

# Opakapaka Simulation Report

Megumi Oshima, Marc Nadon, John Syslo, Hongguang Ma, Felipe Carvalho

February 27, 2026

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Methods</b>	<b>3</b>
2.1	Simulation Approach . . . . .	3
2.2	Operating Model . . . . .	3
2.3	Estimation Model . . . . .	8
2.4	Performance Indices . . . . .	8
<b>3</b>	<b>Results</b>	<b>10</b>
3.1	Normal Recruitment Conditions . . . . .	10
3.2	Poor Recruitment Conditions . . . . .	15
<b>4</b>	<b>Discussion</b>	<b>17</b>

 [GitHub Repository](#)

## 1 Introduction

The crimson jobfish (*Pristipomoides filamentosus*), locally known as 'ōpakapaka, represents one of Hawaii's most culturally and economically significant deep-water bottomfish species, comprising over 50% of the total Deep-7 bottomfish catch and generating close to \$1 million annually in commercial value (Luers, DeMartini, and Humphreys 2017). As a slow-growing, long-lived species attaining up to 43 years of age (Andrews et al. 2012) and inhabiting depths of 100-400 meters around the Main Hawaiian Islands (Luers, DeMartini, and Humphreys 2017), 'ōpakapaka supports both vital commercial and non-commercial fisheries that are deeply embedded in Hawaiian culture and food security. The species is currently managed under the Western Pacific Regional Fishery Management Council's Hawaii Fishery Ecosystem Plan as part of the Deep-7 bottomfish complex (WPRFMC 2009).

The current stock assessment model for the Deep-7 complex combines all seven species together in a surplus production model that relies on combined catch and index of abundance data. A single-species model for 'ōpakapaka has been developed for research and to support the Deep-7 assessment but it has not been used for management. Because 'ōpakapaka is so dominate in the fishery, it is one of the few species out of the seven that we can build an integrated, age-structured model for. The benefit of transitioning to an integrated model is that it allows us to incorporate more data such as length or age compositions and species-specific life-history information. However, collecting the information that goes into an integrated assessment can be resource- and time-intensive. Fishery-independent surveys can provide high-quality, non-biased data that reflect the true population dynamics of targeted species, however, they can be expensive to run on an annual basis. But, if too little effort is put into the survey and the quality of the data collected is poor, it is not informative in a stock assessment model. Therefore, it is important to optimize the sampling procedures of a survey to provide the most informative data while also balancing the costs and time required.

The Bottomfish Fishery-Independent Survey in Hawai'i (BFISH) utilizes a stratified-random sampling design to estimate the abundance of the Deep-7 bottomfish complex—including 'ōpakapaka—across the main Hawaiian Islands. Since its inception in 2016, the survey has historically integrated two sampling modalities: remote stereo-video camera systems and hook-and-line research fishing. However, in response to evolving operational and financial constraints, the survey transitioned to a single-gear approach in 2025, relying exclusively on research fishing. Sampling is conducted within  $500 \times 500$  m Primary Sampling Unit (PSU) grids distributed across 24 distinct strata. A standard research fishing sample consists of 30 minutes of active fishing within a PSU using two four-hook lines baited with squid and fish (Richards 2025). This method provides critical species-level identification, fork-length measurements, and life-history data. Historically, camera deployments complemented this data, with PSU-level metrics derived from the average of two 15-minute deployments (Richards 2025). Collectively, these inputs are synthesized into indices of relative abundance to inform population trends within the Deep-7 stock assessment framework (Ducharme-Barth, N. 2023). The scale of the BFISH survey has fluctuated over time due to logistical and budgetary shifts. Because the total number of sampled grids directly influences the precision of survey-derived indices, it is essential to quantify the relationship between sampling intensity and the reliability of stock assessment outputs. Furthermore, as environmental conditions shift and resource constraints persist, evaluating the implications of alternative sampling designs is necessary to optimize survey efficiency and ensure the long-term stability of management advice.

The objective of this project is to understand how we optimize an operational fishery-independent survey for 'ōpakapaka in the main Hawaiian islands by testing the impact of effort of the survey under two future recruitment conditions on our ability to estimate key stock quantities such as spawning stock biomass, fishing mortality, and management reference points.

## 2 Methods

### 2.1 Simulation Approach

We tested the impact of varying sampling efforts of a fishery-independent survey by simulating abundance trends and data for 'ōpakapaka in the main Hawaiian islands given appropriate process and observational error. We fit the simulated data to an age-structured catch-at-age model and estimated spawning stock biomass, fishing mortality, and management reference points. Both the operating model (OM) and estimation model (EM) were built in Stock Synthesis (Methot and Wetzel 2013) (SS3) and the simulation-estimation process was implemented using *ss3sim* Johnson et al. (2024) in the R statistical software environment (R Core Team 2024). In total, 1,000 iterations were run using the Open Science Grid OSG (2015). Each run involved (1) generating the population dynamics from the OM, with the given F and recruitment deviations timeseries, (2) sampling data with error from the OM, and (3) fitting those data in an EM. For each run, estimated quantities of interest were compared to the “true” values as calculated by the OM. All code for the analysis is available in the [GitHub repository](#).

### 2.2 Operating Model

The operating model was a single-species, single-sex, age-structured model conditioned on 'ōpakapaka life history parameters and historical fishery data trends from the Deep-7 complex. Historical data included 75 years of catch, CPUE, and size data from commercial and non-commercial fisheries, plus seven years of abundance indices and length composition data from the Bottomfish Fishery-Independent Survey in Hawaii (BFISH). The OM simulated population dynamics and generated data with observation error for 100 years. Fishing mortality time series were generated based on historical patterns: low values at fishery inception, increasing through the 1970s and remaining high until the late 1990s, then declining in the early 2000s and stabilizing through the end of the time series (Figure 1). Annual fishing mortality values included stochastic variability (standard deviation of 0.002) to introduce realistic variation between iterations. Following the most recent Deep-7 benchmark stock assessment, we applied time-varying ratios for non-commercial catch: 1.94 from 1948-2003 and 1.09 for 2004-2049 (Syslo et al. 2024).

#### 2.2.1 Recruitment Deviations

The simulation framework was structured to evaluate estimation model performance under varying future conditions. The first 75 years of the 100-year simulation represented the historical period with known data patterns, while the final 25 years constituted “future years” designed to test the robustness of the estimation model against potential changes in recruitment dynamics. Two distinct recruitment scenarios were implemented for these future years

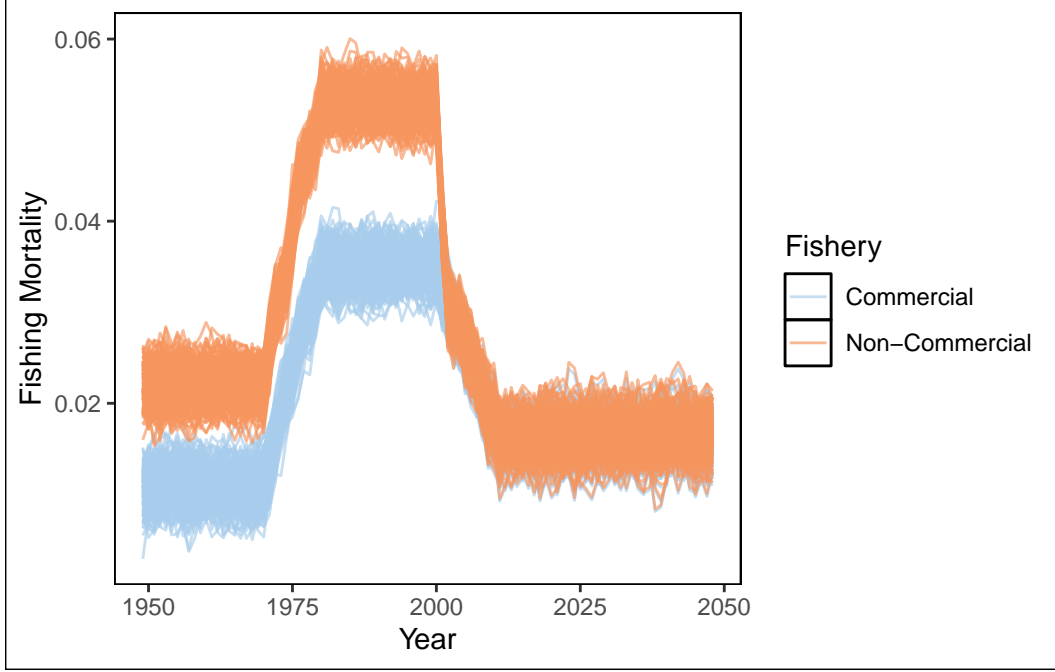


Figure 1: Fishing mortality time series for the commercial (blue lines) and non-commercial (orange lines) fisheries.

to bracket plausible recruitment conditions: (1) a “normal” recruitment scenario that maintained historical mean recruitment levels with no systematic change over time, and (2) a “poor recruitment” scenario that imposed a 40% decline in mean recruitment relative to historical levels. This approach allowed us to assess whether survey sampling intensity affects the estimation model’s ability to detect and respond to recruitment shifts that could significantly impact stock productivity and sustainable harvest levels. A total of 200 recruitment deviation time series were generated, 100 were independent, bias-corrected lognormal random deviates with a standard deviation ( $\sigma_R$ ) of 0.52 for the entire 100 years, and the remaining 100 were the same for the first 75 years and then the last 25 years were calculated to drive a decline in recruitment.

The recruitment decline scenario was structured as a two-phase process spanning 25 years. The first phase consisted of a 10-year gradual decline period, during which recruitment decreased from baseline levels to 60% of the historical mean. The annual decline rate ( $r$ ) was calculated as:

$$r = (r_{reduced}^{\frac{1}{10}}) - 1$$

where  $r_{reduced}$  is the proportion by which the stock was reduced. Because of the non-linear relationship between spawning biomass and recruitment, the reduced level did not actually need to be 60%. We used an iterative process to determine what reduced proportion actually resulted in 60% of the mean recruitment ( $r_{reduced} = 0.71$ ). This formulation ensured a smooth

exponential decline reaching the target reduction after 10 years. The second phase maintained recruitment at the reduced level (60% of baseline) for an additional 15 years, representing a new regime of lower recruitment conditions.

The baseline recruitment level was established using the mean recruitment from the most recent 15-year period (280,000 individuals), representing contemporary recruitment conditions prior to the simulated decline.

For the decline phase, annual multipliers were calculated as:

$$M_t = (1 + r)^t$$

where  $t$  represents the ten years of the decline. For the lower recruitment regime period (years 11-25), multipliers were set to  $r_{reduced}$ , maintaining recruitment at 60% of baseline levels.

Projected recruitment levels were converted to recruitment deviations in log-space to maintain consistency with stock assessment model parameterization. Base recruitment deviations were calculated as:

$$\delta_t = \ln\left(\frac{R_t}{R}\right) - \frac{\sigma_R^2}{2}$$

where  $R_t$  represents projected recruitment in year  $t$ ,  $R$  is the baseline mean recruitment (279.67), and  $\sigma_R$  is the recruitment variability parameter (0.52). The bias correction term ( $\sigma_R^2/2$ ) accounts for the transformation between arithmetic and log scales.

To incorporate realistic recruitment variability around the deterministic decline trend, we generated 100 Monte Carlo iterations of the poor recruitment time series. For each iteration, recruitment deviations were modified by adding random noise drawn from a normal distribution with mean = 0 and standard deviation = 0.15. This additional variability was intentionally set smaller than the base recruitment variability ( $\sigma_R = 0.52$ ) to ensure the underlying decline signal remained detectable above the noise.

The final recruitment deviations for each iteration  $i$  were calculated as:

$$\delta_{t,i} = \delta_t + \varepsilon_{t,i}$$

where  $\varepsilon_{t,i} \sim N(0, 0.15^2)$ .

### 2.2.2 Initial Fishing Mortality and Selectivity

Initial fishing mortality was fixed at 0.012 for the commercial fishery and assumed to start at equilibrium for the non-commercial fishery. Logistic selectivity was modeled for the commercial and non-commercial fisheries and a dome-shaped double normal selectivity model was used for the research fishing survey (Figure 3).

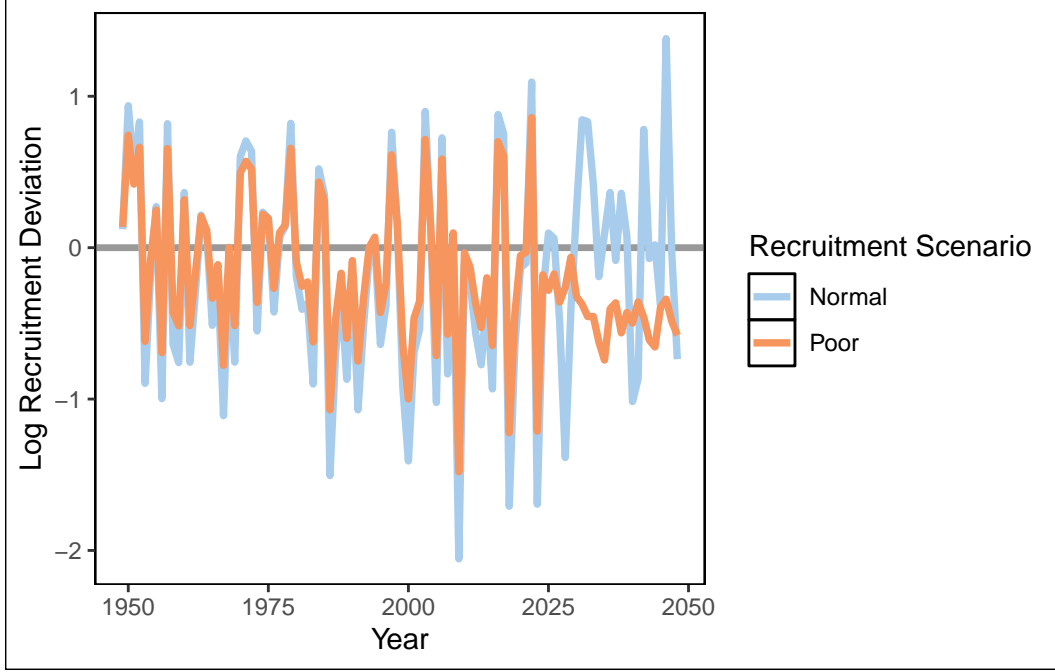


Figure 2: Log recruitment deviations for one iteration used in the normal (blue line) and poor (orange line) recruitment scenarios.

### 2.2.3 Data Generation

We evaluated five sampling scenarios to assess how different levels of survey effort affect stock assessment performance under each recruitment condition (Table 1). Survey effort was defined by the number of sampling grids covered, which directly influenced the precision of the resulting abundance index and length composition data. Catch, CPUE, and length composition data were simulated from the OM based on the population dynamics generated by the specific  $F$  and recruitment deviation time series (Figure 4).

#### Survey Effort Scenarios

The scenarios ranged from high survey effort (HSE) with 632 grids to extra low survey effort (XLSE) with only 257 grids. As expected, higher effort translated to more precise data: index CVs ranged from 15% (HSE) to 30% (XLSE), and effective sample sizes for length compositions ranged from 60 (HSE) to 15 (XLSE). The no survey effort scenario (NSE) represented a baseline case where stock assessments would rely solely on fishery-dependent data.

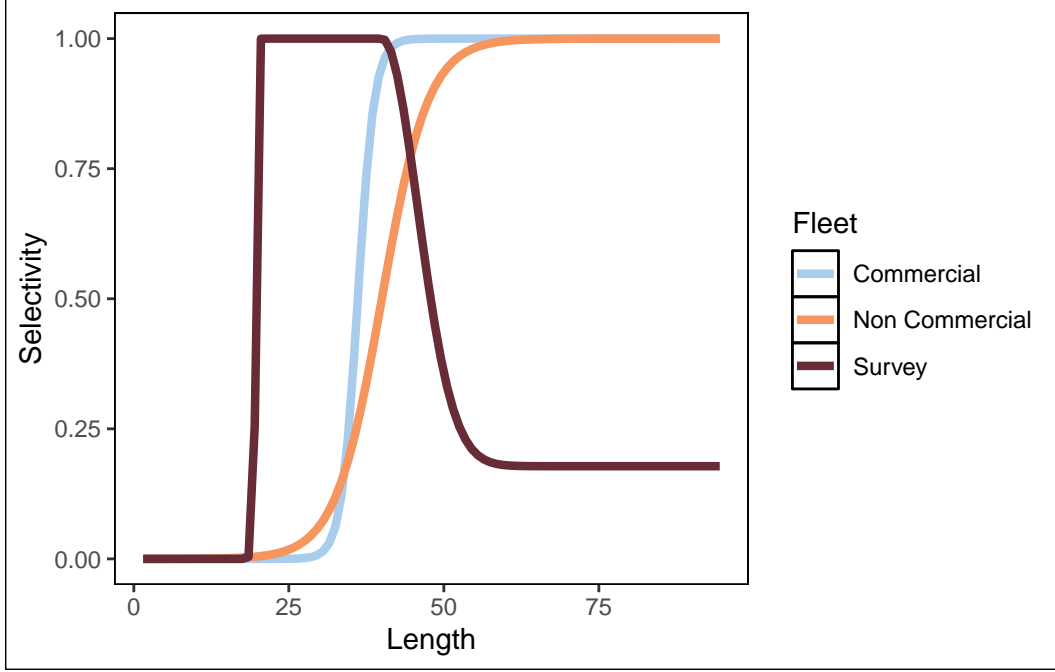


Figure 3: Selectivity functions used for the commercial (blue line), non-commercial (orange line), and survey (red line).

Table 1: The observational error used to sample data from the OM for high survey effort (HSE), intermediate survey effort (ISE), low survey effort (LSE), extra low survey effort (XLSE), and no survey effort (NSE). For each survey effort, the number of sampling grids (Grids) that effort would translate to, the resulting index of abundance CV, and effective sample size for length composition data are shown. All sampling effort scenarios were tested under both normal and poor recruitment conditions.

Sampling Effort	Grids	CV	effN
NSE	0	0.00	0
XLSE	257	0.30	15
LSE	407	0.24	30
ISE	557	0.18	45
HSE	632	0.15	60

### Data Sources by Scenario

All scenarios had some combination of catch, abundance index, and length composition data available for the EM (Figure 4). For all survey scenarios (HSE through XLSE), commercial CPUE and length data were discontinued after 2023, forcing the estimation model to rely entirely on survey-based abundance and length data for the final 25 years. This design allowed

us to isolate the effect of survey data quality on assessment performance. In contrast, the NSE scenario continued using commercial fishery data throughout the entire time series.

### Hyperstability in Fishery Data

For the NSE scenario under poor recruitment conditions, we incorporated hyperstability into the commercial CPUE data to reflect realistic fishery behavior during population decline. The CPUE and catch data were modified to not fully reflect the underlying population decline, simulating the tendency of commercial fisheries to maintain catch rates even as fish abundance decreases through improved fishing efficiency, targeting behavior, or fishing in higher-density areas.

### Observation Error Specification

Sampling error (variability in data sampled from the OM) and observation error specified in the EM were the same, ensuring no misspecification of uncertainty. The CV and effective sample size parameters remained constant across all years within each scenario (Table 1).

## 2.3 Estimation Model

The only structural difference in the EM from the OM was that for the poor recruitment condition, a recruitment regime time block was introduced for the last 15 years. Without the regime, the model was unable to fit the persistent decline in recruitment and the relative errors in outputs were over exaggerated. When fitting the model, all parameters were fixed except for  $R_0$ , the regime parameter for the poor recruitment scenarios, catchability for the commercial and survey fleets, and selectivity parameters for the commercial fishery and survey fleet. All life history and fishery relationships were fixed at their true value (as specified in the OM).

## 2.4 Performance Indices

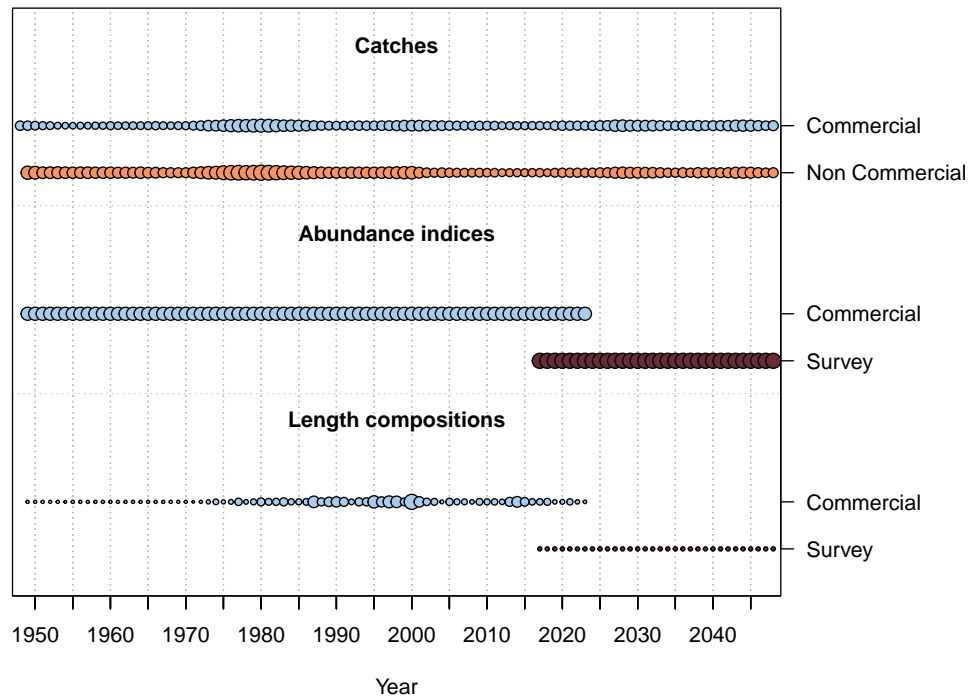
To compare the bias and precision of each sampling strategy, we compared the relative error (RE) between the OM and EM for spawning stock biomass (SSB), fishing mortality (F), age-0 recruitment, and reference points  $SSB/SSB_{MSST}$  and  $F/F_{MSY}$ . Bias was described by the median relative error (MRE)

$$MRE = median(\frac{EM - OM}{OM})$$

where EM is the estimated value from the EM and OM is the true value from the OM, and precision was described by the variability in relative error across iterations. In particular, we report the bias and precision in the terminal year as this provides information on the long-term impacts of sampling effort on assessment and management abilities. We also evaluated the probability that the terminal year reference points ( $SSB/SSB_{MSST}$  and  $F/F_{MSY}$ ) were 10% greater or smaller than the true reference point values. We believe that managers should be cautious of estimates that fall outside of 10% of the true value.



**a) With survey**



**b) Without survey**

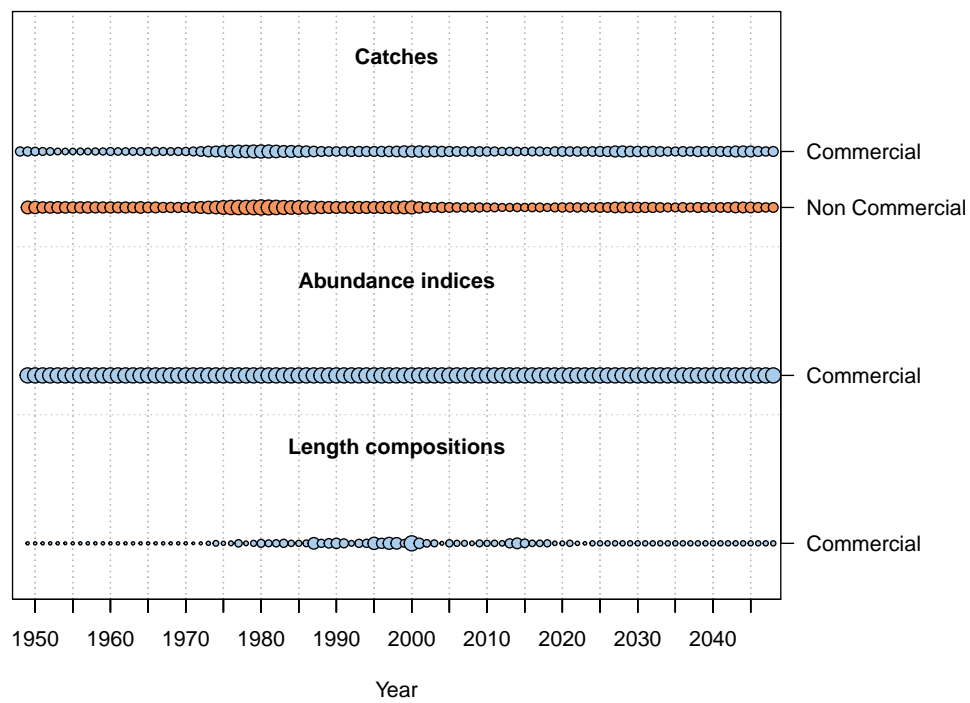


Figure 4: Data generated from operating models and used in the estimation models with the survey (a) and the estimation models without the survey (b). The size of each point represents the relative amount of data available per year and fleet (i.e. large points represent years with more data).

### 3 Results

#### Key results

##### Normal Recruitment Conditions

- All sampling strategies (including no survey effort) showed similar bias for SSB and F estimates (MRE 2-3.6% for SSB, -2.4% to -3.3% for F)
- While bias remained similar, variability increased substantially as sampling effort decreased, with relative error for SSB in the XLSE scenario ranging from -30% to 54%
- Terminal year F and  $F/F_{MSY}$  estimates showed greater sensitivity to extremely low sampling effort (XLSE) compared to SSB-based metrics
- Persistent overestimation of recruitment led to optimistic bias in SSB estimates across all scenarios

##### Poor Recruitment Conditions

- NSE scenario showed dramatically higher error rates (42% MRE for SSB) compared to other scenarios (7.5-10% MRE)
- Performance gaps between high and low/no effort scenarios widened substantially under recruitment failure
- All models failed to accurately capture poor stock conditions, with most runs showing overly optimistic stock status estimates (interquartile ranges rarely included true values)
- While NSE captured true  $F/F_{MSY}$  values more often, it came with much wider uncertainty ranges, potentially leading to management overreactions

Out of 1,000 EM runs, 8 runs did not converge (maximum gradient was greater than  $1e-4$ ). The EM with NSE under poor recruitment conditions had the lowest convergence rate with four iterations that did not converge but overall, model convergence for all sampling scenarios was high.

#### 3.1 Normal Recruitment Conditions

Under normal recruitment, all sampling strategies exhibited similar bias and precision for terminal-year SSB (Table 2). Relative error remained low, ranging from 2% ( $SD = 0.154$ )

Table 2: Median relative error and standard deviation of relative error in the terminal year for spawning biomass, recruitment, and fishing mortality by scenario.

Scenario	Spawning Biomass	Fishing Mortality	Recruitment
<b>Normal Recruitment</b>			
NSE	0.036 (0.100)	-0.031 (0.100)	0.359 (0.931)
XLSE	0.020 (0.154)	-0.024 (0.151)	0.304 (1.400)
LSE	0.036 (0.136)	-0.029 (0.130)	0.351 (1.410)
ISE	0.026 (0.115)	-0.025 (0.110)	0.350 (1.412)
HSE	0.035 (0.109)	-0.033 (0.103)	0.353 (1.418)
<b>Poor Recruitment</b>			
NSE	0.416 (0.278)	0.041 (0.209)	0.611 (0.387)
XLSE	0.102 (0.162)	-0.097 (0.143)	0.109 (0.258)
LSE	0.082 (0.141)	-0.084 (0.130)	0.206 (0.253)
ISE	0.088 (0.120)	-0.083 (0.110)	0.197 (0.230)
HSE	0.075 (0.109)	-0.071 (0.101)	0.202 (0.232)

to 3.6% ( $SD = 0.136$ ). Notably, the NSE scenario—relying solely on commercial CPUE and length data—performed nearly as well as the highest-effort scenarios. This high performance in the NSE scenario likely stems from the extended duration of the CPUE time series and robust effective sample sizes for fishery length data. However, outside of the NSE, precision for SSB over the final 25 years scaled with sampling intensity (Figure 5). For instance, the 95% error interval for the XLSE scenario was wider (−22% to 38%) than that of the HSE (−12% to 28%), a discrepancy that may significantly influence reference point estimation.

Fishing mortality exhibited similar bias and precision across all scenarios in the terminal year, though estimates were consistently lower than true values (Table 2). Terminal-year  $F$  MRE ranged from −2.4% ( $SD = 0.151$ ) to −3.3% ( $SD = 0.103$ ). Notably, while the XLSE scenario showed the lowest bias, it also demonstrated the lowest precision. Conversely, the HSE scenario—despite having slightly higher bias—offered superior precision. The marginal 1% increase in bias in the HSE is a negligible trade-off for the gain in certainty. Consistent with SSB patterns, and excluding the NSE scenario, precision for  $F$  over the final 25 years declined as sampling effort decreased (Figure 6). The 95% error interval for the XLSE scenario [−23%, 31%] was significantly wider than the more constrained [−18%, 18%] range of the HSE.

Over the final 25-year period, the HSE and ISE scenarios provided the most accurate estimates of true age-0 recruitment, while the XLSE scenario performed the poorest—particularly in the most recent years (Figure 7). Although the XLSE recorded the lowest MRE in the terminal year (Table 2), terminal-year recruitment estimates are inherently unreliable due to limited data observations. A persistent overestimation of age-0 recruitment across all models likely ex-

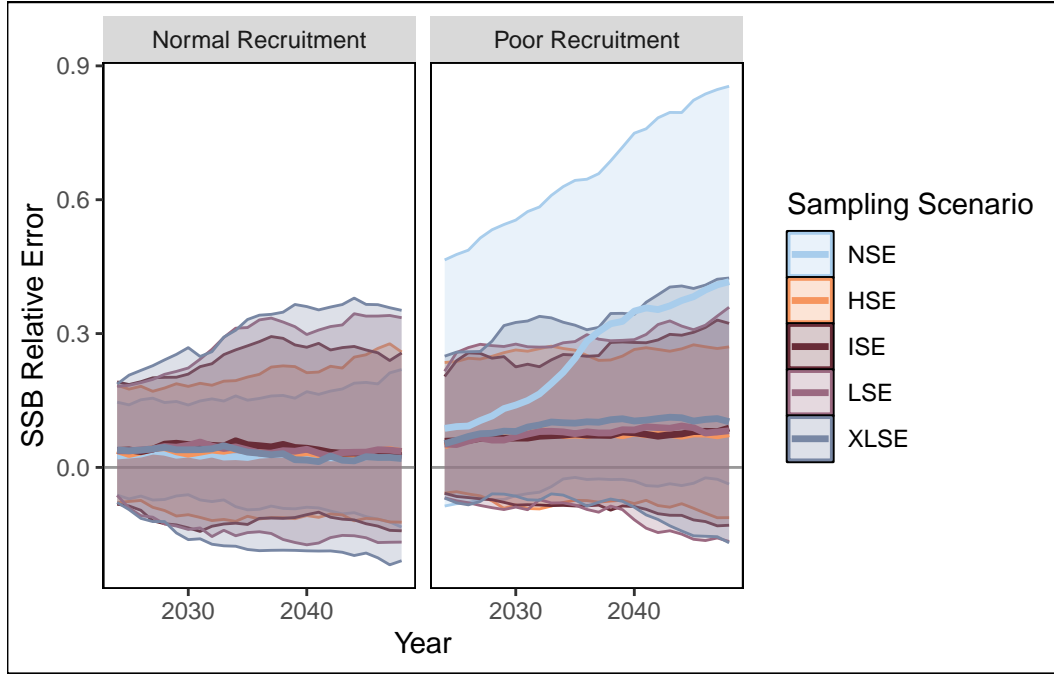


Figure 5: Timeseries of median (line) and 95% quantile (band) relative error of spawning stock biomass for the high (HSE), intermediate (ISE), low (LSE), extra low (XLSE), and no survey effort (NSE) scenarios under both the normal and poor recruitment conditions.

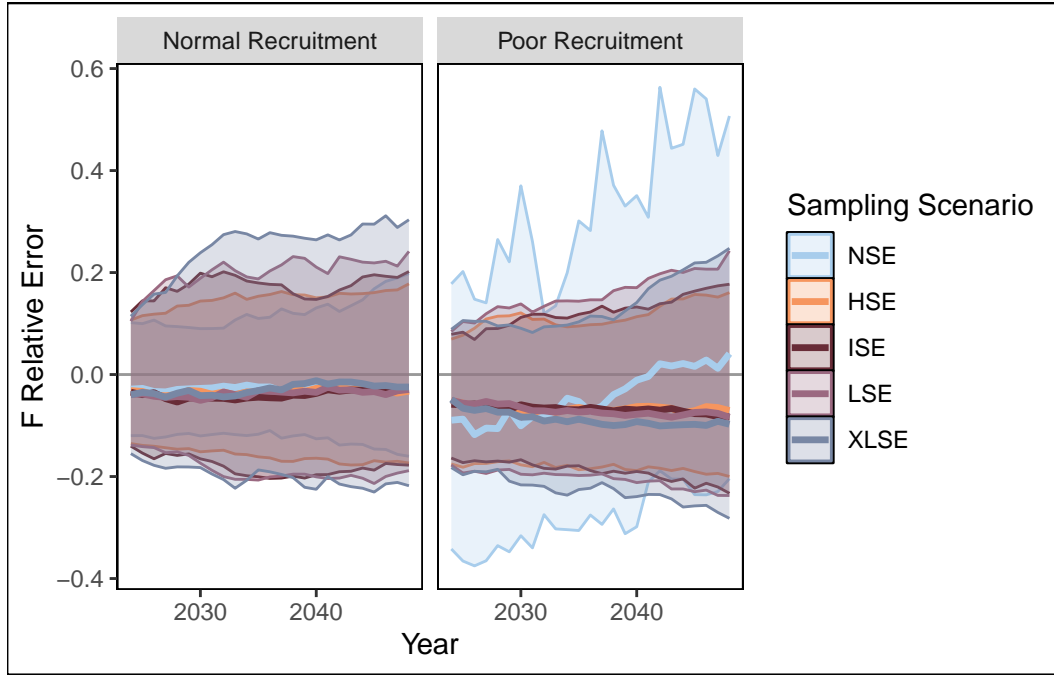


Figure 6: Timeseries of median (line) and 95% interval (band) relative error of fishing mortality for the high (HSE), intermediate (ISE), low (LSE), extra low (XLSE), and no survey effort (NSE) scenarios under both the normal and poor recruitment conditions.

plains the positive bias observed in SSB during the final years of the simulation. Generally, the variability in RE for recruitment was substantially higher than for SSB or  $F$  (Table 2), indicating that the estimation model struggles to precisely capture recruitment dynamics regardless of sampling effort.

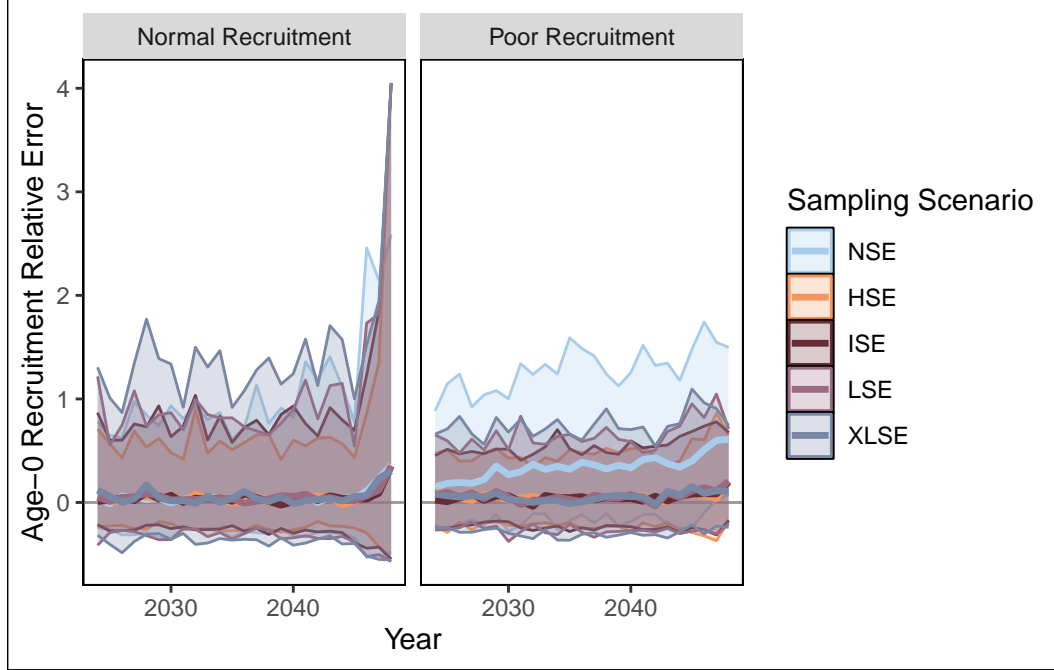


Figure 7: Timeseries of median (line) and 95% interval (band) relative error of age-0 recruitment for the high (HSE), intermediate (ISE), low (LSE), extra low (XLSE), and no survey effort (NSE) scenarios under both the normal and poor recruitment conditions.

Under normal recruitment, all models showed similar bias when estimating stock status ( $SSB/SSB_{MSST}$  and  $F/F_{MSY}$ ), but precision degraded noticeably as sampling effort decreased (Figure 8). The NSE scenario was the sole exception, maintaining bias and precision levels comparable to the HSE. Median relative error for  $SSB/SSB_{MSST}$  ranged from 2.3% to 4% across scenarios (Table 3), with precision (standard deviation) varying from 0.100 (NSE) to 0.153 (XLSE). The 4% precision gap between the XLSE and HSE scenarios represents a meaningful difference that could influence management decisions. Similarly, the MRE for  $F/F_{MSY}$  was tightly clustered between 2.8% and 3.1% for all scenarios (Table 3). However, the standard deviation again revealed a 4% discrepancy between the HSE and XLSE. This suggests that while all sampling intensities yield similar point estimates for terminal-year  $F/F_{MSY}$ , reduced effort compromises precision and increases the risk of significant under- or overestimation.

Lastly, the reliability of key reference point estimates remained consistent across all scenar-

ios under normal recruitment conditions (Table 4). The likelihood of incorrectly estimating  $SSB/SSB_{MSST}$  (defined as a deviation  $> 10\%$ ) was under 50% for nearly all scenarios, ranging from 51% in the LSE to 30% in the NSE. Similarly, the probability of an incorrect  $F/F_{MSY}$  estimate was below 50% for all scenarios, reaching a minimum of 29% in the NSE. These results indicate that under normal recruitment, current sampling efforts generally produce estimates within a 10% margin of the true value the majority of the time.

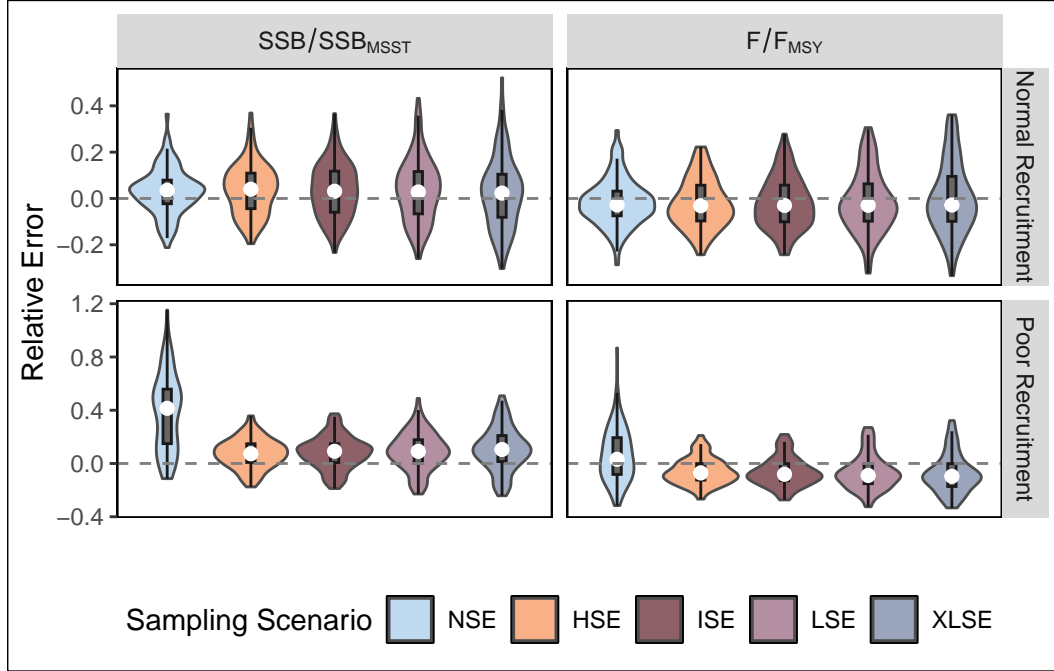


Figure 8: Distribution of relative error for terminal year SSB, spawning stock biomass at minimum stock size threshold ( $SSB_{MSST}$ ), the terminal year stock status ( $SSB/SSB_{MSST}$ ), terminal year  $F$ ,  $F_{MSY}$ , and terminal year fishing status ( $F/F_{MSY}$ ).

### 3.2 Poor Recruitment Conditions

Under poor recruitment conditions, most sampling scenarios exhibited comparable bias and precision for SSB (Figure 5) and  $F$  (Figure 6). However, overall model performance degraded significantly compared to normal recruitment conditions. The NSE scenario was a notable outlier, performing substantially worse than all other configurations with an SSB terminal-year MRE of 42% ( $SD = 0.209$ ) (Table 2). This poor performance is likely driven by the hyperstability of CPUE data inherent in the NSE. For all other scenarios, SSB MRE was more moderate, ranging from 7.5% ( $SD = 0.109$ ) to 10.2% ( $SD = 0.162$ ).

Similarly, estimates of  $F$  were less accurate under poor recruitment (Figure 6). MRE for  $F$

Table 3: Median relative error and standard deviation of relative error in the terminal year for  $SSB/SSB_{MSST}$  and  $F/F_{MSY}$  by scenario.

Scenario	$SSB/SSB_{MSST}$	$F/F_{MSY}$
<b>Normal Recruitment</b>		
NSE	0.035 (0.100)	-0.028 (0.101)
XLSE	0.023 (0.153)	-0.028 (0.152)
LSE	0.028 (0.136)	-0.029 (0.132)
ISE	0.031 (0.115)	-0.030 (0.111)
HSE	0.040 (0.109)	-0.031 (0.105)
<b>Poor Recruitment</b>		
NSE	0.414 (0.284)	0.031 (0.202)
XLSE	0.105 (0.163)	-0.096 (0.145)
LSE	0.091 (0.143)	-0.090 (0.130)
ISE	0.092 (0.122)	-0.081 (0.109)
HSE	0.073 (0.111)	-0.071 (0.099)

Table 4: Probability that reference points  $SSB/SSB_{MSST}$  and  $F/F_{MSY}$  are above or below 10% from the true value for each scenario.

Scenario	$SSB/SSB_{MSST}$	$F/F_{MSY}$
<b>Normal Recruitment</b>		
NSE	0.30	0.29
XLSE	0.49	0.49
LSE	0.51	0.44
ISE	0.43	0.44
HSE	0.39	0.37
<b>Poor Recruitment</b>		
NSE	0.81	0.65
XLSE	0.63	0.59
LSE	0.59	0.57
ISE	0.57	0.52
HSE	0.51	0.48



ranged from  $-7.1\%$  ( $SD = 0.101$ ) in the HSE to  $-9.7\%$  in the XLSE (Table 2). Interestingly, while the NSE scenario yielded the lowest bias for  $F$  over the final decade, it was characterized by much higher variability throughout the projection period, making the estimates less reliable for consistent management.

Age-0 recruitment was estimated with reasonable accuracy in all scenarios except the NSE (Figure 7). Counter-intuitively, terminal-year recruitment estimates were less biased under poor recruitment than under normal recruitment conditions (Table 2). Furthermore, precision was notably higher across all scenarios—including the NSE—throughout the entire projection period (Figure 7).

The HSE and ISE scenarios demonstrated the strongest performance for estimating terminal-year  $SSB/SSB_{MSST}$  and  $F/F_{MSY}$  (Figure 8). These scenarios produced the lowest variability, with MREs between 7% and 9% for  $SSB/SSB_{MSST}$  and  $-7\%$  to  $-8\%$  for  $F/F_{MSY}$  (Table 3). Conversely, the NSE scenario performed the poorest for  $SSB/SSB_{MSST}$ , with an MRE of 41.4% and a standard deviation of 0.280. Critically, the true terminal-year stock status was not recovered in the majority of simulation runs for any scenario, as interquartile ranges failed to include zero. This suggests that under poor recruitment, all tested models tend to provide an overly optimistic view of stock status. The only exception was  $F/F_{MSY}$  under the NSE scenario, where the MRE was near zero; however, this accuracy comes with a trade-off of extreme uncertainty. The wide error range in the NSE increases the risk of severe overestimation of  $F/F_{MSY}$ , potentially leading to unnecessary fishery closures or drastic allocation reductions.

The probability of significant estimation error (deviating  $> 10\%$  from the true value) is markedly higher under poor recruitment conditions (Table 4). The likelihood of incorrectly estimating  $SSB/SSB_{MSST}$  exceeded 50% for all sampling scenarios, reaching 81% in the NSE. Similarly, the probability of an incorrect  $F/F_{MSY}$  estimate was greater than 50% for all scenarios except the HSE, which remained slightly lower at 48%. These results highlight a pervasive risk of mismanagement when recruitment is poor, regardless of the sampling effort employed.

## 4 Discussion

### Key points

- The poor recruitment condition reveals the importance of adequate sampling
- Management reference points become more uncertain as sampling effort decreases
- The NSE scenario suggests that it would be adequate in some situations, however, there are a few critical assumptions that would need to be considered before taking

this approach

- During periods of recruitment failure, enhanced rather than reduced monitoring may be necessary to maintain assessment reliability
- Managers should account for increased uncertainty in stock status during periods of both poor recruitment and reduced survey effort
- While reduced survey effort may seem economically attractive, the substantial increases in assessment uncertainty could lead to suboptimal management decisions with significant long-term costs

Our study demonstrates that the necessity of fishery-independent surveys is highly dependent on recruitment stability. Under normal recruitment, sampling effort primarily dictates the precision rather than the accuracy of biomass and fishing mortality estimates. However, this relationship shifts fundamentally under poor recruitment or regime-type shifts. In these scenarios, survey effort becomes a determinant of both precision and the ability to accurately characterize stock status. This suggests that while reduced effort may be a manageable risk during periods of stability, it compromises assessment reliability if recruitment dynamics begin to falter.

The poor recruitment scenarios revealed a significant optimism bias across all models. When recruitment declined, the interquartile ranges for  $SSB/SSB_{MST}$  and  $F/F_{MSY}$  consistently failed to capture the true values, providing an overly favorable view of stock status. This bias was exacerbated by reduced sampling effort. High and intermediate survey effort scenarios (HSE and ISE) demonstrated comparable performance across all metrics, including median relative error and standard deviation of relative error. Lower effort scenarios (LSE and XLSE) maintained similar central tendency measures but exhibited substantially wider confidence intervals, indicating reduced precision in parameter estimates. This pattern illustrates how poor recruitment conditions amplify the negative consequences of reduced sampling effort, creating a compounding effect where assessment reliability deteriorates precisely when accurate population monitoring becomes most critical for management decisions. The performance degradation observed below intermediate effort levels suggests that survey coverage should be maintained between 550-630 grids to ensure adequate assessment quality for 'ōpākāpāka stock management, particularly given the potential for unexpected recruitment variability in this system.

One interesting finding was the stark contrast of the NSE scenario performances under the two recruitment conditions. The NSE scenario's performance suggests that in some cases, no survey data may be adequate if the fishery CPUE index is accurately tracking the population trends. Having one continuous data source of good quality may be preferable to shorter time series of data or very poor-quality survey data. However, this is only true under the normal recruitment conditions and once a decline in recruitment was introduced, the model performed

far worse than the others in almost every metric. A major assumption of the NSE model is that the fishery data, particularly the CPUE and size data are accurately characterizing the full population. This is likely untrue, as fishers are knowledgeable about where and when to find larger fish and will likely be able to maintain a healthy catch even if the population is declining (i.e. hyperstability). Therefore, we should not assume that fishery-dependent data is an accurate, unbiased representation of the true population dynamics.

Our simulation relied on several key assumptions that may not fully capture the complexity of real-world fishery dynamics and data collection challenges. The NSE scenario under poor recruitment assumed complete hyperstability in fishery data, meaning catch rates remained unaffected by population decline throughout the 25-year projection period. While hyperstability is a documented phenomenon in fishery-dependent CPUE indices, sustained recruitment failure would likely eventually manifest in commercial catch data, suggesting our model may have overestimated the NSE scenario’s limitations. Nevertheless, this represents a plausible “worst-case” scenario that highlights the risks of relying solely on fishery-dependent data during periods of stock decline. Additionally, our model simplified future data availability by excluding fishery CPUE and length composition data after 2023, which artificially increased uncertainty even in high survey effort scenarios. In practice, these fishery data streams would continue but would be appropriately weighted within the assessment framework to balance their influence against survey observations. Finally, the poor recruitment scenarios employed reduced variability in recruitment deviations to clearly establish declining trends, whereas natural recruitment variability typically remains high even during periods of average decline. This simplification may have amplified the apparent impacts of recruitment failure, as realistic variability could obscure declining trends and complicate early detection of regime shifts.

While consistent long-term monitoring remains ideal, a combination of adaptive sampling strategies and improved data collection techniques could help balance assessment reliability with resource constraints. An adaptive approach could involve reducing survey effort during periods when recruitment appears stable and population indicators show no signs of decline, then increasing sampling intensity when early warning signals suggest recruitment failure or population stress. However, such adaptive strategies must be implemented cautiously, as maintaining data continuity over extended periods provides irreplaceable value for detecting long-term trends and regime shifts. Complementary improvements to data collection could enhance assessment quality without necessarily increasing survey effort, such as incorporating annual age data from the fishery-independent survey to better understand population structure and recruitment dynamics, or refining fishing methodology to reduce coefficient of variation in catch estimates.

These results highlight several critical considerations for fishery management. The first is that maintaining adequate survey effort is essential for reliable stock assessments, particularly for fishing mortality estimates used in management decisions. The second is that during periods of recruitment failure, enhanced rather than reduced monitoring may be necessary to maintain assessment reliability. We saw a noticeable decrease in our estimation abilities when recruitment was poor and there was a reduced survey effort. Additionally, scientists and

managers should account for increased uncertainty in stock status during periods of both poor recruitment and reduced survey effort (even under normal recruitment conditions). Lastly, while reduced survey effort may seem economically attractive, the substantial increases in assessment uncertainty could lead to suboptimal management decisions with significant long-term costs for the fishery.

- Anderson, Sean C., Cole C. Monnahan, Kelli F. Johnson, Kotaro Ono, and Juan L. Valero. 2014. “Ss3sim: An R Package for Fisheries Stock Assessment Simulation with Stock Synthesis.” *PLOS ONE* 9 (4): e92725. <https://doi.org/10.1371/journal.pone.0092725>.
- Andrews, Allen H., Edward E. DeMartini, Jon Brodziak, Ryan S. Nichols, and Robert L. Humphreys. 2012. “A Long-Lived Life History for a Tropical, Deepwater Snapper (*Pristipomoides Filamentosus*): Bomb Radiocarbon and Lead–Radium Dating as Extensions of Daily Increment Analyses in Otoliths.” *Canadian Journal of Fisheries and Aquatic Sciences* 69 (11): 1850–69. <https://doi.org/10.1139/f2012-109>.
- Ducharme-Barth, N. 2023. “Model-Based Relative Abundance Indices for the Main Hawaiian Islands Deep 7 Bottomfish Complex, 2016-2022.”
- Johnson, Kelli F., Sean C. Anderson, Kathryn L. Doering, Cole C. Monnahan, and Christine C. Stawitz. 2024. *Ss3sim: Fisheries Stock Assessment Simulation Testing with Stock Synthesis*. <https://github.com/ss3sim/ss3sim>.
- Luers, Meagan A., Edward E. DeMartini, and Robert L. Humphreys. 2017. “Seasonality, Sex Ratio, Spawning Frequency and Sexual Maturity of the Opakapaka *Pristipomoides Filamentosus* (Perciformes: Lutjanidae) from the Main Hawaiian Islands: Fundamental Input to Size-at-Retention Regulations.” *Marine and Freshwater Research* 69 (2): 325–35. <https://doi.org/10.1071/MF17195>.
- Methot, Richard D., and Chantell R. Wetzel. 2013. “Stock Synthesis: A Biological and Statistical Framework for Fish Stock Assessment and Fishery Management.” *Fisheries Research* 142 (May): 86–99. <https://doi.org/10.1016/J.FISHRES.2012.10.012>.
- OSG. 2006. “OSPool.” OSG. <https://doi.org/10.21231/906P-4D78>.
- . 2015. “Open Science Data Federation.” OSG. <https://doi.org/10.21231/0KVZ-VE57>.
- Pordes, Ruth, Don Petravick, Bill Kramer, Doug Olson, Miron Livny, Alain Roy, Paul Avery, et al. 2007. “The Open Science Grid.” In *J. Phys. Conf. Ser.*, 78:012057. 78th Series. <https://doi.org/10.1088/1742-6596/78/1/012057>.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richards, Benjamin. 2025. “Annual Report: 2024 Bottomfish Fishery-Independent Survey in Hawai i.” *NOAA Technical Memorandum NMFS PIFSC ; 177*. <https://doi.org/10.25923/059V-F791>.
- Sfiligoi, Igor, Daniel C Bradley, Burt Holzman, Parag Mhashilkar, Sanjay Padhi, and Frank Wurthwein. 2009. “The Pilot Way to Grid Resources Using glideinWMS.” In *2009 WRI World Congress on Computer Science and Information Engineering*, 2:428–32. 2nd Series. <https://doi.org/10.1109/CSIE.2009.950>.
- Syslo, John, Megumi Oshima, Hongguang Ma, Nicholas Ducharme-Barth, Marc Nadon, and Felipe Carvalho. 2024. “Benchmark Stock Assessment for the Main Hawaiian Islands Deep

7 Bottomfish Complex, with Projections Through 2029.” <https://doi.org/10.25923/5SSG-8D54>.

WPRFMC. 2009. “Fishery Ecosystem Plan for the Hawaii Archipelago.” Western Pacific Regional Fishery Management Council. <http://www.wpcouncil.org/fishery-plans-policiesreports/hawaii-fishery-ecosystem-plan>.