**Introduction**

Pink salmon forecasting in Southeast Alaska began in the mid-1960s and has evolved over the years. To predict returning adult pink salmon populations in Southeast Alaska, ADF&G started a pre-emergent fry indexing program in spring of 1963 (Hoffman 1965). The justification for the relationship between fry and the resultant adult return was based on pink salmon predictions in Prince William Sound (Noerenberg 1961, 1963, 1964). Early predictions of southern and northern Southeast Alaska pink salmon runs, based on a weighted forecast of the escapement-return relationship and the pre-emergent fry indices, were not very accurate (Hoffman 1965, 1966; Smedley and Siebel 1967; Smedley et al. 1968). Throughout the 1970s, separate forecasts were developed for northern Southeast and southern Southeast Alaska adult pink salmon returns due to differences in migration routes and run timing (Nakatani et al. 1975; Alexandersdottir 1987). These forecasts were based on the relationship between fry abundance and adult returns (e.g., Noerenberg and Siebel 1969, ADF&G 1971, 1972). Starting with the 1974 forecasts, environmental variables such as seawater temperature and air temperature, were added to the regression forecasts throughout the mid-1970s and the 1980s (e.g., ADF&G 1973, 1975, 1983). Pre-emergent fry indices were no longer available starting with the 1988 forecast (Eggers and Dean 1988); forecasts based on the fry abundance index were replaced by multiple linear regression forecasts of escapement indices, environmental variables, and fry data collected in the early marine program in Tenakee Inlet in the late 1980s to 1990s (e.g., Geiger and Savikko 1990, 1991, 1992). The 1993 forecast, based on a multiple linear regression of the brood year escapement index, average daily minimum winter air temperatures from NOAA weather stations throughout SEAK, and sum of previous two brood year escapement indices, and harvest rather than adult pink salmon return data as the dependent variable, was the first time that separate forecasts for northern and southern Southeast Alaska were not made (Geiger and Savikko 1993). This change was prompted because ADF&G no longer collected area-specific data; the pre-emergent fry program and the early marine fry program in southern Southeast Alaska were eliminated and pre-emergent fry program in northern Southeast Alaska was eliminated (Geiger and Savikko 1993) in 1992. Early marine fry surveys were completely discontinued in 1997 (Geiger and Hart 1999). The 1994 to 2003 predictions were a subjective combination of statistical forecast models, anecdotal fry abundance data, historic average harvests, environmental variables, and expert opinion (e.g., Scott and Geiger 2000; Eggers 2002, 2003), until this subjective method was replaced by an exponential smoothing model for the 2004 to 2006 forecasts (e.g., Plotnick and Eggers 2004; Eggers 2005, 2006). Starting with the 2007 forecast, the exponential smooth forecasts were adjusted with the NOAA Auke Bay lab June-July pink salmon fry data (e.g., Eggers 2007, Munro and Tide 2014, Brenner and Munro 2017). Recent forecasts (2019 to 2022) have been based on a multiple linear regression model with juvenile pink salmon abundance indices collected by the Southeast Coastal Monitoring project (SECM) project in northern Southeast Alaska inside waters during June and July along with a temperature index, and are produced by the combined efforts of NOAA and ADF&G (Piston et al. 2019).

In 2020 and 2021, the potential use of satellite temperature data (SST; available from the NOAA National Environmental Satellite data and Information Service) was explored to forecast adult pink salmon harvests in 2021 and 2022 (*2021 NOAA Fisheries-Alaska Department of Fish and Game Southeast Alaska pink salmon harvest forecast, Alaska Department of Fish and Game Commercial Fisheries Advisory Announcement 18 November 2020* and *2022 NOAA Fisheries-Alaska Department of Fish and Game Southeast Alaska pink salmon harvest forecast, Alaska Department of Fish and Game Commercial Fisheries Advisory Announcement 16 November 2021*). Satellite data allows for averaging of temperature readings over an almost infinite variety of temporal and geographic units, and may (potentially) be a better predictor of pink salmon returns compared to the temperature data collected during the SECM survey. The temperature index currently used in forecasting pink salmon harvest is based on the overall average 20 m integrated water column temperature during May through July at the 8 stations in Icy Strait collected as part of the annual SECM surveys (Murphy et al. 2019). In addition, a model-averaging approach, as opposed to one best model approach, is being considered. This paper explores the possible methods that would be applied to forecast the 2023 adult pink salmon harvest in 2023 based on weighting alternative multiple linear regression models, with SECM survey temperature or SST covariates, by the inverse-variance (Table 2) in a model-averaging approach.

**Methods**

Biophysical variables based on data from northern Southeast Alaska were used to forecast the 2022 harvest of adult pink salmon in Southeast Alaska, a year in advance using multiple linear regression and a model-averaging approach (Appendix A). Pink salmon harvest data was the response variable and satellite-derived surface temperature data, May through July SECM survey temperature data at a 20 m depth from 8 stations in Icy and Chatham Straits, and SECM survey juvenile abundance data (CPUE) were used as potential predictive biophysical variables in the models.

**Biophysical Predictor Variables**

**Satellite-derived SST data**

Monthly satellite-derived sea surface temperature data from April 1997 through July 2021 were pulled from NOAA Coral Reef Watch (2021a; 2021b) and matched to pre-determined coordinates from four spatial regions to use as potential predictor variables in the pink salmon forecast models. Satellite sea surface temperature data were summarized by region and year (i.e., average of May (May), the average over the months of May, June, and July (MJJ), the average over the months of April through June (AMJ), and the average over the months of April through July (AMJJ)) from 1997 to 2021 (Table 1). The monthly data for July 2021 was not available, so the daily data for July 2021 were summarized by month and region, and then combined with the monthly data from April 1997 through June 2021 to create the SST dataset (April 1997 through July 2021; NOAA Coral Reef Watch 2021b). This satellite-derived SST data set was then matched to pre-determined coordinates from four spatial regions (Icy Strait, Chatham Strait, northern SEAK (NSEAK), SEAK) that corresponded to sixteen variables of interest (four regions and four temporal variables per region; see Satellite-derived SST variables section). 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 (Table 1; Figure 1; Figure 2). The Chatham Strait 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; Table 1; Figure 2; Figure 3). 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; Table 1; Figure 2; Figure 4). The SEAK region encompasses Southeast Alaska from 59.475 to 54.725 degrees north latitude (Table 1; Figure 2; Figure 5). These sixteen variables were then used as potential predictor variables in the pink salmon forecast models.

**Southeast Coastal Monitoring project survey data**

The Southeast Coastal Monitoring project (SECM) has been conducting annual surveys to evaluate the status of the pelagic ecosystem, including juvenile pink salmon, in the northern region of Southeast Alaska since 1997 (Piston et al. 2021). Data collected during the SECM surveys include surface trawl sampling for salmon and other pelagic species, 60 cm bongo net sampling for zooplankton, conductivity-temperature-depth profiles for temperature (°C) and salinity (PSU) data, and a water sample for chlorophyll-a (ug∙L-1). Fish are sampled at each station with a NETS Nordic 264 rope trawl fished for 20 min at each station at least once during June through August with tow speeds of approximately 1.5 m∙sec-1 and typical fishing dimension of 18 m wide by 24 m deep. Data from these surveys are summarized annually, and provide information to forecast pink salmon returns.

***Temperature data***

Survey temperature (°C) data were summarized by year (1997 to 2021), and month (average over the months of May, June, and July) for the 20 m integrated water column in the Icy Strait and Upper Chatham transects combined. The Icy Strait and Upper Chatham transects include 8 sampling stations (stations ISA, ISB, ISC, ISD, UCA, UCB, UCC, UCD). The summarized variable ISTI is the overall average 20 m integrated water column temperature during May through July at the 8 stations in Icy Strait (Table 1; Figure 1; Figure 2). This variable (ISTI) was then used as potential predictor variable in the pink salmon forecast models.

***Index of juvenile abundance (CPUE)***

The index of juvenile abundance of pink salmon is the predictor variable ‘CPUE’ in the pink salmon forecast models. To calculate this variable, the log-transformed juvenile pink salmon catch-per-unit-effort (Ln (CPUE+1); standardized to an effort of a 20 minute trawl set), by haul, from the SECM survey was multiplied by the pooled-species vessel calibration coefficient; calibrated to the NOAA ship *John N. Cobb*  (Wertheimer et al. 2008, 2009, and 2010; Table 1). Then, this value was averaged by month and year; whichever month (June or July) had the highest average catch in a given year is then used as the juvenile abundance index for that particular year (Table 1).

**Response Variable: Harvest Data**

Time series of the annual southeast Alaska adult pink salmon harvest data is downloaded from the ADF&G Fish Ticket Database System in millions of fish. The harvest data was log-transformed in the model.

**Model averaging (multi-model inference)**

A model-averaged approach, based on eighteen multiple regression models (*i* = 1,2,…, 18) with juvenile pink salmon catch-per-unit-effort (CPUE) and temperature data from either the Southeast Alaska Coastal Monitoring Survey (SECM; Piston et al. 2021) or satellite-derived SST data (NOAA Coral Reef Watch 2021a; 2021b) and weighted by the inverse variance, was used to forecast the 2022 adult pink salmon harvest in Southeast Alaska. The simplest regression model consisted of only the predictor variable juvenile pink salmon CPUE , while the other seventeen regression models consisted of the predictor variable juvenile pink salmon CPUE and a temperature index,

(1)

The temperature index was either the SECM survey temperature data integrated over 20 m depth (ISTI) or one of the seventeen satellite-derived SST data (Table 1). The response variable (*Y*; Southeast Alaska adult pink salmon harvest in millions) and CPUE data were log-transformed in the model, but temperature data were not.

The process to weight the model-averaged forecast of Southeast Alaska pink salmon harvest in 2022, and calculate the prediction interval around the model-averaged forecast is as follows:

1. Calculate the bias-corrected (Miller 1984) model-predicted adult pink salmon harvest for 2022 from each of the *i* regression models (18 models in the model set) in log space,

. (2)

1. Calculate the model-averaged forecast () in log space by weighting each of the 18 model forecasts by the one step ahead inverse-variance weight normalized to sum to one,

(3)

The one step ahead inverse-variance (from the last 5 years of predictions) weight, of each model *i* is defined as,

), (4)

where are the observed values (i.e., observed Southeast Alaska adult pink salmon harvest) and are the predicted values (i.e., predicted Southeast Alaska adult pink salmon harvest) from model *i* using dataup to time *t*-1 in the last five years of the time series. In equation 4, *n* = 5 since the last five years are forecasted in the time series. The process for calculating the inverse-variance (one step ahead variance) for each model *i* is as follows:

1. Estimate the regression parameters at time *t* from data up to time *t* – 1 for model *i*.
2. Make a prediction of at time *t* based on the predictor variables at time *t* and the estimate of the regression parameters at time *t* for model *i*.
3. Calculate the inverse-variance weight (equation 4)based on the prediction of at time *t* and the observed value of at time *t*.
4. Repeat this for data up through year 2016 (e.g., data up through year 2016 is *t*−1 and the forecast is for year 2017; *t*), data up through year 2017 (e.g., data up through year 2017 is *t* − 1 and the forecast is for year 2018; *t*), data up through year 2018 to forecast 2019, data up through year 2019 to forecast 2020, and data up through year 2020 to forecast 2021.

For example, based on the CPUE-only regression model (model m1; Table 2) and using data through 2016, the predicted harvest in 2017 is 4.23 (68.39 million fish) and = (3.55-4.23)2 = 0.46; using data through 2017, the predicted harvest in 2018 is 2.67 (14.39 million fish) and = (2.09-2.67)2 = 0.33; using data through 2018, the predicted harvest in 2019 is 2.86 (17.48 million fish) and = (3.05-2.86)2 = 0.04; using data through 2019, the predicted harvest in 2020 is 2.87 (17.67 million fish) and = (2.09-2.87)2 = 0.61; and using data through 2020, the predicted harvest in 2021 is 3.23 (25.23 million fish) and = (3.88-3.23)2 = 0.42. Then the sum of 0.46+0.33+0.04+0.61+0.42 divided by 4 is 0.47. Finally, one divided by 0.47 is 2.14. This value (2.14) is the inverse-variance weight of the CPUE-only model.

To normalize the inverse-variance weights to sum to one ), each individual weight is divided by the sum of all the model weights,  .

1. Calculate the standard error of the model-averaged forecast (i.e., the square root of the unconditional variance estimator; equation 9 in Buckland et al. 1997; derivation in Burnham and Anderson 2002:159-162) as

, (5)

where is the model-averaged forecast (in log space; assumed unbiased), is the individual model bias-corrected forecast for 2022 (in log space), and (i.e., the misspecification bias of model ) is computed as . The confidence interval is then calculated as,

, (6)

where for 80% confidence intervals and the results are then exponentiated.

**Results**

Based on the set of eighteen models and weighting the models by the inverse-variance (Table 2), the model-averaged 2022 forecast was 14.3 million fish (80% CI: 6.75-30.33). The model-averaged forecast is roughly 1.3 million less than the official 2022 forecast that was based on the biophysical predictor variables CPUE and the overall average 20 m integrated water column temperature during May through July at the 8 stations in Icy Strait (i.e., ISTI; model 2 in Table 2) that was presented in the advisory announcement (15.6 million fish rounded to 16 million fish; *2022 NOAA Fisheries-Alaska Department of Fish and Game Southeast Alaska pink salmon harvest forecast, Alaska Department of Fish and Game Commercial Fisheries Advisory Announcement 16 November 2021).*

Weighting the different models aims to adjust the average (i.e., improved inference) so that well-predicting models receive more weight than poorly fit models. There are various approaches to estimating the model-averaged weights (e.g., fixed equal weights (naïve approach), Bayesian model weights, information-theoretic framework based on the Kullback-Leibler divergence (e.g., Akaike’s Information Criterion; Akaike 1973, Burnham and Anderson 2002), cross-validation) (Dormann et al. 2018). “Equal-weight” model averaging was applied in the 2008 forecast of adult pink salmon harvest (Nelson et al. 2008) and Bayesian model averaging was investigated for the 2002 forecast (Adkison 2002) as an alternative to the traditional statistical forecast models implemented at the time. As our goal was to reduce prediction error in the forecast, model weights were based on the one step ahead inverse-variance from the last 5 years of predictions (equation 4).

The confidence interval around the model-averaged forecast was based on Buckland et al. (1997). This method assumes that the average point estimate is unbiased and, thus, the average point estimate can be used to compute the bias of the individual predictions. In addition, the method by Buckland et al. (1997) assumes that predictions from different models are perfectly correlated (i.e., covariance term is large) and, as a result, the variance estimation is more conservative (Dormann et al. 2018). Hjort and Claeskens (2003) criticize the imperfect distributional approximation behind the Buckland et al. (1997) estimator, but this method has proved to work adequately in simulations (Lukacs et al. 2010; Fletcher and Dillingham 2011).

**Discussion**

Model averaging incorporates the uncertainty inherent in model structure and in parameter values into the analysis results (i.e., forecast). The model-averaging approach has been explored in historical forecasts of adult pink salmon returns and in exploratory analyses of alternative methods of forecasting this stock. “Equal-weight” model averaging was applied in the 2008 forecast of adult pink salmon harvest. Due to concerns that the 2007 June-July pink salmon fry CPUE index from the SECM Project was low and outside the range of historical data, the exponential-smooth estimate and the CPUE-error-adjusted estimate were averaged to produce the 2008 forecast (Nelson et al. 2008). Although not implemented for the 2002 forecasts for northern and southern SEAK, Bayesian model averaging (Adkison 2002) was investigated as an alternative approach to the traditional statistical forecast models implemented at the time (Scott and Geiger 2000; Geiger and McNair 2001; Eggers 2002). Spawner-recruit models that included environmental variables (Quinn and Derios 1999) such as average upwelling anomalies, fish scales (growth), average anomaly of Gulf of Alaska SST, average air temperature anomaly in Southeast Alaska from September to August, with and without density-dependent effects, were weighted by the posterior probability of each model and its parameter combination. To implement this method, the joint posterior probability distributions of the parameters of the full Bayesian model (with the environmental predictors included) is first computed. The posterior distribution is then sampled for model parameters and a stochastic forecast is drawn from each sample creating a Bayesian posterior predictive distribution of the 2002 pink salmon forecast for each competing model (Adkison 2002). The approach implemented in this report is a frequentist approach to model averaging (Burnham and Anderson 1998) and opposed to a Bayesian approach (Patterson 1999).

An alternative way to calculate the confidence interval of the model-averaged forecast, as was implemented in the 2008 pink salmon harvest forecast, is to produce hindcast predictions for the last ten years (2012-2021) based on the chosen model-averaging approach, and estimate the sample standard deviation from the sum of the squared errors of the log of the observed values minus the log of the predicted values. The 80% confidence interval is then calculated as equation 6, but the variable is replaced by the ‘estimated’ sample standard deviation (Nelson et al. 2008). To calculate the confidence interval based on this method, hindcast predictions were made for the last five years in the data series for each of the 18 models (18 hindcast predictions per year for 5 years;2017-2021). A forecast was then produced for each of the last 5 years by weighting the 18 models equally within each year. Next, an estimate of the sample standard deviation was calculated from the sum of the squared errors of the log of the observed values minus the log of the predicted values for the five forecasts. This value (0.92) was then used as the variable in equation 6. Based on the set of eighteen models, weighting the models equally (Table 2), and using the last 5 years of hindcast predictions (based on the equally-weighted model average) to calculate the ‘estimated’ standard deviation, the model-averaged 2022 forecast would have been 14.39 million fish (80% CI: 4.45-46.56). This approach could be considered an alternative to the model-averaged approach in this report.

**References**

NOAA Coral Reef Watch (NOAA\_DHW\_monthly dataset). 2021a, updated daily. NOAA Coral Reef Watch Version 3.1 Monthly 5km SST and SST Anomaly, NOAA Global Coral Bleaching Monitoring Time Series Data, May 1997-June 2021. College Park, Maryland, USA: NOAA/NESDIS/STAR Coral Reef Watch program. Data set accessed 2021-10-01 at <https://coastwatch.pfeg.noaa.gov/erddap/griddap/NOAA_DHW_monthly.html>.

NOAA Coral Reef Watch (NOAA\_DHW dataset). 2021b, updated daily. NOAA Coral Reef Watch Daily Near-real-Time Global 5km SST and SST Anomaly, NOAA Global Coral Bleaching Monitoring Time Series Data, July 2021. College Park, Maryland, USA: NOAA/NESDIS/STAR Coral Reef Watch program. Data set accessed 2021-10-01 at <https://coastwatch.pfeg.noaa.gov/erddap/griddap/NOAA_DHW.html>.

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.

Wertheimer, A. C., J. A. Orsi, E. A. Fergusson, and M. V. Sturdevant. 2008. Paired comparisons of juvenile salmon catches between two research vessels fishing Nordic 264 surface trawls in southeastern Alaska, July 2007. NPAFC Doc. 1112.,17 p.

Wertheimer, A. C., J. A. Orsi, E. A. Fergusson, and M. V. Sturdevant. 2009. Calibration of Juvenile Salmon Catches using Paired Comparisons between Two Research Vessels Fishing Nordic 264 Surface Trawls in Southeastern Alaska, July 2008. NPAFC Doc. 1180. 18 pp.

Wertheimer, A. C., J. A. Orsi, E. A. Fergusson, and M. V. Sturdevant. 2010. Calibration of Juvenile Salmon Catches using Paired Comparisons between Two Research Vessels Fishing Nordic 264 Surface Trawls in Southeast Alaska, July 2009. NPAFC Doc. 1177. 19 pp. (Available at http://www.npafc.org).