Agenda Item F.1
Attachment 4
November 2021

Oregon Production Index Hatchery forecast methodology summary

Erik Suring[1] and Michael O'Farrell[2]

[1]Oregon Department of Fish and Wildlife: erik.suring@oregonstate.edu
[2] National Oceanic and Atmospheric Administration

Introduction

The majority of Coho Salmon (*Oncorhynchus kisutch*) in the Oregon Production Index (OPI) return to spawn at three years old, with a variable proportion of precocious males returning at two years old after four to six months in the ocean (Sandercock 1991). The early marine life-stage may be a survival bottleneck for Coho Salmon (Logerwell et al. 2003, Beamish et al. 2004). Jack Coho Salmon return to freshwater spawning areas after experiencing this potential survival bottleneck. Therefore, high jack return rates indicate favorable early life cycle ocean conditions for adults of that brood cycle, and thus jack returns can be a predictor of adult marine survival (Peterman 1982, Briscoe et al. 2005). The current forecast method uses multiple linear regression to incorporate information from sibling regression and hatchery production. Jack returns and smolt production from year *t-1* forecast the adult return in year *t*. Larger hatchery smolts may have higher smolt to adult survival (Holtby et al. 1990). The model incorporates information about smolt size through the "delayed smolt adjustment" metric. This is applied to smolt releases after May 15 which generally have been reared to a larger size.

Most of the Coho Salmon harvested in the OPI area originate from stocks produced in rivers located within the OPI area (Leadbetter Point, Washington to the U.S./Mexico border). These stocks include hatchery and natural production from the Columbia River, Oregon Coast, and northern California, and are divided into the following components: (1) Columbia River, coastal Oregon, and northern California public hatchery (OPIH), (2) Oregon coastal natural (OCN), including river and lake components, and (3) Lower Columbia natural (LCN).

The three OPI components are forecast independently. The adult return for the OPIH component is forecast using fish data from public hatcheries in Washington, Oregon, and California. The present OPIH forecast approach has been used since 1996. The current regression model was first used in 2008, when the jack returns, which had been separated into components from coastal Oregon and California hatcheries and Columbia River hatcheries, were combined into a single OPIH jack return metric. Component datasets have been maintained in OPITT tables 2 and 3. Since 2011 a longer time series beginning with 1969 jack returns and smolt releases was used. Prior to 2015, the model fitting and analysis was conducted by Dr. Pete Lawson from NMFS with the review of the OPITT team. Upon Dr. Lawson's retirement, ODFW staff assumed the main duties for annual updates to the model fit and running the annual forecast, using code provided by Dr. Lawson.

No methodology review documentation can be found for the 2011, 2008, or earlier changes to the OPIH forecast method, but it is documented annually in the Preseason Report I (e.g., PFMC 2021). This methodology summary is created at the request of PFMC members who asked for more detail on the forecast method than is provided in the Preseason Report I. Concerns have been raised at Oregon Production Index Technical Team meetings that the current method may have reduced predictive power in recent years due to factors like changes in the marine environment.

There are many ways one may evaluate the predictive performance of a forecast method. In this short review we evaluated changes in fit of the OPIH model over time by fitting the current method to a 30-year moving window of data from 1969 to present. We also compared relative forecast performance for nine coho stocks that make substantial contributions to Council-area Coho Salmon fisheries.

Methods

OPIH forecast method

OPIH forecast model inputs are Coho Salmon jack returns to all OPI hatcheries, jack returns to Columbia River hatcheries, smolts produced by Columbia River hatcheries, and delayed smolts at Columbia River hatcheries (Table 1). These inputs are reported annually in the Preseason Report I (PFMC 2021) table C-2. The response variable is the adult OPIH abundance which includes harvest impacts and escapement for public hatchery stocks originating in the Columbia River, Oregon coastal rivers, and the Klamath Basin, California. Adult estimates are derived from the Mixed Stock Model (MSM, Packer et al. 2007) from 1986 onwards. Each year a linear model (Eq. 1) is fit to data from 1969 to present, excluding 1983 due to large El Niño impacts.

Eq. 1 $OPIH_t = a(Jack\ OPI_{t-1}) + b(Jack\ CR_{t-1}[SmD_{t-1}/SmCR_{t-1}]) + c$, Where:

Jack OPI = jack returns to all OPI hatcheries in thousands
Jack CR = jack returns to Columbia River hatcheries in thousands
SmD = delayed smolts released from Columbia River hatcheries in millions
SmCR = total smolts released from Columbia River hatcheries in millions

A multivariate linear model is fit to the most current data in R (R Core Team 2021) and the total OPIH adult abundance in year *t* is forecast using jack and smolt production estimates from year *t-1*. Code and input data to replicate the model fit and prediction are available at https://github.com/ErikSuring/OPIH_Evaluation

The total OPIH adult abundance is partitioned into Columbia River early, late, and coastal hatchery stock components. These component proportions are determined using stock-specific jack to adult regressions. The coastal hatchery component is further partitioned into northern and southern coastal components. The northern OPIH coastal component is comprised of hatchery production from the Oregon Coast. The southern OPIH coastal component is comprised of hatchery production from the Rogue River basin in southern Oregon and the Klamath and Trinity basins in northern California. The partition between coastal hatchery stocks is based on the proportion of smolt releases in each area in year *t-1*.

Model fit over time

To evaluate how OPIH model fit has changed over time we fit the model (Eq 1) to a 30-year moving window subset of the Table 1 data (e.g., 1970-1999, 1971-2000, and so on up to 1991-2020); the model was fit to each 30-year period individually and estimated parameters for each 30-year period extracted and summarized. Model summary statistics such as the F-statistic and adjusted R^2 value were recorded for each run. Code and input data to replicate the analysis are available at https://github.com/ErikSuring/OPIH_Evaluation.

Forecast performance

The comparison of forecast performance across different stock groups is being made across stocks which differ in relative abundance and require scale-independent performance measures. We chose two such measures, Mean Absolute Percentage Error (MAPE) and Relative Mean Absolute Error (RelMAE), that are appropriate for these data and this application (Hyndman and Koehler 2006). These measures are also relatively straightforward to interpret.

MAPE is defined as:

$$\text{MAPE} = \text{mean}\left(\left|100\,\frac{Y_t - F_t}{Y_t}\right|\right),$$

where *Y* is the postseason observation, *F* is the preseason forecast and *t* is year. Lower values of MAPE indicate higher forecast accuracy.

RelMAE is defined as the ratio of the Mean Absolute Error (MAE) of the forecasted and observed abundances to the MAE of forecasted and observed abundances assuming a naïve forecast model (MAE_n), where

$$\text{MAE} = \text{mean}(|Y_t - F_t|)$$

and

RelMAE = MAE / MAE_n.

The naïve forecast model applied here was to assume that the forecast abundance in year t is equal to the postseason estimate of abundance in year t-1. RelMAE provides an indication of how much the forecasting approach(es) used in practice have resulted in improvement, or performed worse than, a naïve forecast model. When RelMAE is less than one, the proposed method performs better than the naïve model, and when RelMAE is greater than one, the proposed method performs worse than the naïve model.

The preseason forecasts and postseason estimates of abundance used for the nine coho stocks in this exercise are identical to those used to generate Figures III-1a and III-1b in PFMC (2021).

Results

OPIH forecast method

The parameters used in the 2021 forecast (Eq. 1) are:

a = 19.16

b = 28.18

c = -102.77

Forecast inputs are shown in the last row of Table 1. The adjusted R^2 was 0.94 and the F-statistic was 354.8 on 2 and 47 DF. This forecast was used for the Council adopted 2021 OPIH forecast of 1.6 million fish. Figure 1 shows the model prediction and observed abundance with the model fit to all years of the dataset. Figure 2 shows the predictor error for each year, excluding 1983, when the model is fit to the entire dataset.

Model fit over time

Table 2 shows the regression model statistics for twenty-one 30-year time periods. Change in F-statistic and adjusted R^2 value over time is shown in Figure 3. The figure demonstrates that the model fit to the oldest data has the highest R^2 value and F-statistic, indicating better model fit. The F-statistic declines fairly sharply until the 1977-1998 data subset and shows a slower decline thereafter. The adjusted R^2 value declines over time with the most recent five data subsets at the lowest value (beginning with the 1987-2018 data subset and carrying through the 1991-2020 set). There are differing patterns in t-values for individual regression components. This result indicates that while jack returns remain a strong predictive component of the model in the last five data subsets, the delayed smolt adjustment term is less useful as an explanatory variable ($p > 0.05$) during this timeframe.

Forecast performance

Figure 4 displays the ranked MAPE results and Figure 5 displays the ranked RelMAE results for the nine coho stocks. The OPIH forecast ranks highly relative to the other coho stocks under both measures of performance. For the MAPE assessment of forecast accuracy the OPIH stock ranks third among the nine stocks. For the RelMAE forecast performance metric, OPIH ranks first.

Discussion

The current forecast model fits the entire dataset well and predictor error is not biased. Most Coho Salmon forecast methods assume the early marine life history is the principal population bottleneck and it is not clear that any forecast method would have correctly forecast the large return in 2014 or the low return in 2015 likely associated with the marine heatwave known as The Blob. Excluding 2014 and 2015 we may be concerned about the over prediction in 2019. It is the highest error proportion of the time series but is similar to the magnitude of errors in the 1990s (Figure 2). It does not appear to be of a higher relative magnitude than forecast error in some previous years. Given the ephemeral nature of many environment-based forecasts (Wainwright 2021) the relatively strong predictive power of the current method across the large time frame considered here should be acknowledged.

The current forecast model for OPIH fits recent data subsets less well than earlier data subsets. Hatchery jack returns are still very informative in forecasting adult abundance, with the current model explaining much of the variance in adult abundance even in the most recent data subsets. The delayed smolt adjustment parameter is less informative in recent datasets. This may be due to changes in hatchery practices. The average number of delayed smolts in the early 25 years of the dataset (1971-1995) is more than four times higher than the last 25 years of the dataset (1996-2020). The variation in the number of delayed smolts is also higher in the early part of the dataset.

Forecast performance was assessed for models for nine stocks using all data available (PFMC 2021) and the length of the forecast and observed data series varied by stock. Willapa Bay natural coho had the fewest available pre/post abundance values (10), while Grays Harbor, Hoh, Quilleute, Queets, and Hood

Canal natural coho had the longest data series (34). The OPIH data series was intermediate, with a length of 25 years. While the OPIH model is fit to more than 50 years of data, OPIH predictions have been made since 1996.

Methods used to forecast salmon abundance can and do vary over time, and this is the case for some of the Coho Salmon stocks assessed here. Our assessment of forecast performance did not account for changes in methodology, but rather reflects the forecast methods that were used in each year.

Conclusion

Model evaluation is often conducted to choose between competing forecast methods. At this time there are no alternative method proposed for OPIH abundance forecast. This model review shows that the OPIH linear regression forecast method compares favorably with methods used for other Council-managed Coho Salmon stocks. While model fit has declined, the current model structure still explains a large part of the variation in adult abundance. There is evidence that, perhaps due to environmental change and change in hatchery practices, the current model structure is less informative than it was in the past and improvements may be possible by exploring other forecast indicators. To provide agility in creating the most accurate forecasts available it may be best to specify model evaluation criteria rather than the model to be used. If robust forecast performance criteria are developed it would allow OPITT to confidently select a model that best fits current conditions.

References

Beamish, R.J., C. Mahnken, and C.M. Neville. 2004. Evidence that reduced early marine growth is associated with lower marine survival of coho salmon. Transactions of the American Fisheries Society 133:26-33.

Briscoe, R.J., M.D. Adkison, A. Wertheimer, and S.G. Taylor. 2005. Biophysical factors associated with the marine survival of Auke Creek, Alaska, coho salmon. Transaction of the American Fisheries Society 134:817-878.

Holtby, L.B., B.C. Andersen, and R.K. Kadowaki. 1990. Importance of smolt size and early ocean growth to interannual variability in marine survival of coho salmon (Oncorhynchus kisutch). Canadian Journal of Fisheries and Aquatic Sciences 47.2181-2194.

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

Logerwell, E.A., N. Mantua, P.W. Lawson, R.C. Francis and V.N. Agonstini. 2003. Tracking environmental processes in the coastal zone for understanding and predicting Oregon coho (Oncorhynchus kisutch) marine survival. Fisheries Oceanography 12:6, 554-568.

PFMC (Pacific Fishery Management Council). 2021. Preseason Report I: Stock Abundance Analysis and Environmental Assessment Part 1 for 2021 Ocean Salmon Fishery Regulations. (Document prepared for the Council and its advisory entities.) Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 101, Portland, Oregon 97220-1384.

Packer J.F., J.D. Haymes, and C. Cook-Tabor. 2007. Coho FRAM Base Period Development. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 101, Portland, Oregon 97220-1384.

Peterman R.M. 1982. Model of salmon age structure and its use in preseason forecasting and studies of marine survival. Canadian Journal of Fisheries and Aquatic Sciences 39:1444-1452.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sandercock, F.K. 1991. Life history of coho salmon (Oncorhynchus kisutch). Pages 397-443 in Groot C, and Margolis L. editors. Pacific salmon life histories. University of British Columbia Press, Vancouver

Wainwright, T.C. 2021. Ephemeral relationships in salmon forecasting: A cautionary tale. Progress in Oceanography 193:102522

Table 1. Data sets used in predicting Oregon production index hatchery (OPIH) adult coho. Adults and jacks shown in thousands of fish and smolts in millions of fish.

	Adults (t)		Jacks (t-1)			Columbia River Smolts (t-1)			
Year (t)	OPIH[a/]	MSM[b/]	Total OPI[c/]	Columbia River[d/]	OR Coast/ CA[e/]	Total OPI[f/]	Normal Timed[g/]	Delayed[h/]	Delayed Smolt Adjustment[i/]
1970	2,765.1	-	162.2	148.6	13.6	32.4	27.6	0.0	0.00
1971	3,365.0	-	179.4	172.8	6.6	28.8	24.0	0.0	0.00
1972	1,924.8	-	103.7	100.8	2.9	33.3	28.2	0.0	0.00
1973	1,817.0	-	91.4	85.7	5.7	35.3	28.1	1.8	5.16
1974	3,071.1	-	144.2	132.0	12.1	33.6	25.6	2.9	13.43
1975	1,652.8	-	76.2	75.1	1.1	32.6	26.1	1.8	4.85
1976	3,885.3	-	171.5	146.2	25.3	34.0	27.0	2.0	10.08
1977	987.5	-	53.8	46.3	7.5	33.5	28.7	0.2	0.32
1978	1,824.1	-	103.2	99.2	4.0	35.5	31.4	0.0	0.00
1979	1,476.7	-	72.5	64.1	8.4	37.1	27.6	5.0	9.83
1980	1,224.0	-	57.7	51.6	6.0	34.2	22.2	6.7	11.96
1981	1,064.5	-	48.7	40.6	8.1	32.3	22.5	5.6	8.09
1982	1,266.8	-	61.3	55.0	6.3	37.3	25.7	6.8	11.51
1983[j/]	599.2	-	68.3	61.0	7.2	32.7	22.8	5.0	10.97
1984	691.3	-	31.6	28.0	3.6	30.9	21.9	5.1	5.29
1985	717.5	-	26.0	18.2	7.8	34.5	20.2	9.1	5.65
1986	2,435.8	2,412.0	77.5	64.6	12.9	32.8	16.6	12.2	27.37
1987	887.2	779.4	32.9	24.2	8.7	39.5	23.9	9.0	6.62
1988	1,669.3	1,467.8	85.2	72.3	12.9	35.0	21.1	7.7	19.33
1989	1,720.2	1,922.0	60.8	55.0	5.8	36.0	22.3	7.2	13.42
1990	718.4	713.6	46.6	37.1	9.6	36.1	21.1	8.5	10.65
1991	1,874.8	1,816.5	68.6	60.7	7.9	37.2	23.2	7.1	14.22
1992	543.6	512.6	25.6	19.9	5.7	42.1	29.3	6.0	3.38
1993	261.7	223.3	27.1	19.6	7.5	39.7	27.3	5.5	3.29
1994	202.3	214.1	5.2	3.9	1.3	39.5	28.4	6.0	0.68
1995	147.2	139.4	11.8	9.1	2.7	32.2	23.5	3.1	1.06
1996	185.2	176.5	17.4	14.1	3.2	29.5	20.9	4.2	2.36
1997	200.7	195.6	20.4	15.8	4.6	31.6	24.6	3.4	1.92
1998	207.5	228.3	9.7	6.7	3.0	24.6	18.5	2.5	0.80
1999	334.5	372.5	29.5	23.6	5.9	29.1	23.8	3.0	2.64
2000	673.2	673.1	34.8	31.3	3.5	29.7	23.8	4.1	4.60
2001	1,417.1	1,478.7	87.4	71.7	15.7	32.2	28.7	2.0	4.67
2002	649.8	689.5	25.2	18.9	6.3	26.8	23.9	1.4	1.05
2003	936.6	1,009.9	49.9	41.7	8.2	25.3	23.4	0.3	0.53
2004	622.1	693.6	35.4	29.4	6.0	24.5	21.2	2.0	2.53
2005	443.2	454.0	25.0	21.2	3.8	23.4	21.2	0.8	0.77
2006	440.6	523.4	25.9	20.9	5.0	22.0	20.2	0.4	0.41
2007	476.6	545.3	36.3	34.2	2.2	21.8	20.3	0.1	0.17
2008	565.3	576.9	16.0	14.9	1.2	22.7	20.8	0.6	0.42
2009	1,066.2	1,051.0	60.4	58.4	2.0	22.8	20.8	1.1	2.93
2010	551.3	546.5	25.1	23.8	1.4	21.9	20.7	0.2	0.23
2011	442.3	454.2	23.3	22.2	1.1	19.3	18.2	0.3	0.36
2012	182.3	183.1	17.9	13.9	4.0	19.9	18.1	0.9	0.66
2013	316.9	335.1	26.3	24.1	2.2	19.2	17.1	1.1	1.46
2014	1,263.6	1,316.5	51.4	49.4	2.0	19.6	18.0	0.6	1.59
2015	251.7	268.9	39.6	37.0	2.6	19.4	16.9	1.5	3.02
2016	233.8	247.7	19.7	18.6	1.0	18.9	16.9	1.3	1.33
2017	284.8	291.8	22.9	22.4	0.4	18.4	16.5	1.3	1.64
2018	149.4	182.8	19.2	18.5	0.7	17.2	16.0	0.7	0.78
2019	300.5	340.7	47.4	46.7	0.8	19.7	18.6	0.5	1.28
2020	-	369.6	15.2	14.9	0.3	18.3	17.3	0.5	0.41
2021[k/]	-	1,607.9	86.5	83.3	3.2	18.1	18.1	0.4	1.89

a/ Adult OPIH = Harvest impacts plus escapement for public hatchery stocks originating in the Columbia River, Oregon coastal rivers, and the Klamath River, California.

b/ Adult MSM = Harvest impacts plus escapement for public hatchery stocks originating in the Columbia River, Oregon coastal rivers, and the Klamath River. Estimates derived from the MSM and used for prediction beginning in 2008.

c/ Jack OPI = Total Jack CR and Jack OC.

d/ Jack CR = Columbia River jack returns corrected for small adults.

e/ Jack OC = Oregon coastal and California hatchery jack returns corrected for small adults.

f/ Total OPI = Columbia River (Sm D + Sm CR), Oregon coastal and Klamath Basin.

g/ Sm CR = Columbia River smolt releases from the previous year expected to return as adults in the year listed.

h/ Sm D = Columbia River delayed smolt releases from the previous year expected to return as adults in the year listed.

i/ Correction term for delayed smolts released from Col. R. hatcheries (Col. R. Jacks*(Delayed Smolts/Col. R. Smolts)).

j/ Subsequent to 1983 data not used in predictions due to El Niño impacts.

k/ For MSM: Preseason predicted adults.

Table 2. Model summary statistics for 30-year subsets of data fit to the current OPIH forecast model. Subset start year indicates the beginning year of the series of 30-year moving windows, which begin with 1970-1999 (indicated as 1970) and end with 1991-2020 (indicated as 1991).

Subset Start Year	RMSE	Adjusted R^2	F-statistic	OPI jack t-value	Delayed smolt adjustment t-value
1970	230	0.95	266	21.76	4.54
1971	222	0.95	268	21.12	4.16
1972	227	0.94	212	17.26	3.77
1973	234	0.93	196	15.72	3.31
1974	236	0.93	190	15.34	3.20
1975	236	0.92	156	14.06	3.19
1976	235	0.92	157	13.78	3.24
1977	218	0.87	96	8.73	3.97
1978	217	0.87	98	8.47	4.05
1979	223	0.85	83	5.22	3.54
1980	223	0.85	80	5.36	3.69
1981	223	0.85	78	5.35	3.66
1982	224	0.85	78	5.34	3.65
1983	225	0.85	78	5.42	3.66
1984	222	0.85	81	5.49	3.74
1985	237	0.83	73	5.86	3.13
1986	253	0.81	64	5.35	3.02
1987	243	0.76	47	5.88	1.33
1988	243	0.76	48	5.92	1.23
1989	225	0.78	53	6.47	2.27
1990	220	0.72	38	6.45	1.24
1991	214	0.74	41	6.22	1.85

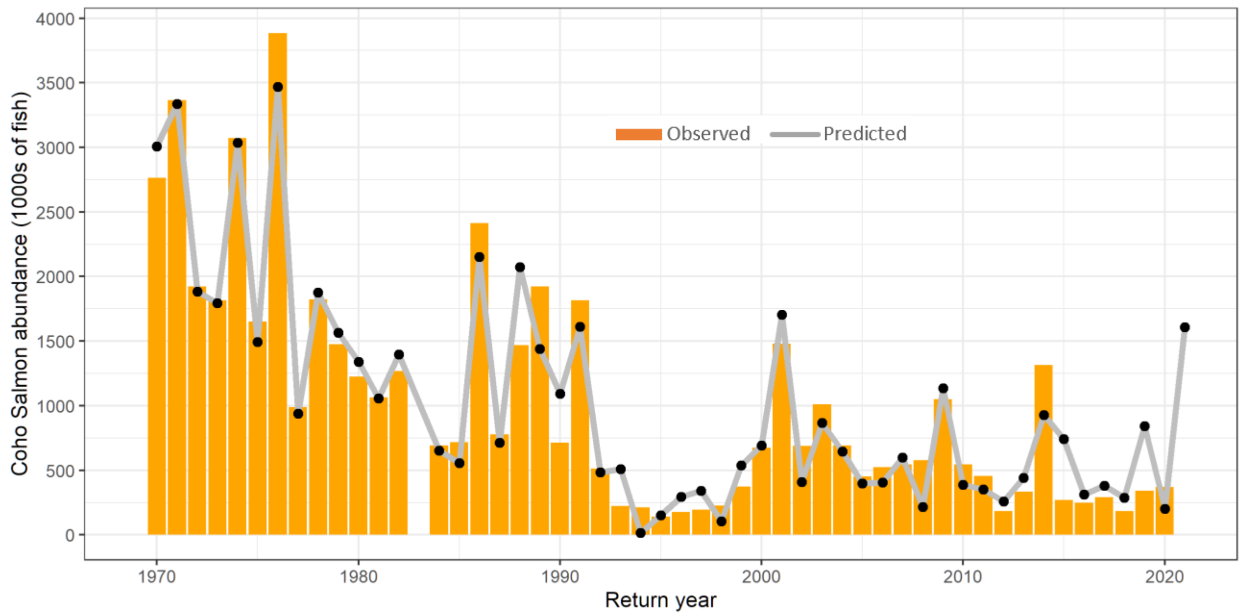

Figure 1. Observed and predicted OPIH Coho Salmon adult abundance. Predicted values are from the model fit to all available data except 1983.

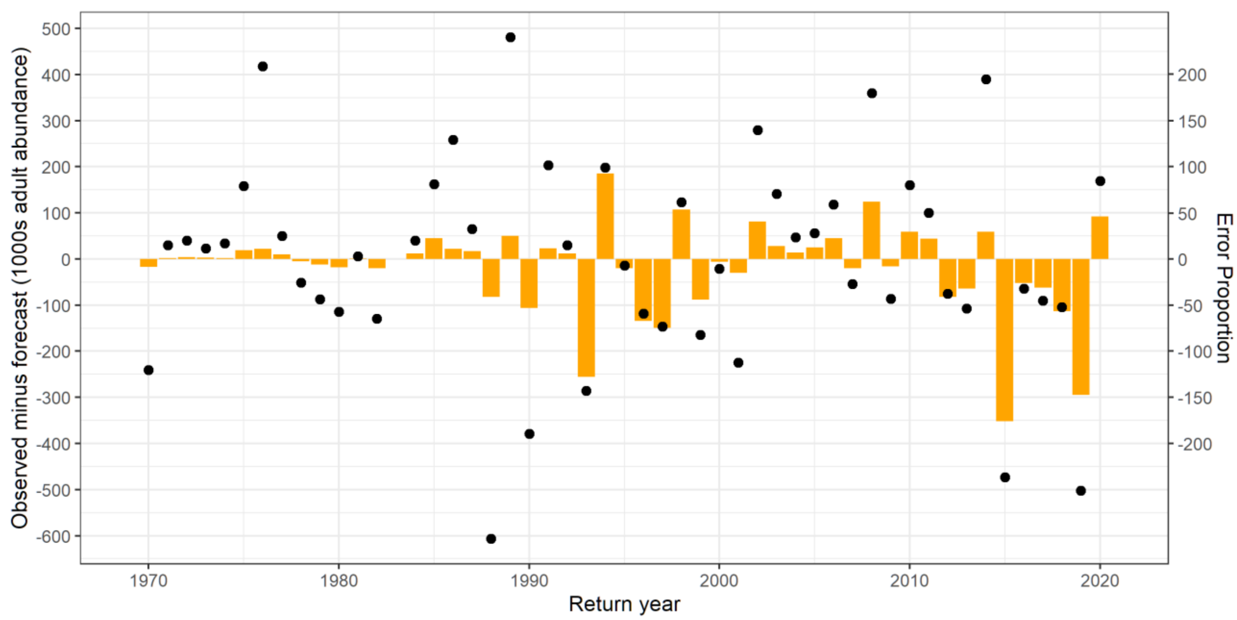

Figure 2. Predictor error (observed minus forecast, points) and error proportion (bars) by year for the current OPIH forecast method from the model fit to all available data except 1983. 1983 prediction error is excluded from the figure due to El Nino impacts.

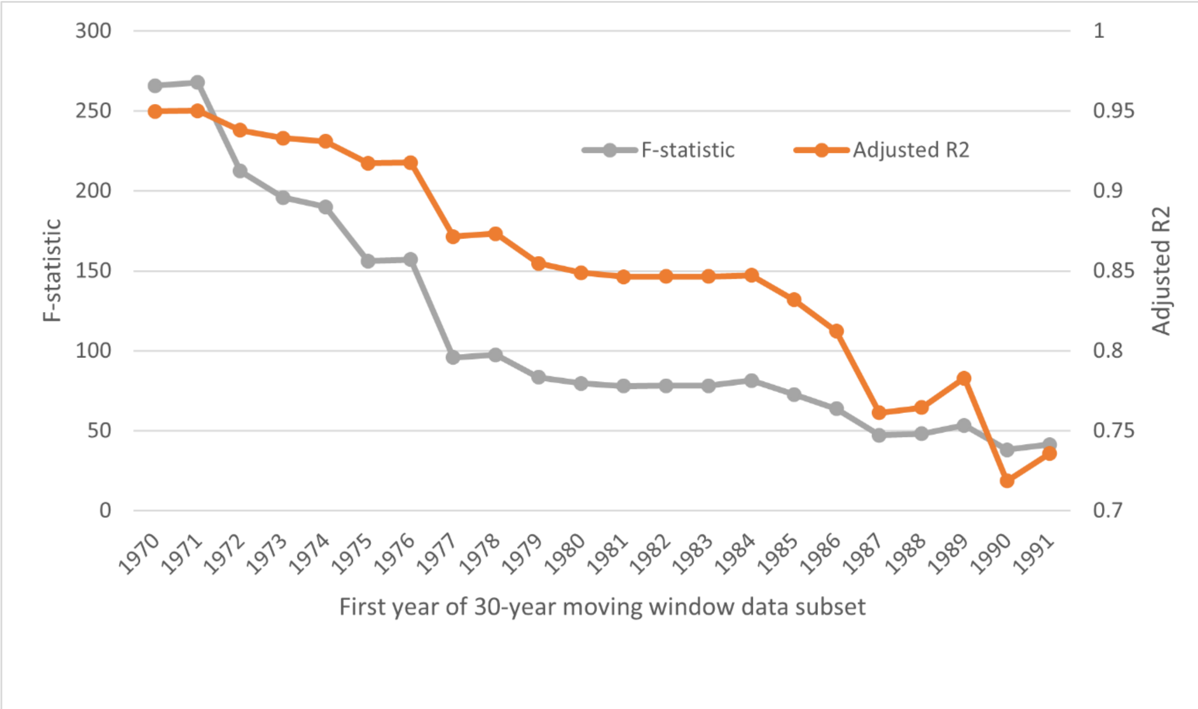

Figure 3. Regression summary statistics for the OPIH predictor model fit to 30-year moving window data subsets. Years on the x-axis indicate the beginning year of the series of 30-year moving windows, which begin with 1970-1999 (indicated as 1970) and end with 1991-2020 (indicated as 1991).

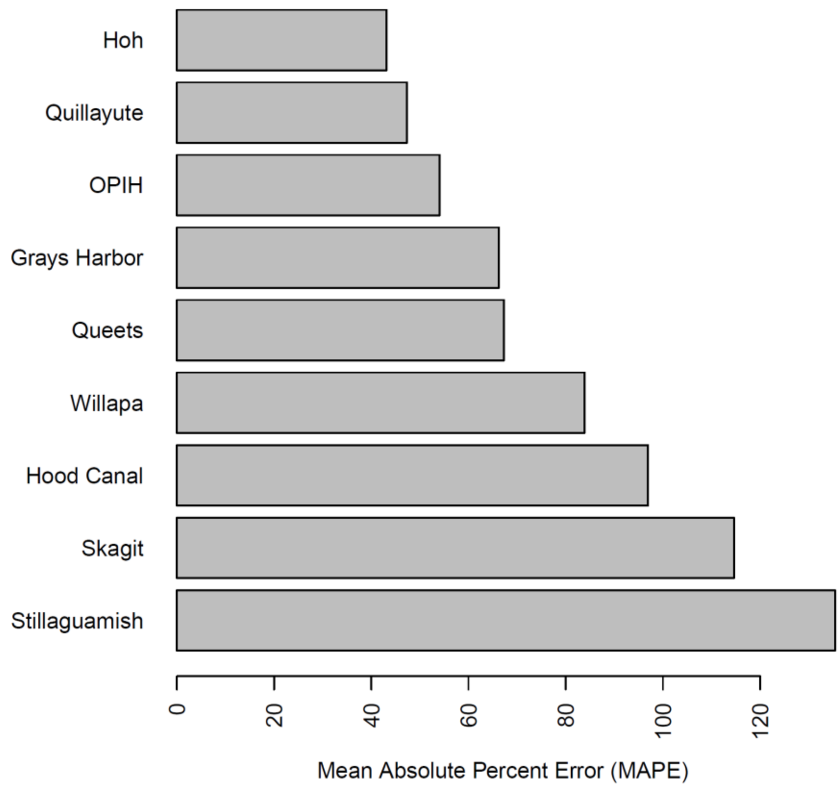

Figure 4. Mean Absolute Percent Error (MAPE) results for nine coho salmon stocks.

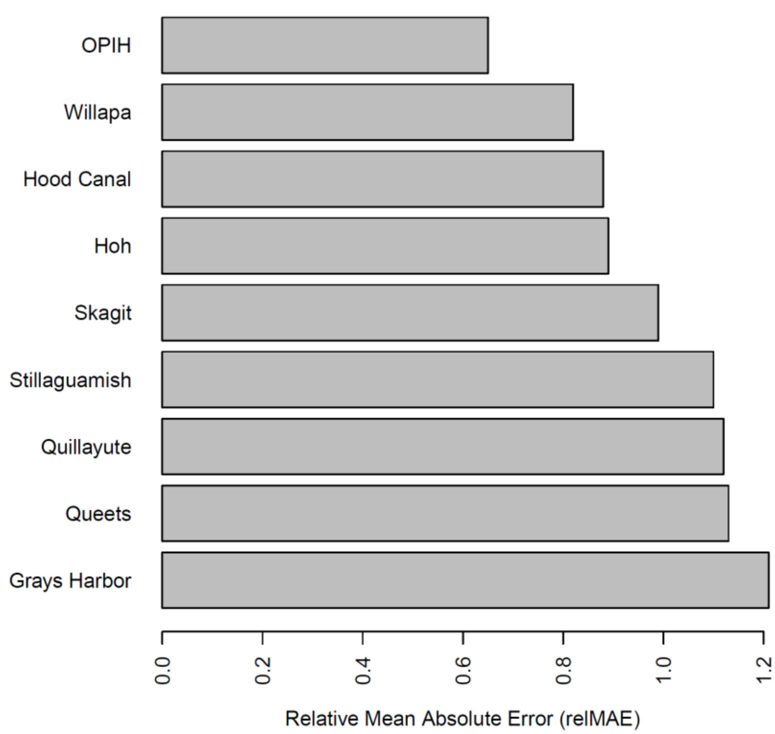

Figure 5. Relative Mean Absolute Error (RelMAE) results for nine coho salmon stocks.