**SUPPLEMENTARY MATERIAL**

### BASA description

BASA uses 16 data types (Table S1), including weight-at-age, fishery-dependent and -independent age compositions, a milt index, two biomass indices from hydroacoustic surveys, catch-at-age from three fisheries, fecundity-at-age, an egg deposition index, and a time series for an ecological covariate. The model starts in 1980 and is projected forward to 2017, although many series have years or sets of years of missing data (Table S1). Full descriptions of the various data types are found in Muradian et al. (2017).

The BASA model is a statistical catch-at-age model that assumes a closed and fully mixed population. There are 10 age classes including a plus group (ages 0-9+). Annual losses in biomass result from fishing or natural mortality, and increases from recruitment and somatic growth (i.e. changes in empirical weight-at-age), all of which are incorporated into equations shown in Table S2. The model year starts with the beginning of spawning in the spring. Catches are treated as instantaneous mortality events in both the spring and fall. Most fishery independent survey data (two hydroacoustic indices, one age-composition dataset, and a milt index) were collected before or during spawning, while egg deposition surveys were conducted post spawning (Muradian et al. 2017).

No stock-recruitment relationship is used to predict recruitment from spawning biomass, and recruitment is directly estimated by BASA. Recruitment estimates are informed by age composition data (i.e. changes in cohort strength from year to year) and surveyed biomass (i.e. how biomass changes in relation to cohort strength), along with assumptions about constant maturity, selectivity, and natural mortality.

Parameters were estimated for initial numbers-at-age, annual recruitment numbers or log deviates with an estimated standard deviation, proportions mature, gear selectivity, survey index scalars and standard errors, additional mortality in 1992-1993, and covariate effects (Table S3). Most parameters had Uniform priors with sufficiently broad priors that avoided boundary issues during model fitting. Prior bounds are shown in Table S3.

To fit the data, likelihood equations were formulated to relate model estimates with the corresponding observations from the fishery-dependent and -independent age compositions, egg deposition survey, two hydroacoustic biomass indices, and mile-days of milt index (Table S4). The effective sample sizes in the Multinomial likelihoods for the age composition data were calculated prior to running MCMC using the iterative reweighting procedure described by Muradian et al. (2017) and based on Stewart and Hamel (2014). The likelihood equations are otherwise nearly identical to those used in Muradian et al. (2017) with the addition of a log-normal likelihood equations for the recruitment deviates (only in the BASA models where the covariates affect recruitment) and latent variables.

### Other modifications to BASA

We modified two other components of BASA to correct for potential model misspecification in Muradian et al. (2017). The first was changing hydroacoustic survey biomass to represent pre-fishery mature biomass instead of age 3+ biomass, as follows:

in which is the acoustic estimate in year *y*, *q* is a scalar for the acoustic estimate (log-link), proportions mature-at-age, numbers-at-age, and weight-at-age. Muradian et al. (2017) previously assumed hydroacoustic surveys captured total adult biomass (all age 3+), although recent data from sampling of the aggregations targeted by acoustic surveys revealed they are mostly mature fish (unpublished data W. Pegau). The second modification was to estimate proportions mature at ages three and four over the entire time period (1980-2017) instead of estimating two sets of proportions for two time periods split at 1997 (Muradian et al., 2017). Sensitivity analysis shown negligible difference in biomass and recruitment estimates, while estimating a maturity for a single time period improved model convergence.

**Equations for model selection criteria**

DIC (Spiegelhalter et al., 2002) is calculated using the following equation:

The first term contains the log-likelihood of the data given the parameter vector () from each posterior draw . The second term is a penalty expressed as the difference between two alternative realizations of deviance and is analogous to the effectivenumber of parameters. denotes the expectation of the quantity within the parentheses which is taken as the mean (for all other criteria presented here as well). To compute DIC, we use the conditional likelihood (i.e. random or process error parameters, such as recruitment, are sampled along with parameters during MCMC) and exclude all prior densities. The best model minimizes DIC and the difference in DIC between each model and the best model (ΔDIC) can be interpreted in the same way as ΔAIC (ΔDIC <2 deserves serious consideration, 2<ΔDIC<7 suggests potential support, ΔDIC>7 is unlikely).

WAIC (Watanabe, 2013) incorporates the posterior distributions of the observations and sums across = 6 data sets fit within BASA as follows:

in which is the number of observations within each data set. The first term is the ln point-wise predictive score which is the logged expectation of posterior point densities of the data (see Figure 1 for calculation steps). The second term is a correction factor for the biased estimate of the true posterior predictive score and follows the form recommended by Gelman et al. (2014b) the sum of the variances of the natural logarithm of the posterior densities for each data point. The best performing model minimizes WAIC.

PPL (Gelfand & Ghosh, 1998) is the sum of a loss and risk function using the posterior predictive distributions of the data:

The first term is loss, which is the squared error between the observed data and the expectation of the posterior predictive distribution (i.e. the posterior predictive mean). The second term is risk, which is the variance of the posterior predictive distributions that acts as a penalty for model complexity (i.e. more parameters induce less precision). As with WAIC, sums over = 6 within BASA to provide a single PPL for each model (see Figure 1). To normalize the differences in scale across data sets (e.g. egg deposition counts in trillions versus hydroacoustic survey data in thousands of tons) while maintaining differences in variance in their observations over time, we define a relative loss and risk function where each term is divided by the squared mean of each observed data set . For the age-composition data, we multiply the observations and posterior predictions of the proportions at each age by the sample size in year for and respectively. In other words, age composition values are converted to numbers at age in each year’s sample and can have decimals. The mean of the observed annual sample sizes across all years is then . The best performing models should minimize both the loss and risk functions across the sum of all data sets, with the minimum leading to the selected best model.

The equation for PSIS-LOO (Vehtari et al., 2017) is the sum of the expected negative log pointwise predictive densities for all observations (, which is the notation used by Vehtari, Gelman, & Gabry, 2017):

where are weights calculated for each individual observation. These weights are derived from a smoothing procedure that uses fits of the generalized Pareto distribution applied to the upper tail (>80th percentile) of raw importance ratios that are calculated from the posterior densities. This procedure is detailed in Fig. 1. Generally, when the difference in between models is less than four, than predictive performance is similar; however, performances can be similar even when differences are greater than four if the standard error of is large. Estimates of and their standard errors are computed for each model using the *loo* function from the ‘loo’ package in R (Vehtari et al., 2020).

Several outputs from *loo* should be considered in the interpretation of PSIS-LOO. First, shape parameters of the generalized Pareto distributions, , are estimated for each posterior predictive distribution (each observation) within the model; in other words, there are estimates. For to have moderate to high accuracy, all (or at least most) points should have . Many instances of may indicate poorly fit outliers, model misspecification or generally flexible models. Another diagnostic is the estimated effective number of parameters () which can be compared to the actual number of estimated parameters () to indicate the predictive ability of an individual model; should approximate or be slightly less than . If not, this may indicate model misspecification, weak priors, or overdispersed data (e.g. when , but especially ). For further details on interpreting PSIS-LOO and cross-validation more generally, we recommend readers refer to Vehtari et al. (2017), the ‘loo’ R package documentation (Vehtari et al., 2020), and their accompanying github page (<https://avehtari.github.io/modelselection/>).

**REFERENCES**

Gelfand, A. E., and Ghosh, S. K. 1998. Model choice: a minimum posterior predictive loss approach. Biometrika, 85: 1-11.

Gelman, A., Hwang, J., and Vehtari, A. 2014b. Understanding predictive information criteria for Bayesian models. Statistics computing, 24: 997-1016.

Muradian, M. L., Branch, T. A., Moffitt, S. D., and Hulson, P.-J. F. 2017. Bayesian stock assessment of Pacific herring in Prince William Sound, Alaska. PloS one, 12: e0172153.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B, 64: 583-639.

Stewart, I. J., and Hamel, O. S. 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. Canadian Journal of Fisheries and Aquatic Sciences, 71: 581-588.

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., and Gelman, A. 2020. loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models.

Vehtari, A., Gelman, A., and Gabry, J. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing, 27: 1413-1432.

Watanabe, S. 2013. A widely applicable Bayesian information criterion. Journal of Machine Learning Research, 14: 867-897.

Table S1. Time series used in the Bayesian ASA model. The first column lists the data type, the second column lists the units, ny refers to number of years that data type was collected, and the final column reports the first and last year of collection. Note some series are discontinuous. Red text denotes updated data from Muradian et al (2017).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data type** | **Units** | **Symbol** | **ny** | **Years** |
| Gillnet catch-at-age | millions |  | 15 | (1980, 1998) |
| Pound utilization catch-at-age | millions |  | 16 | (1980, 1999) |
| Food/bait catch-at-age | millions |  | 17 | (1980, 1998) |
| Fecundity-at-age | no. of eggs per female |  | 7 | (1984, 1993) |
| Weight-at-age of spawning herring | mt/million fish |  | 38 | (1980, 2017) |
| Purse-seine age-composition | proportion |  | 13 | (1980, 1998) |
| Spawner survey age-composition | proportion |  | 36 | (1982, 2017) |
| Female spawners | proportion |  | 38 | (1980, 2017) |
| Total annual purse-seine yield | mt |  | 13 | (1980, 1998) |
| Eggs deposited | trillions |  | 10 | (1984, 1997) |
| C.V. for eggs deposited |  |  | 10 | (1984, 1997) |
| ADF&G hydroacoustic survey biomass | mt |  | 5 | (2005, 2009) |
| PWSSC hydroacoustic survey biomass | mt |  | 24 | (1993, 2017) |
| C.V. for PWSSC hydroacoustic biomass |  |  | 24 | (1993, 2017) |
| Milt | mile/day |  | 38 | (1980, 2017) |
| Ecological covariate | normalized |  | varies | (1980, 2017) |

Table S2. Model formulation. First column gives a description, and the second column gives the mathematical form of the dynamics. Red text denotes the equations we modified from Muradian et al (2017) for this paper.

|  |  |
| --- | --- |
| **Description** | **Equation** |
| **Catch, millions of fish** | |
| Estimated total purse-seine catch |  |
| Spring removals, |  |
| **Survival with covariate as fixed effect, rate** | |
| Half-year survival, ages 0-8, half-year *b=1,2* |  |
| Half-year survival in 1992-1993, ages 3-4 |  |
| Half-year survival in 1992-1993, ages 5-8 |  |
| Half-year survival, 1980, plus group |  |
| Half-year survival, 1981–2017, plus group |  |
| **Survival with covariate as latent variable, rate** | |
| Half-year survival, ages 0-8, half-year *b=1,2* |  |
| **Recruitment with covariate as fixed effect, millions of fish** | |
| Annual recruitment, age 0 |  |
| **Recruitment with covariate as latent variable, millions of fish** | |
| Annual recruitment, age 0 |  |
| **Selectivity, logistic form** |  |
| Purse-seine gear selectivity by age |  |
| **Abundance, millions of fish** |  |
| Pre-fishery total abundance, ages 4–8 |  |
| Pre-fishery total abundance, ages 9+ | + |
| Post-fishery spawning abundance |  |
| **Biomass, mt** |  |
| Pre-fishery total biomass |  |
| Pre-fishery spawning biomass |  |
| Post-fishery spawning biomass |  |
| **Estimates used in the likelihood expressions** |  |
| Estimated ADF&G hydro-acoustic biomass, mt |  |
| Estimated PWSSC hydro-acoustic biomass, mt |  |
| Estimated purse-seine age composition |  |
| Estimated spawning age composition |  |
| Estimated naturally spawned eggs, trillions |  |
| Estimated milt, mile-days |  |

Table S3. Key model parameters and their priors as presented in the main text (excluding sensitivities for certain parameters in Supplementary Material). All mortality is modeled as instantaneous mortality rates. Red text denotes new or modified parameters from Muradian et al (2017).

|  |  |  |
| --- | --- | --- |
| **Parameters** | **Symbols** | **Prior** |
| Background mortality, ages 0–8 |  | Not estimated (0.25) |
| Background mortality, age 9+ |  |  |
| Additional mortality in 1992-1993, ages 3-4 |  |  |
| Additional mortality in 1992-1993, ages 5-8 |  |  |
| Purse-seine gear selectivity |  |  |
| Purse-seine gear selectivity |  |  |
| ADF&G acoustic scalar, log-link |  |  |
| ADF&G acoustic biomass CV |  |  |
| PWSSC acoustic scalar, log-link |  |  |
| PWSSC acoustic biomass add’l error |  |  |
| Egg deposition additional error |  | Not estimated (0.40) |
| Milt scalar, log-link |  |  |
| Milt CV |  |  |
| Proportion mature at age 3 |  |  |
| Proportion mature at age 4 |  |  |
| Age-1 abundance in 1980, log-link |  |  |
| Age-2 abundance in 1980, log-link |  |  |
| Age-3 abundance in 1980, log-link |  |  |
| Age-4 abundance in 1980, log-link |  |  |
| Age-5 abundance in 1980, log-link |  |  |
| Recruitment deviations by year, log-link |  |  |
| Mean recruitment |  |  |
| Recruitment CV |  |  |
| Covariate effect |  |  |
| Error on natural mortality by year, age, and half-year |  |  |
| Scalar for latent variable errors |  | Recruitment:  Mortality: |
| Covariate CV |  | Not estimated (0.3 or 0.7) |

Table S4. Components contributing to the negative of the logarithm of the likelihood expression for the Bayesian ASA model. Red text denotes our new equations we added to BASA from Muradian et al (2017).

|  |  |
| --- | --- |
| **Likelihood component** | **Form** |
| Complete expression, where is the total number of likelihood components (differs between models with either mortality or recruitment covariates, and fixed effect or latent variables) |  |
| Purse-seine age-composition |  |
| Spawner survey age-composition |  |
| Number of eggs deposited |  |
| Total variance for L3 |  |
| ADF&G hydroacoustic biomass |  |
| PWSSC hydroacoustic biomass |  |
| Total variance for L5 |  |
| Milt mile-days |  |
| Mortality latent variables |  |
| Recruitment deviations |  |
| Recruitment latent variables |  |

**A picture containing chart

Description automatically generated**

**Figure S1.** Bar charts of model selection values across recruitment covariates as fixed effects model variants of BASA for time period 1980-2017. Each row represents a different assumption for the parameter : estimated (a-d), fixed at 2.0 (e-h), and fixed at 0.3 (i-l). Each column represents one of the four model selection criteria used (PSIS-LOO, WAIC, PPL, and DIC). Bar lengths measure the difference in the criteria values from the best model (the minimum) in each box. The raw criteria values are labeled next to the bars. The same covariates are shown for all rows and are ordered from the smallest to largest values of in plot a).

**Chart, bar chart

Description automatically generated**

**Figure S2.** Bar charts of model selection values across natural mortality covariates as latent variable model variants of BASA. Each row represents a different assumed prior distribution for the parameter : Normal with standard deviations of 1.0 (a-d) and 5.0 (e-h). Each column represents one of the four model selection criteria used (PSIS-LOO, WAIC, PPL, and DIC). Bar lengths measure the difference in the criteria values from the best model (the minimum) in each box. The raw criteria values are labeled next to the bars. The same covariates are shown for all rows and are ordered from the smallest to largest values of in plot a).