2021 Preseason Pink Salmon Forecast

pink_cal_mixspecies

Sara Miller, Rich Brenner, Jim Murphy

November 3, 2020–draft

Objective

To forecast the Southeast Alaska (SEAK) pink salmon harvest in 2021.

Executive Summary

Forecasts were developed using an approach described in Murphy et al. 2019. A multiple regression model was developed using monthly peak juvenile pink salmon CPUE (standardized catch based on 20 minute trawl set) for the June and July surveys and associated environmental parameters. The model used was:

$$E(y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon,$$

where y is ln(harvest), β_1 is the coefficient for ln(CPUE+1), β_2 is the coefficient for the environmental covariate water temperature (i.e., May through July average temperature in the upper 20 m at eight stations in Icy Strait;'ISTI'), β_3 is the interaction term, and ϵ represents the normal error term that is lognormal.

Leave-one-out cross validation (hindcast) and model performance metrics such as Mean and Median Absolute Percentage Error (MAPE, MEAPE), and mean absolute scaled error (MASE) (Hyndman and Kohler 2006) were then used to evaluate forecast accuracy of alternative models. Statistical analyses were performed with the R Project for Statistical computing version 3.6.3 (R Core Team 2020).

Based on an additive model with CPUE and temperature, the SEAK pink salmon harvest in 2021 is predicted to be in the moderate range with a point estimate of 31.3 million fish (80% prediction interval: 20.9 to 46.9 million fish).

Analysis

Three hierarchical models were investigated. The full model was:

$$E(y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2,$$

where X_1 is ln(CPUE+1) and X_2 is the average temperature in Icy Strait in May, June, and July at eight stations in Icy Strait (Icy Strait and Upper Chatham transects), and β_3 is the interaction term between CPUE and the temperature indice. In the past, the ISTI variable was the average temperature in the upper 20 m during May through August at eight stations in Icy Strait (Icy Strait and Upper Chatham transects).The regression coefficients CPUE and temperature (ISTI) were significant in the first two models (m1, m2). The interaction term was not significant in model m3 (Table 1). Therefore, only the first two models will be considered further.

Table 1: Parameter estimates for the three potential models.

model	term	estimate	std.error	statistic	p.value
m1	(Intercept)	2.264	0.206	11.014	0.000
m1	CPUE	0.464	0.073	6.349	0.000
m2	(Intercept)	6.635	0.993	6.679	0.000
m2	CPUE	0.519	0.055	9.526	0.000
m2	ISTI	-0.496	0.111	-4.451	0.000
m3	(Intercept)	4.162	2.714	1.534	0.142
m3	CPUE	1.510	1.013	1.491	0.152
m3	ISTI	-0.226	0.298	-0.759	0.457
m3	CPUE:ISTI	-0.108	0.110	-0.979	0.340

The model summary results using the metrics AICc, MAPE, MEAPE, and MASE (Hyndman and Kohler 2006) are shown in Table 2. For all of these metrics, the smallest value is the preferred model. The difference (Δ_i) between a given model and the model with the lowest AICc value and the metric MASE were the primary statistics for choosing appropriate models in this analysis. Models with AICc$\Delta_i \leq 2$ have substantial support, those in which $4 \leq$ AICc$\Delta_i \leq 7$ have considerably less support, and models with AICcΔ_i > 10 have essentially no support (Burnham and Anderson 2004). These two metrics (AICc, MASE) suggest that model m2 is the preferred models. If temperature is actually altering how CPUE is related to abundance it makes sense to restrict the temperature data to the CPUE months in the forecast model (June and July). The month of May is included as there are important migratory dynamics prior to the time juveniles are actually sampled in Icy Strait. Model m2 (based on average temperature in May through July) was used to forecast the 2021 pink salmon harvest.

Table 2: Summary of model outputs and forecast error measures

model	AdjR2	AICc	MAPE	MEAPE	MASE
CPUE	0.641	29	0.109	0.089	0.375
CPUE+ISTI	0.811	16	0.079	0.061	0.262

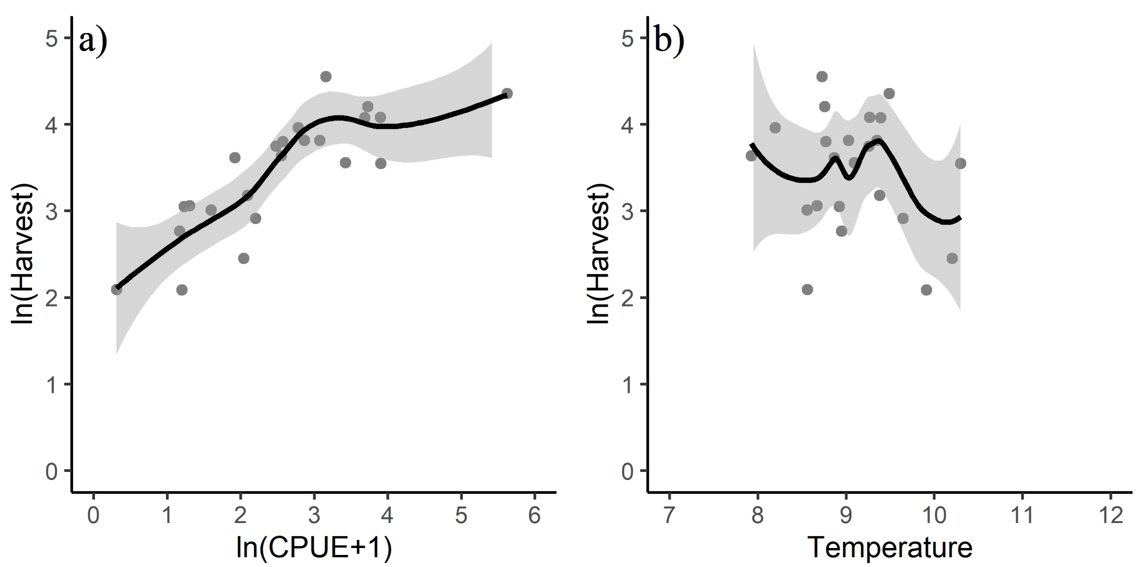

Figure 1: Relationship between a) ln(CPUE+1) and ln(harvest) and b) temperature in May through July (ISTI) and ln(harvest).

Results

The best regression model based on AICc, the MASE metric, significant coefficients in the models, and the argument for restricting temperature to the months when CPUE is sampled was model m2 (i.e. the model containing CPUE, and a May through July temperature variable). The adjusted R^2 value for model m2 was 0.81 indicating overall a good model fit.

Conclusion

The SEAK pink salmon harvest in 2021 is predicted to be in the moderate range with a point estimate of 31.3 million fish (80% prediction interval: 20.9 to 46.9 million fish).

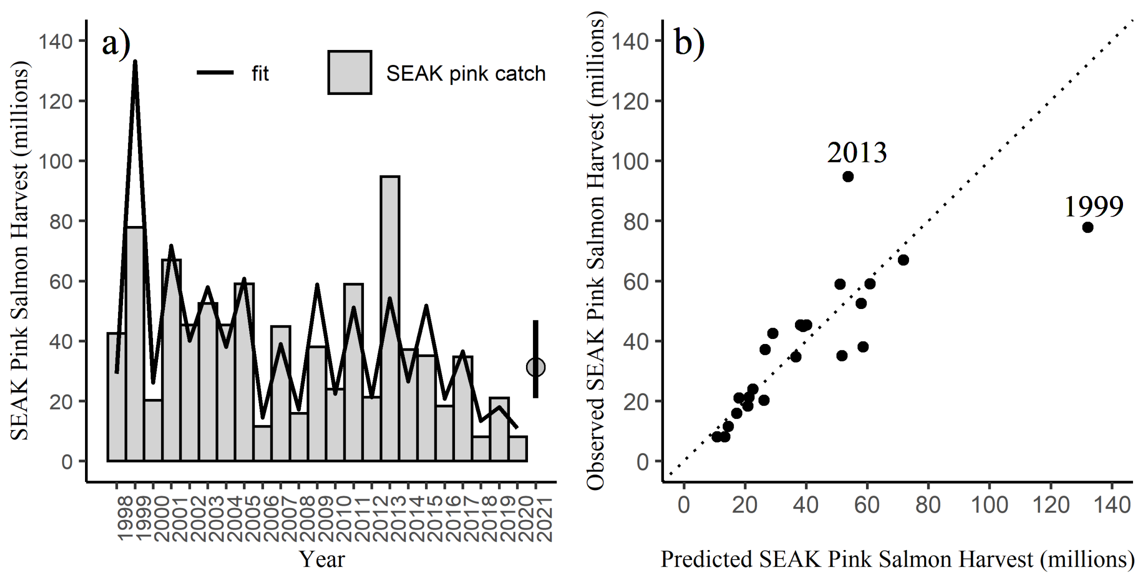

Figure 2: SEAK harvest (millions) a) by year and b) by the fitted values from model m2. The line in figure b is a one to one line. The predicted 2021 forecast is symbolized as a grey circle with an 80% prediction interval (20.9 to 46.9 million fish) in figure a.

Table 3: Detailed output for model m2. Juvenile year 2012 (year 2013) shows the largest standardized residual.

year	SEAKCatch_log	CPUE	resid	hat_values	Cooks_distance	std_resid	fitted
1998	3.750	2.477744	0.421	0.048	0.035	1.455	3.328
1999	4.354	5.622380	-0.493	0.344	0.735	-2.051	4.847
2000	3.011	1.597723	-0.208	0.099	0.020	-0.737	3.218
2001	4.205	3.729985	-0.024	0.119	0.000	-0.085	4.228
2002	3.813	2.868826	0.167	0.048	0.006	0.577	3.646
2003	3.961	2.784664	-0.053	0.165	0.003	-0.197	4.014
2004	3.813	3.077820	0.218	0.057	0.012	0.757	3.595
2005	4.079	3.899407	0.018	0.101	0.000	0.062	4.062
2006	2.451	2.040345	-0.182	0.243	0.053	-0.704	2.633
2007	3.802	2.572781	0.183	0.059	0.008	0.634	3.620
2008	2.766	1.167639	-0.035	0.105	0.001	-0.123	2.801
2009	3.638	2.555111	-0.393	0.237	0.238	-1.515	4.031
2010	3.178	2.094192	0.108	0.066	0.003	0.377	3.070

year	SEAKCatch_log	CPUE	resid	hat_values	Cooks_distance	std_resid	fitted
2011	4.076	3.687782	0.187	0.088	0.014	0.659	3.889
2012	3.059	1.305918	0.046	0.105	0.001	0.165	3.012
2013	4.551	3.161036	0.603	0.082	0.134	2.119	3.948
2014	3.616	1.923429	0.381	0.060	0.037	1.322	3.236
2015	3.558	3.426619	-0.345	0.069	0.036	-1.205	3.903
2016	2.912	2.201588	-0.081	0.096	0.003	-0.287	2.993
2017	3.547	3.905705	-0.008	0.257	0.000	-0.031	3.555
2018	2.092	0.310436	-0.458	0.217	0.280	-1.742	2.550
2019	3.049	1.233765	0.201	0.100	0.019	0.714	2.848
2020	2.088	1.202606	-0.255	0.234	0.098	-0.980	2.343

References

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

Hyndman, R. J. and A. B. Koehler. 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22: 679-688.

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 Commisson Technical Report No. 15: 75-81.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.r-project.org/index.html