Center for Independent Experts (CIE) independent peer review of Three North Pacific flatfish stock assessments: Yellowfin Sole, Northern Rock Sole, and Alaska Plaice

Matthew D. Cieri

June 2018

Executive Summary

From April 16 to 18, 2018, Joe Powers, Matthew Cieri, and Yan Jiao reviewed three flatfish assessments conducted by Jim Ianelli and Tom Wilderbuer of the North Pacific Fishery Management Council (NPFMC) of National Marine Fisheries Service (NMFS). The stocks examined were Yellowfin Sole, Northern Rock Sole, and Alaska Plaice.

Each of these three stocks inhabit much of the same area of the Bering Sea. They also are the subject of a mostly targeted mixed trawl fishery. All three stocks have similar assessment; utilizing age-structured maximum-likelihood modeling with the ability to model the sexes separately for selectivity, growth, maturity, and other features.

Yellowfin Sole is the main targeted stock in this complex. The assessment incorporated a well-thought-out catchability that incorporated temperature as a covariate. This assessment was further strengthened during the workshop by the addition of both start date and temperature as an interaction term. It was recommended that this new model formulation be thoroughly tested and used for providing management advice.

The 2016 assessment for Northern Rock Sole was provided to the reviewers prior to the workshop. The 2017 assessment, however, was subsequently provided near the end of the meeting. Because most of the meeting documents, including the workshop presentation focused on the 2016 assessment, this is what was reviewed. A plethora of alternative runs were provided for this stock. Overall, the model 15.1 performed well, but it was felt that model 16.3, which modeled the sexes separately for natural mortality, should be the one to use moving forward. Other than that, the major recommendation was that other stocks use a similar alternative runs strategy.

Alaska Plaice was the last stock examined during the workshop. Unlike the other Tier 1a assessments, this stock used recruitment deviations rather than a parametrized stock - recruitment relationship. As such a designation of 3a was likely best for this stock, as recommended. In the interest of time, few alternate runs could be performed for this stock. Like with Northern Rock Sole, it was recommended that a formulation which allowed for the estimation of natural mortality, while keeping catchability constant should be explored in the future. As with both Yellowfin Sole and Alaska Plaice, a full suite of alternate model should be conducted at other benchmarks and updates to ensure that stable states of nature are routinely examined.

Overall, the models presented functioned well, and other than Yellowfin Sole, had little retrospective patterns and good fits. As such they were deemed the best available information in hand and could be used for management advice. All three stocks were found to be not overfished nor have overfishing occurring.

Background

From April 16 to April 18, 2018, Center for Independent Experts (CIE) panelists peer reviewed the stock assessments for Yellowfin Sole, Northern Rock Sole, and Alaska Plaice. The meeting was held at the Alaska Fisheries Science Center (AFSC) located in Seattle, Washington.

The purpose of the review was to examine the most recent stock assessments for three species of commercially important flatfish; for Yellowfin Sole, Northern Rock Sole, and Alaska Plaice. All three assessments were conducted by the team at the science center. All three stocks share many similarities; the data streams are all very similar, the fishery targeting these stocks catches each as part of a multi-species fleet (called the Amendment 80 fleet), each assessment was conducted by the same individuals, all used a similar modeling approach, with some variations in assumptions, and all three assessments produced very similar results.

During the first day, about half of the day was devoted to background information including aging, surveys, observer coverage, management, and other information common to all three stocks. While this information was very helpful in supporting the decisions made at the meeting, they are not specifically addressed here as they are outside the Terms of Reference.

To facilitate the reporting, I have elected to deal with the suite of Terms of References (TORs) under each individual stock. The TORs for all stocks outlined below.

Terms of Reference

1. *Evaluate the strengths and weaknesses of the assumptions made in applying the stock assessment model including how survey indices are scaled to the populations. Specifics might include:*
    a. *How natural mortality estimates are estimated/applied*
    b. *Assumptions about survey "catchability"*
    c. *Application of fishery and survey age-specific schedules (maturity, body mass, selectivity)*
    d. *The application (or lack thereof) of a stock-recruitment relationship (and associated parameter estimates)*
2. *Evaluate the stock assessment approach used focusing specifically on how fisheries and survey data are compiled and used to assess the stock status relative to stated management objectives under the Bering Sea and Aleutian Islands Fishery Management Plan (FMP) and the Magnuson-Stevens Act requirements. Elements should consider:*
    a. *The FMP "Tier" designation*
    b. *Fishing rate estimation relative to overfishing definitions*
    c. *Stock status determinations relative to B_{MSY}*
3. *Recommend how assessment data and/or models could be improved.*

Yellowfin Sole

Assessment summary

The model used for Yellowfin Sole is a separable catch-at-age approach using a maximum likelihood approach and implemented in AD model builder. It is similar to other modeling approaches used for other flat fish in the area.

Model specifics include a natural morality estimated outside the model at 0.12 based on a maximum likelihood analysis of Japanese pair trawling data. For this assessment male and female yellowfin were assigned the same natural mortality, despite some evidence of differences in maturity, growth, and life-span.

Maturity was based on two studies, by Nichol (1995) and TenBrink & Wilderbuer (2015). Both gave very similar maturity schedules given that they were nearly two decades apart.

Data inputs to the model included fishery weight and catch at age, fishery removals, survey swept biomass, and survey age composition as outlined below:

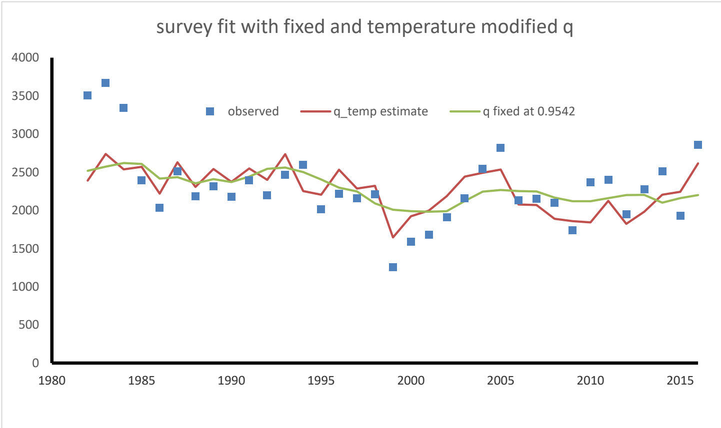
Table 1: Data Sources and Survey Age

Data source	Years
Fishery catch	1954-2017
Fishery age composition	1964-2016
Fishery weight-at-age	Avg wt at age from 2008-16 used for 2008-2016
Survey biomass and standard error, bottom temperature	1982-2017
Survey age composition	1979-2016
Annual length-at-age and weight-at-age from surveys	1979-2016
Maturity at age	Combined 1992 and 2012 samples

Fishery and survey selectivity was modeled with the two sexes separately as a logistic function. For the survey, this was modeled as one selectivity function for each of the two sexes. For the fishery, annual estimates of selectivity by age and sex were used, but constrained to flat-topped.

Stock-recruitment is estimated via a Ricker curve with a preferred data timeframe of 1978-2012.

Survey catchability was estimated annually using temperature as covariates within the model. This dramatically improved the model fit as shown below:



Other sources of data in the model included a bottom trawl survey, with estimates of catchability to scale to a swept areas biomass time series. Additionally, both the survey and the fishery have catch, catch at length, and catch at age information available.

Results are presented below and don't dramatically differ from previous model runs in 2016:

Quantity	As estimated or specified last year for: 2017	2018	As estimated or recommended this year for: 2018	2019
M (natural mortality rate)	0.12	0.12	0.12	0.12
Tier	1a	1a	1a	1a
Projected total (age 6+) biomass (t)	2,290,000	2,202,300	2,553,100	2,460,700
Female spawning biomass (t)	587,300			
Projected	778,600	770,900	895,000	890,000
B_0	1,202,700		1,204,000	
B_{MSY}	424,000		456,000	
F_{OFL}	0.125	0.125	0.12	0.12
$maxF_{ABC}$	0.114	0.114	0.109	0.109
F_{ABC}	0.114	0.114	0.109	0.109
OFL (t)	287,000	276,000	306,700	295,600
maxABC (t)	260,800	250,800	277,500	267,500
ABC (t)	260,800	250,800	277,500	267,500
Status	As determined last year for: 2015	2016	As determined this year for: 2016	2017
Overfishing	No	n/a	No	n/a
Overfished	n/a	No	n/a	No
Approaching overfished	n/a	No	n/a	No

Overall the results seemed reasonable. Model fits to the data were good and the results were not subjected to much uncertainty given the probability plots (see figures below).

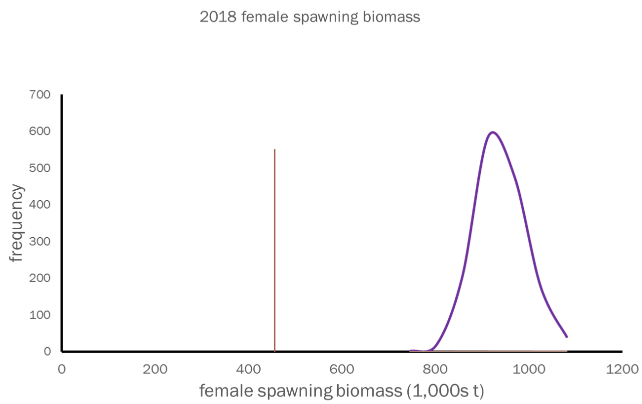
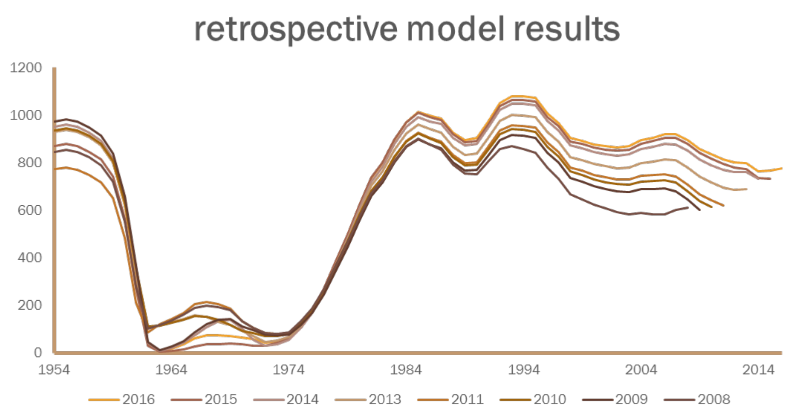


Figure 1: 2018 Female Spawning Biomass

That said, the model does have some issues. In particular, a retrospective pattern is evident suggesting an underestimation of SSB and an overestimation of F relative to the terminal year.



Alternate runs

During the workshop, and at the request of both the lead analyst and the reviewers, a change to the base model was configured to model catchability as both a function of temperature, as in the base run, and start date, given that temperature and start date could be confounded. Hence, the catchability was modeled as:

$$q=e^{-\alpha+\beta T+\gamma S+\mu T:S}$$

Where *T*=survey bottom temperature (averaged per year for all stations <100 m), *S*=survey start date, and *T:S*=interaction of *T* and *S*.

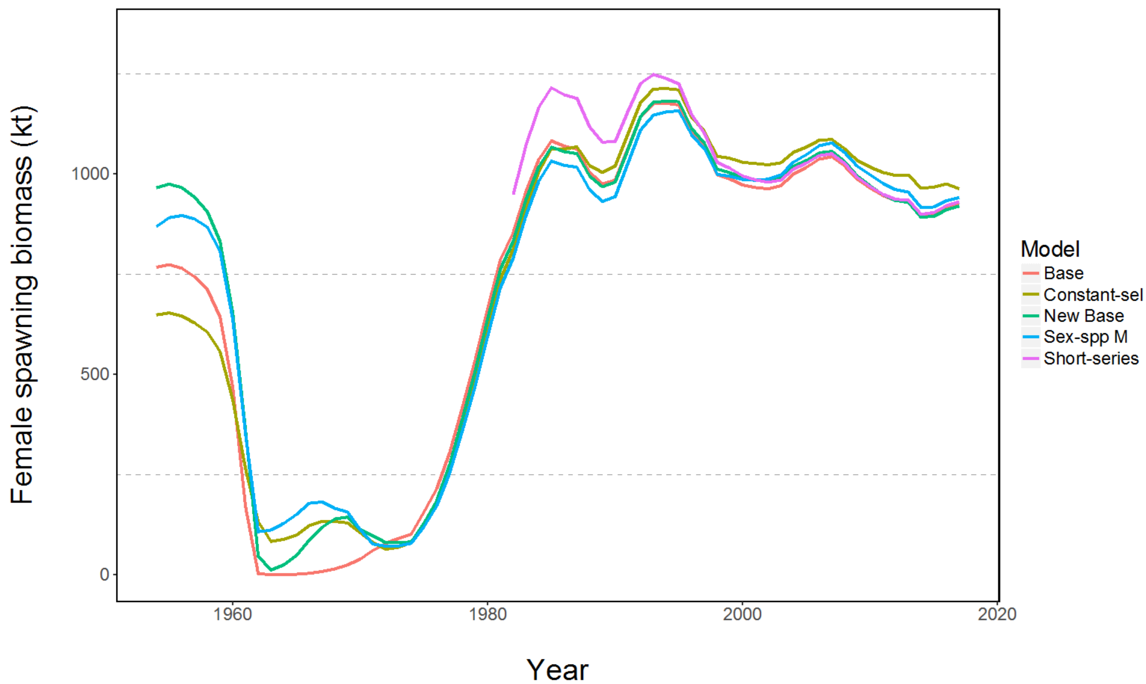
In addition, other alternate runs were suggested including:

- Constant fishery selectivity
- Start the model when all data was available (1982), ignore historical catches and other data

- Examine sex-specific natural mortality
- Time-varying selectivity for survey

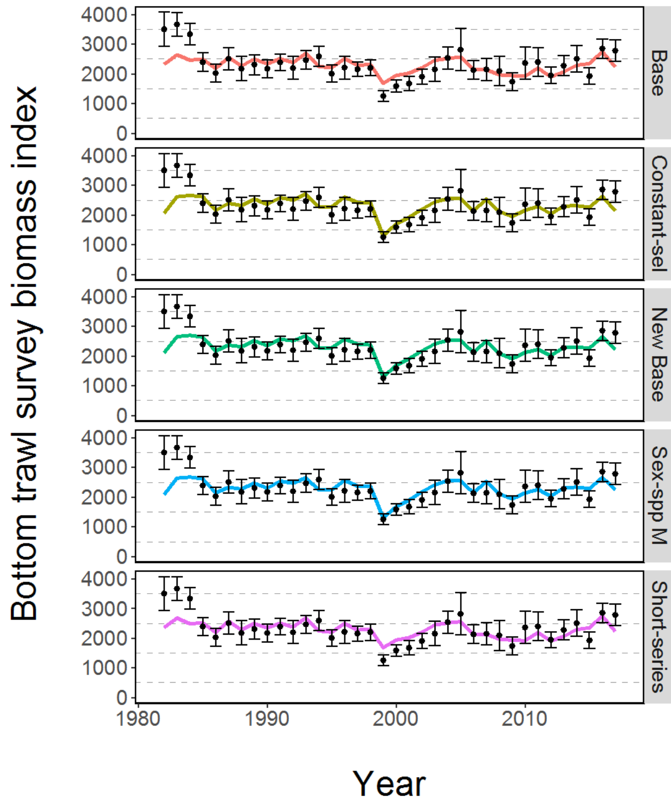
An additional run that allowed for the fishery selectivity to be estimated fully, rather than set as flat-topped was also completed. It resulted in nearly the same selectivity pattern as the one shown in the base run and the new base run, and so is not considered further. Also, the ability to model time varying selectivity for the survey was deemed too labor intensive to accomplish during the meeting, and so was suggested as a future research recommendation.

Results for the runs were rather interesting when compared to both the “old” base run which just used temperature, and the “new” base run which used both temperature and start date.



There was very little change in SSB estimates in the terminal year for each of the runs with the exception of the use of constant selectivity. Further the “new” base run didn’t seem to greatly decrease the retrospective pattern.

Overall, however, the fits to the survey were better with the “New” base run, particularly the time frame between 1999 and 2005.



As such, this "New" base run, which incorporates the interaction between temperature and start date to estimate catchability, was the best run produced so far, though the improvement in fit was only marginal.

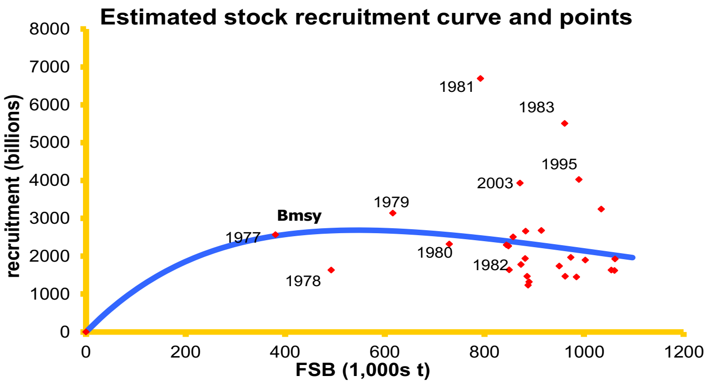
<u>TOR 1</u>

1. *Evaluate the strengths and weaknesses of the assumptions made in applying the stock assessment model including how survey indices are scaled to the populations. Specifics might include:*
    a. *How natural mortality estimates are estimated/applied*
    b. *Assumptions about survey "catchability"*
    c. *Application of fishery and survey age-specific schedules (maturity, body mass, selectivity)*
    d. *The application (or lack thereof) of a stock-recruitment relationship (and associated parameter estimates)*

Overall, the catchability which had temperature in the base run, was further improved when adding in both start date and the interaction term.

The use of natural mortality from a fishery that was prosecuted during the start of the time series (the Japanese trawl fishery) is certainly a concern. Likewise, the use of the same natural mortality value for both sexes, when there seems to be a difference in growth and maturity, is also a concern. However, using sex-specific natural mortality estimates did not seem to improve the fit very much, if at all.

The stock recruitment relation based on Ricker can also be concerning. As shown below, there is a tendency for increased density dependent factors as you move to higher and higher stock sizes. This can be troubling especially if the relationship is not well informed at smaller to moderate stock sizes. Given the biology of the stock however, it seems biologically plausible for such a strongly density dependency to negatively affect recruitment.



In the preferred stock recruitment model, the data used are from 1978-2012; omitting 1954-1977. Given the high productivity of the stock at that time, this use of the time series since 1978 seems more appropriate given the prevailing environmental conditions. As such, use of the preferred model of recruitment is recommended.

Conclusions

Overall, the model "New Base" performs well. It has decent fits and produces reliable results. The use of both time and start date in the catchability is a real strength of this formulation. However, the retrospective pattern is troubling, and points to other forces in the model which might be mis-specified; a serious weakness. Despite this, it represents the best information on stock status relative to reference points.

<u>TOR 2</u>

1. *Evaluate the stock assessment approach used focusing specifically on how fisheries and survey data are compiled and used to assess the stock status relative to stated management objectives under the Bering Sea and Aleutian Islands Fishery Management Plan (FMP) and the Magnuson-Stevens Act requirements. Elements should consider:*
    a. *The FMP "Tier" designation*
    b. *Fishing rate estimation relative to overfishing definitions*
    c. *Stock status determinations relative to B_{MSY}*

For Yellowfin Sole, the information from the bottom trawl survey are incorporated into a statistical model utilizing a maximum likelihood, and catch-at-age approach to generate estimates of population trajectory through time, as well as fishing mortality.

Yellowfin Sole are currently categized in Tier 1A which is defined as:

Tier 1: Reliable point estimates of B and BMSY and pdf of FMSY .

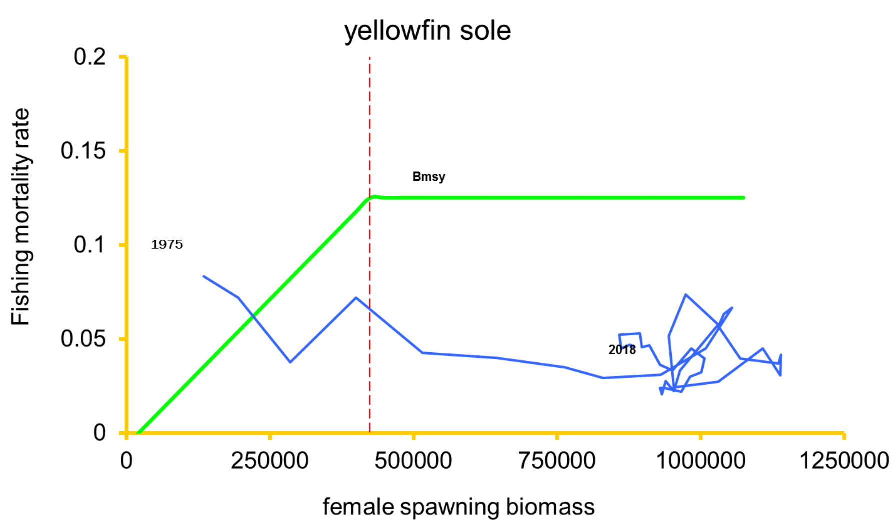
1a) Stock status: B/BMSY > 1 FOFL = mA , the arithmetic mean of the pdf

FABC < mH , the harmonic mean of the pdf

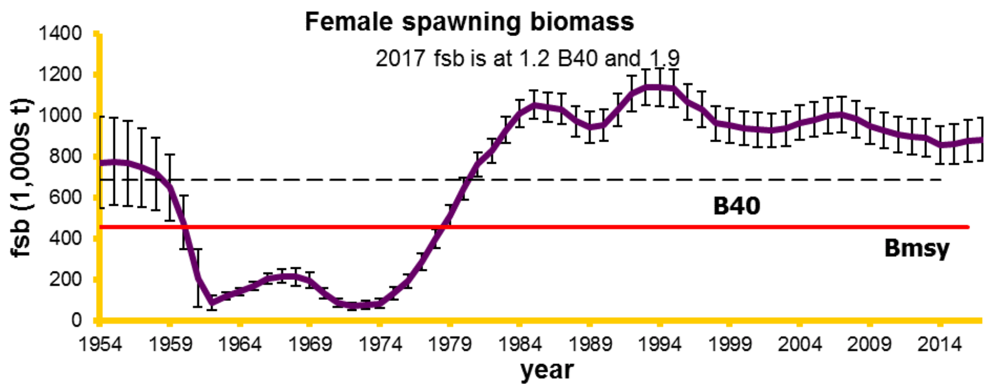
For Yellowfin Sole, as a Tier 1a stock, the harvest is calculated as the product of the harmonic mean of FMSY and the geometric mean (median) of the 2018 biomass. This gives a harvest at FOFL = FMSY 0.12 at 306,700 mt and FABC = 0.109 or 277,500. Note these values are likely to be different than the final values produced by the "New" base run as described above.

Given the amount of info and ability to categorize the uncertainty, Tier 1a seems appropriate for Yellowfin Sole.

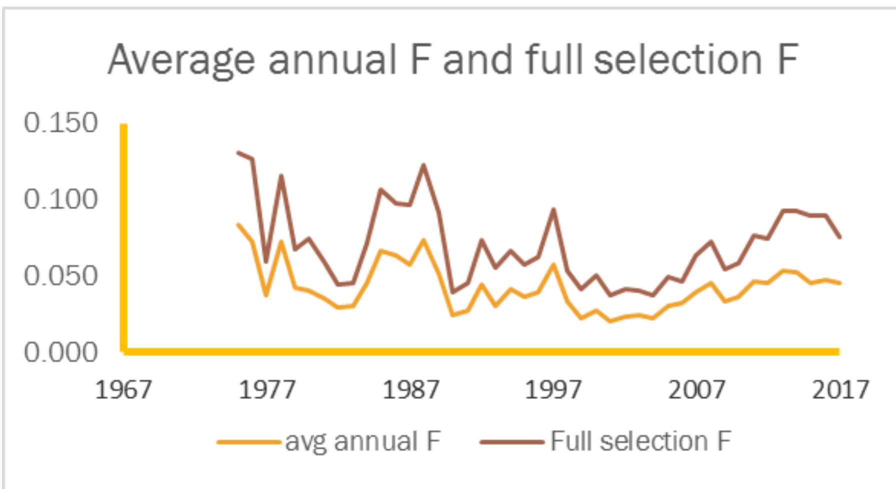
As noted in the stock assessment summary, and in the figure below, the stock is not overfished, and overfishing is not occurring. Fishing mortality is well below F_{MSY} while the SSB is well above SSB_{MSY}. It is reminded that the recent retrospective pattern suggests that the current estimate of SSB is likely an underestimate, while the fishing mortality is likely an overestimate.



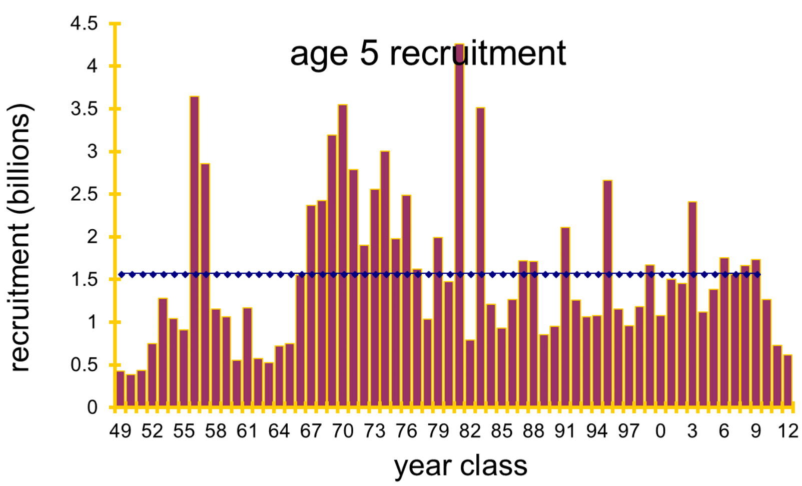
It should be noted that the stock has declined slowly since a recent high SSB in 1994. An increase is seen in the most recent year of the assesment, however.



Fishing mortality has recently declined from a relatively high plateau in the 2014-2016 timeframe. But F is well below F_{MSY}.



Overall, recruitment has been down sharply in the last few years (below) further eroding the SSB to current levels.



TOR 3

Recommend how assessment data and/or models could be improved.

At the Review Workshop, the panel asked to see a number of different alternate runs. Additionally, and given the length of time, some of these longer-term recommendations could not be achieved during the workshop. These included:

- size-based selectivity and potential interactions with growth changes
- Examine plus-group
- Retrospective patterns with full model for survey catchability
- Age-specific natural mortality (e.g., Lorenzen) might be considered (but since fishery ages mostly older, may not matter so much)

It is noted that the assessment team recommend genetic studies to examine isolation, as well as maturity studies in the northern Bering Sea for comparison with recent SE Bering Sea shelf samples.

Northern Rock Sole (NRS)

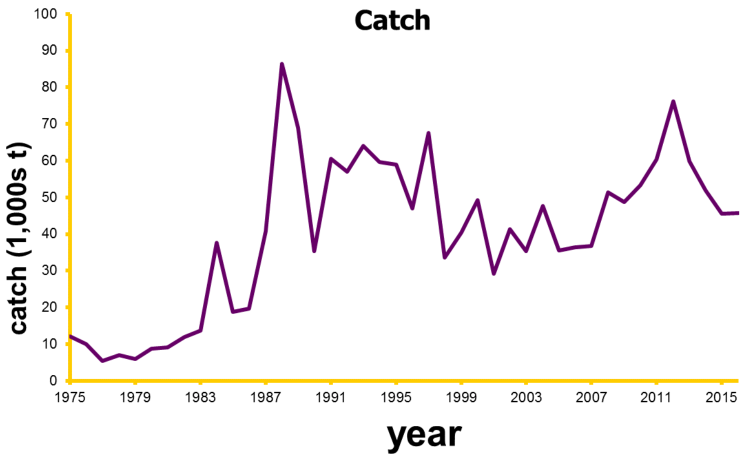
Assessment summary

At the time the materials were available to the peer review panel, as shown in the statement of work, the model for Northern Rock Sole (NRS) was from 2016 (https://www.afsc.noaa.gov/REFM/Docs/2016/BSAIrocksole.pdf).

The 2017 report summary was made available after or near the close of the Review Workshop. During the review, the presentation used the 2016 assessment. As such, comments in this section refer to the 2016 assessment for NRS as presented at the meeting and available prior to the Review Workshop. Because the analysis is virtually the same, differences between the 2016 assessment and report and the 2017 assessment and report are minimal. For completeness, the 2017 summary of the NRS assessment is appended to this report.

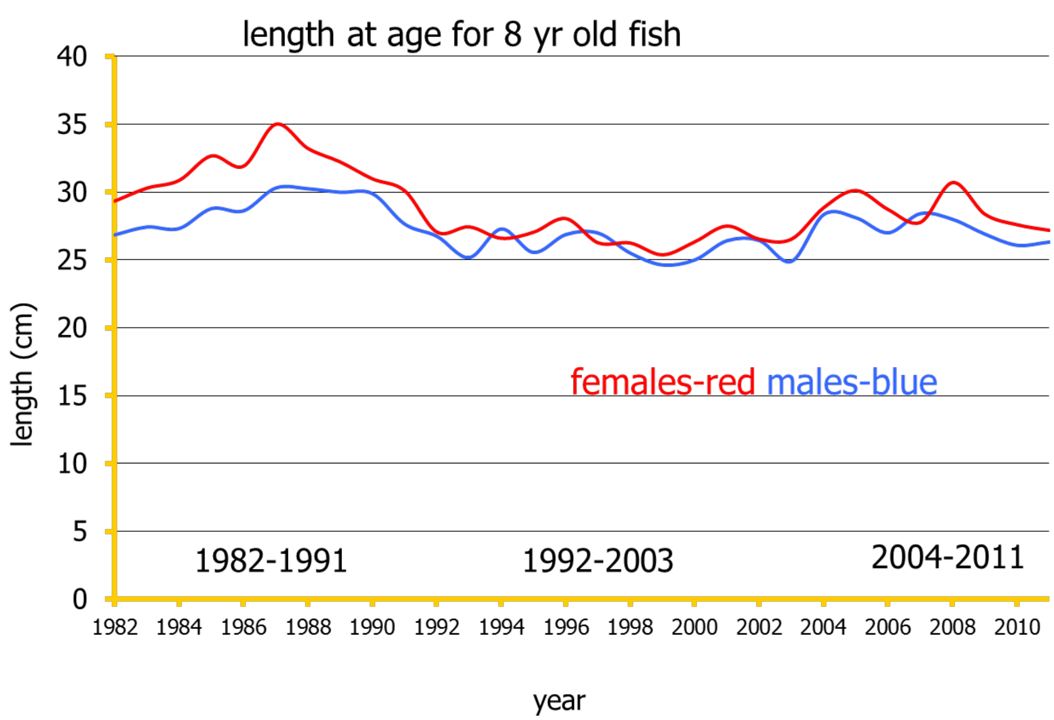
The model used for Northern Rock Sole (NRS) is a separable catch-at-age approach using a maximum likelihood method and implemented in AD model builder. It is similar to other modeling approaches used for other flat fish in the area, as stated in the Yellowfin section (above).

Catch has been variable in this fishery (below). It peaked recently at nearly 77,000 mt in 2012, but has declined to ~45,000 mt in 2016. In 1999, NRS were only 40% retained by the mixed flatfish fishery. This number increased steadily until it achieved an >80% retention by 2008. During February and March, NRS are an important roe fishery, unlike other flatfish in the area. At other times of the year, NRS are less marketable but are still an important part of the overall flatfish removals.

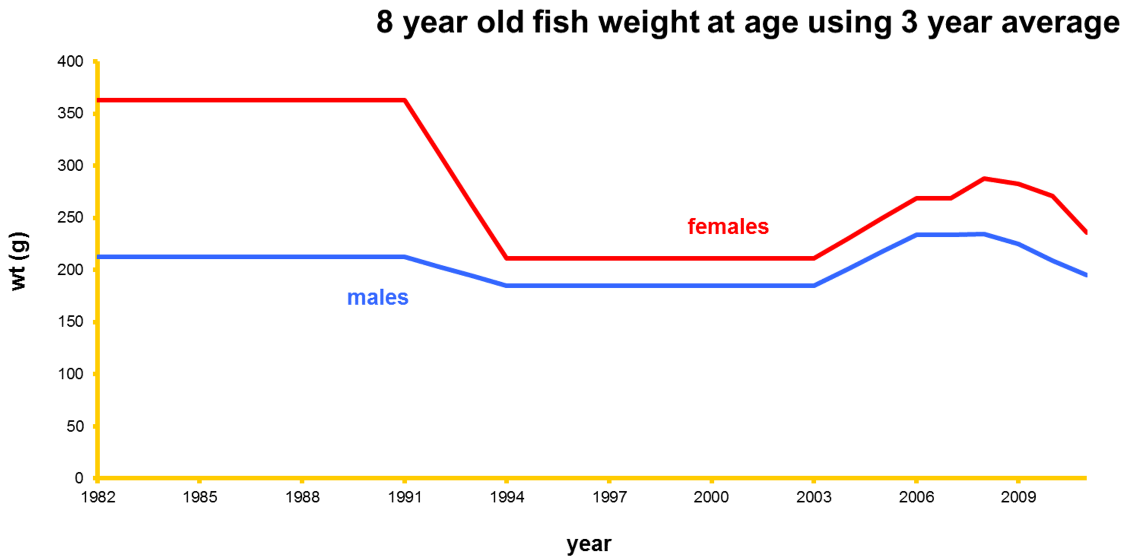


Other sources of data in the model included a bottom trawl survey, with estimates of catchability to scale to a swept areas biomass time series. Additionally, both the survey and the fishery have both catch, catch at length, and catch at age information available.

For weight at age, sex specific weight at age information is available from at-sea observer data across the time series. Results indicate changes over time as shown below:



As with the 2012 assessment, this assessment used a similar averaging approach across multiple years allowing for a ramping between early and mid-years, and then a specification of weight at age since 2003 as shown below.



Maturity was based on Stark (2012) and is unchanged from the previous assessment. NRS appears to be 50% mature at age 8.

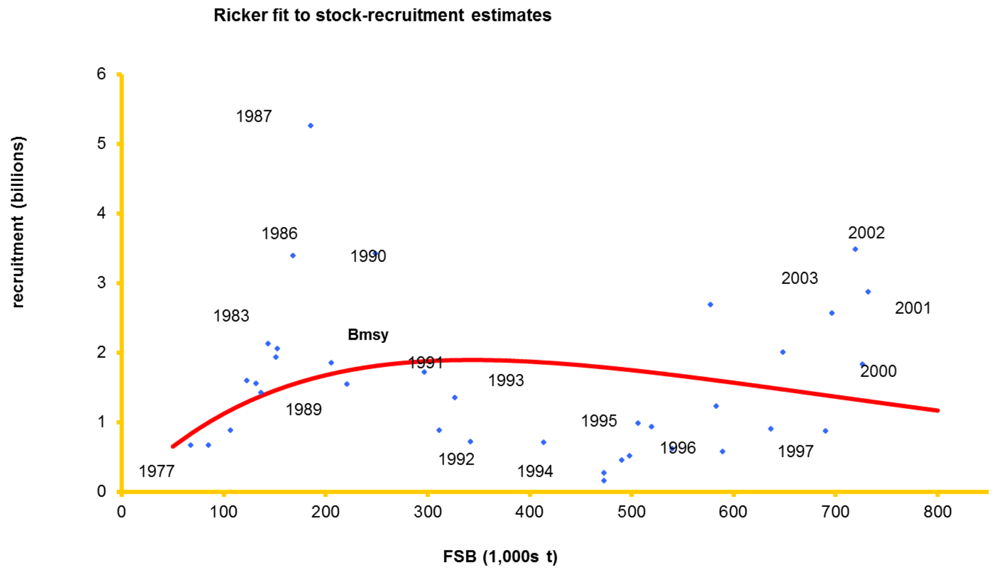
In the base run catchability was fixed at 1.5, similar to a herding experiment done by Somerton and Munro (2001). As outlined below, some alternate runs using estimated catchability were also made. Of particular note, catchability in some alternate runs was estimated in a similar way to Yellowfin Sole; using temperature as a covariate. A second method of estimation was to leave it as a free parameter with a penalty term to keep in close to the 1.4 from Somerton and Munro 2001:

$$qprior = 0.5\left[\frac{q_{\mathrm{exp}} - q_{\mathrm{mod}}}{\sigma_{\mathrm{exp}}}\right]^2$$

Natural mortality in the model was derived by fixing catchability and then estimating sex specific natural mortality. The results suggested 0.159 and 0.19, for males and females. In the final base run, natural mortality was fixed at 0.15 for both sexes, and catchability was fixed also at 1.5. Alternate runs estimating natural mortality were also conducted.

Selectivity for both fishery and survey in the base run was modeled separately by sex as two parameters logistic function given the differences in growth between the two sexes. For the survey, only one selectivity block was used, while for the fishery annual estimates of selectivity were applied. An age 20+ group was used.

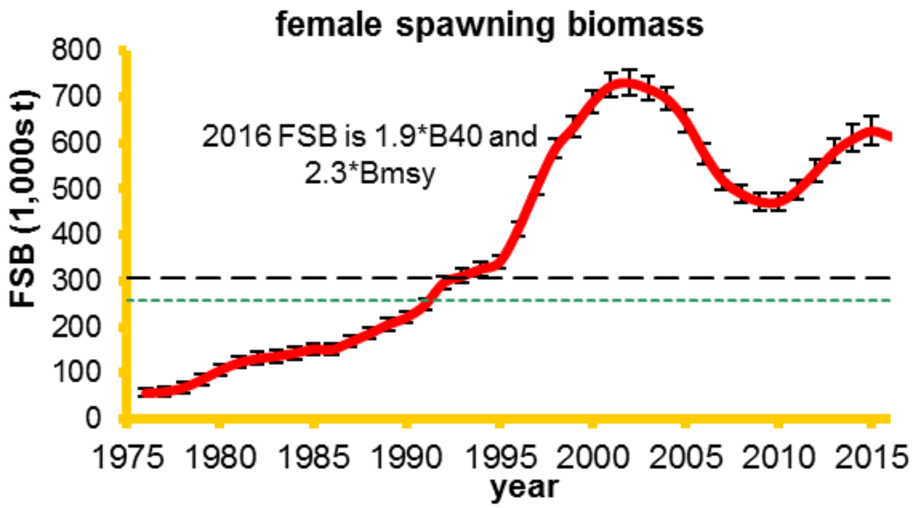
Recruitment was parameterized as a Ricker function with a less than decent fit across all years.



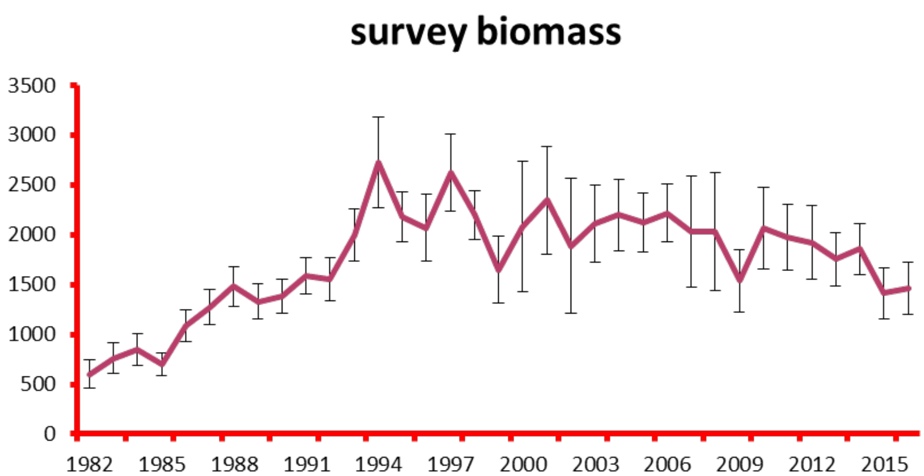
Model results were similar to the formulation used previously as outlined below for the base run.

Quantity	As estimated or specified last year for:		As estimated or recommended this year for:	
	2016	2017	**2017**	2018
M (natural mortality rate)	0.15	0.15	**0.15**	0.15
Tier	1a	1a	**1a**	1a
Projected total (age 6+) biomass (t)	1,085,200	977,200	**1,000,600**	923,200
Female spawning biomass (t)	584,400	522,600	**539,500**	472,200
Projected				
B_0	682,800		**918,500**	
B_{MSY}	257,000	257,000	**257,000**	257,000
F_{OFL}	0.152	0.152	**0.160**	0.160
$maxF_{ABC}$	0.148	0.148	**0.155**	0.155
F_{ABC}	0.148	0.148	**0.155**	0.155
OFL (t)	165,900	149,400	**159,700**	147,300
maxABC (t)	161,000	145,000	**155,100**	143,100
ABC (t)	161,000	145,000	**155,100**	143,100
	As determined last year for:		As determined this year for:	
Status	2014	2015	2015	2016
Overfishing	No	No	No	No
Overfished	No	No	No	No
Approaching overfished	No	No	No	No

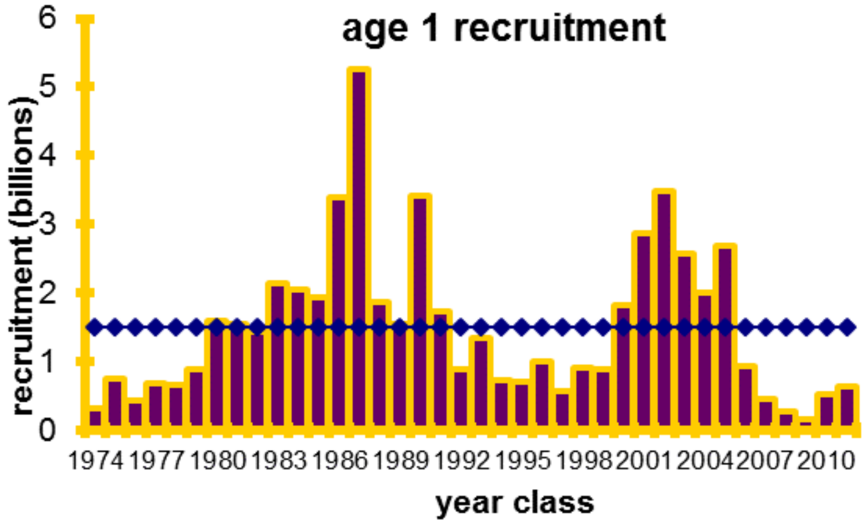
While total biomass has been in decline since 2007, SSB has been variable but recently increasing as show below.



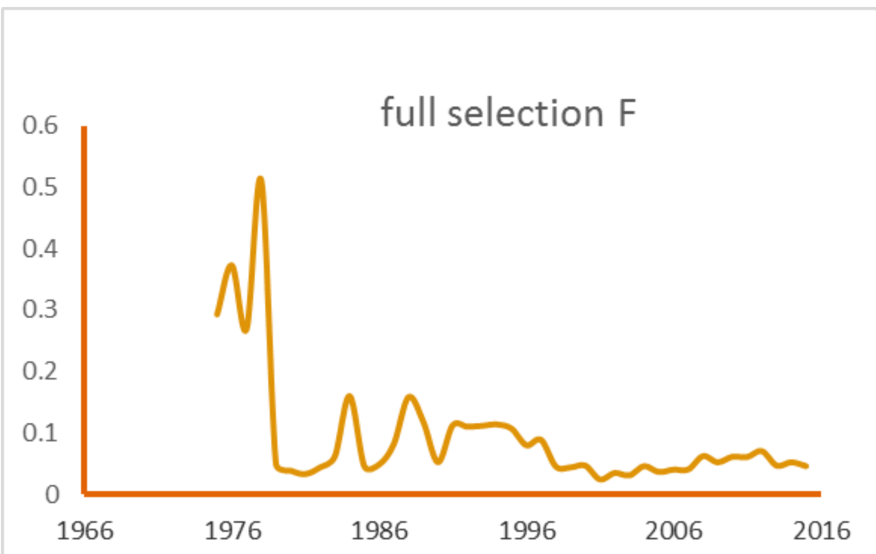
This contrasts with the survey which has been showing a continued decline to near historically low levels. With 2015 the lowest survey biomass since 1990.



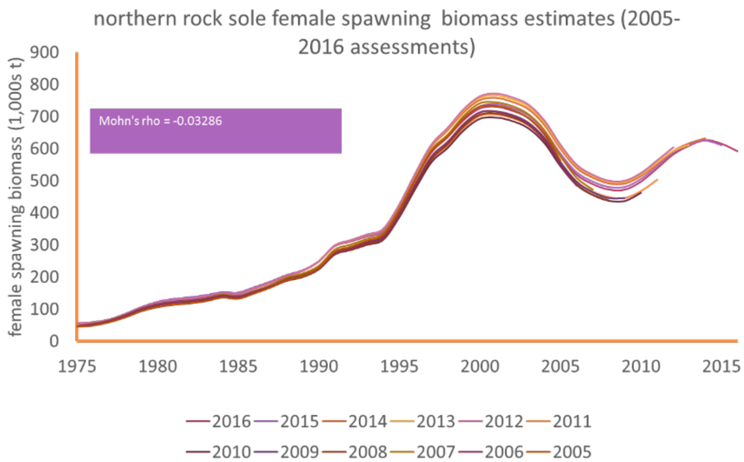
Like with Yellowfin, recruitment has recently been below average.

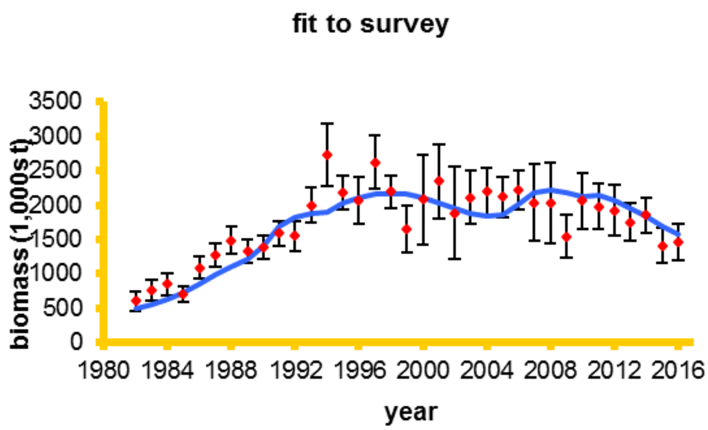


The low recruitment undoubtedly has contributed to the decline as fishing mortality remains quite low; well below F_{MSY}.



Most of the diagnostics for the base run look good. There is little retrospective pattern, unlike with Yellowfin, and the fits to the survey is good to moderate.





Alternate runs

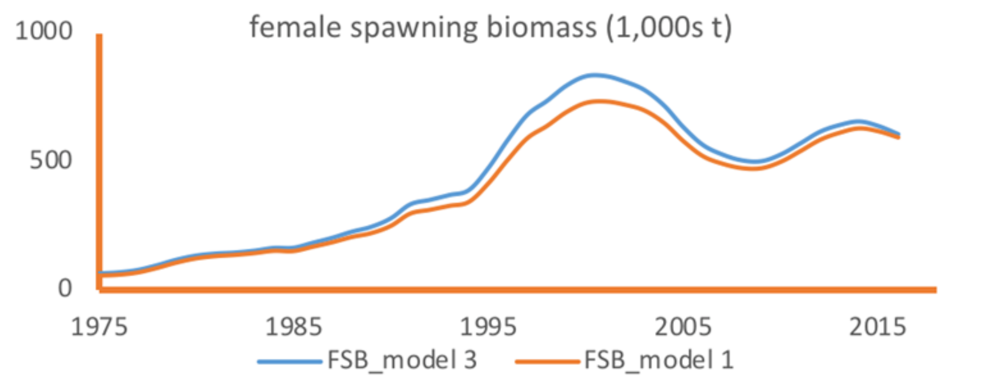
In addition to the base run as outlined above, the analysts also ran several alternate runs ahead of the Review Workshop as outlined below.

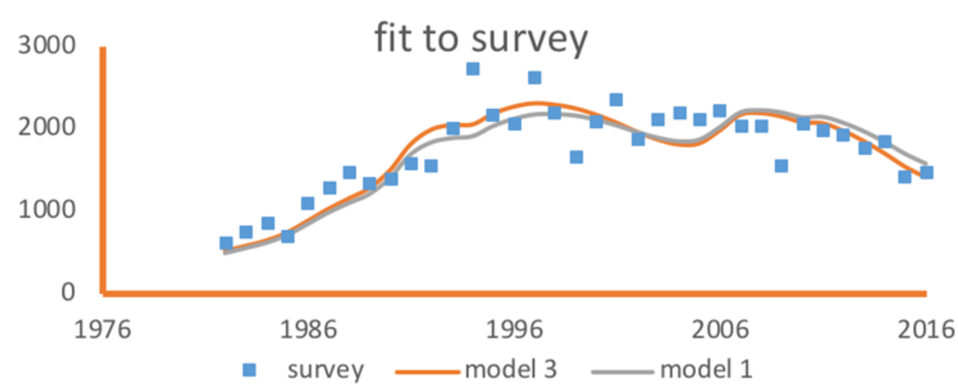
Model exploration	Q	female M	male M	2017 FSB	2017 ABC	FABC
Model 15.1 q fixed at 1.5, male and female M fixed at 0.15	1.5	0.15	0.15	539,500	155,100	0.155
Model 16.2 q fixed at 1.5, female M fixed at 0.15 and male M estimated	1.5	0.15	0.18	594,500	153,000	0.159
Model 16.3 q fixed at 1.5, female M and male M estimated	1.5	0.164	0.192	542,900	141,000	0.157
Model 16.4 q estimated, Female and male M fixed at 0.15	2.22	0.15	0.15	320,300	101,200	0.168
Model 16.5 q estimated, female M fixed at 0.15 and male M estimated	1.98	0.15	0.177	409,500	112,700	0.166
Model 16.6 q, female M and male M all estimated as free parameters	2.17	0.14	0.17	378,800	106,400	0.17
Model 16.7 q fixed at 1.5 but allowed to vary annually with bottom temperature relationship, male and female M fixed at 0.15	1.5	0.15	0.15	540,000	155,300	0.155

Because the analysts presented numerus alternate runs, the Review Panel did not request any additional runs. However, it was agreed that a full work-up of the results of Model 16.3 would be done to compare it with the base run (15.1). It is recommended for this and others that alternate runs be completed each time the assessment is benchmarked or updated as it provides insight into model formulation, diagnostics, and model behavior, which can change among years based on new information. Model runs 16.2 was excluded as it didn’t seem reasonable to fix only one sex’s natural mortality and not the other. The rest of the runs had catchabilities that were out of line with the herding experiment conducted by Somerton and Munro (2001). The last run (16.7) was excluded as there seemed little evidence that temperature plays a role in catchability (see below).

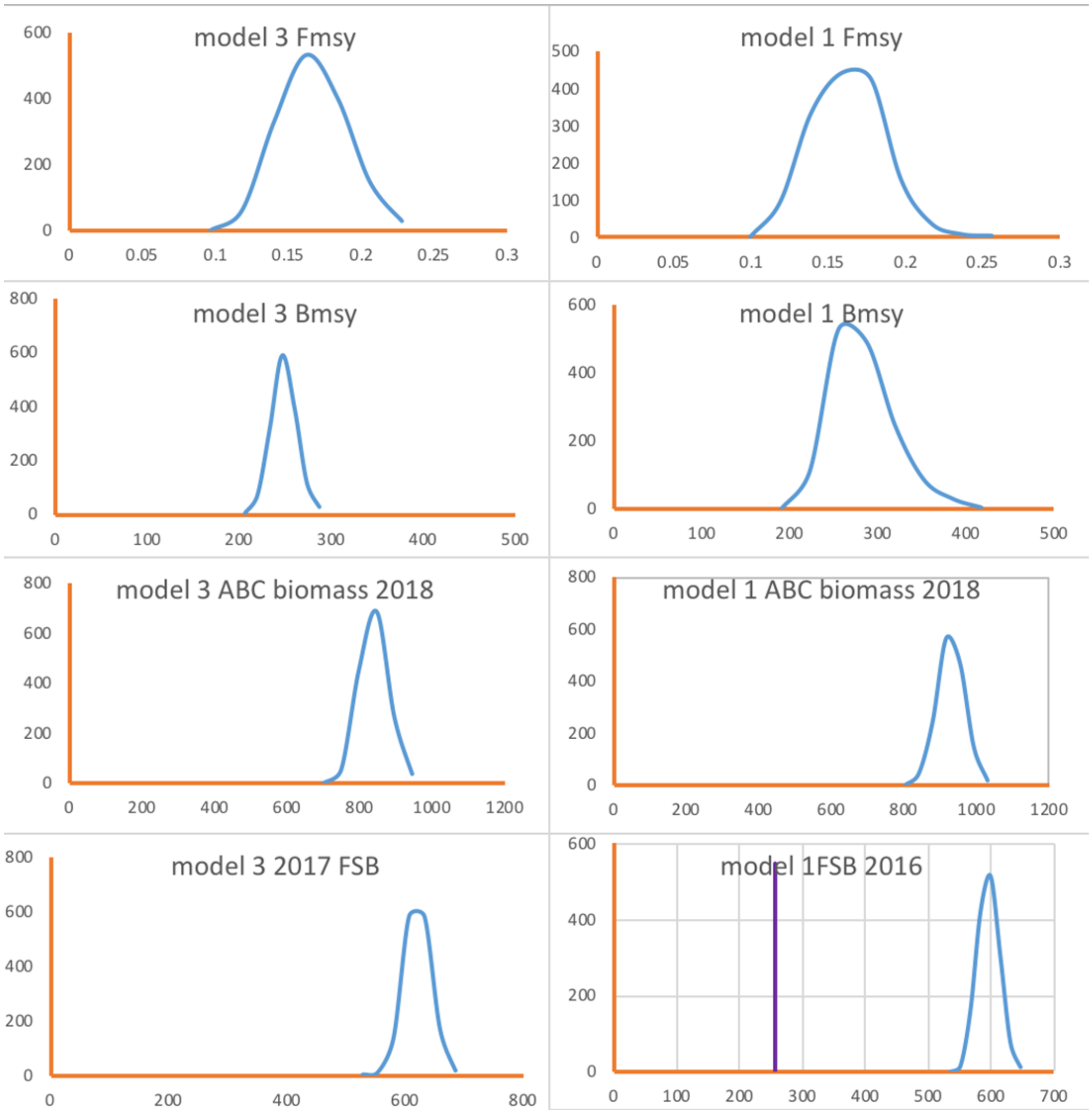
Model 16.3 differed from the base 15.1 as it allowed the natural mortality of NRS to be estimated separately for the sexes. It is often important to not externally fix both natural mortality and catchability, to allow the model for find a solution which is more robust internally. Given that the catchabilities were determined from a herding study, allowing natural mortality to vary by sex seemed to be the most logical choice.

Despite moderate differences in the natural mortality, the trends were very similar when examined over time. Female SSB was higher in the past for model 16.3 when compared to model 15.1, but converged closer to the present time, giving a slightly different picture of the terminal year compared with the recent past. Fits seem mostly similar, but look marginally improved.





As shown below, model 16.3 showed a tighter, more symmetrical distribution in its estimate of F_{MSY}, a much more peaked distribution for B_{MSY} and a slightly different, if lower estimate of 2017 ABC. Both formations resulted in the same stock status. As a result of this work, model 16.3 appears more robust, both in its results as well as its diagnostics. As such, model 16.3 should be the preferred model.



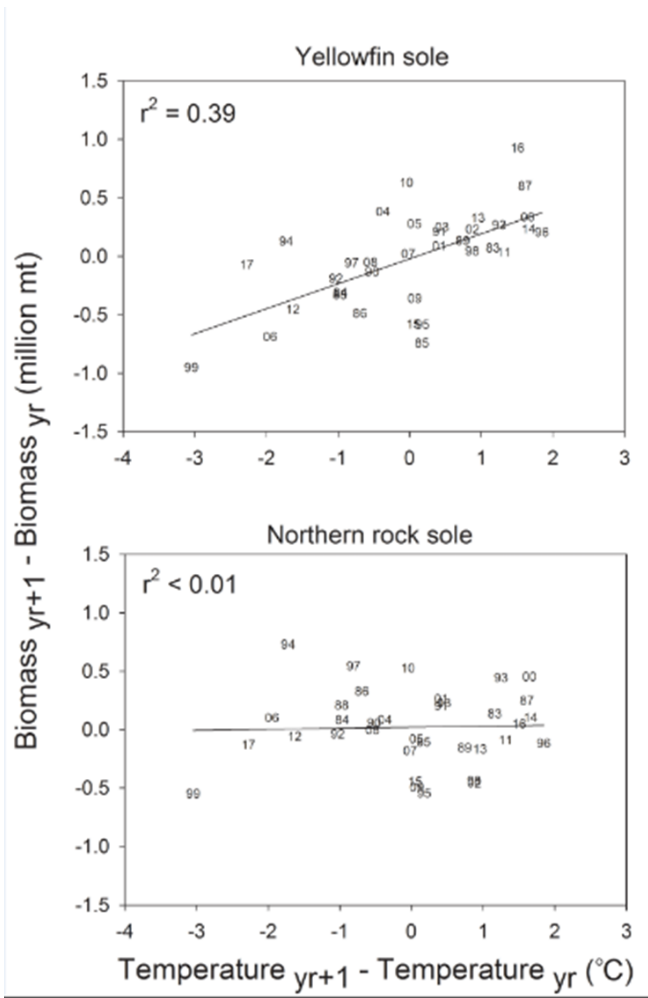
It should be noted, however, that fits for the survey sex ratios are somewhat worse for model 16.3 compared with the others.

<u>TOR 1</u>

1. *Evaluate the strengths and weaknesses of the assumptions made in applying the stock assessment model including how survey indices are scaled to the populations. Specifics might include:*
    a. *How natural mortality estimates are estimated/applied*
    b. *Assumptions about survey "catchability"*
    c. *Application of fishery and survey age-specific schedules (maturity, body mass, selectivity)*
    d. *The application (or lack thereof) of a stock-recruitment relationship (and associated parameter estimates)*

As outlined above, while the natural mortality was fixed for both sexes at 0.15 in the base (15.1) model, allowing these to be estimated appears to be the best option. The differences seen in the growth rates between the sexes seemingly confirms that natural mortality should be different.

Catchability, because natural mortality is estimated, is best used as a fixed value. Unlike Yellowfin Sole, analysis presented at the Review Workshop (below) doesn't indicate the same variation in catchability with changes in temperature.



As such, leaving it as a fixed value using the results of Somerton and Munro (2001) is likely best for this assessment. It should be noted, however, that this should be reexamined when the stock assessment is updated or benchmarked, as a relationship with temperature could develop over time.

The stock recruitment relationship that is based on Ricker can also be concerning. As with Yellowfin, there is a tendency for increased density dependent factors as you move to higher and higher stock sizes. This can be troubling especially if the relationship is not well-informed at smaller to moderate stock sizes. While the use of the Ricker relationship across the time series seems appropriate, a similar pattern of high productivity at smaller stock sizes during the earlier part of the time series is evident. As such, restricting the stock recruitment relationship to a smaller and more recent period, while not warranted currently, could be useful in the near future.

Selectivity was estimated as one block for each of the sexes by age in the survey, and yearly by sex in the fishery. Like with Yellowfin, this seems plausible given the unlikelihood of a change in selectivity by age in the survey by year using the same gear from year to year.

Maturity and weight at age are discussed above. Both seemed appropriate, and where possible, based on actual collected data from the fishery, rather than modeled. This is an important strength in this assessment.

Conclusions

Overall, the strength of this approach is in the model diagnostics and the full suite of alternate runs provided. Not having a retrospect pattern as well as the good fits is also a plus. The weaknesses relate mainly to the base run. It is important that either sex specific natural mortality or catchability be allowed to vary within the model, at least as a method of diagnosing potential problems or states of nature. As an amendment to the model 16.3 functions well and is the best scientific information available.

<u>TOR 2</u>

1. *Evaluate the stock assessment approach used focusing specifically on how fisheries and survey data are compiled and used to assess the stock status relative to stated management objectives under the Bering Sea and Aleutian Islands Fishery Management Plan (FMP) and the Magnuson-Stevens Act requirements. Elements should consider:*
    a. *The FMP "Tier" designation*
    b. *Fishing rate estimation relative to overfishing definitions*
    c. *Stock status determinations relative to B_{MSY}*

Like with Yellowfin Sole, the information for NRS from the bottom trawl survey are incorporated into a statistical model utilizing a maximum likelihood and catch-at-age approach to generate estimates of population trajectory through time, as well as fishing mortality.

NRS are currently categorized in Tier 1A, which is defined as:

Tier 1: Reliable point estimates of B and B_{MSY} and pdf of F_{MSY} .

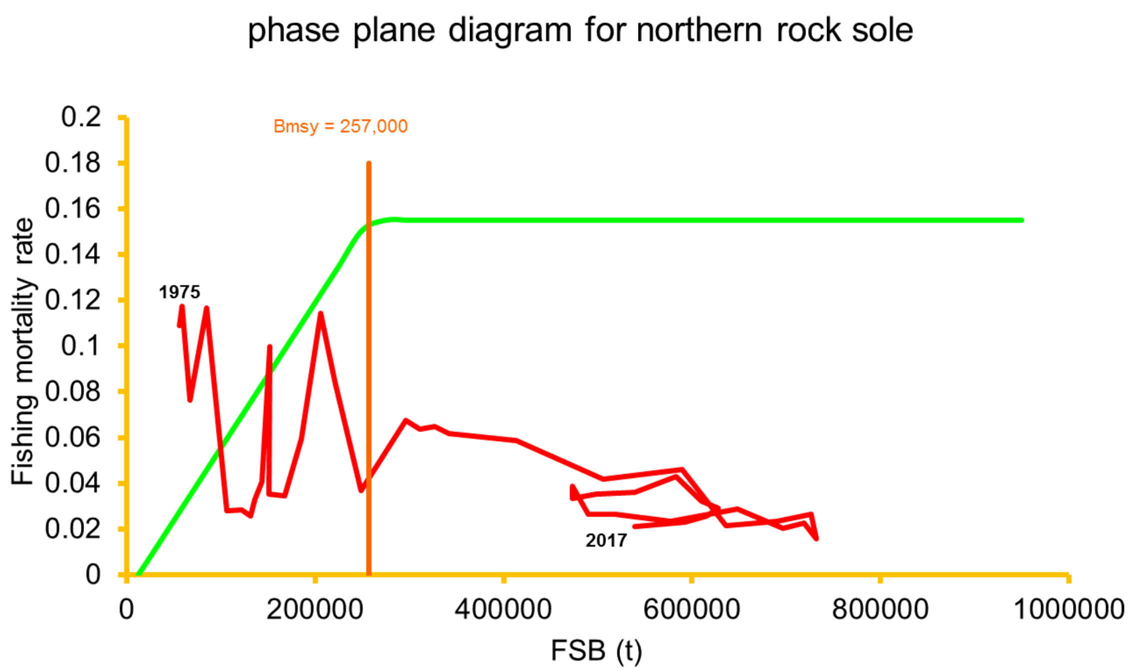
1a) Stock status: $B/B_{MSY} > 1$ F_{OFL} = mA , the arithmetic mean of the pdf

$F_{ABC} < mH$, the harmonic mean of the pdf

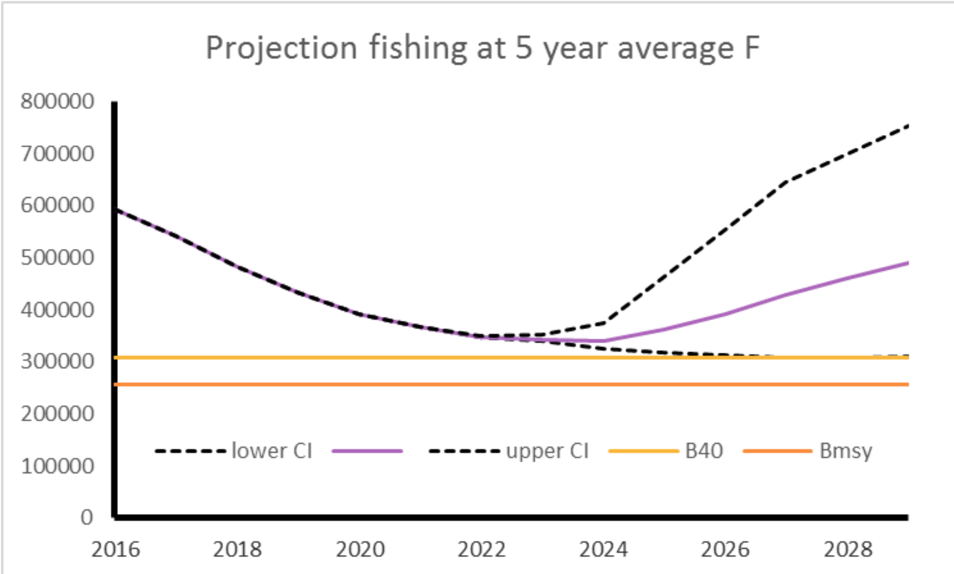
For NRS, as a Tier 1a stock, the harvest is calculated as the product of the harmonic mean of F_{MSY} and the geometric mean (median) of the 2018 biomass. This gives a harvest at FOFL = F_{MSY} 0.16 and an SSB_{MSY} 257,000 mt F_{ABC} is 0.155, giving an OFL of 159,700 mt and an ABC of 155,100 mt. Note these values are likely to be different than the final values produced by the 16.3 model run described above. For example, for model 16.3 the FABC was estimated at 0.157.

Given the amount of info and ability to categorize the uncertainty, Tier 1a seems appropriate for this stock.

Overall, whether using the base 15.1 or the recommended 16.3 model run the results are virtually the same in the terminal year, as shown below for the base 15.1 run, the stock is not overfished, and overfishing is not occurring.

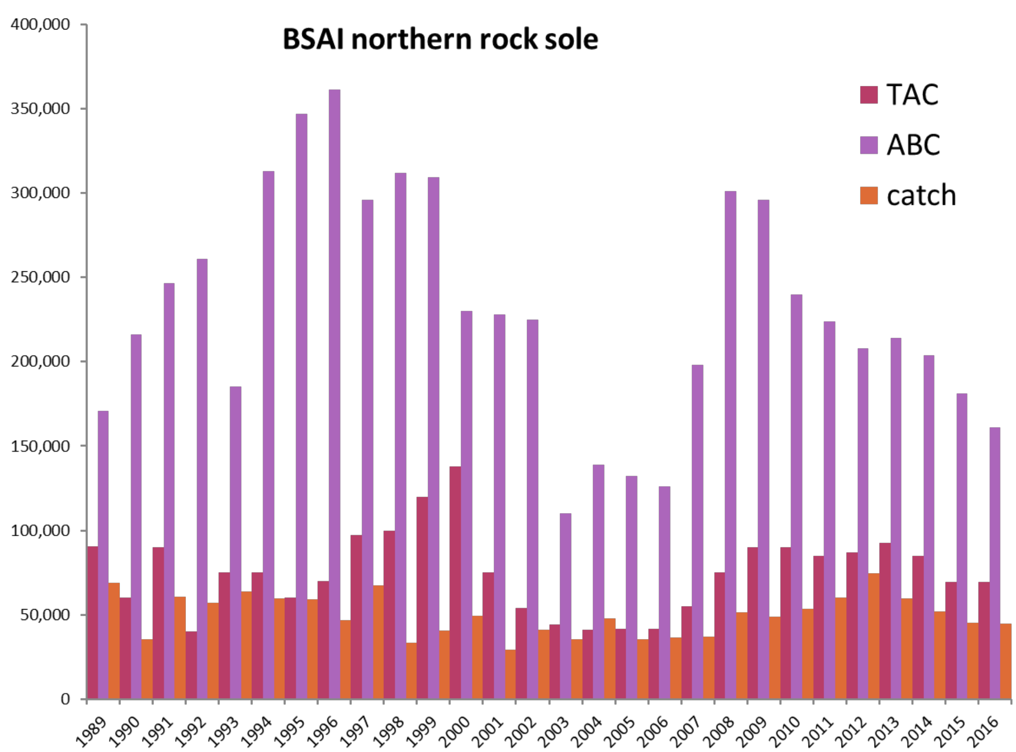


Further short-term projections indicate an increase in stock size.



However, this assumes average recruitment which has not been seen in the last few years. As such, projections are likely optimistic.

It should be noted that this fishery is not close to achieving its ABC on any given year. Removals have been far lower, in fact, than the TAC since 2004.



Despite this, total biomass has declined for this stock. As such, a note of caution should be recognized. This stock has not caught near its estimated ABC, and not achieved its TAC, yet stock size is declining due in part to a reduction in the stock's recruitment.

TOR 3

Recommend how assessment data and/or models could be improved.

During the Review Workshop, recommendations for improvement other than moving to Model 16.3 were not made. However, continued use of alternate runs for this stock should be an important part of the process. Exploration of alternate stable states of nature though alternate runs improves assessments over time; and allows them to adapt to changes in the stock as well as environmental factors.

In general, the stock recruitment relationship used for both the base and model 16.3 is appropriate. In the future, exanimation of restricting the recruitment to only the most recent time frame, like Yellowfin Sole, should be explored for both model outputs as well as reference points.

There were no other recommendations made in the documentation presented by the assessment team.

Alaska Plaice

Model Summary

The model used for Alaska Plaice is a separable catch-at-age model using a maximum likelihood approach and implemented in AD model builder. It is like other modeling approaches used for flat fish in the area and in this report. It used data as outlined below from the catch, survey and other sources of information.

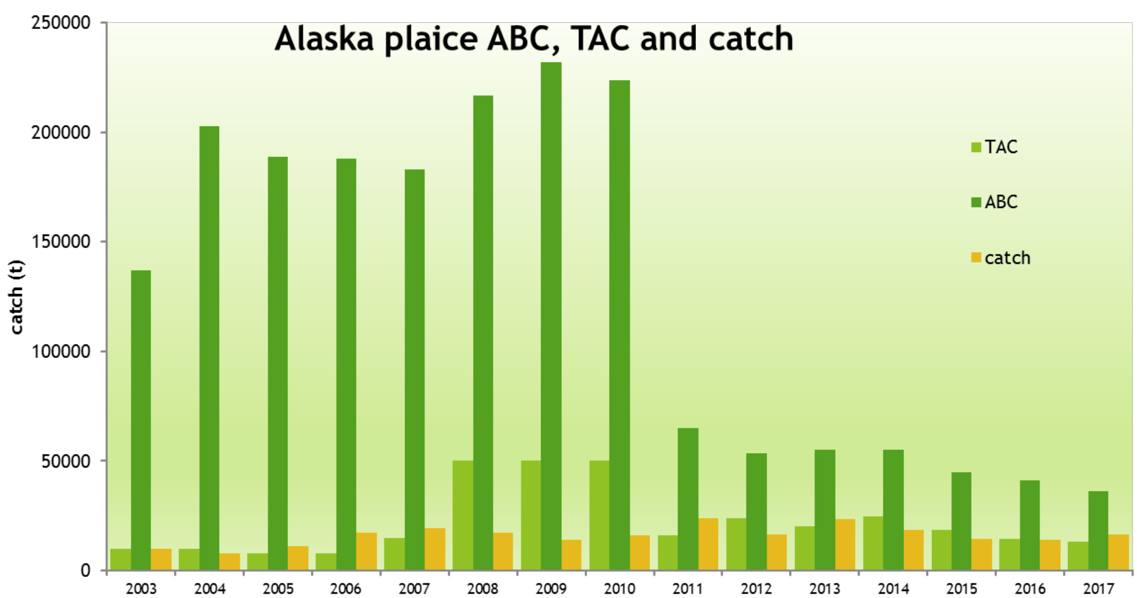
Source	**Data**	**Years**
NMFS Eastern Bering Sea shelf survey	Survey biomass and standard error	1982-2017
	Age Composition (by sex)	1982, 1988, 1992-1995, 1998, 2000-2002, 2005-2014, 2016
	Length Composition (by sex)	1983-1987, 1989-1991, 1996-1997, 1999, 2003, 2004, 2015 and 2017
Fisheries	Catch	1971-2017
	Age Composition (by sex)	2000, 2002 and 2003
	Length Composition (by sex)	1978-89, 1995, 2001 and 2008-2016

Similar to NRS, plaice are generally an incidental catch, with retention rates increasing in recent years due to management changes as well as increased marketing.

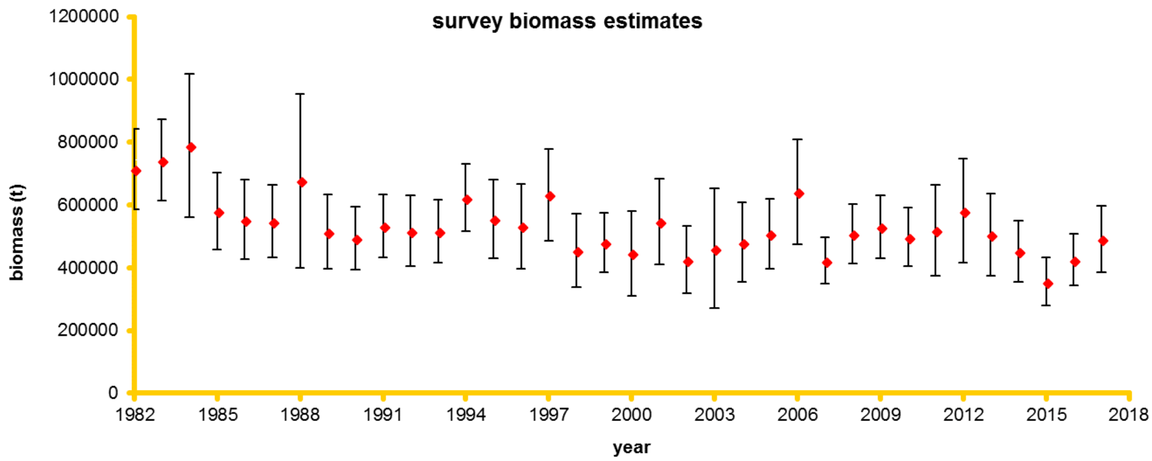
year	Discard	Retained	Total	Proportion discarded
2002	11,806	370	12,176	0.97
2003	9,428	350	9,778	0.96
2004	7,193	379	7,572	0.95
2005	10,293	786	11,079	0.93
2006	14,746	2,564	17,310	0.85
2007	15,481	3,946	19,427	0.8
2008	9,330	8,046	17,376	0.54
2009	5,061	8,882	13,945	0.36
2010	5,845	10,322	16,166	0.36
2011	7,197	16,459	23,656	0.30
2012	3,589	13,023	16,611	0.22
2013	9,053	14,470	23,523	0.38
2014	3,702	15,747	19,449	0.19
2015	1,231	13,382	14,614	0.08
2016	2,070	11,315	13,385	0.15

Removals have recently peaked in 2013, but have declined since.

Overall this fishery does achieve its TAC, but this TAC is set far lower than the ABC.



Fishery independent data in the form of the bottom trawl survey indicate an overall decline in population abundance over the time period, with some indications of increased abundance recently.



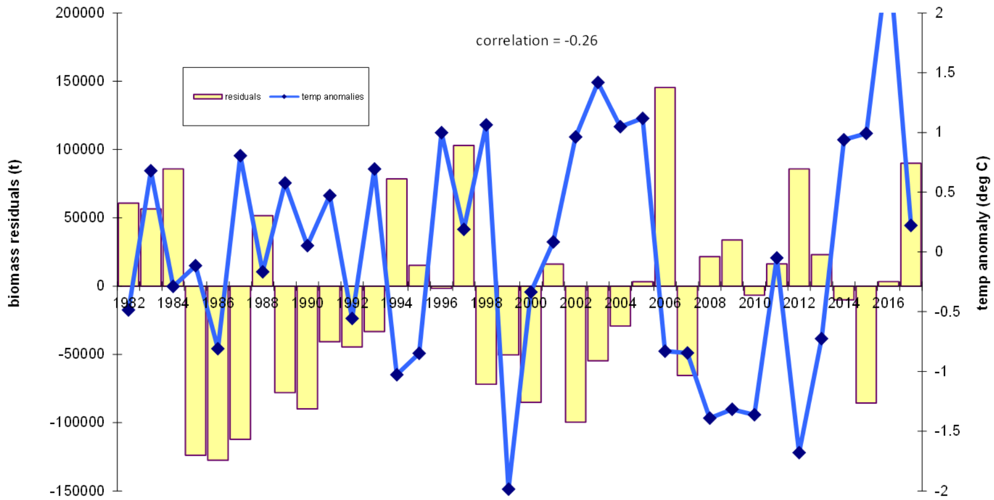
As can be seen in the table above, sex-specific estimates of fishery and survey age composition and length composition, and weight at age are used to allow for a sex-split model to be used for assessing status of plaice.

The 2016 assessment analyzed several different natural mortality estimates for plaice, shown below.

Method	Males	Females
Hoenig (1983)	0.11	0.11
Chapman and Robson (1960)	0.08	0.08
Gislason et al. 2008	0.12	0.29
Model profiling	0.13	0.13

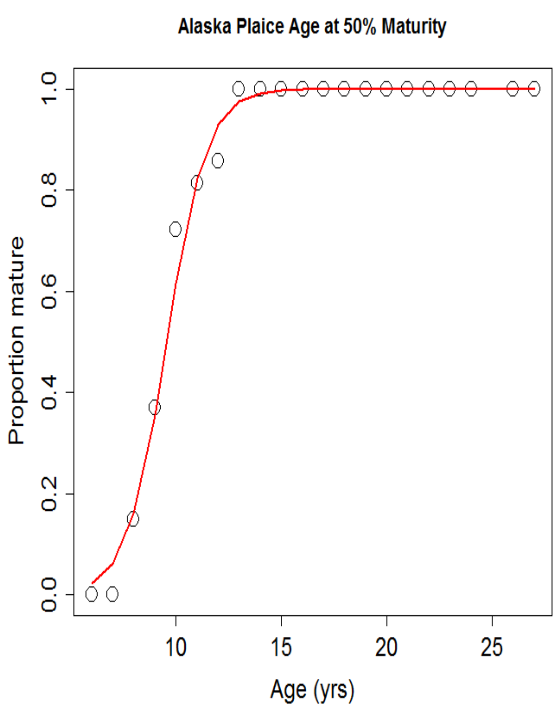
Model likelihood profile indicated that 0.13 was optimal for both sexes, and as such was used in the base model presented.

Unlike yellowfin, plaice seemingly has little correlation between survey catch and temperature as shown below:

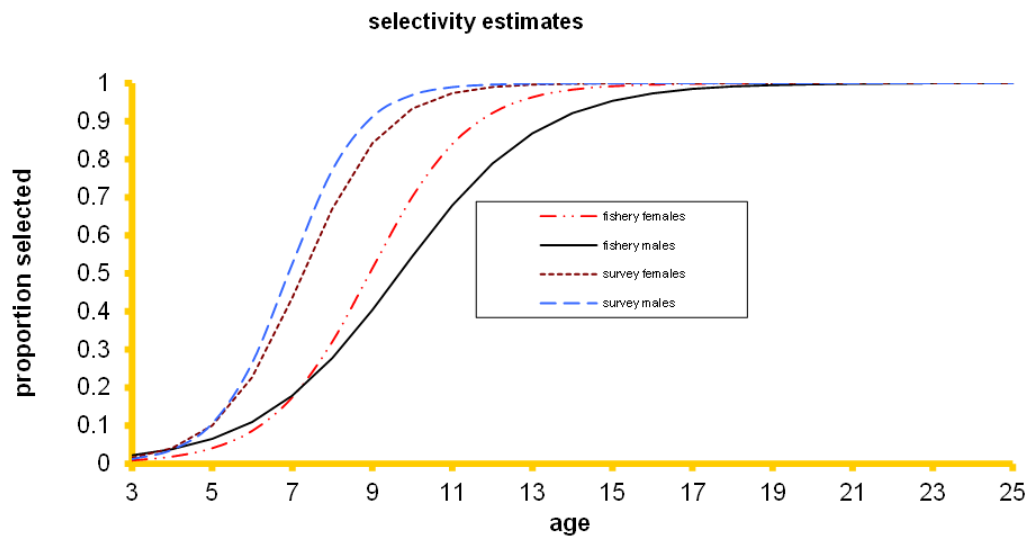


Given the work by Somerton and Munro (2001) as outlined for NRS, a q of 1.2 was used.

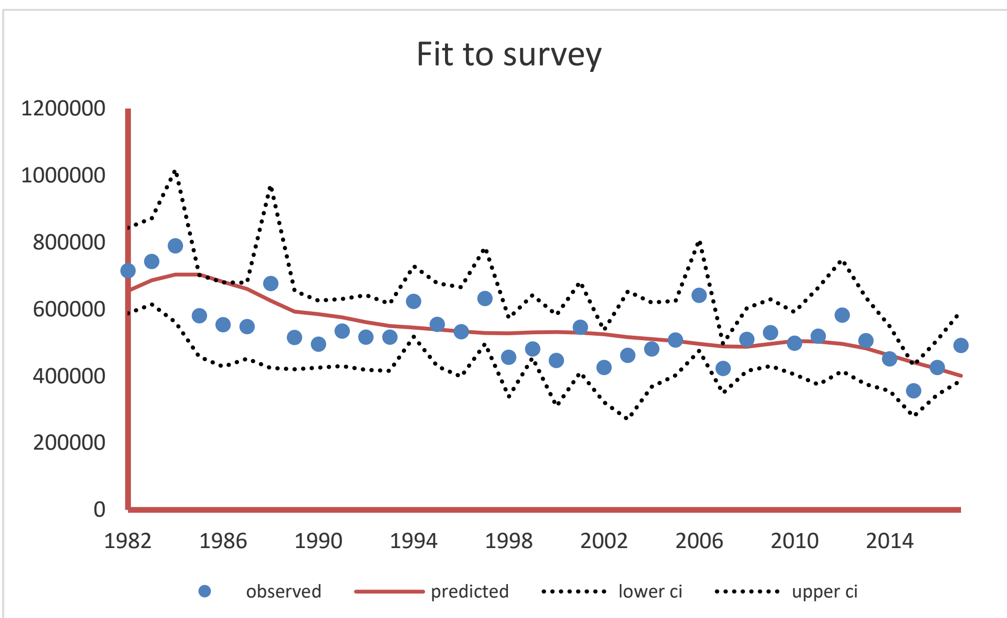
Maturity was also specified outside of the model. Zhang (1987) estimated anatomical estimates, while histological estimates were available from TenBrink and Wilderbuer (2015). Both showed only minor differences, and so the more recent histological work was used as shown below.



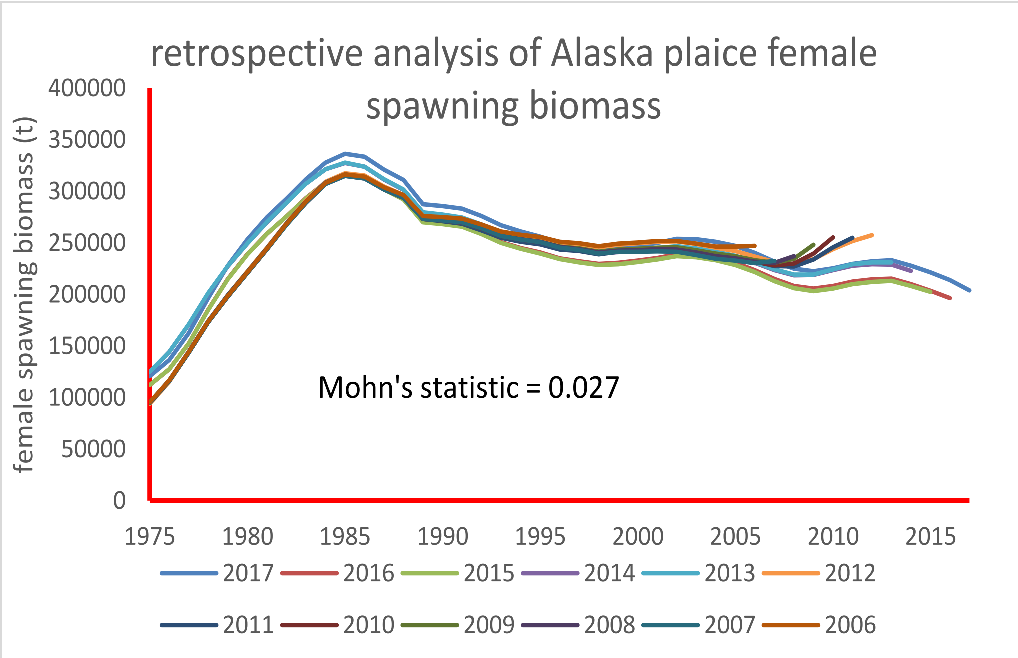
Survey and fishery selectivity was estimated within the model separately for the sexes, by age and by year as shown below.



Model fits to the bottom trawl survey seem appropriate as shown below.



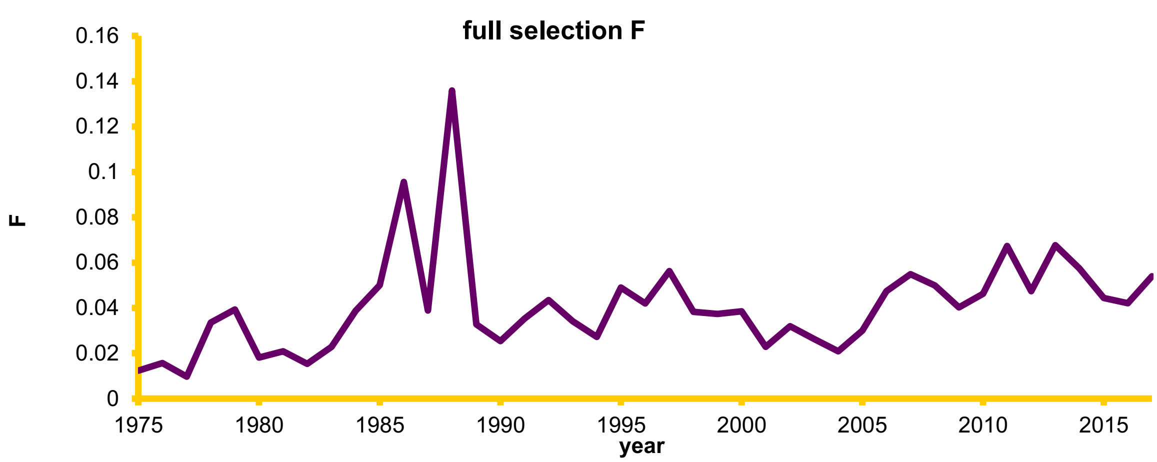
There is only a small retrospective pattern noticed in this assessment.



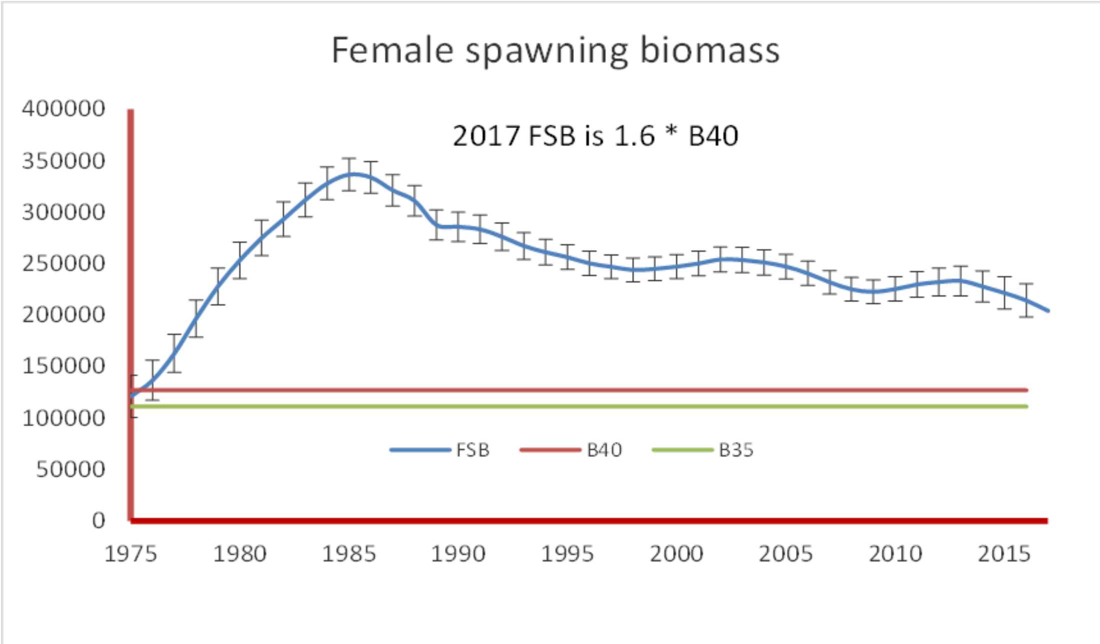
Though it should be recognized that there is some retrospective variability, the pattern seen in the 2009-2012 period seems resolved.

Results for the modeling suggest a similar pattern and results with other flatfish stocks in this review. Because exploitation has been minimal, the stock appears to be above its reference points, and well below possible target fishing mortality.

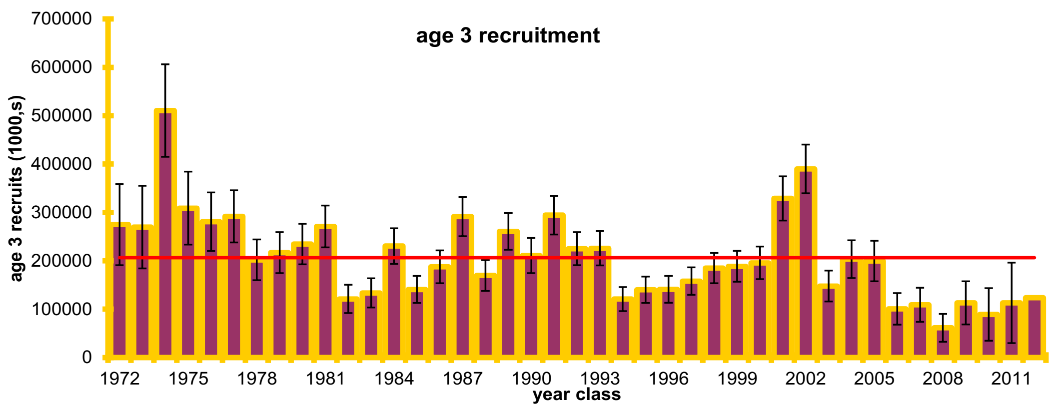
Fishing mortality is lower than the average in the mid-1980's, but increasing since 2005 as shown below.



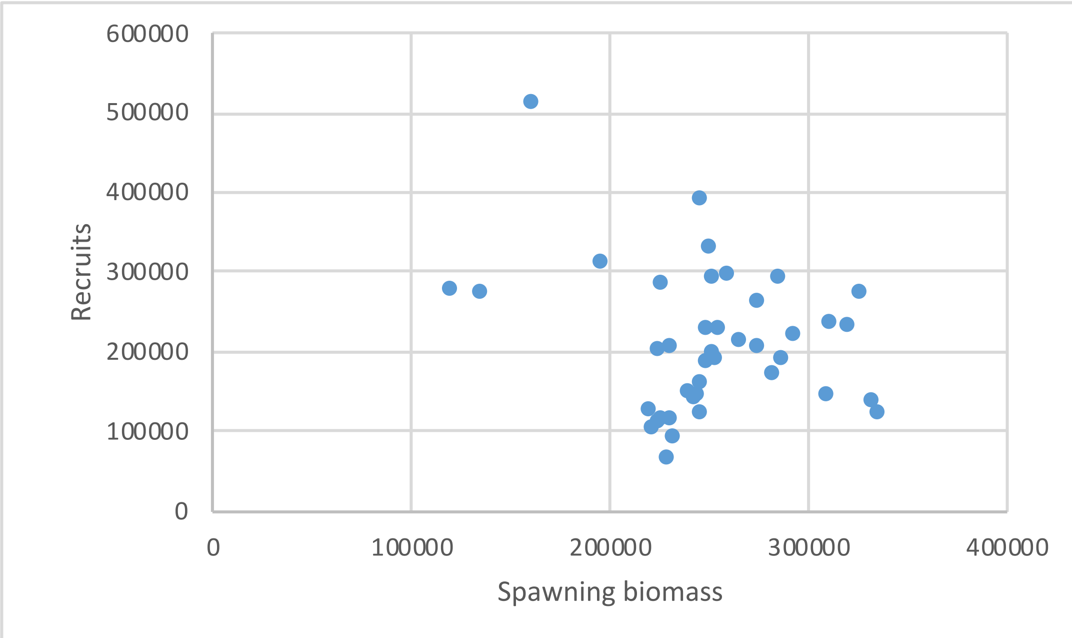
Female SSB has been in slow decline since 1985, but above B40 or B35.



And recruitment has been well below average since 2006, with 2008 being the lowest estimated recruitment on record. This is notable as this weak year class has not fully entered the fishery, being only 70-80% selected in 2018.



Unlike other stock examined in this review, however, plaice has no defined stock-recruitment relationship. Instead recruitment deviations from the median of the time series are used.



This lack has implications for reference points, uncertainty, and ability to provide projections.

Alternate run

In the interest of time, alternate runs were not conducted for this stock. Discussion of possible runs to be completed are given under TOR 3.

TOR 1

1. *Evaluate the strengths and weaknesses of the assumptions made in applying the stock assessment model including how survey indices are scaled to the populations. Specifics might include:*
    a. *How natural mortality estimates are estimated/applied*
    b. *Assumptions about survey "catchability"*
    c. *Application of fishery and survey age-specific schedules (maturity, body mass, selectivity)*
    d. *The application (or lack thereof) of a stock-recruitment relationship (and associated parameter estimates)*

Natural mortality and catchability estimates, which were conducted outside the model, seem appropriately based on either direct study or previous modeling work. However, as discussed under NRS, it is important that some of the parameters not be fixed in future examinations but allow both to be estimated in turn. Like conducted for NRS, a suite of alternate runs fixing natural mortally for either or both sex while allowing catchability to vary, and vice versa, would be ideal.

Selectivity also seems to be interesting. It is noted that there is a large difference in the selectivity at age from the survey vs the fishery. While this is probable, it might be best if the document contained some explanation as to the differences between the two gears that could lead to such differences.

Conclusions

Overall, the model appears to be performing well relative to the diagnostics. While the use of externally set natural mortality and catchability, and the ability to estimate a stock recruitment relationship is an important weakness. The model fits and lack of retrospective pattern is a real strength to the formulation. It is important to recognize, however, that the refence points lie outside the historic stock abundance in terms of biomass. Essentially, this mean that the stock would have to go near its historical low abundance to trigger management response. This doesn't seem to be very tenable. Alternative reference points can and should be explored in future assessments.

TOR 2

1. *Evaluate the stock assessment approach used focusing specifically on how fisheries and survey data are compiled and used to assess the stock status relative to stated management objectives under the Bering Sea and Aleutian Islands Fishery Management Plan (FMP) and the Magnuson-Stevens Act requirements. Elements should consider:*
    a. *The FMP "Tier" designation*
    b. *Fishing rate estimation relative to overfishing definitions*
    c. *Stock status determinations relative to B_{MSY}*

Overall, the results, presented below, seems reasonable.

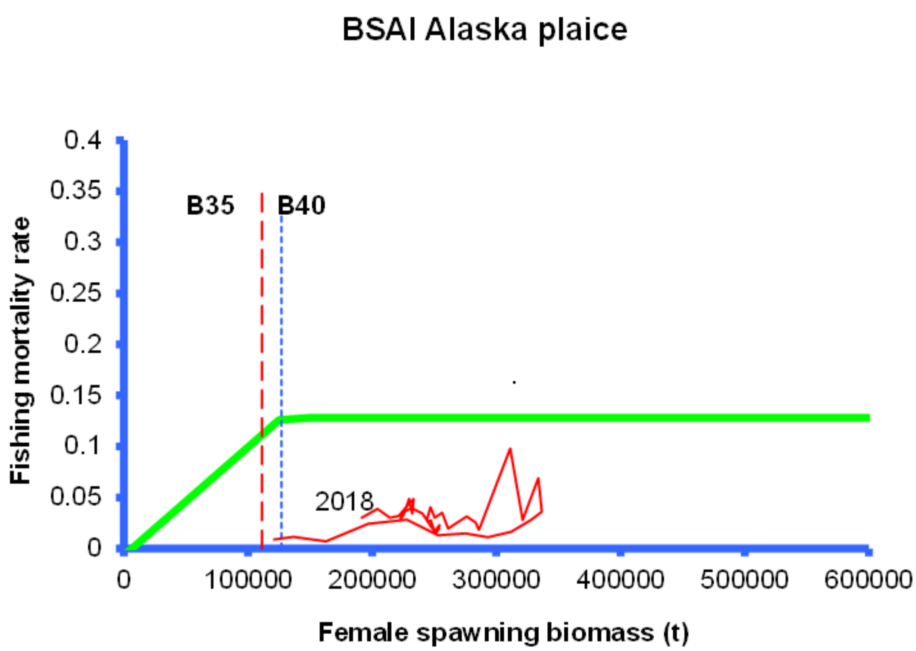
Quantity	As estimated or specified last year for:		As estimated or recommended this year for:	
	2017	2018	2018	2019
M (natural mortality rate)	0.13	0.13	0.13	0.13
Tier	3a	3a	3a	3a
Projected total (3+) biomass (t)	412,600	407,300	417,300	412,000
Female spawning biomass (t)	186,300	177,500	191,460	181,730
$B_{100\%}$	276,250	276,500	317,360	317,360
$B_{40\%}$	110,500	110,500	126,900	126,900
$B_{35\%}$	96,700	96,700	111,100	111,100
F_{OFL}	0.154	0.154	0.149	0.149
$maxF_{ABC}$	0.128	0.128	0.124	0.124
F_{ABC}	0.128	0.128	0.124	0.124
OFL (t)	42,800	36,900	41,170	38,800
maxABC (t)	36,000	32,100	34,590	32,700
ABC (t)				
	As determined last year for:		As determined this year for:	
Status	2015	2016	2016	2017
Overfishing	no	n/a	No	n/a
Overfished	n/a	no	n/a	no
Approaching overfished	n/a	no	n/a	no

But a large change in the B_{100} and B_{40} is notable from the previous assessment. The use of B_{40} or B_{35} rather than B_{MSY} is appropriate given that a stock-recruitment relationship cannot be estimated.

2018 SSB estimate (B) = 191,400 t

$B_{40\%}$	=	126,900 t
$F_{40\%}$	=	0.124
F_{ABC}	=	0.124
$F_{35\%}$	=	0.149
F_{OFL}	=	0.149

A phase plot of this stocks status is given below.



The Tier designation for this stock is 3a; as:

Tier 3: Reliable point estimates of B, $B_{40\%}$, $F_{35\%}$, and $F_{40\%}$.

3a) Stock status: $B/B_{40\%} > 1$ $F_{OFL} = F_{35\%}$; FABC < $F_{40\%}$

Given the lack of a stock recruitment relationship forcing the use of B40 or B35, a Tier of 3a is appropriate for this stock.

As noted previously, however, this translates to biomass reference points that are nearly below the historical abundance for this stock, and so re-examination in future assessments is recommended.

Also of note is the stock's status. While fishing mortality has been low, recruitment has also been in decline. This results in a slow decline, or erosion of the stocks status since. Clearly, something is not quite right, as the stock's TAC has been set to low levels, but the SSB has been in decline for decades.

TOR 3

Recommend how assessment data and/or models could be improved.

In the interests of time, alternate runs could not be conducted at the Review workshop. As such a number of recommendations for alternate runs are suggested. These include:

- Allowing either sex-specific natural mortality or catchability to vary
- Setting one sex's natural mortality while exploring either catchability, or the other sex's natural mortality (or both!)
- Changing the start year to 1982 when more data are present in the model

These alternate runs would look like, and be as complete as, those done for NRS. Such an approach could elucidate model behavior, as well as reveal sensitivities of the assessment to alternate stable states of nature.

A second, more broad recommendation is to examine the stock recruitment relationship in more detail. In particular, attempting to perhaps investigate environmental covariates is recommended. Additionally, using shorter time-frames of the stock-recruit relationship, as outlined for Yellowfin Sole, could also be productive. Clearly, the lack of a stock-recruitment relationship is hampering management of this stock.

Additionally, the reference points used for this stock should also be investigated. B_{40} or B_{35} is highly sensitive to estimates of B_{100}, which as can be seen, varies considerably from year to year.

In the documentation made available to the reviewers, the assessment also recommended genetic studies to further define the unit stock. Such a recommendation seems appropriate given the changes and recent exposure of this stock to fishing pressure given the reduction in sea ice.

Concluding Remarks and Process

I generally like to make comments more informally outside of addressing the TORs; to make comments on issues beyond the scope of the formal review. Please take with a grain of salt as it were, as they are just opinions of the overall meeting.

The meeting was held in the North Pacific Research Center of NMFS. It is a beautifully distracting venue, and I am glad there were few windows in the meeting room itself; it might have been quite a bit less productive if there were! While there was an issue with Security the first day, staff around the building were very helpful, and the accommodations for a meeting were quite generous. It is a nice place to do such a review.

Our hosts Tom Wilderbuer and Jim Ianelli were very helpful during the review and accommodating to our requests for further runs. They helped quite a bit with both logistics and overall support.

The meeting was much less formal than in other reviews; and it was refreshing. I thought it was productive and collaborative. It allowed the review to go unhampered by a lot of other issues found in more formal settings. I thought it increased efficiency greatly.

I'm unclear why NRS materials prior to the Review Workshop were from 2016 and why most of the presentation was on the 2016 assessment. Towards the end, if not after the review, we got a summary document containing the 2017 assessment. In the future it might be best if more recent assessments were made available to the panel prior to the meeting. In the end, it mattered very little as the assessments were virtually the same. For completeness, the 2017 assessment summary is attached below in the addendum below.

Likewise, I found it annoying that the Adobe documents made available prior to the meeting were password protected. These protections prevent commenting directly on the document or copying any portion into one's notes. Jim explained that this was due to a clerical snafu; but in the future it would be best if this was avoided.

I found it especially useful to have the half day of presentations on background materials for the NPFMC. Presentations on the survey, observer coverage, age and growth, as well as the management were very helpful to myself, and I believe the rest of the panel, as we moved through the review. I recommend doing so, if possible, in other reviews.

Overall, it was a fun and productive meeting. It was a real pleasure to work with Tom, Jim and the other panel members. All the staff, as well as the other reviewers, were quite knowledgeable and very friendly during the work. The process seemed to work very well for this assessment.

Reference cited

Chapman, D. G., and D. S. Robson. 1960. The analysis of a catch curve. Biometrics 16:354-368.

Gislason, H., Pope, J. G., Rice, J. C., and Daan, N. 2008. Coexistence in North Sea fish communities: implications for growth and natural mortality. – ICES Journal of Marine Science, 65: 514–530.

Hoenig, J. 1983. Empirical use of longevity data to estimate mortality rates. Fish. Bull. 82:898-903.

Nichol, D. R . 1995. Spawning and maturation of female Yellowfin Sole in the eastern Bering Sea. In Proceedings of the international flatfish symposium, October 1994, Anchorage, Alaska, p. 35-50. Univ. Alaska, Alaska Sea Grant Rep. 95-04.

Somerton, D. A. and P. Munro. 2001. Bridle efficiency of a survey trawl for flatfish. Fish. Bull. 99:641-652(2001).

Stark, J. W. 2009. Contrasting maturation and growth of Northern Rock Sole in the eastern Bering Sea and Gulf of Alaska for the purpose of stock management. NAJFM, 32:1, 93-99.

TenBrink, T. T. and T. K. Wilderbuer. 2015. Updated maturity estimates for flatfishes (Pleuronectidae) in the eastern Bering Sea, with notes on histology and implications to fisheries management. Coastal and Marine Fisheries: Dynamics, Management and Ecosystem Science. O:1-9. 2015. DOI: 10.1080/19425120.2015.1091411.

Zhang, C. I., T.K. Wilderbuer, and G.E. Walters. 1998. Biological characteristics and fishery assessment of Alaska Plaice, Pleuronectes quadrituberculatus, in the Eastern Bering Sea. Marine Fisheries Review 60(4), 16-27.

Addendum: Northern Rock Sole 2017 assessment.

Assessment of the Northern Rock Sole stock in the Bering Sea and Aleutian Islands

Thomas Wilderbuer and Daniel Nichol

November 2017

Executive Summary

Northern Rock Sole (*Lepidopsetta polyxystra*) are assessed on a biennial stock assessment schedule as part of the National Marine Fisheries Service assessment prioritization plan implemented in 2017. For Bering Sea/Aleutian Islands partial assessments, an executive summary is presented to recommend harvest levels for the next two years. Please refer to last year's full stock assessment report for further information regarding the stock assessment model (Wilderbuer and Nichol, 2016, available online at https://www.afsc.noaa.gov/REFM/Docs/2016/BSAIrocksole.pdf). A full stock assessment document with updated assessment and projection model results is scheduled to be presented in next year's SAFE report.

A statistical age-structured model is used as the primary assessment tool for the Bering Sea/Aleutian Islands Northern Rock Sole assessment, a Tier 1 stock. This assessment consists of a population model, which uses survey and fishery data to generate a historical time series of population estimates, and a projection model, which uses results from the population model to predict future population estimates and recommended harvest levels. The data sets used in this assessment include total catch biomass, fishery age compositions, trawl survey abundance estimates and trawl survey age compositions. In a partial assessment year, the full assessment model is not rerun but instead a Tier 1 projection model with an assumed future catch is run to estimate the stock level in the next two years. This incorporates the most current catch information without re-estimating model parameters and biological reference points. A Tier 1 partial projection rule is implemented to estimate the 2019 ABC and OFL.

The Tier 1 projection operates within the full assessment model by projecting estimates of the female spawning biomass, age 6+ total biomass, ABC and OFL ahead two years. Since the full assessment model is not rerun in this assessment, only the projected values from the 2016 assessment are available (2017 and 2018) whereby values for 2019 are not estimated. The 2019 values are determined by a linear fit to the 2017 and 2018 estimates. If the trend is increasing, then the 2019 values are a roll-over of 2018. If the trend is decreasing, the projected proportional decrease from 2017 to 2018 is applied to 2018 to get 2019.

Summary of Changes in Assessment Inputs

Changes in the input data: There were no changes made to the assessment model inputs since this was not a full assessment year. New data added to the Tier 3 projection model, used to forecast stock condition out to year 2030, included an updated 2016 catch estimate (45,006 t) and new catch estimates for 2017. The 2017 catch was estimated by setting the catch as of October 21, 2017 as the final 2017 catch (35,069 t). To estimate future catches through 2030, the catches that corresponded to the average F of the most recent 5 years were used.

Changes in the assessment methodology: There were no changes in assessment methodology since this was an off-cycle year.

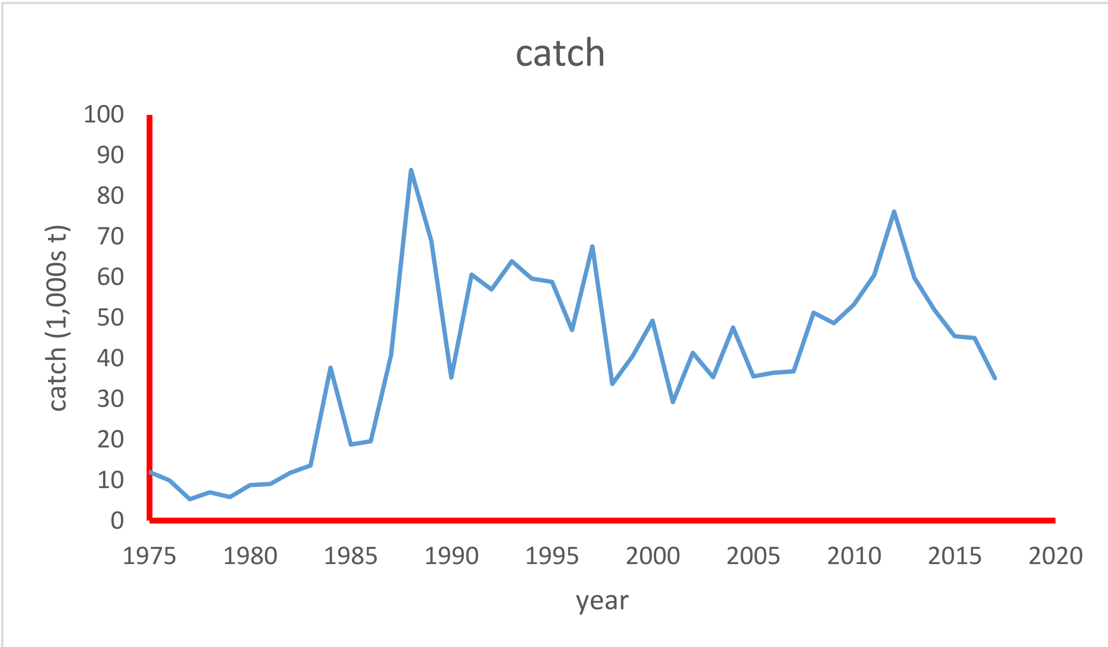
Summary of Results

For the 2018 fishery, the recommend harvest is the maximum allowable ABC of 143,100 t from the Tier 1 projection model. This ABC is 14% less than year's ABC of 155,100 t. Reference values for BSAI RE/BS rockfish are summarized in the following table, with the recommended ABC and OFL values for 2018 in bold.

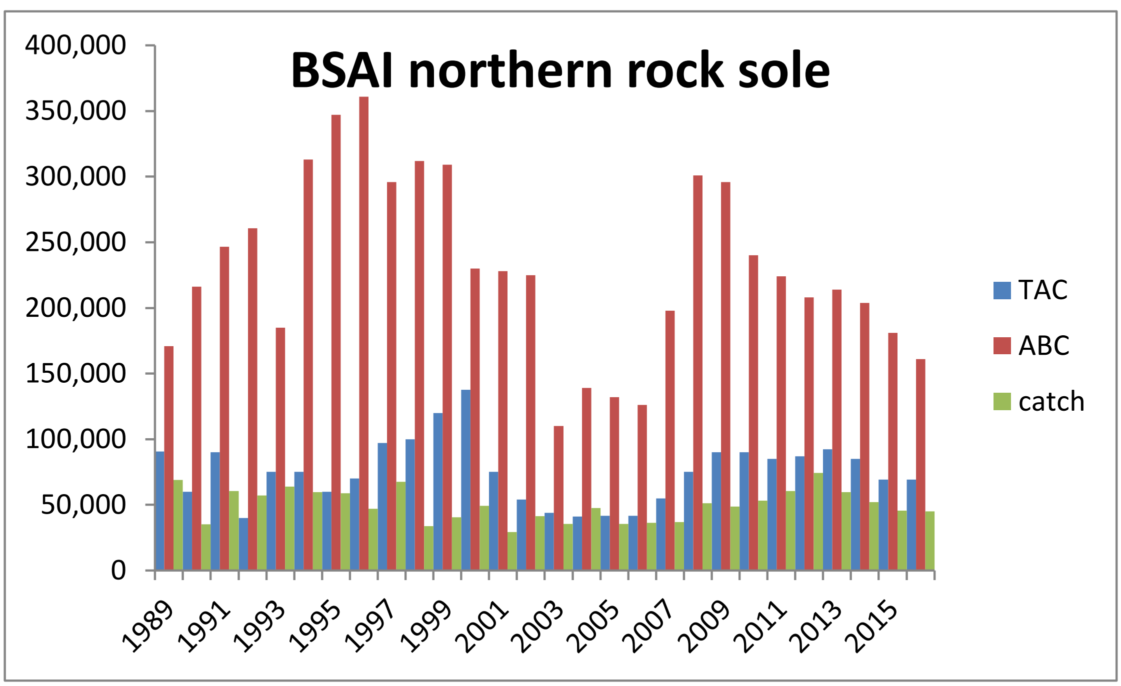
Quantity	As estimated or *specified last* year for:		As estimated or *recommended this* year for:	
	2017	2018	2018	2019
M (natural mortality rate)	0.15	0.15	0.15	0.15
Tier	1a	1a	1a	1a
Projected total (age 6+)	1,000,600	923,200	923,200	852,000
Female spawning biomass (t)	539,500	472,200	472,200	413,300
Projected				
B_0	678,310		678,310	
B_{MSY}	257,000	257,000	257,000	257,000
F_{OFL}	0.160	0.160	0.160	0.160
$maxF_{ABC}$	0.155	0.155	0.155	0.155
F_{ABC}	0.155	0.155	0.155	0.155
OFL (t)	159,700	147,300	**147,300**	136,000
maxABC (t)	155,100	143,100	**143,100**	132,000
ABC (t)	155,100	143,100	**143,100**	132,000
Status	As determined *last* year for:		As determined *this* year	
	2015	2016	2016	2017
Overfishing	No	n/a	No	n/a
Overfished	n/a	No	n/a	No
Approaching overfished	n/a	No	n/a	No

The stock is not being subject to overfishing, is not currently overfished, nor is it approaching a condition of being overfished. The tests for evaluating these three statements on status determination require examining the official total catch from the most recent complete year and the current model projections of spawning biomass relative to $B_{MSY\%}$ for 2017 and 2018. The estimated total catch for 2017 is 35,069 t, far below the 2017 OFL of 159,700 t; therefore, the stock is not being subjected to overfishing. The estimates of spawning biomass for 2017 and 2018 from the 2016 stock assessment are 539,500 t and 472,200 t, respectively. Both estimates are well above the estimate of $B_{MSY\%}$ at 257,000 t and, therefore, the stock is not currently overfished nor approaching an overfished condition.

Fishery Trends

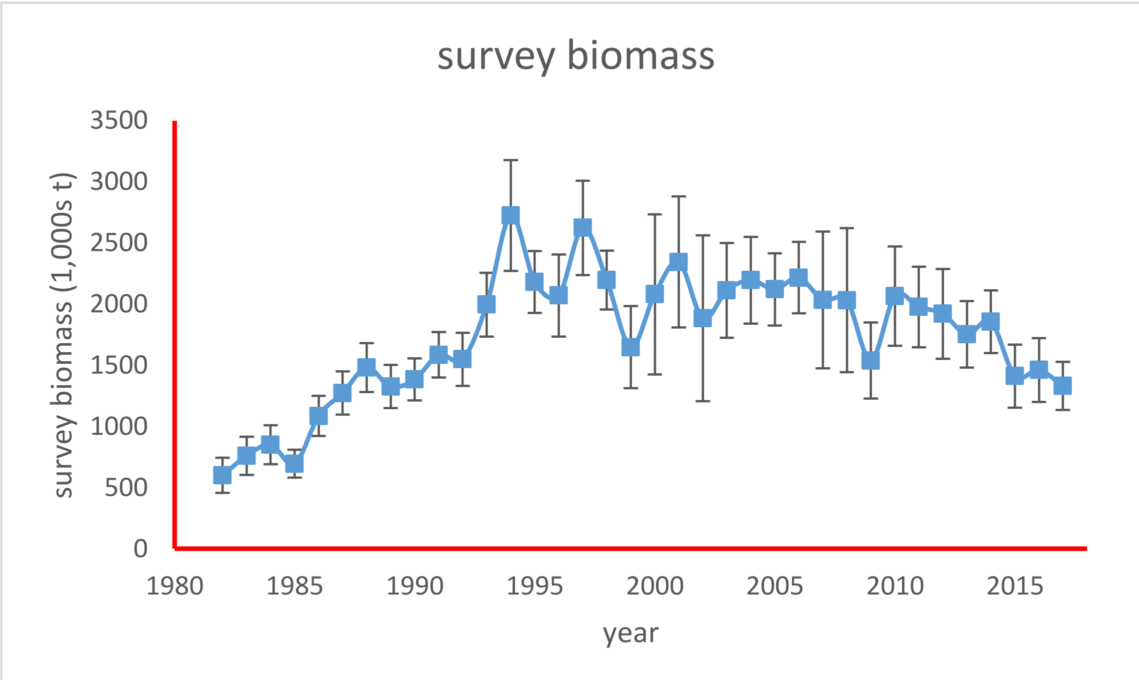


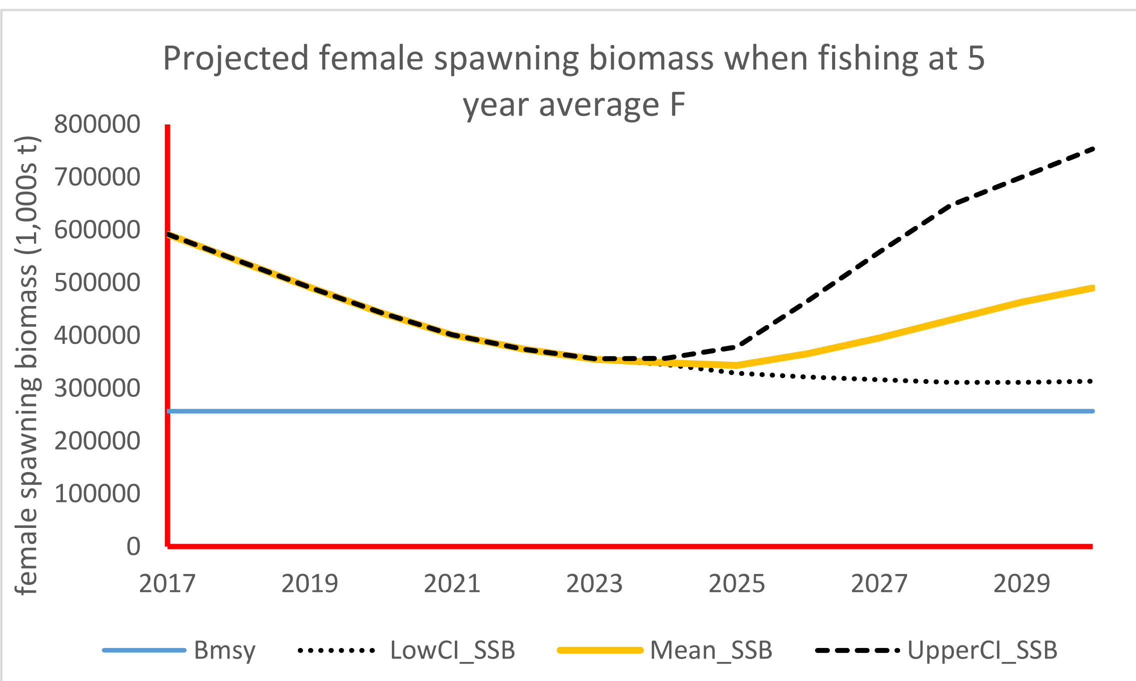
The Northern Rock Sole catch in 2017 of 35,069 t is below the 1975-2017 long term average of 40,000 t, and well below the annual ABC in every year. Catches primarily are made during a late-winter/early spring roe fishery and also as bycatch in the Yellowfin Sole fishery. Retention rates are high, estimated at 98% in 2015.



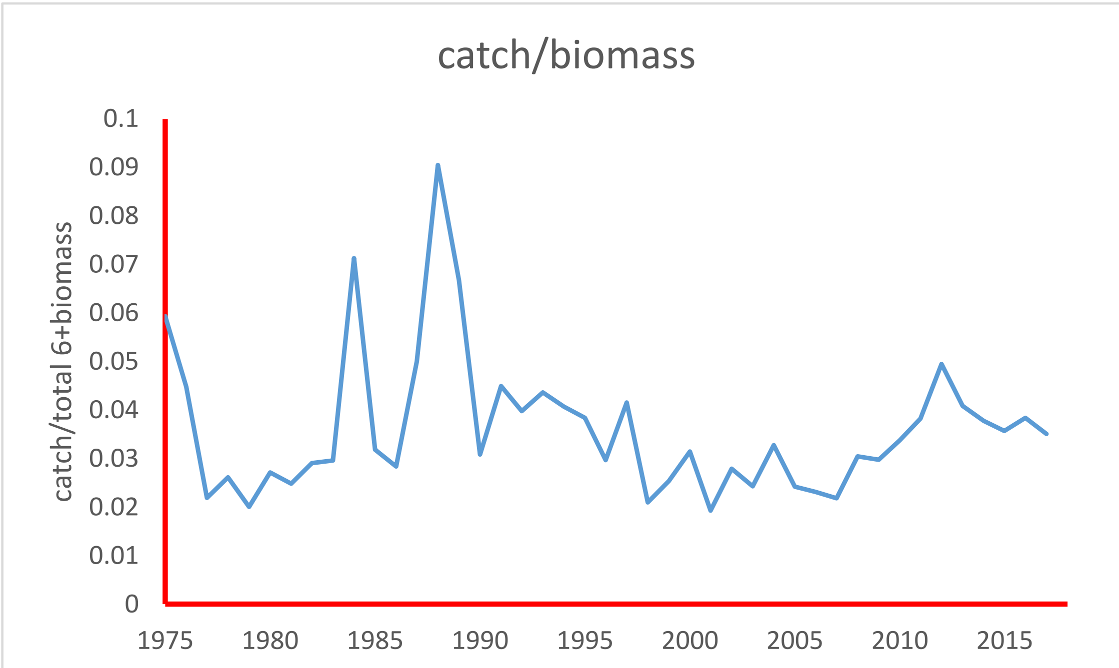
Survey Trends

The 2017 shelf trawl survey abundance estimate decreased about 11% from the 2016 estimate and has been in a downward trend since about 2008, currently about half of the peak value estimated for 1994.





The Northern Rock Sole stock is projected to remain above the B_{MSY} level of female spawning biomass while declining through 2024.



Appendix

Estimating Northern Rock Sole recruitment in the last (most recent) 6 years of the assessment using environmental covariates

Dan Cooper, Lauren Rogers and Tom Wilderbuer

Difficulties exist in estimating Northern Rock Sole recruitment at young ages since they do not appear in BSAI survey catches until age 3 and not in survey age sampling until age 4 or 5. They are estimated to be 25 and 40% selected by the survey trawl (males and females respectively) at age 3 and 95 and 98% selected at age 5. The age 4 and 5 fish that do end up in the age samples are quite rare, typically only 7 fish out of 500 on an annual basis. Therefore, there is not a lot of information to inform the stock assessment model estimates of year class strength for the last (most recent) 6 years. Some assessments provide estimates for the last 3 years by using an average of the estimated values to provide more credible values of year class strength. Here we propose to use two environmental covariates in regression modeling to estimate the unknown recruitment, and then compare those estimates with future estimates derived from fitting full age composition data in the stock assessment model.

Studies on the influence of environmental variables on BSAI Northern Rock Sole recruitment have shown that both on-shelf springtime winds (Wilderbuer et al. 2002, Wilderbuer et al. 2013) and above average water-temperatures in nursery areas (Cooper et al. 2014, Cooper and Nichol 2016) are positively correlated with Northern Rock Sole recruitment. Spring wind direction was obtained from the Ocean Surface Current Simulation Model (OSCURS) and was classified as either on- or across-shelf or off-shelf, depending on the ending longitude position after 90 days of drift starting from a locale in a known spawning area. Water temperature effects were calculated from the percent of the known Northern Rock Sole nursery area (Cooper et al. 2014) that is in the cold pool each year from annual trawl survey bottom temperature data. For most models, percentage of the northern nursery area covered by the cold pool was used as a continuous variable. In one model, the percent cold pool was used a categorical variable, dividing years into cold and not-cold categories under the hypothesis that there is some amount of cold pool coverage of the northern nursery area that inhibits use of the northern nursery area and precluded high overall recruitment for the EBS in that year. Both indices extend back to 1982 for this analysis. Estimates of female spawning stock biomass were also included in the analysis for model runs when recruitment was estimated from a Ricker stock-recruitment model with environmental variables.

The analysis seeks to answer the following questions using multiple models.

Q1: Do onshore winds and the size of the cold pool (as a percentage of the nursery area) affect recruitment of Northern Rock Sole?

Q2: Does the effect of the cold pool on recruitment depend on the presence of favorable winds? (i.e. is there a significant interaction?)

Q3: Does including wind and cold pool covariates in the stock-recruitment model improve predictions of age-4 recruitment?

We assessed the performance of a suite of models, ranging from a simple Ricker stock-recruit model, to Ricker models with environmental covariates, to models with only environmental covariates. For parsimony, we also assessed simpler forecasting models that used the previous year recruitment or running mean recruitment. We also tested for an interaction between the cold pool effect and winds, because nursery habitat conditions may only matter if winds were favorable for onshore transport (i.e. the fish have to get there in the first place).

We assessed 14 models. Thirteen are the same models from the 2016 stock assessment appendix, and we present one new model, the categorical model.

1) Ricker model
2) Ricker model with % cold pool covariate
3) Ricker model with wind covariate
4) Ricker model with % cold pool covariate + wind covariate
5) Ricker model with an interaction between % cold pool and wind (hypothesis is that the thermal conditions on the nursery grounds only matter if winds are favorable)
6) Same as above, but cold pool slope set to 0 if unfavorable winds
7) Regression model with % cold pool
8) Regression model with wind
9) Regression model with % cold pool + wind
10) Regression model with interaction between % cold pool and wind
11) Same as above, but cold pool slope set to 0 if unfavorable winds
12) Categorical model with threshold low temperature for recruitment success (hypothesis is that there is a some amount of coverage by the cold pool which inhibits use of the northern nursery area and precludes high recruitment)
13) Previous year recruitment (t-1)
14) Running mean recruitment (t:(t-1))

We compared model performance using traditional statistical methodology on all data (AIC), as well as by using two prediction methods. First we used a leave-one-year out analysis: we left out one year of data, fit the model to the remaining 27 years of data, and then compared the prediction for the left-out year to the observed value. Second, we did a one-step-ahead forecast: beginning with year 11 (1992), we used the data collected up to that year to fit the model, and then compared the prediction for that year with the observation. We repeated for all remaining years. We calculated the mean squared error for each prediction: (Observed – Predicted)^2. Models were initially fit using log(recruitment) as the response, so the mean squared error is for the difference between the observed and predicted log(recruitment). However, the mean squared error can also be calculated based on the predicted recruitment on the real scale. In this case, Duan's smearing estimate for the lognormal re-transformation bias was used to adjust the mean of the exponentiated log(recruitment) to be equal to the mean recruitment. Both results are given in Table 1.

In the 2016 Northern Rock Sole SAFE appendix, we presented modeled and observed recruitment from 1982 through 2009. In this assessment, we also use models #1-12 to predict recruitment for the 2010 through 2016 year classes using the environmental covariates and estimated spawning stock biomass (Figure 1).

The environmental-factors based recruitment models with the lowest prediction errors included both the winds and cold pool indices (Table 1). The Categorical Model had the lowest AIC score and the lowest MSE in both the LOYO log scale and LOYO real scale prediction methods (Table 1). Other environmental-factors based models with the best predictive scores include the Coldpool + Wind model and the Coldpool*Wind model. While the model with an interaction between the Coldpool and Wind had reasonable predictive ability, the interaction term was not statistically significant. The Previous Year Model had the lowest (best) MSE for the 1 step ahead prediction method for both log and real scales, and had the second best score in the LOYO log scale, indicating some autocorrelation in recruitment; however, the Previous Year Model is capable of predicting recruitment only one year class into the future, limiting its utility. The six models including a Ricker spawning biomass term had the highest (worst) AIC scores and generally had poor MSE scores relative to the other models.

Recruitment predictions from models with environmental covariates suggest that conditions were conducive to relatively strong recruitment in 2011, 2014, and 2015, and moderate to weak recruitment in 2010, 2012, 2013 and 2016 (Figure 1). As recruitment estimates become available from the stock assessment model, we will continue to assess the suitability of these models for forecasting Northern Rock Sole recruitment.

Table 1: Mean squared error (MSE) is the mean of the squared prediction errors for each model. LOYO = Leave one year out. Lower values for MSE indicate lower prediction errors. The three best (lowest) AIC and MSE scores are in bold.

	Model	df	AICc	MSE (LOYO, log-scale)	MSE (1 step ahead, log-scale)	MSE (LOYO, real scale)	MSE (1 step ahead, real scale)
1	Ricker	3	67.8	0.69	0.85	725	635
2	Ricker + coldpool	4	66.5	0.73	0.86	678	539
3	Ricker + wind	4	67.9	0.67	0.81	703	625

4	Ricker + coldpool + wind	5	63.9	0.66	0.78	635	509
5	Ricker + coldpool*wind	6	65.0	0.64	0.85	622	514
6	Ricker + coldpool*wind (slope=0)	5	66.0	0.69	0.81	655	536
7	coldpool	3	57.8	0.60	0.69	545	600
8	wind	3	60.9	0.58	0.69	585	631
9	coldpool + wind	4	**54.9**	0.55	0.61	**531**	**504**
10	coldpool*wind	5	**55.7**	**0.53**	0.71	**522**	570
11	coldpool*wind (slope=0)	4	57.2	0.58	0.64	552	533
12	Categorical	4	**45.5**	**0.41**	**0.47**	**456**	**412**
13	Previous Year	NA	NA	**0.46**	**0.45**	533	**382**
14	Running Mean	NA	NA	0.62	0.74	637	638

Literature Cited

Cooper, D.W. and Nicol, D. 2016. Juvenile Northern Rock Sole spatial distribution and abundance are correlated in the eastern Bering Sea: spatially-dependent production linked to temperature. ICES Journal of Marine Science, 73, 1136-1146.

Cooper D, Duffy-Anderson J.T., Norcross B.L., Holladay B.A., Stabeno P.J. 2014. Northern Rock Sole (*Lepidopsetta polyxystra*) juvenile nursery areas in the eastern Bering Sea in relation to hydrography and thermal regimes. ICES Journal of Marine Science 72, 515-527.

Wilderbuer, T., A., Hollowed, A., Ingraham, J., Spencer, P., Conner, L., Bond, N., Walters, G. 2002. Flatfish recruitment response to decadal climatic variability and ocean conditions in the eastern Bering Sea. Progress in Oceanography, 55, 235-247.

Wilderbuer, T., W. Stockhausen, N. Bond. 2013. Updated analysis of flatfish recruitment response to climate variability and ocean conditions in the Eastern Bering Sea. Deep Sea Research II, 94, 157-164.

Appendix 1: Bibliography of materials provided for review

Wilderbuer, T.K., D.G. Nichol, and J. Ianelli, 2017. Assessment of the Yellowfin Sole stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions.
https://www.afsc.noaa.gov/REFM/stocks/assessments.htm

Wilderbuer, T.K. and D. G. Nichol. 2017. Assessment of the Northern Rock Sole stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions.
https://www.afsc.noaa.gov/REFM/Docs/2016/BSAIrocksole.pdf

Wilderbuer, T.K. and D. Nichol. 2017. Assessment of the Alaska Plaice stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK.
https://www.afsc.noaa.gov/REFM/stocks/assessments.htm

Introduction to Bering Sea/Aleutian Islands SAFE
https://www.afsc.noaa.gov/REFM/stocks/assessments.htm

Somerton, D., K. Weinberg, P. Munro, L. Rugolo and T. Wilderbuer. 2017. The effects of wave-induced vessel motion on the geometry of a bottom survey trawl and the herding of Yellowfin Sole. Fish. Bull. 116:21–33 (2018)
https://www.st.nmfs.noaa.gov/spo/FishBull/1161/somerton.pdf

Appendix 2: A copy of this Statement of Work

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)

National Marine Fisheries Service (NMFS)

Center for Independent Experts (CIE) Program

External Independent Peer Review of Fisheries stock assessments for Yellowfin Sole, Northern Rock Sole and Alaska Plaice

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[1]. Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Alaska Fisheries Science Center's (AFSC) Resource Ecology and Fisheries Management Division (REFM) requests an independent review of the integrated stock assessments that have been developed for three Bering Sea flatfish species; Yellowfin Sole, Northern Rock Sole and Alaska Plaice. The fishery for these species is managed by the North Pacific Fisheries Management Council. The sum of the ABCs for these three species is 455,200 metric tons (t) in 2018, with catch levels annually set lower than the ABC due to a 2.0 million t harvest cap for all species and constraints due to Pacific halibut bycatch limits and

[1] http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf

markets. The catch limits are established using automatic differentiation (AD) Model software that uses survey abundance data and survey and fishery age and length composition data with a harvest control rule to model the status and productivity of these stocks and set quotas. Having these assessments vetted by an independent expert review panel is a valuable part of the AFSC's review process. The Terms of Reference (TORs) of the peer review and the tentative agenda of the meeting are below.

Requirements

NMFS requires three reviewers to conduct an impartial and independent peer review in accordance with the SOW, OMB Guidelines, and the TORs below. The reviewers shall have working knowledge and recent experience in the application of fisheries stock assessment processes and results, including population dynamics, separable age-structured models, harvest strategies, survey methodology, and the AD Model Builder programming language. They should also have experience conducting stock assessments for fisheries management.

Tasks for reviewers

1. Review the following background materials and reports prior to the review meeting:

Wilderbuer, T.K., D.G. Nichol, and J. Ianelli, 2017. Assessment of the Yellowfin Sole stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions.
https://www.afsc.noaa.gov/REFM/stocks/assessments.htm

Wilderbuer, T.K. and D. G. Nichol. 2017. Assessment of the Northern Rock Sole stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions.
https://www.afsc.noaa.gov/REFM/Docs/2016/BSAIrocksole.pdf

Wilderbuer, T.K. and D. Nichol. 2017. Assessment of the Alaska Plaice stock in the Bering Sea and Aleutian Islands. *In* Stock Assessment and Fishery Evaluation Report for the Groundfish Resources of the Bering Sea/Aleutian Islands Regions. North Pacific Fisheries Management Council, Anchorage, AK.
https://www.afsc.noaa.gov/REFM/stocks/assessments.htm

Introduction to Bering Sea/Aleutian Islands SAFE
https://www.afsc.noaa.gov/REFM/stocks/assessments.htm

Somerton, D., K. Weinberg, P. Munro, L. Rugolo and T. Wilderbuer. 2017. The effects of wave-induced vessel motion on the geometry of a bottom survey trawl and the herding of Yellowfin Sole. Fish. Bull. 116:21–33 (2018). doi: 10.7755/FB.116.1.3
https://www.st.nmfs.noaa.gov/spo/FishBull/1161/somerton.pdf

2. Attend and participate in the panel review meeting
   - The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers
3. After the review meeting, reviewers shall conduct an independent peer review in accordance with the requirements specified in this SOW, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus

4. Each reviewer may assist the Chair of the meeting with contributions to the summary report, if required by the TORs
5. Deliver their reports to the Government according to the specified milestone dates

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 40 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place and Period of Performance

The place of performance shall be at the contractor's facilities, and at the Alaska Fisheries Science Center, Seattle, Washington, USA. The period of performance shall be from the time of the award through June 1, 2018. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables

The contractor shall complete the tasks and deliverables in accordance with the following schedule:

Within two weeks of award	Contractor selects and confirms reviewers
No later than April 12, 2018	Contractor provides the pre-review documents to the reviewers
April 16-18, 2018	Panel review meeting
May 11, 2018	Contractor receives draft reports
May 25, 2018	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

1. The reports shall be completed in accordance with the required formatting and content
2. The reports shall address each TOR as specified
3. The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $11,000.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

NMFS Project Contact

Thomas K Wilderbuer
7600 Sand Point Way NE, Bldg. 4
Seattle, WA 98115-6349
tom.wilderbuer@noaa.gov

Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.

2. The report must contain a background section, description of the individual reviewers' roles in the review activities, summary of findings for each TOR in which the weaknesses and strengths are described, and conclusions and recommendations in accordance with the TORs.

   a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.
   b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.
   c. Reviewers should elaborate on any points raised in the summary report that they believe might require further clarification.
   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
   e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:
   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of this Statement of Work
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Terms of Reference for the Peer Review

For the three assessments (Bering Sea and Aleutian Islands Yellowfin Sole, Northern Rock Sole, and Alaska Plaice) consider the following:

1. Evaluate the strengths and weaknesses of the assumptions made in applying the stock assessment model including how survey indices are scaled to the populations. Specifics might include:
    a. How natural mortality estimates are estimated/applied
    b. Assumptions about survey "catchability"
    c. Application of fishery and survey age-specific schedules (maturity, body mass, selectivity)
    d. The application (or lack thereof) of a stock-recruitment relationship (and associated parameter estimates)
2. Evaluate the stock assessment approach used focusing specifically on how fisheries and survey data are compiled and used to assess the stock status relative to stated management objectives under the Bering Sea and Aleutian Islands Fishery Management Plan (FMP) and the Magnuson-Stevens Act requirements. Elements should consider:
    a. The FMP "Tier" designation
    b. Fishing rate estimation relative to overfishing definitions
    c. Stock status determinations relative to B_{MSY}
3. Recommend how assessment data and/or models could be improved.

CIE Flatfish Assessment Review Agenda

NMFS Alaska Fisheries Science Center

Room 2079

7600 Sand Point Way NE, Building 4

Seattle, Washington

AGENDA *INITIAL VERSION* April 16-18, 2018

Monday April 16th

9:00	Welcome and Introductions, adopt agenda, TORs	
9:15	Overview (species, biology, surveys, fishery, catch levels, ABCs, TACs, bycatch) **Tom**	
10:00	Bering Sea shelf trawl survey	**Dan Nichol**
10:30	Observer Program	**Marlon Conception**
11:00	Coffee break	
11:20	Age Determination	**Tom Helser**
11:50	Effect of rationalization on flatfish fisheries	**Alan Haynie**
12:30	Lunch	
1:30	BSAI Yellowfin Sole	**Tom and Jim**

Tuesday April 17th

9:00	BSAI Northern Rock Sole	**Tom and Jim**
11:00	Coffee break	
11:20	BSAI Northern Rock Sole (continued)	**Tom and Jim**
12:30	Lunch	
1:30	BSAI Alaska Plaice	**Tom and Jim**

Wednesday April 18th

9:00	CIE panel discussion (assessment authors will be available)
11:00	Coffee break
11:20	Discussions continue
12:30	Lunch

Appendix 3: Panel membership

Attendees:

- Dan Nichol
- Marlon Conception
- Tom Helser
- Alan Haynie
- Tom Wilderbuer
- Jim Ianelli
- One unnamed industry liaison
- Matthew Cieri (CIE panelist)
- Joe Powers (CIE panelist)
- Yan Jiao (CIE panelist)