Estimating Historical Salvage via Negative Binomial Regression

Background

The following document provides a summary of an analysis focused on evaluating the effects of CVP operations (i.e., exports and OMR) on historical monthly salvage (1993-2013) of Winter-Run salmon. The factors examined in this analysis include OMR, exports, prior year escapement, Sacramento River discharge at Verona, and month (December – April). We included prior year Winter-Run escapement and Sacramento River discharge as potential variables to account for potential effects of population size and flow in the analysis.

Analytical Framework

Historical salvage data were analyzed via negative binomial regression. This family of regression models is used to analyze count data (i.e., salvage count). Negative binomial regression is similar to Poisson regression except that it requires estimation of a dispersion parameter rather than assuming the variance is equal to the mean. In doing so, negative binomial regression can account for overdispersion, which is common in ecological data, and reduce the likelihood of biased coefficient estimation. Data for continuous predictor variables were centered and scaled (i.e., standard score or z-values) to facilitate model convergence.

A candidate set of models were developed ranging from a null model (i.e., intercept only model) to a full model that included all factors noted above and one, two-way interaction. All models were compared via Akaike Information Criterion corrected for small sample size (i.e., AICc). The goal of applying this model selection criterion was to identify the most parsimonious model (i.e., lowest AICc score) that best balanced the model’s complexity and fit to the data. Within this framework a difference of 4 AICc units is evidence of model superiority (Akaike 1974, Burnham and Anderson 2002).

Cross Validation

Leave-one-out cross validation (LOOCV) was applied to the top supported model to provide a measure model predictive performance. LOOCV involves removal of a single record from the dataset, refitting the top model to the remaining data, estimating the expected salvage count for the ‘out-of-sample’ data, and comparing the predicted vs. observed salvage count. This process is repeated for all records in the dataset. Linear regression is used to compare the relationship between observed and predicted salvage counts and the resulting R2 from this regression is a measure agreement between observed and predicted observations.

Data Availability

We used monthly total salvage records of Winter Run salmon collected between 1993 and 2013 (Figure 1). Overdispersion was apparent during initial inspection of these data (mean ≠ variance) supporting the use of negative binomial regression in this analysis. Less than 2% of the records from this dataset included counts of 0 salvage, and thus there was no need to explore the use of zero-inflated regression models.

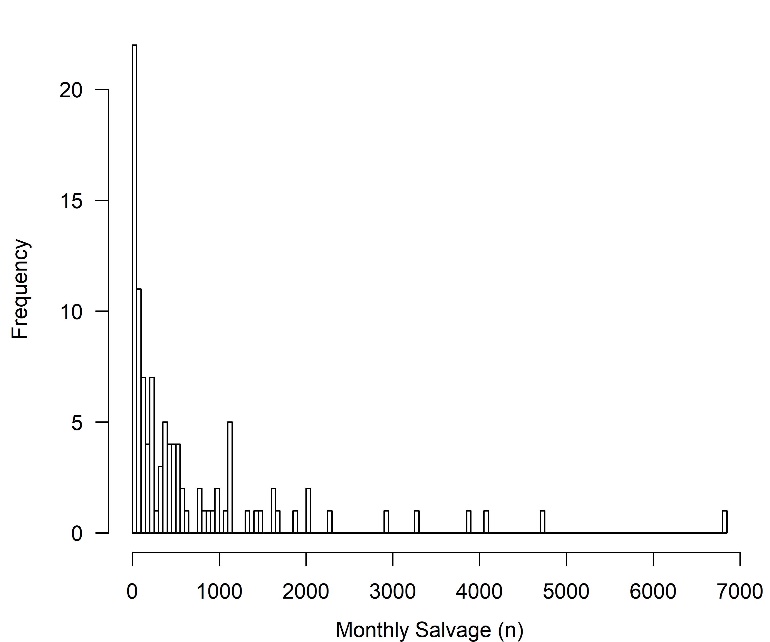


Figure 1. Frequency distribution of Winter Run salvage between 1993 and 2013.

Estimates of monthly average exports and OMR were included in the analysis to estimate hydrodynamic effects of SWP and CVP operations on salvage count. Monthly exports from SWP and CVP was measured as the sum of the average monthly export from SWP and CVP facilities. OMR was calculated as the monthly average OMR flow.

Winter-Run salmon escapement from the previous year was included in the analysis to account for differences in salvage count emerging from differences in juvenile fish production. Here it is assumed that the number of juvenile salmon outmigrating from the Sacramento River scales positively with the number of spawning adults in the wild and hatchery (i.e., total escapement). Data on spawning adults were obtained from the GrandTab database (https://www.wildlife.ca.gov/Conservation/Fishes/Chinook-Salmon/Anadromous-Assessment).

Measures of monthly average Sacramento River discharge (cfs) were also included in the model to account for upstream flow effects on salvage count. Here it is assumed that the volume of water flowing from the Sacramento effects routing and therefore impacts the likelihood juvenile salmon will end up at the pumping facilities and salvage. Data from historical Sacramento River discharge were taken from the USGC gauging station 11425500 at Verona (<https://waterdata.usgs.gov/usa/nwis/uv?site_no=11425500>).

Results

A total of 111 candidate models were constructed from different combinations of the factors described above and one possible two-way interaction. From this candidate set 103 models converged on a stable solution. There were four models that received similar quantitative support (i.e., ΔAICc < 4 units). The top supported model was the most parsimonious and included factors for Month, OMR, Exports and an interaction between Month and OMR (Table 1). The top-ranked model was substantially more supported than the null model (ΔAICc = 76.5) and had a corrected likelihood-ratio-based R2 for generalized linear models of 0.62. Collectively these results suggest the top-ranked model reasonably fit the observed data.

Table 1: Top-ranked negative binomial regression models from the candidate set based on lowest AICc score. All other models were >10 AICc units from the most supported model. Predictor variables included in the candidate models were OMR, Exports, Month, Escapement, and Sacramento River Discharge (SRD). The addition of fixed factors are noted by “+” and an interaction between terms is noted with “:”.

|  |  |  |
| --- | --- | --- |
| Model Rank | ΔAICc | Model Parameters |
| 1 | 0.00 | Month + OMR + Month:OMR + Exports |
| 2 | 0.45 | Month + OMR + Month:OMR + Exports + Escapement |
| 3 | 2.57 | Month + OMR + Month:OMR + Exports + SRD |
| 4 | 3.04 | Month + OMR + Month:OMR + Exports + Escapement + SRD |
| 5 | 6.36 | Month + OMR + Month:OMR |
| 6 | 7.37 | Month + OMR + Exports + Month:Exports |
| 7 | 7.54 | Month + OMR + Month:OMR + SRD |
| 8 | 8.56 | Month + OMR + Month:OMR + Escapement |
| 9 | 8.73 | Month + OMR + Exports + Month:Exports + SRD |
| 10 | 9.51 | Month + OMR + Month:OMR + SRD + Escapement |
| 11 | 9.79 | Month + OMR + Exports + Month:Exports + Escapement |

There were strong and significant effects of Month, OMR, and Exports on salvage (Table 2). In general, salvage is higher in February and March compared to December, January, and April (Figures 2, 3). There was a significant effect of increasing exports on salvage where for every ~3,200 cfs increase in exports salvage would increase approximately 1.6 times from the baseline (~150 fish). Decreasing OMR flows (i.e., more negative flow) also significantly increased the expected salvage count (Figures 2, 3; Table 2). However, it is important to note the significant interaction between OMR and Month, which means the relative impact of altering OMR will be significantly different across months.

Table 2: Top-ranked negative binomial regression model parameter coefficients, coefficient estimate standard error, Z-value, and P value. The intercept in this model is fit to average salvage in April with average OMR and Exports. Coefficients ‘adjust’ the expected salvage for the corresponding month, OMR, Export combination. Interaction terms are noted with “:”. Note data for OMR and Exports were centered and scaled and thus the parameter estimates are based on units of standard deviation for these parameters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Estimate | Std. Error | Z value | Pr (> | z |) |
| Intercept | 5.0552 | 0.2689 | 18.797 | < 0.001 |
| December | -0.5101 | 0.3658 | -1.395 | 0.163 |
| January | 1.0824 | 0.3688 | 2.935 | 0.003 |
| February | 1.5046 | 0.3568 | 4.217 | < 0.001 |
| March | 1.7985 | 0.3570 | 5.038 | < 0.001 |
| OMR | 0.2944 | 0.2046 | 1.439 | 0.150 |
| Exports | 0.4910 | 0.1554 | 3.160 | 0.002 |
| December:OMR | -1.3654 | 0.4828 | -2.828 | 0.005 |
| January:OMR | -1.5884 | 0.3279 | -4.844 | < 0.001 |
| February:OMR | -0.8414 | 0.2531 | -3.325 | < 0.001 |
| March:OMR | -0.9264 | 0.3348 | -2.767 | 0.006 |

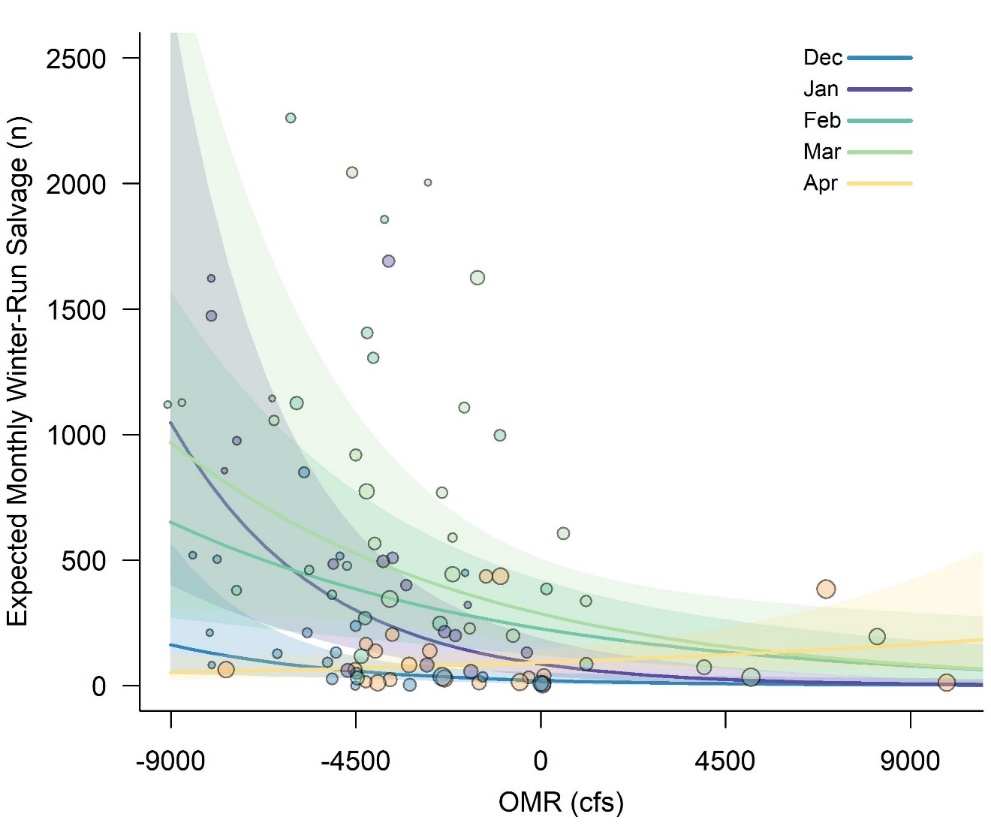


Figure 2. Negative binomial model fit when average SWP and CVP exports are approximately 1500 cfs. Each month is color coded where lines represent the mean model estimate and polygons encompass the 95% confidence intervals. Observed salvage data are plotted by color-coded points, where point size scales with export level for that data record.

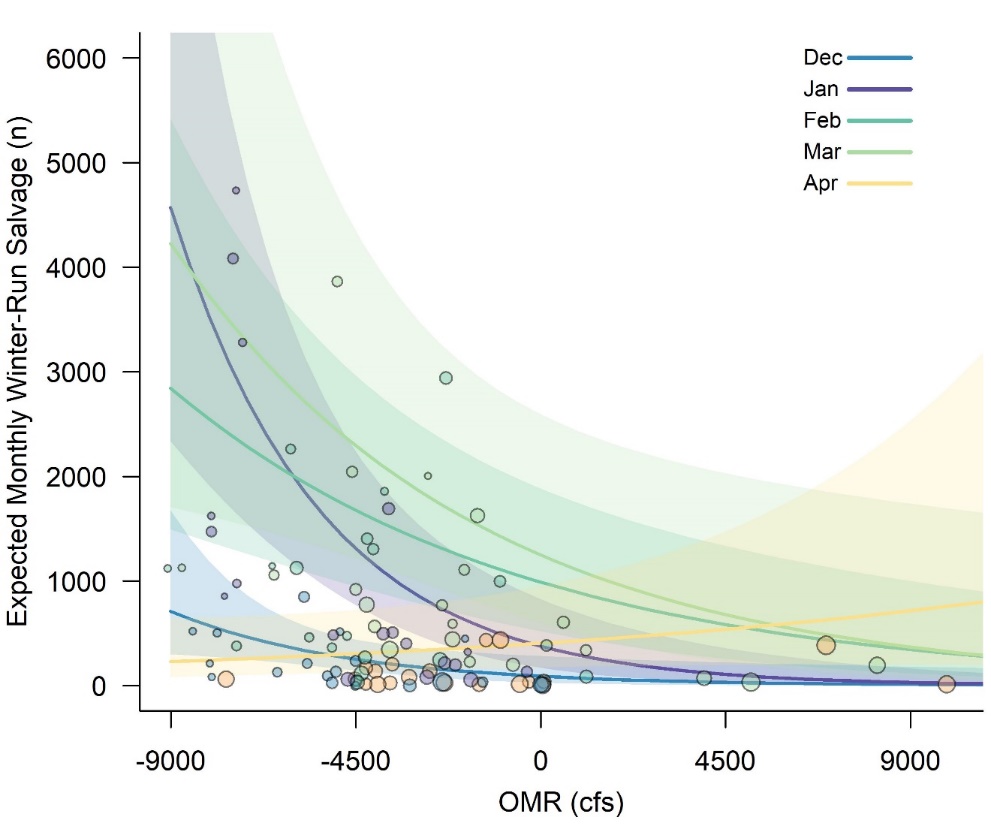


Figure 3. Negative binomial model fit when average SWP and CVP exports are approximately 12,000 cfs. Each month is color coded where lines represent the mean model estimate and polygons encompass the 95% confidence intervals. Observed salvage data are plotted by color-coded points, where point size scales with export level for that data record.

The top-ranked model has limited predictive capability despite fitting the observed data reasonably well. The correlation between observed and predicted data (log10 transformed) was positive but relatively weak (Adjusted R2 = 0.49). Across the range of observed data, the model could possibly over or under estimate salvage by up to an order of magnitude (Figure 4). Perhaps more importantly, the error in model predictions was not uniform with the model tending to overestimate salvage at low numbers and possibly under predict salvage at high numbers (Figure 4).

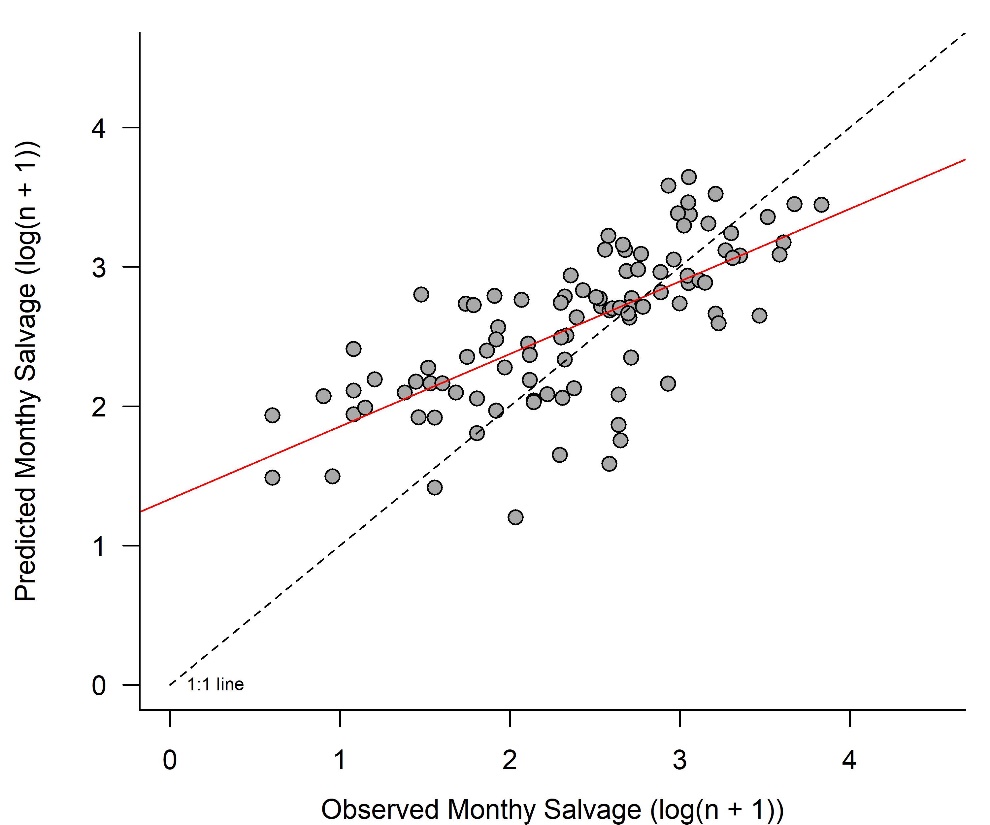


Figure 4: Observed and predicted monthly salvage based on leave-one-out cross validation. Note both x and y axis are on a 1og10 scale. The dashed 1:1line marks were observed and predicted estimates are equal. The linear model fit between observed and predicted data is illustrated by the red line.

Conclusion

Over 100 different models were evaluated to examine the effects of CVP operations (i.e., exports and OMR) on historical monthly salvage of Winter-Run salmon. There was very strong support for a model including Month, OMR, Exports, and an interaction between Month and OMR to explain past variability in salvage counts. Examining the fit of this model to historical data provides useful information for interpreting how changing OMR and/or Exports within different months will impact salvage number (Figures 2, 3). However, the model does a relatively poor job predicting “out of sample” salvage counts (Figure 4). As such, substantial caution is advised prior to using this model for making predictions regarding how CVP operations will impact salvage counts.

References

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716–723.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: a practical-theoretic approach. 2nd edition. Springer-Verlag, New York.