SEAK Pink Salmon 2024 Forecast Process

Sara Miller

October 31, 2023

Objective

To forecast the Southeast Alaska (SEAK) pink salmon commercial harvest in 2024.

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 juvenile pink salmon catch-per-unit-effort (CPUE; a proxy for abundance), 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), and some additional biophysical variables (condition (July), energy density (July), average zooplankton total water column (May, June, or July), zooplankton density (May, June, or July), the North Pacific Index (NPI), and the Southeast Alaska (SEAK) pink salmon escapement index; see Appendix A). Based on prior discussions, the index of juvenile abundance (i.e., CPUE) was based on the pooled-species vessel calibration coefficient.

There were 37 individual models considered:

- CPUE-only model (m1);
- CPUE model with temperature data from the the SECM survey (m2);
- 16 CPUE models with satellite SST data (m3-m18);
- biophysical variable model based on a backward/forward stepwise regression (m19);
- CPUE model with the SEAK pink salmon escapement index (m1a); and
- 17 CPUE models with a temperature variable (based on the SECM survey or satellite SST data) and the SEAK pink salmon escapement index (m2a-m18a).

A backward/forward stepwise regression with an alpha value of $p < 0.05$ was performed on the full model that included CPUE, all the temperature variables, and the additional biophysical variables. Next, the Akaike Information Criterion (AIC values; Burnham and Anderson 1998) for each significant step of the stepwise regression was calculated, to prevent over-parameterization of the model. The model with the lowest AIC value was added as an additional forecasting model (m19) along with the other 36 models. Next, the model performance metrics one-step ahead mean absolute percent error (MAPE) for the last five years (forecast years 2019 through 2023) and for the last ten years (forecast years 2014 through 2023) were used to evaluate the forecast accuracy of the 37 individual models, the AICc values were calculated for each model to prevent over-parameterization of the model, and the adjusted R-squared values were used to the determine fit. Based

upon the performance metric the 10-year MAPE, the AICc values, significant parameters in the models, and the adjusted R-squared values, model m11 (a model that included CPUE and the satellite SST variable from northern SEAK in May; Appendix B) was the best performing model and the 2024-forecast using this model would be in the average range with a point estimate of 19.2 million fish (80% prediction interval: 11.7 to 31.6 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 m1–m19 and models m1a-m18a). The simplest regression model (model m1) consisted of only the predictor variable juvenile pink salmon CPUE (X_1), 17 regression models (models m2-m18) consisted of the predictor variable juvenile pink salmon CPUE (X_1) and a temperature index (X_2), one regression model (model m19) consisted of the predictor variables juvenile pink salmon CPUE, the ISTI temperature variable, zooplankton density in June, and fish condition in July ($X_1 ... X_n$), one model consisted of predictor variable juvenile pink salmon CPUE (X_1) and the Southeast Alaska escapement index (X_2) (model m1a), and 17 models (models m2a-m18a) consisted of the predictor variable juvenile pink salmon CPUE (X_1), a temperature index (X_2), and the Southeast Alaska escapement index (X_3). The general models structure was

$$E(Y_i) = \hat{\alpha}_i + \hat{\beta}_{1_i} X_1 + ... \hat{\beta}_{n_i} X_n. \quad (1)$$

The temperature index for models m2-m18 and for models m2a-m18a 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 only contained CPUE (model m1), 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), CPUE data, and the Southeast Alaska pink salmon escapement index were natural log transformed in the model, but temperature data and zooplankton data were not. The forecast ($\hat{Y}_i$), and 80% prediction intervals (based on output from program R; R Core Team 2023) from the 37 regression models were exponentiated and bias-corrected (Miller 1984),

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

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 1; 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,

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

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

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

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. For the 10 year one-step-ahead MAPE for model m1, the process would be repeated, but the first forecast year would be 2012.

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. For example, model m3 and model m3a were compared to determine if the escapement index term explains more of the variation in the SEAK harvest, then just a model with CPUE and the Chatham Strait SST in May. The AICc value for model m3 was 26 and the AICc value for model m3a was 28. The difference between the two AICc values is 2 and model m3 has a lower AICc value. This suggests that the escapement index term does not explain an additional significant amount of the variation in the model, and the simpler model should be chosen (model m3).

Results

Based upon the 5-year MAPE, the best performing models were m2, m2a, and m19 and based upon the 10-year MAPE, the best performing models were m3, m3a, m7, m7a, m11, and m11a. When comparing the hierarchical models m2 to m2a, m3 to m3a, m7 to m7a, and m11 to m11a, the AICc value is lower for models without the escapement index term (m2, m3, m7, m11), the adjusted R-squared value is higher for models without the escapement index term (m2, m3, m7, m11), and the escapement index term is not significant in any of the four models (m2a, m3a, m7a, m11a) at the $p < 0.05$ significance level. Therefore, only models m2, m3, m7, m11, and m19 should be considered candidate models (Appendix C).

Based upon the 10-year MAPE performance metric, the AICc values, significant parameters in the models, and the adjusted R-squared values, model m11 (a model that included CPUE and the satellite SST variable from northern SEAK in May; Table 1 and Table 2; Appendix D) was the best performing model and the 2024-forecast using this model would be in the average range with a point estimate of 19.2 million fish (80% prediction interval: 11.7 to 31.6 million fish).

Table 1: Summary of the 5-year and 10-year one-step ahead MAPEs for the 37 regression models.

Model	Terms	5-year MAPE	10-year MAPE
m1	CPUE	51%	64%
m1a	CPUE + esc_index_log	46%	63%
m2	CPUE + ISTI20_MJJ	32%	39%
m2a	CPUE + ISTI20_MJJ + esc_index_log	32%	41%
m3	CPUE + Chatham_SST_May	38%	28%
m3a	CPUE + Chatham_SST_May + esc_index_log	36%	28%
m4	CPUE + Chatham_SST_MJJ	44%	42%
m4a	CPUE + Chatham_SST_MJJ + esc_index_log	42%	42%
m5	CPUE + Chatham_SST_AMJ	38%	31%
m5a	CPUE + Chatham_SST_AMJ + esc_index_log	37%	34%
m6	CPUE + Chatham_SST_AMJJ	41%	36%
m6a	CPUE + Chatham_SST_AMJJ + esc_index_log	39%	37%
m7	CPUE + Icy_Strait_SST_May	40%	27%
m7a	CPUE + Icy_Strait_SST_May + esc_index_log	37%	27%
m8	CPUE + Icy_Strait_SST_MJJ	44%	42%
m8a	CPUE + Icy_Strait_SST_MJJ + esc_index_log	41%	42%
m9	CPUE + Icy_Strait_SST_AMJ	41%	32%
m9a	CPUE + Icy_Strait_SST_AMJ + esc_index_log	39%	35%
m10	CPUE + Icy_Strait_SST_AMJJ	41%	37%
m10a	CPUE + Icy_Strait_SST_AMJJ + esc_index_log	39%	40%
m11	CPUE + NSEAK_SST_May	39%	27%
m11a	CPUE + NSEAK_SST_May + esc_index_log	36%	27%
m12	CPUE + NSEAK_SST_MJJ	43%	37%
m12a	CPUE + NSEAK_SST_MJJ + esc_index_log	41%	38%
m13	CPUE + NSEAK_SST_AMJ	39%	31%
m13a	CPUE + NSEAK_SST_AMJ + esc_index_log	38%	34%
m14	CPUE + NSEAK_SST_AMJJ	41%	32%
m14a	CPUE + NSEAK_SST_AMJJ + esc_index_log	39%	36%
m15	CPUE + SEAK_SST_May	41%	31%
m15a	CPUE + SEAK_SST_May + esc_index_log	39%	32%
m16	CPUE + SEAK_SST_MJJ	45%	40%
m16a	CPUE + SEAK_SST_MJJ + esc_index_log	42%	40%
m17	CPUE + SEAK_SST_AMJ	41%	34%
m17a	CPUE + SEAK_SST_AMJ + esc_index_log	40%	36%
m18	CPUE + SEAK_SST_AMJJ	43%	35%
m18a	CPUE + SEAK_SST_AMJJ + esc_index_log	41%	37%
m19	CPUE + ISTI20_MJJ + zoo_density_June + condition_July	28%	43%

Table 2: Summary of the adjusted R-squared values, and the AICc values for the 37 regression models.

Model	Terms	AdjR2	AICc	Change in AICc
m1	CPUE	54%	37	15
m1a	CPUE + esc_index_log	55%	38	17
m2	CPUE + ISTI20_MJJ	74%	24	2
m2a	CPUE + ISTI20_MJJ + esc_index_log	73%	27	5
m3	CPUE + Chatham_SST_May	72%	26	4
m3a	CPUE + Chatham_SST_May + esc_index_log	71%	28	7
m4	CPUE + Chatham_SST_MJJ	64%	32	11
m4a	CPUE + Chatham_SST_MJJ + esc_index_log	64%	34	13
m5	CPUE + Chatham_SST_AMJ	72%	26	4
m5a	CPUE + Chatham_SST_AMJ + esc_index_log	71%	28	7
m6	CPUE + Chatham_SST_AMJJ	68%	29	8
m6a	CPUE + Chatham_SST_AMJJ + esc_index_log	67%	32	10
m7	CPUE + Icy_Strait_SST_May	70%	27	6
m7a	CPUE + Icy_Strait_SST_May + esc_index_log	69%	30	8
m8	CPUE + Icy_Strait_SST_MJJ	64%	33	11
m8a	CPUE + Icy_Strait_SST_MJJ + esc_index_log	64%	35	13
m9	CPUE + Icy_Strait_SST_AMJ	69%	28	7
m9a	CPUE + Icy_Strait_SST_AMJ + esc_index_log	68%	31	10
m10	CPUE + Icy_Strait_SST_AMJJ	66%	31	9
m10a	CPUE + Icy_Strait_SST_AMJJ + esc_index_log	66%	33	11
m11	CPUE + NSEAK_SST_May	70%	28	6
m11a	CPUE + NSEAK_SST_May + esc_index_log	69%	30	9
m12	CPUE + NSEAK_SST_MJJ	63%	33	12
m12a	CPUE + NSEAK_SST_MJJ + esc_index_log	63%	35	14
m13	CPUE + NSEAK_SST_AMJ	70%	28	6
m13a	CPUE + NSEAK_SST_AMJ + esc_index_log	68%	31	9
m14	CPUE + NSEAK_SST_AMJJ	66%	31	10
m14a	CPUE + NSEAK_SST_AMJJ + esc_index_log	65%	34	12
m15	CPUE + SEAK_SST_May	68%	29	7
m15a	CPUE + SEAK_SST_May + esc_index_log	68%	32	10
m16	CPUE + SEAK_SST_MJJ	62%	34	12
m16a	CPUE + SEAK_SST_MJJ + esc_index_log	62%	36	14
m17	CPUE + SEAK_SST_AMJ	68%	29	8
m17a	CPUE + SEAK_SST_AMJ + esc_index_log	67%	32	11
m18	CPUE + SEAK_SST_AMJJ	64%	32	11
m18a	CPUE + SEAK_SST_AMJJ + esc_index_log	64%	35	13
m19	CPUE + ISTI20_MJJ + zoo_density_June + condition_July	80%	22	0

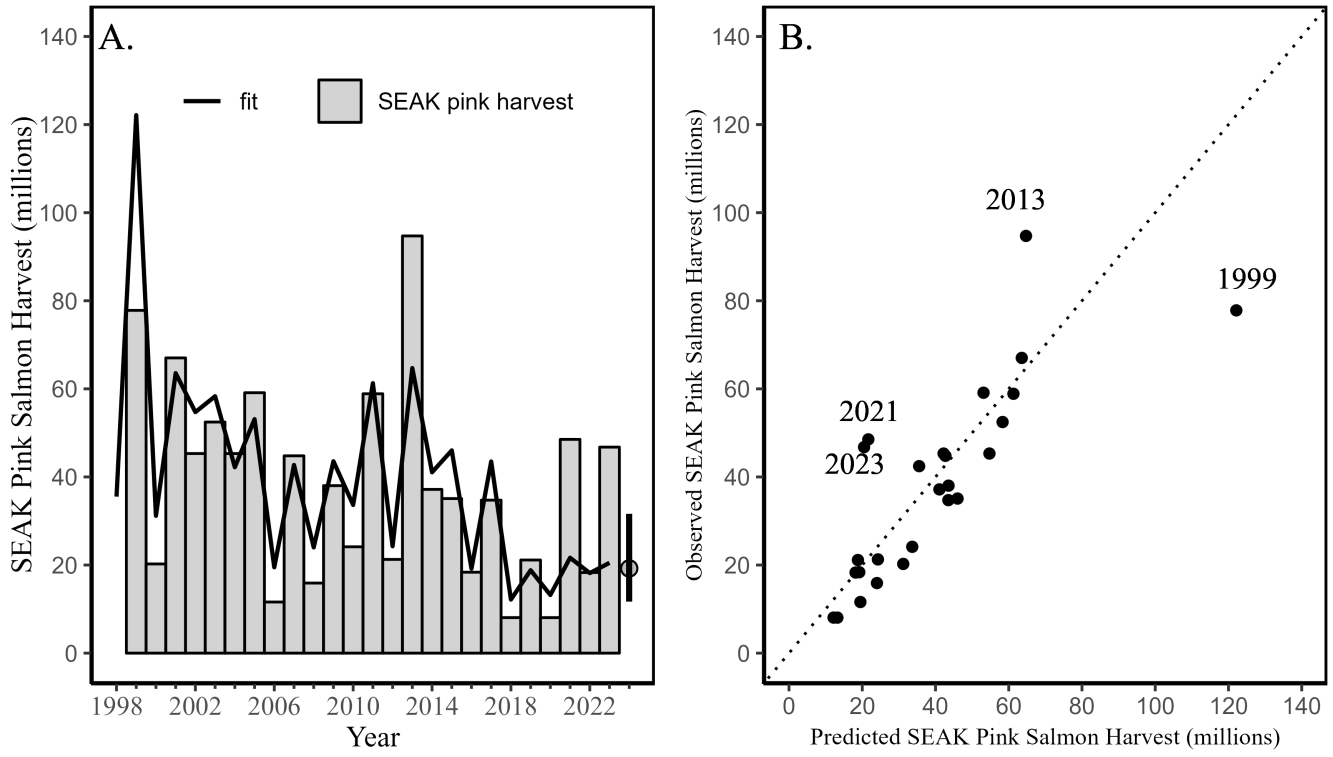

Figure 1: A. SEAK pink salmon harvest (millions) by year with the model fit (line) based upon the best performing model (model m11). The predicted 2024 forecast is symbolized as a grey circle with an 80% prediction interval (11.7 to 31.6 million fish). B. SEAK pink salmon harvest (millions) against the fitted values from model m11 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

Table 3: Annual adult pink salmon harvest data from Southeast Alaska (SEAK; millions of fish), the juvenile pink salmon CPUE data collected from the SECM project for years 1998-2024 (juvenile years 1997 to 2023), and the SEAK pink salmon escapement index.

Year	Juvenile year	Harvest	CPUE	Escapement index
1998	1997	42.45	2.48	18090337
1999	1998	77.82	5.62	14681062
2000	1999	20.25	1.60	14248275
2001	2000	67.02	3.73	27195944
2002	2001	45.32	2.87	10699025
2003	2002	52.47	2.78	18566341
2004	2003	45.31	3.08	16549670
2005	2004	59.12	3.90	19867278
2006	2005	11.61	2.04	15601572
2007	2006	44.80	2.58	19820824
2008	2007	15.91	1.17	10194617
2009	2008	38.02	2.32	17582000
2010	2009	24.14	2.33	9484485
2011	2010	58.88	4.11	12626432
2012	2011	21.28	1.51	11134757
2013	2012	94.72	3.52	14178881
2014	2013	37.17	2.14	11025094
2015	2014	35.09	3.80	25153162
2016	2015	18.37	2.45	13764123
2017	2016	34.73	4.35	12342523
2018	2017	8.07	0.35	10076959
2019	2018	21.14	1.17	13882487
2020	2019	8.06	1.14	8145577
2021	2020	48.53	2.15	8809780
2022	2021	18.30	0.88	9734280
2023	2022	46.77	1.45	15665251
2024	2023	NA	1.22	10032916

Table 4: Zooplankton density in June, and fish condition index in July for years 1998-2024 (juvenile years 1997 to 2023).

Year	Juvenile year	Zooplankton density (June)	Condition index (July)
1998	1997	1190	0.0438
1999	1998	392	-0.0247
2000	1999	802	-0.0330
2001	2000	1452	-0.0556
2002	2001	2083	-0.0026
2003	2002	1065	-0.0507
2004	2003	2736	-0.0078
2005	2004	2606	-0.0050
2006	2005	746	-0.0360
2007	2006	3120	-0.0231
2008	2007	2066	-0.0214
2009	2008	2749	-0.0172
2010	2009	3765	0.0219
2011	2010	1815	-0.0047
2012	2011	1357	-0.0143
2013	2012	1553	-0.0046
2014	2013	1025	-0.0046
2015	2014	1479	-0.0032
2016	2015	1651	0.0343
2017	2016	1422	0.0406
2018	2017	1959	0.0198
2019	2018	2024	-0.0891
2020	2019	1339	-0.0280
2021	2020	1520	-0.0749
2022	2021	1228	-0.0739
2023	2022	4744	-0.0696
2024	2023	358	0.1369

Variable definitions

CPUEcal: 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 calibration factors to account for differences in fishing power among vessels. The last time the CPUEcal variable was incorporated in the forecasting process was the 2023 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.

Condition: The average annual residuals derived from the regression of all paired Ln(weights) and Ln(lengths) for pink salmon collected during SECM sampling since 1997 in June and July. The last time the condition residuals were incorporated in the forecasting process was the 2019 forecast.

Energy Density: The average energy content (kJ/g dry weight, determined by bomb calorimetry) of subsamples of juvenile pink salmon captured in June or July of each year. The last time the energy density variables were incorporated in the forecasting process was the 2017 forecast (Wertheimer et al. 2018).

Zooplankton metric (**Average Zooplankton Total Water Column**; ml/m^3): The average May, June, or July 333-μm bongo net standing crop (displacement volume divided by water volume filtered, ml/m^3), and index of integrated mesozooplankton to 200-m depth (i.e., May, June, or July average zooplankton total water column). The last time the zooplankton metric variables were incorporated in the forecasting process was the 2017 forecast (Wertheimer et al. 2018).

Zooplankton metric (**zooplankton density**; $number/m^3$): The average density ($number/m^3$) of prey available in May, June, or July; an index computed from total density of six zooplankton taxa typically utilized by planktivorous juvenile salmon in summer (Sturdevant et al. 2004) and present in integrated 333-μm bongo net samples (June Preferred Prey). The last time the zooplankton metric variables were incorporated in the forecasting process was the 2015 forecast (Wertheimer et al. 2015).

North Pacific Index (**NPI**): June, July, August average of the NPI; a measure of atmospheric air pressure in the GOA thought to affect upwelling and downwelling oceanographic conditions (Trenberth and Hurrell 1994); higher values indicate a relaxation of downwelling along the Alaska coast adjacent to the eastern GOA and a widening of the Alaska Coastal Current. Source: https://climatedataguide.ucar.edu/climate-data/north-pacific-np-index-trenberth-and-hurrell-monthly-and-winter

Southeast Alaska pink salmon index (**esc_index_log**): Annual index of the pink salmon escapement in Southeast Alaska based on peak aerial survey counts (Heinl and Piston 2009; Piston and Heinl 2020).

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 5: Parameter estimates for the 37 individual models.

Model	Term	Estimate	Standard Error	Statistic	p value
m1	(Intercept)	2.4592202	0.201	12.224	0.000
m1	CPUE	0.3978377	0.072	5.534	0.000
m2	(Intercept)	7.3341828	1.126	6.516	0.000
m2	CPUE	0.4586351	0.056	8.186	0.000
m2	ISTI20_MJJ	-0.5552944	0.127	-4.371	0.000
m3	(Intercept)	5.6334942	0.807	6.978	0.000
m3	CPUE	0.4548245	0.058	7.828	0.000
m3	Chatham_SST_May	-0.4395394	0.110	-4.009	0.001
m4	(Intercept)	6.1837933	1.376	4.494	0.000
m4	CPUE	0.4153462	0.064	6.476	0.000
m4	Chatham_SST_MJJ	-0.3838292	0.141	-2.730	0.012
m5	(Intercept)	6.3638371	0.979	6.502	0.000
m5	CPUE	0.4421142	0.057	7.728	0.000
m5	Chatham_SST_AMJ	-0.5234331	0.130	-4.042	0.001
m6	(Intercept)	6.5068332	1.228	5.299	0.000
m6	CPUE	0.4262103	0.061	6.995	0.000
m6	Chatham_SST_AMJJ	-0.4718977	0.142	-3.328	0.003
m7	(Intercept)	5.2379040	0.767	6.829	0.000
m7	CPUE	0.4665508	0.061	7.650	0.000
m7	Icy_Strait_SST_May	-0.4100416	0.111	-3.707	0.001
m8	(Intercept)	5.8806523	1.284	4.579	0.000
m8	CPUE	0.4214650	0.065	6.520	0.000
m8	Icy_Strait_SST_MJJ	-0.3460245	0.129	-2.690	0.013
m9	(Intercept)	5.9243032	0.987	6.001	0.000
m9	CPUE	0.4506290	0.061	7.412	0.000
m9	Icy_Strait_SST_AMJ	-0.4815940	0.135	-3.560	0.002
m10	(Intercept)	6.0738051	1.179	5.151	0.000
m10	CPUE	0.4332171	0.063	6.907	0.000
m10	Icy_Strait_SST_AMJJ	-0.4226186	0.136	-3.099	0.005
m11	(Intercept)	5.2576601	0.775	6.783	0.000
m11	CPUE	0.4305538	0.059	7.316	0.000
m11	NSEAK_SST_May	-0.3838592	0.104	-3.693	0.001
m12	(Intercept)	5.7550290	1.294	4.447	0.000
m12	CPUE	0.3991158	0.065	6.167	0.000
m12	NSEAK_SST_MJJ	-0.3335425	0.130	-2.572	0.017
m13	(Intercept)	5.9214143	0.975	6.073	0.000
m13	CPUE	0.4200576	0.059	7.116	0.000
m13	NSEAK_SST_AMJ	-0.4646829	0.129	-3.603	0.002
m14	(Intercept)	6.0160325	1.198	5.020	0.000
m14	CPUE	0.4087033	0.062	6.554	0.000
m14	NSEAK_SST_AMJJ	-0.4119235	0.137	-3.000	0.006
m15	(Intercept)	5.2523123	0.834	6.300	0.000
m15	CPUE	0.4283586	0.060	7.086	0.000
m15	SEAK_SST_May	-0.3542807	0.104	-3.420	0.002
m16	(Intercept)	5.5832431	1.292	4.322	0.000
m16	CPUE	0.3936765	0.065	6.015	0.000
m16	SEAK_SST_MJJ	-0.2997873	0.123	-2.443	0.023
m17	(Intercept)	5.8273538	1.024	5.690	0.000

Model	Term	Estimate	Standard Error	Statistic	p value
m17	CPUE	0.4168743	0.061	6.884	0.000
m17	SEAK_SST_AMJ	-0.4192035	0.126	-3.334	0.003
m18	(Intercept)	5.8209504	1.217	4.782	0.000
m18	CPUE	0.4027926	0.063	6.345	0.000
m18	SEAK_SST_AMJJ	-0.3653761	0.131	-2.792	0.010
m19	(Intercept)	6.2047832	1.072	5.788	0.000
m19	CPUE	0.4977254	0.051	9.689	0.000
m19	ISTI20_MJJ	-0.4784279	0.116	-4.113	0.000
m19	zoo_density_June	0.0001485	0.000	2.384	0.027
m19	condition_July	-3.3019573	1.805	-1.829	0.082
m1a	(Intercept)	-3.7405005	5.170	-0.723	0.477
m1a	CPUE	0.3508431	0.081	4.316	0.000
m1a	esc_index_log	0.3841676	0.320	1.200	0.242
m2a	(Intercept)	6.8340093	4.820	1.418	0.170
m2a	CPUE	0.4546306	0.068	6.642	0.000
m2a	ISTI20_MJJ	-0.5503304	0.138	-3.990	0.001
m2a	esc_index_log	0.0282930	0.265	0.107	0.916
m3a	(Intercept)	2.3634188	4.438	0.533	0.600
m3a	CPUE	0.4288127	0.068	6.292	0.000
m3a	Chatham_SST_May	-0.4231968	0.113	-3.751	0.001
m3a	esc_index_log	0.1953179	0.261	0.750	0.461
m4a	(Intercept)	1.4726987	5.064	0.291	0.774
m4a	CPUE	0.3801715	0.074	5.150	0.000
m4a	Chatham_SST_MJJ	-0.3646485	0.142	-2.564	0.018
m4a	esc_index_log	0.2803911	0.290	0.967	0.344
m5a	(Intercept)	3.9623736	4.628	0.856	0.401
m5a	CPUE	0.4234106	0.068	6.231	0.000
m5a	Chatham_SST_AMJ	-0.5055950	0.136	-3.723	0.001
m5a	esc_index_log	0.1405620	0.265	0.531	0.601
m6a	(Intercept)	2.8209181	4.902	0.575	0.571
m6a	CPUE	0.3983786	0.071	5.600	0.000
m6a	Chatham_SST_AMJJ	-0.4498295	0.146	-3.085	0.005
m6a	esc_index_log	0.2166696	0.279	0.777	0.445
m7a	(Intercept)	2.1400687	4.598	0.465	0.646
m7a	CPUE	0.4411145	0.072	6.122	0.000
m7a	Icy_Strait_SST_May	-0.3932246	0.115	-3.432	0.002
m7a	esc_index_log	0.1848966	0.270	0.684	0.501
m8a	(Intercept)	1.1512981	5.041	0.228	0.821
m8a	CPUE	0.3857325	0.074	5.180	0.000
m8a	Icy_Strait_SST_MJJ	-0.3285492	0.130	-2.527	0.019
m8a	esc_index_log	0.2823487	0.291	0.970	0.342
m9a	(Intercept)	3.6136889	4.912	0.736	0.470
m9a	CPUE	0.4320714	0.073	5.926	0.000
m9a	Icy_Strait_SST_AMJ	-0.4626599	0.143	-3.232	0.004
m9a	esc_index_log	0.1347363	0.280	0.481	0.636
m10a	(Intercept)	2.2587101	4.977	0.454	0.654
m10a	CPUE	0.4038899	0.073	5.506	0.000
m10a	Icy_Strait_SST_AMJJ	-0.4009940	0.140	-2.860	0.009
m10a	esc_index_log	0.2249429	0.285	0.789	0.438
m11a	(Intercept)	2.0647281	4.594	0.449	0.658
m11a	CPUE	0.4058761	0.069	5.880	0.000
m11a	NSEAK_SST_May	-0.3680133	0.107	-3.424	0.002

Model	Term	Estimate	Standard Error	Statistic	p value
m11a	esc_index_log	0.1906927	0.270	0.705	0.488
m12a	(Intercept)	1.4142754	5.205	0.272	0.788
m12a	CPUE	0.3676940	0.075	4.928	0.000
m12a	NSEAK_SST_MJJ	-0.3126938	0.133	-2.357	0.028
m12a	esc_index_log	0.2562106	0.297	0.861	0.398
m13a	(Intercept)	3.5922921	4.877	0.737	0.469
m13a	CPUE	0.4025590	0.070	5.756	0.000
m13a	NSEAK_SST_AMJ	-0.4466243	0.136	-3.277	0.003
m13a	esc_index_log	0.1359874	0.279	0.488	0.631
m14a	(Intercept)	2.4787461	5.103	0.486	0.632
m14a	CPUE	0.3827770	0.073	5.261	0.000
m14a	NSEAK_SST_AMJJ	-0.3892305	0.142	-2.734	0.012
m14a	esc_index_log	0.2070472	0.290	0.714	0.483
m15a	(Intercept)	2.0908920	4.768	0.439	0.665
m15a	CPUE	0.4039667	0.071	5.683	0.000
m15a	SEAK_SST_May	-0.3380646	0.108	-3.143	0.005
m15a	esc_index_log	0.1879764	0.279	0.674	0.507
m16a	(Intercept)	1.1788849	5.261	0.224	0.825
m16a	CPUE	0.3621583	0.075	4.813	0.000
m16a	SEAK_SST_MJJ	-0.2796802	0.126	-2.227	0.036
m16a	esc_index_log	0.2599337	0.301	0.864	0.397
m17a	(Intercept)	3.1947002	5.004	0.638	0.530
m17a	CPUE	0.3972137	0.072	5.551	0.000
m17a	SEAK_SST_AMJ	-0.4001458	0.133	-3.019	0.006
m17a	esc_index_log	0.1536450	0.286	0.538	0.596
m18a	(Intercept)	2.1284865	5.200	0.409	0.686
m18a	CPUE	0.3760557	0.074	5.093	0.000
m18a	SEAK_SST_AMJJ	-0.3430846	0.136	-2.528	0.019
m18a	esc_index_log	0.2160957	0.296	0.731	0.473

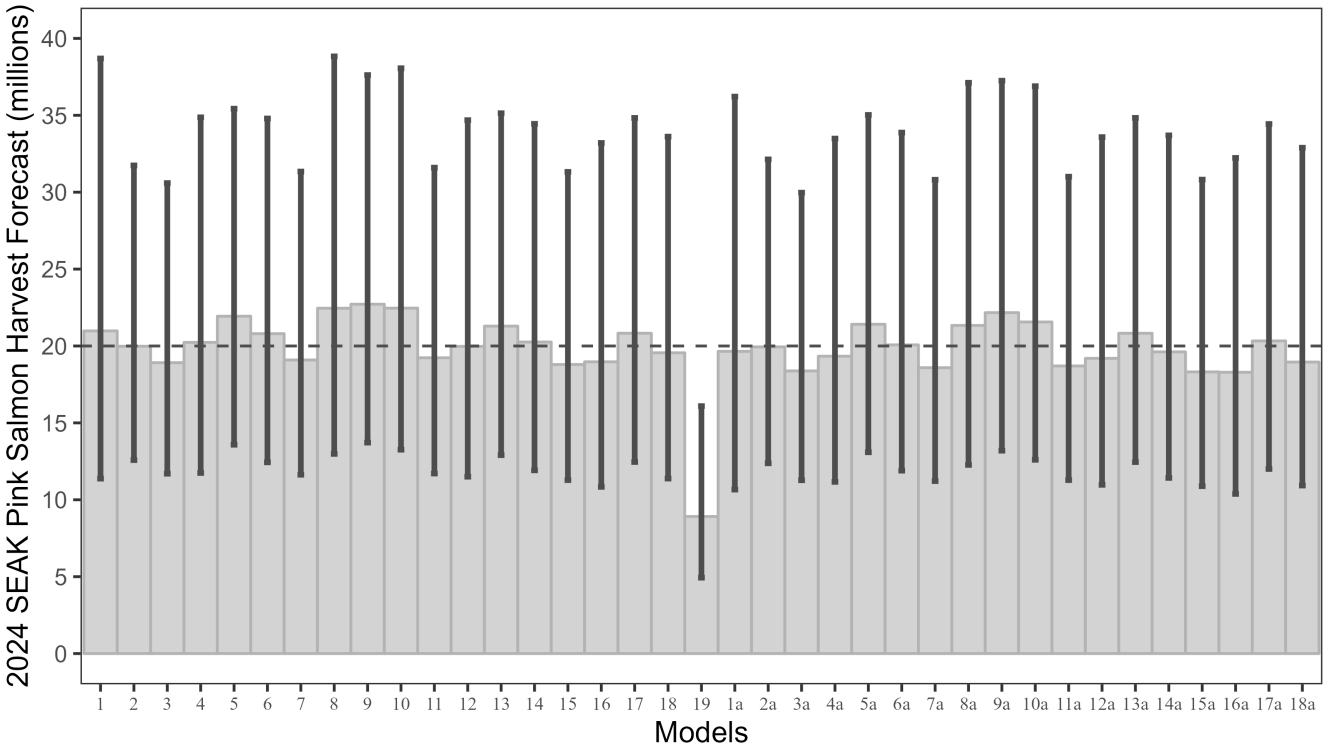

Figure B1: Bias-corrected forecasts (grey bars) for the 37 regression models with 80% prediction intervals (vertical grey lines). Based upon the performance metrics, the best performing models were model m2, m3, m7, m11, and m19. A horizontal dotted line at 20,00,000 fish is placed on the figure for reference only. The 2024-forecast using the model m11 would be in the average range with a point estimate of 19.2 million fish (80% prediction interval: 11.7 to 31.6 million fish).

Table 6: Forecasts (in millions of fish; with 80% prediction intervals; Lower and Upper columns) for the Southeast Alaska pink salmon commercial harvest in 2024 based on the 37 individual models.

Model	Terms	Fit	Lower	Upper
m1	CPUE	21	11	39
m5	CPUE + Chatham_SST_AMJ	22	14	35
m5a	CPUE + Chatham_SST_AMJ + esc_index_log	21	13	35
m6	CPUE + Chatham_SST_AMJJ	21	12	35
m6a	CPUE + Chatham_SST_AMJJ + esc_index_log	20	12	34
m3	CPUE + Chatham_SST_May	19	12	31
m3a	CPUE + Chatham_SST_May + esc_index_log	18	11	30
m4	CPUE + Chatham_SST_MJJ	20	12	35
m4a	CPUE + Chatham_SST_MJJ + esc_index_log	19	11	33
m1a	CPUE + esc_index_log	20	11	36
m9	CPUE + Icy_Strait_SST_AMJ	23	14	38
m9a	CPUE + Icy_Strait_SST_AMJ + esc_index_log	22	13	37
m10	CPUE + Icy_Strait_SST_AMJJ	22	13	38
m10a	CPUE + Icy_Strait_SST_AMJJ + esc_index_log	22	13	37
m7	CPUE + Icy_Strait_SST_May	19	12	31
m7a	CPUE + Icy_Strait_SST_May + esc_index_log	19	11	31
m8	CPUE + Icy_Strait_SST_MJJ	22	13	39
m8a	CPUE + Icy_Strait_SST_MJJ + esc_index_log	21	12	37
m2	CPUE + ISTI20_MJJ	20	13	32
m2a	CPUE + ISTI20_MJJ + esc_index_log	20	12	32
m19	CPUE + ISTI20_MJJ + zoo_density_June + condition_July	9	5	16
m13	CPUE + NSEAK_SST_AMJ	21	13	35
m13a	CPUE + NSEAK_SST_AMJ + esc_index_log	21	12	35
m14	CPUE + NSEAK_SST_AMJJ	20	12	34
m14a	CPUE + NSEAK_SST_AMJJ + esc_index_log	20	11	34
m11	CPUE + NSEAK_SST_May	19	12	32
m11a	CPUE + NSEAK_SST_May + esc_index_log	19	11	31
m12	CPUE + NSEAK_SST_MJJ	20	12	35
m12a	CPUE + NSEAK_SST_MJJ + esc_index_log	19	11	34
m17	CPUE + SEAK_SST_AMJ	21	12	35
m17a	CPUE + SEAK_SST_AMJ + esc_index_log	20	12	34
m18	CPUE + SEAK_SST_AMJJ	20	11	34
m18a	CPUE + SEAK_SST_AMJJ + esc_index_log	19	11	33
m15	CPUE + SEAK_SST_May	19	11	31
m15a	CPUE + SEAK_SST_May + esc_index_log	18	11	31
m16	CPUE + SEAK_SST_MJJ	19	11	33
m16a	CPUE + SEAK_SST_MJJ + esc_index_log	18	10	32

Appendix C

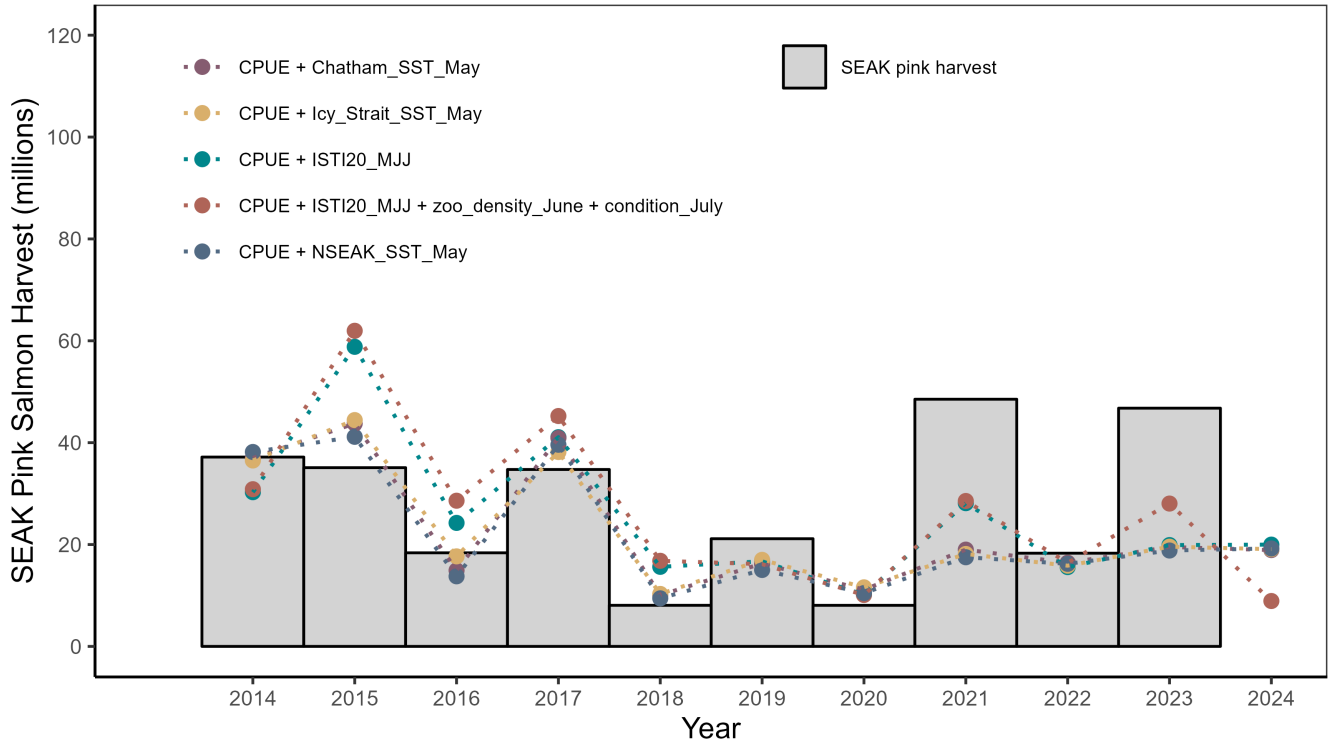

Figure C1: Southeast pink salmon harvest (millions of fish) by year with one step ahead forecasts based upon the the best performing models using the 5-year and 10-year MAPEs: models m2, m3, m7, m11, and m19.

Appendix D

Model Diagnostics

Model diagnostics for model m11 included residual plots, the curvature test, and influential observation diagnostics using Cook's distance (Cook 1977), the Bonferroni outlier test, and leverage plots (Table 7; Figure D1; Figure D2). Model diagnostics were used to identify observations that were potential outliers, had high leverage, or were influential (Zhang 2016).

Table 7: Detailed output for model m11. Juvenile years 1998, 2005, 2012, 2019, 2020, and 2022 (years 1999, 2006, 2013, 2020, 2021, 2023) show the largest standardized residual (Std. residuals). Fitted values (in millions of fish) are bias-corrected.

Year	Harvest	Residuals	Hat values	Cooks distance	Std. residuals	Fitted values
1998	42.45	0.24	0.04	0.01	0.69	35.53
1999	77.82	-0.39	0.29	0.22	-1.26	122.14
2000	20.25	-0.37	0.10	0.04	-1.07	31.19
2001	67.02	0.12	0.09	0.00	0.34	63.56
2002	45.32	-0.12	0.11	0.01	-0.36	54.73
2003	52.47	-0.04	0.15	0.00	-0.12	58.32
2004	45.31	0.14	0.05	0.00	0.39	42.21
2005	59.12	0.17	0.09	0.01	0.50	53.11
2006	11.61	-0.45	0.12	0.08	-1.34	19.49
2007	44.80	0.11	0.06	0.00	0.32	42.71
2008	15.91	-0.35	0.11	0.04	-1.01	24.01
2009	38.02	-0.07	0.10	0.00	-0.21	43.56
2010	24.14	-0.27	0.04	0.01	-0.75	33.65
2011	58.88	0.03	0.11	0.00	0.07	61.28
2012	21.28	-0.07	0.07	0.00	-0.19	24.27
2013	94.72	0.45	0.10	0.06	1.30	64.71
2014	37.17	-0.03	0.11	0.00	-0.10	41.07
2015	35.09	-0.21	0.11	0.01	-0.60	46.02
2016	18.37	0.02	0.21	0.00	0.07	19.19
2017	34.73	-0.16	0.26	0.03	-0.51	43.51
2018	8.07	-0.35	0.18	0.08	-1.05	12.17
2019	21.14	0.18	0.09	0.01	0.53	18.82
2020	8.06	-0.43	0.18	0.12	-1.31	13.20
2021	48.53	0.87	0.09	0.22	2.54	21.65
2022	18.30	0.07	0.11	0.00	0.21	18.18
2023	46.77	0.89	0.07	0.17	2.56	20.50

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 0.18 was used; observations with a Cook's distance greater than 0.18 may be influential observations (Figure D1a).

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., 3), and n is the number of observations in the model (i.e., 25); $p/n = 3/25 = 0.12$ for this study). Therefore, a leverage cut-off of 0.24 was used; observations with a leverage value greater than 0.24 may affect the model properties (e.g., summary statistics, standard errors, predicted values) (Figure D1b).

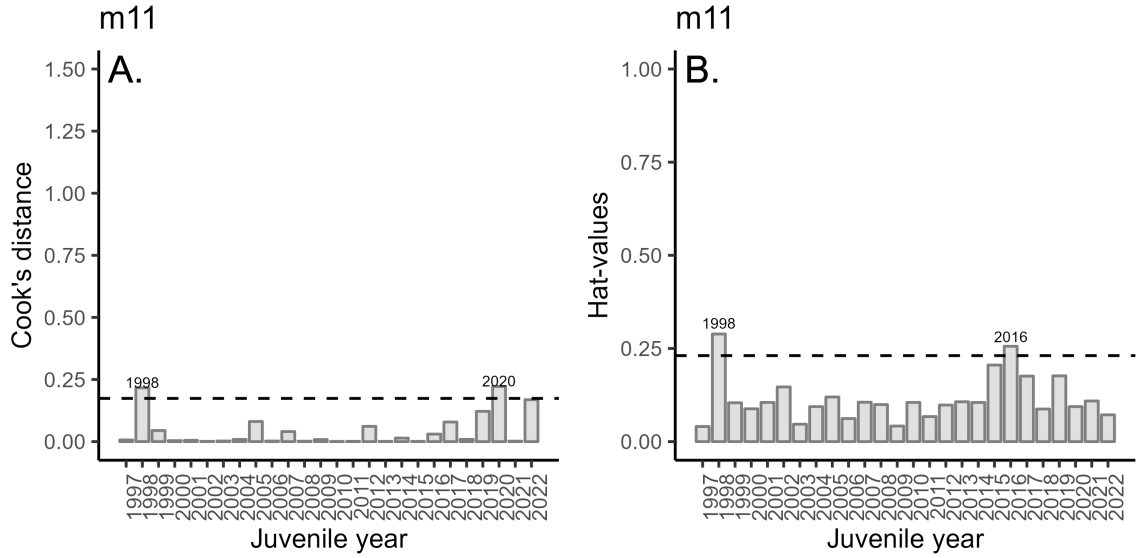

Figure D1: Diagnostics plots of influential observations including A. Cook's distance (with a cut-off value of 0.18), and B. leverage values (with a cut-off value of 0.24) from model m11.

Influential datapoints

To determine if a variable has a relationship with residuals, a lack-of fit curvature test was performed. In this test, terms that are non-significant suggest a properly specified model. No terms were significant in the lack-of-fit curvature test (P<0.05) (Figure D2a; Figure D2b). Diagnostics indicated that two of the data points were above the cut-off value for the Cook's distance (Figure D1a; 1998 and 2020). Two observations had high leverage values (Figure D1b; 1998 and 2016). 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 observations 2, 9, 16, 23, 24, and 26 were the most extreme (juvenile years 1998, 2005, 2012, 2019, 2020, and 2022) based on standardized residuals (Figure D2c; Table 7). Based on the lightly curved fitted lines in the residual versus fitted plot (Figure D2d), the fitted plot shows some lack of fit of the model.

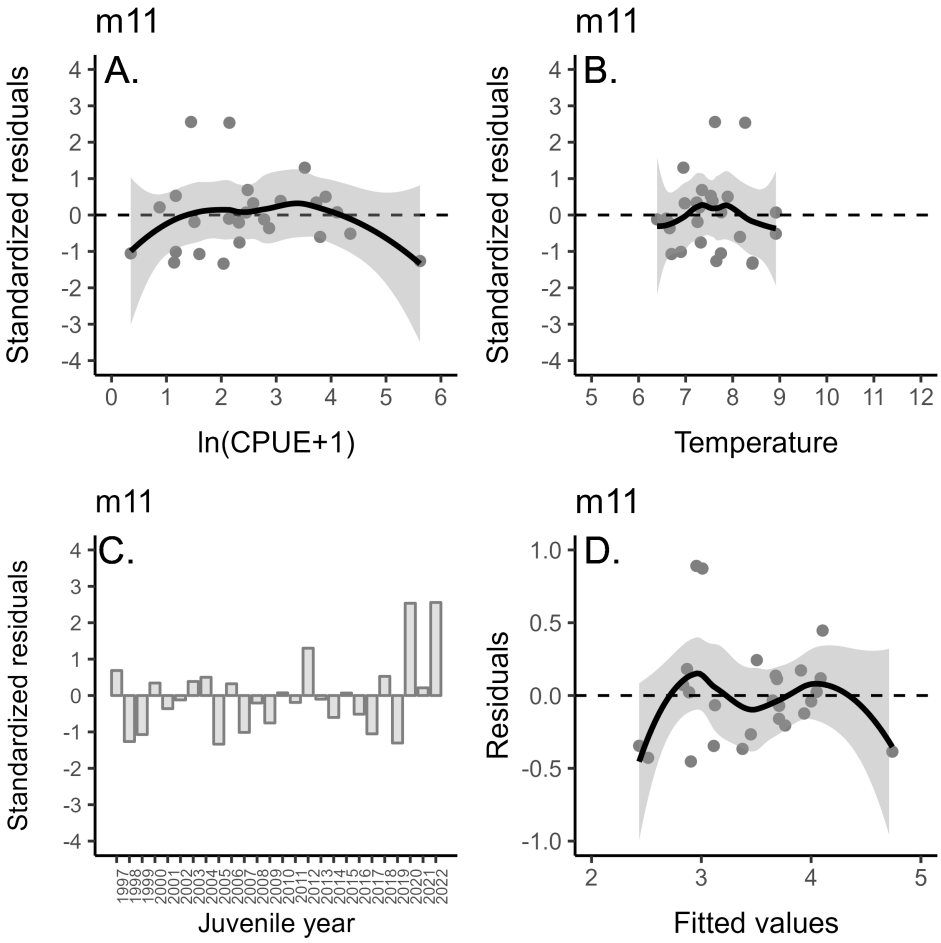

Figure D2: Standardized residuals versus predicted plots for A. CPUE and B. temperature (average May SST in northern Southeast Alaska) for model m11. C. Standardized residuals versus juvenile year and D. residuals versus fitted values for model m11. Positive residuals indicate that the observed harvest was larger than predicted by the model.