Evaluation and analysis of the Gulf of Alaska Pacific Ocean Perch stock assessment

Pete Hulson and Curry Cunningham

September 2017

Introduction

Since 2014 the NPFMC Groundfish Plan Team (PT) and Science and Statistical Committee (SSC) have provided several suggestions and comments to be investigated for potential improvement of and implementation into the Gulf of Alaska (GOA) Pacific Ocean Perch (POP) assessment. Responses to a number of these comments are intended to be included within the full assessment presented in November 2017. The purpose of this document is to provide the PT with preliminary analyses at the September 2017 meeting for discussion in preparation for the full assessment. The following PT and SSC comments will be addressed in this document:

- *"The Plan Team recommends evaluation of how the data weights given to the various fishery and survey age and length composition data affect the estimates of recruitment and age composition."* (Plan Team, September 2014)
- *Many assessments are currently exploring ways to improve model performance by re-weighting historic survey data. The SSC encourages the authors and PTs to refer to the forthcoming CAPAM data-weighting workshop report.* (SSC, December 2015)
- *The SSC recommends that the Gulf of Alaska Groundfish Plan Team (GOA GPT), BSAI GPT, and CPT encourage the continued use of multiple approaches to data weighting (not just the Francis (2011) method, but also including the harmonic mean and others).* (SSC, October 2016)
- *The Team recommends increasing the plus group for the length compositions to evaluate model performance.* (Plan Team, November 2015)
- *In September (2014), the PT and SSC recommended evaluating data weighting for fishery and survey age and length compositions with respect to estimates of recruitment and age compositions. The authors note that this issue pertains to all GOA rockfish assessments and plan to do a more thorough evaluation of this issue for future assessments. The SSC agrees and would recommend a broader look at the issue across all GOA rockfish species, and to consider relevant recommendations from the 2015 CAPAM workshop on data weighting. Further, the SSC concurs with the PT recommendations for the next full POP assessment to investigate 1) increasing the plus group for length compositions to evaluate model performance, 2) using an alternate trawl survey index, 3) using alternative length bins, 4) including sample sizes for composition data, and 5) relating fishery selectivity to average depth fished.* (SSC, December 2015)

Following from the comment made by the SSC at its December 2015 meeting there are 4 general categories in which the first five PT and SSC comments above can be grouped: (1) analysis of length composition data, (2) analysis of the input sample sizes used for age and length composition data, (3) analysis of fishery selectivity, and (4) analysis of a GLMM alternative to the design based estimates currently used for the bottom trawl survey index. In this document, analyses and evaluations are provided for each of these categories. Within each category the historical usage and changes leading to the current methods used in the POP assessment is provided first as background. Then, the results of the influence of the different scenarios or methods selected for analysis on the assessment model from 2015 is reported. Statistics of model performance to evaluate the influence of the alternative methods investigated include 1) the root mean squared error (RMSE) of model fit to the bottom trawl survey biomass index, the bottom trawl survey age composition, and the fishery age composition; 2) the estimates of spawning biomass from the assessment model, the coefficient of variation (CV) in spawning biomass, and the estimates of recruitment from 1961-2015; and 3) the percent difference compared to the 2015 assessment model for key parameter estimates, including $F_{40\%}$, bottom trawl survey catchability (q), natural mortality (M), recruitment variation (σ_r), and the log of mean recruitment (ln$R$). In some instances, additional results are provided for further discussion and exploration of the models and methods tested.

Category 1: Analysis of length composition data

Fishery length compositions from 1963-1977, 1991-1992, and 1995-1997 are the only length compositions fit by the POP assessment model. The length bins for POP fishery data are ≤12 cm, 13-15 cm, 16-34 cm in 1 cm increments, 35-38 cm, and ≥39 cm. These length bins have been historically utilized since the inception of the current age-structured model within AD model builder (ADMB) in 2001 (Heifetz et al. 2001). In 2003, it was noted that the mean length of fishery catch samples 1963-1977 was smaller compared with catch samples after 1990, and using a single size-age transition matrix for the entire time series of fishery length composition was problematic. It was hypothesized that this could have been due to a density dependent increase in growth resulting from smaller population sizes after the 1980s compared to the larger abundances in the 1960s and 1970s. In response, a second size-age transition matrix was added that took into account a 6% reduction in growth for the 1963-1977 fishery length composition data (Hanselman et al. 2003). Since 2003, the only change in the assessment model as it pertains to fitting the fishery length composition data, was implementation of growth estimates within the size-age transition matrices that took into account the length-stratified sampling design of the bottom trawl survey in the last full assessment (Hulson et al. 2015).

Both the PT and SSC have requested that the bin structure and the plus length group be investigated for the fishery length composition data fit by the POP assessment. To respond to these requests, 4 bin structures and an additional plus length group are investigated. Analysis of the length composition begins by first defining the starting bin for the fishery length composition data. Figure 1 shows the proportion at length by age (ranging from age-2 to age-14+) observed from the bottom trawl survey from 1984-2013. The current starting bin is ≤12 cm, however, a very small proportion of age-2 fish (the starting age for the assessment) have ever been observed to be smaller than 12 cm (top row Figure 1). The current starting bin for POP in the Bering Sea/Aleutian Island (BSAI) assessment is ≤15 cm (Spencer et al. 2016), and from the observed bottom trawl survey data in the GOA the lower 95% confidence interval for age-3 is 16 cm (Figure 1). Therefore, so that the starting bin would consist of primarily age-2 fish the starting bin for the alternative bin structures investigated was set at ≤16 cm. To test the influence of bin structure on the POP assessment 4 alternative bin structures were evaluated that binned the fishery length frequencies into 1 cm, 2 cm, 3 cm, and 4 cm bins from ≤16 cm to the plus length group. The current plus length group for both the GOA and BSAI POP assessments is ≥39 cm. In the bottom row of Figure 1 the observed data from the bottom trawl survey was pooled for ages greater than 14. The upper 95% confidence interval of length for ages greater than 14 is 45 cm. To respond to the PT and SSC request to evaluate the influence of the plus length group this length was selected as the additional plus length group investigated. In summary, 4 bin structures ranging from 1 to 4 cm in width and two plus length groups (≥39 cm and ≥45 cm) were investigated in comparison to the 2015 assessment results.

In general, the influence of alternative bin structures and plus length groups compared to the 2015 assessment was minor (Tables 1-2, Figure 2). The RMSEs for the bottom trawl survey biomass index, the bottom trawl survey age composition, and the fishery age composition compared to the 2015 assessment were very similar, and, were slightly smaller with a plus length group of ≥45 cm, with the exception of the survey age composition (Table 1). The CV in spawning biomass was generally larger for the alternative length bins and plus groups investigated compared to the 2015 assessment, although, only by about 3% (Figure 2, middle panel). The estimated spawning biomass and recruitment (Figure 2, top and bottom panels) was nearly identical to the 2015 assessment, with some slight differences in recruitment estimates at the beginning of the time series. The percent difference in key parameter estimates compared to the 2015 assessment were small, ranging from -1.1% to 3.7% (Table 2). The only improvement that resulted when compared to the 2015 assessment was a slight reduction in the CV in spawning biomass in recent years (after the early 2000s) upon setting the length bins to either 1 or 2 cm with the current plus length bin of ≥39 cm (Figure 2, middle panel). Overall, because of the negligible differences between the 2015 assessment and the alternative length bins and plus length group, any changes to the current structure of the length composition data for the full assessment in 2017 are difficult to recommend. The PT is requested to recommend an alternative bin structure to the current structure that they would like to see further investigated for the full 2017 assessment in November.

Category 2: Analysis of the input sample sizes used for age and length composition data

Currently, the input sample sizes for the fishery and bottom trawl survey age compositions are set at the square root of sample size, and for the fishery length composition are set at the number of hauls scaled to a maximum of 100. In the original formulation of the GOA POP assessment in ADMB the input samples sizes for composition data were set at the number of hauls from which age and length observations were taken, scaled to a maximum of 100 (Heifetz et al. 2001), which followed from the original ADMB formulation of the generic rockfish assessment template (Courtney et al. 1999). In 2005, the bottom trawl survey and fishery age composition sample sizes were set at the square root of sample size, which is the current convention.

Both the PT and SSC have recommended that the input sample sizes used for composition data within the GOA POP assessment be evaluated. This is in response to a number of recent studies that have been published on effective samples size and data weighting in age-structured assessment models (e.g., Maunder 2011, Francis 2011, Hulson et al. 2012) and a workshop performed by the Center for the Advancement of Population Assessment Models in 2015 (CAPAM, e.g., Maunder et al. 2017). Further, a number of assessments at the Alaska

Fishery Science Center (AFSC) have been actively evaluating data weighting and input sample sizes (e.g., McGilliard et al. 2013, Spencer and Ianelli 2016). Response to this request by the PT and SSC is provided by evaluating different methods of estimating input sample size for composition data.

Two general classifications of input sample size estimation were evaluated for the GOA POP assessment: iterative estimation, in which input sample sizes are estimated through repeated iterative processes, and internal estimation, in which input sample sizes are estimated as parameters within the assessment model. Within each classification, iterative and internal estimation, two methods were selected for evaluation. The first method selected within the iterative classification was the method of McAllister and Ianelli (1997) and was adopted in the BSAI POP assessment in 2016 (Spencer and Ianelli 2016, hereafter called the 'McAllister-Ianelli' method). In this method the weights of the input sample sizes used in the 2015 assessment for GOA POP were computed as the harmonic mean of the ratio of effective sample size to the original input sample size and are iteratively estimated until these weights converge. The second method selected within the iterative classification was the method presented in Francis (2011, method TA1.8, hereafter called the 'Francis' method), which has been adopted in several AFSC assessments (e.g., McGilliard et al. 2013). In this method, the weights of the input sample sizes used in the 2015 assessment for GOA POP were computed as the inverse of the variance of the standardized residuals between the means of the observed and predicted ages (or lengths). Within the internal estimation classification, the first method employed the Dirichlet distribution and estimated the mean input sample size (across years) as a parameter within the assessment model (Maunder 2011, Hulson et al. 2012, hereafter called the 'Dirichlet' method). The second method used the Dirichlet-Multinomial compound distribution and through a linear relationship with observed sample size, estimated a parameter that related the observed sample sizes to the input sample sizes (Thorson et al. 2017, hereafter called the 'Dirichlet-Multinomial' method).

Compared to the 2015 assessment of GOA POP, alternative input sample size estimation methods generally resulted in increased RMSE for the bottom trawl survey biomass, while the RMSE in the bottom trawl survey age composition decreased (Table 3). The RMSE for the fishery age composition was either smaller or the same compared to the 2015 assessment when input sample size was iterated, and was larger than the 2015 assessment when input sample size was estimated (Table 3). Relative to the 2015 assessment, RMSE of the fishery length composition were larger when the iterative methods were used, and either slightly smaller (Dirichlet) or slightly larger (Dirichlet-Multinomial) when sample size was estimated (Table 3).

The percent difference in key parameter estimates compared to the 2015 assessment ranged from -9.1% to 21.3% (Table 4). The largest percent difference occurred with the trawl survey catchability parameter (q) when input sample size was iteratively estimated with the McAllister-Ianelli method, catchability increased by 21.3% compared to the 2015 assessment (Table 4). Generally when sample size was estimated (i.e. Dirichlet and Dirichlet-Multinomial methods), estimates of catchability (q) decreased, natural mortality (M) increased, recruitment variability (σ_r) increased, and mean recruitment ($\ln R$) decreased compared to the 2015 assessment. For the majority of the time series, the Francis, Dirichlet, and Dirichlet-Multinomial methods produced estimates of spawning biomass that were similar to the 2015 assessment (top panels Figure 3). The McAllister-Ianelli method also resulted in spawning biomass estimates that were similar to the 2015 assessment from 1961 to the mid-1990s, after which this method resulted in spawning biomass estimates that were smaller than the 2015 assessment. For each of the input sample size estimation methods the CV in spawning biomass was smaller than the 2015 assessment estimates (middle panels Figure 3). The method that resulted in the smallest CVs in spawning biomass was the McAllister-Ianelli method, although, this method also resulted in the smallest estimates of spawning biomass. The methods that estimate input sample size as a parameter resulted in the next two smallest CVs in spawning biomass across the time series and on average (middle panels Figure 3), and also resulted in spawning biomass estimates that were similar to the 2015 assessment (top panels Figure 3). Trends in estimated recruitment were generally similar, although some differences in the scale of recruitment between the estimation methods tested and the 2015 assessment were observed (bottom panels Figure 3). On average, when the input sample size was iteratively estimated, estimated recruitment decreased, while when input sample size was estimated as a parameter, recruitment increased. In general, compared to the input sample sizes used in the 2015 GOA POP assessment each of the methods investigated resulted in an increase in the input sample size (Tables 5-7). The only method that resulted in a smaller input sample size compared to the 2015 assessment was the Francis method for the fishery length composition data (Table 7). The largest sample size increase compared to the 2015 assessment was for the fishery age composition using the McAllister-Ianelli method (Table 5). Visually, the fits to the survey age compositions were very similar (Figure 4).

Each of the methods investigated to estimate input sample size has strengths and weaknesses. The McAllister-Ianelli and Francis methods have both been employed in assessments at AFSC in the past, which provides the benefit of familiarity and allows for ease of communication when presenting results. However, the McAllister-Ianelli method resulted in a 21.3% increase in bottom trawl survey catchability to 2.37, which is conflicting with the recent results in the assessment of catchability estimates being below 2, which the authors have perceived to be a desirable result. This result could be the consequence of this method weighting the fishery age composition twice as high as the bottom trawl survey age composition. The Francis method seemed to perform well, but the reduction in spawning biomass CV was the smallest compared to all of the alternative methods investigated. A theoretical strength of the methods that estimate input sample size as a parameter is that the uncertainty in this parameter is then propagated through the uncertainty in the assessment through MCMC and Hessian derived uncertainty estimates. Both the Dirichlet and Dirichlet-Multinomial performed well; an advantage of the Dirichlet-Multinomial above the Dirichlet is that it estimates the uncertainty in annual input sample sizes (as opposed to just a mean) and the estimates are bounded by the sample size, whereas the Dirichlet sample size can range from 0 to infinity. An interesting result from these two methods is that the catchability parameter decreased compared to the 2015 assessment, however, there was not an associated increase in the spawning biomass estimates that one would expect. Rather, the estimates of spawning biomass was similar to the 2015 assessment. In terms of author recommendations, the Francis and Dirichlet-Multinomial methods look like promising methods to pursue investigations for the full assessment in 2017. For further investigations of the input sample size methods to be pursued in the full assessment in November, the PT is also requested to provide their input as to which methods they would like to be further investigated.

Category 3: Analysis of fishery selectivity

In the 2015 assessment for GOA POP, fishery selectivity was estimated for three time blocks with the logistic function (asymptotic), the gamma function (dome-shaped), and an average of the two. The three time blocks are from 1961-1976 (logistic/asymptotic fishery selectivity), 1977-1995 (average of logistic and gamma fishery selectivity), and 1996-present (gamma/dome-shaped fishery selectivity). In the original formulation of the POP assessment in ADMB, fishery selectivity was modeled using non-parametric differencing to produce smoothness from age to age and penalized a steeply descending right limb (Heifetz et al. 2001). In the 2009 assessment it was noted that the fit of the time-invariant fishery selectivity to fishery length and age data, as well as the influence on bottom trawl survey catchability, was such that time-dependent fishery selectivity was warranted (Hanselman et al. 2009). Thus, in the 2009 assessment the current convention of 3 time-blocks for fishery selectivity was adopted and has been utilized since.

In recent assessments for GOA POP it has been noted that the catch-weighted fishery depth has been decreasing (e.g., Hulson et al. 2014) and both the PT and SSC have requested that fishing depth as related to estimates of fishery selectivity be investigated. To relate the time-series of average depth fished to fishery selectivity in the GOA POP assessment a covariate approach was used (e.g., Maunder and Watters 2003). The time-series of catch-weighted average depth fished is available from 1987 to present; the data from 1996 to present was used with the gamma function estimate of fishery selectivity within the most recent time-block of fishery selectivity. There are two parameters estimated within the gamma function, these two parameters can be interpreted as (1) the age at which fishery selectivity is maximized, and (2) the slope of the dome-shaped curve (controlling both the slope before and after the age at maximum selectivity). To relate average fishing depth to estimates of fishery selectivity, a linear relationship was estimated between the mean depth and each gamma parameter. Relationships that included age at maximum selectivity were not significantly better than a model that used just the slope parameter, thus, the results for only the slope parameter are shown. Two additional fishery selectivity cases were also tested for comparison with the 2015 assessment. The first additional case utilized the bi-cubic spline function in ADMB as adopted in the 2014 BSAI POP assessment (Spencer and Ianelli 2014) with the same age and year nodes and weightings for penalties. The second additional case removed the time-blocks in the 2015 assessment and used time-invariant gamma selectivity for fishery selectivity.

The RMSEs resulting from use of the time-invariant gamma selectivity for all years was the smallest for the bottom trawl survey biomass index, trawl survey age composition, and fishery size composition compared to the alternative fishery selectivity models tested and the 2015 assessment (Table 8). The smallest RMSE for the fishery age composition resulted from the bi-cubic spline selectivity (Table 8). The percent difference in key parameter estimates compared to the 2015 assessment was, for the majority of parameters, the largest from the bi-cubic spline selectivity (Table 9). The percent difference in key parameter estimates from the 2015 assessment resulting from the gamma selectivity for all years was the opposite of the other selectivity methods tested, and notably, it reduced the bottom trawl survey catchability parameter (q) by 14.3% (Table 9). The estimates of spawning biomass were nearly identical to the 2015 assessment when fishing depth was used as a covariate to fishery selectivity, was estimated to be larger when the gamma selectivity function was used for all years, and smaller when the bi-cubic spline selectivity was used for fishery selectivity (top panels Figure 5). The CV of spawning biomass was, on average, larger than the 2015 assessment from the covariate selectivity, and smaller for both the bi-cubic and time-invariance gamma selectivity (middle panels Figure 5).

The smallest CV for spawning biomass was from using the bi-cubic spline function for selectivity, although, this method also resulted in the smallest estimates of spawning biomass. The most notable differences in recruitment compared to the 2015 assessment occurred when the gamma selectivity was used for all years and, on average, the bi-cubic spline selectivity and the covariate selectivity were nearly the same as the 2015 assessment (bottom panels Figure 5).

The model fit to the fishery age and length compositions (Figures 6 and 7) were very similar across the selectivity methods selected and the 2015 assessment; the most notable difference was that the bi-cubic spline selectivity fit the plus age group in the fishery age composition more precisely in recent years (Figure 6). In terms of the models' likelihood values, with a reduction of 2 parameters using the gamma selectivity for all years compared to the 2015 assessment, the data likelihood (models fit to the data without penalty functions) was smaller than any other method tested (Table 10). The data likelihood resulted in an increase when using the bi-cubic spline selectivity and adding 16 parameters compared to the 2015 assessment, and was nearly the same when adding depth as a covariate to the fishery selectivity (Table 10). In general, adding depth as a covariate to the fishery selectivity did not improve the model fit to data compared to the 2015 assessment. However, further investigation will continue in other areas as the decreasing fishing depth may be an indication of distribution expansion of the GOA POP stock in recent years. Model comparison criteria could be computed for these different models, like DIC, however, it seems clear that the time-invariant gamma selectivity would be preferred. For the full 2017 assessment the time-invariant gamma selectivity will be investigated with updated data, as well as possibly investigating the bi-cubic spline selectivity with difference nodes and weightings than the BSAI POP assessment. The PT's recommendation on which fishery selectivity methods to investigate for the full assessment in November is also requested.

Category 4: Analysis of a spatial temporal alternative to the design based estimates of bottom trawl survey biomass

Currently, the design based estimates of biomass from the AFSC bottom trawl survey from 1984 to 2015 are used as the primary abundance index for GOA POP. The design-based estimates of biomass have been used since the original inception of this assessment in ADMB (Heifetz et al. 2001). Recently, a spatial temporal GLMM based abundance index (VAST, Thorson and Barnett 2017) has been investigated and was used in the last full assessment cycle for GOA Dusky rockfish (Lunsford et al. 2015). The VAST index was also investigated in the 2016 GOA pollock assessment (Dorn et al. 2016). The PT and SSC recommended that a working group investigate the utility of the VAST model based alterative index as well as to investigate its utility in the GOA POP assessment. In this section, a new VAST index was constructed and investigated within the 2015 assessment for POP. Additionally, in recent AFSC assessments for stocks in the GOA the 1984 and 1987 bottom trawl survey estimates have been increasingly omitted from the time series due to uncertainties in standardizing the surveys since Japanese vessels with different gears were used (e.g., Dorn et al 2014). In this section, in addition to investigating a new model-based index, we also evaluated the removal of the 1984 and 1987 bottom trawl survey estimates from both the design-based and VAST indices for comparison with the 2015 assessment.

The bottom trawl survey biomass estimates from the VAST model prior to 2001 are comparable to the design-based estimates; during this same period, the bottom trawl survey uncertainty as estimated by the VAST model is smaller than from the design-based estimates (Figure 8). After 2001 the bottom trawl survey biomass estimated from the VAST model is larger than the design-based, and estimates an increase in POP biomass that is not reflected as significantly in the design-based estimates. Interestingly, the VAST model estimates an increase in bottom trawl survey biomass in 2015 compared to 2013, whereas the design-based estimates a decrease in 2015 compared to 2013. The RMSE to fitted data sets (besides the bottom trawl survey biomass) were largely unaffected by the inclusion of the VAST index or whether the 1984 and 1987 bottom trawl survey data were omitted or included (Table 11, note that the survey age composition RMSE is smaller for the cases without the 1984 and 1987 data because those years were omitted from this dataset). The largest difference in the key parameter estimates when comparing between the design-based and VAST bottom trawl survey biomass was in the catchability parameter (Table 12). Using the VAST index increased the estimate of catchability by 28%, from 1.95 in the 2015 assessment to 2.49 for the full time series of the VAST index and 2.58 when the 1984 and 1987 biomass estimates were omitted. The increase in catchability coincides with an increase in spawning biomass since the 1980s from the model that utilized the VAST index compared to the 2015 assessment (top panels Figure 9). Omitting the 1984 and 1987 surveys from the design-based index increased the estimates of spawning biomass slightly towards the end of the time series compared to the 2015 assessment (top panels Figure 9). Omitting the 1984 and 1987 surveys from both the design-based and VAST bottom trawl survey biomass' resulted in CVs in spawning biomass that were, on average, larger than the 2015 assessment (middle panels Figure 9). On average, the CV in spawning biomass from using the entire time series of the VAST index were similar to the 2015 assessment overall, and smaller at the end of the time series (middle panels Figure 9). Estimated recruitment, on average, from each of the survey biomass alternatives investigated was larger than the 2015 assessment (bottom panels Figure 9). On the whole, the POP assessment model fits the time series of trawl survey biomass well prior to 2013, whether from the design-based or VAST methods, but does not fit the recent increase in trawl survey biomass that have been observed in 2013 and 2015 well for either method, but particularly not for the VAST method (Figure 10).

In general, the VAST survey biomass decreased the uncertainty in historical estimates prior to 2001 relative to design-based estimates, but in the recent time series the uncertainty from both methods has been nearly the same. The increase in estimated biomass in 2013 and 2015 do not seem biologically reasonable from the perspective of a long-lived species model framework, and similar concern was always stated with the large increases in the design-based estimates in the 1990s. An additional noteworthy result was the increase in the estimate of catchability to around 2.5 or larger using the VAST index. These concerns warrant further investigation before this index is implemented into the assessment. The PT is requested to suggest whether they would like to see the assessment results fit to the VAST index with the 2017 survey data in November as part of the assessment evaluation, or if they would rather the authors defer to a future date upon which the working group has made recommendations. We also request the PT to note whether they would like to see an alternative model that removes the 1984 and 1987 bottom trawl survey biomass from the time series.

Categories 1 – 4 Bridging analysis

If a bridging analysis were to be conducted for all of the alternative methods and models investigated in categories 1 – 4 there would be a total of 720 models to evaluate. There were several alternatives identified by the authors for potential investigation in the full 2017 assessment in November that could reduce this number. As a reminder, for each category these included:

1. Setting the length bins at either 1 or 2 cm with a plus length group of ≥39 cm
2. Exploring the Francis iterative method and Dirichlet-Multinomial estimation method of estimating input sample size
3. Evaluate the time-invariant gamma fishery selectivity as an alternative and potentially further explore the nodes and weighting used in the bi-cubic spline selectivity
4. Potentially include the VAST index into the model for comparison with results based on the design-based bottom trawl survey biomass.

Performing a bridging analyses on these identified alternatives would result in 54 models for investigation, ultimately leading to a recommended model for the full assessment in November 2017. However, prior to implementing and digesting a full bridging analysis the PT is also requested to identify alternatives that they would like to see investigated in the full assessment in November.

References

Courtney, D. L., J. Heifetz, M. F. Sigler, and D. M. Clausen. 1999. An age structured model of northern rockfish, *Sebastes polyspinis*, recruitment and biomass in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2000. Pp. 361-404. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Dorn, M., K. Aydin, D. Jones, W. Palsson, and K. Spalinger. 2014. Assessment of the walleye pollock stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2015. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Dorn, M., K. Aydin, B. Fissel, D. Jones, W. Palsson, K. Spalinger, and S. Stienessen. 2016. Assessment of the walleye pollock stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2017. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Francis, R. I. C. C., 2011. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci. 68:1124-1138.

Hanselman, D. H., J. Heifetz, J. Fujioka, and J. N. Ianelli. 2003. Gulf of Alaska Pacific ocean perch. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2004. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Hanselman, D. H., S. K. Shotwell, J. Heifetz, J. Fujioka, and J. N. Ianelli. 2009. Assessment of Pacific ocean perch in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2010. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Heifetz, J., D. L. Courtney, D. M. Clausen, J. T. Fujioka, and J. N. Ianelli, 2001. Slope rockfish. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2002. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Hulson, P.-J. F., Hanselman, D. H., and Quinn II, T. J., 2012. Determining effective sample size in integrated age-structured assessment models. ICES J. Mar. Sci. 69(2): 281-292.

Hulson, P.-J. F., D. H. Hanselman, S. K. Shotwell, C. R. Lunsford, and J. N. Ianelli, 2014. Assessment of the Pacific ocean perch stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2015. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Hulson, P.-J. F., D. H. Hanselman, S. K. Shotwell, C. R. Lunsford, and J. N. Ianelli, 2015. Assessment of the Pacific ocean perch stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2016. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Lunsford, C. R., P.-J. F. Hulson, S. K. Shotwell, and D. H. Hanselman, 2015. Assessment of the dusky rockfish stock in the Gulf of Alaska. In: Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2016. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Maunder, M. N., 2011. Review and evaluation of the likelihood functions for composition data in stock-assessment models: estimating the effective sample size. Fish. Res. 109: 311–319.

Maunder, M.N., and Watters, G.M. 2003. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and example. Fish. Bull. 101: 89–99.

Maunder, M. N., P. R. Crone, A. E. Punt, J. L. Valeor, and B. X. Semmens, 2017. Data conflict and weighting, likelihood functions and process error. Fish. Res. 192:1-4.

McAllister, M.K., and Ianelli, J.N. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Can. J. Fish. Aq. Sci. 54(2): 284-300.

McGilliard, C. R., W. Palsson, W. Stockhausen, and J. N. Ianelli. 2013. Assessment of the flathead sole stock in the Gulf of Alaska. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Gulf of Alaska as projected for 2014. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Spencer, P. D., and J. N. Ianelli, 2014. Assessment of the Pacific ocean perch stock in the Bering Sea/Aleutian Islands. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands as projected for 2015. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Spencer, P. D., and J. N. Ianelli, 2016. Assessment of the Pacific ocean perch stock in the Bering Sea/Aleutian Islands. *In* Stock assessment and fishery evaluation report for the groundfish resources of the Bering Sea/Aleutian Islands as projected for 2017. North Pacific Fishery Management Council, 605 W 4th Ave, Suite 306 Anchorage, AK 99501.

Thorson, J. T., K. F. Johnson, R. D. Methot, and I. G. Taylor, 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-Multinomial distribution. Fish. Res. 192:84-93.

Thorson, J. T., and L. A. K. Barnett. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science 75:1311-1321.

Tables

Table 1. RMSE of the bottom trawl survey biomass index, fishery age composition, and survey age composition from the 2015 assessment and the alternative bins and plus length groups investigated for the fishery length composition.

		Trawl survey biomass	Fishery age composition	Survey age composition
Base 2015		0.31781	0.01618	0.02004
≥39	1 cm bin	0.31737	0.01621	0.01998
≥39	2 cm bin	0.31735	0.01619	0.01997
≥39	3 cm bin	0.31751	0.01617	0.02000
≥39	4 cm bin	0.31693	0.01617	0.01994
≥45	1 cm bin	0.31684	0.01617	0.01995
≥45	2 cm bin	0.31688	0.01616	0.01995
≥45	3 cm bin	0.31698	0.01615	0.01995
≥45	4 cm bin	0.31709	0.01615	0.01993

Table 2. Percent difference in key parameter estimates across the alternative bins and plus length groups investigated for the fishery length composition compared to the 2015 assessment (Base 2015). Parameter estimates from the 2015 assessment are shown.

		$F_{40\%}$	q	M	σ_r	ln$R$
Base 2015		0.102	1.95	0.061	0.877	3.97
≥39	1 cm bin	-0.3%	1.0%	-0.2%	0.1%	-0.1%
≥39	2 cm bin	-0.5%	1.5%	-0.5%	0.2%	-0.3%
≥39	3 cm bin	-0.7%	0.9%	-0.6%	0.0%	-0.2%
≥39	4 cm bin	-1.1%	3.2%	-1.0%	0.7%	-0.8%
≥45	1 cm bin	0.6%	3.4%	1.7%	1.4%	-0.7%
≥45	2 cm bin	0.3%	3.7%	1.4%	1.4%	-0.8%
≥45	3 cm bin	0.0%	2.7%	1.0%	1.0%	-0.6%
≥45	4 cm bin	-0.7%	3.5%	0.3%	1.2%	-1.0%

Table 3. RMSE of the bottom trawl survey biomass index, fishery age composition, and survey age composition from the 2015 assessment and the alternative methods selected to evaluate input sample size for composition data.

		Trawl survey biomass	Fishery age composition	Survey age composition	Fishery size composition
Base 2015		0.3178	0.0162	0.0200	0.0254
Iterated	McAllister-Ianelli	0.3224	0.0150	0.0190	0.0265
	Francis	0.3213	0.0162	0.0184	0.0297
Estimated	Dirichlet	0.3248	0.0176	0.0198	0.0250
	Dirichlet-Multinomial	0.3415	0.0172	0.0199	0.0260

Table 4. Percent difference in key parameter estimates across the alternative methods selected to evaluate input sample size for composition data compared to the 2015 assessment (Base 2015). Parameter estimates from the 2015 assessment are shown.

		$F_{40\%}$	q	M	σ_r	ln$R$
Base 2015		0.102	1.954	0.061	0.877	3.965
Iterated	McAllister-Ianelli	-4.4%	21.3%	-9.1%	12.9%	-7.7%
	Francis	-2.7%	-1.5%	-7.3%	3.9%	0.2%
Estimated	Dirichlet	0.6%	-7.5%	18.5%	11.7%	-0.6%
	Dirichlet-Multinomial	-2.3%	-11.3%	15.1%	11.9%	-2.5%

Table 5. Input sample sizes for the fishery age composition from the 2015 GOA POP assessment (Base 2015) and the alternative methods selected to estimate input sample size. The bottom row is the ratio of the average input sample size for each alternative method compared to the 2015 assessment.

	Base 2015	McAllister-Ianelli	Francis	Dirichlet	Dirichlet-Multinomial
1990	24	171	42	117	150
1998	23	164	40	117	133
1999	19	135	33	117	98
2000	27	192	47	117	190
2001	23	164	40	117	135
2002	19	135	33	117	96
2004	28	200	49	117	208
2005	27	192	47	117	188
2006	27	192	47	117	190
2008	25	178	44	117	158
2010	25	178	44	117	163
2012	32	228	56	117	265
Ratio	1.00	7.13	1.75	4.70	6.60

Table 6. Input sample sizes for the bottom trawl survey age composition from the 2015 GOA POP assessment (Base 2015) and the alternative methods selected to estimate input sample size. The bottom row is the ratio of the average input sample size for each alternative method compared to the 2015 assessment.

	Base 2015	McAllister-Ianelli	Francis	Dirichlet	Dirichlet-Multinomial
1984	38	131	98	132	172
1987	43	148	111	132	219
1990	42	145	108	132	211
1993	37	128	95	132	166
1996	25	86	64	132	78
1999	30	104	77	132	108
2003	31	107	80	132	119
2005	32	110	82	132	122
2007	34	117	88	132	142
2009	20	69	51	132	51
2011	28	97	72	132	96
2013	30	104	77	132	106
Ratio	1.00	3.45	2.57	4.06	4.08

Table 7. Input sample sizes for the fishery size composition from the 2015 GOA POP assessment (Base 2015) and the alternative methods selected to estimate input sample size. The bottom row is the ratio of the average input sample size for each alternative method compared to the 2015 assessment.

	Base 2015	McAllister-Ianelli	Francis	Dirichlet	Dirichlet-Multinomial
1963	27	34	13	149	53
1964	39	49	19	149	110
1965	76	95	37	149	417
1966	100	125	49	149	726
1967	84	105	41	149	518
1968	81	101	40	149	481
1969	86	108	42	149	532
1970	73	91	36	149	391
1971	77	96	38	149	432
1972	83	104	41	149	499
1973	24	30	12	149	43
1974	43	54	21	149	132
1975	42	53	21	149	131
1976	41	51	20	149	124
1977	35	44	17	149	88
1991	29	36	14	149	62
1992	28	35	14	149	56
1995	20	25	10	149	30
1996	26	33	13	149	51
1997	30	38	15	149	66
Ratio	1.00	1.25	0.49	2.85	4.73

Table 8. RMSE of the bottom trawl survey biomass index, fishery age composition, survey age composition, and fishery length composition from the 2015 assessment and the alternative methods selected for fishery selectivity.

	Trawl survey biomass	Fishery age composition	Survey age composition	Fishery length composition
Base 2015	0.3178	0.0162	0.0200	0.0254
Depth as covariate	0.3221	0.0157	0.0200	0.0252
Bi-Cubic spline	0.3195	0.0155	0.0201	0.0244
Time-invariant gamma	0.3174	0.0170	0.0198	0.0236

Table 9. Percent difference in key parameter estimates across the alternative methods selected for fishery selectivity compared to the 2015 assessment (Base 2015). Parameter estimates from the 2015 assessment are shown.

	$F_{40\%}$	q	M	σ_r	$\ln R$
Base 2015	0.102	1.954	0.061	0.877	3.965
Depth as covariate	-8.9%	10.1%	-2.2%	0.8%	-1.4%
Bi-Cubic spline	-24.9%	34.1%	-2.1%	6.4%	-2.7%
Time-invariant gamma	4.3%	-14.3%	0.7%	-5.0%	6.3%

Table 10. Comparison of the number of parameters, data and total likelihood across the alternative methods selected for fishery selectivity compared to the 2015 assessment (Base 2015).

	Number of parameters	Data Likelihood	Total Likelihood
Base 2015	152	117.96	256.29
Depth as covariate	153	117.48	255.47
Bi-Cubic spline	168	119.64	274.71
Time-invariant gamma	150	110.17	247.11

Table 11. RMSE of the fishery age composition, survey age composition, and fishery length composition from the 2015 assessment, the 2015 assessment with 1980s survey data omitted, the assessment with the VAST survey biomass, and the assessment with the VAST survey biomass with 1980s survey data omitted. Note that the RMSE for the survey age composition decreases in the cases when 1980s data is omitted because of the omission of data, but is still comparable between the design-based and VAST survey biomass for the same time-series.

	Fishery age composition	Survey age composition	Fishery length composition
Base 2015	0.0162	0.0200	0.0254
Base 2015 w/o 80s	0.0165	0.0150	0.0252
VAST	0.0160	0.0205	0.0261
VAST w/o 80s	0.0164	0.0153	0.0256

Table 12. Percent difference in key parameter estimates across the alternative bottom trawl survey biomass methods selected compared to the 2015 assessment (Base 2015). Parameter estimates from the 2015 assessment are shown.

	$F_{40\%}$	q	M	σ_r	$\ln R$
Base 2015	0.102	1.954	0.061	0.877	3.965
Base 2015 w/o 80s	0.86%	0.02%	3.35%	-5.08%	2.26%
VAST	1.95%	27.46%	3.98%	0.65%	4.00%
VAST w/o 80s	1.54%	31.89%	4.39%	-5.02%	4.14%

Figures

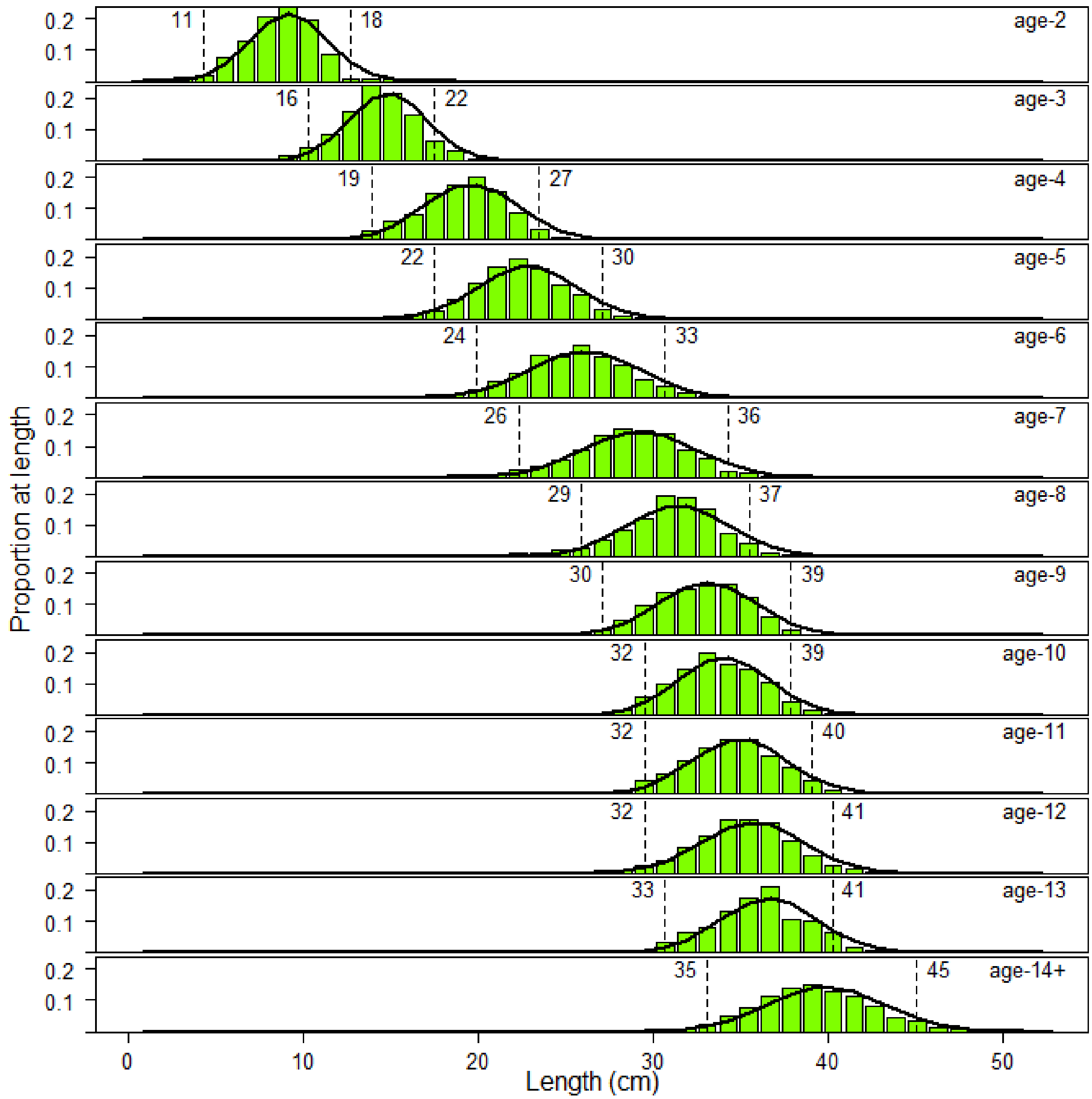

Figure 1. Observed (bars) and estimated (lines) proportion at length for age-2 to age-14+ from the bottom trawl survey (data pooled from 1984-2013). Vertical dashed lines with text represent the 95% confidence intervals of length for each age (rounded to the nearest cm).

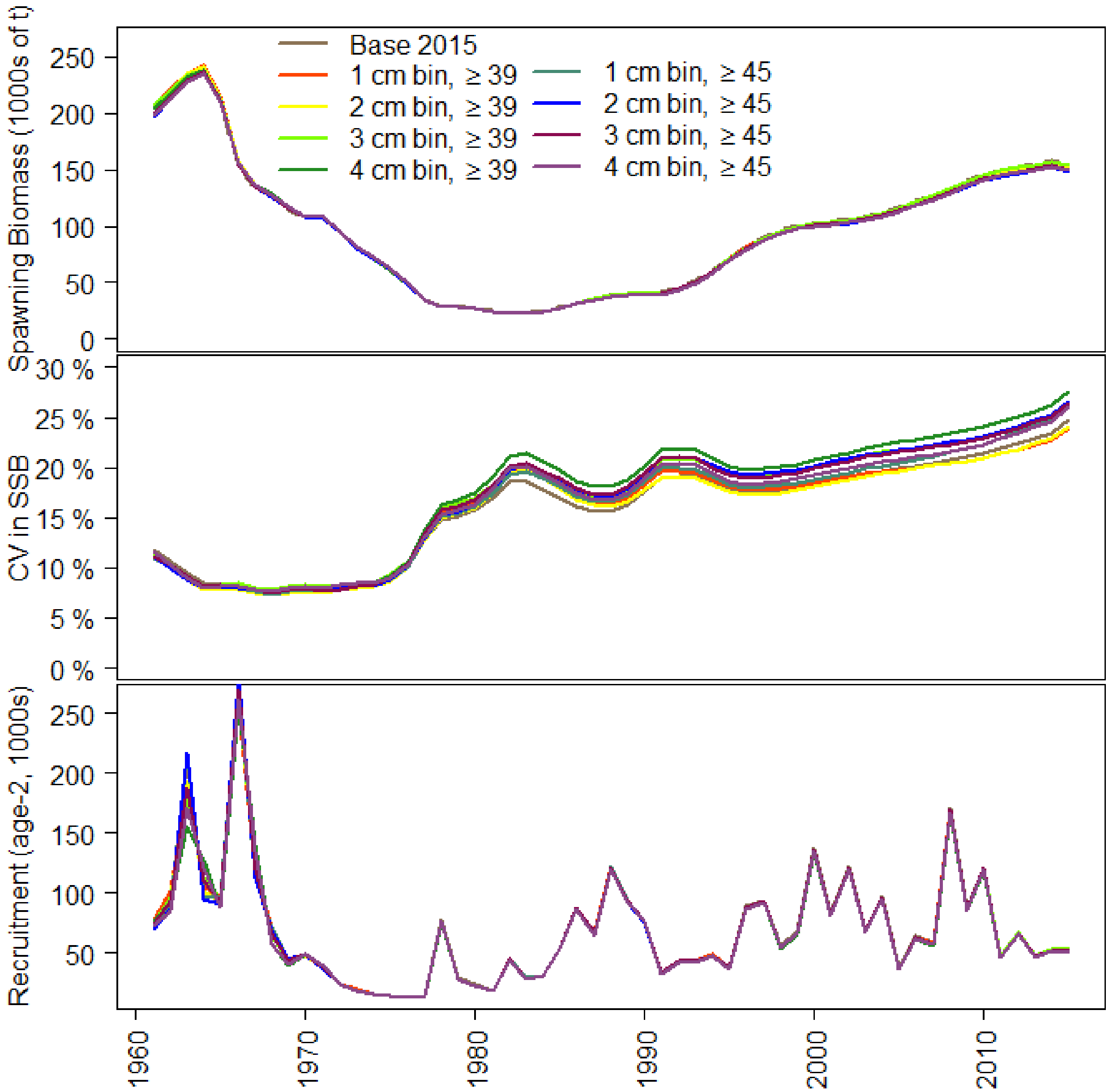

Figure 2. Estimated spawning biomass (top panel), CV of spawning biomass (middle panel) and estimated recruitment (bottom panel) across the alternative bins and plus length groups investigated for the fishery length composition in comparison with the 2015 assessment (Base 2015).

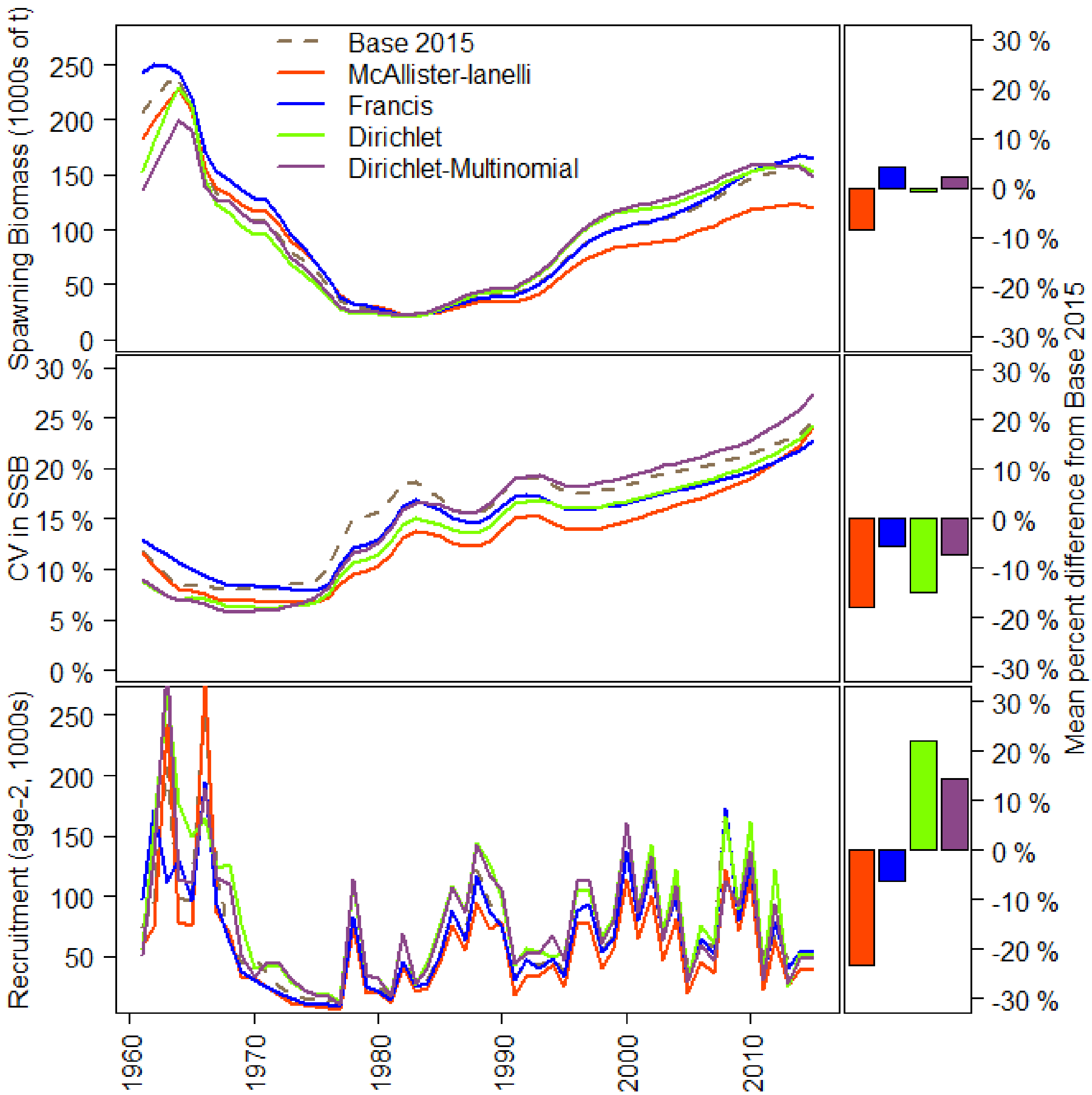

Figure 3. Estimated spawning biomass (top panels), CV of spawning biomass (middle panels) and estimated recruitment (bottom panels) with mean percent difference compared to the 2015 assessment (right panels) across the alternative methods selected to evaluate input sample size for composition data.

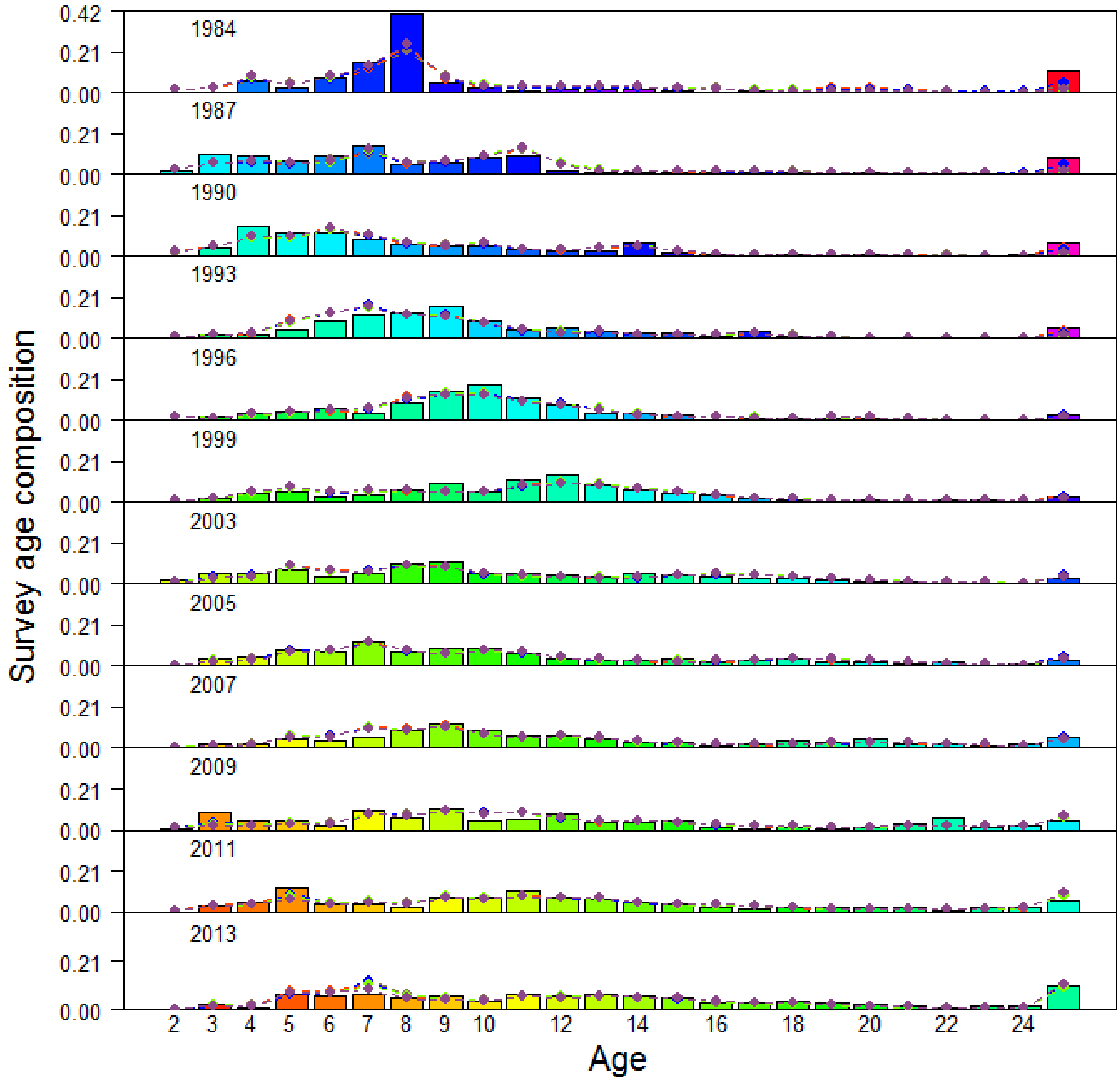

Figure 4. Fit of the 2015 assessment and alternative methods selected to evaluate input sample size for composition data to the observed bottom trawl survey age composition (using the same colors for each model as in the legend of Figure 3).

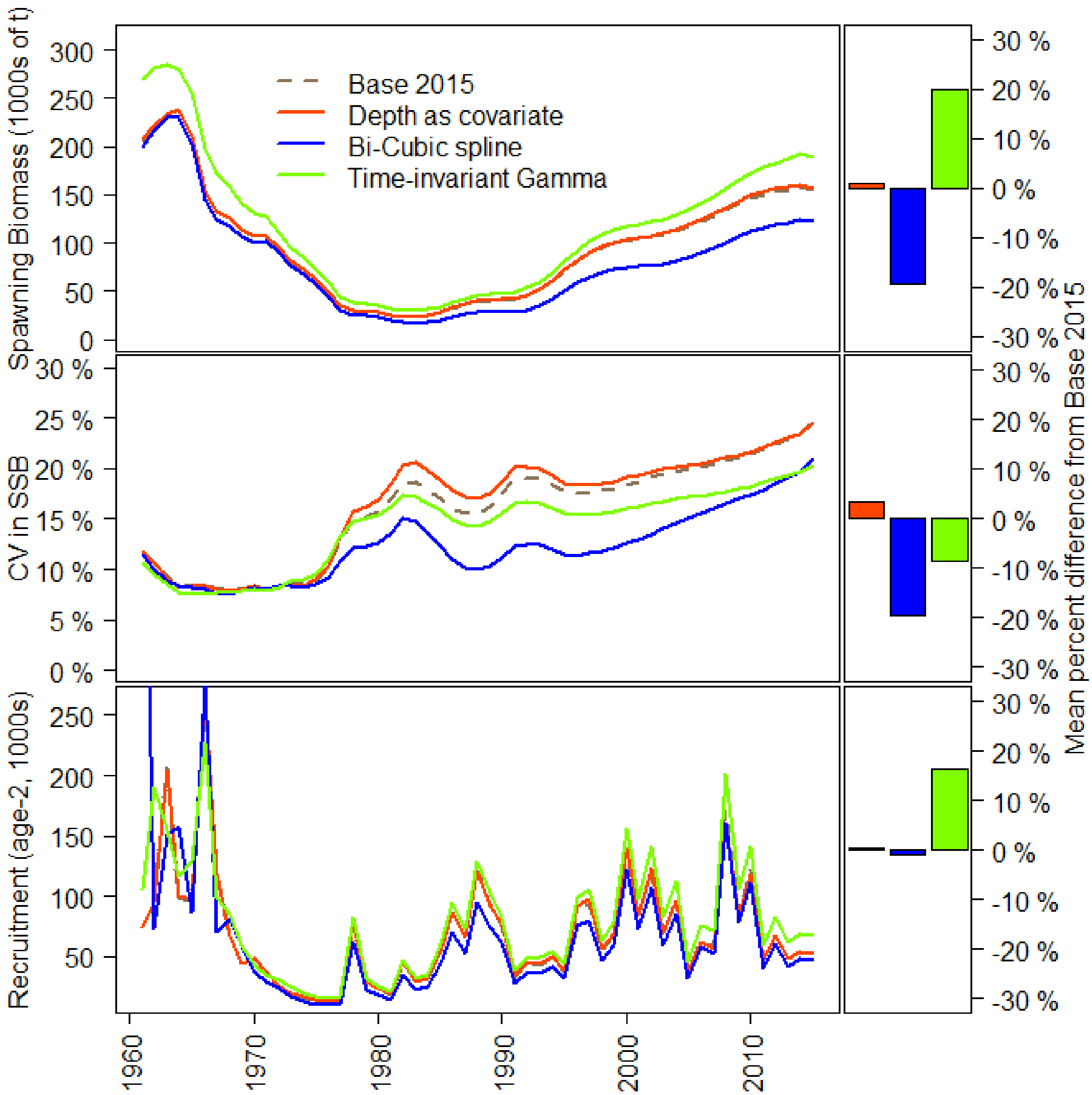

Figure 5. Estimated spawning biomass (top panels), CV of spawning biomass (middle panels) and estimated recruitment (bottom panels) with mean percent difference compared to the 2015 assessment (right panels) across the alternative methods selected to evaluate fishery selectivity.

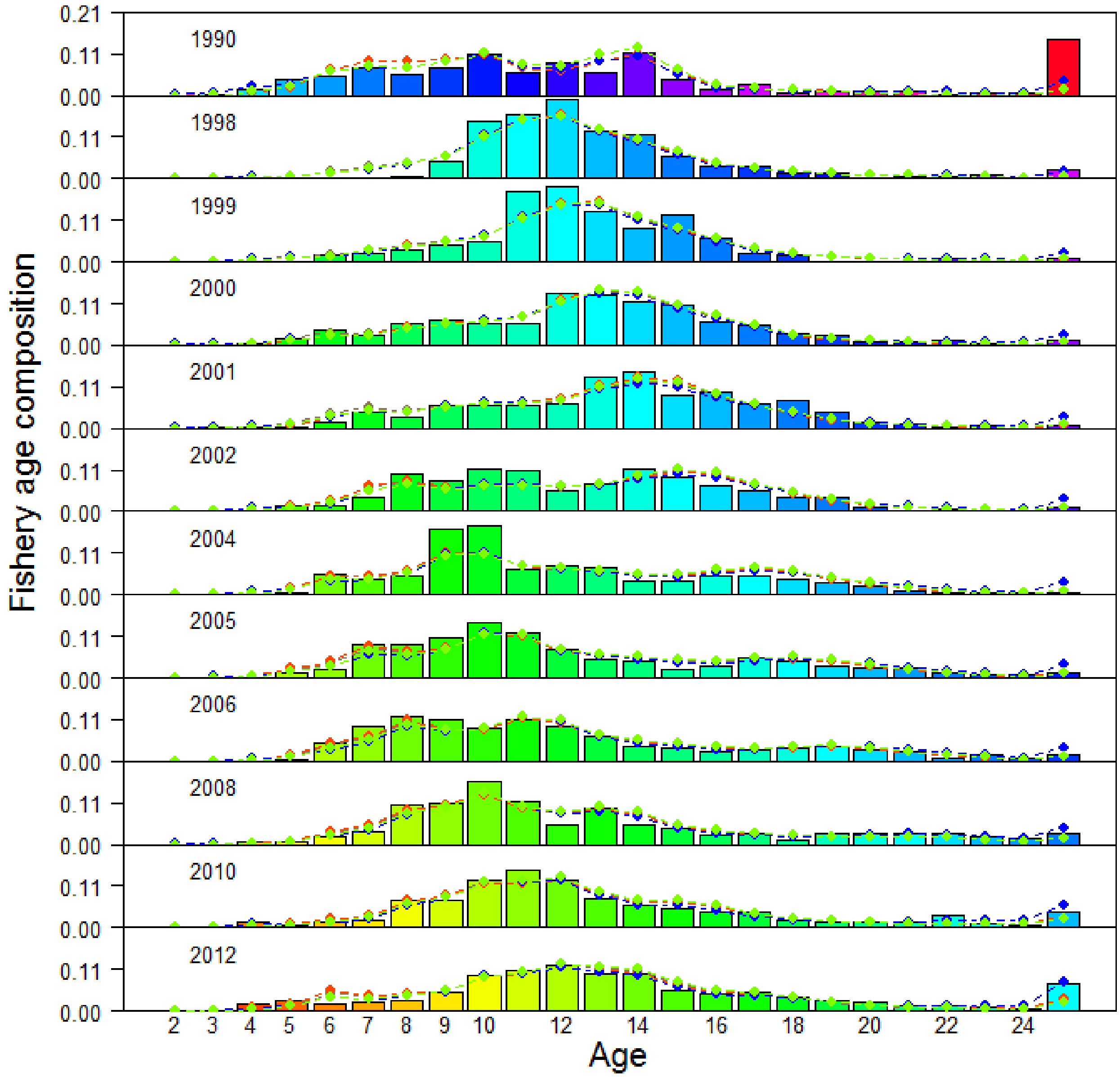

Figure 6. Fit of the 2015 assessment and alternative methods selected for fishery selectivity to the observed fishery age composition (using the same colors for each model as in the legend of Figure 5).

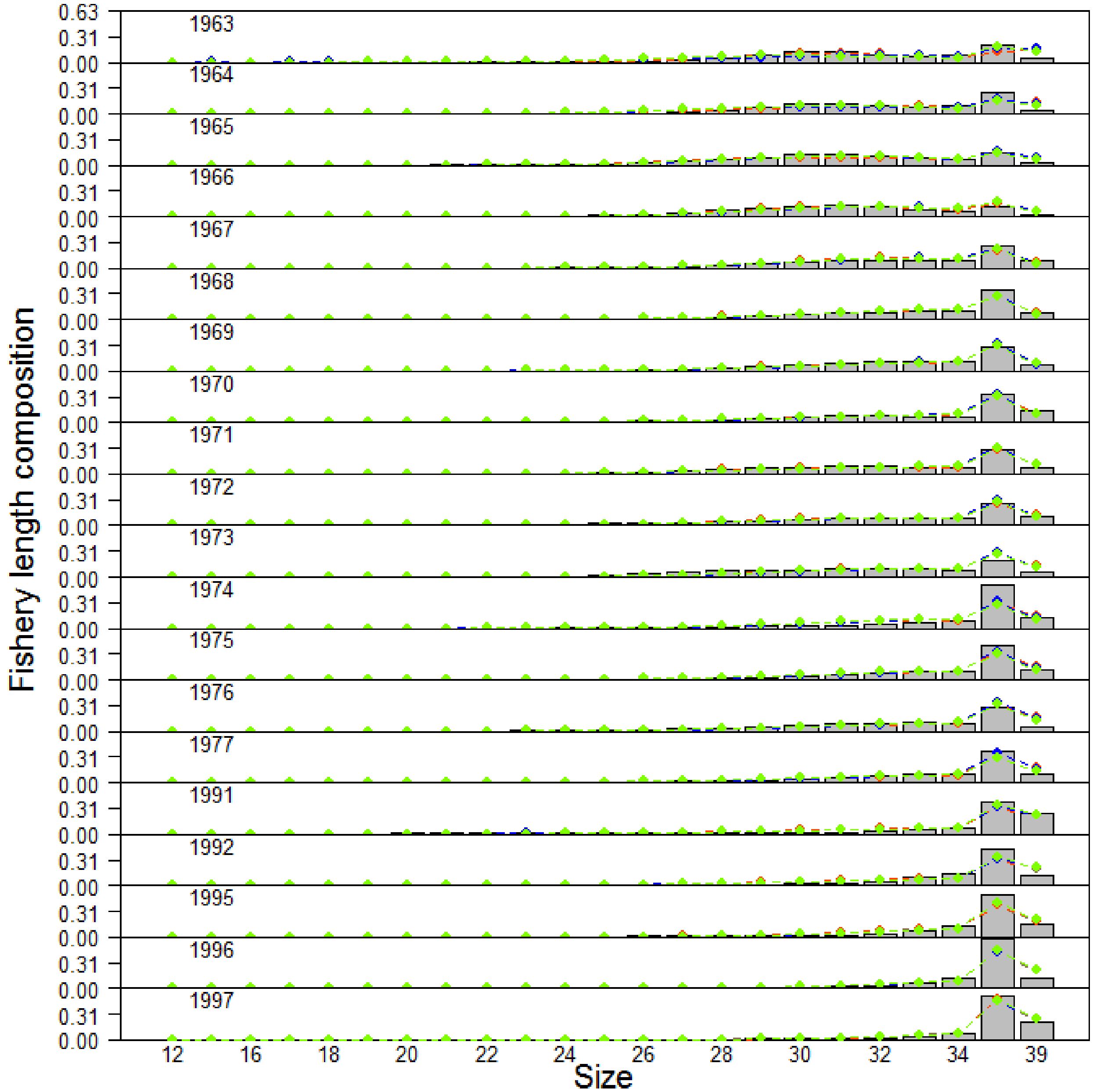

Figure 7. Fit of the 2015 assessment and alternative methods selected for fishery selectivity to the observed fishery size composition (using the same colors for each model as in the legend of Figure 5).

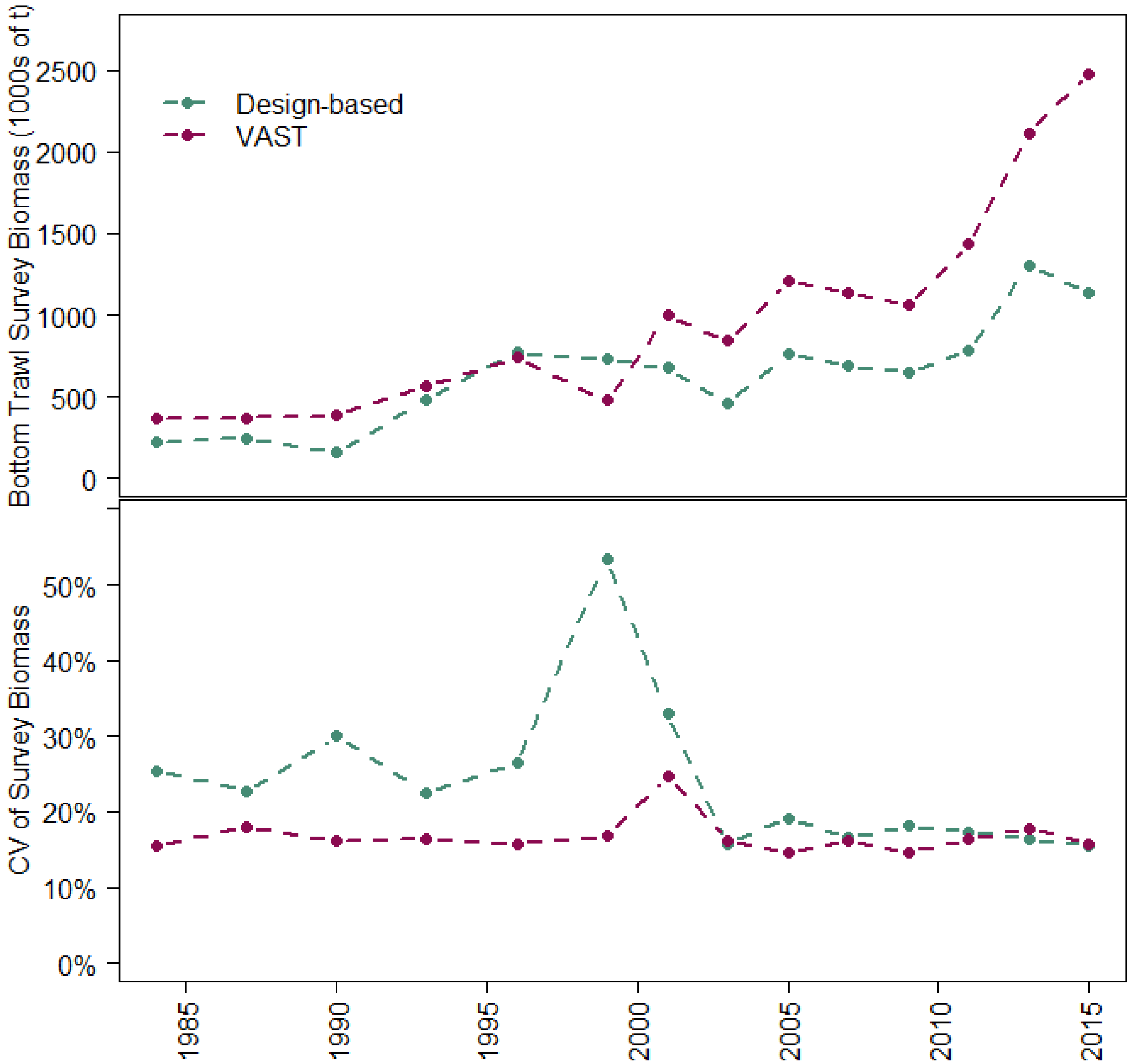

Figure 8. Bottom trawl survey biomass estimates (top panel) with associated CV (bottom panel) from the design-based method and VAST.

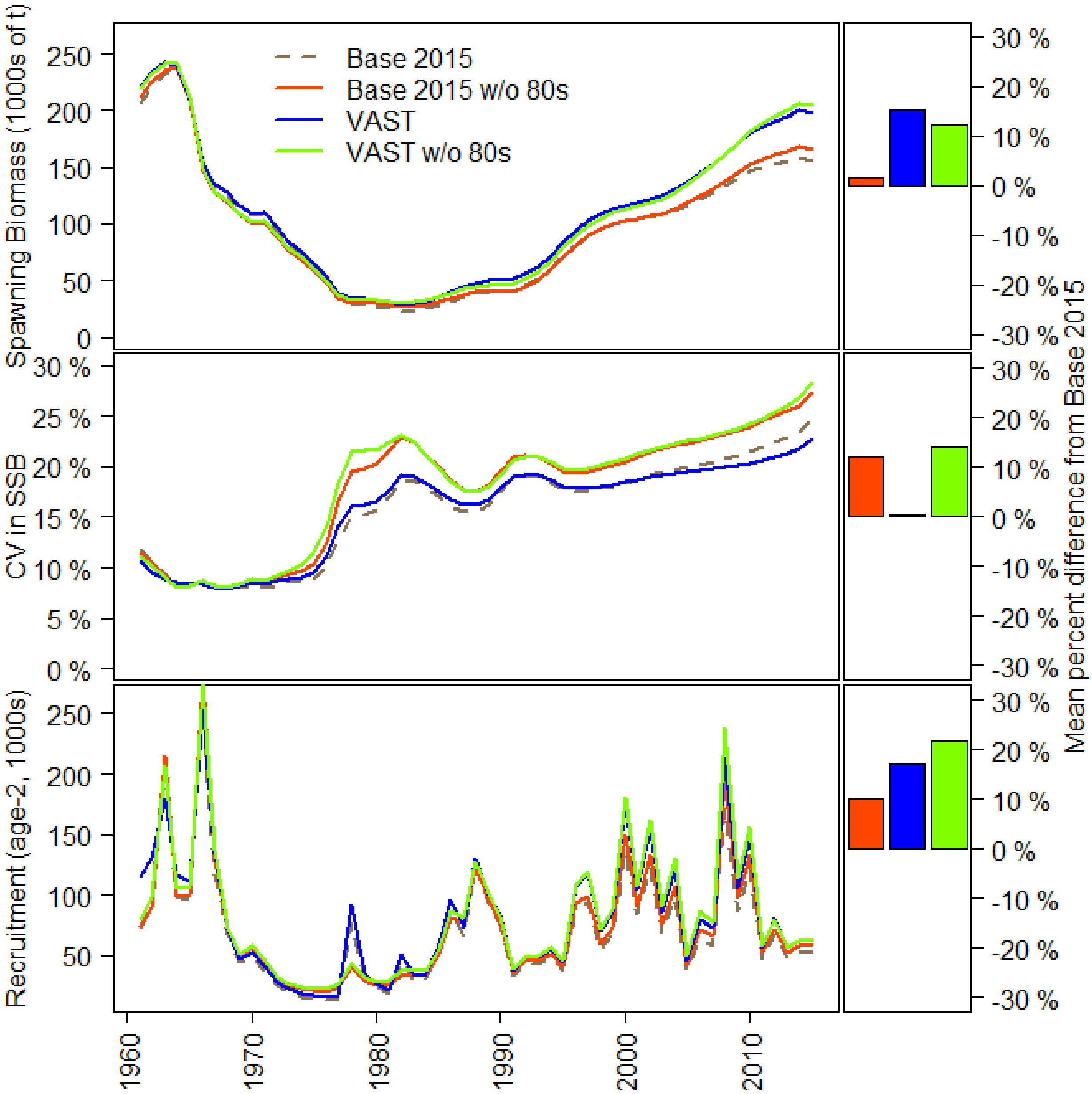

Figure 9. Estimated spawning biomass (top panels), CV of spawning biomass (middle panels) and estimated recruitment (bottom panels) with mean percent difference compared to the 2015 assessment (right panels) across models that incorporate the VAST estimates of bottom trawl survey biomass and/or omit the 1980s cooperative survey data.

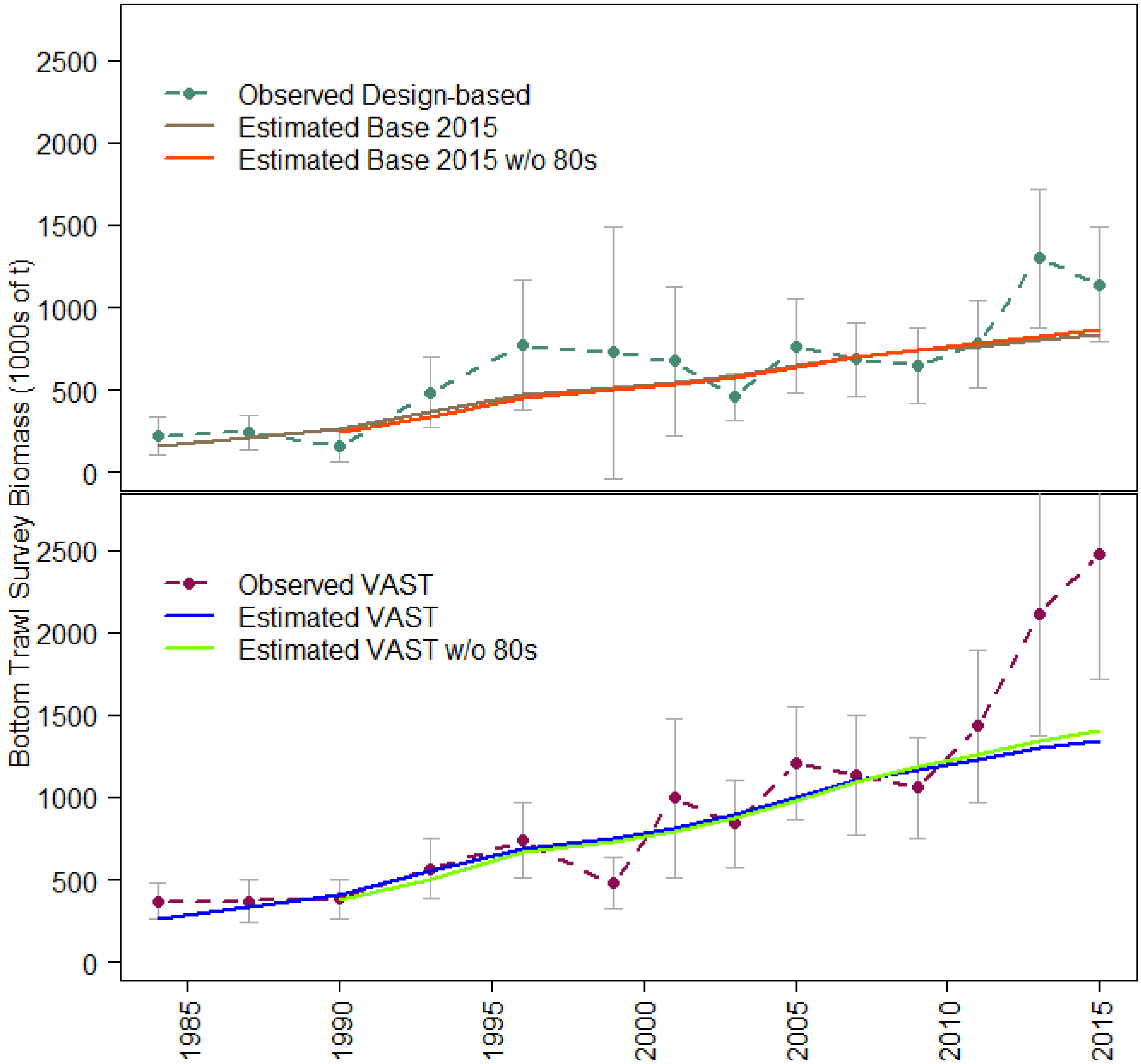

Figure 10. Fit to observed design-based (top panel) and VAST (bottom panel) bottom trawl survey biomass by POP assessment model including or not include the 1980s survey data.