Random effects on numbers-at-age transitions implicitly account for movement dynamics and improve stock assessment and management

Chengxue Li[1,6,*] Jonathan J. Deroba[1] Aaron M. Berger[2]

Daniel R. Goethel[3] Brian J. Langseth[4] Amy M. Schueller[5]

Timothy J. Miller[1]

[1] NOAA, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA

[2] NOAA, Northwest Fisheries Science Center, 2032 SE OSU Drive, Newport, OR 97365, USA

[3] NOAA, Alaska Fisheries Science Center, 17109 Point Lena Loop Road, Juneau, AK 99801, USA

[4] NOAA, Northwest Fisheries Science Center, 2725 Montlake Blvd E, Seattle, WA 98112, USA

[5] NOAA, Southeast Fisheries Science Center, 101 Pivers Island Road, Beaufort, NC 28516, USA

[6] Saltwater Inc., 733 N Street, Anchorage, AK 99501, USA

* *Corresponding author*: chengxue.li@noaa.gov

Abstract

Implementing operational assessment models that account for spatial structure and movement dynamics is challenging, especially with limited tagging data. Random effects on numbers-at-age (NAA) transitions in state-space models offer a potential solution to circumvent direct movement estimation by attributing movement variation to NAA random effects. However, whether this approach reliably achieves desirable management outcomes remains unclear. In this study, we conducted a management strategy evaluation that emulated a generic medium-lived fish that exhibit natal homing dynamics, using assessment models with varying levels of spatial complexity. We compared the performance of each spatial implementation with and without NAA random effects to evaluate their effectiveness in achieving management outcomes. Our results showed that models with NAA random effects consistently outperformed those without, although the benefits of NAA random effects degraded at high rates of movement. Therefore, NAA random effects could serve as a practical intermediate solution when explicit movement modeling is not feasible due to insufficient movement information. Our findings suggest that incorporating NAA random effects should be a default starting point in state-space stock assessments. (word count: 174)

Keywords: state-space models, management strategy evaluation, movement dynamics, numbers-at-age transitions, random effects

1 Introduction

Random effects on numbers-at-age (NAA) transitions introduce flexibility into a state-space stock assessment that can account for a wide range of population (e.g., movement, natural mortality) and fishery processes (e.g., selectivity, misreported catch) (Nielsen and Berg, 2014; Cadigan, 2016; Albertsen et al., 2018; Perretti et al., 2020; Stock and Miller, 2021; Stock et al., 2021). Simulation-estimation studies have consistently concluded that incorporating NAA random effects results in less biased estimates of management quantities (e.g., biomass, reference points), even when NAA random effects were included in the assessment model but absent in the operating model that generated the data (Li et al., 2024). Previous studies suggested that NAA deviations can be interpreted as migrations or irregular natural mortality (Frisk et al., 2010; Gudmundsson and Gunnlaugsson, 2012). However, the extent to which NAA random effects improve assessment results and management outcomes when accounting for movement has not been explicitly tested within the state-space modeling framework, despite claims that they can do so (Nielsen and Berg, 2014; Cadigan, 2016; Stock and Miller, 2021).

The benefits of accounting for spatial structure and movement dynamics in stock assessments have become increasingly apparent (Goethel et al., 2021; Bosley et al., 2022; Goethel et al., 2023; Berger et al., 2024) and are necessary as climate change continues to drive fish movement, alter fish biology and ecology, and affect habitat availability and suitability (Pörtner and Peck, 2010; Ciannelli et al., 2013). Acknowledging biocomplexity and spatial heterogeneity in fish distribution and demographics challenges the panmictic assumptions of traditional stock assessment models, which consider the modeled stock as a single population that is uniformly distributed, demographically homogeneous, closed to migration, and reproductively isolated (Beverton, 1957; Cadrin, 2020). Assessment models that do not account for spatial structure and movement dynamics, when in fact they are present, can produce biased estimates of population quantities, potentially failing to achieve management objectives (Bosley et al., 2019; Goethel et al., 2023). Including spatial structure and movement dynamics in stock assessments, however, can be challenging (Punt, 2019; Punt et al., 2020; Goethel et al., 2022; Berger et al., 2021, 2024), and in some cases may not be practical due to jurisdictional or management constraints (Kerr et al., 2017), low quality and quantity of geo-partitioned data (Cadrin, 2020; Punt et al., 2020), inadequate tagging information (Goethel et al., 2011, 2019), or limited knowledge about stock origin and connectivity (Berger et al., 2017).

The "fleets-as-areas" approach is often used as a practical starting point in spatial assessments (Berger et al., 2012; Waterhouse et al., 2014; Lee et al., 2017; Bosley et al., 2022). This approach assumes the population is homogeneously distributed across its range and implicitly accounts for spatial heterogeneity in distribution through area-specific fleets by attributing differences in age compositions to fishing selectivity. However, true fishing selectivity may not varying spatially, and observed differences may instead result from population processes such as movement (Cope and Punt, 2011; Berger et al., 2012; Hurtado-Ferro et al., 2014; O'Boyle et al., 2016). The implicit mismatch between assumption (spatially varying selectivity) and true process (spatially varying distribution) can sometimes undermine the performance of the assessment-management framework (Punt et al., 2016a,b).

Explicitly specifying population structure and spatial units, but ignoring movement, provides an alternative approach for improving population estimates and derived management parameters in a spatial context, particularly when adult movement is infeasible to estimate (Cadrin et al., 2019; Goethel et al., 2023). Evidence suggests that incorporating spatially referenced parameters within an assessment model to account for the net effects of movement on observations of population structure at the local scale is both practical and broadly applicable to many fisheries (Berger et al., 2012; Goethel et al., 2023). This approach is particularly useful in cases where species biology varies geographically due to local biophysical conditions, habitat availability and suitability, or harvest patterns (Goethel et al., 2023), and when adult movement is limited, with larval dispersal driving spatial structure (Kerr et al., 2017).

Incorporating NAA random effects in state-space models represents a step forward in model complexity, where spatial structure is implicitly or explicitly defined, and movement is handled implicitly through NAA random effects. Evaluating whether NAA random effects can improve management outcomes in stock assessments that exclude explicit movement modeling is crucial. Using NAA random effects as an "add-on" proxy for movement simplifies movement dynamics and reduces the number of parameters estimated in the assessment model. However, despite simplifying parameter estimation, NAA random effects do not mechanistically capture the underlying movement dynamics of fish populations. Instead, movement variability is absorbed into the NAA random effects structure, which can lead to misinterpretations of population processes, model misspecification, and biased estimates of stock status and productivity.

Thus, NAA random effects may provide a more tractable solution, but the trade-off between model complexity and biological realism must be carefully considered, especially in the context of management performance. One effective tool to evaluate whether management procedures achieve management objectives (e.g., sustainable catch) is through management strategy evaluation (MSE) (Punt et al., 2017; Keymer et al., 2000). The MSE approach in a spatial context involves considering various candidate spatial population structures, data collection methods, estimation model structures, harvest control rules, and their interactions and feedbacks (Kerr and Goethel, 2014; Goethel et al., 2016).

In this study, we use a MSE with closed-loop simulations in a controlled, research-based environment to evaluate the performance of several spatial assessment model configurations that do not explicitly account for movement but include NAA random effects, while the operating model defining the true population dynamics includes movement. We examine the utility of incorporating NAA random effects across increasing levels of assessment model complexity (panmictic, spatially implicit, and spatially explicit) and assess their influence on estimated assessment quantities and management outcomes under varying movement dynamics and historical fishing pressure.

2 Methods

2.1 Model Overview

The Woods Hole Assessment Model (WHAM, https://timjmiller.github.io/wham), a state-space assessment model, was used as the basis of the operating model (OM) and estimation model (or assessment model; EM) in the MSE framework. WHAM can incorporate time- and/or age-varying process errors on population and fishery processes, including recruitment, NAA, natural mortality, movement, fishing selectivity, and survey catchability, and it has ability to link environmental covariates to specific processes (Stock and Miller, 2021). In this study, we used a multi-stock and multi-region extension of WHAM (hereafter referred to as Multi-WHAM) that incorporates functionalities for modeling populations with spatial heterogeneity in fish biology, movement dynamics, and fleet dynamics (Miller et al., In review). Here we give a minimal overview necessary for the present study (see Methods in the supplementary file for details). We used the whamMSE package (https://lichengxue.github.io/whamMSE/), a comprehensive MSE toolbox developed upon Multi-WHAM, to conduct MSE and explore potential implications of management strategies.

2.2 Operating Models (OMs)

Two operating models (OMs) were developed that represent the true dynamics of the system. Each was conditioned using parameters for fish with a medium-lived (maximum age = 12) life history (Wiedenmann et al., 2015) and differed only in the rates of movement assumed (high and low) between populations and regions (Table 1). The spatial domain included two regions, each of which is the natal home of one population but can also be occupied by the other population outside the spawning season. Population and fishery parameters were consistent across the domain, with exceptions for stock-recruitment relationship and fishing selectivity (Table 1).

2.2.1 Fishing History

For each OM, three alternative fishing histories (see Table 2) were used during the simulation historical period: 1) $\mathrm{F_A}$: both regions experience overfishing ($2.5F_{MSY}$) for the first 15 years and fishing at F_{MSY} for the last 15 years; 2) $\mathrm{F_U}$: both regions experience constant overfishing ($2.5F_{MSY}$) for the entire 30 years; and 3) $\mathrm{F_C}$: region 1 experiences overfishing ($2.5F_{MSY}$) for the first 15 years and fishing at F_{MSY} for the last 15 years, while region 2 follows the opposite fishing history.

2.2.2 Biological Information

The Beverton-Holt stock-recruitment relationship was assumed for both populations, with slightly different productivity (α) parameters to create some discrepancy in biology between the two populations. Weight-at-age, maturity-at-age, and natural mortality were assumed to be the same between the two populations. Initial numbers-at-age for each population were specified at the equilibrium distribution fishing at $F = 0.1$ in each region with an initial recruitment of e^{10}. Recruitment and NAA transitions for each population were treated as independent random effects (i.e. IID), with standard deviations (σ) of 0.5 and 0.2, respectively.

2.2.3 Movement Dynamics

Four seasons (spring, summer, autumn, winter) were included in the OMs to model seasonal migrations typical of natal homing dynamics. Specifically, 2 populations with reproductive isolation exhibit natal homing migrations during the summer, spawn in the autumn, and may migrate to other regions during the spring and winter (e.g., feeding migration). Two different movement scenarios low movement and high movement were included in this study (see Figure 1). The high movement scenario ($\mathrm{OM_{high}}$) allows both populations to move between regions outside the spawning season. The movement rate from region 1 to region 2 was assumed to be much higher than the rate in the opposite direction, reflecting a degree of source-sink dynamics, that could reflect habitat suitability (Keymer et al., 2000). The low movement scenario ($\mathrm{OM_{low}}$) allows population 1 to move from region 1 to region 2 and back from region 2 to region 1, whereas population 2 cannot move between regions. This type of movement represents strict source-sink movement dynamics typical of some species (e.g., black sea bass, *Centropristis striata*, Miller et al., In review). The movement rate outside the spawning season was assumed to be time-varying and treated as a random effect, with a standard deviation of 0.5 and an AR1 autocorrelation coefficient of 0.5. A realization of these specified movement rates can be found in Figure S1 and Figure S16.

2.2.4 Fleet and Survey

A fleet was assumed to operate in each region, with logistic selectivity patterns differing slightly between the two fleets to reflect spatial heterogeneity (Table 1). A survey was assumed to be conducted in each region during the spawning season to provide the most accurate index of SSB for each population. Both surveys share the same selectivity curve, with a lower a_{50} than that of the fleet, ensuring that adequate information on young fish can be collected. A multinomial distribution with an effective sample size of 100 was assumed for the fleet and survey composition data, and a log-normal distribution with a coefficient of variation (CV) of 0.1 was used for aggregate catch and survey indices. Weight-at-age was assumed to be the same across populations, regions, fleets, and surveys.

2.3 Estimation Models (EMs)

EMs with varying population structures, complexities, and the inclusion or exclusion of NAA random effects were explored (Figure 2; Table 3). For simplicity, EMs that included NAA random effects are hereafter referred to as $\mathrm{EMs_{NAA}}$, while EMs without NAA random effects are referred to as $\mathrm{EMs_{noNAA}}$.

Of the 10 EMs, the first 8 were paired (with and without NAA random effects) and included a single spatially aggregated panmictic assessment model ($\mathrm{PAN_{NAA}}$ and $\mathrm{PAN_{noNAA}}$), a single spatially implicit assessment model ($\mathrm{FAA_{NAA}}$ and $\mathrm{FAA_{noNAA}}$), two separate individual assessment models by region ($\mathrm{SEP_{NAA}}$ and $\mathrm{SEP_{noNAA}}$), and a spatially disaggregated assessment model with no movement ($\mathrm{SpD_{NAA}}$ and $\mathrm{SpD_{noNAA}}$). The difference between SEP and

SpD EMs is that SpD EMs use global SPR-based reference points weighted by regional recruitment, while SEP EMs separately calculate regional SPR-based reference points. The remaining two EMs were spatially explicit with movement either fixed ($\text{SpE}_{\text{NAA,Fix}}$) at the true value or with mean movement estimated ($\text{SpE}_{\text{NAA,Est}}$) using a prior distribution (with mean = μ_{True} and standard deviation = 0.5). When NAA random effects were included in the EM, they were assumed to be distributed as independent and identically distributed as normal (N~IID), in log space. All EMs modeled recruitment as N~IID random effects in log space. We did not apply the log-normal bias adjustment for process errors or observations (Li et al., In review).

The EM with fixed movement ($\text{SpE}_{\text{NAA,Fix}}$) was considered as the best possible EM and was treated as the baseline for comparison.

2.4 Management Strategy Evaluation (MSE)

A MSE was conducted to evaluate the performance of the EMs. The simulation included a 30-year historical period followed by a 30-year feedback period. Preliminary results indicated that a 30-year feedback period was sufficient to capture differences in management outcomes among the EMs compared to a longer feedback period. Assessments were conducted every 3 years, with each assessment using the most recent 20 years of data available at that time. Sensitivity analysis indicated small differences in convergence rates and parameter estimates of the EMs when the number of years of input data was increased beyond 20 (Li et al., 2018).

In the interim years between assessments, the population was projected in the EM using the average selectivity-at-age, maturity-at-age, weight-at-age, and natural mortality from the last 5 model years of the most recent assessment, as well as the average recruitment across all model years. For EMs with processes (e.g., recruitment, NAA, movement) treated as random effects, only recruitment random effects were assumed to continue while other process random effects were set to zero during the projection years (Stock and Miller, 2021).

Target catch in those interim years, was based on a constant fishing mortality harvest control rule at 75% of $F_{40\%}$. The SEP EM used region-specific SPR-based reference points, whereas all other EMs used global SPR-based reference points. For SpD and SpE EMs, additional weights based on the mean recruitment of each population estimated by the EM were applied when calculating the global SPR-based reference points. In contrast, no such weights were used for PAN and FAA EMs, as they assumed a single population and relied on a total estimated recruitment. For the PAN EM, the total catch target was allocated to region-specific catches using weights derived from the relative biomass catch observed in each survey during the terminal year of the assessment (Bosley et al., 2019). All other EMs provided region-specific catch advice that was inherently based on regional F-at-age matrices, all of which collectively contribute to the calculation of the global F. See Table S1 for more details about catch apportionment.

The annual catch advice from each EM was sequentially fed into the OMs to update fishing mortality and population dynamics. This closed-loop simulation was performed iteratively until the end of the feedback period. If any of the 10 EMs failed to converge during a replicate, the results for that entire replicate, across all EMs, were discarded. A new random seed was selected, and the replicate was attempted again until 100 replicates successfully converged for all 10 EMs in the study. In total, the MSE evaluated 60 scenarios comprising two states of nature (high or low movement), three fishing histories, and 10 EMs.

2.4.1 EM Diagnostics

Performance was summarized over 100 converged runs to ensure equal sample sizes across scenarios. Convergence rates for each EM were calculated based on the first 100 realizations (including replicates that failed to converge) for each OM and fishing history. In addition to convergence rates, all 10 EMs were evaluated by comparing their estimates of mean recruitment, recruitment standard deviation, and NAA standard deviation, using all 1,000 converged models (100 replicates per EM) for each OM and fishing history.

2.4.2 Simulation-Estimation Results

To evaluate estimation bias of each EM, the relative bias of estimated annual recruitments and SSB was calculated for each EM compared to the true values from the simulated pseudo-dataset i. The median relative bias for each EM from the first assessment was calculated as:

$$\text{Relative Error}_i = \text{Median}\left(\frac{\hat{\theta}_{i,y}}{\theta_{i,y}} - 1\right) \tag{1}$$

where $\theta(i, y)$ represents the true value for year y from the simulated pseudo-dataset i, and $\hat{\theta}(i, y)$ is the estimated value obtained from the EM fitted to the pseudo-dataset.

To compare parameter estimates associated with recruitment and NAA processes, estimates of mean recruitment (μ_{Rec}), recruitment standard deviation (σ_{Rec}), and NAA standard deviation (σ_{NAA}) were also collected from the EMs at the last iteration of the closed-loop.

2.4.3 MSE Performance Metrics

Short-term (first 5 years of the feedback period) and long-term (last 5 years of the feedback period) performance metrics at both global and regional scales were evaluated to quantify the performance of each EM. These performance metrics include:

- Average catch: $\frac{1}{n}\sum_{y=1}^{n} \text{Catch}_y$;
- Average fully selected F: $\frac{1}{n}\sum_{y=1}^{n} F_y$;
- Average SSB: $\frac{1}{n}\sum_{y=1}^{n} SSB_y$;

In our study, $\text{SpE}_{\text{NAA,Fix}}$ was used as the baseline EM because it represents performance with correctly specified movement, thereby reflecting the best possible performance. For replicate i, the relative difference between the performance metrics of the baseline EM and other EMs was calculated as:

$$\text{Relative Difference } (i) = \frac{\text{Metric}_{EM_x}(i)}{\text{Metric}_{\text{SpE}_{\text{NAA,Fix}}}(i)} - 1 \tag{2}$$

This approach standardizes differences across performance metrics, providing better visualization and comparison on the same scale.

Following performance metrics were also collected:

- Probability of $F > F_{MSY}$: $\mathrm{P}_{\mathrm{F>F_{MSY}}} = \frac{1}{n}\sum_{y=1}^{n} I\left(\frac{F_y}{F_{MSY_y}} > 1\right)$;
- Probability of $SSB < SSB_{MSY}$: $\mathrm{P}_{\mathrm{SSB<SSB_{MSY}}} = \frac{1}{n}\sum_{y=1}^{n} I\left(\frac{SSB_y}{SSB_{MSY_y}} < 1\right)$;
- Overfishing status in the terminal year of the feedback period: $\mathrm{OFG}_{\mathrm{terminal}} = \frac{F_T}{F_{MSY_T}}$;
- Overfished status in the terminal year of the feedback period: $\mathrm{OFD}_{\mathrm{terminal}} = \frac{SSB_T}{SSB_{MSY_T}}$;
- Average annual catch variation: $\mathrm{AACV} = \frac{\sum_{y=1}^{n}|\mathrm{Catch}_y - \mathrm{Catch}_{y-1}|}{\sum_{y=1}^{n}\mathrm{Catch}_{y-1}}$.

The annual global reference points (F_{MSY_y} and SSB_{MSY_y}) were used for calculating some of the above performance metrics (Stock et al., 2021; Miller et al., In review), using weighted average where the weights were based on the mean of realized annual recruitments for each population within each replicate.

All performance metrics for the 10 EMs at the global scale were collected to provide a holistic view of model performance. Metrics such as catch, F, SSB, terminal-year SSB/SSB_{MSY} for each OM and fishing history were normalized to scores between 0 and 1 using:

$$\text{Score } (i) = \frac{\text{Metric}_{EM_x}(i) - \min(\text{Metric}_{EM_A}(i))}{\max(\text{Metric}_{EM_A}(i)) - \min(\text{Metric}_{EM_A}(i))}, \text{with } A = \{1, 2, \dots, 10\} \tag{3}$$

For metrics where higher values indicated poorer performance (i.e., $\mathrm{P}_{\mathrm{F>F_{MSY}}}$, $\mathrm{P}_{\mathrm{SSB<SSB_{MSY}}}$, $\mathrm{OFG}_{\mathrm{terminal}}$, and AACV), normalization was adjusted to ensure higher values consistently indicate better performance, as follows:

$$\text{Score } (i) = 1 - \frac{\text{Metric}_{EM_x}(i) - \min(\text{Metric}_{EM_A}(i))}{\max(\text{Metric}_{EM_A}(i)) - \min(\text{Metric}_{EM_A}(i))}, \text{with } A = \{1, 2, \dots, 10\} \tag{4}$$

The median, 25th quantile, and 75th quantile of the above performance metrics and scores were calculated and presented in the results.

3 Results

Our simulation-estimation results indicated the $\text{SpE}_{\text{NAA,Est}}$ EM and the $\text{SpE}_{\text{NAA,Fix}}$ EM (baseline EM) performed similarly, both producing accurate estimates of recruitment (Figure S2 and Figure S17) and SSB (Figure S3 and Figure S18), as well as comparable estimates of model parameters (Figure S4 and Figure S19). In addition, our MSE results showed that $\text{SpE}_{\text{NAA,Est}}$ performed similarly to the baseline EM for the most of MSE performance metrics presented below and performed relatively better than the other EMs. For simplicity, no further discussion of the results for the $\text{SpE}_{\text{NAA,Est}}$ EM is provided.

Convergence rate was high (85%-100%) for all cases. Results of OM_{low} generally aligned with those of OM_{high}, but the performance metrics among EMs were less distinct due to lower movement rates. Additionally, EMs_{NAA} under the OM_{low} scenario generally performed similarly to the baseline EM in most performance metrics. For simplicity, detailed results were presented only for OM_{high}, while a brief summary of the results for OM_{low} was provided at the end of this section, with additional details available in the supplementary file (Figures S16-S35).

3.1 Simulation-estimation Results

Total recruitment estimate from PAN and FAA EMs were similar to the sum of region-specific recruitment from SEP and SpD EMs, all of which were slightly overestimated compared to the total recruitment estimated from the baseline EM (Figure S4). The mean recruitment parameter (μ_{Rec}), recruitment standard deviation (σ_{Rec}) and NAA standard deviation (σ_{NAA}) from EMs without movement included, appeared to compensate to account for the absence of movement dynamics (Figure S4).

3.2 MSE Performance Metrics

3.2.1 Relative Difference in Annual Catch and SSB

Relative difference in annual catch over the 30-year feedback period revealed larger discrepancies among EMs during the first 10 years compared to later years. As time progressed, EMs_{NAA} performed more similarly to the baseline EM (Figure S5). This trend was more pronounced at the regional scale (Figures S6-S7). Relative differences in annual SSB at both global (Figure S8) and regional scales (Figures S9-S10) tended to diverge from zero over time. EMs_{NAA} maintained higher SSB that was more similar to the baseline EM than $\text{EMs}_{\text{noNAA}}$.

3.2.2 Short-term Catch, F, and SSB

In the first five years of the feedback period, EMs_{NAA} generally produced catch, F, and SSB more similar with smaller variations (i.e., inter-quantile ranges) to the baseline EM than $\text{EMs}_{\text{noNAA}}$ at both regional and global scales (Figure 3). SEP and SpD EMs with NAA random effects generally produced results more similar to the baseline EM than other EMs, and this was especially true for the SpD_{NAA} EM (Figure 3).

Performance in catch, F, and SSB also exhibited difference across EMs with difference in model complexity. For example, PAN EMs generally produced higher catch (with higher F) in region 1 and lower catch (with lower F) in region 2, and as a result lower SSB in region 1 and higher SSB in region 2, although the global performance metrics were similar to the baseline EM (Figure 3). Despite high contrast in F between regions when using PAN EMs, SSB showed only marginal difference compared to the baseline (5% less) in the short term (Figure 3).

3.2.3 Long-term Catch, F, and SSB

In the last five years of the feedback period, there was marginal difference in median catch (despite variations being higher for $\text{EMs}_{\text{noNAA}}$) at both regional and global scales between EMs_{NAA} and $\text{EMs}_{\text{noNAA}}$ across differing levels of model complexity (Figure 4). However, $\text{EMs}_{\text{noNAA}}$ often produced significantly high F, particularly at the regional scale (Figure 4). The consistently high F in $\text{EMs}_{\text{noNAA}}$ resulted in lower SSB compared to the baseline EM at both regional and global scales (Figure 4).

Spatially implicit and explicit EMs (i.e., FAA, SEP, and SpD EMs) performed similarly, particularly at the global scale, all of which produced ~10% lower SSB when NAA random effects were included and ~20% lower SSB when NAA random effects were not included, compared to the baseline EM (Figure 4). Interestingly, the simplest PAN EMs performed slightly better than other EMs at the global scale, particularly under the unsustainable and contrasting fishing histories (5-10% lower global SSB) (Figure 4). However, the corresponding F for region 1 produced by PAN EMs was 50%-75% higher than the F in the baseline EM (Figure 4).

3.2.4 Average Annual Catch Variation (AACV)

Including NAA random effects exhibited potential to lower AACV, particularly at the regional scale, regardless of model complexity and fishing histories (Figure 5). Although AACV at the global scale was similar across EMs for each fishing history, some benefits of including NAA random effects in reducing AACV were still observed under the adaptive fishing history (Figure 5). Interestingly, AACV at the global scale was particularly low when the PAN_{NAA} EM was used (Figure 5). At the regional scale, relatively low AACV occurred when using FAA EMs, which even outperformed the baseline EM, regardless of fishing histories (Figure 5). In contrast, PAN and SEP EMs generally exhibited high AACV at the regional scale (Figure 5).

3.2.5 Probability of Overfished and Overfishing

In most cases, EMs_{NAA} outperformed $\text{EMs}_{\text{noNAA}}$ in preventing populations from becoming overfished and overfishing at the global scale (Figure 6). EMs_{NAA} exhibited similar performance in terms of the probability of $SSB < SSB_{MSY}$ compared to the baseline EM, but performed worse in preventing overfishing (Figure 6). Interestingly, the probability of overfishing was found to be lower when using the PAN_{NAA} EM compared to other EMs_{NAA} (Figure 6).

3.2.6 Terminal-year Status

Populations were not overfished in the terminal year at the global scale, regardless of NAA random effects, model complexity, and fishing histories (Figure S12). EMs_{NAA} performed better than $\text{EMs}_{\text{noNAA}}$ in preventing global overfishing in the terminal year (Figure S12). Notably, none of $\text{EMs}_{\text{noNAA}}$ could prevent global overfishing (Figure S12).

3.2.7 Holistic View of Performance

In general, EMs_{NAA} outperformed $\text{EMs}_{\text{noNAA}}$ across most metrics, except for short-term catch (Figure 7). $\text{SEP}_{\text{noNAA}}$ produced the highest global catch under adaptive and unsustainable fishing histories, while $\text{PAN}_{\text{noNAA}}$ produced the highest global catch under the contrasting fishing history (Figure 7). Interestingly, relatively complex $\text{EMs}_{\text{noNAA}}$, such as $\text{SEP}_{\text{noNAA}}$ and $\text{SpD}_{\text{noNAA}}$ EMs, exhibited the lowest global performance among all EMs, performing even worse than the simplest $\text{PAN}_{\text{noNAA}}$ EM (Figure 7). Similar results were found at the regional scale (Figures S14-S15).

3.2.8 Performance under the Low Movement Scenario

Short-term and long-term performance showed that EMs_{NAA} performed more similarly to the baseline EM (Figures S26-S27). Specifically, SEP_{NAA} and SpD_{NAA} EMs exhibited nearly identical performance to the baseline EM in terms of catch, F, and SSB (Figures S26-S27). In contrast, $\text{EMs}_{\text{noNAA}}$, although achieving higher catch in the short term, led to significant declines in SSB in the long term at both regional and global scales, regardless of model complexity or fishing history (Figures S26-S27). Similar patterns in AACV, $\text{P}_{\text{SSB}<\text{SSB}_{\text{MSY}}}$, $\text{P}_{\text{F}>\text{F}_{\text{MSY}}}$, $\text{OFD}_{\text{terminal}}$, and $\text{OFG}_{\text{terminal}}$ were observed under the high movement scenario (Figures S28-S32). These results consistently demonstrate that EMs_{NAA} outperformed $\text{EMs}_{\text{noNAA}}$, with the most complex EMs with NAA generally showing superior performance compared to other EMs (Figures S33-S35).

4 Discussion

Our findings confirm that management outcomes generally improve when NAA random effects are included, whereas excluding them from assessment models leads to undesirable outcomes both regionally and globally. Earlier claims have suggested that NAA deviations can represent fish migration in stock assessment models (Gudmundsson and Gunnlaugsson, 2012; Cadigan, 2016). Our study suggests that the improved management outcomes likely result from the ability of NAA random effects to implicitly account for movement variation, in accordance with the understanding that NAA random effects serve as a 'catch-all' to absorb process variation from confounding processes such as fishing selectivity and natural mortality (Li et al., 2024).

4.1 Improved Parameter Estimation with NAA Random Effects

It is common for one modeled process to alias others in state-space models (Stock et al., 2021). Our study demonstrates that NAA random effects not only alias unmodeled movement dynamics but also mitigate the consequences of model misspecification caused by the lack of a movement component in the assessment model. We found that NAA random effects absorbed movement process variation. Without NAA random effects, mean recruitment and recruitment variability increased significantly to compensate for movement. When NAA random effects are included, the movement-related uncertainty is partitioned between recruitment and NAA. This partitioning reduces recruitment overestimation and attributes some of the movement uncertainty to NAA random effects, as evidenced by decreased recruitment variability and increased NAA variability. Consequently, EMs_{NAA} provided more accurate recruitment estimates, both regionally and globally. For example, estimates of global recruitment and recruitment in region 2 showed significant improvements in EMs_{NAA} compared to models without them. A similar pattern was observed for SSB, with global SSB and region 2 SSB estimates being more accurate in EMs_{NAA} compared to $\text{EMs}_{\text{noNAA}}$. This improvement may indirectly enhance the accuracy of other parameter estimates, thereby improving management outcomes.

4.2 Improved Management Performance with NAA Random Effects

We showed that including NAA random effects generally improved management outcomes in nearly every aspect of performance metrics. By contrast, excluding NAA random effects impeded management outcomes by producing F that induced overfishing, partially due to recruitment overestimation. This overestimation led to higher catch but decreased SSB, both regionally and globally, in the short term. The absence of NAA random effects caused more pronounced issues in the long term. For instance, consistently high F driven by optimistic estimation of population size and productivity led to significant reductions in SSB, while the gain in catch diminished. Compared to the baseline EM, we found global SSB declined by up to 20% and 15% under high and low movement scenarios, respectively, without corresponding increases in catch. Furthermore, mean recruitment and recruitment variability was amplified in $\text{EMs}_{\text{noNAA}}$, leading to inaccurate estimation of population size and productivity, thereby impacting the stability of catch and SSB in both the short term and long term. Issues of using $\text{EMs}_{\text{noNAA}}$ were further compounded when assessments were conducted every 3 years, as errors of population estimates, reference points, and catch advice accumulated during the interim years, ultimately degrading long-term management performance.

Under low rates of movement, our results indicate that NAA random effects have the potential to account for movement and achieve satisfactory management outcomes. This is echoed by the nearly identical performance of SEPɴᴀᴀ and SpDɴᴀᴀ EMs compared to the baseline EM. However, under high movement rates, NAA random effects may struggle to fully compensate for the consequences of model misspecification due to the absence of explicitly modeled movement. For instance, while EMs_{NAA} achieved desirable management outcomes in the short term, in the long term global SSB still decreased by 5-10%, with region 1 experiencing a 0-25% reduction and region 2 a 0-30% reduction, depending on historical fishing pressure and the spatial complexity of the assessment model. In addition, $\mathrm{EMs}_{\mathrm{NAA}}$ exhibited limited ability to prevent overfishing, as reflected by even relatively complex EMs approaching overfishing thresholds in the terminal year.

4.3 Benefits of Correct Spatial Structure

Our study demonstrates that the benefits of NAA random effects are maximized when combined with correct spatial structures in stock assessment models. Accurate spatial boundaries and sufficient data quantity and quality for spatial disaggregation are critical for achieving management objectives (Berger et al., 2021). Conversely, mismatches between population structures and management units can result in adverse management outcomes (Cope and Punt, 2011; Hintzen et al., 2015; Kerr et al., 2017; Berger et al., 2021). Spatial heterogeneity can arise from spatially heterogeneous biological characteristics such as recruitment, growth, maturity, and movement dynamics, or fishery characteristics such as fishing mortality and selectivity patterns. In our study, such heterogeneity created barriers for assessment models with incorrect spatial structure to achieve successful management, as demonstrated by increased risks of localized depletion due to inappropriate allocation of catch.

Limitations of incorrect spatial structure were exemplified by the poor performance of the $\mathrm{PAN}_{\mathrm{NAA}}$ EM in our study, even when catch allocation was informed using the proportion of total survey biomass in each management area, a method regarded as best practice for maximizing system yield when the underlying spatial structure is unknown (Bosley et al., 2019). The primary challenge in our scenario arises from the interaction of natal homing and source-sink dynamics under high movement rates, which reduces the effectiveness of this approach. In our high movement scenario, region 1 (source region with high productivity) disproportionately contributes to global recruitment but does not retain a large portion of its recruits. Instead, many recruits and subsequent fish at age 2+ migrate to region 2 (sink region with low productivity) outside the spawning season and contribute to the fishery there. If catch allocation is based solely on SSB or recruitment within each region, this approach will likely lead to an underestimation of the contribution of the source region to the fishery in the sink region and overly optimistic estimation about the productivity from the source region. However, due to the substantial emigration of recruits to the sink region, the actual biomass available to the fishery in the source region should be lower than the survey suggests. This mismatch can result in overharvesting in the source region and underharvesting in the sink region, ultimately degrading management outcomes and reducing the likelihood of achieving maximum sustainable yield, as evidenced by significantly high F (combined with risk of overfishing) in region 1 with low F in region 2 (combined with not overfishing) in our study. In addition, fishing selectivity patterns differ between regions in our study, while assuming a 'global' fishing selectivity pattern in a PAN EM may lead to overestimation in one region and underestimation in another. This mismatch likely perturbs model performance, resulting in incorrect estimation of fishing status and suboptimal management advice. Even though management performance may be sustained, PAN EMs should be applied with caution, as they may lead to localized depletion (Punt et al., 2015; Bosley et al., 2019, 2022) and loss of biocomplexity, ultimately affecting population resilience and stability, particularly in the face of climate change and human perturbations (Kerr et al., 2017).

The FAA_{NAA} EM outperformed the PAN_{NAA} EM in our study by providing robust estimates of population status and productivity, with regional catch advice better reflecting the underlying spatial heterogeneity in fishing patterns. Compared to the PAN_{NAA} EM, the FAA_{NAA} EM with spatial structure modeled implicitly delivered more desirable management outcomes in our study. We found that the FAA_{NAA} EM also outperformed the baseline EM by providing more stable catches at both regional and global scales, consistent with Punt et al. (2017). This highlights their potential as a cost-effective approach for balancing model complexity and practicality. The fleets-as-areas approach assumes a single homogenous population while attributing regional differences in age compositions to fishing selectivity. In some cases, this flexibility in modeling spatially distinct fleet patterns helps absorb movement process variation by attributing differences in age compositions resulting from movement dynamics to fishing selectivity (Cope and Punt, 2011; Berger et al., 2012; O'Boyle et al., 2016). For example, previous studies have reported dome-shaped selectivity as a result of model behavior that helps capture spatial heterogeneity in population structure and fishing with low rates of movement (Waterhouse et al., 2014). Incorporating time-varying process error into selectivity has been shown to provide precise estimates of derived and management quantities when spatial patterns due to age-specific movement are unmodeled (Lee et al., 2017). This highlights an avenue for future research to explore modeling selectivity as random effects (in addition to NAA random effects) to better account for movement dynamics and improve model accuracy.

SEPNAA and SpDNAA EMs, which explicitly model spatial structure, demonstrate promising potential in improving management outcomes. This is evident from their comparable management performance to the baseline EM in this study. However, their performance was degraded under the high movement scenario, as they do not fully capture the movement dynamics. Incorporating spatially referenced parameters in an assessment model is generally ideal when assumptions of homogeneity in biology and fisheries are violated, as it accounts for the spatial structure more effectively. With increasing model complexity and flexibility, assessment models may better account for the net effects of movement on observations of population structure at local scales, effectively addressing spatial structure and movement dynamics. The primary difference between SEP and SpD EMs lies in how biological reference points were determined. SpD EMs rely on global SPR-based reference points, which are weighted by the estimated recruitment within each region, whereas SEP EMs calculate reference points independently for each region. In most cases, our results showed minimal differences in management performance between these two EMs. However, under the contrasting fishing history, distinct patterns of fishing mortality impacted recruitment estimates in each region, thereby affecting the recruitment-based weights that determined the influence of each region on the global SPR-based reference point. Consequently, the F reference points provided by SEP and SpD EMs differed, resulting in variations in regional management performance, while global management performance remained largely identical. For instance, source regions with high recruitment may dominate the global SPR-based reference point, even if their actual contribution to the fishery is diminished by emigration. This can lead to inappropriate catch allocations, causing underharvesting in source regions (due to overestimated recruitment weights) and overharvesting in sink regions (due to underestimated recruitment weights), as evidenced in our case where the $\mathrm{SpD}_{\mathrm{NAA}}$ EM resulted in underfishing for region 1 and overfishing for region 2 when movement rates were high.

Our results demonstrate that estimating movement in spatial state-space assessment models by treating movement as random effects with prior information (i.e. $\mathrm{SpE}_{\mathrm{NAA,Est}}$) could be a promising solution to better meet management objectives. This approach showed improved accuracy in parameter estimation and yielded more favorable management outcomes compared to EMs that excluded movement. However, assessment models estimating movement often rely on strong priors, which may be challenging to obtain due to the limited availability of tagging, genetic, or stable isotope data. While testing the effects of prior distributions on movement parameter estimation is beyond the scope of this study, we recommend that future research investigate the sensitivity of model performance to varying qualities of prior information in movement estimation.

4.4 Caveats and Future Research Recommendations

We note that several assumptions made in our study could influence the overall findings. For instance, all fish were assumed to be reproductively isolated and to exhibit natal homing spawning migrations. This assumption ensured that surveys conducted during the spawning season had the best opportunity to collect the information of SSB for each population. However, mismatches between the timing of surveys and the spawning season could potentially affect management outcomes. Second, movement rates were assumed to be uniform across all ages. While ontogenetic movement may be biologically plausible in certain cases (Waterhouse et al., 2014; Hanselman et al., 2015), it remains unclear whether a more complex structure for NAA random effects (e.g., age-specific movement) could effectively capture differences in age-based movement dynamics. Further investigation is needed to assess the broader applicability of NAA random effects under such scenarios.

Another key assumption was that selectivity pattern in our estimation models was constrained to a logistic curve. While this assumption provided NAA random effects with the best opportunity to account for spatial heterogeneity caused by movement between regions, it restricted the ability of selectivity patterns to reflect movement dynamics, which are often characterized by a dome-shaped selectivity pattern in EMs that do not account for movement (O'Boyle et al., 2016). Moreover, incorporating random effects into the selectivity process may offer greater flexibility to account for the underlying movement dynamics, likely maximizing the benefits of using selectivity to absorb the movement process variation, despite a potential increased risk of bias in population estimates and reference points due to model misspecification (Fisch et al., 2023; Li et al., 2024). Thus, the management outcomes of these more flexible frameworks within state-space models remain poorly understood and require further investigation (O'Boyle et al., 2016).

Our approach to biological reference points and catch apportionment may also influence MSE performance by affecting management outcomes rather than assessment bias. Apportioning global biological reference points to regions introduces inherent challenges in comparing single-region and multi-region model outputs or harvest control rules. Thus, further work is needed to explore pros and cons of using regionally explicit versus globally apportioned biological reference points (e.g., separate vs. spatially disaggregated approaches), particularly under varying assumptions of movement dynamics among regions.

Finally, fish biology (stock-recruitment relationship) and fleet characteristics (e.g., fishing selectivity) were assumed to be slightly different. However, relative impacts of these discrepancies, along with other biological and fishery characteristics, on model performance remain unclear and warrant further investigation to fully understand the potential of random effects in management outcomes. Our results are indicative of data-rich stocks with a known source for connectivity dynamics (natal homing), while alternative data quality and population dynamic assumptions may improve or degrade the utility of NAA random effects.

5 Conclusion

While claims have been made that NAA random effects can account for other processes including movement, our study found that assessment models with NAA random effects outperformed their counterparts without NAA random effects. However, NAA random effects could not fully account for movement dynamics, particularly under high levels of movement. Thus, NAA random effects can be considered an intermediate estimation approach when explicitly modeling movement dynamics is not possible. The results of this study add to the growing body of literature suggesting that incorporating NAA random effects should be the default starting point in state-space stock assessments and that the addition leads to more accurate stock assessment and management relevant estimates.

6 Acknowledgements

This work was funded by NOAA Fisheries Northeast Fisheries Science Center (50%), Northwest Fisheries Science Center (25%), and Alaska Fisheries Science Center (25%). We acknowledge the support provided by Saltwater Inc. for facilitating Chengxue Li's contribution to this project. We are grateful to Christopher Legault for conducting NOAA internal review of the manuscript and to Dana Hanselman for his insightful contributions to the conceptualization of this study design. We also thank Microsoft Azure for providing high-performance computing resources for this study. We would also like to thank the Quahog Republic and the Midway Trap and Skeet Club for keeping the wheels adequately greased during the study design development.

7 Competing interests statement

One co-author, Daniel R. Goethel, serves as an Associate Editor for CJFAS, and another co-author, Timothy J. Miller, serves as a Guest Editor for CJFAS for this special issue.

8 CRediT authorship contribution statement

Chengxue Li: Conceptualization, Methodology, Software, Writing - original draft, Formal analysis, Visualization.

Jonathan J. Deroba: Conceptualization, Funding acquisition, Supervision, Writing - review & editing.

Aaron M. Berger: Conceptualization, Writing - review & editing.

Brian J. Langseth: Conceptualization, Writing - review & editing.

Daniel R. Goethel: Conceptualization, Writing - review & editing.

Amy M. Schueller: Conceptualization, Writing - review & editing.

Timothy J. Miller: Software, Writing - review & editing.

9 Funding statement

This work was funded by NOAA Fisheries Northeast Fisheries Science Center (50%), Northwest Fisheries Science Center (25%), and Alaska Fisheries Science Center (25%).

10 Data availability statement

No empirical data were collected for this study. Data used for this study were produced through simulation. Code for all simulations is available at https://github.com/lichengxue/MSE_NAA_Project.

11 Tables

Table 1. Biological and fishery parameter values used in the operating models (OM_{high} and OM_{low}). When a single value is provided, it applies to both populations/regions. When two values are provided, the first corresponds to either population 1 (bold) or region 1, and the second value corresponds to either population 2 (bold) or region 2. Settings for the closed-loop management strategy evaluation are also provided.

Parameters	Description	Value
Biological information		
α_{SR}	Productivity parameter	**7, 5.5**
β_{SR}	Density-dependent recruitment parameter	1×10^{-4}
σ_R	Standard deviation of recruitment random effects (age 1)	0.5

σ_{NAA}	Standard deviation of numbers-at-age (age 2+) random effects	0.2
a_{max}	Maximum age (yr)	12
M	Instantaneous natural mortality	0.2
a_0	Age at length = 0	0
L_∞	Maximum length (cm)	90
k	Growth rate (yr^{-1})	0.13
b_1	L – W scalar (kg cm^{-1})	3×10^{-6}
b_2	L – W exponent	3
m_{50}	Age at 50% maturity	3.5
m_{slope}	Slope of maturity	1
Spatial structure and movement dynamics		
n_R	Number of regions	2
n_{season}	Number of seasons	4
f_{SSB}	Fraction of the year that passes before spawning occurs	0.625
μ_1	Movement rate outside the spawning and spawning migration seasons	0.3, 0.1 ($\mathrm{OM_{high}}$); **0.1, 0.0 ($\mathbf{OM_{low}}$)**
μ_2	Movement rate during the spawning migration season	1
μ_3	Movement rate during the spawning season	0
σ_μ	Standard deviation of movement rate	0.1
ρ_μ	Autocorrelation in movement	0.5
Fishery information		
n_F	Number of fleets in each region	1
γ_{50}	Age at 50% selectivity in fishery	5.5, 5
γ_{slope}	Slope of logistic selectivity for fishery	1
ESS_C	Effective sample size for fleet catch	100
CV_C	Coefficient of variation for fleet catch	0.1

LL_C	Likelihood for age compositional data	Multinomial
Survey information		
n_I	Number of indices in each region	1
ζ_{50}	Age at 50% selectivity in survey	2
ζ_{slope}	Slope of logistic selectivity for survey	1
ESS_I	Effective sample size for survey index	100
CV_I	Coefficient of variation for survey index	0.1
q	Survey catchability	0.2
f_I	Proportion of the year elapsed when index is observing the population	0.625
LL_I	Likelihood for age compositional data	Multinomial
MSE closed-loop		
$\tau_{hist.}$	Number of years in the historical period	30
$\tau_{feedback}$	Number of years in the feedback period	30
Δt	Assessment interval (yr)	3
A_{total}	Number of total assessments	10
Φ	A fraction of $F_{40\%}$ used as the harvest control rule	75%

Table 2. Description of fishing scenarios in the historical period used for operating models ($\mathrm{OM_{high}}$ and $\mathrm{OM_{low}}$). The "Adaptive" scenario ($\mathrm{F_A}$) represents fishing mortality rates for both regions transitioning from high to sustainable levels ($2.5F_{MSY}$–F_{MSY}). The "Unsustainable" scenario ($\mathrm{F_U}$) represents constantly high fishing mortality rates ($2.5F_{MSY}$), and the "Contrasting" scenario ($\mathrm{F_C}$) represents contrasting fishing mortality rates in each region, with one region experiencing sustainable fishing (F_{MSY}) and the other unsustainable ($2.5F_{MSY}$).

Fishing scenarios	**Region 1**	**Region 2**
Adaptive ($\mathrm{F_A}$)	H – L ($2.5F_{MSY} - F_{MSY}$)	H – L ($2.5F_{MSY} - F_{MSY}$)
Unsustainable ($\mathrm{F_U}$)	H ($2.5F_{MSY}$)	H ($2.5F_{MSY}$)
Contrasting ($\mathrm{F_C}$)	H – L ($2.5F_{MSY} - F_{MSY}$)	L – H ($F_{MSY} - 2.5F_{MSY}$)

Table 3. Descriptions of 10 estimation models

Model	**Type**	**Move**	**Random effects**	**Description**
$\mathrm{PAN}_{\mathrm{NAA}}$	Panmictic (catch aggregated)	No	NAA, Rec	Catch data aggregated across regions
$\mathrm{PAN}_{\mathrm{noNAA}}$	Panmictic (catch aggregated)	No	**Rec only**	Catch data aggregated across regions
$\mathrm{FAA}_{\mathrm{NAA}}$	Spatially implicit (fleets-as-areas)	No	NAA, Rec	Multiple fleets account for spatial difference in fleet structure
$\mathrm{FAA}_{\mathrm{noNAA}}$	Spatially implicit (fleets-as-areas)	No	**Rec only**	Multiple fleets account for spatial difference in fleet structure
$\mathrm{SEP}_{\mathrm{NAA}}$	Separate	No	NAA, Rec	Separate single stock assessment model
$\mathrm{SEP}_{\mathrm{noNAA}}$	Separate	No	**Rec only**	Separate single stock assessment model
$\mathrm{SpD}_{\mathrm{NAA}}$	Spatially disaggregated	No	NAA, Rec	Spatially disaggregated without movement
$\mathrm{SpD}_{\mathrm{noNAA}}$	Spatially disaggregated	No	**Rec only**	Spatially disaggregated without movement
$\mathrm{SpE}_{\mathrm{NAA,Est}}$	Spatially explicit	Yes (Est)	NAA, Rec, Move	Spatially explicit with movement estimated using a prior
$\mathrm{SpE}_{\mathrm{NAA,Fix}}$	Spatially explicit	Yes (Fix)	NAA, Rec	Spatially explicit with movement fixed as known

615 # 12 Figures

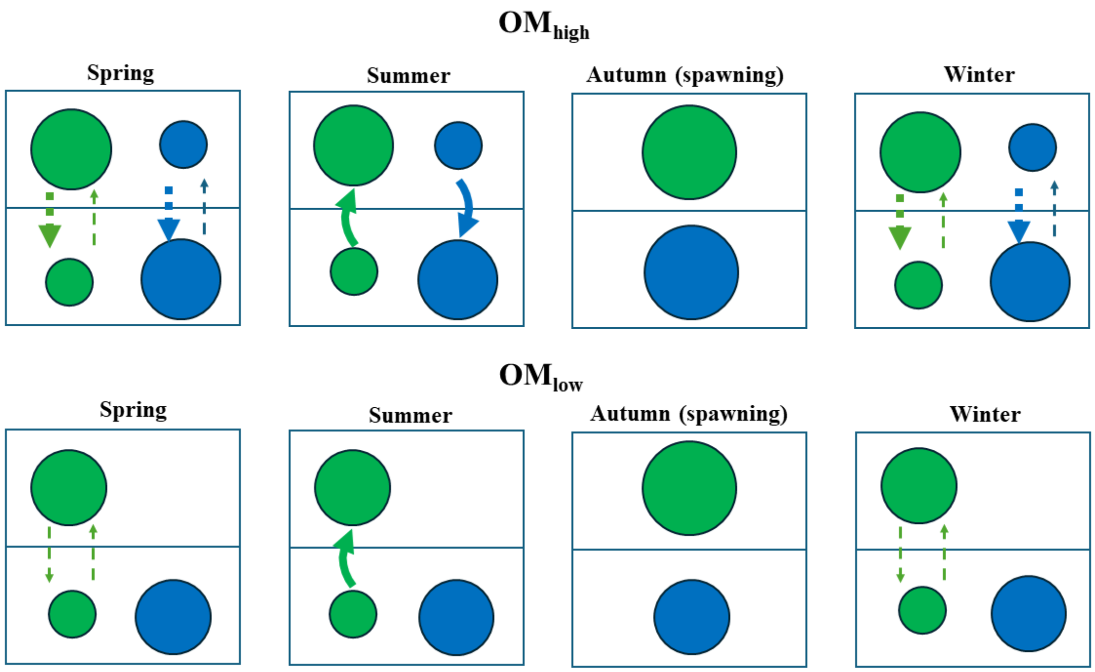

Figure. 1. Movement dynamics for high (OM$_{high}$) and low (OM$_{low}$) movement scenarios. Blue represents the population originating from the upper region and red represents the population originating from the lower region. The dashed arrow represents movement outside the spawning season, with thickness indicating the magnitude of movement rate, while the solid curved arrow indicates natal homing movement.

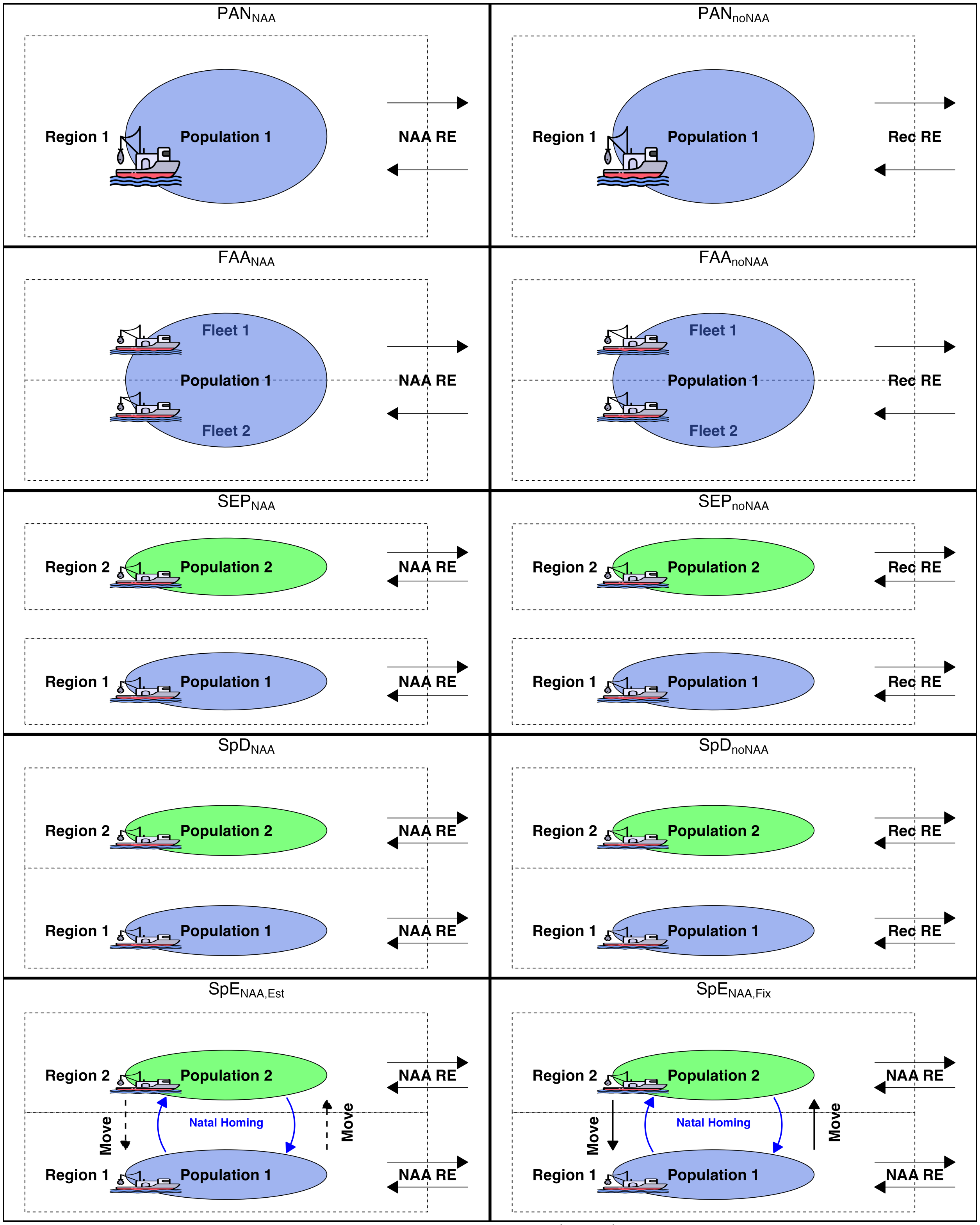

Figure. 2. Overview of the 10 estimation models (EMs) used in the present study. The description of the model structure is shown in Table 2. Note that the dashed black arrow in the $SpE_{NAA,Est}$ EM indicates the movement outside the spawning season is estimated using a prior, while the solid black arrow in the $SpE_{NAA,Fix}$ EM indicates the movement outside the spawning season is fixed at the true value. The blue curve arrow in $SpE_{NAA,Est}$ and $SpE_{NAA,Fix}$ EMs indicates the natal homing movement.

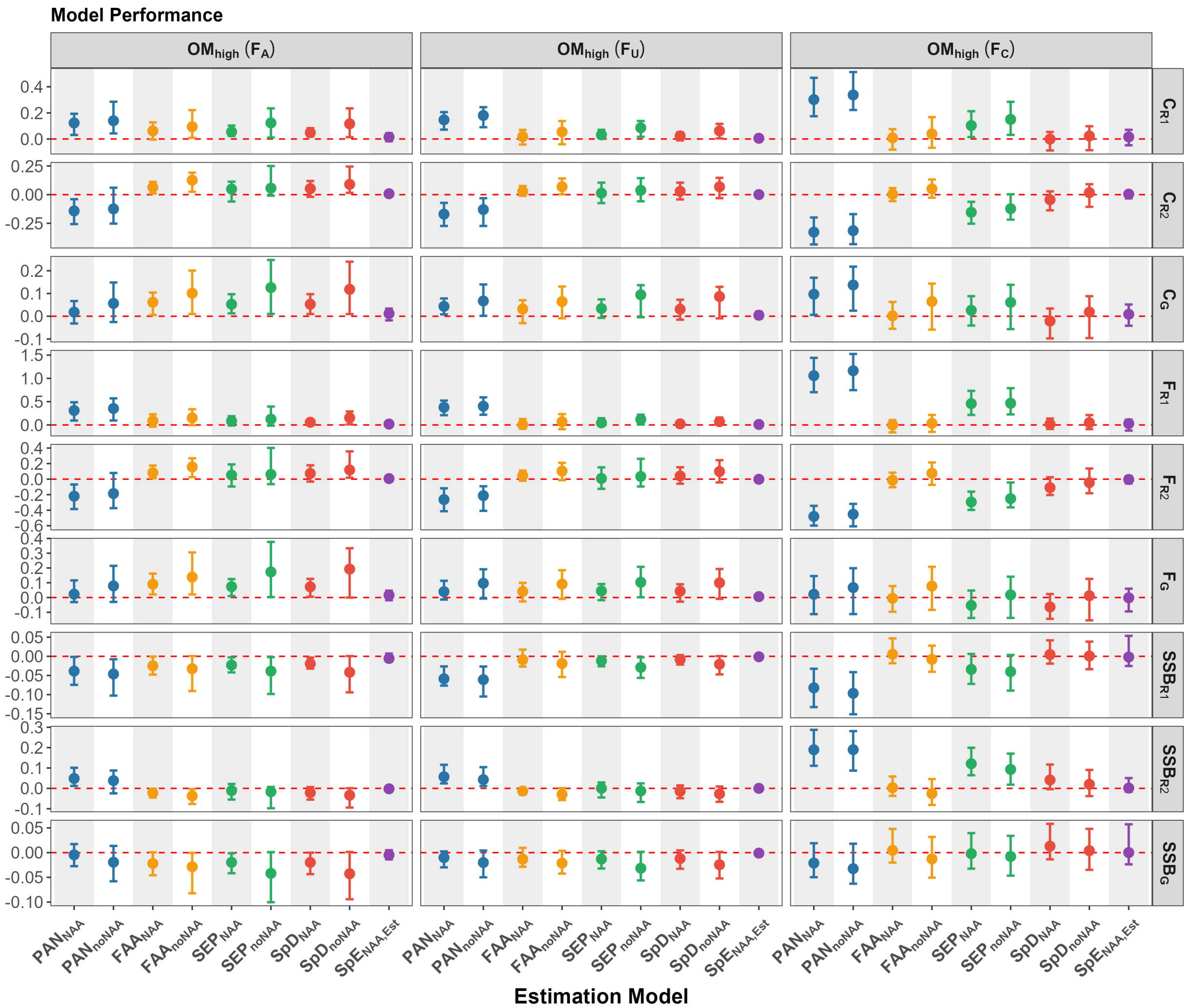

Figure. 3. Relative difference in catch, F, and SSB between the baseline EM ($\mathrm{SpE}_{\mathrm{NAA,Fix}}$) and other EMs over the first 5 years of the feedback period under the high movement scenario ($\mathrm{OM}_{\mathrm{high}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

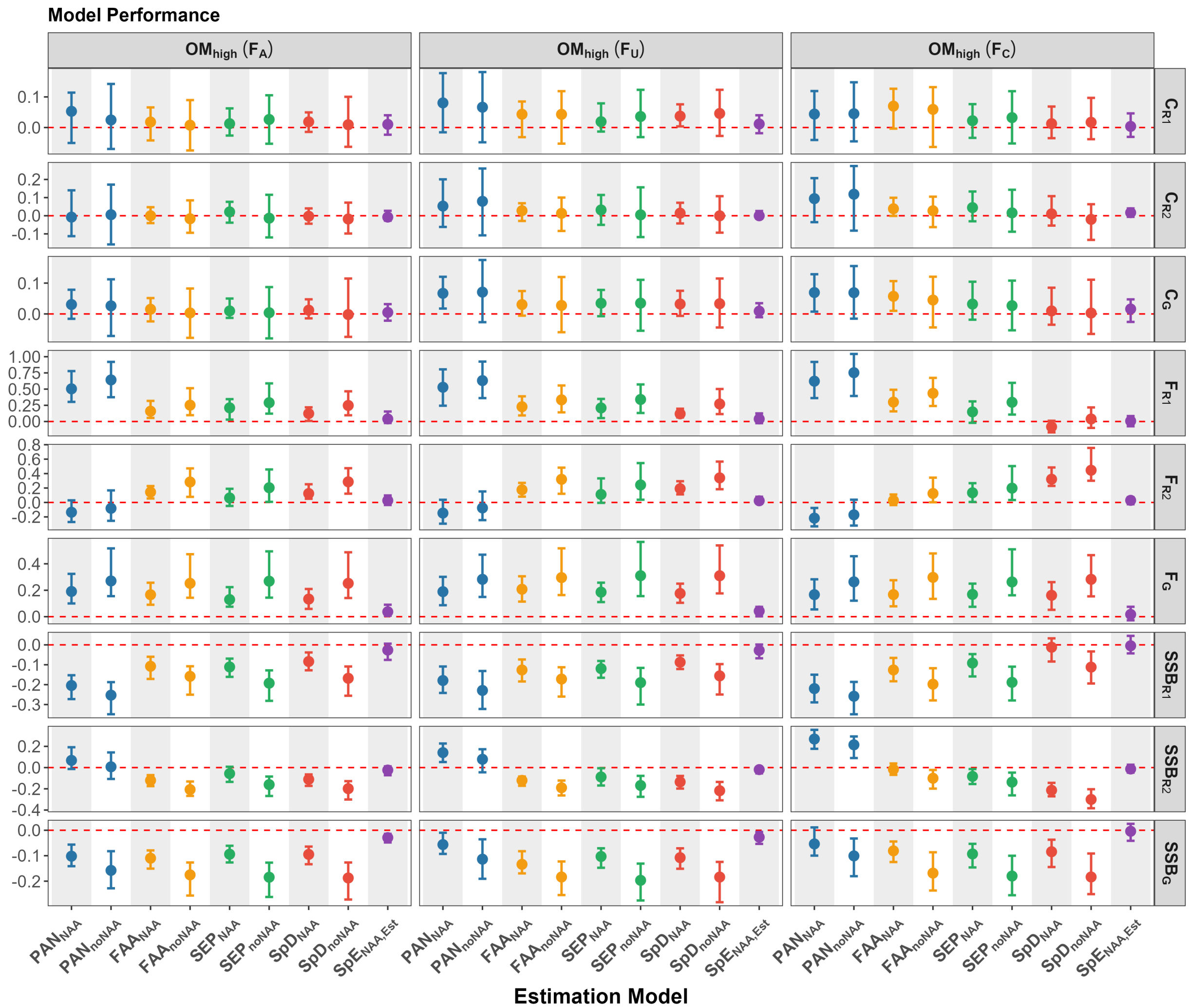

Figure. 4. Relative difference in catch, F, and SSB between the baseline EM ($\mathrm{SpE}_{\mathrm{NAA,Fix}}$) and other EMs over the last 5 years of the feedback period under the high movement scenario ($\mathrm{OM}_{\mathrm{high}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

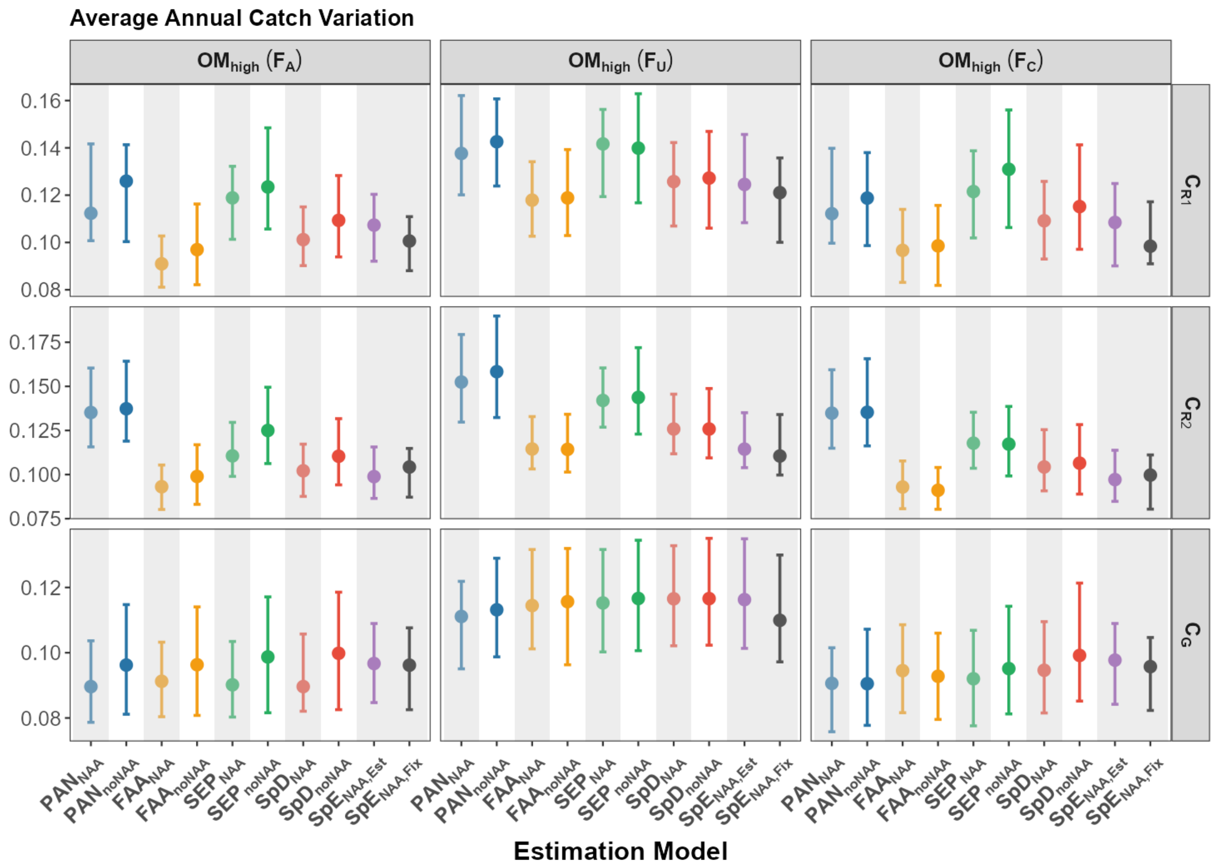

Figure. 5. Average annual catch variation (AACV) of each EM at both regional and global scales under the high movement scenario (OM_{high}). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

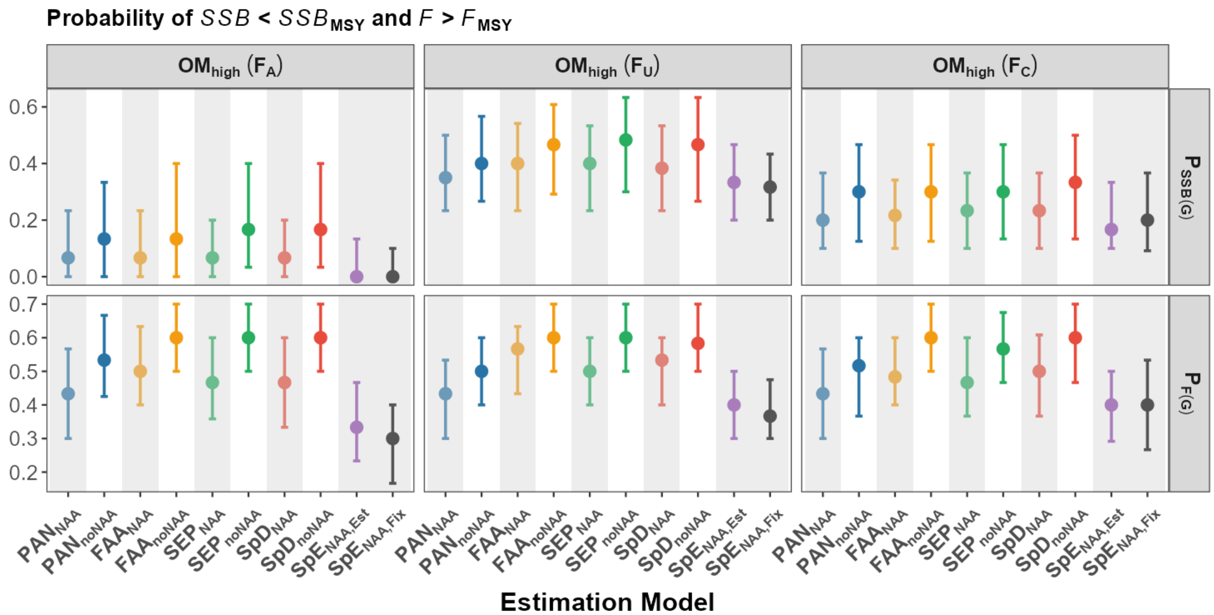

Figure. 6. Probability of $SSB < SSB_{MSY}$ and $F > F_{MSY}$ for each EM over the 30-year feedback period under the high movement scenario ($\mathrm{OM_{high}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

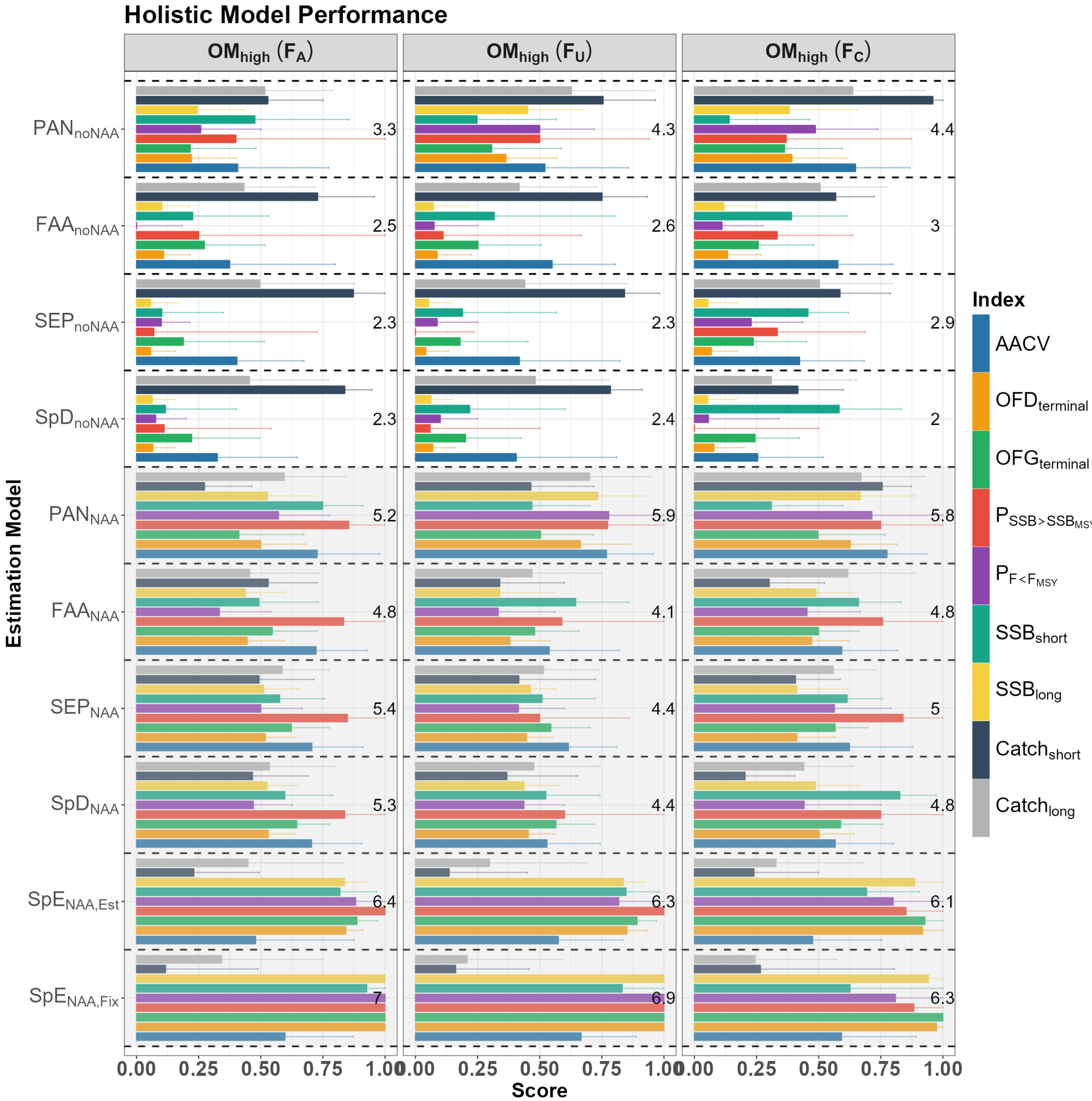

Figure. 7. Overall performance of each EM at the global scale under the high movement scenario ($\mathrm{OM_{high}}$), including the following relative performance metrics: 1) average annual catch variation (AACV); 2) overfished status in the terminal year ($\mathrm{OFD_{terminal}}$); 3) overfishing status in the terminal year ($\mathrm{OFG_{terminal}}$); 4) probability of $SSB > SSB_{MSY}$ ($\mathrm{P_{SSB>SSB_{MSY}}}$); 5) probability of $F < F_{MSY}$ ($\mathrm{P_{F<F_{MSY}}}$); 6) short-term SSB ($\mathrm{SSB_{short}}$); 7) long-term SSB ($\mathrm{SSB_{long}}$), 8) short-term catch ($\mathrm{Catch_{short}}$); and 9) long-term catch ($\mathrm{Catch_{long}}$). All the indices were standardized to scores between 0 and 1, with higher values indicating better performance. The total score for each EM was provided for each fishing scenario. The black dashed line separates the performance of each EM. EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

13 Supplementary file

Methods

Population Dynamics

For population s, the number-at-age a that survived ($N_{S,s,a,t+\delta}$), were harvested from the fishery ($N_{H,s,a,t+\delta}$), and died from natural mortality ($N_{D,s,a,t+\delta}$) over the time interval δ are given by:

$$\begin{aligned} N_{S,s,a,t+\delta} &= N_{s,a,t}S(s,a,\delta) = N_{s,a,t}e^{-Z_{s,a}\delta} \\ N_{H,s,a,t+\delta} &= N_{s,a,t}H(s,a,\delta) = N_{s,a,t}\frac{F_{s,a}}{Z_{s,a}}\left(1-e^{-Z_{s,a}\delta}\right) \\ N_{D,s,a,t+\delta} &= N_{s,a,t}D(s,a,\delta) = N_{s,a,t}\frac{M_{s,a}}{Z_{s,a}}\left(1-e^{-Z_{s,a}\delta}\right) \end{aligned} \tag{1}$$

where $N_{s,a,t}$ is the number-at-age a at time t for population s; S, H, and D represent the proportions (fractions) of individuals that survived, were harvested, or died from natural mortality, respectively, over the time interval δ; $F_{t,a}$ and $M_{t,a}$ denote instantaneous rates of fishing mortality and natural mortality at age a and time t, and $Z_{s,a}$ denotes total mortality rate, where $Z_{s,a} = F_{s,a} + M_{s,a}$.

The probability transition matrix, when there is movement between regions, can be generalized as:

$$\mathbf{P}_{s,a,t,\delta} = \begin{bmatrix} \mathbf{O}(s,a,\delta) & \mathbf{H}(s,a,\delta) & \mathbf{D}(s,a,\delta) \\ 0 & \mathbf{I}_H & 0 \\ 0 & 0 & \mathbf{I}_D \end{bmatrix} \tag{2}$$

where $\mathbf{O}$ is the $n_R \times n_R$ matrix representing the combined probability of survival and movement between regions. Assuming movement occurs sequentially after survival, $\mathbf{O}$ is the product of the survival probability matrix ($\mathbf{S}$) and movement probability matrix μ, such that $\mathbf{O} = \mathbf{S}\mu$. $\mathbf{H}$ is the $n_R \times n_F$ matrix representing proportions of individuals harvested by each fleet in each region, $\mathbf{D}$ is the $n_R \times n_R$ matrix representing proportions of individuals that died from natural mortality within each region, and $\mathbf{I}_H$ and $\mathbf{I}_D$, which have the same dimensions as $\mathbf{H}$ and $\mathbf{D}$), are identity matrices that ensure the overall dimensions of the probability transition matrix remain consistent when combined. If there is only one region and one fleet, these identity matrices reduce to scalars of 1.

For population s, the total number-at-age a at the end of time interval δ can be generalized as:

$$\mathbf{N}_{s,a,t+\delta} = \mathbf{P}(s,a,\delta)'\mathbf{N}_{s,a,t} \tag{3}$$

where the numbers of individuals that survived, were harvested, and died from natural mortality are decomposed as:

$$\begin{aligned}\mathbf{N}_{S,s,a,t+\delta} &= \mathbf{O}(s,a,\delta)'\mathbf{N}_{s,a,t}\\ \mathbf{N}_{H,s,a,t+\delta} &= \mathbf{H}(s,a,\delta)'\mathbf{N}_{s,a,t}\\ \mathbf{N}_{D,s,a,t+\delta} &= \mathbf{D}(s,a,\delta)'\mathbf{N}_{s,a,t}\end{aligned} \tag{4}$$

Population Structure and Seasonality

A natal homing population structure is assumed in Multi-WHAM. Each population is associated with a natal region, and all fish originating in that region must return there for spawning. Outside of the spawning season, fish can move between regions (see details in the sections on OM or EM movement dynamics). Seasonality is included to support the assumption of natal homing. Specifically, a number of time intervals within a year ($\Sigma_{i=1}^{n}\delta_i = 1$) are used to indicate movement dynamics prior to the spawning season, during the spawning season, and after the spawning season. The probability transition matrix for the entire year is a product of all "seasonal" matrices:

$$\mathbf{P}_{s,a,y} = \prod_{t=1}^{n} \mathbf{P}_{s,a,y,t} \tag{5}$$

Fishing Mortality (F)

F at age a for fleet f is calculated as a product of fully selected F and selectivity-at-age (i.e., the separability assumption) for fleet f: $\tilde{F}_{f,a} = \tilde{F}Sel_{f,a}$. In Multi-WHAM, a fully selected, total F is the maximum of the total F-at-age summed across all fleets:

$$\begin{aligned}F_{\text{total},a,y} &= \sum_{i=1}^{n_f} F_{f_i,a,y}\\ F_{\text{total},y} &= \operatorname*{argmax}_a F_{\text{total},a,y}\end{aligned} \tag{6}$$

Reference Points

The fully selected total F defined above was used to estimate a single universal F reference point applicable to all populations and regions. A 'static' SPR-based reference point was calculated by averaging all of the inputs (e.g., selectivity, natural mortality, movement, weight-at-age, maturity-at-age) over the most recent 5 model years for each region, except for recruitment, which was averaged over all model years for each region.

For $X\%$ SPR-based reference points, Multi-WHAM uses a Newton method to solve for the universal $F_{X\%}$ reference point:

$$\tilde{F}^{(i)} = \tilde{F}^{(i-1)} - \frac{g\left(\tilde{F}^{(i-1)}\right)}{\tilde{F}^{(i-1)} g'\left(\tilde{F}^{(i-1)}\right)} \tag{7}$$

where $g(F)$ is the difference between the weighted spawning biomass per recruit at F and $X\%$ of the weighted unfished spawning biomass per recruit:

$$g(F) = \sum_{s=1}^{S} w_s \left(\frac{S\tilde{\tilde{S}}B}{\tilde{R}(F)} - \frac{X}{100} \frac{S\tilde{\tilde{S}}B}{\tilde{R}(F=0)} \right) \tag{8}$$

Here weights are determined by the average recruitment for different regions/populations:

$$w_i = \frac{\bar{R}_i}{\sum_{i=1}^{n_s} \bar{R}_i} \tag{9}$$

The fully selected total F reference point can be disaggregated to regional F reference points, given the relative contribution of selectivity (Sel) at age a from each fleet in each region R:

$$F_{X\%_{a,R}} = F_{X\%} Sel_{a,R} \tag{10}$$

The regional $F_{X\%}$ (e.g., $F_{40\%}$) does not guarantee achieving 40% of region-specific SPR, but serve to collectively achieve 40% of SPR at the global scale. Regional F reference points are not currently implemented, as a unique solution is not guaranteed in the presence of movement between regions. An analogous Newton method is used to solve for F that maximizes yield for MSY-based reference points.

Table S1. Biological reference point (BRP) and harvest control rule (HCR) used for 10 estimation models.

Model	**Global BRP**	**Rec Weights**	**Catch Apportionment**
$\mathrm{PAN}_{\mathrm{NAA}}$	Yes	No	Regional survey biomass
$\mathrm{PAN}_{\mathrm{noNAA}}$	Yes	No	Regional survey biomass
$\mathrm{FAA}_{\mathrm{NAA}}$	Yes	No	Disaggregated regional F from global F
$\mathrm{FAA}_{\mathrm{noNAA}}$	Yes	No	Disaggregated regional F from global F
$\mathrm{SEP}_{\mathrm{NAA}}$	No	No	Independent regional F
$\mathrm{SEP}_{\mathrm{noNAA}}$	No	No	Independent regional F
$\mathrm{SpD}_{\mathrm{NAA}}$	Yes	Yes	Disaggregated regional F from global F
$\mathrm{SpD}_{\mathrm{noNAA}}$	Yes	Yes	Disaggregated regional F from global F
$\mathrm{SpE}_{\mathrm{NAA,Est}}$	Yes	Yes	Disaggregated regional F from global F
$\mathrm{SpE}_{\mathrm{NAA,Fix}}$	Yes	Yes	Disaggregated regional F from global F

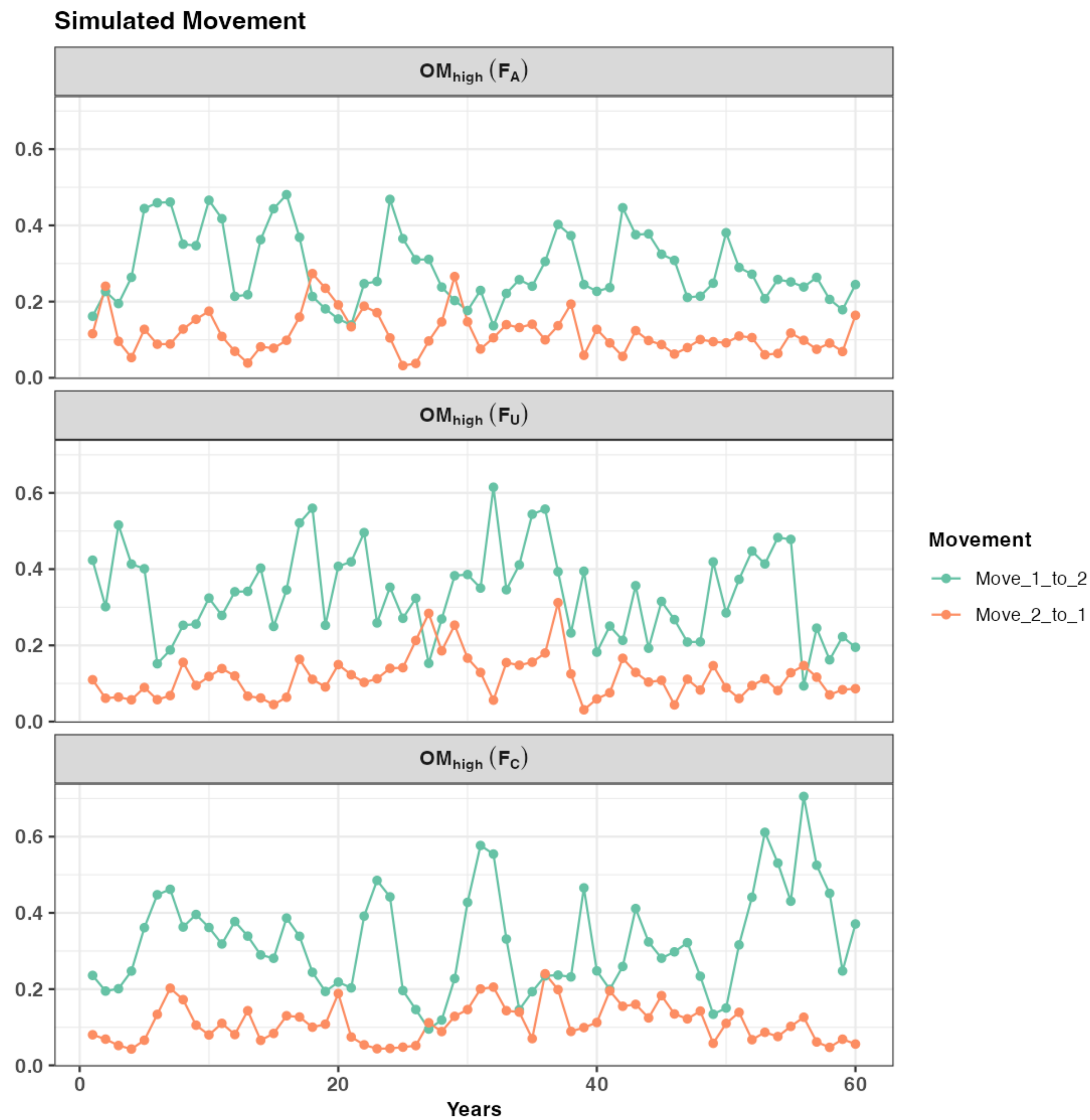

Figure. S1. Simulated AR1 movement rates (from one replicate) under the high movement scenario ($\mathrm{OM}_{\mathrm{high}}$).

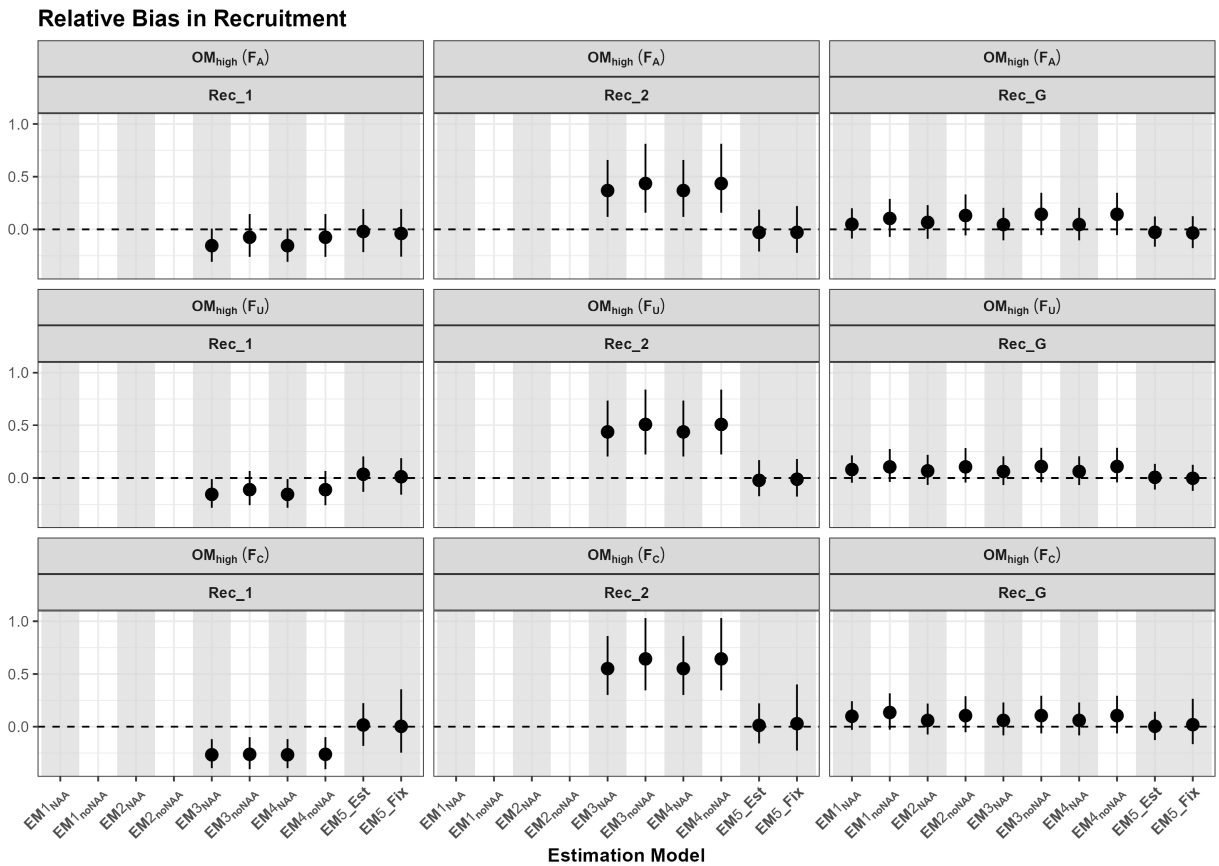

Figure. S2. Relative bias in recruitment for region 1, region 2, and global recruitment from the first assessment model during the feedback period under the high movement scenario ($\mathrm{OM_{high}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

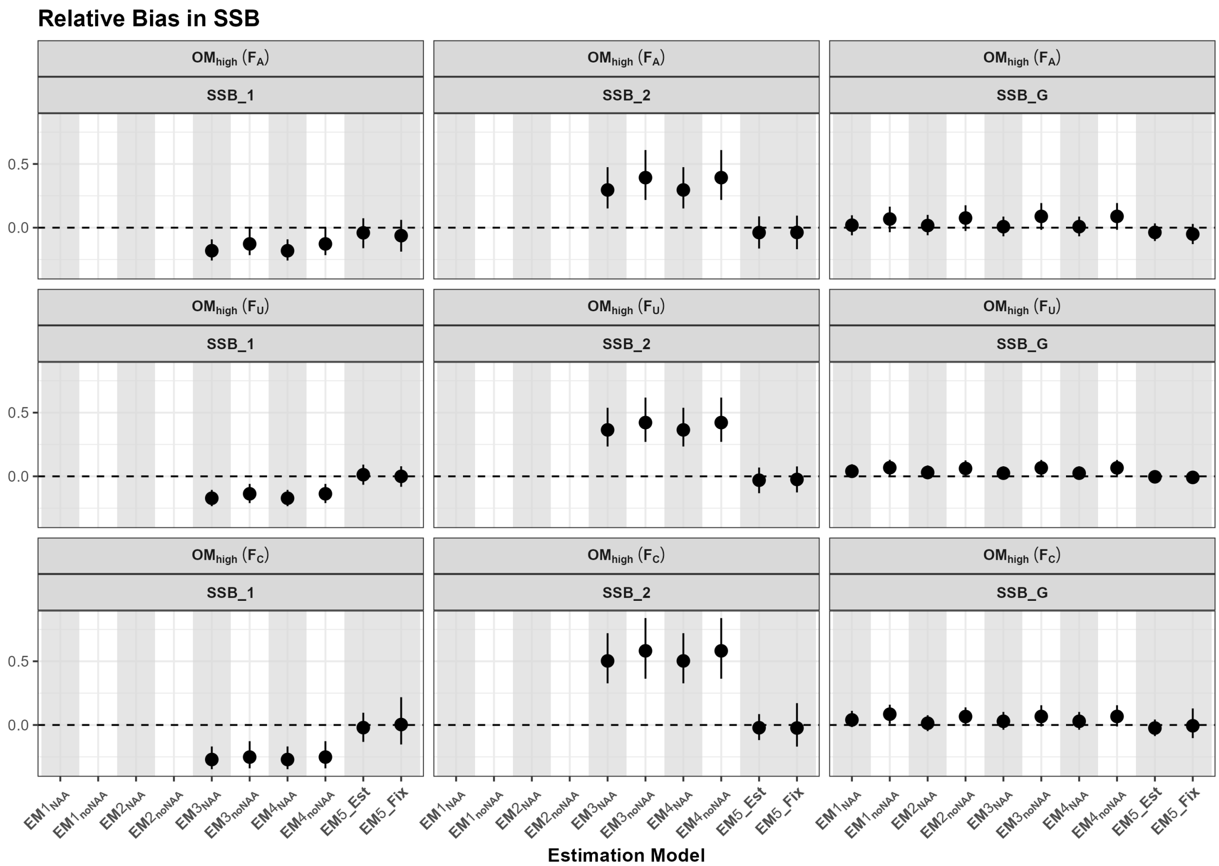

Figure. S3. Relative bias in SSB for region 1, region 2, and global SSB from the first assessment model during the feedback period under the high movement scenario ($\mathrm{OM}_{\mathrm{high}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

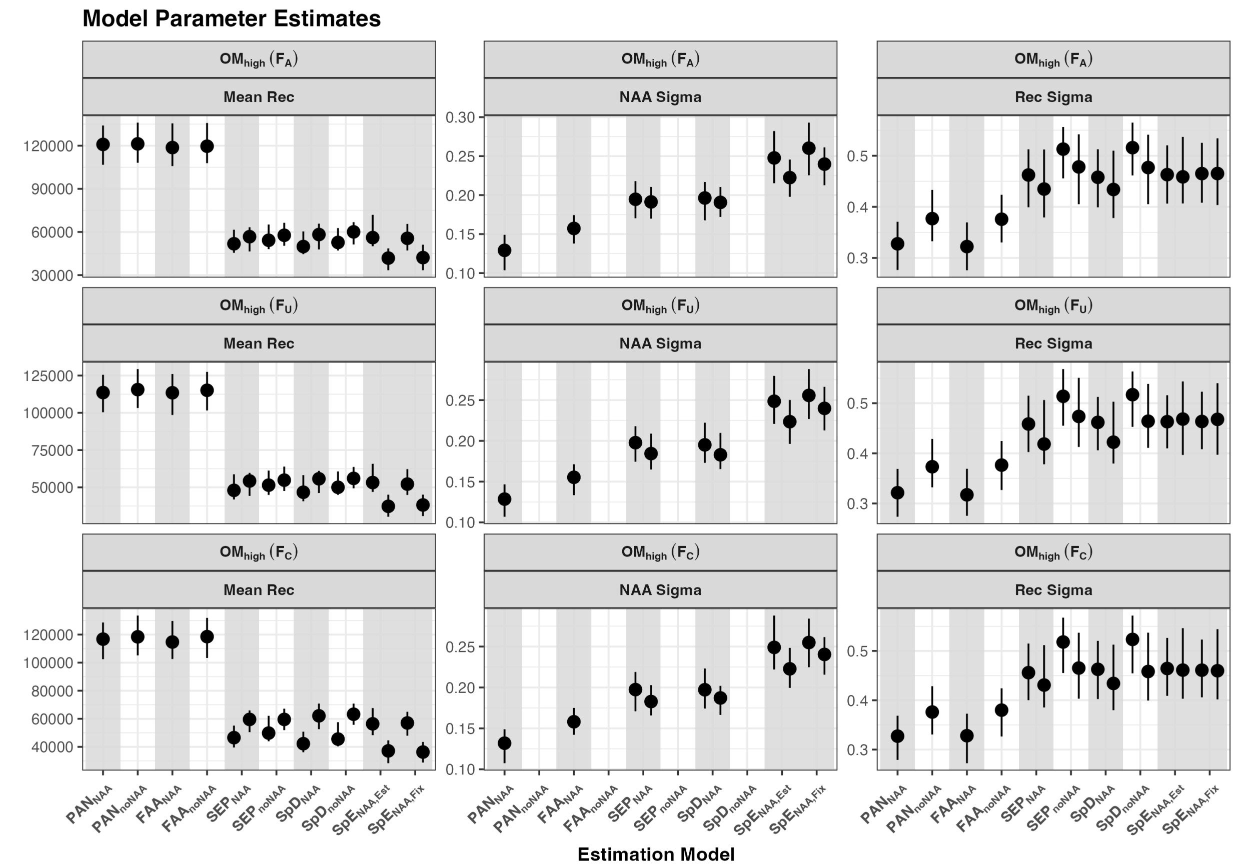

Figure. S4. The mean recruitment parameter, recruitment standard deviation and NAA standard deviation from the last assessment model during the feedback period under the high movement scenario (OM_{high}). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

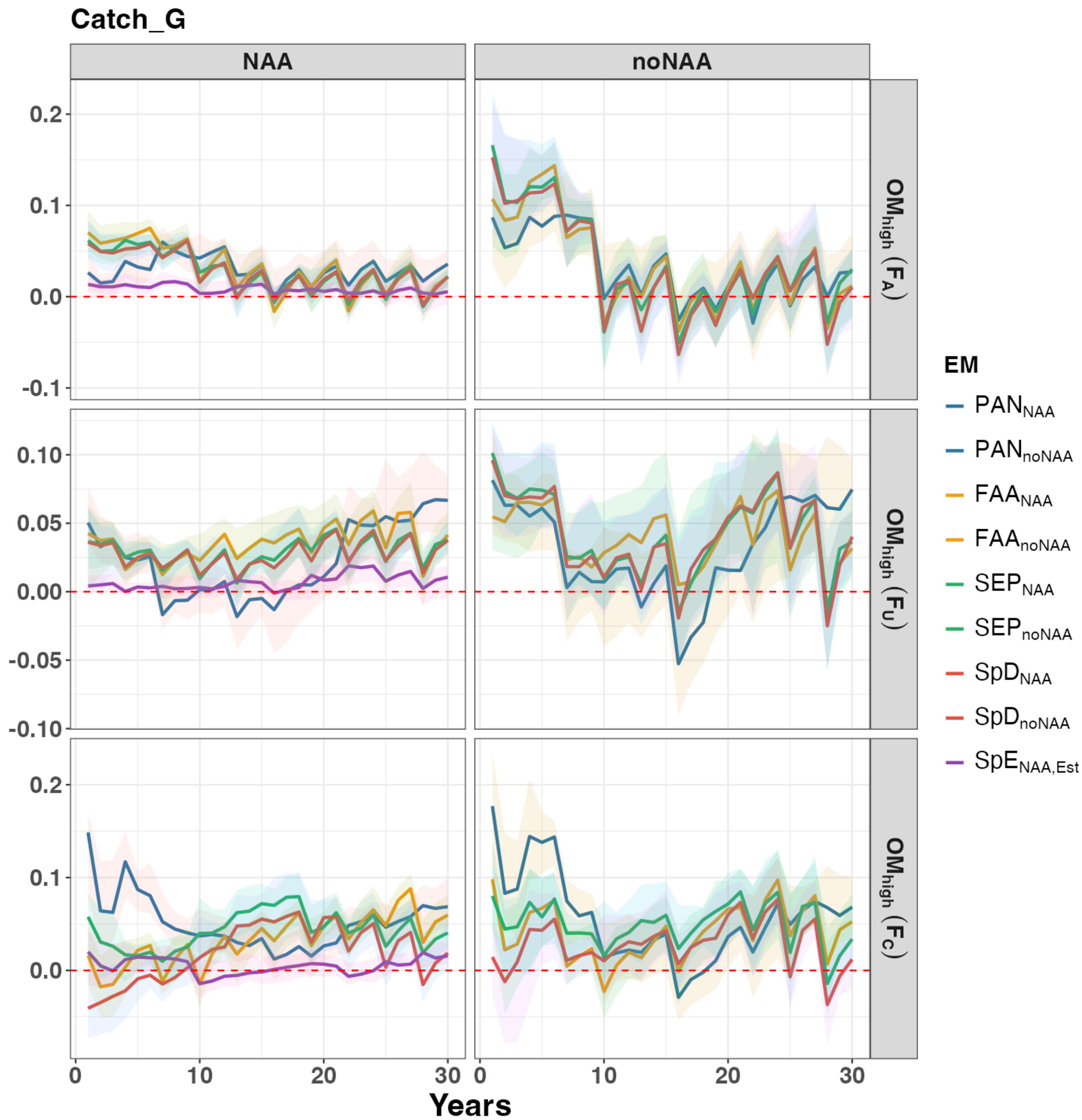

Figure. S5. Relative difference in annual total catch between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the high movement scenario (OM_{high}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

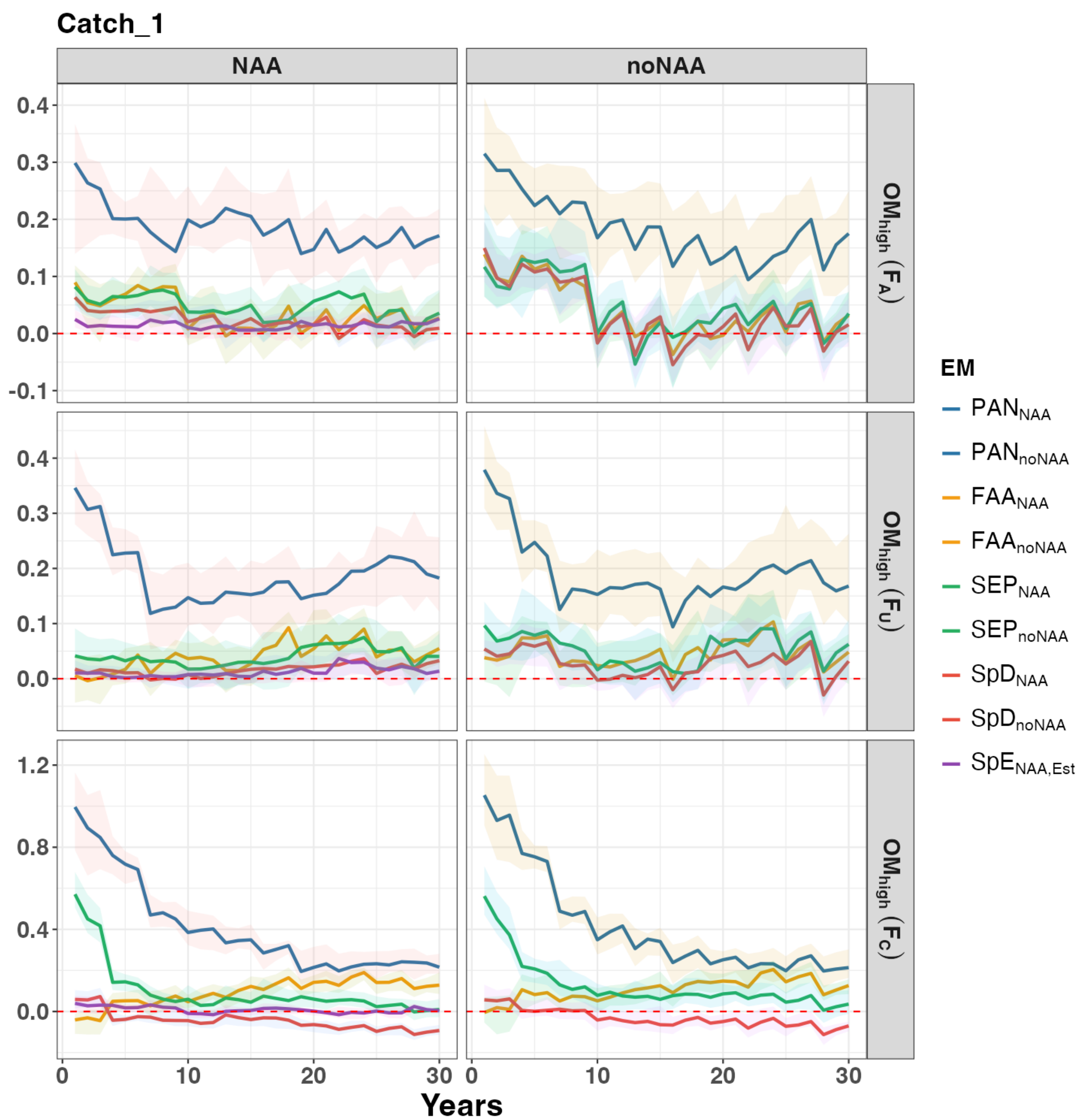

Figure. S6. Relative difference in annual catch in region 1 between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the high movement scenario (OM_{high}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

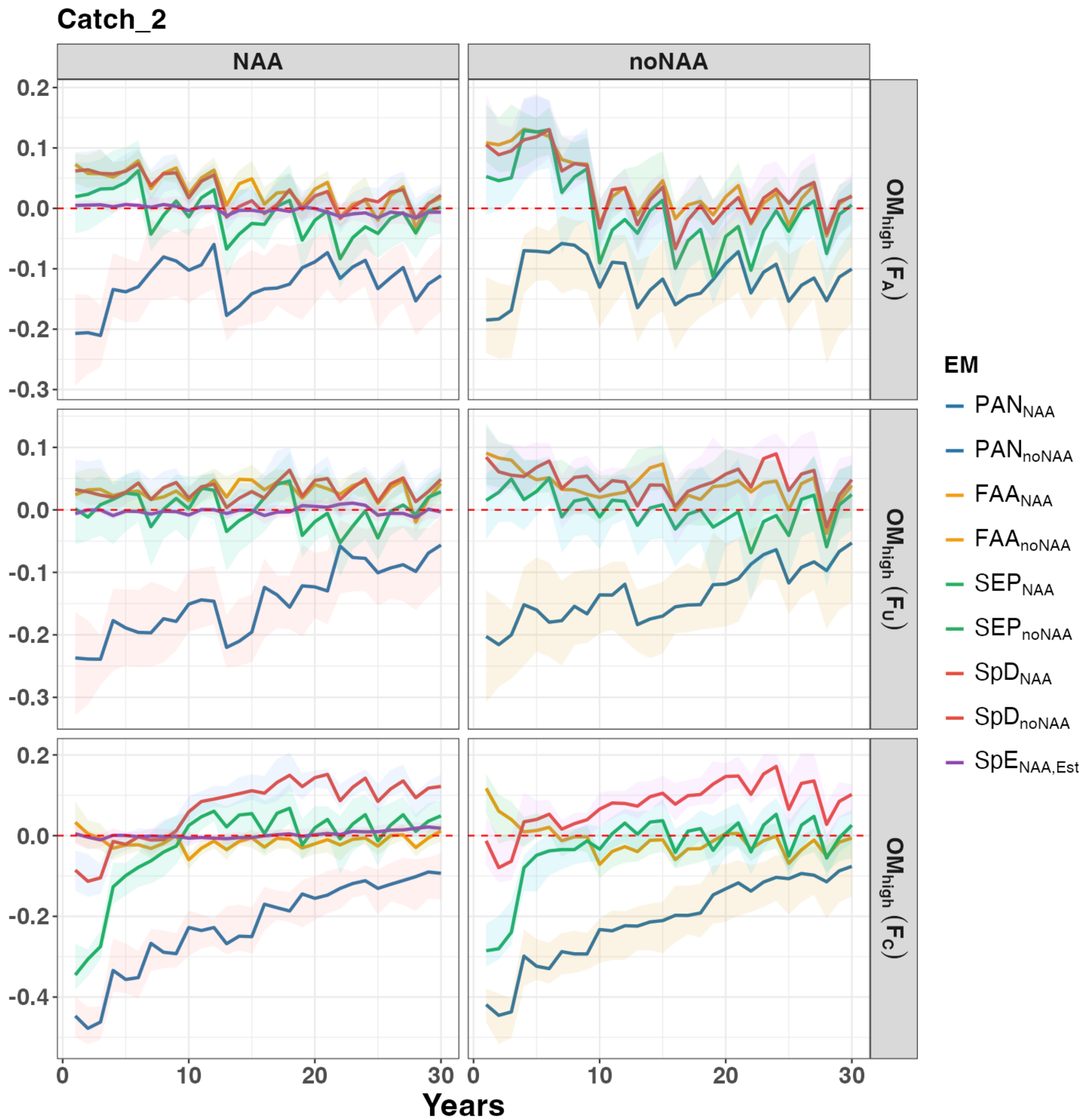

Figure. S7. Relative difference in annual catch in region 2 between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the high movement scenario (OM_{high}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

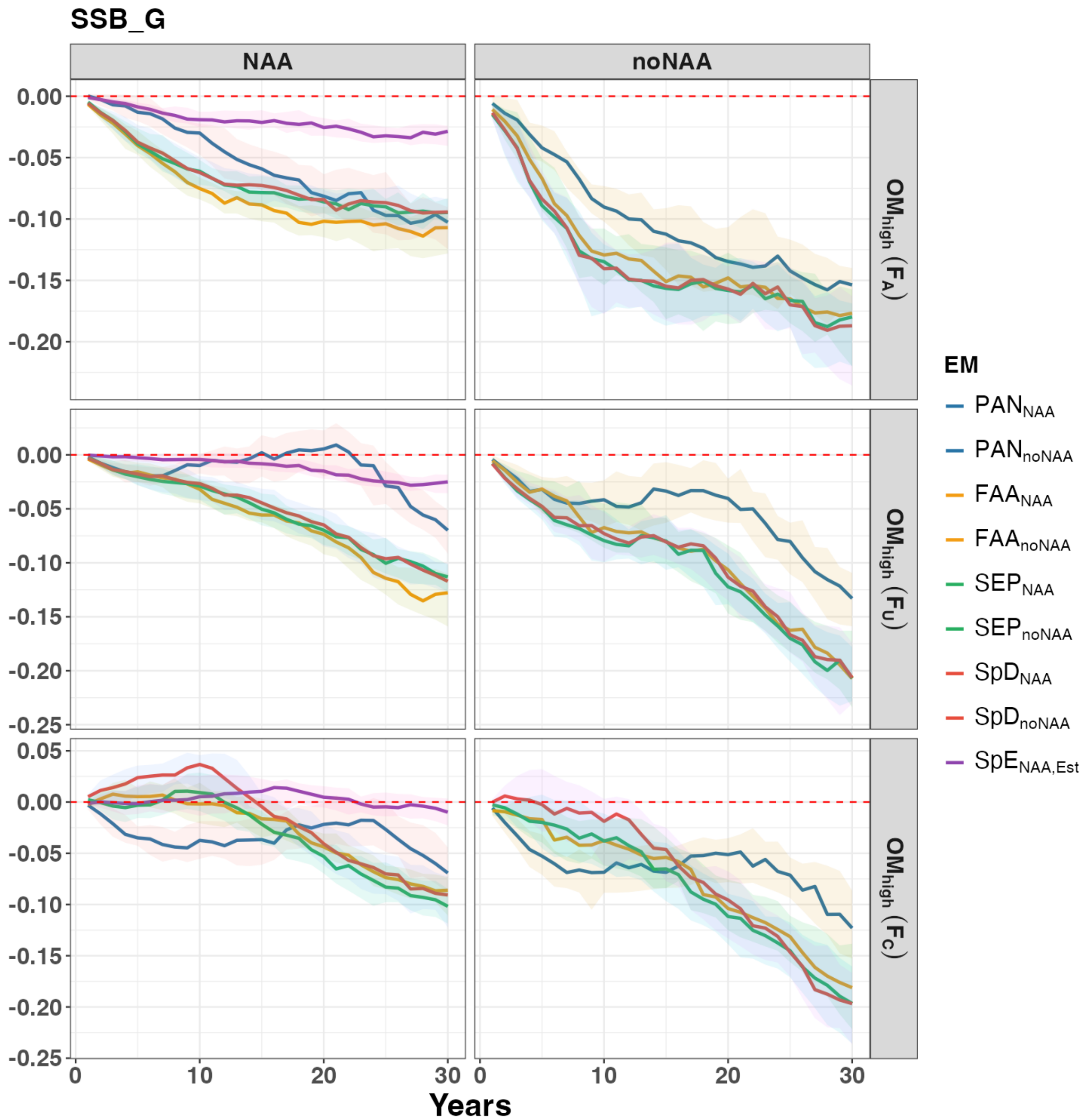

Figure. S8. Relative difference in annual total SSB between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the high movement scenario (OM_{high}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

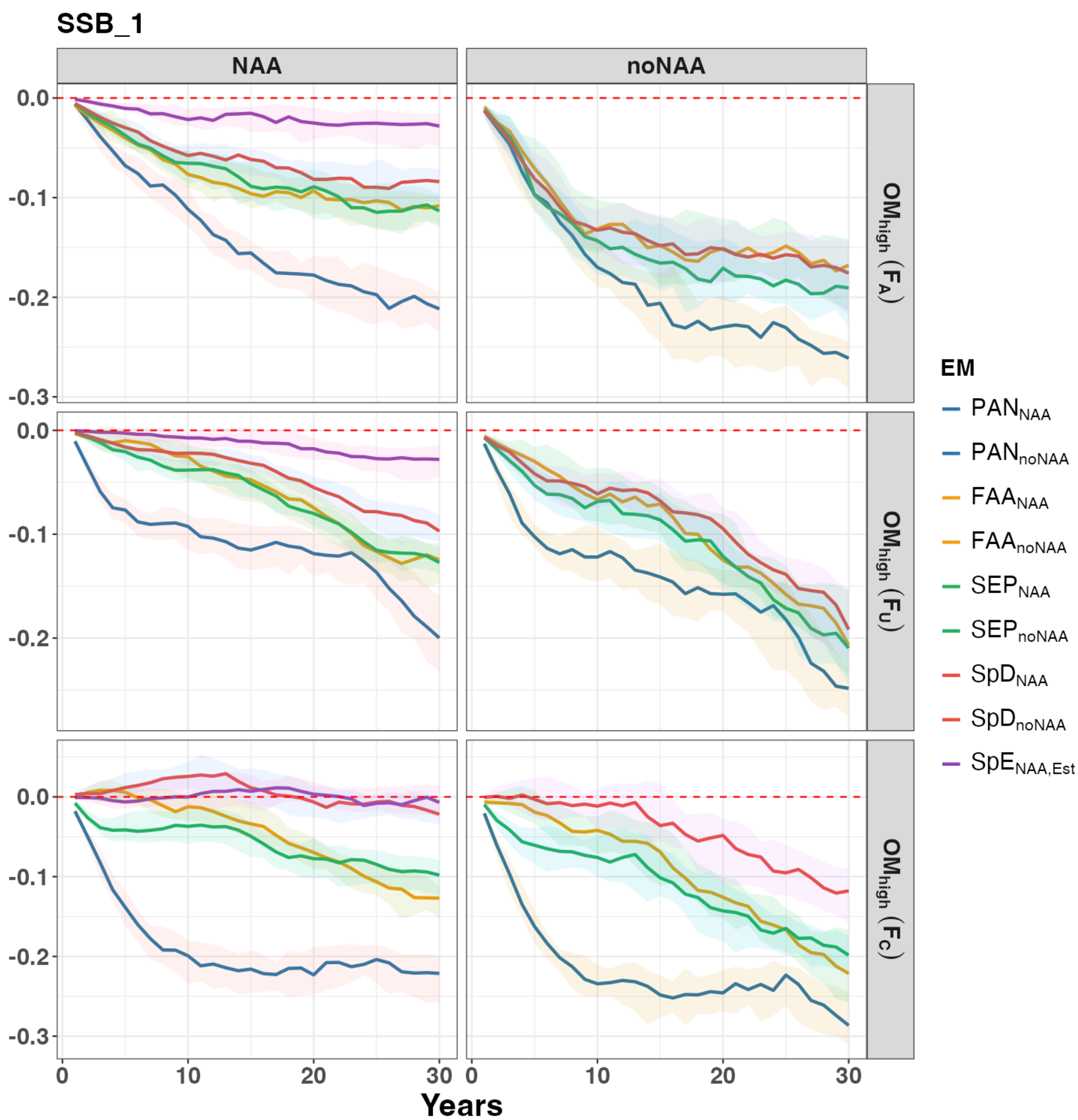

Figure. S9. Relative difference in annual SSB for region 1 between the baseline EM ($\mathrm{SpE}_{\mathrm{NAA,Fix}}$) and other EMs for each fishing history under the high movement scenario ($\mathrm{OM}_{\mathrm{high}}$). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

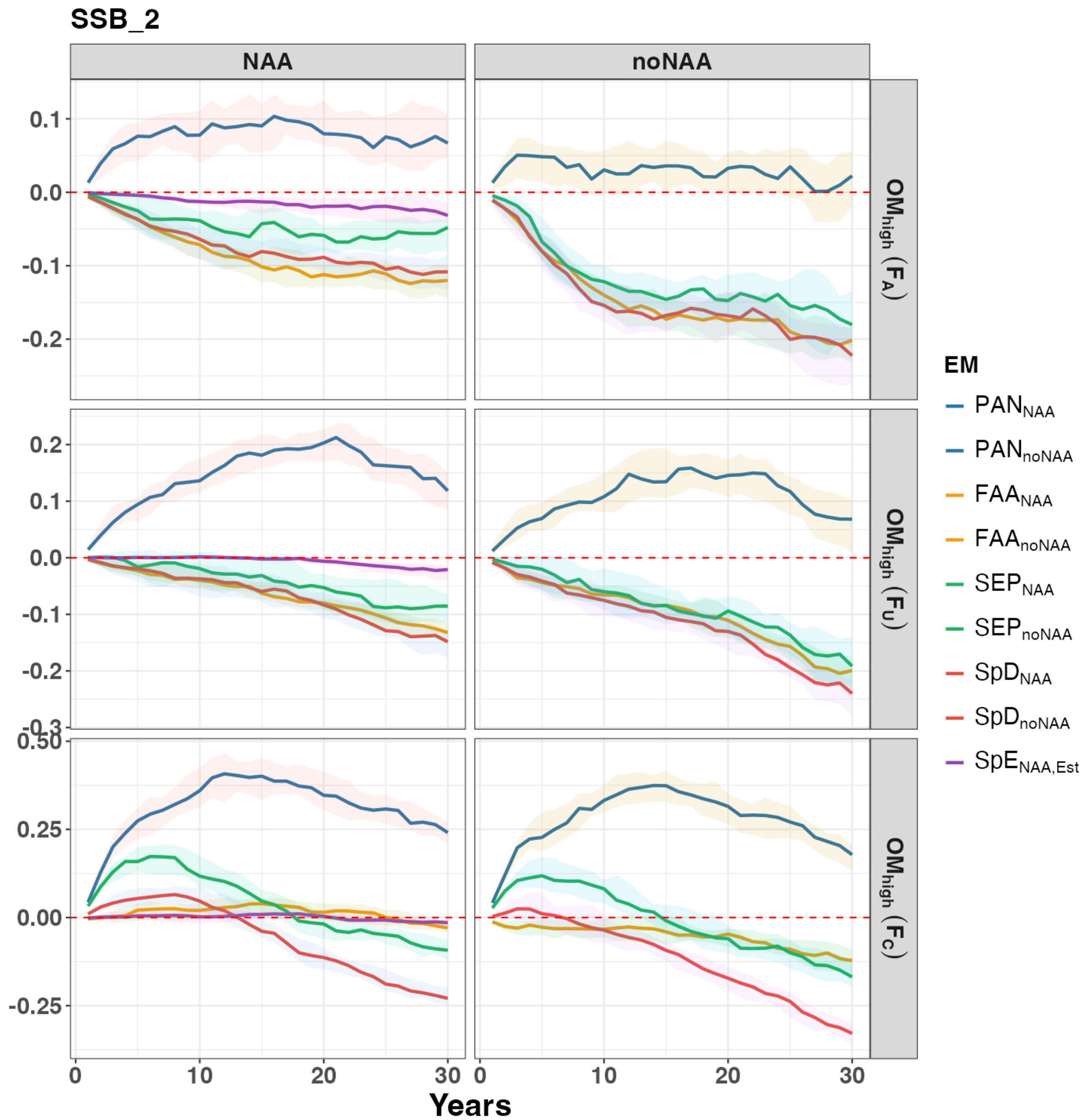

Figure. S10. Relative difference in annual SSB for region 2 between the baseline EM ($\mathrm{SpE}_{\mathrm{NAA,Fix}}$) and other EMs for each fishing history under the high movement scenario ($\mathrm{OM}_{\mathrm{high}}$). The line represents the median. TThe shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

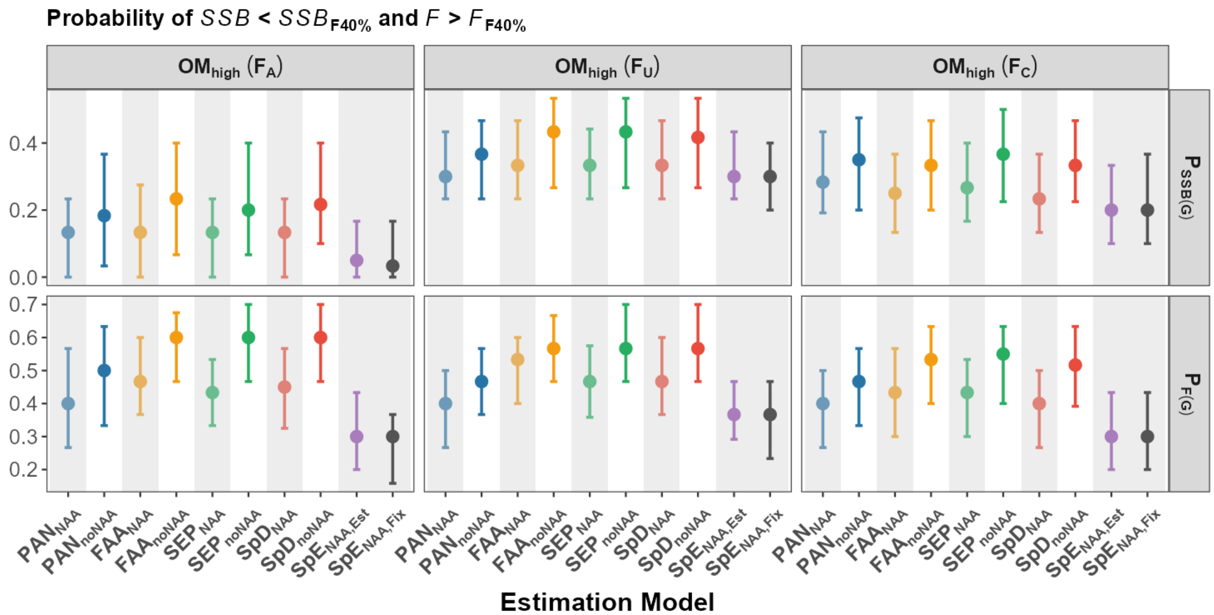

Figure. S11. Probability of $SSB < SSB_{40\%}$ and $F > F_{40\%}$ for each EM over the 30-year feedback period under the high movement scenario ($\mathrm{OM_{high}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

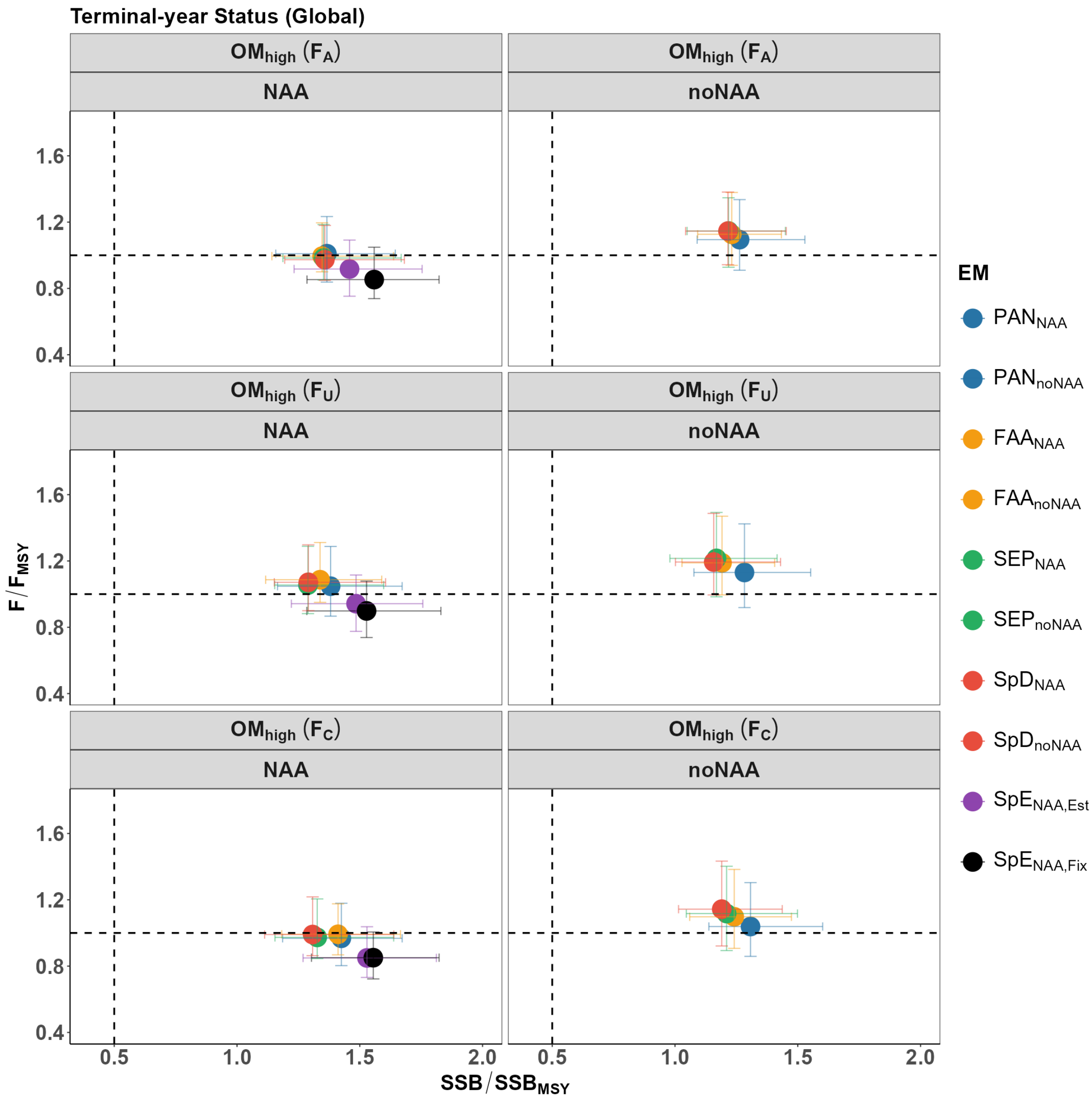

Figure. S12. Terminal-year status of F and SSB at the global scale for each EM under the high movement scenario ($\mathrm{OM}_{\mathrm{high}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). The vertical dashed line indicates the threshold of overfished ($SSB_T/SSB_{MSY_T} < 0.5$), and the horizontal dashed line indicates the threshold of overfishing ($F_T/F_{MSY_T} > 1$). The performance of EMs with and without NAA random effects is separated for better comparison.

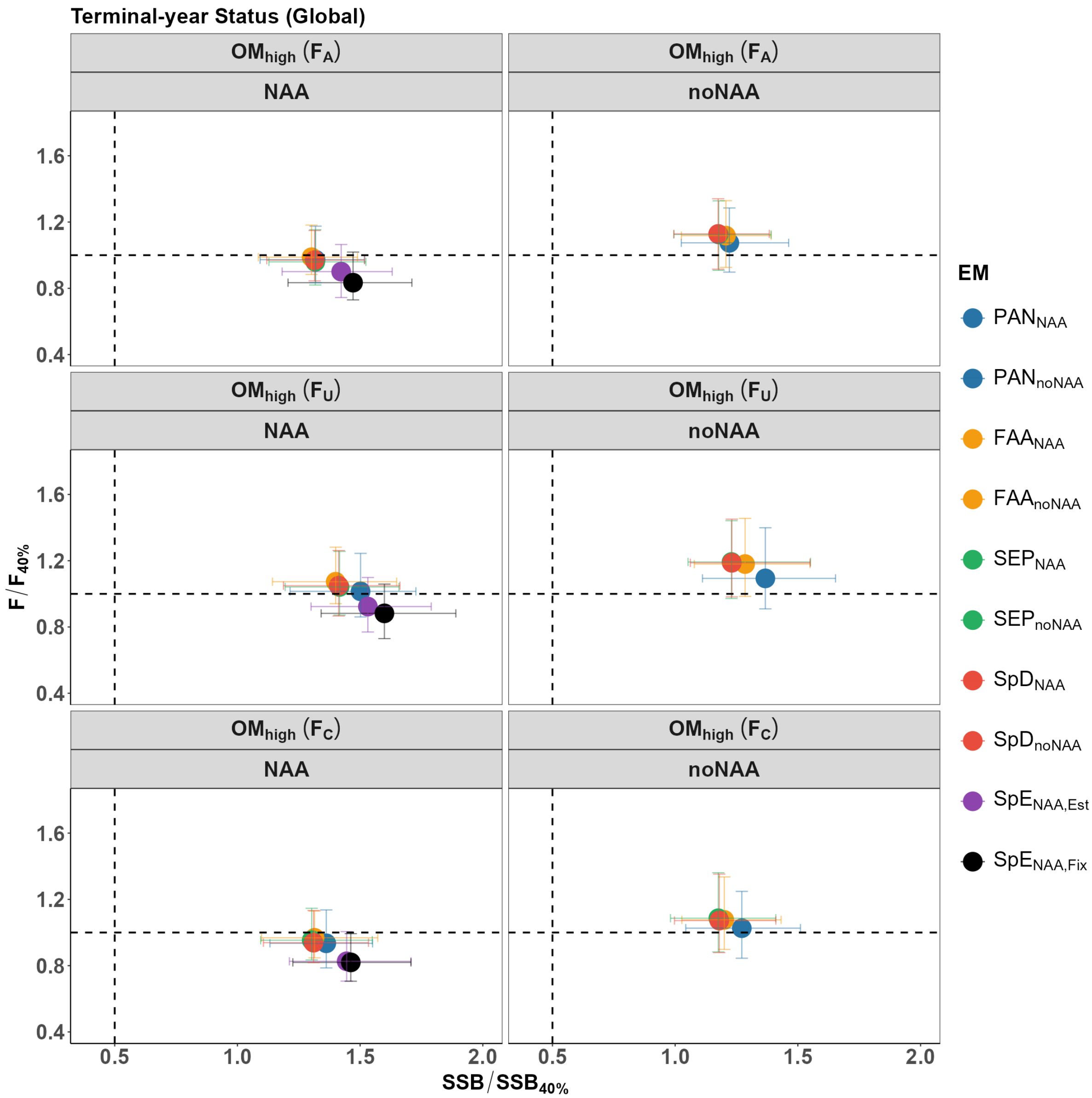

Figure. S13. Terminal-year status of F and SSB at the global scale for each EM under the high movement scenario (OM_{high}). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). The vertical dashed line indicates the threshold of overfished ($SSB_T/SSB_{40\%_T} < 0.5$), and the horizontal dashed line indicates the threshold of overfishing ($F_T/F_{40\%_T} > 1$). The performance of EMs with and without NAA random effects is separated for better comparison.

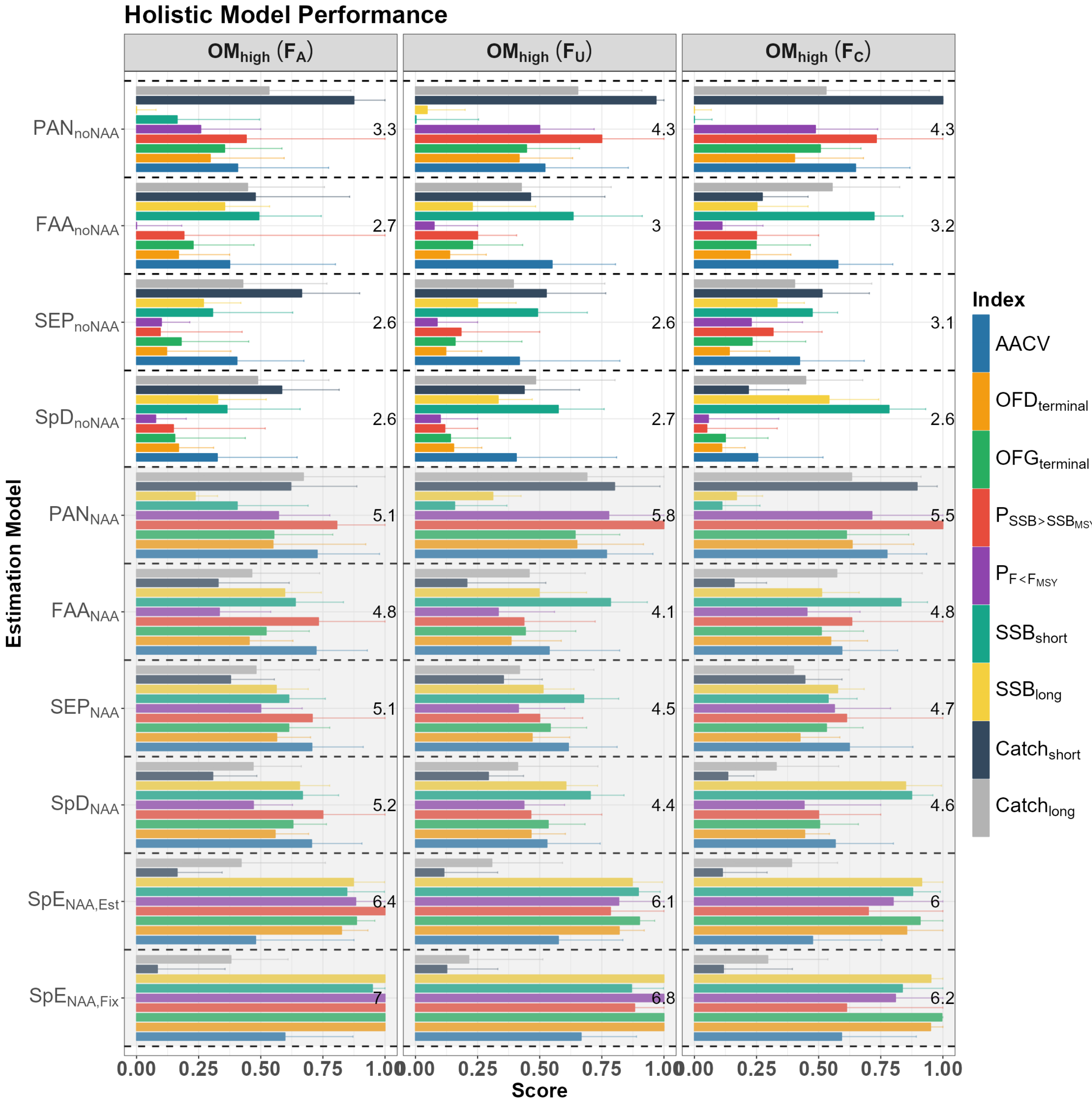

Figure. S14. Overall performance of each EM in region 1 under the high movement scenario (OM_{low}), including the following relative performance metrics: 1) average annual catch variation (AACV); 2) overfished status in the terminal year ($OFD_{terminal}$); 3) overfishing status in the terminal year ($OFG_{terminal}$); 4) probability of $SSB > SSB_{MSY}$ ($P_{SSB>SSB_{MSY}}$); 5) probability of $F < F_{MSY}$ ($P_{F<F_{MSY}}$); 6) short-term SSB (SSB_{short}); 7) long-term SSB (SSB_{long}), 8) short-term catch ($Catch_{short}$); and 9) long-term catch ($Catch_{long}$). All the indices were standardized to scores between 0 and 1, with higher values indicating better performance. The total score for each EM was provided for each fishing scenario. The black dashed line separates the performance of each EM. EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

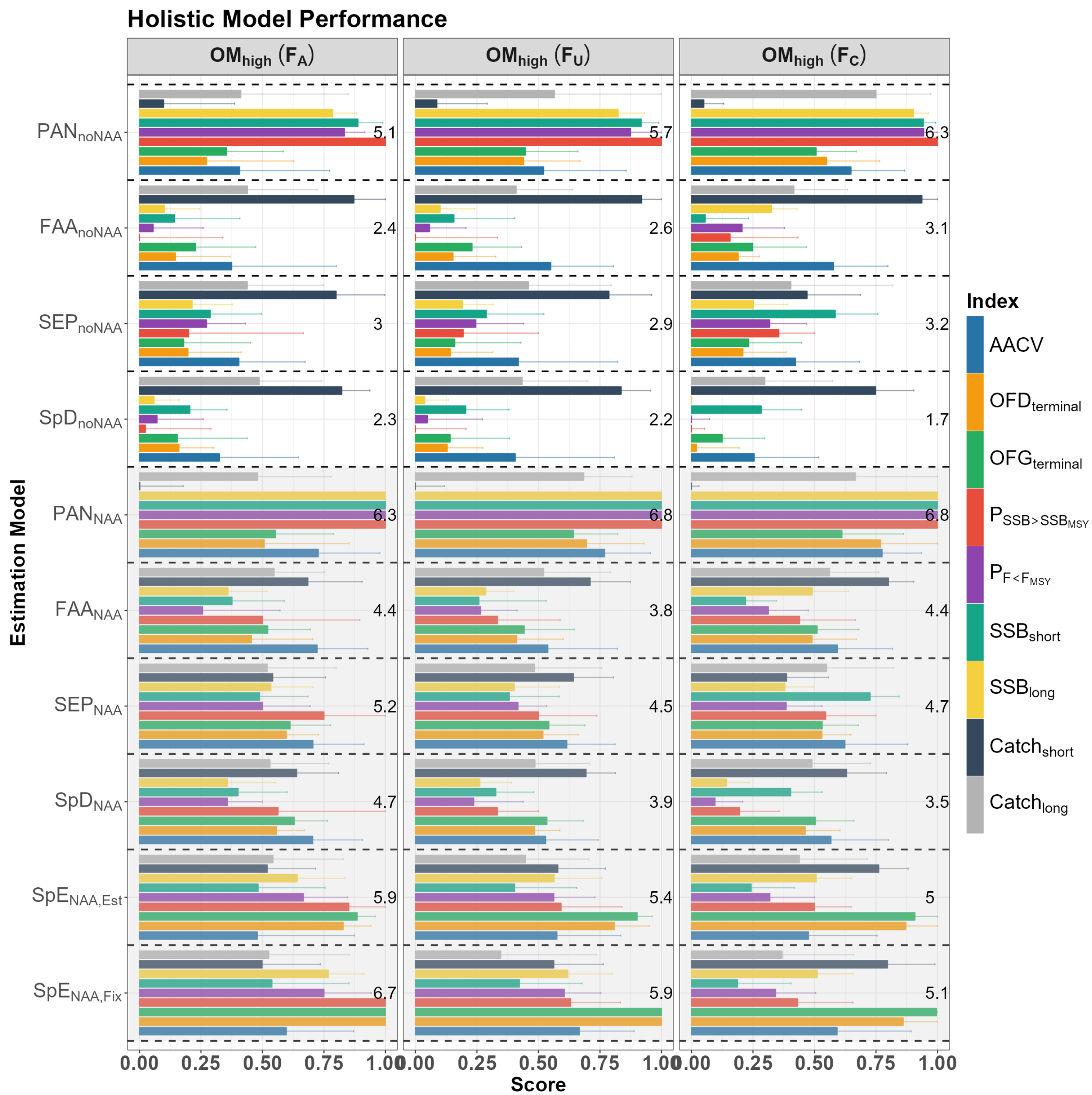

Figure. S15. Overall performance of each EM in region 2 under the high movement scenario (OM_{low}), including the following relative performance metrics: 1) average annual catch variation (AACV); 2) overfished status in the terminal year ($\text{OFD}_{\text{terminal}}$); 3) overfishing status in the terminal year ($\text{OFG}_{\text{terminal}}$); 4) probability of $SSB > SSB_{MSY}$ ($\text{P}_{\text{SSB}>\text{SSB}_{\text{MSY}}}$); 5) probability of $F < F_{MSY}$ ($\text{P}_{\text{F}<\text{F}_{\text{MSY}}}$); 6) short-term SSB ($\text{SSB}_{\text{short}}$); 7) long-term SSB (SSB_{long}), 8) short-term catch ($\text{Catch}_{\text{short}}$); and 9) long-term catch ($\text{Catch}_{\text{long}}$). All the indices were standardized to scores between 0 and 1, with higher values indicating better performance. The total score for each EM was provided for each fishing scenario. The black dashed line separates the performance of each EM. EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

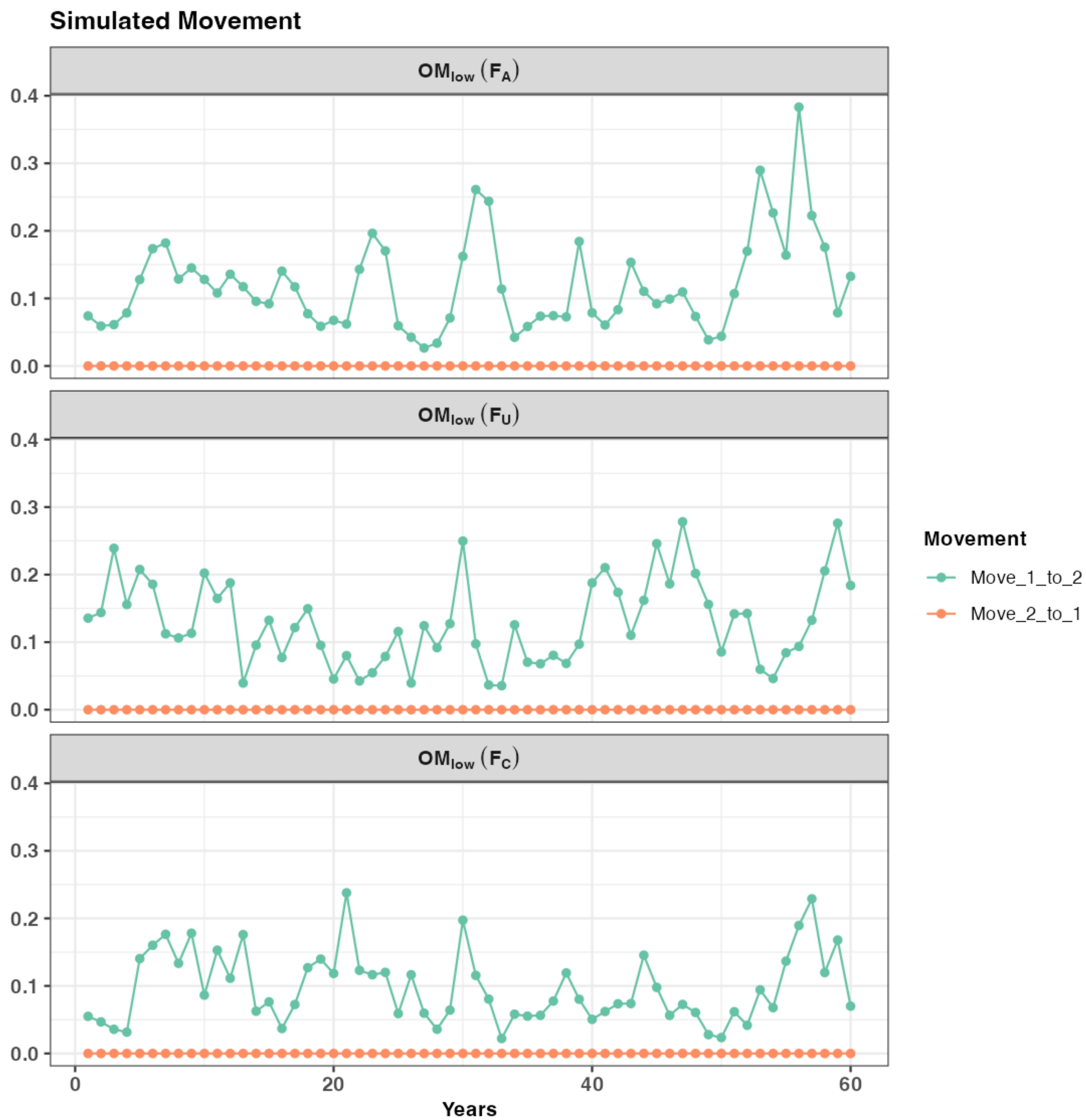

Figure. S16. Simulated AR1 movement rates (from one replicate) under the low movement scenario (OM_{low}).

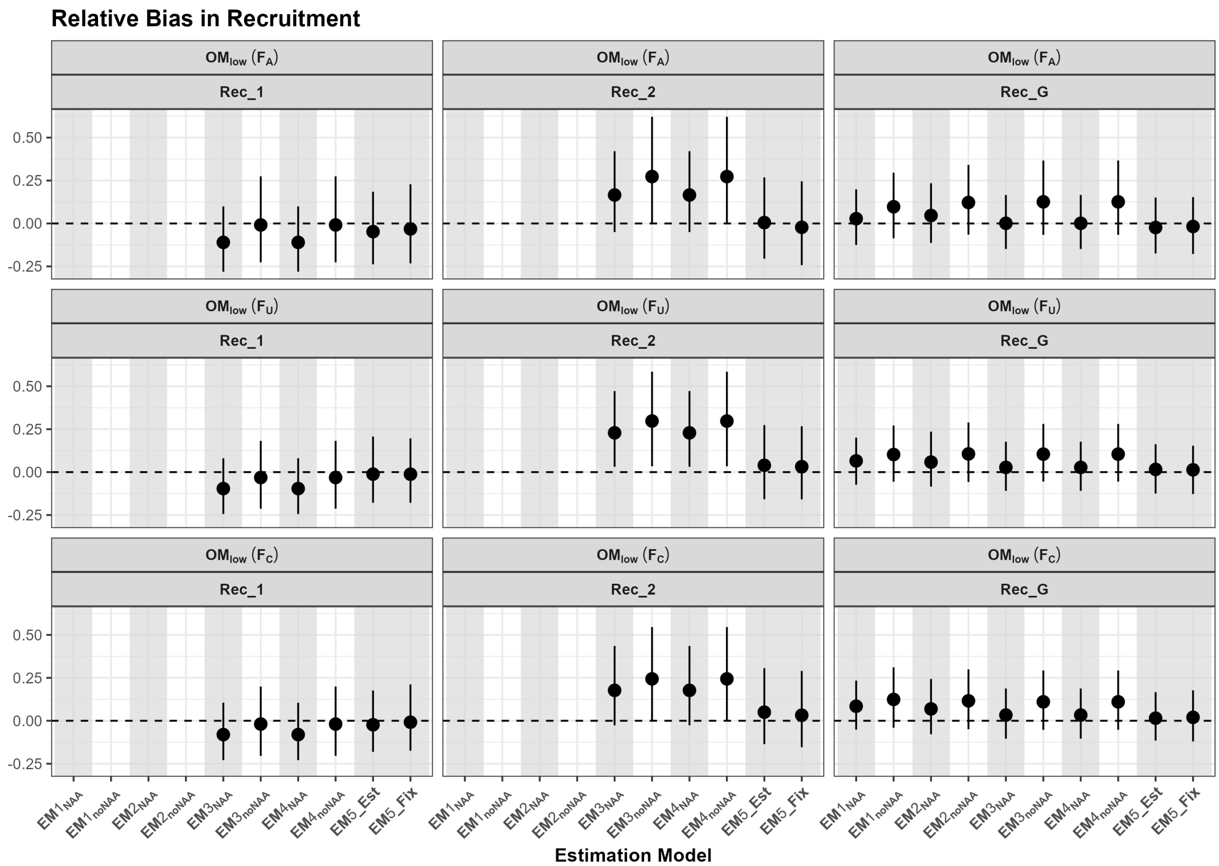

Figure. S17. Relative bias in recruitment for region 1, region 2, and global recruitment from the first assessment model during the feedback period under the low movement scenario (OM_{low}). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

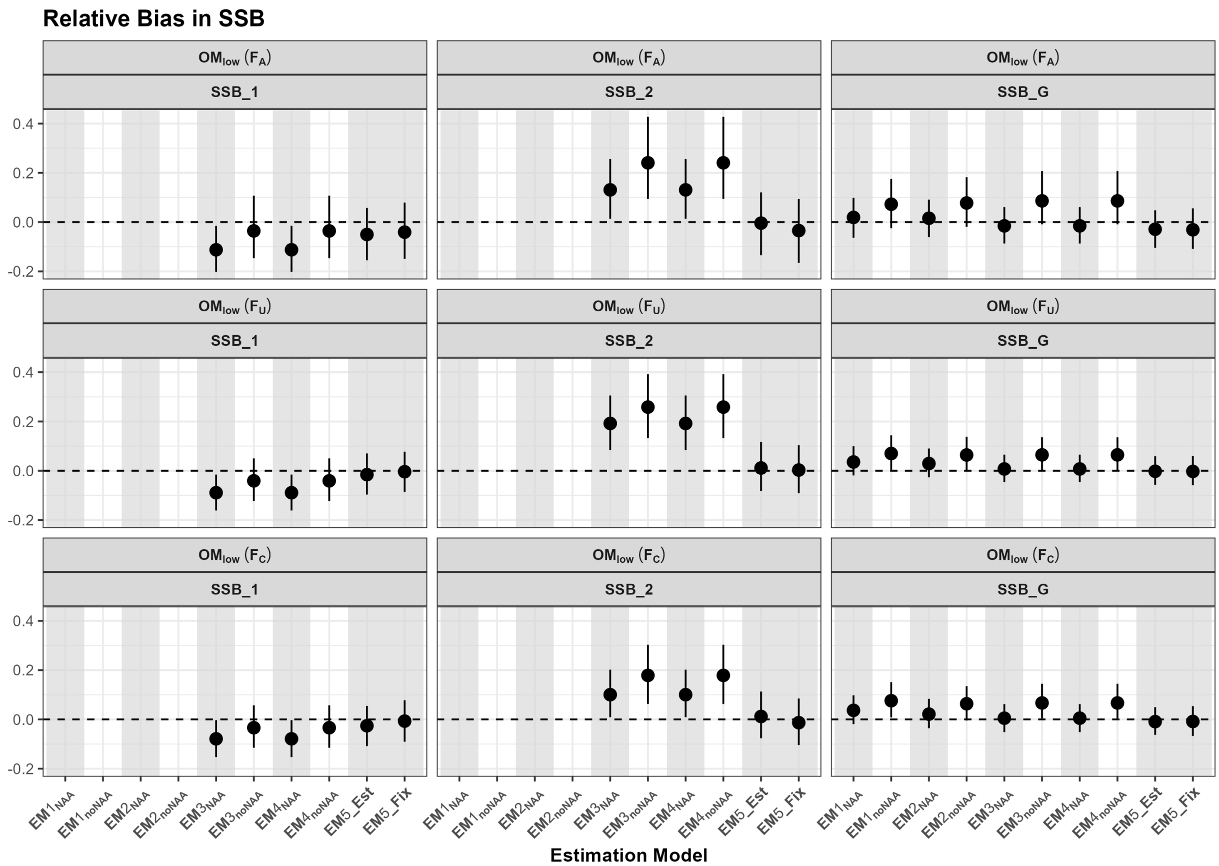

Figure. S18. Relative bias in SSB for region 1, region 2, and global SSB from the first assessment model during the feedback period under the low movement scenario ($\mathrm{OM}_{\mathrm{low}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

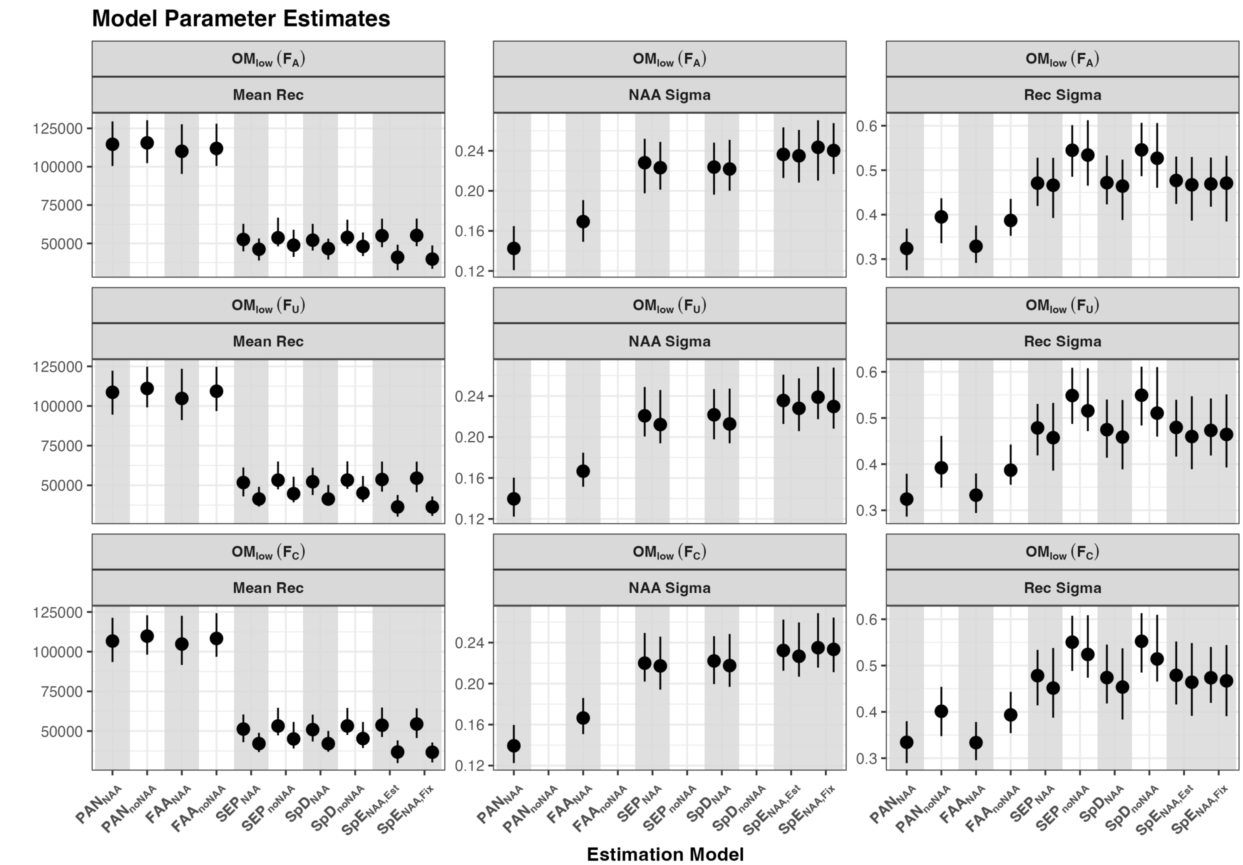

Figure. S19. The mean recruitment parameter, recruitment standard deviation and NAA standard deviation from the last assessment model during the feedback period under the low movement scenario (OM_{low}). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

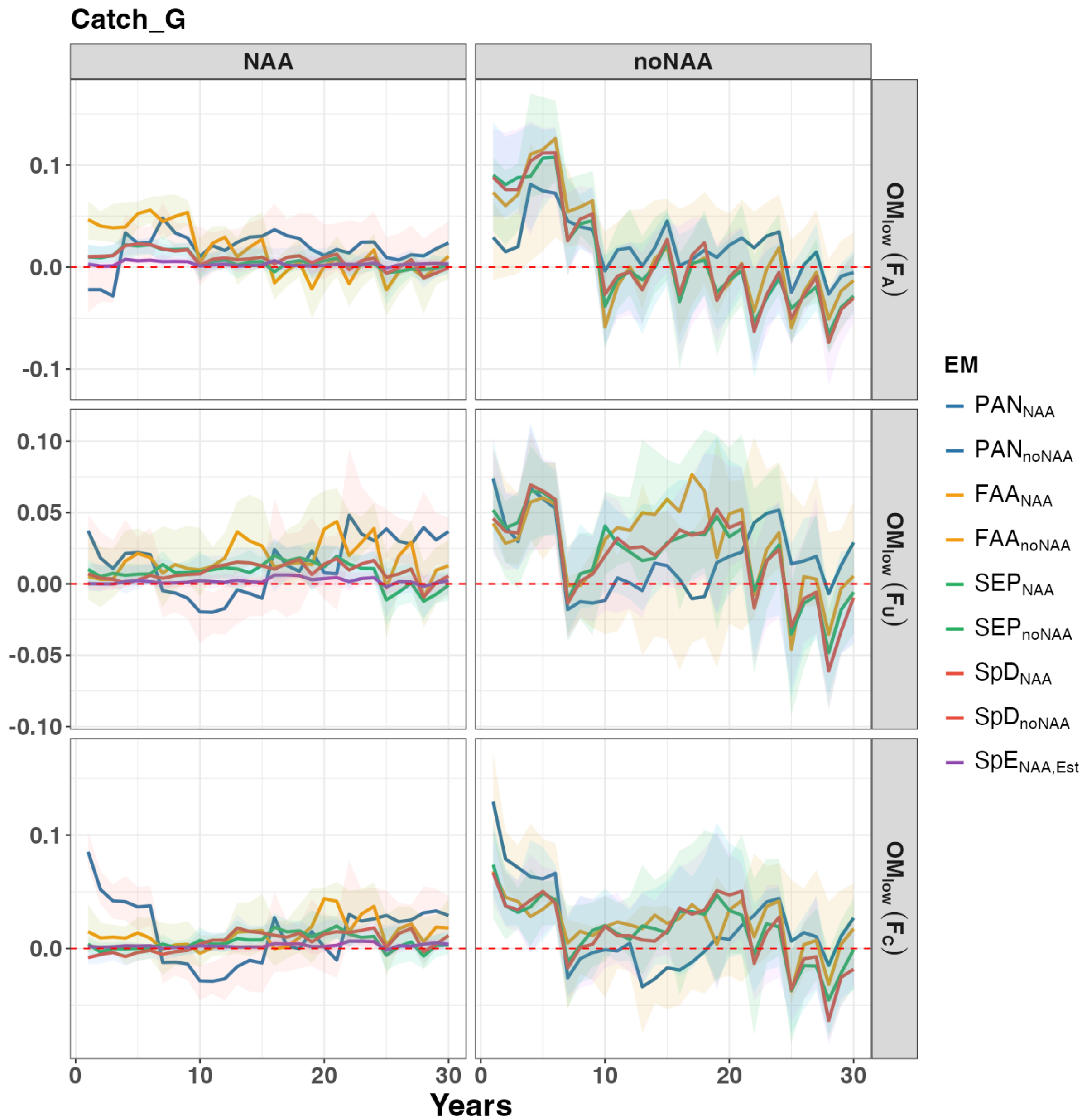

Figure. S20. Relative difference in annual total catch between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the low movement scenario (OM_{low}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

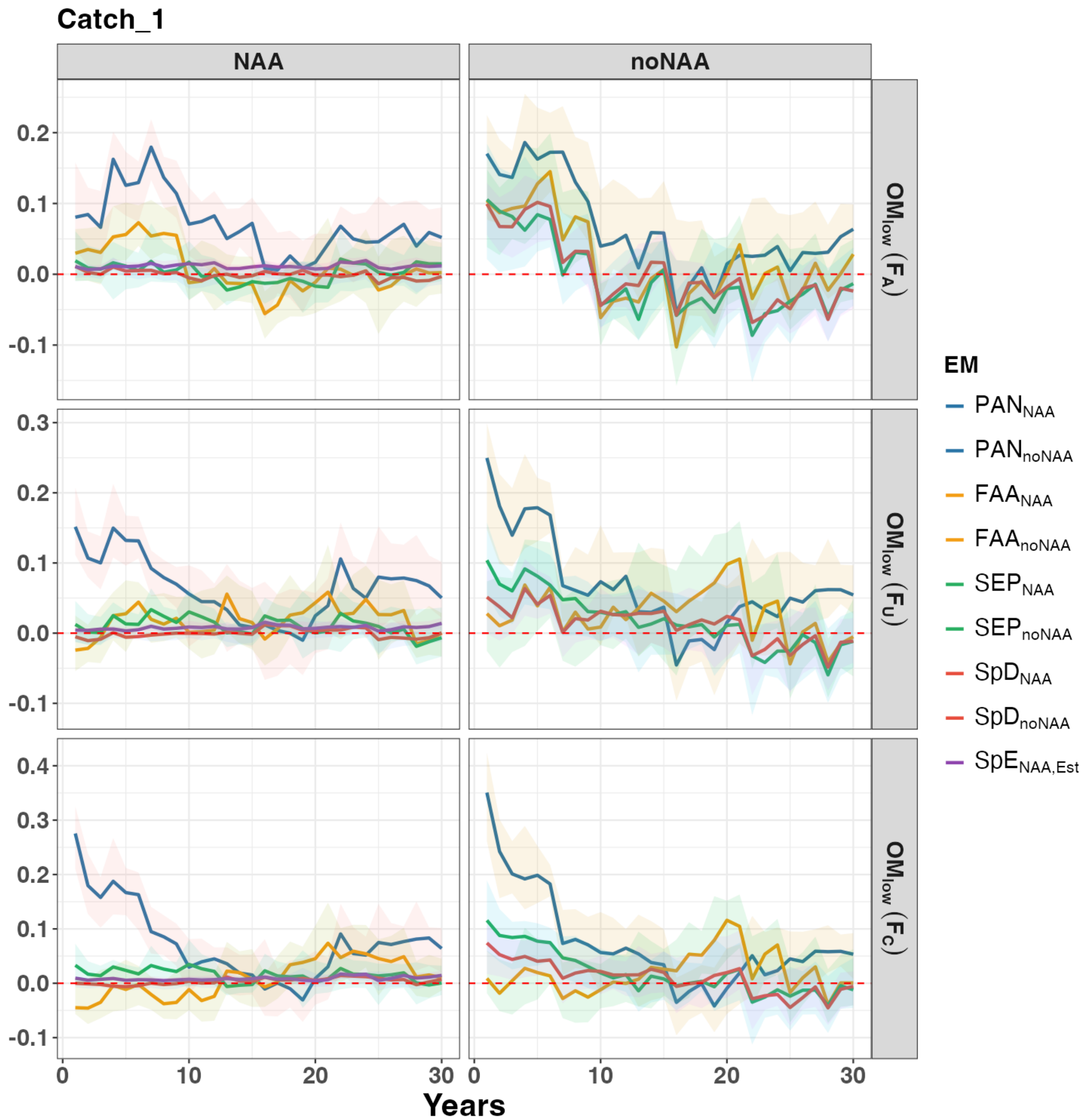

Figure. S21. Relative difference in annual catch in region 1 between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the low movement scenario (OM_{low}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

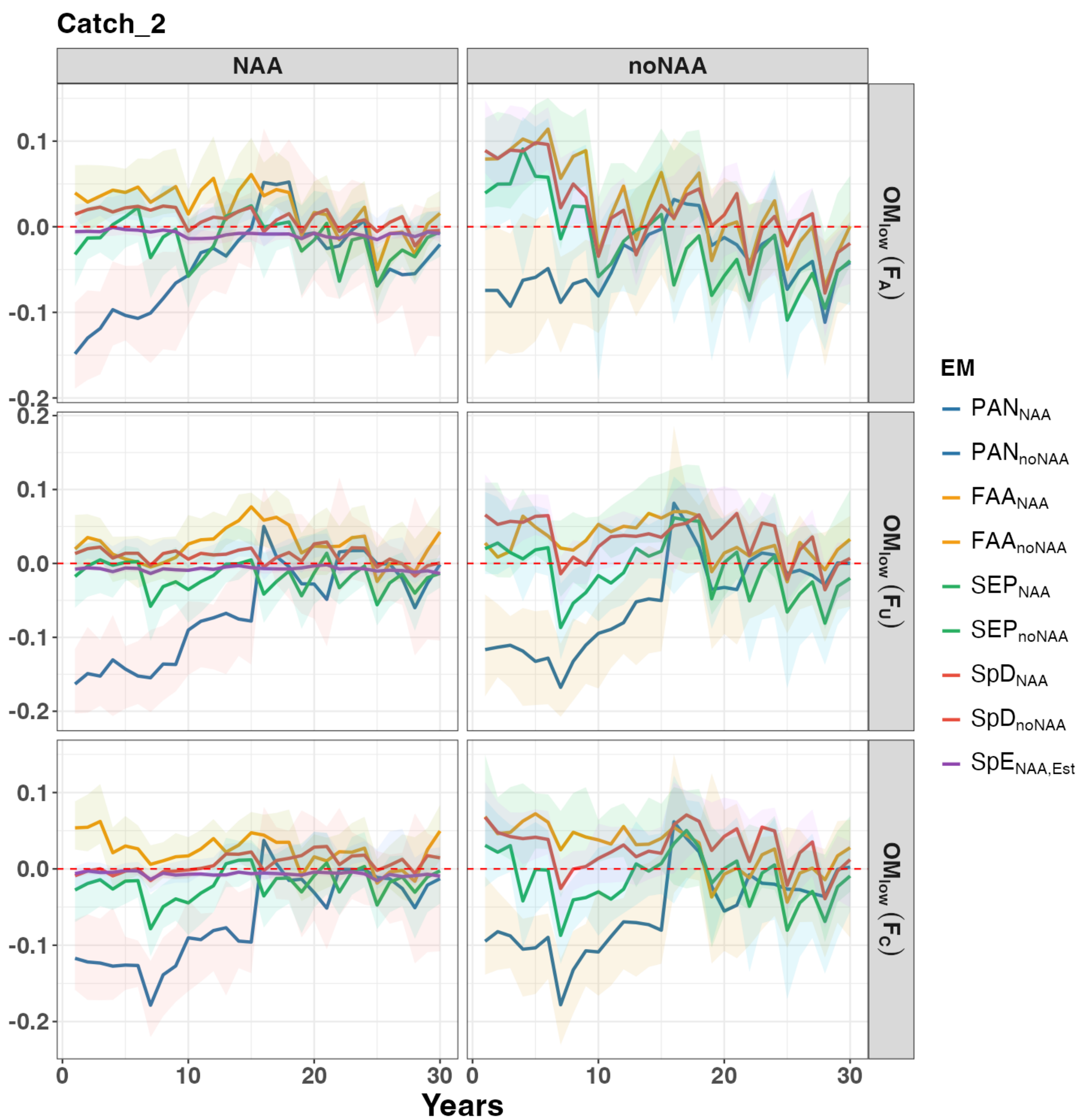

Figure. S22. Relative difference in annual catch in region 2 between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the low movement scenario (OM_{low}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

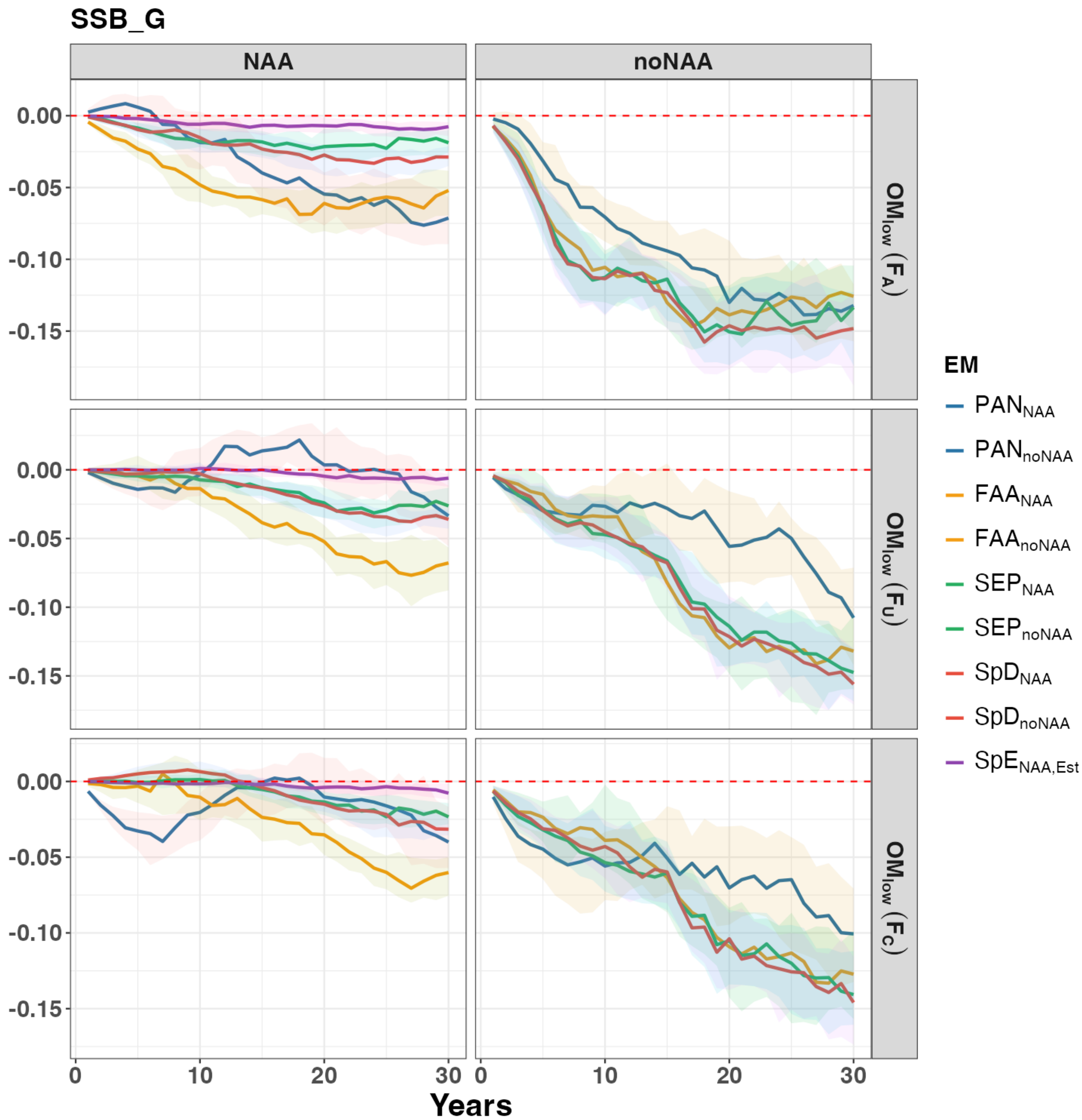

Figure. S23. Relative difference in annual total SSB between the baseline EM (SpE$_{\text{NAA,Fix}}$) and other EMs for each fishing history under the low movement scenario (OM$_{\text{low}}$). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

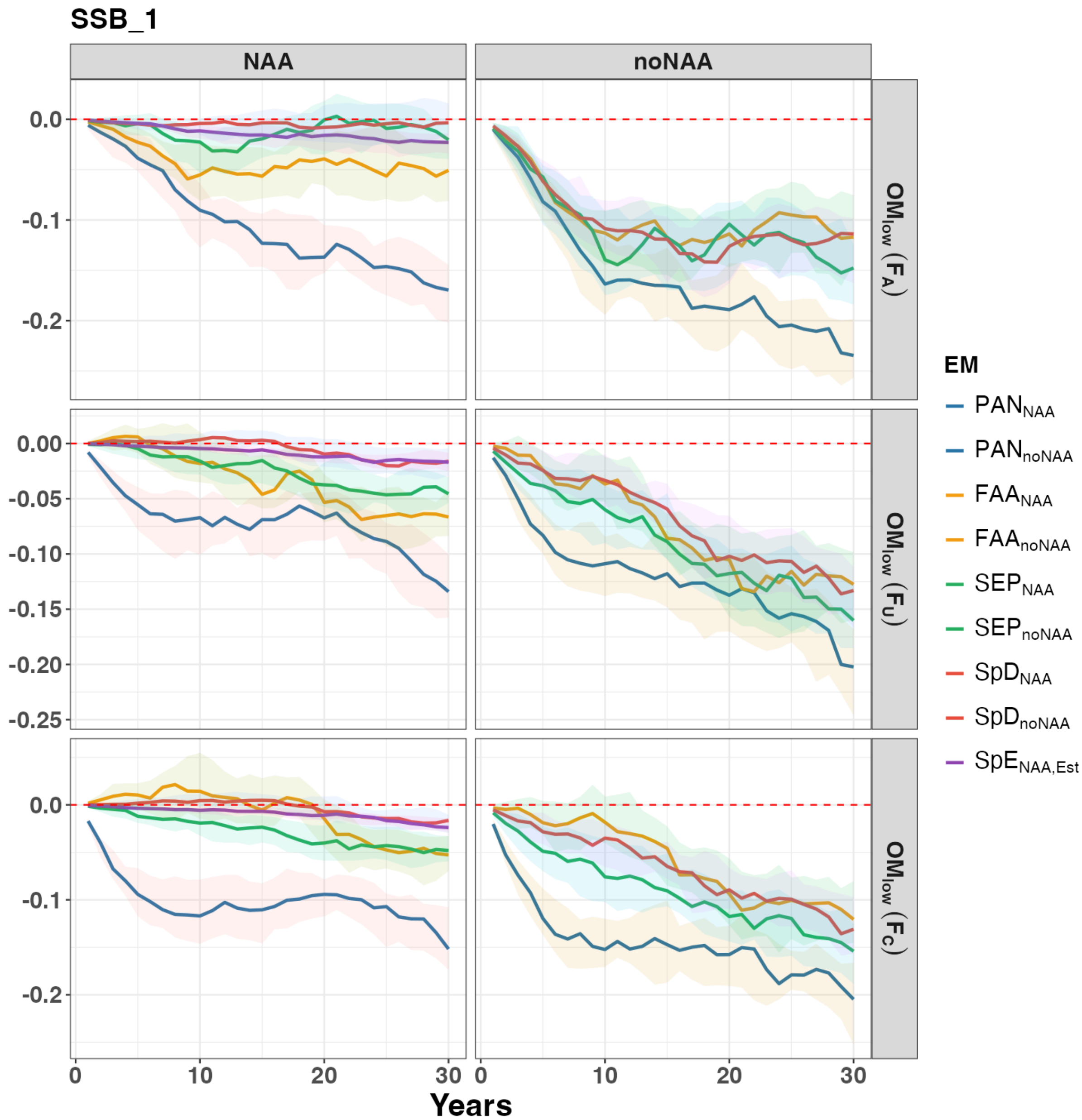

Figure. S24. Relative difference in annual SSB for region 1 between the baseline EM ($\text{SpE}_{\text{NAA,Fix}}$) and other EMs for each fishing history under the low movement scenario (OM_{low}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

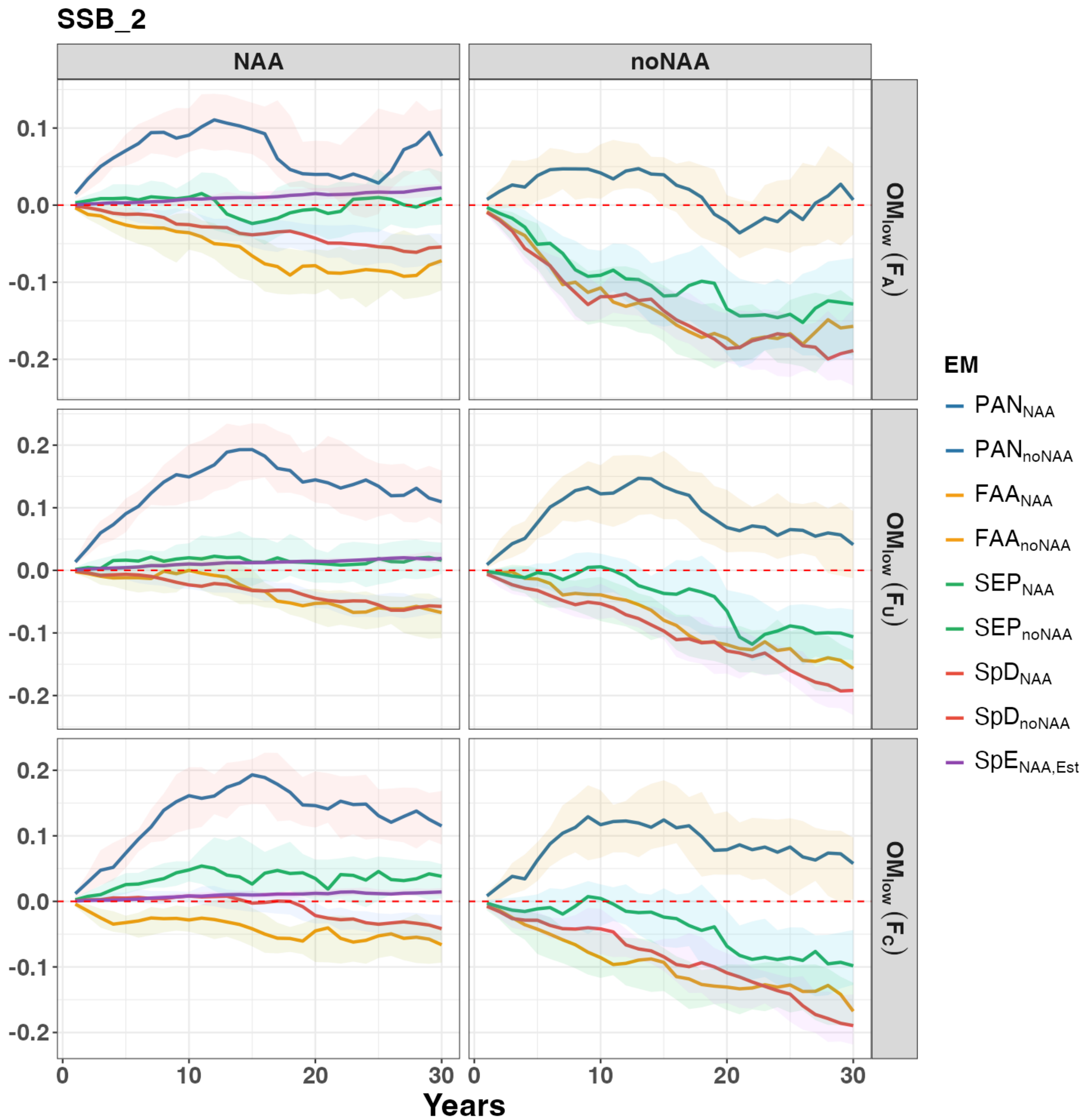

Figure. S25. Relative difference in annual SSB for region 2 between the baseline EM ($SpE_{NAA,Fix}$) and other EMs for each fishing history under the low movement scenario (OM_{low}). The line represents the median. The shade area represents the inter-quantile range (the 40th and 60th quantiles). The performance of EMs with and without NAA random effects is separated for better comparison.

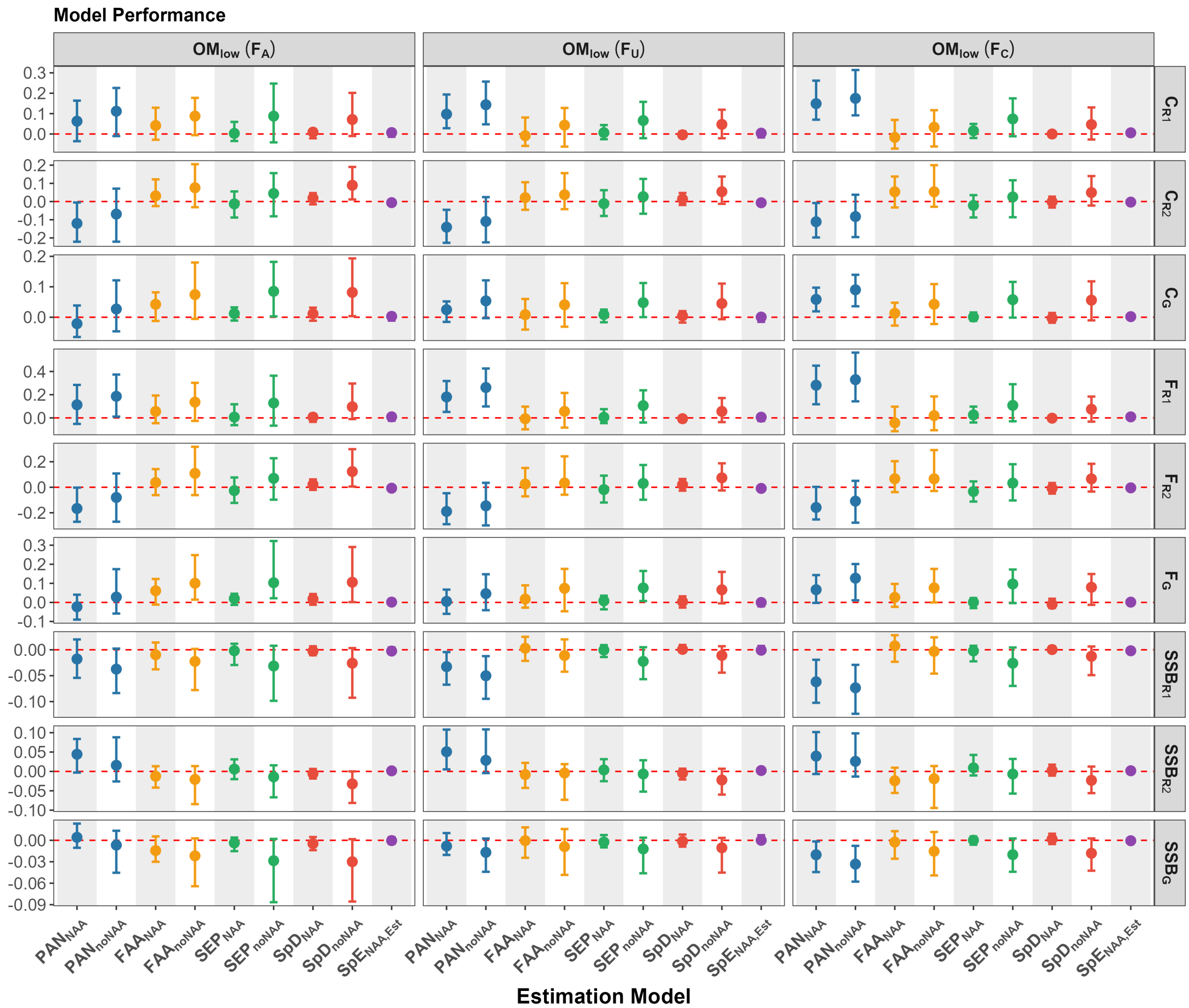

Figure. S26. Relative difference in catch, F, and SSB between the baseline EM ($\mathrm{SpE}_{\mathrm{NAA,Fix}}$) and other EMs over the first 5 years of the feedback period under the low movement scenario ($\mathrm{OM}_{\mathrm{low}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

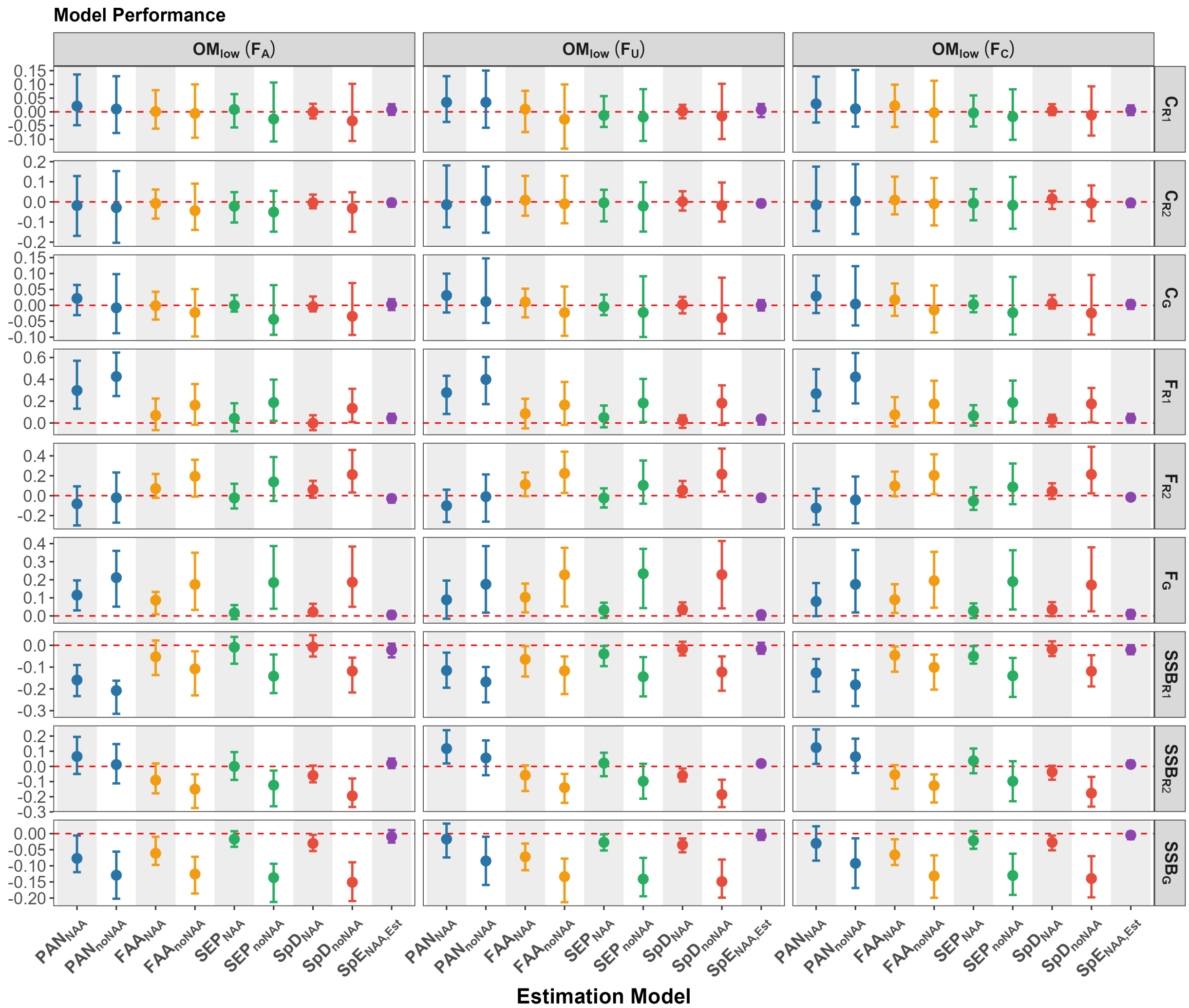

Figure. S27. Relative difference in catch, F, and SSB between the baseline EM ($\mathrm{SpE}_{\mathrm{NAA,Fix}}$) and other EMs over the last 5 years of the feedback period under the low movement scenario ($\mathrm{OM}_{\mathrm{low}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

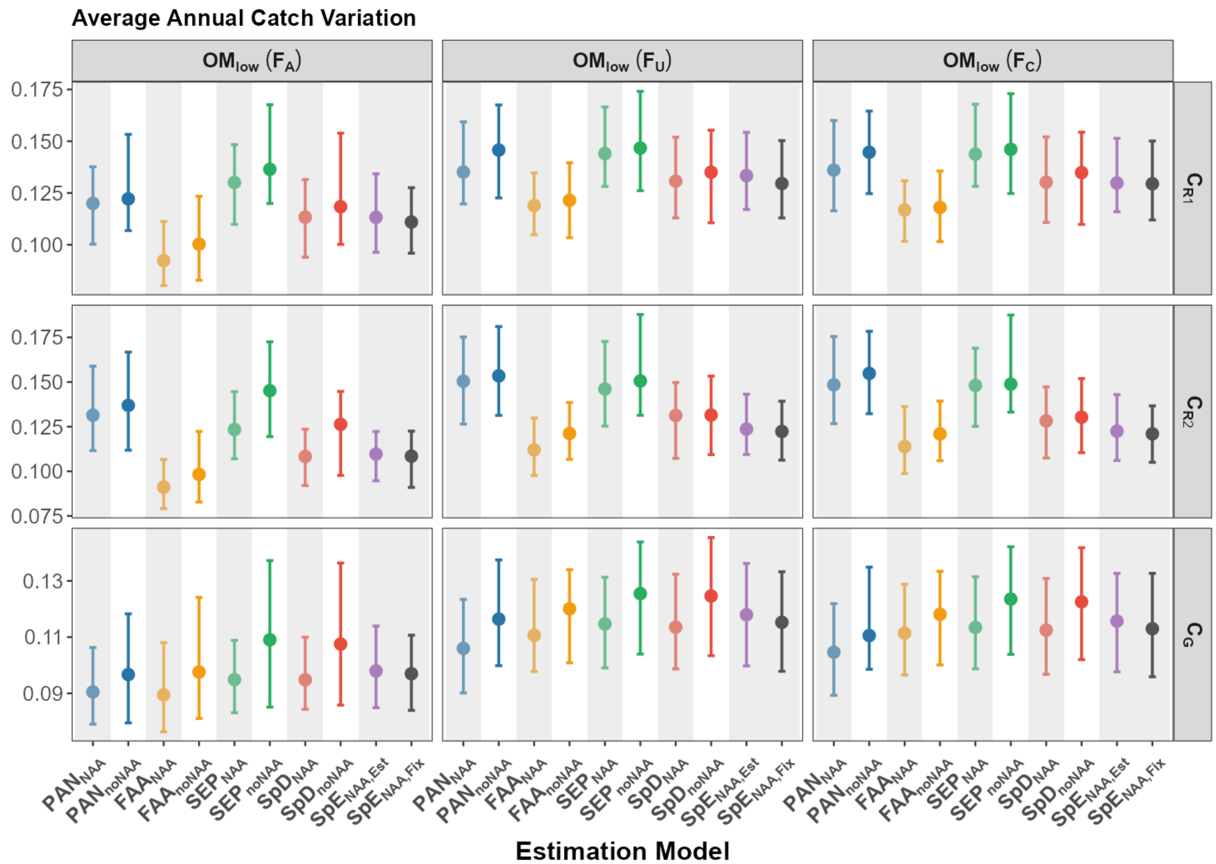

Figure. S28. Average annual catch variation (AACV) of each EM at both regional and global scales under the low movement scenario (OM$_{low}$). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

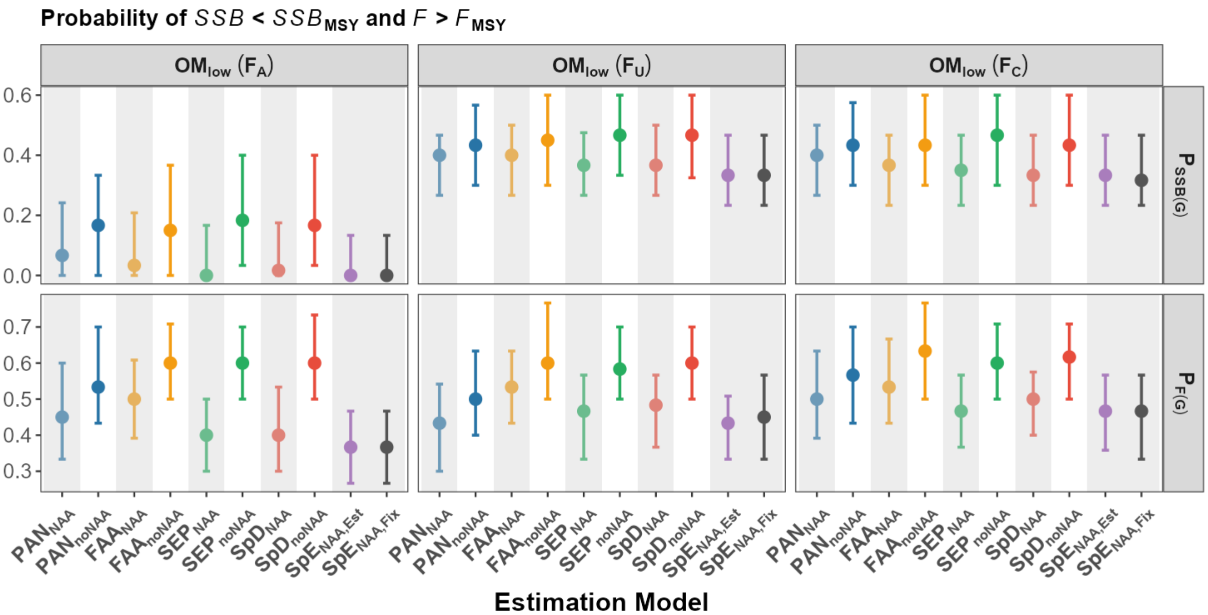

Figure. S29. Probability of $SSB < SSB_{MSY}$ and $F > F_{MSY}$ for each EM over the 30-year feedback period under the low movement scenario ($\mathrm{OM}_{\mathrm{low}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

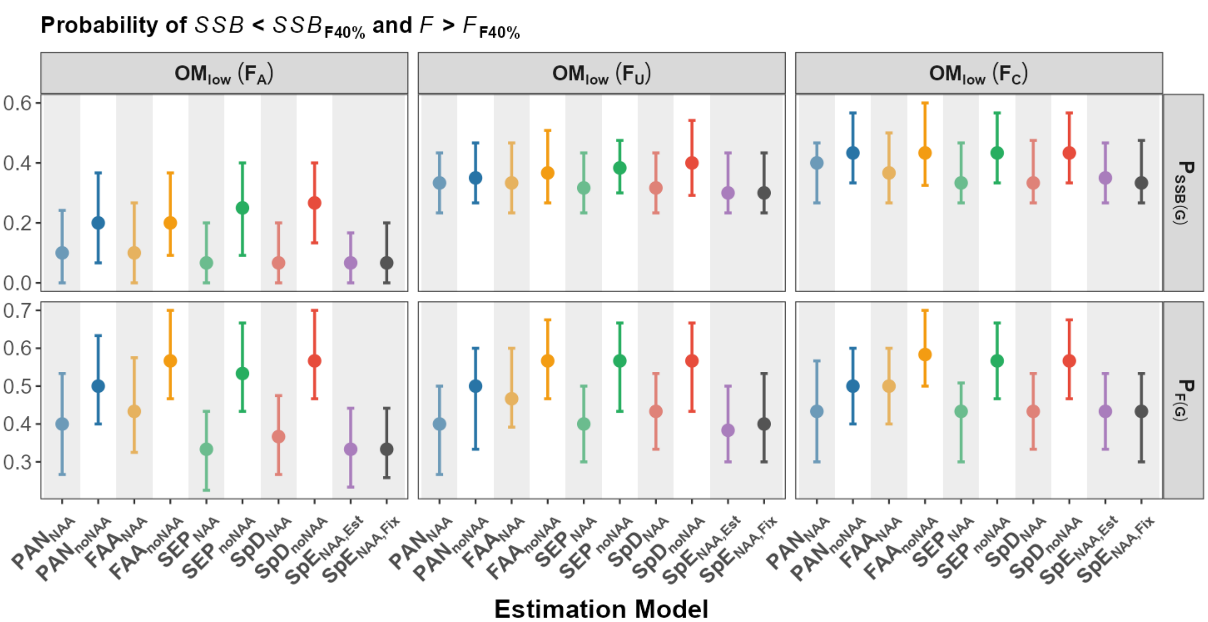

Figure. S30. Probability of $SSB < SSB_{40\%}$ and $F > F_{40\%}$ for each EM over the 30-year feedback period under the low movement scenario ($\mathrm{OM}_{\mathrm{low}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

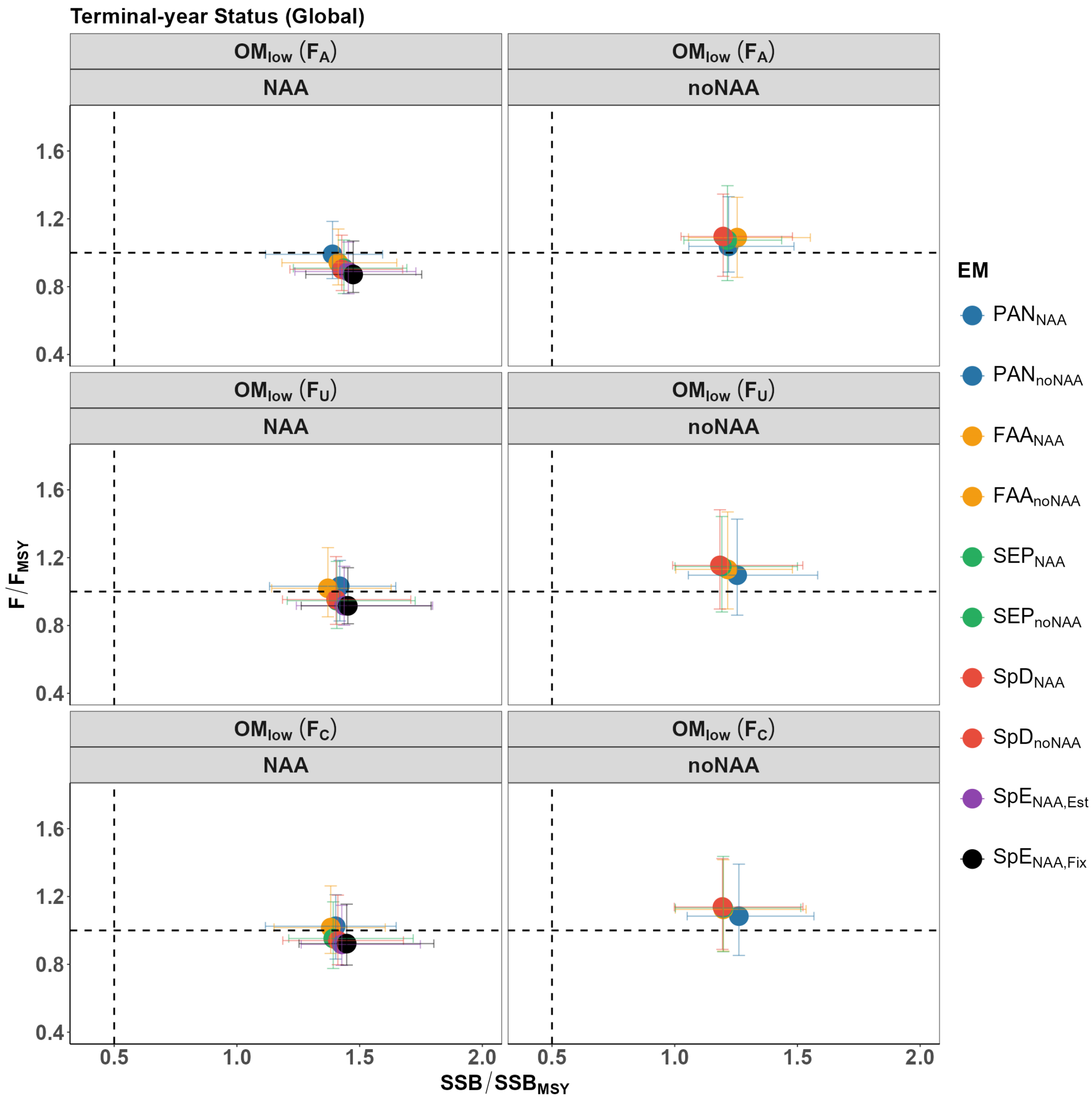

Figure. S31. Terminal-year status of F and SSB at the global scale for each EM under the low movement scenario ($\mathrm{OM_{low}}$). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). The vertical dashed line indicates the threshold of overfished ($SSB_T/SSB_{MSY_T} < 0.5$), and the horizontal dashed line indicates the threshold of overfishing ($F_T/F_{MSY_T} > 1$). The performance of EMs with and without NAA random effects is separated for better comparison.

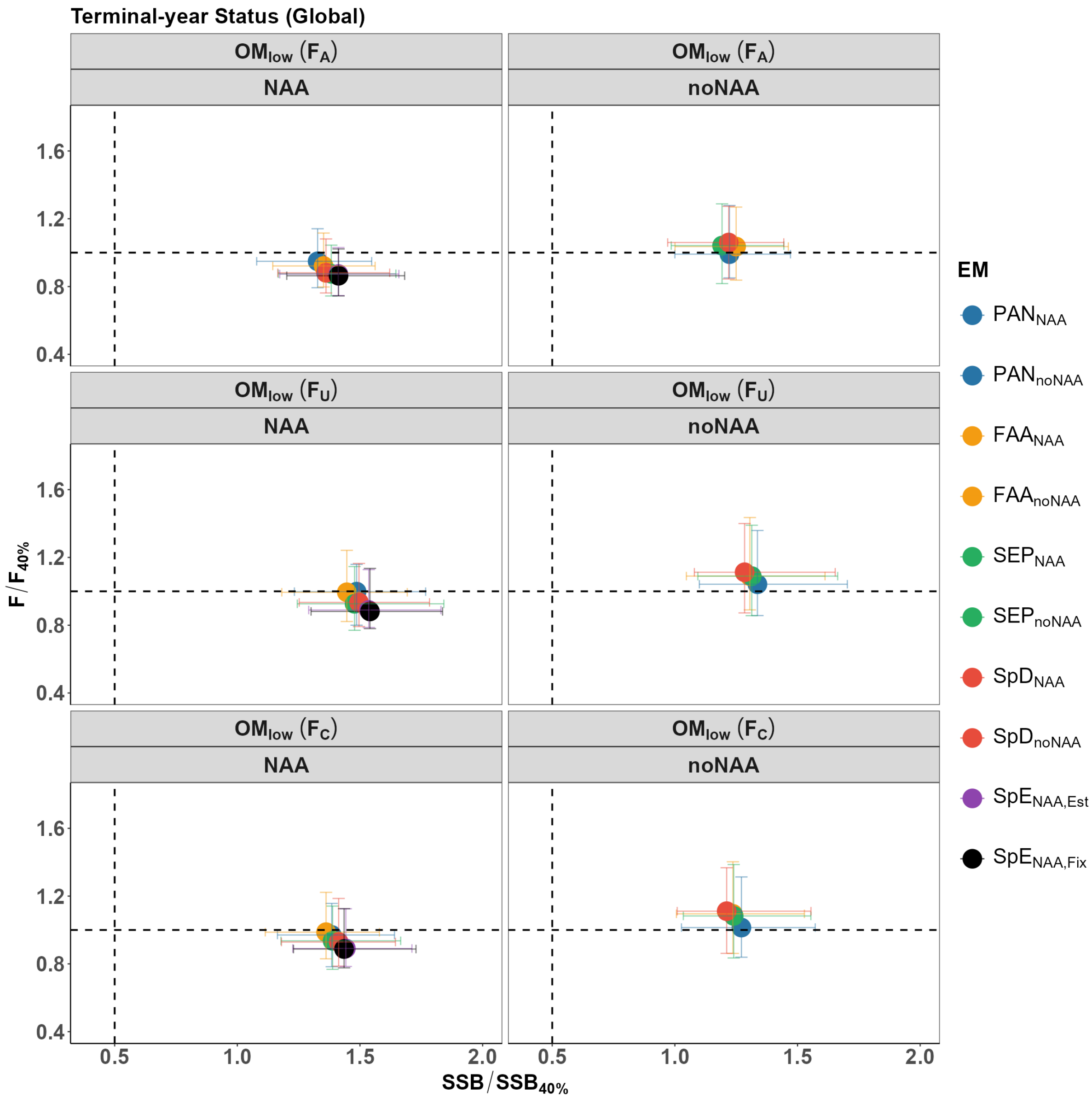

Figure. S32. Terminal-year status of F and SSB at the global scale for each EM under the low movement scenario (OM_{low}). The point indicates the median. The error bar indicates the inter-quantile range (the 25th and 75th quantiles). The vertical dashed line indicates the threshold of overfished ($SSB_T/SSB_{40\%_T} < 0.5$), and the horizontal dashed line indicates the threshold of overfishing ($F_T/F_{40\%_T} > 1$). The performance of EMs with and without NAA random effects is separated for better comparison.

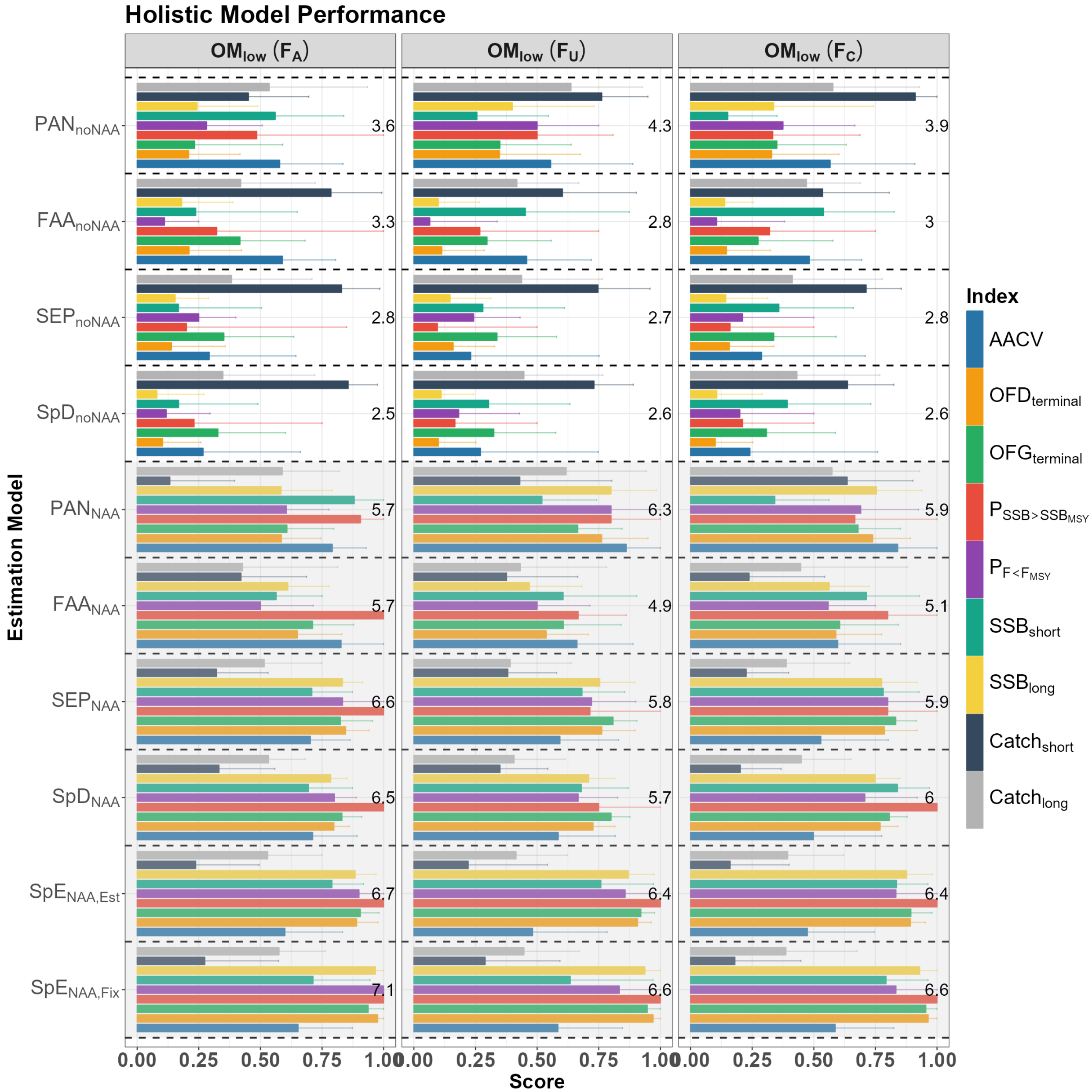

Figure. S33. Overall performance of each EM at the global scale under the low movement scenario ($\mathrm{OM_{low}}$), including the following relative performance metrics: 1) average annual catch variation (AACV); 2) overfished status in the terminal year ($\mathrm{OFD_{terminal}}$); 3) overfishing status in the terminal year ($\mathrm{OFG_{terminal}}$); 4) probability of $SSB > SSB_{MSY}$ ($\mathrm{P_{SSB>SSB_{MSY}}}$); 5) probability of $F < F_{MSY}$ ($\mathrm{P_{F<F_{MSY}}}$); 6) short-term SSB ($\mathrm{SSB_{short}}$); 7) long-term SSB ($\mathrm{SSB_{long}}$), 8) short-term catch ($\mathrm{Catch_{short}}$); and 9) long-term catch ($\mathrm{Catch_{long}}$). All the indices were standardized to scores between 0 and 1, with higher values indicating better performance. The total score for each EM was provided for each fishing scenario. The black dashed line separates the performance of each EM. EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

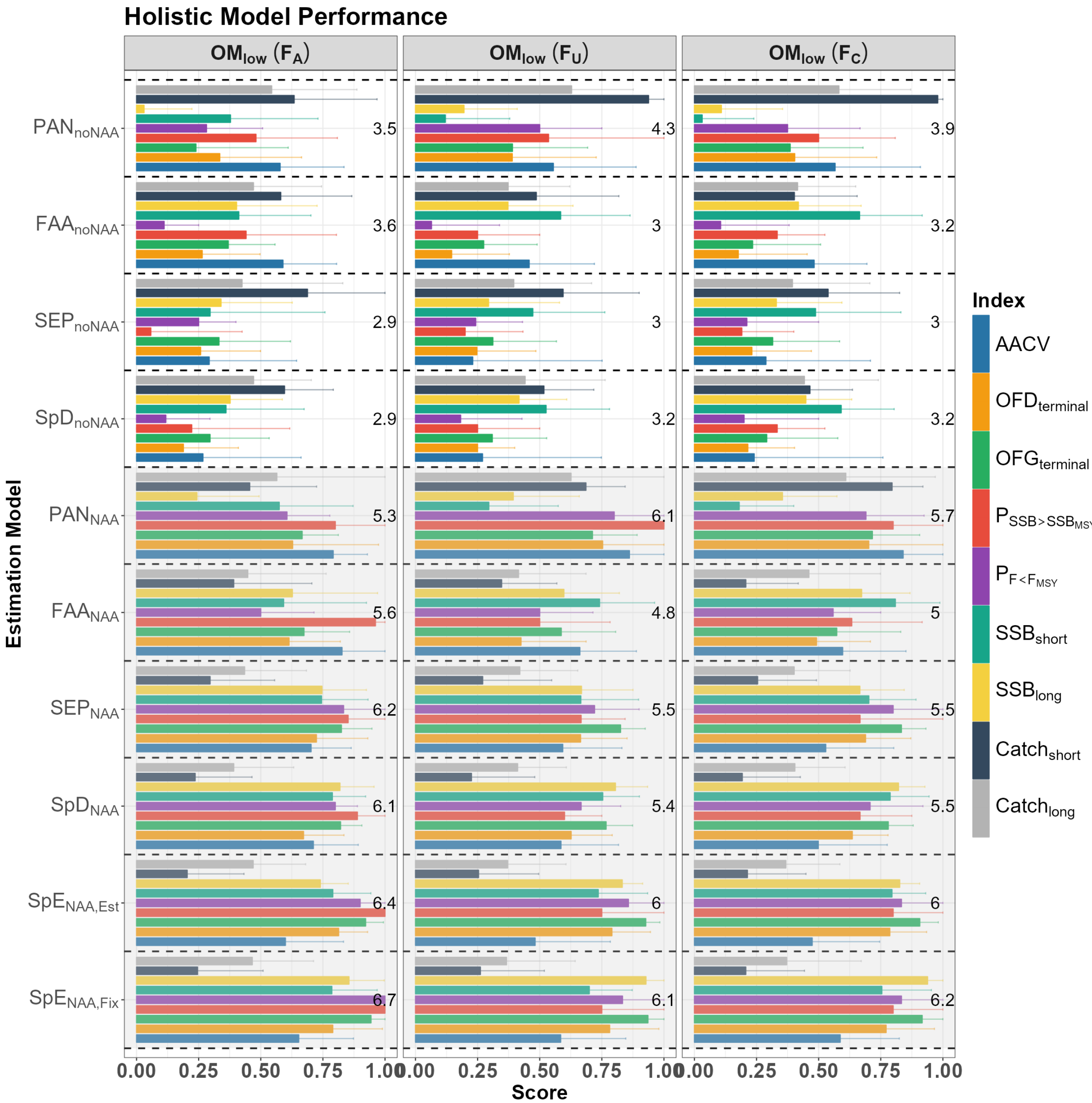

Figure. S34. Overall performance of each EM in region 1 under the low movement scenario ($\mathrm{OM_{low}}$), including the following relative performance metrics: 1) average annual catch variation (AACV); 2) overfished status in the terminal year ($\mathrm{OFD_{terminal}}$); 3) overfishing status in the terminal year ($\mathrm{OFG_{terminal}}$); 4) probability of $SSB > SSB_{MSY}$ ($\mathrm{P_{SSB>SSB_{MSY}}}$); 5) probability of $F < F_{MSY}$ ($\mathrm{P_{F<F_{MSY}}}$); 6) short-term SSB ($\mathrm{SSB_{short}}$); 7) long-term SSB ($\mathrm{SSB_{long}}$), 8) short-term catch ($\mathrm{Catch_{short}}$); and 9) long-term catch ($\mathrm{Catch_{long}}$). All the indices were standardized to scores between 0 and 1, with higher values indicating better performance. The total score for each EM was provided for each fishing scenario. The black dashed line separates the performance of each EM. EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

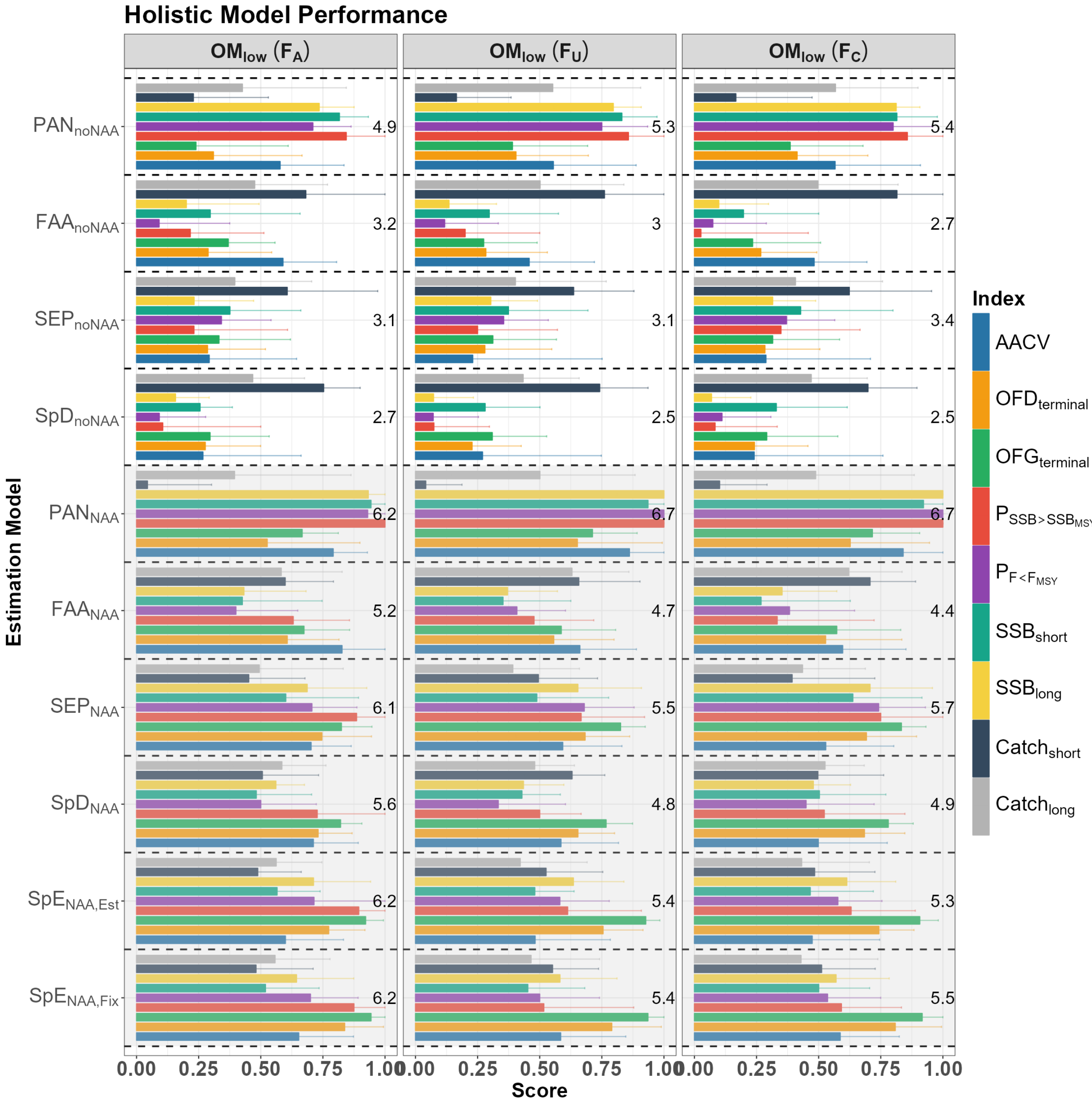

Figure. S35. Overall performance of each EM in region 2 under the low movement scenario (OM_{low}), including the following relative performance metrics: 1) average annual catch variation (AACV); 2) overfished status in the terminal year ($OFD_{terminal}$); 3) overfishing status in the terminal year ($OFG_{terminal}$); 4) probability of $SSB > SSB_{MSY}$ ($P_{SSB>SSB_{MSY}}$); 5) probability of $F < F_{MSY}$ ($P_{F<F_{MSY}}$); 6) short-term SSB (SSB_{short}); 7) long-term SSB (SSB_{long}), 8) short-term catch ($Catch_{short}$); and 9) long-term catch ($Catch_{long}$). All the indices were standardized to scores between 0 and 1, with higher values indicating better performance. The total score for each EM was provided for each fishing scenario. The black dashed line separates the performance of each EM. EMs with a grey background indicate that NAA random effects were included, while EMs with a white background indicate that NAA random effects were not included.

References

Albertsen, C. M., Nielsen, A., and Thygesen, U. H. 2018. Connecting single-stock assessment models through correlated survival. ICES Journal of Marine Science **75**(1): 235–244.

Berger, A. M., Barceló, C., Goethel, D. R., Hoyle, S. D., Lynch, P. D., McKenzie, J., Dunn, A., Punt, A. E., Methot, R. D., Hampton, J., et al. 2024. Synthesizing the spatial functionality of contemporary stock assessment software to identify future needs for next generation assessment platforms. Fisheries Research **275**: 107008.

Berger, A. M., Deroba, J. J., Bosley, K. M., Goethel, D. R., Langseth, B. J., Schueller, A. M., and Hanselman, D. H. 2021. Incoherent dimensionality in fisheries management: consequences of misaligned stock assessment and population boundaries. ICES Journal of Marine Science **78**(1): 155–171.

Berger, A. M., Goethel, D. R., Lynch, P. D., Quinn, T., Mormede, S., McKenzie, J., and Dunn, A. 2017. Space oddity: the mission for spatial integration. Canadian Journal of Fisheries and Aquatic Sciences **74**(11): 1698–1716.

Berger, A. M., Jones, M. L., Zhao, Y., and Bence, J. R. 2012. Accounting for spatial population structure at scales relevant to life history improves stock assessment: the case for lake erie walleye sander vitreus. Fisheries Research **115**: 44–59.

Beverton, R. 1957. On the dynamics of exploited fish population. Fish. Invest. IV .

Bosley, K. M., Goethel, D. R., Berger, A. M., Deroba, J. J., Fenske, K. H., Hanselman, D. H., Langseth, B. J., and Schueller, A. M. 2019. Overcoming challenges of harvest quota allocation in spatially structured populations. Fisheries Research **220**: 105344.

Bosley, K. M., Schueller, A. M., Goethel, D. R., Hanselman, D. H., Fenske, K. H., Berger, A. M., Deroba, J. J., and Langseth, B. J. 2022. Finding the perfect mismatch: Evaluating misspecification of population structure within spatially explicit integrated population models. Fish and Fisheries **23**(2): 294–315.

Cadigan, N. G. 2016. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences **73**(2): 296–308.

Cadrin, S. X. 2020. Defining spatial structure for fishery stock assessment. Fisheries Research **221**: 105397.

Cadrin, S. X., Goethel, D. R., Morse, M. R., Fay, G., and Kerr, L. A. 2019. ”so, where do you come from?” the impact of assumed spatial population structure on estimates of recruitment. Fisheries Research **217**: 156–168.

Ciannelli, L., Fisher, J. A., Skern-Mauritzen, M., Hunsicker, M. E., Hidalgo, M., Frank, K. T., and Bailey, K. M. 2013. Theory, consequences and evidence of eroding population spatial structure in harvested marine fishes: a review. Marine Ecology Progress Series **480**: 227–243.

Cope, J. M. and Punt, A. E. 2011. Reconciling stock assessment and management scales under conditions of spatially varying catch histories. Fisheries Research **107**(1-3): 22–38.

Fisch, N., Shertzer, K., Camp, E., Maunder, M., and Ahrens, R. 2023. Process and sampling variance within fisheries stock assessment models: estimability, likelihood choice, and the consequences of incorrect specification. ICES Journal of Marine Science **80**(8): 2125–2149.

Frisk, M., Martell, S., Miller, T., and Sosebee, K. 2010. Exploring the population dynamics of winter skate (leucoraja ocellata) in the georges bank region using a statistical catch-at-age model incorporating length, migration, and recruitment process errors. Canadian Journal of Fisheries and Aquatic Sciences **67**(5): 774–792.

Goethel, D. R., Berger, A. M., and Cadrin, S. X. 2023. Spatial awareness: good practices and pragmatic recommendations for developing spatially structured stock assessments. Fisheries Research **264**: 106703.

Goethel, D. R., Bosley, K. M., Hanselman, D. H., Berger, A. M., Deroba, J. J., Langseth, B. J., and Schueller, A. M. 2019. Exploring the utility of different tag-recovery experimental designs for use in spatially explicit, tag-integrated stock assessment models. Fisheries Research **219**: 105320.

Goethel, D. R., Bosley, K. M., Langseth, B. J., Deroba, J. J., Berger, A. M., Hanselman, D. H., and Schueller, A. M. 2021. Where do you think you're going? accounting for ontogenetic and climate-induced movement in spatially stratified integrated population assessment models. Fish and Fisheries **22**(1): 141–160.

Goethel, D. R., Kerr, L. A., and Cadrin, S. X. 2016. Incorporating spatial population structure into the assessment-management interface of marine resources. *In* Management Science in Fisheries, pp. 339–367, Routledge.

Goethel, D. R., Omori, K. L., Punt, A. E., Lynch, P. D., Berger, A. M., de Moor, C. L., Plagányi, É. E., Cope, J. M., Dowling, N. A., McGarvey, R., et al. 2022. Oceans of plenty? challenges, advancements, and future directions for the provision of evidence-based fisheries management advice. Reviews in Fish Biology and Fisheries **33**(2): 375–410.

Goethel, D. R., Quinn, T. J., and Cadrin, S. X. 2011. Incorporating spatial structure in stock assessment: movement modeling in marine fish population dynamics. Reviews in Fisheries Science **19**(2): 119–136.

Gudmundsson, G. and Gunnlaugsson, T. 2012. Selection and estimation of sequential catch-at-age models. Canadian Journal of Fisheries and Aquatic Sciences **69**(11): 1760–1772.

Hanselman, D. H., Heifetz, J., Echave, K. B., and Dressel, S. C. 2015. Move it or lose it: movement and mortality of sablefish tagged in alaska. Canadian Journal of Fisheries and Aquatic Sciences **72**(2): 238–251.

Hintzen, N., Roel, B., Benden, D., Clarke, M., Egan, A., Nash, R., Rohlf, N., and Hatfield, E. 2015. Managing a complex population structure: exploring the importance of information from fisheries-independent sources. ICES Journal of Marine Science **72**(2): 528–542.

Hurtado-Ferro, F., Punt, A. E., and Hill, K. T. 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fisheries Research **158**: 102–115.

Kerr, L. A. and Goethel, D. R. 2014. Simulation modeling as a tool for synthesis of stock identification information. *In* Stock identification methods, pp. 501–533, Elsevier.

Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L. W., Dickey-Collas, M., Goethel, D. R., Hatfield, E. M., Kritzer, J. P., and Nash, R. D. 2017. Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. ICES Journal of Marine Science **74**(6): 1708–1722.

Keymer, J. E., Marquet, P. A., Velasco-Hernández, J. X., and Levin, S. A. 2000. Extinction thresholds and metapopulation persistence in dynamic landscapes. The American Naturalist **156**(5): 478–494.

Lee, H.-H., Piner, K. R., Maunder, M. N., Taylor, I. G., and Methot Jr, R. D. 2017. Evaluation of alternative modelling approaches to account for spatial effects due to age-based movement. Canadian Journal of Fisheries and Aquatic Sciences **74**(11): 1832–1844.

Li, C., Deroba, J. J., Miller, T. J., Legault, C. M., and Perretti, C. T. 2024. An evaluation of common stock assessment diagnostic tools for choosing among state-space models with multiple random effects processes. Fisheries Research **273**: 106968.

Li, C., Deroba, J. J., Miller, T. J., Legault, C. M., and Perretti, C. T. In review. Guidance on bias-correction of log-normal random effects and observations in state-space assessment models. For Canadian Journal of Fisheries and Aquatic Sciences, State-space special collection .

Li, Y., Bence, J. R., and Brenden, T. O. 2018. Can spawning origin information of catch or a recruitment penalty improve assessment and fishery management performance for a spatially structured stock assessment model? Canadian Journal of Fisheries and Aquatic Sciences **75**(12): 2136–2148.

Miller, T. J., Curti, K., and Hansell, A. In review. Space for wham: a multi-region, multi-stock generalization of the woods hole assessment model with an application to black sea. For Canadian Journal of Fisheries and Aquatic Sciences, State-space special collection .

Nielsen, A. and Berg, C. W. 2014. Estimation of time-varying selectivity in stock assessments using state-space models. Fisheries Research **158**: 96–101.

O'Boyle, R., Dean, M., and Legault, C. M. 2016. The influence of seasonal migrations on fishery selectivity. ICES Journal of Marine Science **73**(7): 1774–1787.

Perretti, C. T., Deroba, J. J., and Legault, C. M. 2020. Simulation testing methods for estimating misreported catch in a state-space stock assessment model. ICES Journal of Marine Science **77**(3): 911–920.

Pörtner, H. O. and Peck, M. A. 2010. Climate change effects on fishes and fisheries: towards a cause-and-effect understanding. Journal of fish biology **77**(8): 1745–1779.

Punt, A. E. 2019. Spatial stock assessment methods: a viewpoint on current issues and assumptions. Fisheries Research **213**: 132–143.

Punt, A. E., Dunn, A., Elvarsson, B. Þ., Hampton, J., Hoyle, S. D., Maunder, M. N., Methot, R. D., and Nielsen, A. 2020. Essential features of the next-generation integrated fisheries stock assessment package: a perspective. Fisheries Research **229**: 105617.

Punt, A. E., Haddon, M., Little, L. R., and Tuck, G. N. 2016a. Can a spatially-structured stock assessment address uncertainty due to closed areas? a case study based on pink ling in australia. Fisheries Research **175**: 10–23.

Punt, A. E., Haddon, M., Little, L. R., and Tuck, G. N. 2016b. The effect of marine closures on a feedback control management strategy used in a spatially aggregated stock assessment: a case study based on pink ling in australia. Canadian Journal of Fisheries and Aquatic Sciences **74**(11): 1960–1973.

Punt, A. E., Haddon, M., Little, L. R., and Tuck, G. N. 2017. The effect of marine closures on a feedback control management strategy used in a spatially aggregated stock assessment: a case study based on pink ling in australia. Canadian Journal of Fisheries and Aquatic Sciences **74**(11): 1960–1973.

Punt, A. E., Haddon, M., and Tuck, G. N. 2015. Which assessment configurations perform best in the face of spatial heterogeneity in fishing mortality, growth and recruitment? a case study based on pink ling in australia. Fisheries Research **168**: 85–99.

Stock, B. C. and Miller, T. J. 2021. The woods hole assessment model (wham): A general state-space assessment framework that incorporates time- and age-varying processes via random effects and links to environmental covariates. Fisheries Research **243**: 105967.

Stock, B. C., Xu, H., Miller, T. J., Thorson, J. T., and Nye, J. A. 2021. Implementing two-dimensional autocorrelation in either survival or natural mortality improves a state-space assessment model for southern new england-mid atlantic yellowtail flounder. Fisheries Research **237**: 105873.

Waterhouse, L., Sampson, D. B., Maunder, M., and Semmens, B. X. 2014. Using areas-as-fleets selectivity to model spatial fishing: asymptotic curves are unlikely under equilibrium conditions. Fisheries Research **158**: 15–25.