SEAK Pink Salmon 2025 Forecast Process

Sara Miller

November 12, 2024

Objective

To forecast the Southeast Alaska (SEAK) pink salmon commercial harvest in 2025. This document is for guidance as to what was done for the current forecast year. It is for internal use only.

Executive Summary

Forecasts were developed using an approach originally described in Wertheimer et al. (2006), and modified in Orsi et al. (2016) and Murphy et al. (2019), but assuming a log-normal error structure (Miller et al. 2022). This approach is based on a multiple regression model with the raw juvenile pink salmon catch-per-unit-effort (adj_raw_pink; a proxy for abundance), a vessel factor to account for the survey vessels through time, an odd and even year factor to account for potential odd and even year cycles of abundance, and temperature data from the Southeast Alaska Coastal Monitoring Survey (SECM; Piston et al. 2021; ISTI20_MJJ) or from satellite sea surface temperature (SST) data (Huang et al. 2017). The adj_raw_pink variable is not the same as the CPUE variable used in prior years that was adjusted by the pooled-species vessel calibration coefficient for the Cobb and was the maximum average from either June or July (whichever was higher). Instead, the CPUE term used in the 2025 forecast, the adj_raw_pink variable, is the natural logarithm of the maximum untransformed catch, adjusted to a 20 minute haul, from either June or July. The Stellar and Chellissa vessels were only used for one year each, 2008 and 2009, respectively, and so these two years are not used in the assessment (Table 1).

There were 18 individual models considered:

- adj_raw_pink model with a vessel interaction and an odd/even year factor (m1d);
- adj_raw_pink model with a vessel interaction, an odd/even year factor, and with temperature data from the the SECM survey (m2d);
- 16 adj_raw_pink models with a vessel interaction, an odd/even year factor, and with satellite SST data (m3d-m18d);

The model performance metrics one-step ahead mean absolute percent error (MAPE) for the last five years (MAPE5; forecast years 2020 through 2024) was used to evaluate the forecast accuracy of the 18 individual models, the AICc values were calculated for each model to prevent over-parameterization of the model, and the adjusted R-squared values, significant terms, and overall p-value of the model were used to the determine fit. Based upon the performance metric the 5-year MAPE, the AICc values, significant parameters in the models, and the adjusted R-squared values, model m11d (a model that included NSEAK_SST_May; Appendix B) was the best performing model and the 2025-forecast using this model has a point estimate of 28.9 million fish (80% prediction interval: 15.7 to 52.9 million fish).

Analysis

Individual, multiple linear regression models

Biophysical variables based on data from Southeast Alaska were used to forecast the harvest of adult pink salmon in Southeast Alaska, one year in advance, using individual, multiple linear regression models (models m1d–m18d). The simplest regression model (model m1d) consisted of the predictor variable juvenile adj_raw_pink (X_1) with a vessel factor interaction (X_V), and an odd/even year factor(X_B). The other 17 regression models (models m2d-m18d) consisted of the predictor variable juvenile adj_raw_pink (X_1) with a vessel factor interaction (X_V), an odd/even year factor (X_B), and a temperature index (X_2). The general model structure was

$$E(Y) = \hat{\beta}_0 + \hat{\beta}_{1V} X_V X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_B X_B$$

.

The odd/even year factor adjusts the model intercept by $\hat{\beta}_0 + \beta_B$ as $X_B = 0$ for even years (no adjustment to intercept, defaults to model intercept) and $X_B = 1$ for odd years (adjustment for odd years based on β_B). The vessel interaction adjusts the model slope. For example, during years when the survey vessel was the Cobb, the slope is adjusted by the $\hat{\beta}_{1COBB}$ which is then multiplied by the adj_raw_pink term in that year.

The temperature index for models m2d-m18d was either the SECM survey Icy Strait temperature Index (ISTI20_MJJ; Murphy et al. 2019) or one of the 16 satellite-derived SST data (Huang et al. 2017). Although the simplest model did not contain a temperature variable, including temperature data with CPUE has been shown to result in a substantial improvement in the accuracy of model predictions (Murphy et al. 2019). The response variable (Y; Southeast Alaska adult pink salmon harvest in millions), and the adj_raw_pink (CPUE) data were natural log transformed in the model, but temperature data were not. The forecast ($\hat{Y}_i$), and 80% prediction intervals (based on output from program R; R Core Team 2023) from the 18 regression models were exponentiated and bias-corrected (Miller 1984),

$$\hat{F}_i = \exp(\hat{Y}_i + \frac{{\sigma_\mathrm{i}}^2}{2}), \quad (2)$$

where $\hat{F}_i$ is the preseason forecast (for each model i) in millions of fish, and σ_i is the variance (for each model i).

Performance metric: One-step ahead MAPE

The model summary results using the performance metric one-step ahead MAPE are shown in Table 2; the smallest value is the preferred model (Appendix C). The performance metric one-step ahead MAPE was calculated as follows.

1. Estimate the regression parameters at time t-1 from data up to time t-1.
2. Make a prediction of $\hat{Y}_t$ at time t based on the predictor variables at time t and the estimate of the regression parameters at time t-1 (i.e., the fitted regression equation).
3. Calculate the MAPE based on the prediction of $\hat{Y}_t$ at time t and the observed value of Y_t at time t,

$$\mathrm{MAPE} = |\frac{\exp(Y_t) - \exp(\hat{Y}_t + \frac{{\sigma_\mathrm{t}}^2}{2})}{\exp(Y_t)}|. \quad (3)$$

4. For each individual model, average the MAPEs calculated from the forecasts,

$$\frac{1}{n}\sum_{t=1}^{n}|\frac{\exp(Y_t) - \exp(\hat{Y}_t + \frac{\sigma_t{}^2}{2})}{\exp(Y_t)}|, \tag{4}$$

where n is the number of forecasts in the average (5 forecasts for the 5-year MAPE and 10 forecasts for the 10-year MAPE). For example, to calculate the five year one-step-ahead MAPE for model m1 for the 2022 forecast, use data up through year 2016 (e.g., data up through year 2016 is t -1 and the forecast is for t, or year 2017). Then, calculate a MAPE based on the 2017 forecast and the observed pink salmon harvest in 2017 using equation 3. Next, use data up through year 2017 (e.g., data up through year 2017 is t -1 and the forecast is for year 2018; t) and calculate a MAPE based on the 2018 forecast and the observed pink salmon harvest in 2018 using equation 3. Repeat this process for each subsequent year through year 2020 to forecast 2021. Finally, average the five MAPEs to calculate a five year one-step-ahead MAPE for model m1. As the results of the 5-year MAPEs with or without the forecast bias adjustment have been similar (i.e., the model performance order did not change whether the five year one-step-ahead MAPE or the bias-corrected five year one-step-ahead MAPE was compared), for simplicity, the bias adjustment for the forecast was not used in the calculation of the five year one-step-ahead MAPE for model comparison.

Akaike Information Criterion corrected for small sample sizes (AICc)

Hierarchical models were compared with the AICc criterion. The best fit models, according to the AICc criterion, is one that explains the greatest amount of variation with the fewest independent variables (i.e., the most parsimonious; Table 2). The lower AICc values are better, and the AICc criterion penalizes models that use more parameters. Comparing the AICc values of two hierarchical models, a $\Delta_i \leq 2$ suggests that the two models are essentially the same, and the most parsimonious model should be chosen (Burnham and Anderson 2004). If the $\Delta_i > 2$, the model with the lower AICc value should be chosen.

Results

Based upon the 5-year MAPE performance metric, the AICc values, significant parameters in the models, and the adjusted R-squared values, model m11d (a model that included CPUE (i.e., adj_raw_pink) and the satellite SST variable from northern SEAK in May; Table 1 and Table 2; Appendix C) was the best performing model and the 2025-forecast using this model has a point estimate of 28.9 million fish (80% prediction interval: 15.7 to 52.9 million fish).

Table 1: Summary of the data used for the 2025-forecast.

JYear	Year	SEAKCatch	CPUE	Vessel	adj_raw_pink
1997	1998	42.448749	2.4777444	Cobb	5.332719
1998	1999	77.818105	5.6223800	Cobb	7.141245
1999	2000	20.248977	1.5977233	Cobb	5.049856
2000	2001	67.023653	3.7299847	Cobb	6.385194
2001	2002	45.315401	2.8688260	Cobb	5.501258
2002	2003	52.466735	2.7846641	Cobb	6.109248
2003	2004	45.309744	3.0778204	Cobb	5.318120
2004	2005	59.121487	3.8994067	Cobb	6.804614
2005	2006	11.606500	2.0403454	Cobb	5.062595
2006	2007	44.796397	2.5727807	Cobb	6.144186
2007	2008	15.908924	1.1676386	Cobb	3.850148
2008	2009	38.024357	2.3234731	Steller	NA
2009	2010	24.141671	2.3330031	Chellissa	NA
2010	2011	58.882054	4.1083181	NW Explorer	9.431963
2011	2012	21.277018	1.4548381	NW Explorer	5.707110
2012	2013	94.719422	3.5215052	NW Explorer	8.515792
2013	2014	37.173865	2.1427677	NW Explorer	6.800170
2014	2015	35.092568	3.8173733	NW Explorer	7.673223
2015	2016	18.374200	2.4526465	NW Explorer	6.720220
2016	2017	34.734774	4.3510925	NW Explorer	7.937732
2017	2018	8.067700	0.3458366	NW Explorer	3.555348
2018	2019	21.141910	1.1715584	Medeia	3.931826
2019	2020	8.062989	1.1419709	Medeia	5.318120
2020	2021	48.528192	2.1475023	Medeia	4.406719
2021	2022	18.299158	0.8707362	Medeia	4.462619
2022	2023	47.839511	1.4463812	Medeia	4.330733
2023	2024	20.117000	1.1932412	Medeia	3.401197
2024	2025	NA	1.6601690	Medeia	3.850148

Table 2: Summary of the performance metrics for the 18 regression models.

terms	Model	Fit	Fit_LPI	Fit_UPI	AdjR2	AICc	MAPE5
NSEAK_SST_May	m11d	28.872	15.744	52.945	0.646	39.23816	101.305
Icy_Strait_SST_May	m7d	28.240	15.216	52.412	0.632	40.16175	101.942
NSEAK_SST_AMJ	m13d	26.801	14.728	48.773	0.657	38.42185	102.573
Icy_Strait_SST_AMJ	m9d	27.622	15.110	50.497	0.651	38.89029	106.620
NSEAK_SST_AMJJ	m14d	29.545	16.001	54.556	0.637	39.82400	107.200
Chatham_SST_May	m3d	28.027	15.158	51.820	0.637	39.84129	107.380
Icy_Strait_SST_AMJJ	m10d	29.723	16.065	54.990	0.635	39.98447	108.792
Chatham_SST_AMJ	m5d	26.053	14.247	47.641	0.653	38.68780	109.687
SEAK_SST_AMJ	m17d	27.176	14.799	49.904	0.646	39.19713	111.520
Icy_Strait_SST_MJJ	m8d	32.481	17.267	61.099	0.618	41.15368	112.711
NSEAK_SST_MJJ	m12d	32.481	17.275	61.073	0.618	41.12003	113.291
Chatham_SST_AMJJ	m6d	29.103	15.685	53.998	0.632	40.19896	113.323
SEAK_SST_AMJJ	m18d	29.739	16.023	55.196	0.631	40.24411	114.618
SEAK_SST_May	m15d	28.861	15.536	53.613	0.631	40.28240	114.944
Chatham_SST_MJJ	m4d	32.262	16.995	61.247	0.606	41.89097	120.243
SEAK_SST_MJJ	m16d	32.640	17.275	61.671	0.613	41.47587	120.676
no temp	m1d	30.467	15.503	59.872	0.558	42.11397	131.599
ISTI20_MJJ	m2d	23.004	12.230	43.269	0.637	39.84092	133.840

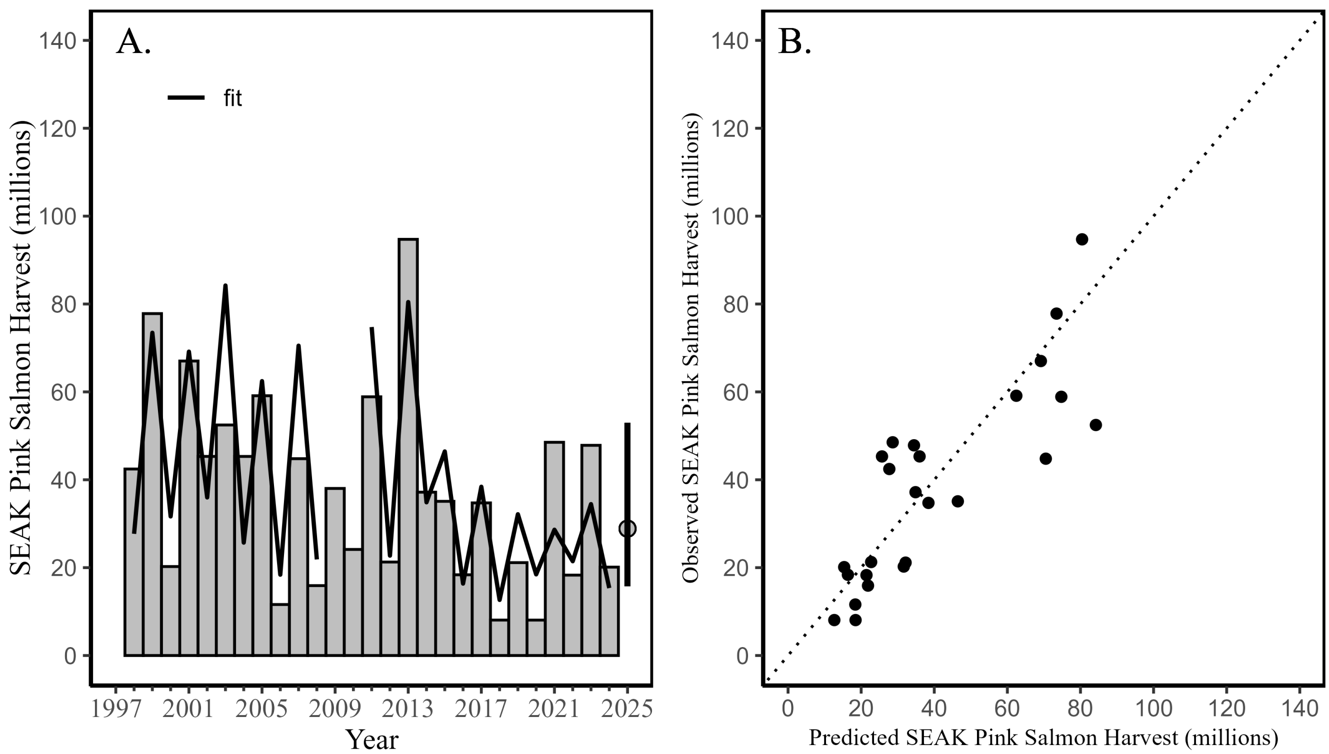

Figure 1: A. SEAK pink salmon harvest (millions) by year with the model fit (line) based upon the best performing model (model m11d). The predicted 2025 forecast is symbolized as a grey circle with an 80% prediction interval (15.7 to 52.9 million fish). B. SEAK pink salmon harvest (millions) against the fitted values from model m11d by year. The dotted line is a one to one reference line.

References

Burnham, K. P., and D. R. Anderson (1998) Model Selection and Inference. Springer, New York. 353 pp.

Burnham, K. P., and D. R. Anderson (2004) Multimodel inference: Understanding AIC and BIC in model selection. Sociological Methods & Research, Vol. 33(2): 261-304.

Cook, R. D. (1977) Detection of influential observations in linear regression. Technometrics 19: 15-18.

Fox, J. and S. Weisburg (2019) An R Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage Publications, Inc.

Heinl, S. C., and A. W. Piston (2009) Standardizing and automating the Southeast Alaska pink salmon escapement index. Alaska Department of Fish and Game, Division of Commercial Fisheries, Regional Information Report No. 1J09-06, Douglas.

Huang, B., P. W. Thorne, V. F. Banzon, T. Boyer, G. Chepurin, J. H. Lawrimore, M. J. Menne, T. M. Smith, R. S. Vose, and H. M. Zhang (2017) Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. Journal of Climate 30:8179–8205.

Miller, D. M. (1984) Reducing transformation bias in curve fitting. The American Statistician 38: 124-126.

Miller, S. E., J. M. Murphy, S. C. Heinl, A. W. Piston, E. A. Fergusson, R. E. Brenner, W. W. Strasburger, and J. H. Moss (2022) Southeast Alaska pink salmon forecasting models. Alaska Department of Fish and Game, Fishery Manuscript No. 22-03, Anchorage.

Murphy, J. M., E. A. Fergusson, A. Piston, A. Gray, and E. Farley (2019) Southeast Alaska pink salmon growth and harvest forecast models. North Pacific Anadromous Fish Commission Technical Report No. 15: 75-81.

Orsi, J. A., E. A. Fergusson, A. C. Wertheimer, E. V. Farley, and P. R. Mundy (2016) Forecasting pink salmon production in Southeast Alaska using ecosystem indicators in times of climate change. N. Pac. Anadr. Fish Comm. Bull. 6: 483–499. (Available at https://npafc.org)

Piston, A. W., and S. C. Heinl (2020) Pink salmon stock status and escapement goals in Southeast Alaska through 2019. Alaska Department of Fish and Game, Special Publication No. 20-09, Anchorage

Piston, A. W., J. Murphy, J. Moss, W. Strasburger, S. C. Heinl, E. Fergusson, S. Miller, A. Gray, and C. Waters (2021) Operational Plan: Southeast coastal monitoring, 2021. ADF&G, Regional Operational Plan No. ROP.CF.1J.2021.02, Douglas.

R Core Team (2023) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.r-project.org/index.html

Ren, Y. Y., L. C. Zhou, L. Yang, P. Y. Liu, B. W. Zhao and H. X. Liu (2016) Predicting the aquatic toxicity mode of action using logistic regression and linear discriminant analysis, SAR and QSAR in Environmental Research, DOI: 10.1080/1062936X.2016.1229691

Sturdevant, M.V., E.A. Fergusson, J.A. Orsi, and A.C. Wertheimer (2004) Diel feeding and gastric evacuation of juvenile pink and chum salmon in Icy Strait, Southeastern Alaska, May-September 2001. NPAFC Tech. Rep. 5. (Available at http://www.npafc.org).

Trenberth, K. E., and J. W. Hurrell (1994) Decadal atmosphere-ocean variations in the Pacific Climate Dynamics, Berlin 9(6):303-319.

Wertheimer A. C., J. A. Orsi, M. V. Sturdevant, and E. A. Fergusson (2006) Forecasting pink salmon harvest in Southeast Alaska from juvenile salmon abundance and associated environmental parameters. In Proceedings of the 22nd Northeast Pacific Pink and Chum Workshop. Edited by H. Geiger (Rapporteur). Pac. Salmon Comm. Vancouver, British Columbia. pp. 65–72.

Wertheimer, A. C., J. A. Orsi, and E. A. Fergusson (2015) Forecasting pink salmon harvest in southeast Alaska from juvenile salmon abundance and associated biophysical parameters: 2014 returns and 2015 forecast. NPAFC Doc. 1618. 26 pp. National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service (NMFS), Alaska Fisheries Science Center, Auke Bay Laboratories, Ted Stevens Marine Research Institute (Available at http://www.npafc.org).

Wertheimer, A. C., J. A. Orsi, E. A. Fergusson, and J.M. Murphy (2018) Forecasting pink salmon harvest in Southeast Alaska from juvenile salmon abundance and associated biophysical parameters: 2016 returns and 2017 forecast. NPAFC Doc. 1772. 25 pp. National Oceanic and Atmospheric Administration (NOAA), National Marine Fisheries Service (NMFS), Alaska Fisheries Science Center, Auke Bay Laboratories, Ted Stevens Marine Research Institute (Available at http://www.npafc.org).

Zhang, Z. (2016) Residuals and regression diagnostics: focusing on logistic regression. Annals of Translational Medicine 4: 195.

Appendix A

Variable definitions

adj_raw_pink: First, the raw untransformed catch by year, month, haul, and area was adjusted to a 20 minute haul (raw_pink_catch). Then, these values were log-transformed (i.e., ln(raw_pink_catch + 1). Next, the maximum value in either June or July, whichever month had the highest in a given year, was used as the adj_raw_pink value for the year. Note: the adj_raw_pink value data was not adjusted using vessel calibration factors. This value was used as the 'CPUE' value in the 2025 forecast (i.e., adj_raw_pink = maximum(ln(raw_pink_catch + 1)) in either June or July).

CPUE: The average Ln(CPUE+1) for catches in either June or July, whichever month had the highest average in a given year, where effort was a standard trawl haul. The CPUE data was adjusted using vessel calibration factors to account for differences in fishing power among vessels. The last time the CPUE variable was incorporated in the forecasting process was the 2024 forecast.

ISTI20_MJJ: The average 20-m integrated water column temperature at the eight stations in Icy Strait (Icy Strait and Upper Chatham transects) sampled during the SECM surveys in May, June, and July of each year (in degrees Celsius). The last time the ISTI variable was incorporated in the forecasting process was the 2023 forecast.

Satellite SST variables

Icy_Strait_SST_May: The Icy Strait region encompasses waters of Icy Strait from the east end of Lemesurier Island to a line from Point Couverden south to Point Augusta. This variable is the average SST in May. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Icy_Strait_SST_MJJ: The Icy Strait region encompasses waters of Icy Strait from the east end of Lemesurier Island to a line from Point Couverden south to Point Augusta. This variable is the average SST in May through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Icy_Strait_SST_AMJ: The Icy Strait region encompasses waters of Icy Strait from the east end of Lemesurier Island to a line from Point Couverden south to Point Augusta. This variable is the average SST in April through June. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Icy_Strait_SST_AMJJ: The Icy Strait region encompasses waters of Icy Strait from the east end of Lemesurier Island to a line from Point Couverden south to Point Augusta. This variable is the average SST in April through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Chatham_SST_May: The Chatham and Icy Straits region encompasses waters of Chatham and Icy Straits east of Lemesurier Island to Point Couverden, and south to the approximate latitude of 56.025 degrees north (roughly Cape Decision off Kuiu Island). This variable is the average SST in May. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Chatham_SST_MJJ: The Chatham and Icy Straits region encompasses waters of Chatham and Icy Straits east of Lemesurier Island to Point Couverden, south to the approximate latitude of 56.025 degrees north (roughly Cape Decision off Kuiu Island). This variable is the average SST in May through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Chatham_SST_AMJ: The Chatham and Icy Straits region encompasses waters of Chatham and Icy Straits east of Lemesurier Island to Point Couverden, south to the approximate latitude of 56.025 degrees north (roughly Cape Decision off Kuiu Island). This variable is the average SST in April through June. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Chatham_SST_AMJJ: The Chatham and Icy Straits region encompasses waters of Chatham and Icy Straits east of Lemesurier Island to Point Couverden, south to the approximate latitude of 56.025 degrees north (roughly Cape Decision off Kuiu Island). This variable is the average SST in April through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

NSEAK_SST_May: The NSEAK region encompasses northern Southeast Alaska from 59.475 to 56.075 degrees north latitude (approximately Districts 9 through 15, and District 13 inside area only; northern Southeast Inside subregion for Southeast Alaska (NSEI)). This variable is the average SST in May. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

NSEAK_SST_MJJ: The NSEAK region encompasses northern Southeast Alaska from 59.475 to 56.075 degrees north latitude (approximately Districts 9 through 15, and District 13 inside area only; northern Southeast Inside subregion for Southeast Alaska (NSEI)). This variable is the average SST in May through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

NSEAK_SST_AMJ: The NSEAK region encompasses northern Southeast Alaska from 59.475 to 56.075 degrees north latitude (approximately Districts 9 through 15, and District 13 inside area only; northern Southeast Inside subregion for Southeast Alaska (NSEI)). This variable is the average SST in April through June. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

NSEAK_SST_AMJJ: The NSEAK region encompasses northern Southeast Alaska from 59.475 to 56.075 degrees north latitude (approximately Districts 9 through 15, and District 13 inside area only; northern Southeast Inside subregion for Southeast Alaska (NSEI)). This variable is the average SST in April through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

SEAK_SST_May: The SEAK region encompasses Southeast Alaska from 59.475 to 54.725 degrees north latitude. This variable is the average SST in May. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

SEAK_SST_MJJ: The SEAK region encompasses northern Southeast Alaska from 59.475 to 54.725 degrees north latitude. This variable is the average SST in May through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

SEAK_SST_AMJ: The SEAK region encompasses Southeast Alaska from 59.475 to 54.725 degrees north latitude. This variable is the average SST in April through June. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

SEAK_SST_AMJJ: The SEAK region encompasses Southeast Alaska from 59.475 to 54.725 degrees north latitude. This variable is the average SST in April through July. The last time this variable was incorporated in the forecasting process was the 2023 forecast.

Appendix B

Table 3: Parameter estimates for the 18 individual models.

Model	Term	Estimate	Standard Error	Statistic	p value
m1d	(Intercept)	1.9912203	0.571	3.488	0.002
m1d	as.factor(odd_even_factor)odd	0.6160075	0.230	2.682	0.014
m1d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2399567	0.112	2.150	0.044
m1d	as.factor(vessel)Medeia:adj_raw_pink_log	0.1839251	0.150	1.223	0.235
m1d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.1638610	0.090	1.827	0.083
m2d	(Intercept)	5.3270340	1.533	3.475	0.003
m2d	as.factor(odd_even_factor)odd	0.4329940	0.223	1.944	0.067
m2d	ISTI20_MJJ	-0.4255504	0.184	-2.311	0.032
m2d	as.factor(vessel)Cobb:adj_raw_pink_log	0.3456410	0.111	3.113	0.006
m2d	as.factor(vessel)Medeia:adj_raw_pink_log	0.3290965	0.150	2.193	0.041
m2d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2536124	0.090	2.816	0.011
m3d	(Intercept)	4.1959348	1.085	3.867	0.001
m3d	as.factor(odd_even_factor)odd	0.5929511	0.208	2.845	0.010
m3d	Chatham_SST_May	-0.3203209	0.139	-2.311	0.032
m3d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2691563	0.102	2.640	0.016
m3d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2462952	0.139	1.773	0.092
m3d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2079838	0.083	2.491	0.022
m4d	(Intercept)	5.0007657	1.713	2.919	0.009
m4d	as.factor(odd_even_factor)odd	0.5777599	0.218	2.652	0.016
m4d	Chatham_SST_MJJ	-0.3247573	0.175	-1.851	0.080
m4d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2695953	0.107	2.529	0.020
m4d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2490667	0.146	1.703	0.105
m4d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.1956742	0.086	2.265	0.035
m5d	(Intercept)	4.9387544	1.263	3.912	0.001
m5d	as.factor(odd_even_factor)odd	0.5459356	0.205	2.659	0.015
m5d	Chatham_SST_AMJ	-0.4216897	0.166	-2.548	0.020
m5d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2870243	0.101	2.854	0.010
m5d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2684859	0.137	1.956	0.065
m5d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2221011	0.083	2.687	0.015
m6d	(Intercept)	5.2279771	1.539	3.397	0.003
m6d	as.factor(odd_even_factor)odd	0.5523270	0.212	2.610	0.017
m6d	Chatham_SST_AMJJ	-0.4001149	0.179	-2.235	0.038
m6d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2848838	0.104	2.743	0.013
m6d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2690256	0.142	1.889	0.074
m6d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2103180	0.084	2.490	0.022
m7d	(Intercept)	3.9396862	1.013	3.890	0.001
m7d	as.factor(odd_even_factor)odd	0.6233829	0.210	2.975	0.008
m7d	Icy_Strait_SST_May	-0.3056860	0.136	-2.243	0.037
m7d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2725515	0.103	2.650	0.016
m7d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2487522	0.140	1.775	0.092
m7d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2119986	0.085	2.507	0.021
m8d	(Intercept)	4.8547956	1.511	3.213	0.005
m8d	as.factor(odd_even_factor)odd	0.5889460	0.214	2.750	0.013
m8d	Icy_Strait_SST_MJJ	-0.3077262	0.152	-2.024	0.057
m8d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2789870	0.106	2.641	0.016
m8d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2554634	0.144	1.770	0.093
m8d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2017628	0.086	2.359	0.029
m9d	(Intercept)	4.7283396	1.204	3.927	0.001

Model	Term	Estimate	Standard Error	Statistic	p value
m9d	as.factor(odd_even_factor)odd	0.5977309	0.204	2.924	0.009
m9d	Icy_Strait_SST_AMJ	-0.4108472	0.164	-2.507	0.021
m9d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2890404	0.101	2.857	0.010
m9d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2736832	0.138	1.977	0.063
m9d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2243148	0.083	2.693	0.014
m10d	(Intercept)	4.9480911	1.396	3.544	0.002
m10d	as.factor(odd_even_factor)odd	0.5766986	0.210	2.753	0.013
m10d	Icy_Strait_SST_AMJJ	-0.3724685	0.163	-2.281	0.034
m10d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2927844	0.104	2.813	0.011
m10d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2753920	0.142	1.934	0.068
m10d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2145028	0.084	2.539	0.020
m11d	(Intercept)	4.2127501	1.045	4.030	0.001
m11d	as.factor(odd_even_factor)odd	0.6103952	0.206	2.967	0.008
m11d	NSEAK_SST_May	-0.3177307	0.130	-2.437	0.025
m11d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2553825	0.100	2.550	0.020
m11d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2441305	0.137	1.783	0.090
m11d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.1986937	0.082	2.436	0.025
m12d	(Intercept)	5.1785877	1.656	3.127	0.006
m12d	as.factor(odd_even_factor)odd	0.5809702	0.214	2.711	0.014
m12d	NSEAK_SST_MJJ	-0.3404367	0.168	-2.032	0.056
m12d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2645502	0.104	2.532	0.020
m12d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2651131	0.145	1.823	0.084
m12d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.1952628	0.085	2.303	0.033
m13d	(Intercept)	4.9535063	1.245	3.978	0.001
m13d	as.factor(odd_even_factor)odd	0.5659653	0.203	2.784	0.012
m13d	NSEAK_SST_AMJ	-0.4242135	0.163	-2.601	0.018
m13d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2739327	0.099	2.761	0.012
m13d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2723961	0.137	1.992	0.061
m13d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2143598	0.081	2.635	0.016
m14d	(Intercept)	5.2897395	1.516	3.490	0.002
m14d	as.factor(odd_even_factor)odd	0.5676896	0.209	2.714	0.014
m14d	NSEAK_SST_AMJJ	-0.4059872	0.175	-2.315	0.032
m14d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2757206	0.102	2.695	0.014
m14d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2776597	0.142	1.954	0.066
m14d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2050484	0.083	2.465	0.023
m15d	(Intercept)	4.1778878	1.116	3.744	0.001
m15d	as.factor(odd_even_factor)odd	0.6042856	0.210	2.876	0.010
m15d	SEAK_SST_May	-0.2901060	0.131	-2.217	0.039
m15d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2566226	0.102	2.507	0.021
m15d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2464497	0.140	1.756	0.095
m15d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.1985297	0.083	2.378	0.028
m16d	(Intercept)	5.0781468	1.671	3.039	0.007
m16d	as.factor(odd_even_factor)odd	0.5790526	0.216	2.681	0.015
m16d	SEAK_SST_MJJ	-0.3122964	0.160	-1.950	0.066
m16d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2583076	0.105	2.461	0.024
m16d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2635742	0.147	1.798	0.088
m16d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.1935404	0.085	2.268	0.035
m17d	(Intercept)	4.9143041	1.300	3.780	0.001
m17d	as.factor(odd_even_factor)odd	0.5650080	0.207	2.735	0.013
m17d	SEAK_SST_AMJ	-0.3894226	0.159	-2.445	0.024
m17d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2746771	0.101	2.723	0.014
m17d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2752872	0.140	1.972	0.063

Model	Term	Estimate	Standard Error	Statistic	p value
m17d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2133403	0.083	2.578	0.018
m18d	(Intercept)	5.2500131	1.554	3.378	0.003
m18d	as.factor(odd_even_factor)odd	0.5697285	0.211	2.701	0.014
m18d	SEAK_SST_AMJJ	-0.3760316	0.169	-2.226	0.038
m18d	as.factor(vessel)Cobb:adj_raw_pink_log	0.2697126	0.103	2.622	0.017
m18d	as.factor(vessel)Medeia:adj_raw_pink_log	0.2767213	0.144	1.928	0.069
m18d	as.factor(vessel)NW Explorer:adj_raw_pink_log	0.2033564	0.084	2.426	0.025

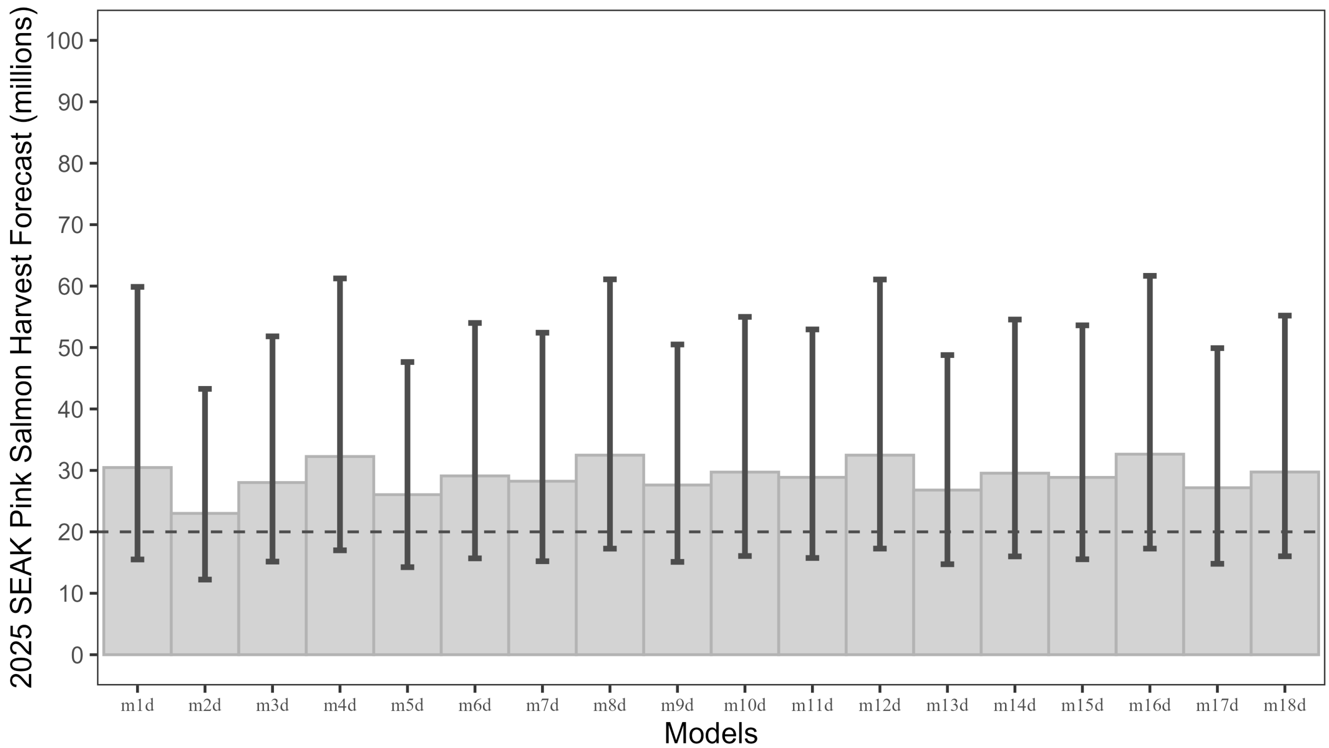

Figure B1: Bias-corrected forecasts (grey bars) for the 18 regression models with 80% prediction intervals (vertical grey lines). Based upon the performance metrics, the best performing model was m11d. A horizontal dotted line at 20 million fish is placed on the figure for reference only. The 2025-forecast using the model m11d has a point estimate of 28.9 million fish (80% prediction interval: 15.7 to 52.9 million fish).

Appendix C

Model Diagnostics

Model diagnostics for model m11d included residual plots, the lack of fit test, and influential observation diagnostics using Cook's distance (Cook 1977), the Bonferroni outlier test, and leverage plots (Table 4; Figure C1; Figure C2). Model diagnostics were used to identify observations that were potential outliers, had high leverage, or were influential (Zhang 2016).

Table 4: Detailed output for model m11d. Fitted values (in millions of fish) are bias-corrected.

Year	Harvest	Residuals	Hat values	Cooks distance	Std. residuals	Fitted values	vessel
1998	42.45	0.51	0.12	0.04	1.35	27.70	Cobb
1999	77.82	0.14	0.21	0.01	0.39	73.45	Cobb
2000	20.25	-0.37	0.14	0.03	-0.98	31.66	Cobb
2001	67.02	0.05	0.15	0.00	0.13	69.16	Cobb
2002	45.32	0.31	0.17	0.02	0.85	35.98	Cobb
2003	52.47	-0.39	0.23	0.06	-1.11	84.21	Cobb
2004	45.31	0.65	0.13	0.08	1.73	25.70	Cobb
2005	59.12	0.03	0.21	0.00	0.07	62.45	Cobb
2006	11.61	-0.38	0.26	0.07	-1.09	18.40	Cobb

Year	Harvest	Residuals	Hat values	Cooks distance	Std. residuals	Fitted values	vessel
2007	44.80	-0.37	0.16	0.03	-1.01	70.51	Cobb
2008	15.91	-0.24	0.19	0.02	-0.66	21.87	Cobb
2011	58.88	-0.16	0.31	0.02	-0.47	74.76	NW Explorer
2012	21.28	0.02	0.17	0.00	0.04	22.73	NW Explorer
2013	94.72	0.24	0.30	0.04	0.72	80.44	NW Explorer
2014	37.17	0.15	0.31	0.01	0.43	34.87	NW Explorer
2015	35.09	-0.20	0.19	0.01	-0.55	46.44	NW Explorer
2016	18.37	0.20	0.31	0.03	0.59	16.37	NW Explorer
2017	34.73	-0.02	0.30	0.00	-0.06	38.41	NW Explorer
2018	8.07	-0.37	0.41	0.17	-1.19	12.65	NW Explorer
2019	21.14	-0.34	0.28	0.06	-0.99	32.15	Medeia
2020	8.06	-0.75	0.51	1.23	-2.65	18.47	Medeia
2021	48.53	0.61	0.26	0.18	1.76	28.63	Medeia
2022	18.30	-0.08	0.28	0.00	-0.23	21.46	Medeia
2023	47.84	0.41	0.24	0.07	1.17	34.46	Medeia
2024	20.12	0.35	0.17	0.03	0.96	15.35	Medeia

Cook's distance

Cook's distance is a measure of influence, or the product of both leverage and outlier. Cook's distance,

$$D_i = \frac{e^2_{PSi}}{k+1} * \frac{h_i}{1-h_i}, \tag{5}$$

where e^2_{PSi} is the standardized Pearson residuals, h_i are the hat values (measure of leverage), and k is the number of predictor variables in the model, is a measure of overall influence of the i_{th} data point on all n fitted values (Fox and Weisburg 2019). A large value of Cook's distance indicates that the data point is an influential observation. Cook's distance values greater than $4/(n-k-1)$, where n is the number of observations (i.e., 25), was used as a benchmark for identifying the subset of influential observations (Ren et al. 2016). Therefore, a Cook's distance cut-off of 0.21 was used; observations with a Cook's distance greater than 0.21 may be influential observations (Figure C1a).

Leverage

An observation that is distant from the average covariate pattern is considered to have high leverage or hat-value. If an individual observation has a leverage value h_i greater than 2 or 3 times p/n (Ren et al. 2016), it may be a concern (where p is the number of parameters in the model including the intercept (i.e., 6), and n is the number of observations in the model (i.e., 25); $p/n = 6/25 = 0.24$ for this study). Therefore, a leverage cut-off of 0.48 was used; observations with a leverage value greater than 0.48 may affect the model properties (e.g., summary statistics, standard errors, predicted values) (Figure C1b).

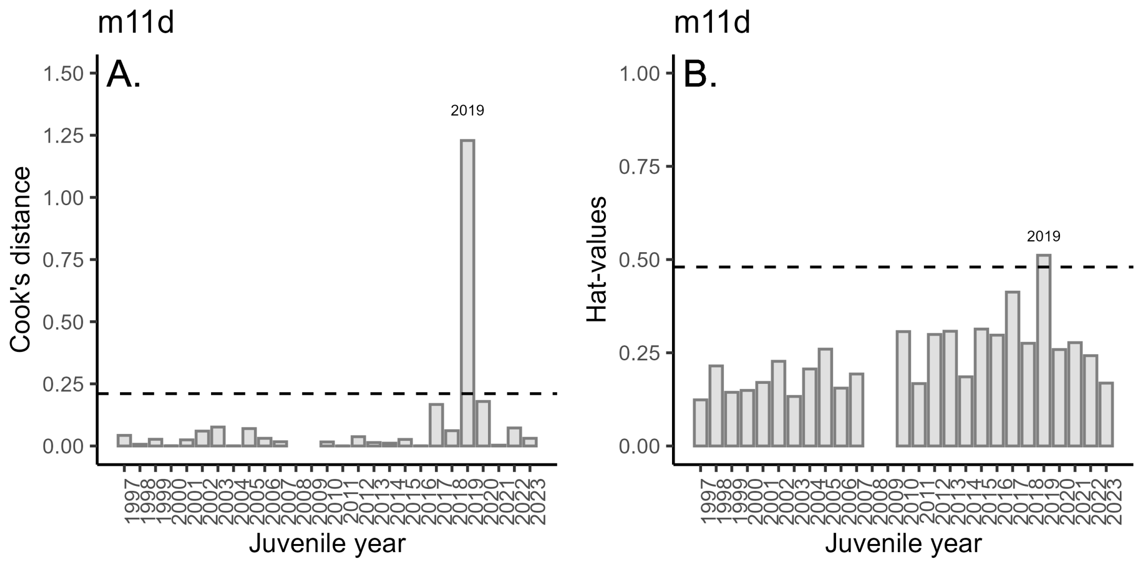

Figure C1: Diagnostics plots of influential observations including A. Cook's distance (with a cut-off value of 0.24), and B. leverage values (with a cut-off value of 0.48) from model m11d.

Influential datapoints

A lack-of fit test was performed between model m11d (full model) and a reduced model with just the adj_raw_pink ('CPUE' term) and vessel interaction term (reduced model). Based on an anova between the full and reduced models, the p-value was less than .05, the null hypothesis of the test is rejected, and it can be concluded that the full model offers a statistically significantly better fit than the reduced model. Diagnostics indicated that one of the data points was above the cut-off value for the Cook's distance (Figure C1a; 2019). One observation had high leverage values (Figure C1b; 2019). Based on the Bonferroni outlier test, none of the data points had studentized residuals with a significant Bonferroni P-value suggesting that none of the data points impacted the model fitting; although observation 21 was the most extreme (juvenile years 2019) based on standardized residuals (Figure C2c; Table 5). Based on the lightly curved fitted lines in the residual versus fitted plot (Figure C2d), the fitted plot shows some lack of fit of the model.

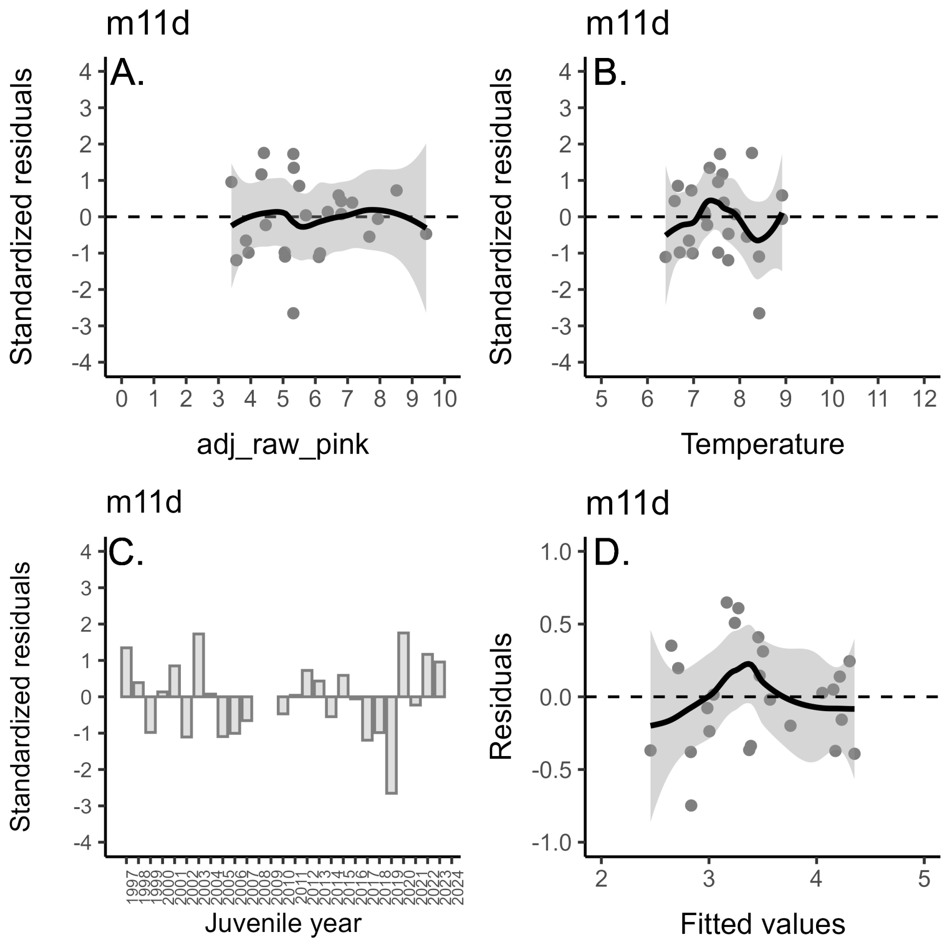

Figure C2: Standardized residuals versus predicted plots for A. adj_raw_pink ('CPUE' term) and B. temperature (average May SST in northern Southeast Alaska) for model m11d. C. Standardized residuals versus juvenile year and D. residuals versus fitted values for model m11d. Positive residuals indicate that the observed harvest was larger than predicted by the model.