**Appendix A: Detailed description of case study**

**Case study methods**

We explored the behavior of percent excess variance (PEV) as model diagnostic using a case study from our work in Alaska, the Gulf of Alaska walleye pollock (Monnahan et al., 2021), hereafter called “pollock”. The operational model was ported to the WHAM framework (Stock and Miller, 2021) and matched as closely as possible. Briefly, this model runs from 1971 to 2021 and uses an empirical weight at age approach so that all processes are age-based. It has one fishery with time-varying double-logistic selectivity, and 4 distinct surveys (Fig. S1), two of which have time-varying catchabilities.

PEV is calculated is a function of the effective $n\_{effective}$ and input $n\_{input}$ sample sizes. In the operational model, the multinomial effective sample sizes for the age compositions uses Francis tuning, but the input sample sizes are an area of active research and complicates the demonstration of this metric. We therefore updated the $n\_{input}$ for each fleet to the best of our knowledge. For example, we updated the survey age-composition $n\_{input}$ to a value derived from new software to conduct a nonparametric bootstrap that mimics the multi-stage sampling design used at the Alaska Fisheries Science Center (Hulson et al., 2023). However, we do not have software to do this calculation for the fishery age-composition data, which follows a more complicated multi-level design. We therefore set the fishery age compositions to have $n\_{input}$ of 1000 for all years. We then used the fitted pollock model in WHAM to simulate a data set so that the $n\_{input}$ was known to be correct. The following procedure was then followed:

1. Condition the model on the original data set, with only a change to an age-based selectivity curve that follows a first-order autoregressive process across time.
2. Update $n\_{input}$ using best information available;
3. Simulate a single new data set with age composition using the multinomial distribution. Process errors were not updated so the underlying dynamics remained the same as the model in #1;
4. Refit two new WHAM models to the simulated data that use the linear Dirichlet-multinomial age composition likelihood (Thorson et al., 2017). The linear DM estimates $n\_{effective}$, and the model also provides maximum marginal likelihood estimates for process errors. We specifically fit two estimation models:
	1. A model with no time variation in fishery selectivity or survey catchabilities.
	2. A model with time variation turned on.
5. For models 4a and 4b, calculate $PEV=1-\frac{n\_{effective}}{n\_{input}}$.

**Case study results**

The estimated PEV for the fishery was 77.1% when fitted with a model that did not include time-varying fishery selectivity (Table S1). This PEV was substantially larger than for any other fleet and correctly identified which fleet was likely to have some mis-specified process. When refitting with a model that included time-varying fishery selectivity, PEV reduced to 0.0%. One survey had no excess variance for either version, while the other three had some that only decreased marginally (Table S1). Model 4b matches the operating model and thus we compared estimates of the variance (0.256) and correlation (0.989) of the AR(1) process for fishery selectivity. The confidence interval in untransformed space for the variance term contained the truth (0.275), but not for the correlation (0.898).

# **Works cited**

Hulson, P.-J.F., Williams, B., Bryan, M., Conner, J., Siskey, M.R., Stockhausen, W.T., McDermott, S., Long, W.C., 2023. Subsampling catches to determine sex-specific length frequency in Alaska Fisheries Science Center bottom trawl surveys (NOAA Technical Memorandum No. NMFS-AFSC-464). Alaska Fisheries Science Center.

Monnahan, C.C., Dorn, M.W., Deary, A.L., Ferriss, B.E., Fissel, B.E., Honkalehto, T., Jones, D.T., Levine, M., Rogers, L., Shotwell, S.K., 2021. Assessment of the Walleye Pollock Stock in the Gulf of Alaska.

Stock, B.C., Miller, T.J., 2021. The Woods Hole Assessment Model (WHAM): A general state-space assessment framework that incorporates time-and age-varying processes via random effects and links to environmental covariates. Fish. Res. 240, 105967.

Thorson, J.T., Johnson, K.F., Methot, R.D., Taylor, I.G., 2017. Model-based estimates of effective sample size in stock assessment models using the Dirichlet-multinomial distribution. Fish. Res. 192, 84–93. https://doi.org/10.1016/j.fishres.2016.06.005

**Table S1.** Percent excess variance (PEV) calculated using equation 8 for the pollock case study. A reduction in excess variance is expected for the fishery but not the surveys when adding time variation to the model. See Fig. S1 for definition and temporal extent of data inputs.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model version | Fishery | Winter AT | NMFS BT | ADF&G BT | Summer AT |
| No time variation | 77.1 | 0 | 11.1 | 19.1 | 18.0 |
| With time variation | 0 | 0 | 8.2 | 13.4 | 13.8 |

Figure S1. Overview of the data inputs to the pollock stock assessment. Size of circle represents size of catch, age composition effective sample size, or inverse index CV, all relative within a row. The Shelikof acoustic (“Shelikof”), summer acoustic trawl (“Summer AT”), National Marine Fisheries Service bottom trawl (“NMFS BT”) and Alaska Department of Fish and Game bottom trawl surveys (“ADF&G BT”) have time-varying catchability.



Figure 2. Comparing time-varying age-based selectivity in the operating model (OM) to the two estimation models (each line is a different year), and the resulting estimates of spawning stock biomass (SSB).

