United States Department of the Interior

FISH AND WILDLIFE SERVICE

1655 Heindon Road
Arcata, California, 95521
Phone: (707) 822-7201 FAX: (707) 822-8411

In Reply Refer To:
AFWO

Technical Memorandum

TO: Dave Hillemeier, Yurok Tribal Fisheries; Craig Tucker, Karuk Tribe

FROM: Nicholas A. Som and Nicholas J. Hetrick, Arcata Fish and Wildlife Office

SUBJECT: Response to Request for Technical Assistance – Predictive Model for Estimating 80% Outmigration Threshold of Natural juvenile Chinook Salmon Past the Kinsman Trap Site, Klamath River

DATE: May 15, 2017

Purpose. In August 2016, the Yurok and Karuk tribes made a formal request for technical assistance of the Arcata Fish and Wildlife Office Fish and Aquatic Habitat Conservation Program (AFWO), to develop a predictive tool for estimating when 80% of the natural juvenile Chinook Salmon had migrated downstream of the Kinsman trap site on the Klamath River. Development of the model was discussed during several subsequent “Disease Management Planning Team” meetings and conference calls hosted by NOAA Fisheries and Reclamation and, in general, was well supported by the group. In responding to the request, we shared our professional opinion that it was likely that a predictive model for estimating when 80% of the outmigration had occurred would be possible. We also stated, however, that we were unclear as to whether the tool would be a statistical model constructed using physical and/or biological variables, or instead rely on a simplistic approach based on the observed occurrence of the 80% outmigration threshold during a similar water year(s).

Model Development. The Kinsman trap site is located just upstream of the confluence of the Klamath River mainstem and the Scott River (Figure 1). The site is historically the most downstream location along the mainstem Klamath River where the abundance of juvenile salmonid outmigrants has been estimated using catch and capture efficiencies estimated from mark/recapture sample events.

In order for an 80% outmigration predictive tool to be effective, it needed to rely on data readily available by the onset of potential *Ceratonova shasta* disease risk. For this reason, neither weekly-stratified abundance estimates nor within-season weekly catch numbers were considered as predictor variables. Instead, we opted to consider measures of peak spawn timing and egg and fry development as candidate predictor variables. These measures included the weeks of both 50% and 80% (timing relative to seasonal total of each metric) carcass counts or redd counts from the preceding autumn; water temperature degree-day accumulations for the periods November through January, January through March, or November through March; and the mean fork length of juvenile Chinook Salmon measured during a survey encompassing the first week of April in the Kinsman reach of the mainstem Klamath River.

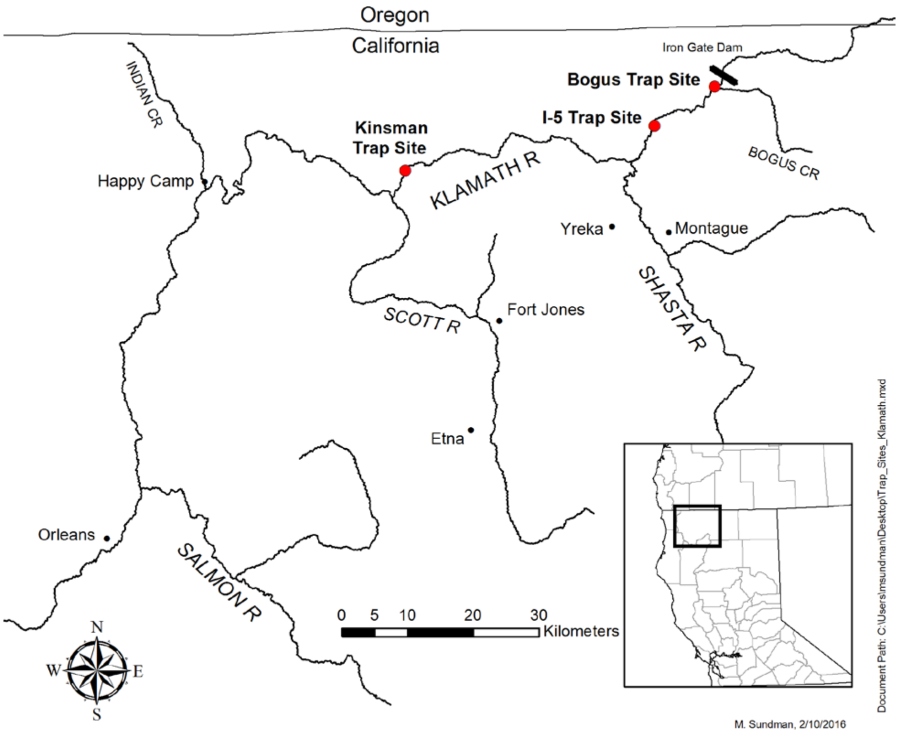


Figure 1. Location of the Kinsman Trap site within the lower section of the Kinsman (Shasta to Scott or "K4") fish heath monitoring study reach (taken from David et al. 2017).

The spawn timing predictor variables were calculated over the spatial extents of each survey, which are restricted to the Klamath River mainstem (Gough and Som 2015; Magneson and Colombano 2014). The degree day calculations were generated using water temperature data collected by the AFWO at a long-term monitoring station located in the mainstem Klamath River near Seiad (Magneson 2015). Water temperature data from this station were not available for the complete periods of degree day calculation for water years 2002 and 2004. To fill-in the missing water temperature data for the period of record analyzed, we used the RBM10 water temperature model constructed for the mainstem Klamath River (Perry et al. 2011). Fish size data were summarized from data collections associated with juvenile abundance (David et al. 2017) or fish health (True et al. 2016) monitoring programs. For numeric stability and multicollinearity prevention, all predictor variables were centered and scaled before model fitting.

For the response variable, we relied on the 80% outmigration timing estimates generated by the AFWO's juvenile Chinook Salmon outmigrant monitoring program (David et al. 2017). Weekly and total abundance estimates are generated using a Bayesian p-spline method (Bonner and Schwarz 2011) that couples weekly-stratified counts of outmigrating juveniles with mark-recapture experiments. This analysis procedure also produces proportional run-timing estimates, and the 80% values for each year were extracted as the response variable for development of this estimation tool.

Multiple linear regression modeling was chosen as the platform for predictive tool development. After a candidate set of predictive variables was narrowed to avoid potential multicollinearity issues, a model selection procedure evaluated the evidence for inclusion of single variables and interaction terms. It was established *a priori* that the simplest model (containing the fewest estimated parameters) among those within 2 Akaike Information Criteria (AIC) units of the best-fitting model would be selected and applied to estimate 80% outmigration timing.

Results. A total of 13 years had the requisite response variable information for fitting the predictive model. After visual assessment of the pairwise correlation between each candidate predictive variable and the response, the model selection procedure commenced considering the week that 50% of a year's carcasses had been observed (C50), the degree-day accumulation for January 1 through March 31 (ddJM), and the mean fork-length of fish measured on (or near) April 1 (FL).

The AIC model selection procedure resulted in the retention of the three univariate terms (C50, ddJM, FL) as well as interactions of C50 with FL and C50 with ddJM. This model had the lowest AIC value, and was at least 3 AIC units better than any simpler model (Appendix A). Additionally, analysis of residuals to assess the assumptions of this linear model demonstrated no evidence that the constant variance or normality of residuals assumptions were violated (Appendix B). Finally, the condition index for the selected model was 2.9, and therefore multicollinearity was not a concern for this model.

When model coefficients are examined individually, there is not strong evidence of statistical significance (Appendix C). However, the sample size is limited, there are a relatively high number of parameters being estimated, and the R^2 value for the model (0.63) suggests the model is explaining a majority of the variation in the response variable. Further, a visual assessment of predicted vs. observed values tracks a 1-to-1 line well, and the correlation between predicted and observed values is 0.79 (Figure 2).

Discussion. As requested by the Yurok and Karuk tribes, and later generally supported by the Disease Management Planning Team, a predictive tool was developed to estimate when 80% of naturally-produced Chinook Salmon had migrated downstream past the Kinsman trap site. Developing this decision support tool was challenging for several reasons, including the limited sample size of years having weekly-stratified outmigrant abundance estimates, and multiple source populations of Chinook Salmon that migrate through the Kinsman reach of the Klamath mainstem. One could argue that with a sample size of only 13, the model is over fit and contains too many estimated parameters. While this is a valid concern, spurious statistical results were not observed in the results of the analysis. Retrospective comparisons of R^2 values and the AIC ranking demonstrate a consistent pattern in regards to lack of fit and variance explained. Additionally, over fit concerns can be assed in future years by evaluating the out-of-sample prediction performance of the model. Finally, imprecision in outmigration timing estimates is well reflected in standard measures of regression model estimation and prediction uncertainty (Figure 3).

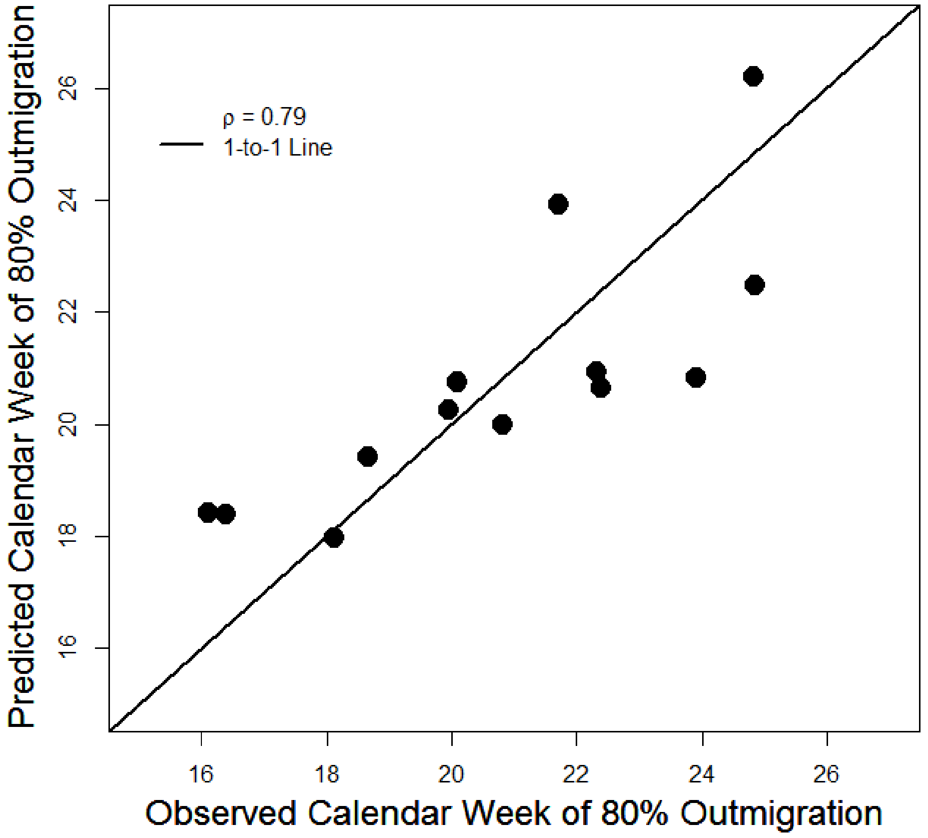


Figure 2. Plots of model predicted vs. observed values of 80% outmigration timing. The 1-to-1 line is provided for visual assessment of model performance, noting that a perfectly predicting model would place all points along the 1-to-1 line, and ρ is a measure of pairwise linear correlation.

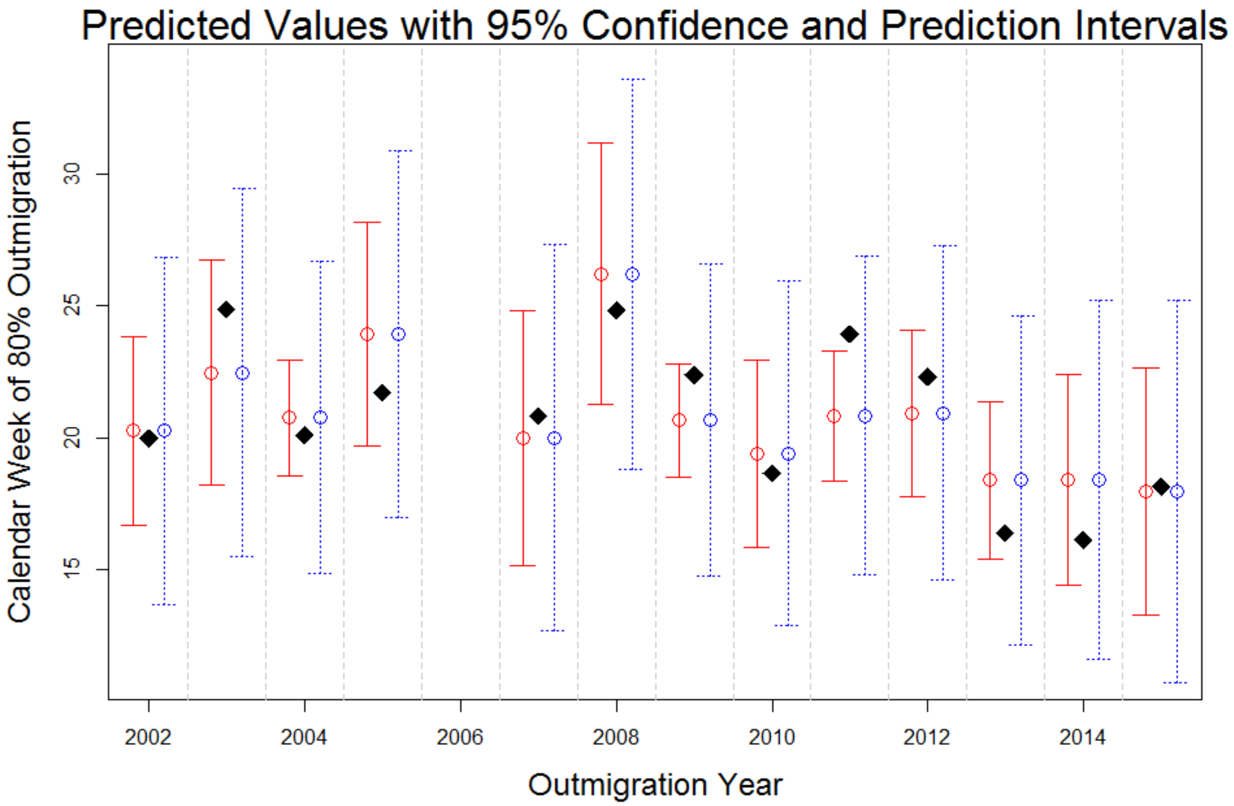


Figure 3. Observed (solid diamonds) and predicted (open circles) values of 80% outmigration timing with 95% confidence and prediction intervals, plotted by year. Confidence (solid red, slight left) and prediction (dotted blue, slight right) are offset to aid in visual inspection. Dashed light gray lines are provided to visually separate years. Confidence intervals express uncertainty related to the mean value given a set of predictor variable values, and prediction intervals express the uncertainty related to a single new observation given a set of predictor variable values.

References

Bonner, S.J. and C.J. Schwarz. 2011. Smoothing Population Size Estimates for Time-Stratified Mark-Recapture Experiments Using Bayesian P-Splines. Biometrics 67(4): 1498-1507.

David, A.T., S.A. Gough, and W.D. Pinnix. 2017. Summary of Abundance and Biological Data Collected During Juvenile Salmonid Monitoring on the Mainstem Klamath River Below Iron Gate Dam, California, 2015. U.S. Fish and Wildlife Service. Arcata Fish and Wildlife Office, Arcata Fisheries Data Series Report Number DS 2017-48.

Gough, S.A., and N.A. Som. 2015. Fall Chinook Salmon Run Characteristics and Escapement for the Mainstem Klamath River, 2012. U.S. Fish and Wildlife Service. Arcata Fish and Wildlife Office, Arcata Fisheries Data Series Report Number DS 2015–46, Arcata, CA.

Magneson, M.D. 2015. Klamath River flow and water temperature, Water Year 2012. U.S. Fish and Wildlife Service. Arcata Fish and Wildlife Office, Arcata Fisheries Technical Report Number DS 2015-42, Arcata, CA.

Magneson, M.D., and P. Colombano. 2014. Mainstem Klamath River Fall Chinook Salmon Redd Survey 2013. U. S. Fish and Wildlife Service, Arcata Fish and Wildlife Office, Arcata Fisheries Data Series Report Number DS 2014-40, Arcata, CA.

Perry, R.W., J.C. Risley, S.J. Brewer, E.C. Jones, and D.W. Rondorf. 2011. Simulating daily water temperatures of the Klamath River under dam removal and climate change scenarios: U.S. Geological Survey Open-File Report 2011-1243, 78 p.

True, K., A. Voss, and J.S. Foott. 2016. Myxosporean Parasite (Ceratonova shasta and Parvicapsula minibicornis) Prevalence of Infection in Klamath River Basin Juvenile Chinook Salmon, March - August 2016. U.S. Fish & Wildlife Service California – Nevada Fish Health Center, Anderson, CA. http://www.fws.gov/canvfhc/reports.asp.

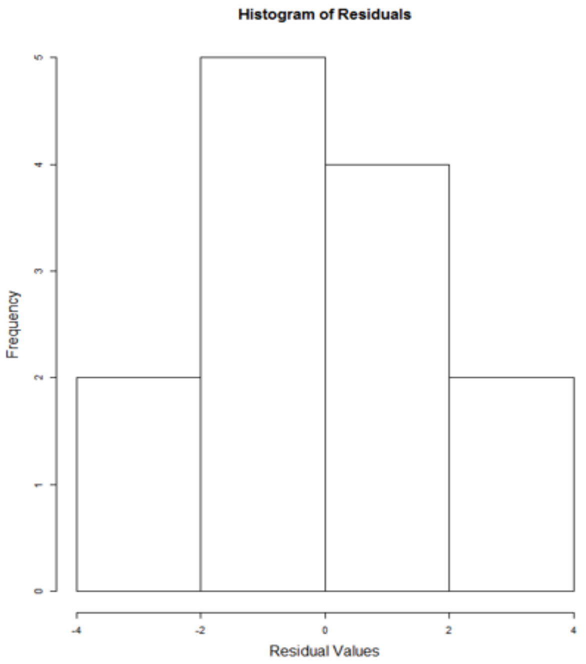
Appendix A. Model Selection Table

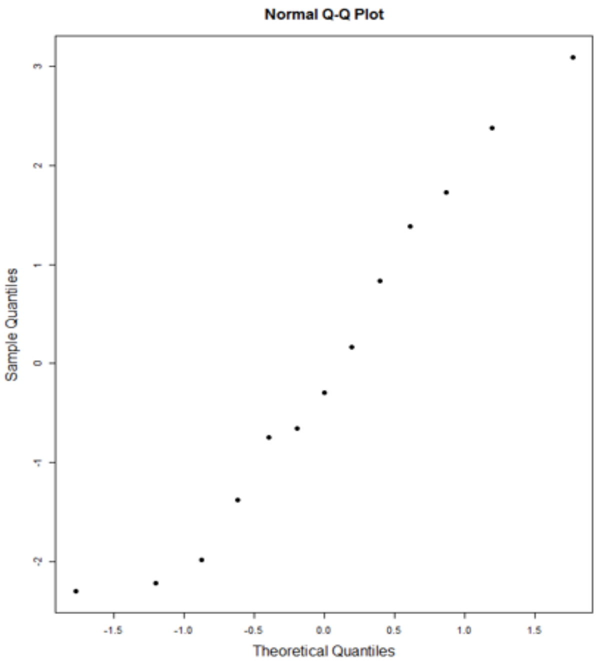
Model selection table used to select predictive tool. All models included an estimate of an overall intercept parameter, a variance parameter, and 3 slope parameters associated with the week that 50% of a year's carcasses had been observed (C50), the degree-day accumulation for January 1 through March 31 (ddJM), and the mean fork-length of fish measured on (or near) April 1 (FL). The terms listed in the "Interactions" column designate which interactions terms each model contained, and a ":" denotes interacting terms. The last cell of this column is empty because it did not contain any interaction terms. The "k" column represents the total number of estimated parameters, the "AIC" column lists the AIC value for each model, and the "ΔAIC" column displays each model's AIC value relative to the best overall value.

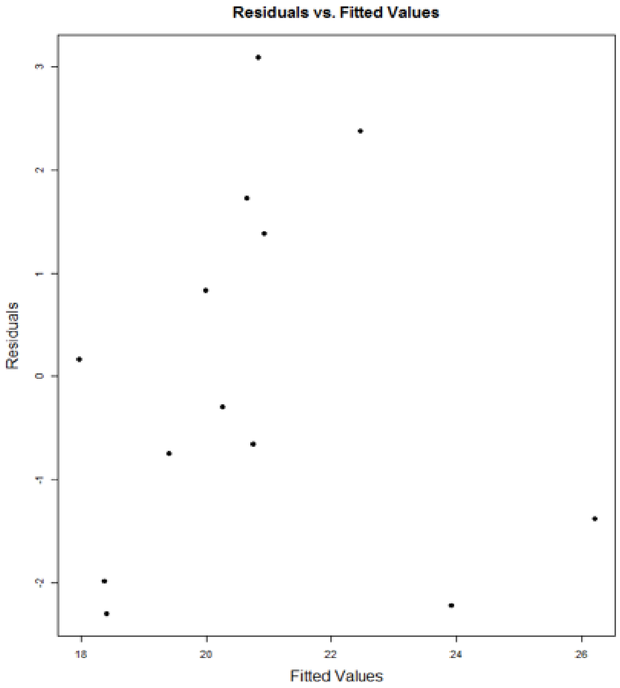
Interactions	k	AIC	ΔAIC
C50:FL; C50:ddJM; FL:ddjM	8	66.14	1.31
C50:FL; C50:ddJM	7	64.83	0
C50:ddJM; FL:ddjM	7	68.09	3.26
C50:FL; FL:ddjM	7	70.99	6.16
FL:ddjM	6	71.22	6.39
C50:FL	6	69.8	4.97
	5	70.7	5.87

Appendix B. Residual Diagnostics

Figures to visually assess the regression assumptions of constant, normally-distributed residuals.


Histogram of Residuals


Normal Q-Q Plot


Residuals vs. Fitted Values

Appendix C: Estimated Coefficients Table

Estimated coefficients, standard errors, t-statistics, and p-values for all coefficients in the model retained after the model selection procedure. Coefficients include slopes associated with the week that 50% of a year's carcasses had been observed (C50), the degree-day accumulation for January 1 through March 31 (ddJM), and the mean fork-length of fish measured on (or near) April 1 (FL). The ":" denotes interacting terms.

Coefficient	Estimate	St. Error	t-value	p-value
Intercept	20.00	0.76	26.37	<0.01
C50	-0.10	0.90	-0.11	0.92
FL	0.72	0.82	0.88	0.41
ddJM	0.63	0.79	0.80	0.45
C50:FL	-1.15	0.80	-1.44	0.19
C50:ddJM	-1.88	0.84	-2.23	0.06