P1437-17

Review of ADF&G's 2017 Copper River Chinook Salmon Forecast

A review of relevant sections of Brenner and Munro (2017)
and the Excel file called "2017 Copper River Chinook forecast to NVE_5_15_2017.xlsx"

David Robichaud

LGL Alaska Research Associates, Anchorage, AK, USA

Prepared for:
Native Village of Eyak,
Cordova, AK

August 1, 2017

Overview

I have thoroughly reviewed the Excel file provided ("2017 Copper River Chinook forecast to NVE_5_15_2017.xlsx"). While I found quite a few process errors (formulas pointing to wrong cells), my corrections did not result in substantive changes to any of the forecasts that were generated for the 2017 Copper River Chinook Salmon run. I did take issue with the method for selection among the competing models. Specifically, I found evidence in support of the 4-yr and 5-yr Average models, which would have allowed a harvest of 17,000 to 18,000 fish; and I did not find support in the literature for either of ADF&G's selection methods (which selected a model that allowed a harvest of 5,000 fish).

This document outlines the methods that ADF&G used to arrive at their 2017 forecast, and how they selected one model from among the competing models. I have also commented on alternative models and alternative methods of model selection. I extracted any available information about data sources that was embedded in ADF&G's Excel file. And I produced a list of corrections that I have incorporated into my copy of the Excel file, which I can provide to ADF&G if the NVE wishes me to do so.

Introduction

There are many different models that could have been used to forecast the 2017 Copper River Chinook Salmon run. The ones that were put forward by ADF&G (Table 1) in their forecast report (Brenner and Munro 2017) and in their Excel spreadsheet include:

1) PY (Previous Year) Model: The 2017 total run was forecast to equal the total run from the previous year.
2) 2-Yr Average Model: The 2017 total run was forecast to equal the average of the previous two years' total run.
3) 3-Yr Average Model: The 2017 total run was forecast to equal to the average of the previous three years' total run.
4) 4-Yr Average Model: The 2017 total run was forecast to equal to the average of the previous four years' total run.
5) 5-Yr Average Model: The 2017 total run was forecast to equal the average of the previous five years' total run.
6) 10-Yr Average Model: The 2017 total run was forecast to equal to the average of the previous ten years' total run.
7) Hybrid Sibling model: The Age 1.3 run was forecast based on a regression of same-brood-year Age 1.2 returns, and all other ages were forecast to equal the age-specific run from the previous year.

Each of these seven models produced different forecasts, had different prediction intervals, and some performed better than others (Table 1). ADF&G's main goal was to generate forecasts from a suite of reasonable models, and to use an objective method to pick the best from among them. In the end, the model that ADF&G selected to forecast the total 2017 Copper River Chinook Salmon run was the PY Model, as it had the second-smallest mean absolute percentage error (MAPE) and the smallest standard deviation (SD) of the MAPE (Brenner and Munro 2017) as compared to the other forecast models (Table 1). Below, I have outlined the methods used for each of the seven forecasting models, and discussed alternative methods. I have also outlined and critiqued ADF&G's model selection methods.

Table 1. Summary of 2017 forecast model results for Copper River Chinook Salmon. The Previous Year Model (yellow highlighting) was selected by ADF&G as the best forecast for 2017. Values in this table are similar to those in the "2017 Copper River Chinook forecast to NVE_5_15_2017.xlsx" Excel file, but differ from the file provided to NVE because errors have been corrected (see Appendix B).

		Total Run [a]				SD of
Forecast Model	n	L 80% [b]	Point	U80% [b]	MAPE	MAPE
Previous Year	17	3,000	29,000	55,000	29%	21%
Previous 2–years average	16	18,000	43,000	67,000	28%	23%
Previous 3–years average	15	15,000	40,000	66,000	31%	28%
Previous 4–years average	14	15,000	41,000	67,000	30%	34%
Previous 5–years average	13	14,000	42,000	70,000	35%	33%
Previous 10-years average	8	8,000	48,000	89,000	63%	39%
Hybrid Sibling: (Ln1.3=Ln1.2) and previous year return of remaining age classes [c]	18	30,000	44,000	68,000	35%	33%

Note: MAPE = mean absolute percentage error, SD = Standard Deviation. All naive models used 2000–2016 run data. The sibling relationship used 1977–2016 data.

a) Total run includes commercial harvest of fish that did not originate from Copper River stocks.

b) The total run 80% prediction intervals for all forecasts except the Hybrid Sibling Model were calculated using the mean performance of retrospective forecasts (Fried and Yuen 1987).

c) The sibling model was calculated using natural log transformed data. Point estimate and 80% prediction intervals were back transformed using methods from Hayes et al. (1995). The hybrid model prediction intervals for all age classes were calculated as the sum of the forecast point estimates for all age classes plus/minus the sum of squared differences between the individual age class point estimates and lower and upper prediction intervals.

Forecast Methods

PY (Previous Year) Model

For the PY Model, the 2017 total-run forecast (F_{2017}) was expected to be equal to the 2016 year total run (TR_{2016}):

$$F_{2017} = TR_{2016} \quad \text{Equation 1}$$

Since the total run for 2016 was determined to be 29,211 fish, the forecast for 2017, rounded to the thousand fish, was 29,000 (see "point" column in Table 1).

The 80% prediction interval (80%PI)[1] around this forecast was calculated using the data from n = 17 years (2000 to 2016). The method evaluated how well the model actually worked over the years. Figure 1 shows a plot of observed vs. forecasted values, where each dot is a year (the position in vertical space

[1] The 80% prediction interval is the range of values between which we are 80% sure that the observed value will lie.

is based on the actual run for that year, and the position in horizontal space is based on the forecast that would have been made for that year, had the PY model been used to make the forecast). The formula for the 80%PI around the 2017 forecast was:

$$80\%PI = F_{2017} \pm RMSE_{ADF\&G} \times t_{\{0.2,df\}} \qquad \text{Equation 2}$$

The term $RMSE$ (root mean square error) in Equation 2 was calculated as:

$$RMSE_{ADF\&G} = \sqrt{\frac{\sum_y^n (F_y - TR_y)^2}{n-1}} \qquad \text{Equation 3}$$

where F_y = is the forecasted total run value for year y, and TR_y = is the actual observed total-run value for year y. $RMSE$ is a measure of the average 'error' of the forecast. Basically, you calculate the difference between the forecast and the actual run for each year, square all those differences, sum up all the squared differences, divide the sum by *n*-1, and then take the square root. The differences

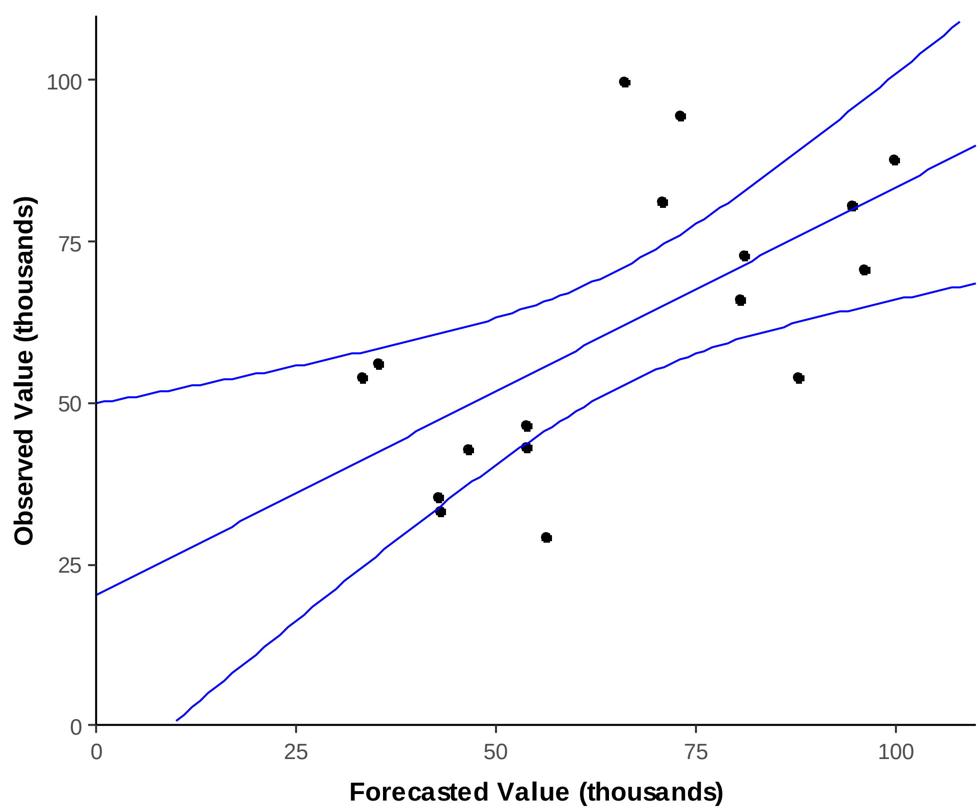

Figure 1. Observed vs. forecasted Total Run values for Copper River Chinook Salmon, using the PY Model to forecast. The blue line is a linear regression, and the two curved blue lines show the 95% confidence interval of the slope (we are 95% certain that the true slope will lie somewhere in the space between the two curved lines).

between actual and forecast values are squared before they are summed in order to preserve the overall magnitude of the forecasting errors. Otherwise, large overestimates in some years would be offset to some extent by large underestimates in others, which would tend to produce a sum that is close to zero (falsely indicating that the forecasts in each year are generally very good). Thus, to reflect the magnitude of forecasting errors encountered each year regardless of sign, the differences are squared (squares are always positive, so the sum of the squared differences does not have the problem). The squared differences are averaged to give an overall measure of model success, but the 'average error' is in units of 'squared fish'. By taking the square-root, the 'average error' is put back into terms of 'numbers of fish' (rather than 'squared fish'). This method is described in detail in Haeseker et al. (2008), except that in that paper, the RMSE denominator was *n*, instead of *n*-1.

$$RMSE_{\ Haeseker} = \sqrt{\frac{\sum_i^n (F_i - TR_i)^2}{n}} \qquad \text{Equation 4}$$

I could not find a reason why the ADF&G analysts used *n*-1, although the ADF&G method would result in wider prediction intervals (i.e., ADF&G would err on the side of caution), and the differences were small.

The other term in Equation 2 is the t-distribution, $t_{\{0.2,df\}}$ which depends on two parameters: the probability (determines the size of bounds being calculated); and the degrees of freedom (df).

for 80% PI bounds, the probability = 1 - 80% = 0.2

and *df* = *n* - 1 = 17 - 1 = 16.

Thus $t_{\{0.2,df\}} = t_{\{0.2,16\}}$ = 1.3368. In the olden days, we would have had to use statistical tables to look up the value of t (e.g., Rohlf and Sokal 1995). Now, we can derive it directly in Excel using the function `=T.INV.2T(0.2,16)`.

The complete dataset that was used to calculate the PI bounds of the 2017 forecast is shown in Table 2. Specifically the bounds were calculated as:

$$80\%PI = F_{2017} \pm RMSE_{ADF\&G} \times t_{\{0.2,df\}}$$
$$80\%PI = 29{,}221 \pm 19{,}385 \times 1.3368$$
$$80\%PI = 29{,}221 \pm 25{,}913$$
$$80\%PI = 29{,}221 - 25{,}913 \ \ to \ \ 29{,}221 + 25{,}913$$
$$80\%PI = 3{,}308 \ \ to \ \ 55{,}135$$

which, rounded to the thousand-fish, is 3 to 55 thousand (see the L80% and U80% columns in Table 1).

It should be noted that the prediction interval methodology was not actually described in the ADF&G report (Brenner and Munro 2017). In Appendix C, Brenner and Munro (2017) said that the forecast range was calculated using the methods described for Coghill Lake Sockeye Salmon. However, the Coghill Lake methods (described in their Appendix B) involved multiple age classes which were summed to calculate the total forecast, thus some re-interpretation was required to try to guess how the Lake Coghill methods would have applied to the non-age-structured Copper River Chinook Salmon model.

It is also important to note that Brenner and Munro (2017) described the prediction interval methodology for natural-run Pink Salmon in Prince William Sound as "calculated using the squared deviations between the 1995–2015 odd-broodline retrospective forecasts and actual runs as the forecast variance: $\hat{y} \pm MSE \times t_{\{0.2,df\}}$ where $\hat{y}$ is the forecast prediction from the average of the recent 3 odd-year returns, t is the critical value, and MSE is the mean squared error". This sounded more like

Table 2. Complete dataset for the calculation of 80%PI bounds of the 2017 Copper River Chinook Salmon forecast using the PY Model.

Year (*y*)	$\boldsymbol{F_y}$	$\boldsymbol{TR_y}$	$\boldsymbol{(F_y - TR_y)^2}$
2000	95,909	70,749	633,025,600
2001	70,749	81,063	106,378,596
2002	81,063	72,960	65,658,609
2003	72,960	94,404	459,845,136
2004	94,404	80,479	193,905,625
2005	80,479	66,080	207,331,201
2006	66,080	99,639	1,126,206,481
2007	99,639	87,683	142,945,936
2008	87,683	53,847	1,144,874,896
2009	53,847	42,992	117,831,025
2010	42,992	33,184	96,196,864
2011	33,184	53,889	428,697,025
2012	53,889	46,442	55,457,809
2013	46,442	42,886	12,645,136
2014	42,886	35,322	57,214,096
2015	35,322	56,207	436,183,225
2016	56,207	29,221	728,226,205
Total Squared Error, $\sum_y^n (F_y - TR_y)^2$			6,012,623,465
$RMSE_{ADF\&G} = \sqrt{\frac{\sum_y^n (F_y - TR_y)^2}{n-1}}$			19,385
$t_{\{0.2,df\}}$			1.3368
1/2 Width 80%PI., $RMSE_{ADF\&G} \times t_{\{0.2,df\}}$			25,913

the prediction interval methodology that I inferred from ADF&G's Excel spreadsheet, except, importantly, that MSE (not defined in Brenner and Munro 2017) is *not* the same thing as RMSE (i.e., it is not square-rooted, thus would produce *much* larger bounds). Also, no information was given about how MSE was calculated, thus I was still not able to determine why *n*-1 was used in the denominator, instead of *n*.

<u>x-yr Average Model</u>

Another model that ADF&G explored, the '2-yr Average' Model, is one which forecasted the 2017 total-run forecast (F_{2017}) to be equal to the average of the total runs in 2015 (TR_{2015}) and 2016 (TR_{2016}):

$$F_{2017} = (TR_{2015} + TR_{2016})/2 \qquad \text{Equation 5}$$

Since the total run for 2015 and 2016 were determined to be 56,207 and 29,211 fish, respectively, the forecast for 2017 was calculated to be 42,714, which was rounded to 43,000 (see "point" column in Table 1).

The 80% prediction interval (80%PI) around this forecast was calculated using *n* = 16 data points. The reason that *n* was 1 lower than for the PY Model is because 2 years of data were required to calculate each average. Since the historical run data that were used for these forecasts spanned from 1999 to 2016, the first possible year that a 2-yr Average forecast could be made was for 2001 (based on the average of the 1999 and 2000 run sizes). Whereas the PY Model had a 2000 forecast (it was based on the 1999 run size), and the 2-yr Average Model did not, the *n* was one lower for the 2-yr Average Model (*n*=16) than for the PY Model (*n*=17).

Other than having one less data point, the method used to calculate the 80%PI was the same as for the PY Model. A new RMSE was calculated (the difference between the forecast and the actual run for each year all those differences were squared, all the squared differences were summed and then divided by *n*-1, and then square rooted). The value for the t-distribution, $t_{\{0.2,df\}}$ was different, since there were fewer degrees of freedom (*df* = *n* - 1 = 16 - 1 = 15): $t_{\{0.2,15\}}$ = 1.3406. Thus,

$$80\%PI = F_{2017} \pm RMSE_{ADF\&G} \times t_{\{0.2,df\}}$$
$$80\%PI = 42{,}417 \pm 18{,}365 \times 1.3406$$
$$80\%PI = 42{,}417 \pm 24{,}608$$
$$80\%PI = 42{,}417 - 24{,}608 \ to \ 29{,}221 + 24{,}608$$
$$80\%PI = 18{,}106 \ to \ 67{,}322$$

Rounded to the thousand fish, the 80% Predication Interval spanned from 18 to 67 thousand (see the L80% and U80% columns in Table 1).

The theory is identical for the 3-yr Average, 4-yr Average, 5-yr Average, and 10-yr Average models. Sample sizes for these models were 15, 14, 13, and 8, respectively (Table 1). Again, the differences in sample size among models resulted from the need to average over differing numbers of years.

<u>Hybrid Sibling Model</u>

For this model, the Age 1.3 forecast came from a regression on the previous year's Age 1.2 run; and the remaining ages were forecasted directly from the previous year's return for that age, in the same manner as the PY Model (but this time each age was forecasted separately). In order to make these forecasts, it was necessary to first determine the age distributions of the Copper River Chinook Salmon in each run year.

Annual Age Distributions

Each year, scale samples were collected from the commercial fishery (date ranges, sources, and some sample sizes were provided in the Excel spreadsheet), and the year-specific relative proportion of each age-class was assessed (where sum of the proportions over all the age-classes was 100%). The commercial-fishery proportion (pc) for each age (a) and year (y) was:

$$pc_{a,y} = {nc_{a,y}}/{\sum nc_{a,y}} \qquad \text{Equation 6}$$

where $nc_{a,y}$ was the number of fish of age a in the commercial sample for year y. The total Lower Copper River harvest (which includes total commercial harvest, homepack, personal use, CR District State subsistence, CR Delta Federal subsistence, and educational harvests) for each year (HC_y) was then 'partitioned' among the age classes according to their year-specific relative abundance in the scale samples. Thus numbers of fish harvested by age and year ($HC_{a,y}$) was:

$$HC_{a,y} = HC_y \cdot pc_{a,y} \qquad \text{Equation 7}$$

Similarly, there were scale samples collected each year from the upper river (date ranges, sources, and some sample sizes were provided in the Excel spreadsheet); and the year-specific relative proportion of each age was assessed. The upper-river proportion (pu) for each age (a) and year (y) was:

$$pu_{a,y} = {nu_{a,y}}\Big/{\sum nu_{a,y}} \qquad \text{Equation8}$$

where $nu_{a,y}$ was the number of fish of age a in the upper-river sample for year y. The estimated total in-river abundance (determined until 1998 from a proportion of sonar counts, and thereafter from mark-recapture results) for each year (NU_y) was then 'partitioned' among the age classes according to their year-specific relative abundance in the scale samples. Thus numbers of fish in the upper river, by age and year ($NU_{a,y}$), was:

$$NU_{a,y} = NU_y \cdot pu_{a,y} \qquad \text{Equation 9}$$

The total run for each age class in each year was calculated as the sum of the lower-river harvest and the in-river abundance as:

$$TR_{a,y} = HC_{a,y} + NU_{a,y} \qquad \text{Equation 10}$$

Age 1.3 Forecast

The forecast for Age 1.3 fish was calculated from a regression equation, where the independent variable (X-axis) was the return of the Age 1.2 fish in the previous year. Thirty-five years of data from brood years 1977 to 2011 (i.e., from age-5 return years 1982 to 2016) were used to generate a linear regression (Figure 2) of the ln-transformed Age 1.3 run in a given year vs. the ln-transformed Age 1.2 run from the previous run year (i.e., from the same brood year). Given that the 2016 run of Age 1.2 fish was 4,125 fish (ln-transformed to 8.32, red line in Figure 2), the regression predicts a ln-transformed value of 10.314, which is equivalent[2] to 32,752 Age 1.3 fish in 2017. The 80%PI (9.774-10.854), once back-transformed, spanned 19,093 to 56,188 Age 1.3 fish.

$$F_{a=1.3,2017} = 32{,}752$$
$$80\%PI_{a=1.3,2017} = 19{,}093 \ to \ 56{,}188$$

Forecast for All Other Ages

The forecasts for all age-classes except Age 1.3 were equal to the previous year's age-specific total run, as:

$$F_{a,2017} = TR_{a,2016} \qquad \text{Equation 11}$$

where $a \in \{0.1, 0.2, 1.1, 0.3, 1.2, 2.1, 0.4, 2.2, 3.1, 0.5, 1.4, 2.3, 3.2, 1.5, 2.4, 3.3, 6.1, 2.5, 3.4\}$.

[2] To correct for bias, Hayes et al. (1995) recommended adding half of the residual regression variance to the predicted values before back-transforming them into linear space. The residual regression variance is the residual SS (Sum of Squares) divided by residual df, all values that are outputted when one runs a linear regression using Excel or any standard statistical package.

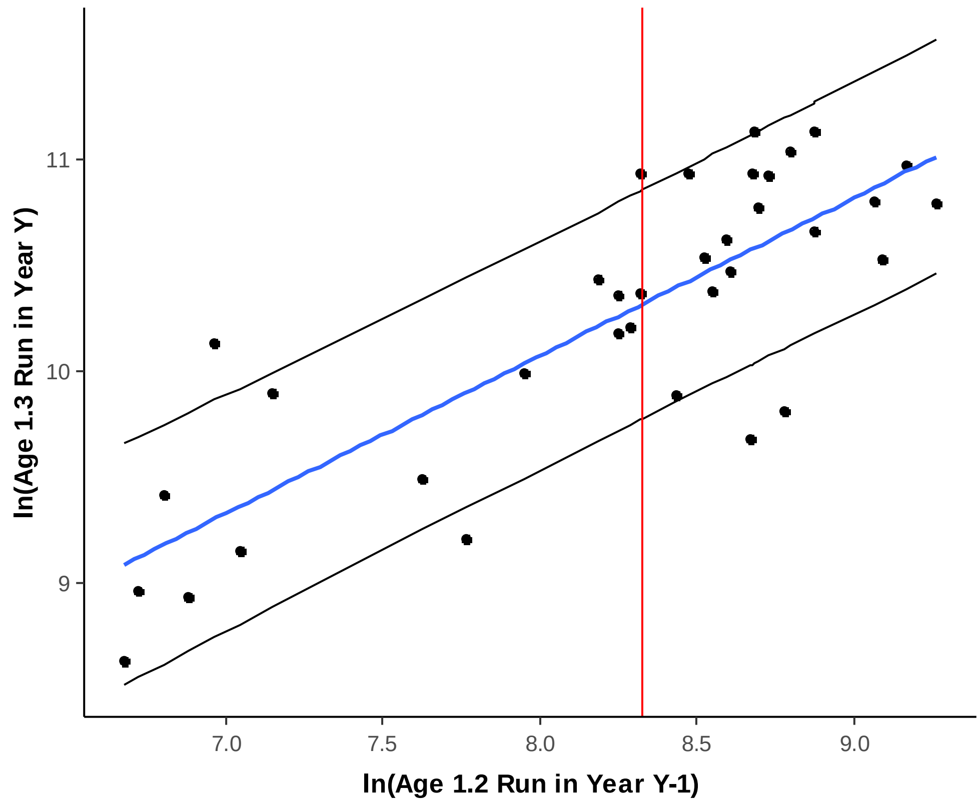

Figure 2. Age 1.3 Copper River Chinook Salmon in 35 survey years (run years 1982-2016; which corresponds to brood years 1977-2011) vs. Age 1.2 estimated from the previous run year (i.e., from the same brood year). The blue line is a linear regression. The vertical red line shows the 2016 age 1.2 value, thus the Y value where it intersects with the blue line is the forecasted 2017 value for Age 1.3 fish. The two black lines show the 80% prediction interval (we are 80% certain that the true value will lie somewhere in the space between the two black lines). All values are ln-transformed.

For each of these forecasts, the 80%PI was calculated using the standard deviation (σ) of the last five years of data, as:

$$80\%PI_{a,2017} = F_{a,2017} \pm \frac{\sigma_{a,y}}{\sqrt{5}} \times t_{\{0.2,4\}}$$ Equation 12

where the standard deviation was the square-root of the variance, and was determined using the usual formula, specifically:

$$\sigma_{a,y} = \sqrt{\frac{\sum_{y-1}^{y-5}\left(TR_{a,y} - \frac{\sum_{y-1}^{y-5} TR_{a,y}}{5}\right)^2}{5-1}}$$ Equation 13

Overall 2017 Hybrid Sibling Forecast

The 2017 total run forecast, derived from the Hybrid Sibling Model, was the sum of the forecasts for the individual ages,

Equation 14

$$F_{2017} = \sum_{a} F_{a,2017} + F_{a=1.3,2017}$$

and the 80%PI was calculated in a way that was suggested by Dave Eggers and Xinxian Zhang (both of ADF&G) in 2009 :

Equation 15

$$80\%PI_{2017} = \left(F_{2017} - \sqrt{\Sigma_a\left(F_{a,2017} - Lower80\%PI_{a,2017}\right)^2}\right) \text{ to } \left(F_{2017} + \sqrt{\Sigma_a\left(Upper80\%PI_{a,2017} - F_{a,2017}\right)^2}\right)$$

where $a \in$ all age-classes, including Age 1.3.

For 2017, the forecasts for Age 1.3 was 32,752. That for all the other ages combined was 11,140. Thus the total 2017 forecast was 43,892 which rounded to 44,000 (see the Point column in Table 1), and the 80%PI ranged from 29,984 to 67,505 fish, rounded to 30 to 68 thousand (see the L80% and U80% columns in Table 1).

Alternative Models

There are many other possible models that were not considered by ADF&G. Even within the scope of the seven models considered above, there are infinite variations as to how each could be executed. Take the 2-yr Average Model as an example: instead of averaging the values for the last two years, you could take the harmonic mean; you could take the geometric mean; you could account for the certainty in each value and use an average that is weighted by the inverse of the variance; you could take the sum of the squared observations, divide by two and then take the square-root; etc. Literally, anything is possible, and some may even be justified. Beyond the seven models considered by ADF&G, there is indeed *no limit* to the other possible models that could be developed. For example, we could develop models that use weather, and/or sea state, and/or productivity data as covariates to help with forecasting. Indeed, anything could be included.

In scrutinizing the ADF&G spreadsheet, I found that they had been exploring some models in addition to the main suite of seven that were included in Table 1. For example, they developed a set of models that used a two to five year lag between observation and forecast. Whereas in the PY Model the 2017 forecast was equal to the observed value from 2016, the 2-yr Lag Model instead used the observed value from 2015. The 3-yr, 4-yr and 5-yr Lag models used the observed values from 2014, 2013, and 2012, respectively. This type of model would be very appealing if the Copper River Chinook Salmon run was highly cyclical, such as the local Pink Salmon run (2-yr lag would be better than the PY model), or the Fraser River Sockeye Salmon run (4-yr lag), whose stocks are overwhelmingly made up of salmon with 4-yr life histories (the Fraser River Sockeye Salmon run is famously large every four years). However, in the specific case of the scrutinized ADF&G spreadsheet, their exploration was limited to predictions of Age 1.2 salmon from lagged Age 1.2 data; predictions of Age 1.3 salmon from lagged Age 1.3 data; and to predictions of Age 1.4 salmon from lagged Age 1.4 data: no attempt was made to predict the whole run from lagged total-run data.

Although models could have been explored that included covariates to help with forecasting, it should be noted that adding predictor variables to a model does not always help. Here, it is important to distinguish between model fit and forecasting performance. It should be kept in mind that the R^2 of a model (the model fit) *will* increase (it has to, mathematically) whenever predictor variables are added (it may not go up by much, but it will go up!). In fact, any model will have a perfect fit (R^2 =1) when it has a number of predictors equal to one less than the number of data points, regardless of how dumb the predictor variables might seem (ice cream sales in Timbuktu?). As an illustration, I just generated a model that fits the 1999-2016 Copper River Chinook Salmon returns (n=18) with a perfect R^2, using 17 sets of 18 random numbers as the predictors. Of course, there was never any hope that my model had anything but dismal forecasting performance. When I generated 17 new random numbers to forecast the 2017 Copper River Chinook Salmon returns, my forecast for 2017 was *negative* 14,664! "Garbage in – garbage out" is a common modelling expression: one really has to think carefully about what is going to be included in the predictive models. For example, it may sound like a good idea to include egg counts in a forecast model, but if the method used to count the eggs is hopelessly inaccurate, then it might be better just to not use it.

Model Selection

Once ADF&G's suite of seven forecast models had been developed, there was a need to select the best from among them. Generally, the best model is the one that produces forecasts that most closely match the actual observed values, and that does so most consistently. This section describes the methods that ADF&G used to determine which model to use for their official 2017 forecast.

There is no single method to characterize the central tendency and variability in the distribution of annual forecasting errors. While many performance measures have been proposed (see Lo and Gao 1997), each has its pros and cons. Some standard measures include the mean raw error (MRE), mean absolute error (MAE), the mean percent error (MPE), and the mean absolute percent error ($MAPE$), all of which have been used in the past to evaluate salmon forecasting models (Haeseker et al. (2008).

Raw error (RE) for a given year is the difference between the forecasted total run value (F_y) and the eventually- observed total run (TR_y) for that year:

$$RE_y = F_y - TR_y \quad \text{Equation 16}$$

Positive values of RE result when forecasts are too high, and negative values occur when forecasts are too low. For the MRE, the raw errors are averaged over the number of years (*n*) that were forecasted within each stock:

$$MRE = \frac{\sum_{y=1}^{n} RE_y}{n} \quad \text{Equation 17}$$

While the MRE reflects the overall bias of the forecasts, it is not particularly useful for describing how good any single forecast estimate might be, since large overestimates in some years would be offset to some extent by large underestimates in other years, thus producing an MRE near 0 (giving the false impression that the forecasts might be very accurate). So, in order to keep information about the magnitude of forecasting errors, regardless of sign, the absolute value of the errors is used:

$$AE_y = |F_y - TR_y| \quad \text{Equation 18}$$

$$MAE = \frac{\sum_{y=1}^{n} AE_y}{n} \quad \text{Equation 19}$$

The issue with MAE is that it tends to be larger for more abundant stocks, thus it makes more sense to express the differences in term of percent deviances. Thus, the percent error (PE) is used to calculate MPE; and more appropriately, for the reason discussed above, the absolute percent error (APE) is used to calculate $MAPE$:

$$PE_y = \frac{F_y - TR_y}{TR_y} \quad \text{Equation 20}$$

$$MPE = \frac{\sum_{y=1}^{n} PE_y}{n} \quad \text{Equation 21}$$

$$APE_y = \frac{|F_y - TR_y|}{TR_y} \quad \text{Equation 22}$$

$$MAPE = \frac{\sum_{y=1}^{n} APE_y}{n} \quad \text{Equation 23}$$

$MAPE$, a standardized measure of how the forecast performs on average, was selected by ADF&G as their measure of forecast performance. Along with $MAPE$, ADF&G also calculated the standard deviation (σ) of the $MAPE$ (a measure of how much the APE values varied among years):

$$\sigma_{MAPE} = \sqrt{\frac{\sum_{y=1}^{n}(APE_y - MAPE)^2}{n-1}} \quad \text{Equation 24}$$

Models that do the best job at forecasting actual returns will have the smallest $MAPE$ values. Models that work consistently will have the smallest σ_{MAPE} values (but note that models that are consistently wrong by the same amount will have very low σ_{MAPE}).

<u>Calculation of $MAPE$ for 2017</u>

As shown in Equations 23 and 24, $MAPE$ and σ_{MAPE} were calculated directly from the absolute percent error for each run year (APE_y), which was calculated using Equation 22. The APE for a given year (APE_y) was calculated as the difference between two values: TR_y (the observed values for that year) and F_y (the forecast that *would have been made* prior to that year, given the model in question and using *<u>only</u> the data that would have been in hand at that time*). These calculations were straightforward for the 2017 PY Model, x-yr Average models, and x-yr Lag models.

Table 3 illustrates the data used to calculate $MAPE$ and σ_{MAPE} for the 2017 PY Model forecast. Although only one data point was needed to make the 2017 forecast (only the 2016 total run is needed), a long time series (of historical forecasts paired with the eventually-observed total run values) was needed to assess the forecasting performance of the model.

The process was more complicated for the Hybrid Sibling Model. This model included two steps: 1) forecasting the individual ages; and 2) summing over all ages to calculate total run. Again, it was

Table 3. Complete dataset for the calculation of $\boldsymbol{MAPE}$ and $\boldsymbol{\sigma_{MAPE}}$ for the 2017 Copper River Chinook Salmon forecast under the PY Model. See text for formulas for $\boldsymbol{MAPE}$ and $\boldsymbol{\sigma_{MAPE}}$.

Year (y)	$\boldsymbol{F_y}$	$\boldsymbol{TR_y}$	$\boldsymbol{RE_y} = \boldsymbol{F_y - TR_y}$	$\boldsymbol{PE_y} = \boldsymbol{(F_y - TR_y)/TR_y}$	$\boldsymbol{APE_y} = \boldsymbol{\lvert(F_y - TR_y)\rvert/TR_y}$
2000	95,909	70,749	25,160	36%	36%
2001	70,749	81,063	-10,314	-13%	13%
2002	81,063	72,960	8,103	11%	11%
2003	72,960	94,404	-21,444	-23%	23%
2004	94,404	80,479	13,925	17%	17%
2005	80,479	66,080	14,399	22%	22%
2006	66,080	99,639	-33,559	-34%	34%
2007	99,639	87,683	11,956	14%	14%
2008	87,683	53,847	33,836	63%	63%
2009	53,847	42,992	10,855	25%	25%
2010	42,992	33,184	9,808	30%	30%
2011	33,184	53,889	-20,705	-38%	38%
2012	53,889	46,442	7,447	16%	16%
2013	46,442	42,886	3,556	8%	8%
2014	42,886	35,322	7,564	21%	21%
2015	35,322	56,207	-20,885	-37%	37%
2016	56,207	29,221	26,986	92%	92%
				Mean (i.e., $MAPE$)	29%
				Standard Deviation (i.e., σ_{MAPE})	21%

relatively straightforward to make the forecast estimate, but in order to calculate $MAPE$ and σ_{MAPE} it required a historical dataset whose generation required some attention to detail.

For the Hybrid Sibling Model, the first step was to calculate the historical forecasts of Age 1.3 fish. To make each annual forecast, special care was taken to use only the data that would have been in hand during the year in question. As can be seen in Table 4, the 1999 forecast of Age 1.3 fish (highlighted yellow) was generated based on the regression formula $\ln(Y_y) = 3.21 + (0.87 \cdot \ln(X_{y-1}))$, which was built using the 17 rows of data that would have been available at that time (brood years 1977-1993; corresponding to Age 1.2 data from years 1981-1997 and to Age 1.3 data from years 1982-1998). Data from the 1994 brood year would not be included in this regression model since, at the end 1998 (when the 1999 forecast was being made) the return of Age 1.3 fish would not yet be known (it would not be known until the end of 1999). Once the correct regression models were built, the Age 1.3 forecasts were calculated for each year y, by replacing X_{y-1} in each regression equation with the observed number of Age 1.2 fish that returned during year y-1. The Age 1.3 forecasts (blue column in Table 4) were put into a table with all the other age-class forecasts, and summed to calculate the retrospective total run forecast for each year (Table 5).

Table 4. Complete dataset for the forecast of Age 1.3 Copper River Chinook Salmon using a regression-based sibling model.

Brood Year	Age 1.2 Return Year	Age 1.3 Return Year	Observed Number of Age 1.2 fish	Observed Number of Age 1.3 fish	Forecast Formula	Sample Size	Brood Years included	Residual Regression Variance[2]	Forecast $\ln(Y_y)$	Back-transformed[2] Age 1.3 fish Forecast
1977	1981	1982	1,059	25,211						
1978	1982	1983	3,577	33,918						
1979	1983	1984	2,052	13,184						
1980	1984	1985	900	12,240						
1981	1985	1986	2,829	21,884						
1982	1986	1987	2,352	9,999						
1983	1987	1988	829	7,775						
1984	1988	1989	970	7,598						
1985	1989	1990	793	5,597						
1986	1990	1991	1,269	19,787						
1987	1991	1992	1,146	9,375						
1988	1992	1993	3,824	31,348						
1989	1993	1994	4,111	31,721						
1990	1994	1995	4,110	55,953						
1991	1995	1996	7,143	42,452						
1992	1996	1997	5,839	56,034						
1993	1997	1998	7,114	68,371						
1994	1998	1999	6,161	55,667	$\ln(Y_y) = 3.21 + (0.87 \cdot \ln(X_{y-1}))$	17	1977-1993	0.161695	10.80	53,278
1995	1999	2000	10,515	48,502	$\ln(Y_y) = 3.13 + (0.88 \cdot \ln(X_{y-1}))$	18	1977-1994	0.152424	11.29	86,438
1996	2000	2001	4,773	55,981	$\ln(Y_y) = 3.54 + (0.83 \cdot \ln(X_{y-1}))$	19	1977-1995	0.155185	10.53	40,611
1997	2001	2002	8,615	49,174	$\ln(Y_y) = 3.42 + (0.84 \cdot \ln(X_{y-1}))$	20	1977-1996	0.154709	11.06	68,906
1998	2002	2003	6,611	61,859	$\ln(Y_y) = 3.56 + (0.82 \cdot \ln(X_{y-1}))$	21	1977-1997	0.149646	10.81	53,359
1999	2003	2004	5,977	47,494	$\ln(Y_y) = 3.48 + (0.84 \cdot \ln(X_{y-1}))$	22	1977-1998	0.144419	10.75	49,910
2000	2004	2005	5,020	37,628	$\ln(Y_y) = 3.47 + (0.84 \cdot \ln(X_{y-1}))$	23	1977-1999	0.137564	10.60	43,049

2001	2005	2006	5,892	68,219	$\ln(Y_y) = 3.48 + (0.83 \cdot \ln(X_{y-1}))$	24	1977-2000	0.131497	10.73	48,876
2002	2006	2007	9,564	57,979	$\ln(Y_y) = 3.38 + (0.85 \cdot \ln(X_{y-1}))$	25	1977-2001	0.132263	11.17	75,724
2003	2007	2008	5,160	32,190	$\ln(Y_y) = 3.47 + (0.84 \cdot \ln(X_{y-1}))$	26	1977-2002	0.128259	10.63	44,139
2004	2008	2009	3,974	27,037	$\ln(Y_y) = 3.51 + (0.83 \cdot \ln(X_{y-1}))$	27	1977-2003	0.125538	10.40	35,048
2005	2009	2010	4,591	19,725	$\ln(Y_y) = 3.52 + (0.83 \cdot \ln(X_{y-1}))$	28	1977-2004	0.122142	10.51	39,148
2006	2010	2011	5,401	41,073	$\ln(Y_y) = 3.58 + (0.82 \cdot \ln(X_{y-1}))$	29	1977-2005	0.131489	10.62	43,836
2007	2011	2012	5,474	35,349	$\ln(Y_y) = 3.58 + (0.82 \cdot \ln(X_{y-1}))$	30	1977-2006	0.126793	10.63	44,221
2008	2012	2013	3,822	26,390	$\ln(Y_y) = 3.61 + (0.82 \cdot \ln(X_{y-1}))$	31	1977-2007	0.123270	10.33	32,705
2009	2013	2014	5,821	15,967	$\ln(Y_y) = 3.61 + (0.81 \cdot \ln(X_{y-1}))$	32	1977-2008	0.119916	10.67	45,776
2010	2014	2015	8,845	37,232	$\ln(Y_y) = 3.79 + (0.79 \cdot \ln(X_{y-1}))$	33	1977-2009	0.146497	10.96	61,808
2011	2015	2016	6,472	18,133	$\ln(Y_y) = 3.94 + (0.77 \cdot \ln(X_{y-1}))$	34	1977-2010	0.147386	10.69	47,189
2012	2016	2017	4,125		$\ln(Y_y) = 4.12 + (0.74 \cdot \ln(X_{y-1}))$	35	1977-2011	0.165489	10.314	32,752

Table 5. Complete dataset for the calculation of $\boldsymbol{MAPE}$ and $\boldsymbol{\sigma_{MAPE}}$ for the 2017 Copper River Chinook Salmon forecast under the Hybrid Sibling Model. Forecasts of Age 1.3 fish (blue highlighting) came from a sibling regression model (see Table 4 and Figure 2). All other forecasts are equal to the age-specific run from the year prior. Total forecast (F) is the sum of all ages for that run year. Total Returns (TR) are the eventually-observed values (including total system harvest and spawning escapement) that correspond to the forecast year. See text for formulas for Raw Error, Percent Error, APE, $\boldsymbol{MAPE}$ and $\boldsymbol{\sigma_{MAPE}}$.

Run Year	0.1	0.2	1.1	0.3	1.2	2.1	0.4	1.3	2.2	3.1	0.5	1.4	2.3	3.2	1.5	2.4	3.3	1.6	2.5	3.4	F	TR	Raw Error	Percent Error	APE
1999	0	44	180	44	6,161	44	0	53,278	192	0	0	29,707	349	0	0	643	0	0	0	0	90,643	95,951	-5,308	-6%	6%
2000	0	0	86	0	10,515	0	0	86,438	253	0	0	27,830	964	0	461	175	0	0	0	0	126,721	70,754	55,967	79%	79%
2001	0	0	247	0	4,773	0	0	40,611	241	0	0	16,005	541	0	77	368	0	0	0	0	62,862	81,139	-18,277	-23%	23%
2002	0	43	243	0	8,615	98	0	68,906	113	0	0	15,065	762	0	82	136	0	0	0	0	94,063	72,974	21,089	29%	29%
2003	0	10	160	0	6,611	25	0	53,359	452	0	0	14,144	1,411	0	28	959	0	0	0	0	77,159	94,555	-17,396	-18%	18%
2004	0	38	395	76	5,977	32	0	49,910	198	0	0	24,229	1,276	0	112	363	0	0	0	0	82,606	80,566	2,040	3%	3%
2005	0	0	371	115	5,020	20	0	43,049	206	0	0	26,018	1,056	0	0	266	0	0	0	0	76,121	66,357	9,764	15%	15%
2006	0	17	351	73	5,892	18	0	48,876	109	0	0	20,305	1,382	0	14	450	0	118	0	0	77,605	99,877	-22,272	-22%	22%
2007	0	0	364	0	9,564	0	0	75,724	851	0	0	16,360	3,549	0	414	556	0	0	0	0	107,382	87,771	19,611	22%	22%
2008	0	0	108	26	5,160	93	0	44,139	502	0	0	22,480	830	0	215	379	0	0	0	0	73,931	53,893	20,038	37%	37%
2009	0	16	7	0	3,974	16	0	35,048	199	0	0	17,049	396	0	22	25	0	0	0	0	56,751	43,007	13,744	32%	32%
2010	0	0	193	0	4,591	52	0	39,148	358	0	0	9,625	353	0	36	763	0	0	0	0	55,118	33,184	21,934	66%	66%
2011	0	0	358	0	5,401	128	0	43,836	222	0	0	6,759	452	0	0	140	0	0	0	0	57,296	53,890	3,406	6%	6%
2012	0	0	0	0	5,474	23	0	44,221	131	0	0	6,003	1,169	0	0	17	0	0	0	0	57,038	46,443	10,595	23%	23%
2013	0	12	3	0	3,822	12	0	32,705	69	0	0	5,329	984	0	832	31	0	0	0	0	43,799	42,902	897	2%	2%
2014	0	0	4,456	0	5,821	26	0	45,776	163	0	0	4,736	1,304	0	0	0	6	0	0	0	62,288	35,322	26,966	76%	76%
2015	0	0	491	0	8,845	25	0	61,808	437	0	0	8,848	586	0	4	120	0	0	0	0	81,164	56,220	24,944	44%	44%
2016	0	26	92	26	6,472	9	0	47,189	142	0	0	11,696	197	0	288	41	0	0	0	0	66,177	29,273	36,903	126%	126%
																								Mean (i.e., $MAPE$)	35%
																								Standard Deviation (i.e., σ_{MAPE})	33%

Alternative Methods

As stated above, $MAPE$ and σ_{MAPE} were selected by ADF&G as their measure of forecast performance. I have not seen anything in the literature to suggest a reason for using σ_{MAPE} as a measure of performance. Indeed, a model that performs consistently badly will have a very low σ_{MAPE}. And while $MAPE$ sounds very simple and convincing (and is a very widely used measure of forecast accuracy, Gneiting 2011), it has major drawbacks in practical application. One main criticism, that it cannot be used if there are zero values (because there would be a division by zero), would hopefully never apply to total run forecasts. More importantly, when $MAPE$ is used to compare the accuracy of prediction methods, it is biased in that it will *systematically select a method whose forecasts are too low* (Kolassa and Martin 2011; see also references cited in Tofallis 2015). This is because of asymmetry in the formula (Tofallis 2015): for forecasts which are too low the percentage error cannot exceed 100%, but for forecasts which are too high there is no upper limit to the percentage error. Consider two forecasts. In the first, a forecast of 10 is made, and the actual return was 100 fish (large underestimate). In the second, a forecast of 100 is made, and the actual return was 10 fish (large overestimate). Logically, these seem like they should have the same magnitude of error, but the two forecasts have APE values of 90% and 900%, respectively. By minimizing APE, the first forecast, that which underestimated the run, would be seen as the preferred forecast by a large margin.

Tofallis (2015) demonstrated that this serious asymmetry issue can be overcome by using an accuracy measure based on the ln of the Accuracy Ratio (Accuracy Ratio is the ratio of the predicted to actual value):

$$\ln(\mathrm{Q}) = \ln(F_y/TR_y). \qquad \text{Equation 25}$$

To demonstrate symmetry, take the example of 10 and 100 fish from the previous paragraph: the $\ln(\mathrm{Q})$ of the two forecasts are symmetrical in that they have the same magnitude (i.e., $|\ln(10/100)| = |\ln(100/10)|$). Tofallis (2015) demonstrated that models fit using the least squares ($\sum(\ln(Q)^2)$) outperformed many other methods, particularly when data were strictly positive and heteroscedastic (as are total run data). To use AR measures to evaluate different Copper River Chinook Salmon forecasts, one could use Median Symmetric Accuracy (Morely 2016):

$$\zeta = \exp(Median(|\ln(\mathrm{Q})|) - 1 \qquad \text{Equation 26}$$

Using ζ, it would appear that the 5-yr Average model had the best performance, followed closely by the 4-yr Average Model (Table 6).

Other alternatives to $MAPE$ have been proposed in literature. Mean Absolute Scaled Error ($MASE$, Hyndman and Koehler 2006) is another viable alternative. Hyndman and Koehler (2006) defined a scaled error as one which clearly independent of the scale of the data. For scale, they used the average difference between the observed value and that from the previous year):

$$Scaler = \frac{1}{n-1}\sum_{y=2}^{n}|TR_y - TR_{y-1}| \qquad \text{Equation 27}$$

such that

$$MASE = \frac{1}{n}\sum_{y=1}^{n}\frac{AE_y}{Scaler} \qquad \text{Equation 28}$$

Table 6. Performance of seven Copper River Chinook Salmon forecast models, as evaluated using four Model Performance metrics. Data sources as in Table 1. The best model under each selection method highlighted in green.

Forecast Model	$MAPE$	σ_{MAPE}	ζ (MSA)	$MASE$	Point Est. Total Harv.
Previous Year	29%	21%	25.2%	1.03	5,000
Previous 2–years average	28%	23%	29.4%	0.88	19,000
Previous 3–years average	31%	28%	25.0%	0.89	16,000
Previous 4–years average	30%	34%	24.6%	0.84	17,000
Previous 5–years average	35%	33%	24.2%	0.94	18,000
Previous 10-years average	63%	39%	62.5%	1.74	24,000
Hybrid Sibling	35%	33%	25.2%	1.03	20,000

Thus, $MASE$ is easily interpretable: values of $MASE$ less than one indicate that the forecasts are better, on average, than in-sample one-step forecasts from the naive method values (and $MASE$ greater than one indicate that the forecasts are worse than the one-step naive method). When applied to Copper River Chinook Salmon forecasts, the $MASE$ method showed that the 4-yr Average Model had the best performance (Table 6).

Other alternatives to $MAPE$ are less appealing: Symmetric Mean Absolute Percentage Error (sMAPE, Armstrong 1985) was shown by Tofallis (2015) to have directional bias; and Mean Directional Accuracy (MDA, e.g., Greer 2003) indicates direction only, ignoring magnitudes of errors altogether.

Summary

Based on the above, I recommend that ADF&G revise their method for selection among the competing models. The use of σ_{MAPE} does not appear to have any founding. The use of $MAPE$ is biased toward methods whose forecasts routinely underestimate observed values, resulting in lost harvest opportunities. Combined use of σ_{MAPE} and $MAPE$ resulted in the selection of the PY Model, one which allowed a total of harvest of 5,000 fish (point estimate) for the 2017 Copper River Chinook Salmon run. Methods that should be considered include $MASE$ and ζ (MSA), methods that are not biased by asymmetry. For the 2017 Copper River Chinook Salmon run, these last methods would have selected either the 4-yr or 5-yr Average Model, with total harvests (17-18,000 fish) more than triple that of the PY Model.

Literature Cited

Armstrong, J.S. 1985. Long-range forecasting: from crystal ball to computer, 2nd. ed. John Wiley & Sons, New York, NY.

Brenner, R.E., and A.R. Munro. 2017. Run forecasts and harvest projections for 2017 Alaska salmon fisheries and review of the 2016 season. ADF&G Special Publication 17-08.

Fried, S.M., and H.J. Yuen. 1987. A synopsis and critique of forecasts of Sockeye Salmon (*Oncorhynchus nerka*) returning to Bristol Bay, Alaska, in 1987. ADF&G Fishery Research Bulletin 87-01.

Gneiting, T. 2011. Making and evaluating point forecasts. Journal of the American Statistical Association 106: 746-762.

Greer, M. 2003. Directional accuracy tests of long-term interest rate forecasts. International Journal of Forecasting 19: 291-298.

Haeseker, S.L., R.M. Peterman, Z. Su, C.C. Wood. 2008. Retrospective evaluation of preseason forecasting models for Sockeye and Chum salmon. North American Journal of Fisheries Management 28: 12-29.

Hayes, D.B., J.K.T. Brodziak, and J.B. O'Gorman. 1985. Efficiency and bias of estimators and sampling designs for determining length-weight relationships of fish. Canadian Journal of Fisheries and Aquatic Sciences 52: 84-92.

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

Kolassa, S. and R. Martin. 2011. Percentage errors can ruin your day (and rolling the dice show how). Foresight 23: 21-27.

Lo, B.W.N., AND X. Gao. 1997. Assessing software cost estimation models: criteria for accuracy, consistency and regression. Australasian Journal of Information Systems 5: 30-44.

Morely, S.K. 2016. Alternatives to accuracy and bias metrics based on percentage errors for radiation belt modeling applications. Report for Los Alamos National Laboratory, LA-UR-16-24592.

Rolpf, F.J., and R.R. Sokal. 1995. Statistical Tables. 3rd Edition. W.H. Freeman and Company, New York, NY.

Tofallis, C. 2015. A better measure of relative prediction accuracy for model selection and model estimation. Journal of the Operational Research Society 66: 1352-1362.

Appendix A: Data Sources

I am not in a position to validate any of the data, nor assess whether alternative sources should have been used. Based on ADF&G's Excel file, data sources were as follows:

1. Harvests

1.1. Copper River District (below Baird Canyon)

LCR Total Harvest was calculated as a sum of Commercial Harvest (includes Copper and Bering rivers), Homepack / Personal Use, CR District State Subsistence, CR Delta Federal Subsistence, and Educational harvests.

1.1.1. Copper River Harvest

1893-1971: Annual Management Report 1972-73. Pirtle, Fridgen, and Bailey. Appendices 6 and 7. Citing:

- 1896-1926: Rich, Willis H & Edward M Ball. 1932. Statistical Review of the Fisheries, Part III: PWS, CBR. US Dept Commerce, BCF, Bulletin No 7. Mostly calculated from case pack and avg size.
- 1928-1951: Seton H Thompson. 1964. The red salmon of the Copper R, AK.
- 1952-1959: INPFC publications. Unknown which?
- 1960-1971: This AMR ("ADF&G Field Data")

1972-1973: PWS Area Annual Finfish Management Report, 1981. Randall, Fridgen, McCurdy, and Roberson. 1982

1974-1996: PWS Management Area 1996 Annual Finfish Report. Morstad, Sharp, Wilcock, and Johnson. Mark Sommerville database from December 2010 for upriver harvests, 1960-2009 or 2010.

1987 -2015: OceanAK FT query as of 12-15-2015. The 1987–2015 data include harvest codes 11 (state managed), 18 (confiscated), and 36 (donated). Delivery code 95 (homepack) and 35 (Educational) are in separate fields and are not included here.

Post 2015: source not provided

1.1.2. Berring River Harvest

Pre 1987: source not provided

1987 -2015: OceanAK FT query as of 12-15-2015. The 1987–2015 data include harvest codes 11 (state managed), 18 (confiscated), and 36 (donated). Delivery code 95 (homepack) and 35 (Educational) are in separate fields and are not included here.

Post 2015: source not provided

1.1.3. Homepack

1994-2015: OceanAK FT query as of 12-15-2015.

Post 2015: source not provided

1.1.4. CR District State Subsistence

1960-2003: Data from Mark Somerville database, AMRs, and Access database. Includes both state and federal subsistence harvests.

2004–2015: data are from Acess database as of 12-15-2015.

Post 2015: source not provided

1.1.5. CR Delta Federal Subsistence

2005–2014: Federal reports are all 0 Chinook Salmon.

Post 2014: source not provided

1.1.6. Educational

1987-2016: All data from commercial harvest reports. Some years entered in FT database, but most are not and are from separate reporting.

1.2. Upper River Subsistence and/or personal use

Includes State and Federal harvest from Chitina and Glennallen sub-districts.

1.2.1. Chitina SubDistrict State

1960-2015: From Mark Sommerville Database, December 2015. The 1960-2015 Chinook harvests are expanded to reflect unreported state harvest and include reported federal harvest (2002-2004) and expanded federal harvest beginning in 2005.

Post 2015: source not provided

1.2.2. Chitina SubDistrict Federal

1952-59: Data need to be source checked.

2002-2004: Per Mark Sommerville database December 2015. Reported Only

2005-2015: Per Mark Sommerville database December 2015. Data are expanded.

1.2.3. Glennallen SubDistrict State

1960-2015: From Mark Sommerville Database, December 2015. All expanded harvests.

1.2.4. Glennallen SubDistrict Federal

2002-2004: Per Mark Sommerville database December 2015. Reported Only

2005-2015: Per Mark Sommerville database December 2015. Data are expanded.

1.3. Sport

1966-76: source not provided

1977-2014: From Mark Sommerville database, 2015.

2015-2016: source not provided

1.4. Total Harvest

Total Harvest was calculated as a sum of the 'Copper River District (below Baird Canyon)' + 'Upper River Subsistence and/or personal use' + 'Sport'.

2. Escapements

2.1. Total Sonar Count (all Salmon)

1978-2011: Updated to DIDSON sonar equivalent numbers in 2011.

2012-2016: source not provided

2.2. Reported UCR Sub/PU harvests (For Chinook, Sockeye ad Coho salmon)

1978-2011: Reported harvests from Mark Sommerville database, December 2011. Includes all state and federal (2002 to present) harvests in the Chitina and Glennallen subdistricts.

2011-2013: From here: O:\DCF\SALMON\CATCH\111_Research_Catch_Analysis\111_CBR\113_UCR\UCR PU_SU Harvests\2013\UCR_Harvest_by_dy_and_week_State_Federal_2013

2.3. UCR proportion Chinook

Was calculated by dividing 'Reported UCR Sub/PU harvests for Chinook Salmon,' by the total for all three species.

2.4. In-river Abundance

2.4.1. Chinook in Sonar Count

Was calculated by multiplying 'UCR proportion Chinook' by the 'Total Sonar Counts'

2.4.2. Mark-recapture

1999-2014: See "O:\DCF\SALMON\ESCAPEMENT\116_TAGGING\111_CBR\111_Mark_Recap[CR Chinook salmon mark-recapture estimates 1999-2010.xls]Inriver Estimates and methods" for citations of estimates.

2015: NVE preliminary as of December 2015

2016: source not provided

2.4.3. Best In-river Abundance Estimates

ADF&G appeared to use the Mark-recapture data when available, otherwise the sonar proportions.

2.5. Estimated Spawning Escapement

2.5.1. Derived from Sonar Counts

1978-2014: Calculated by subtracting upper river harvest ('Upper River Subsistence and/or personal use' + 'Sport') from 'Chinook in Sonar Count'

2.5.2. Derived from Mark-Recapture

1999-2016: Calculated by subtracting upper river harvest ('Upper River Subsistence and/or personal use' + 'Sport') from 'Mark Recapture' numbers

2.5.3. Derived from Modelling

1980-1999: Estimated by age-structured model published in CJFAS by James Saveride and Dr. Terry Quinn.

2.5.4. Best Spawning Escapement Estimates

ADF&G appeared to use the Mark-recapture data when available, otherwise the modelling results.

3. Total Run Size

3.1. Total System harvest

1890-present: From “Harvests” sheet, ‘Total Harvest’ column.

3.2. Spawning Escapement

1980-1998: Spawning escapement from age structured model. From “Escapements” sheet, ‘Modeled’ column.

1999-present: Spawning escapement from mark-recaptured projects (1999- present). From “Escapements” sheet, ‘Mark-Recapture’ column.

3.3. Total Run

1980-present: Sum of ‘Total System Harvest’ and ‘Spawning Escapement’

4. Age Data

Dates, sources, and sometimes sample sizes provided in the “UCRPercentage” and “CCPercentage” sheets.

4.1. UCRPercentage

Prior to 1998: Brenner obtained from a pivot table within the Historical AWL file. These may not have been apportioned by harvest (1980-1982 from AMR; 1983-1997 from C&E).
1999-2013: APPB0213CRChin. Strata combined. Years back to at least 1998 are also APPB02. Need to convert years prior to 1998 and enter here. Source: C&E
2014: O:\DCF\SALMON\AWL\CURRENT\REPORTS\2014C&E. Source: C&E

4.2. CCPercentage

1980-1982: Source: AMR
1983-2015: Source: C&E
2005: APPA0105.xls
2008: APPA0108.xls
2009: APPA0109.xls This is the CR District Commercial Common Property Harvest.
2010: APPA0110.xls
2011: APPA0111.xls
2012: APPA0112.xls
2013: APPA0113CRChin. Strata combined

Appendix B: List of Corrected Errors and Suggested Improvements

1. Harvests Sheet

Cell D146: error in formula: I changed 130 to 134 (or could be changed to 135)

Cells B137-O141: fixed formulas to make them consistent within each row. Resulted in changes to min, max and n values. Warning: "offset" formulas can give wrong result if there are any non-numeric values in the time series (hence need for different start value for each column – BUT WATCH OUT FOR MISSING VALUES as they can give the wrong block of ten observations).

Cells Q8:Q134: updated formulas so they don't need to be manually updated each decade.

Cells S137:141: updated cells with Offset formula, so they don't need to be updated. Values in R137-141 dictate how many years the average is drawn over.

Figure (cell W94): changed figure title to show date range ending in 2016 instead of 2015.

2. Escapements Sheet

Cell B3: changed figure title to show date range ending in 2016 instead of 2015.

Added a "best spawning escapement" column that selects MR results when available, but switches to model results otherwise.

Added a "best inriver" column that selects MR results when available, but switches to sonar results otherwise.

3. Total Run Size Sheet

Cells M136-N137: updated cells with Offset formula, so they don't need to be updated. Values in L136-137 dictate how many years the average is drawn over.

Figure (cell T115): added figure title and years to X axis.

Column E: last two values are biased: the second-to-last does not include age 8 (they have not yet retuned), and the last does not include age 7 or 8 (not yet retuned). After that, I see that NAs are rightly used as values missing too many age classes would be unreliable.

4. Exploitation Rates Sheet

Cell A3: fixed a typo. RiverChinook -> River Chinook.

Cell A28 says "last 5 years (2008-2016)", but the 2008-2016 period is 9 years. Should say "2012-2016", and the formulas should be adjusted accordingly.

5. Forecast Summary Sheet

Sample sizes (column B) for 1-5 year predictors are wrong for rows 7:11 as they include the prediction year. Changed count formulas from looking at the 'forecast' columns (Sheet 'Mean Run Forecast' cols S:W) to looking at the 'error' columns (Sheet 'Mean Run Forecast' cols Z:AD). Sample size for 10-year predictor is correct (already pointing to Sheet 'Mean Run Forecast' cols AE).

Cumulative % error column is biased. There are more samples in the 1 year average than in the 2 year average than in the 3 year average etc. To sum over the full range of data means more values are being added to some predictors than to others. This could be valid if cumulative sums were only calculated over years for which there are values for *all* predictors (i.e., 1999 onwards).

The L80 and U80 columns may be mildly incorrect. Formulas used a n-1 denominator instead of n (see below). I can't find a reason why the n-1 was used instead of n, (although using n-1 produces wider PIs).

I don't see the "a", "b" or "d" footnotes within the table.

In the table notes, I moved the superscripts into their own column, thus allowing the 5-year average harvest (0.703) to be pulled from the Harvests sheet, instead of having to be typed in manually each year.

Columns M and N are not robust to edge cases like Column L is. I changed formulas to prevent negative values.

In Column T, I've added a selection box, so the user can select a model. Once selected, the appropriate row in the table gets coloured green, and the previously hard-coded table in cells X25:AB29 are automatically updated.

Cells X25:AB29 are strangely hard-coded. I replaced them with cell references to keep updated.

Fixed a typo in footnote b: performance -> performance.

Footnote b contains a reference to Fried and Yuen 1989. As far as I can tell, there is no paper by Fried and Yuen from 1989. Stormy Haught suggested that 1989 might have been a typo, and that 1987 was meant instead.

Cell E13: - round to nearest 1000 (formula pointing to wrong cell; change to C35)

Cell C13: - formula ended at 2015 instead of 2016.

Cells D13 and F13 – not calculated using the "blue cell" method (see Brood_Forecasts sheet for details. Blue cell method tightened up the bounds quite a bit.

Columns L80 and U80. These formulas add the PI to an already rounded point estimate. It would be better to add the PI to the exact prediction, and then round.

6. Mean Run Forecasts Sheet

Column Q formula mistakenly leaves out 2016 data.

AG3:AL3 – updated the sparklines

AP31:AP37: cumulative Percent error: this metric is biased. There are more samples in the 1 year average than in the 2 year average than in the 3 year average etc. To sum over the full range of data means more values are being added to some predictors than to others. This could be valid if cumulative sums were only calculated over years for which there are values for *all* predictors (i.e., 1999 onwards). I added these sums in cells AG27:AH27.

7. Mean Run Performance Sheet

Column AG ('Actual') referred to 1 year prediction (col B), instead of actual (col C). Replaced with more appropriate formula. Affects squared error (col AH) and percent error (col AI)

Cell D36 contains a note saying "Use this to calculate the total run 80% CI. Total run estimate plus/minus t(n-1,0.2)*mean total run forecast error See Fried and Yuen 1989 for details.". As far as I can tell, there is no paper by Fried and Yuen from 1989. Stormy Haught suggested that 1989 might have been a typo, and that 1987 was meant instead. Fried and Yuen 1987 does contain these formulas:

$$80\%\ \text{C.L.} = F_0 \mp [t_{0.20[df]} \times SE_0]\text{ , where}$$

$t_{0.20[df]}$ = Student's t value with a probability of type I error of 0.20 and df degrees of freedom;

df = sum of degrees of freedom of variance terms = n(N-1),

where N = number of years examined for each of the n methods used in the pooled forecast.

In the original paper, they used n(N-1) degrees of freedom. For Chinook, there is only one method (nothing is being pooled), thus n = 1 and df reduces to N-1. Where N is the number years, here 17, thus df = 17-1 = 16. We note, however, that SE_o is calculated very differently from what is used in the Chinook forecast.

Row 32: Mean Error. Can't find reason why denominator is n-1 instead of n. See below.

8. Brood Forecasts Sheet

Missing Column for Age 3.1 and for 1.6.. Makes many references to RunXAge incorrect. E.g., 1.4 should be an important age, but is showing up on this sheet as weak (with 2.3 showing as of major import).

Cells I27:I28: values were off due to apparent error on sheet Ln1.3=Ln1.2_3, but updated themselves automatically once the latter sheet was fixed.

Rows 27-28, simplified formula.

Top right "n=" formula wasincorrect.

Mean Error (was Y35, now AA35). Formula was incorrect (formula providing the square root of the sum of squared errors. Try dividing by n (or by n-1??), and if you still want to take sqrt, then call it RMSE. Drops value from 95k to 22k!!

Lower and upper 80% CI are not calculated the way the sheet says they should be (i.e., they don't follow the "blue cells rules"). I've redone the math and it tightens up the CI bounds quite a bit.

9. Sibling Forecast Sumry Sheet

Not used in analysis.

Footnote "a' was wrong. Should say 2010-2014.

Was not able to check validity of cells I15:O18 due to a missing macro, see below.

10. Naïve Age Specific Sheet

Not used in analysis. But includes lagged models that I thought could have been good.

C94:G94 had not been updated.

Cells I51:S51 referred to different cells than the formulas on rows 52 and 53. Same issue for rows 96-98. Same issue applied to 1.3 and 1.4 models. Fixed.

Cells O51:S51 and O96:S96 are biased – the same number of years must be compared for all models. Same issue for 1.3 and 1.4 models.

BU96:BY96: formulas pointed to subset of cells required. Fixed.

SD of MAPE is calculated for the 'Averaged' models (rows 54:100) but not for 'Lagged' models (rows 1:53). Not sure why.

11. RunXAge Sheet

Missing column for age 1.6. Results in mistake when referring to sheet that has this column (e.g., UCRNumber).

K46:K51: formulas missing.

Row 50: some formulas are pointing to 2011-2015 instead of 2012-2016.

Row 51: replaced lookup with native Excel TINV formula.

Cell A48: indicated that % is for last 5 years.

12. Ricker Sheet

Not used in analysis. Not called by any other sheet.

C11:C29: using proportion sonar data (Escapements sheet Col M), rather than model data (Escapements sheet Col O) – not sure why because elsewhere (e.g., the total run size sheet), they use model data instead of proportion sonar data.

Cells E40:E41 – data here a bit biased since the age 8 (cell E40) and age 7 and 8 (E41) fish have not yet been accounted for.

Cell Q14_ I cannot fathom a reason why this would be the correct formula.

Regression (output P22:X76) was done on the wrong data. This can be seen from the #Obs of 30, when there are actually 31 rows of data, and from the named X variable (cell P39) which should say something like “X-Variable”, but instead shows the first value [which implies that it used the first data-row as a label, which explains why the sample size is one less than it should be].

Col Z:AB – I don’t know why certain rows are highlighted. Perhaps these have not been updated recently?

13. UCR Percentage Sheet

Cell AA46, formula pointing to wrong cells.

Cell AA47 – should be 100%, can’t include non-100% rows as they are obviously incomplete.

Cells AA7:AA43 – if value is not 100%, then there is something wrong with the cells being summed. Changed formula to so that an error is shown in sum<>100, which has the added benefit of fixing CellAA47.

Rows 46 and 47: formulas are incorrect, as they ignore zeros. For example, Age 0.3 made up, for the last 10 years, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, 0%, and 0.73% of the return. The average thus is (0+0+0+0+0+0+0+0+0+0.73)/10 = 0.073%. But the formula that is in there is doing 0.73/1 = 0.73 – off by an order of magnitude. This can be fixed by filling-in the missing zeros from rows 7 to 43. I used conditional formatting to set zero values to be written in light gray so that they don’t cloud the overall pattern of data.

Footnote b is missing. I added it in. Footnote “a” is not cited in the table – Consider removing it.

14. UCR Number Sheet

Updated Column B so that it points to Escapements Sheet “Best Inriver” column, allowing a consistent formula to be filled down.

15. CC Percentage Sheet

Missing column for age 1.6. Results in mistake in sheets (e.g., RunXAge) that refer to this sheet.

Cell AA46, formula pointing to wrong cells.

Cells AA7:AA43 – if value is not 100%, then there is something wrong with the cells being summed. Changed formula to so that an error is shown in sum<>100.

Rows 45 pointing to 2006-2015 instead of 2007-2016. Fixed.

Rows 45 and 46: formulas are incorrect, as they ignore zeros. This can be fixed by filling-in the missing zeros from rows 7 to 43. I used conditional formatting to set zero values to be written in light gray so that they don’t cloud the overall pattern of data.

Rows 53:58: no idea what this is.

16. CC Number Sheet

Missing column for age 1.6. Results in mistake in sheets (e.g., RunXAge) that refer to this sheet.

Rows 47 pointing to 2006-2016 instead of 2007-2016. Fixed.

17. FBrood Sheet

Cols C, G and K refer to 1.2, 1.3 and 1.4 columns on Brood_Tab Sheet, but one value was manually changed from 0 to 1 (such that its ln drops down to 0 instead of #NUM). I changed the formulas to automatically change 0 values to 1.

The figure only uses data from BY 1995-2010, even though adequate data from 2011 exist. I changed the figure to extend into the most recent year of data.

Regression output in the lower right was hard-coded for an unknown set of years. These cells have been cleared, and replaced with formulaic treatments at the bottom of the data rows on the bottom left.

18. Brood_Tab Sheet

Added columns for ages 3.1 and 1.6.

Age 8 data referred to wrong columns in source material.

Cell W5, added "to date" to the heading "Total Brood Year Returns", since some are a work in progress (until all 8 ages have fully been accounted for).

Regression output in the lower right was hard-coded for an unknown set of years. These cells have been cleared

Row 54: replaced lookup with native Excel TINV formula.

19. Ln1.3=Ln1.2_3 Sheet

It looks like the regression output (hard-coded on right part of sheet was done incorrectly [can't reproduce result], hence the MSE value is wrong, thus value in cell J43 is wrong. Hence all other values in column J are suspect. Replaced the regression output, although I don't know that it is very useful. Included R script for the generation of the Regression Residual values. Replaced all values in Col J accordingly.

Cells H48:I50 (80% prediction intervals) were wrong (I could not reproduce them, so don't know what was done). Apparently done in SAS, but can't understand how SAS produced values that were so far off [note that the point estimates didn't even match those calculated on row 43 above!]. I have provided the R script to calculate the 80%CIs for the current-year prediction, and replaced the values in cells H48:I50.

20. Ln1.3=Ln1.2 Sheet

Not used in analysis. Called by Sibling_Forecast_Sumry Sheet. Appears to need a Jackknife Macro which was not included.

21. Ln1.3=Ln1.2_2 Sheet

Not used in analysis. Another version of Ln1.3=Ln1.2 Sheet (the latter is called by Sibling_Forecast_Sumry Sheet). Appears to need a Jackknife Macro which was not included.

22. Ln1.4=Ln1.3 Sheet

Not used in analysis. Called by Sibling_Forecast_Sumry Sheet. Appears to need a Jackknife Macro which was not included.

23. RegOut Sheet

Not used in analysis. Not sure what this is.

24. ttab Sheet

This is fine, but no longer needed as Excel has native TINV functions.

25. Major Age Brood Sheet

This sheet is hidden and does not appear to have been updated since 2015. Is probably no longer in use. Ignored.