Model-Averaged Predictions

Sara Miller

July 22, 2021

Contents

1 Objective

To determine the weighting and variance for the model averaging approach which will be applied to the 2022 preseason Southeast Alaska pink salmon harvest forecast. As data was not available, the examples and methods in this report are based on data used for the 2021 preseason Southeast Alaska pink salmon harvest forecast. Two issues still remain:

1. How should the model-averaged forecast be weighted by *MAPE_one_step_ahead* because the lower the *MAPE_one_step_ahead*, the better the model forecast performance (see *Model averaging (multi-model inference)* section below).

2. Is the variance (and 80% prediction interval) for the model-averaged forecast calculated corrected in the excel template (see step 3 in *Model averaging (multi-model inference)* section below)?

2 Model 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). We used a similar approach to Murphy et al. (2019), but assumed a log-normal error. This approach is based on a multiple regression model with juvenile pink salmon catch-per-unit-effort (CPUE) and temperature data from the Southeast Alaska Coastal Monitoring Survey (SECM; Piston et al. 2021) or satellite sea surface temperature data (SST and SST Anomaly, NOAA Global Coral Bleaching Monitoring, 5km, V.3.1, Monthly, 1985-Present' time series (https://coastwatch.pfeg.noaa.gov/erddap/griddap/NOAA_DHW_monthly.html). Based on prior discussions, the index of juvenile abundance (i.e., CPUE) was based on the pooled-species vessel calibration coefficient.

2.1 Hierarchical models

Forty nine hierarchical models were investigated (i.e., 24 additive models, CPUE only model, and 24 interaction models). 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 the average CPUE for catches in either the June or July survey, whichever month had the highest average catches in a given year, and was based on the pooled-species vessel calibration coefficient, X_2 is a temperature index, and β_3 is the interaction term between CPUE and the temperature index. The CPUE data were log-transformed in the model, but temperature data were not. The simplest model did not contain a temperature variable (model m1). None of the interactions were significant at $\alpha <= 0.05$; therefore only additive models (25 models; Figure 1 and Table 1) were considered further.

2.2 Performance metrics

The performance metrics used to assess the forecast models were:

1. Akaike Information Criterion corrected for small sample sizes (AICc values; Akaike 1973; Burnham and Anderson 2004);
2. the mean absolute scaled error (*MASE* metric; Hyndman and Kohler 2006);
3. the weighted mean absolute percentage error (*wMAPE*);
4. leave one out cross validation *MAPE* (*MAPE_LOOCV*); and
5. one step ahead forecasts (*MAPE_one_step_ahead*) (Table 2).

For all of these metrics, the smallest value is the preferred model. Models with Δ_iAICc ≤ 2 have substantial support, those in which $4 \leq \Delta_i$AICc ≤ 7 have considerably less support, and models with Δ_iAICc > 10 have essentially no support (Burnham and Anderson 2004). The performance metric *MAPE* was calculated as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}|\frac{A_t - F_t}{A_t}|$$

where A_t is the observed value and F_t is the predicted value at time t. The performance metric *wMAPE* was calculated as:

$$wMAPE = \sum_{t=1}^{n}\frac{1}{w_t}\sum_{t=1}^{n}|\frac{A_t - F_t}{A_t}|w_t$$

where w_t is the weight for each year. For the *wMAPE* metric, the last 5 years (juvenile years 2015-2019) were given a weight of 1 and all other years, a weight of 0.001. Therefore, compared to the performance metric *MAPE_LOOCV*, the performance of the model in the last 5 years was given more weight in the *wMAPE* metric.

The performance metric *MAPE_LOOCV* was calculated using five steps.

1. The data set is split into a training set. The training set uses all but one observation of the full data set.
2. The regression model is run based on the training set.

3. The regression model, based on the training set, is used to predict F_t for the one observation left out of the model.

4. The process is repeated n times based on the number of observations in the data set, leaving out a different observation from the training set each time.

5. The MAPE is calculated based on the average of all the training data sets (i.e., one MAPE is calculated for each training set and then these are averaged).

The performance metric *MAPE_one_step_ahead* was calculated using three steps.

1. Estimate the regression parameters at time t from data up to time $t-1$.

2. Make a prediction of F_t at time t based on the predictor variables at time t and the estimate of the regression parameters at time t.

3. Calculate the MAPE based on the prediction of F_t at time t and the observed value of A_t at time t.

4. The *MAPE_one_step_ahead* will then be an average of the MAPE calculated from data up through juvenile year 2013 (e.g., juvenile year 2013 is $t-1$ and the forecast is for juvenile year 2014; t), data up through juvenile year 2014, data up through juvenile year 2015, data up through juvenile year 2016, data up through juvenile year 2017, data up through juvenile year 2018, and data up through juvenile year 2019. Therefore, the MAPE from six years was averaged for the *MAPE_one_step_ahead* performance metric.

Table 1: Summary of model forecasts including the 80 percent prediction intervals (corrected for log transformation bias in a linear-model).

model	terms	fit	fit_LPI	fit_UPI
m1	CPUE	27.568	15.659	48.536
m2	CPUE + ISTI3_May	23.986	15.670	36.716
m3	CPUE + ISTI10_May	23.592	15.619	35.635
m4	CPUE + ISTI15_May	23.599	15.731	35.403
m5	CPUE + ISTI20_May	23.994	16.111	35.736
m6	CPUE + ISTI3_MJJ	31.465	19.960	49.603
m7	CPUE + ISTI10_MJJ	30.411	19.990	46.264
m8	CPUE + ISTI15_MJJ	29.191	19.709	43.234
m9	CPUE + ISTI20_MJJ	28.281	19.461	41.097
m10	CPUE + Chatham_SST_May	19.015	13.628	26.531
m11	CPUE + Chatham_SST_MJJ	23.573	15.463	35.937
m12	CPUE + Chatham_SST_AMJ	22.884	15.939	32.855
m13	CPUE + Chatham_SST_AMJJ	23.972	16.206	35.461
m14	CPUE + Icy_Strait_SST_May	18.218	12.872	25.783
m15	CPUE + Icy_Strait_SST_MJJ	22.064	14.389	33.833
m16	CPUE + Icy_Strait_SST_AMJ	20.319	13.911	29.677
m17	CPUE + Icy_Strait_SST_AMJJ	21.979	14.654	32.966
m18	CPUE + NSEAK_SST_May	17.566	12.575	24.537
m19	CPUE + NSEAK_SST_MJJ	21.710	14.479	32.553
m20	CPUE + NSEAK_SST_AMJ	20.686	14.464	29.585
m21	CPUE + NSEAK_SST_AMJJ	21.877	14.876	32.173
m22	CPUE + SEAK_SST_May	17.680	12.328	25.356
m23	CPUE + SEAK_SST_MJJ	22.540	14.750	34.445
m24	CPUE + SEAK_SST_AMJ	20.924	14.365	30.478

model	terms	fit	fit_LPI	fit_UPI
m25	CPUE + SEAK_SST_AMJJ	22.275	14.856	33.399

Table 2: Summary of model outputs and forecast error measures. These metrics included Akaike Information Criterion corrected for small sample sizes (AICc values), the mean absolute scaled error (MASE metric), the weighted mean absolute percentage error (wMAPE; based on the last 5 years), leave one out cross validation MAPE (MAPE_LOOCV), and one step ahead forecasts (MAPE_one_step_ahead).

model	AdjR2	AICc	MASE	wMAPE	MAPE_LOOCV	MAPE_one_step_ahead
m1	0.627	30.21	0.399	0.184	0.116	0.208
m2	0.790	18.77	0.284	0.134	0.087	0.145
m3	0.803	17.28	0.276	0.128	0.084	0.139
m4	0.810	16.52	0.278	0.119	0.084	0.131
m5	0.816	15.74	0.274	0.111	0.082	0.121
m6	0.763	21.57	0.308	0.130	0.094	0.160
m7	0.797	17.99	0.286	0.119	0.087	0.149
m8	0.821	15.08	0.263	0.105	0.079	0.135
m9	0.838	12.85	0.246	0.092	0.073	0.122
m10	0.876	6.76	0.234	0.074	0.069	0.084
m11	0.795	18.29	0.300	0.093	0.089	0.124
m12	0.849	11.22	0.261	0.072	0.076	0.090
m13	0.822	14.95	0.274	0.069	0.079	0.097
m14	0.866	8.40	0.212	0.064	0.062	0.073
m15	0.791	18.66	0.305	0.093	0.090	0.123
m16	0.838	12.87	0.255	0.071	0.075	0.092
m17	0.812	16.26	0.292	0.079	0.085	0.105
m18	0.877	6.47	0.224	0.070	0.066	0.078
m19	0.813	16.18	0.286	0.078	0.084	0.100
m20	0.854	10.39	0.250	0.066	0.073	0.088
m21	0.830	14.00	0.270	0.063	0.078	0.081
m22	0.858	9.87	0.246	0.083	0.074	0.095
m23	0.794	18.39	0.297	0.088	0.088	0.110
m24	0.839	12.68	0.262	0.073	0.077	0.094
m25	0.812	16.28	0.281	0.073	0.082	0.090

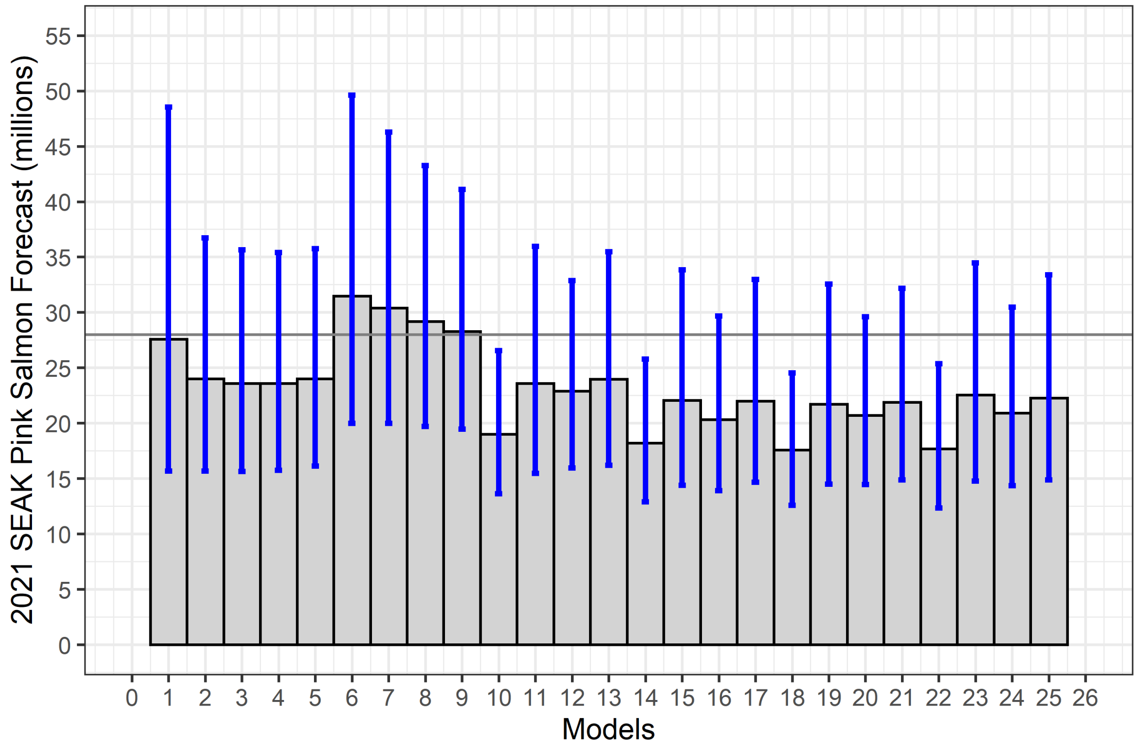

Figure 1: The 2021 SEAK pink salmon harvest (millions) forecast by model. The 80% prediction intervals (corrected for log transformation bias in a linear-model) around each forecast were calculated using the car package (Fox and Weisberg 2019) in program R (R Core Team 2020). The SEAK pink salmon harvest in 2021 (based on the November 18, 2020 advisory announcement) was a point estimate of 28 million fish (80% prediction interval: 19–42 million fish; grey horizontal line).

2.3 Model averaging (multi-model inference)

The process to weight the model-averaged forecast of Southeast Alaska pink salmon harvest in 2021, $\hat{\theta}$, and calculate the prediction interval around the model-averaged forecast is as follows:

1. Determine the contending models, $M_k, k = 1, ..., K$, that will be weighted in the model-averaged forecast. In this case, the 25 additive models were included as contending models.

2. Determine weights associated with the estimates, w_k, derived from each contending model and scaled so that $\sum w_k = 1$ (see *Model weighting approaches*). These weights are then multiplied by the estimate of the parameter $\hat{\theta}_k$, an estimate of θ (i.e., the harvest forecast for 2021) under model k (corrected for log transformation bias in a linear-model),

$$\hat{\theta} = \sum_k w_k \hat{\theta}_k.$$

Log transformation bias in a linear model is corrected by (Miller 1984),

$$\hat{E}(Y) = \exp(\hat{\alpha} + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \frac{\hat{\sigma}^2}{2}).$$

3. The prediction interval around the model-averaged forecast, $\hat{\theta}$, is then based on equation 9 in Buckland et al. 1997 (derivation in Burnham and Anderson 2002:159-162), where the variance of the model-

averaged prediction $\hat{\theta}$, is

$$\mathrm{var}(\hat{\theta}) = \left(\sum_{k=1}^{K} w_k \sqrt{\mathrm{var}(\hat{\theta}_k) + \gamma_k^2} \right)^2 ,$$

and where $\gamma_k = \hat{\theta}_k - \hat{\theta}$. Although this approach has been criticized (Claeskens and Hjort 2008:207), it has been shown to work well in simulations (Lukacs et al. 2010; Fletcher and Dillingham 2011).

2.3.1 Model weighting approaches

Approaches to estimating model-averaged weights fall into four broad categories (Dormann et al. 2018).

a. Bayesian approach (i.e., model weights are probabilities sampled from the joint posterior of the models or approximate marginal likelihoods);

b. Information-theoretic model weighting based on AIC where $\Delta_i = \mathrm{AICc}_i - \min \mathrm{AICc}$ where min AICc is the smallest AICc value in the contending models k. The Akaike weights are then calculated as (Burnham and Anderson 2002),

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum_{k=1}^{K} \exp(-\Delta_k/2)}.$$

c. Model weights are chosen to "achieve the best predictive performance of the average (Dormann et al. 2018)."

d. Assigning fixed, equal weights to all contending models.

It was decided to weight the models by the performance metric *MAPE_one_step_ahead*.

3 References

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. In B. N. Pet rov & F. Caski (Eds.), Proceedings of the Second International Symposium on Information Theory (pp. 267-281). Budapest: Akademiai Kiado.

Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. Model selection: an integral part of inference. Biometrics: 603-618.

Burnham, K. P., and D. R. Anderson, D. R. 2002. Model selection and multimodel inference: A practical information-theoretic approach. Second Edition. New York: Springer-Verlag

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

Claeskens, G., and N. L. Hjort. 2008. Model selection and model averaging. Cambridge University Press, Cambridge, UK.

Dormann, C. F., J. M. Calabrese, G. Guillera-Arroita, E. Matechou, V. Bahn, K. Bartoń, C. M. Beale, S. Ciuti, J. Elith, K. Gerstner, J. Guelat,P. Keil, J. J. Lahoz-Monfort, L. J. Pollock, B. Reineking, D. R. Roberts, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, S. N. Wood, R. O. Wüest, and F. Hartig. 2018. Model averaging in ecology: A review of Bayesian, information-theoretic, and tactical approaches for predictive inference. Ecological Monographs, 88: 485-504.

Fletcher, D., and P. W. Dillingham. 2011. Model-averaged confidence intervals for factorial experiments. Computational Statistics and Data Analysis 55:3041–3048. Fox, J. and S. Weisburg. 2019. An R Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage Publications, Inc.

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

Lukacs, P. M., K. P. Burnham, and D. R. Anderson. 2010. Model selection bias and Freedman's paradox. Annals of the Institute of Statistical Mathematics 62:117–125.

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

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., 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. Alaska Department of Fish and Game, Regional Operational Plan No. ROP.CF.1J.2021.02, Douglas.

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

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.