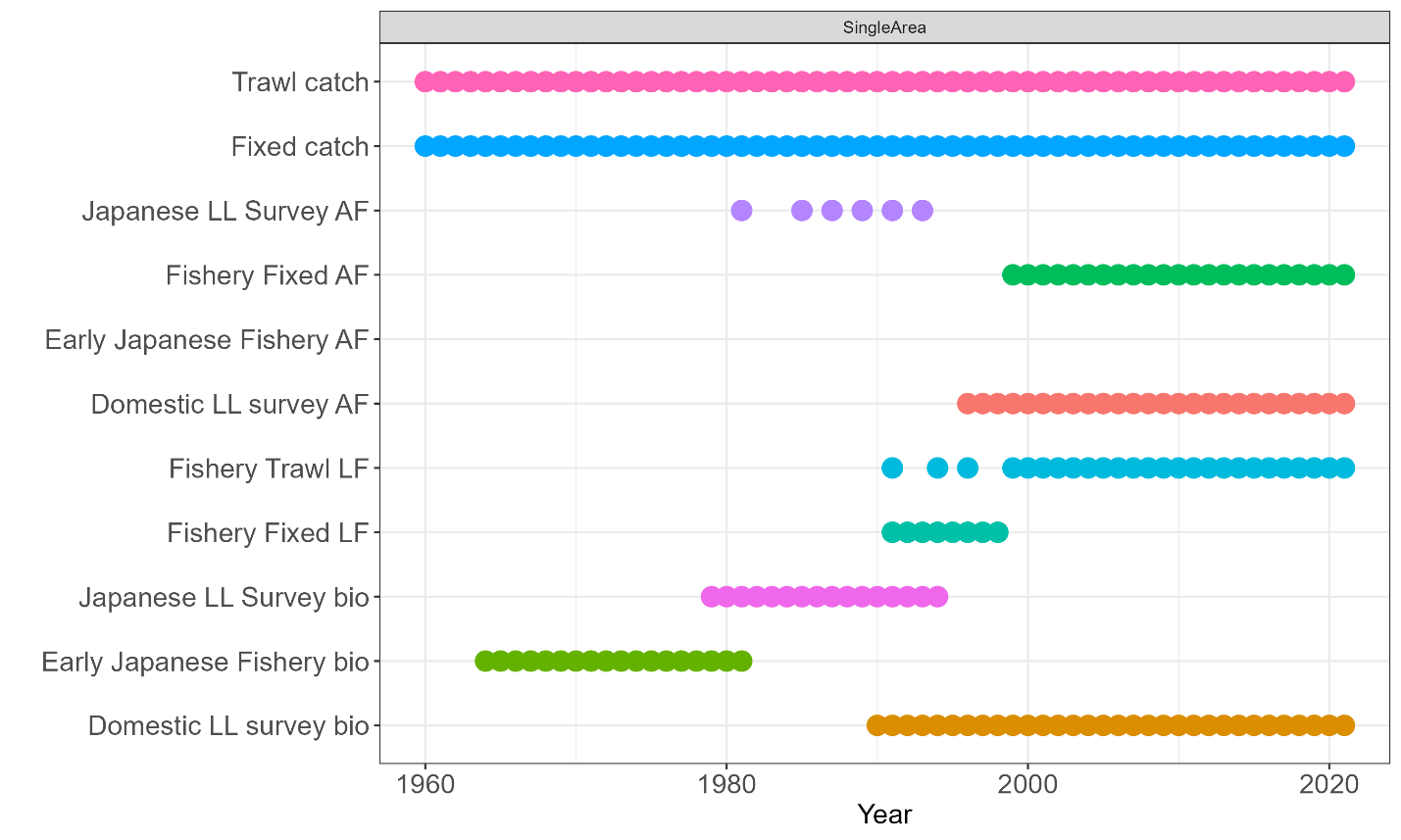


Figure :Observation frequency for 1-Area model.

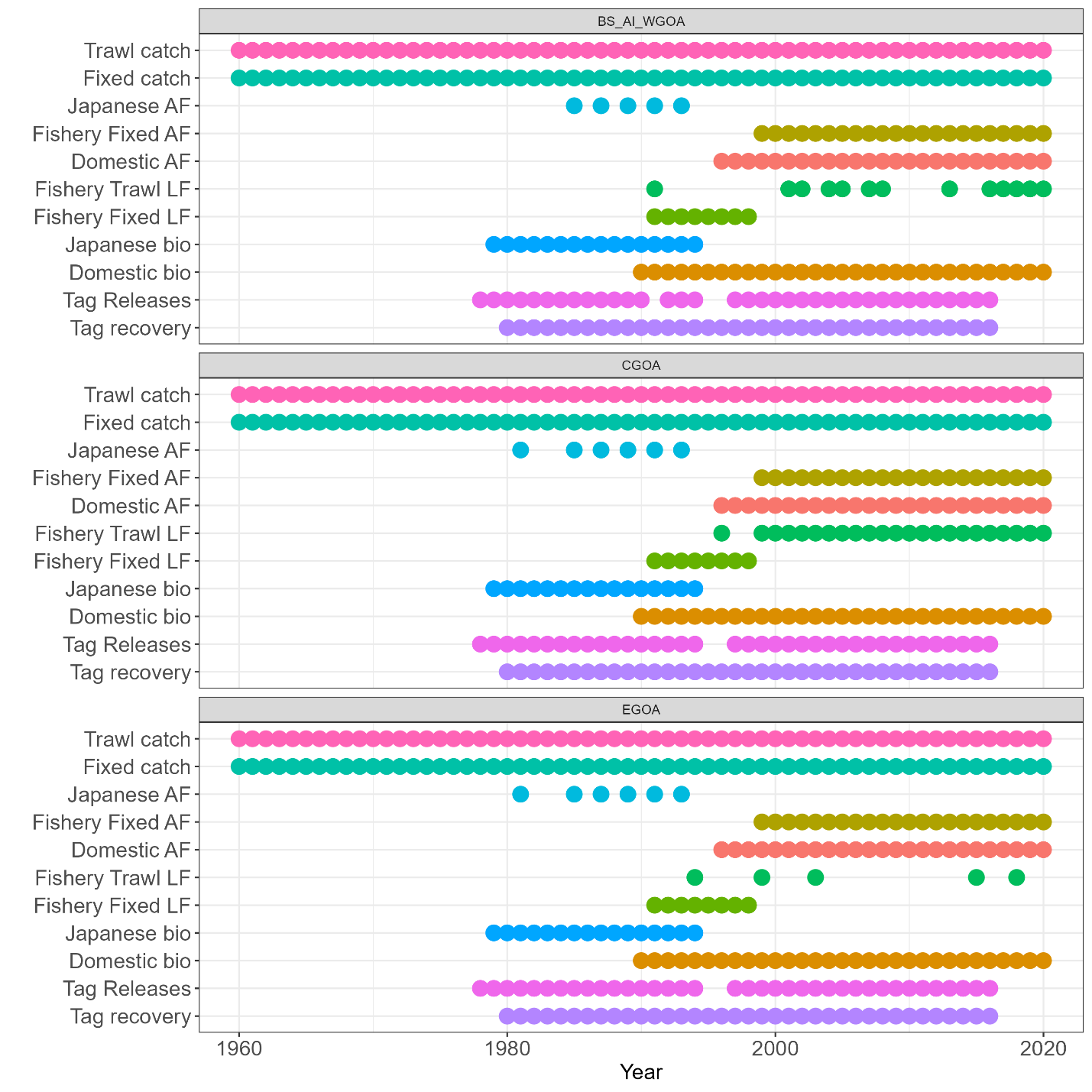


Figure : Observation frequency of 3 area model.

# Appendix C

## Fits to the 1960 Single area model

Fits to the 1960 Three area model



