Salmon Salvage NegBin Model

Brian Mahardja

2022-09-22

# Estimating Salvage via Negative Binomial Regression

Authors: Brian Mahardja, Catarina Pien, Michael Beakes

## Background

To evaluate potential changes to the number of length-at-date winter-run and spring-run Chinook Salmon salvaged at the CVP and SWP pumping facilities based on the alternatives, Reclamation analyzed historical salvage data via negative binomial regression. Negative binomial regression 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 (e.g., the salvage dataset) as well as reduce the likelihood of biased coefficient estimation.

monthly\_data <- read.csv("negbinmodel\_monthly\_dataset.csv")

## Methods

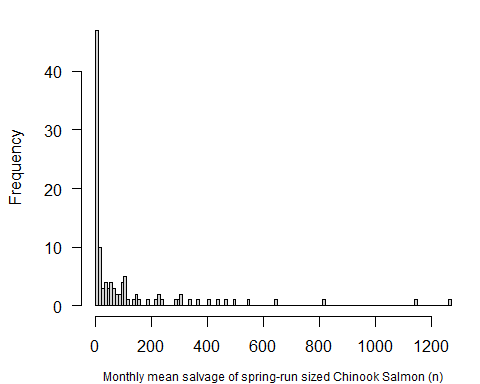
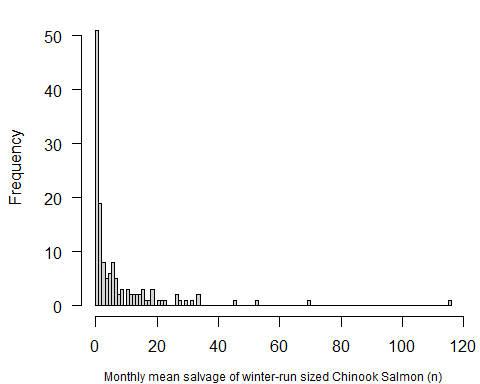
### Data

Winter-run and spring-run length-at-date Chinook Salmon daily expanded salvage records from January 1st 1993 to December 31st 2020 were gathered from the California Department of Fish and Wildlife salvage database posted at the SacPAS website (<http://www.cbr.washington.edu/sacramento/data/query_loss_detail.html>). To incorporate hydrodynamic effects on salvage count into the models, Delta export (QEXPORT), Sacramento River flow (QSAC), and San Joaquin River (QSJR) were extracted from the California Department of Water Resources Dayflow data (<https://data.cnra.ca.gov/dataset/dayflow>). Additionally, combined Old and Middle River flow (OMR) data were pulled from the U.S. Geological Survey National Water Information System website (<https://nwis.waterdata.usgs.gov/nwis>; stations 11313405 and 11312676). Because data gaps exist in the Old and Middle River flow data, ordinary least squares regressions were conducted so that each dataset can be used to predict, and therefore fill, missing data in the other (adjusted : 0.97). Lastly, to account for the variable numbers of juvenile Chinook Salmon entering the Delta by year and month, Sacramento Trawl data were acquired from the Delta Juvenile Fish Monitoring Program through the ‘deltafish’ package available on GitHub (<https://github.com/jeanetteclark/deltaFish>). Sacramento Trawl catch per unit trawl for each day was calculated for both winter-run sized and spring-run sized Chinook Salmon.

Monthly average value summary table

| month | mean\_winter\_lad\_loss | sd\_winter\_lad\_loss | mean\_spring\_lad\_loss | sd\_spring\_lad\_loss |
| --- | --- | --- | --- | --- |
| 1 | 16.4082604 | 27.1491258 | 0.1326613 | 0.3537573 |
| 2 | 24.5842804 | 41.2832423 | 0.9332653 | 1.8764361 |
| 3 | 45.2405876 | 86.5127744 | 78.8869585 | 121.2230005 |
| 4 | 4.9828452 | 9.4312297 | 477.1068452 | 674.3915428 |
| 5 | 0.1888940 | 0.4313571 | 231.7682143 | 366.1717628 |
| 6 | 0.0092976 | 0.0491984 | 25.7662262 | 67.2633749 |
| 7 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| 8 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| 9 | 0.0000000 | 0.0000000 | 0.0310000 | 0.1640366 |
| 10 | 0.0000000 | 0.0000000 | 0.0453456 | 0.2399465 |
| 11 | 0.0000000 | 0.0000000 | 0.0000000 | 0.0000000 |
| 12 | 7.8927304 | 14.4122838 | 0.0000000 | 0.0000000 |

For each variable, data were averaged by month and year with missing data removed. Because monthly salvage values tend to be low or mostly zeroes in for most months out of the year, only December to April period was used for winter-run Chinook Salmon analysis, and only March to June period was used for spring-run Chinook Salmon analysis. Overdispersion was apparent during initial inspection the response variable data (mean variance) supporting the use of negative binomial regression in this analysis.



### VIF analysis

To avoid collinearity, variance inflation factor (VIF) analyses were conducted for all predictor variables mentioned above. A full negative binomial regression model with all predictor variables was constructed for each Chinook Salmon race (winter-run and spring-run), followed by an assessment of VIF values. Per Zuur et al. (2010), variable with the highest VIF value was removed and models were re-run until all VIF values are below 3. For both winter-run and spring-run Chinook Salmon models, OMR had the highest VIF value (>25) and had to be removed from further analysis along with Sacramento River flow. For the final model selection, covariates included were: San Joaquin River flow, Delta export flow value, Sacramento Trawl catch per unit effort (specific to each race), and monthly categorical variable. Each continuous covariate was standardized to z-score prior to the model selection process.

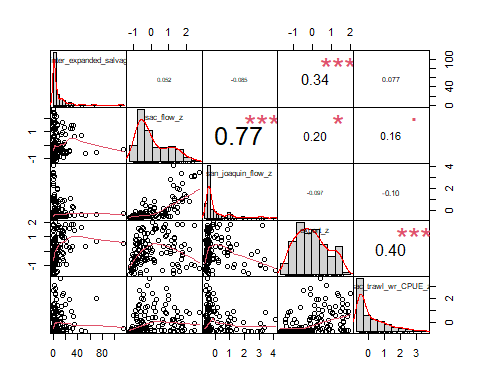
## Winter-run model VIF screening  
wr\_full\_model<-glm.nb(winter\_expanded\_salvage~ sac\_flow\_z + san\_joaquin\_flow\_z + export\_z + omr\_flow\_extrap\_z + sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
car::vif(wr\_full\_model)

## sac\_flow\_z san\_joaquin\_flow\_z export\_z omr\_flow\_extrap\_z   
## 3.119323 25.239984 10.549042 31.946145   
## sac\_trawl\_wr\_CPUE\_z   
## 1.405114

#sac\_flow\_z san\_joaquin\_flow\_z export\_z omr\_flow\_extrap\_z sac\_trawl\_wr\_CPUE\_z   
#3.119324 25.238260 10.548374 31.944030 1.405092   
#Removed OMR flow due to excess collinearity (VIF > 10)  
  
wr\_full\_model<-glm.nb(winter\_expanded\_salvage~ sac\_flow\_z + san\_joaquin\_flow\_z + export\_z + sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
car::vif(wr\_full\_model)

## sac\_flow\_z san\_joaquin\_flow\_z export\_z sac\_trawl\_wr\_CPUE\_z   
## 3.062806 2.921772 1.341683 1.281029

#sac\_flow\_z san\_joaquin\_flow\_z export\_z sac\_trawl\_wr\_CPUE\_z   
# 3.062806 2.921772 1.341683 1.281029   
  
#Plot correlation matrix just for winter-run  
##  
chart.Correlation(data\_WR %>% select(winter\_expanded\_salvage, sac\_flow\_z, san\_joaquin\_flow\_z, export\_z, sac\_trawl\_wr\_CPUE\_z), histogram=TRUE, pch=19)



#0.77 correlation for Sac and SJ flows  
  
#We will go with 3 VIF threshold instead then  
wr\_full\_model<-glm.nb(winter\_expanded\_salvage~ san\_joaquin\_flow\_z + export\_z + sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
car::vif(wr\_full\_model)

## san\_joaquin\_flow\_z export\_z sac\_trawl\_wr\_CPUE\_z   
## 1.015018 1.191279 1.203709

#san\_joaquin\_flow\_z export\_z sac\_trawl\_wr\_CPUE\_z   
#1.015018 1.191279 1.203709   
  
  
## Spring-run model VIF screening  
sr\_full\_model<-glm.nb(spring\_expanded\_salvage~ sac\_flow\_z + san\_joaquin\_flow\_z + export\_z + omr\_flow\_extrap\_z + sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
car::vif(sr\_full\_model)

## sac\_flow\_z san\_joaquin\_flow\_z export\_z omr\_flow\_extrap\_z   
## 3.382410 26.182536 9.399063 29.224090   
## sac\_trawl\_sr\_CPUE\_z   
## 1.098350

#sac\_flow\_z san\_joaquin\_flow\_z export\_z omr\_flow\_extrap\_z sac\_trawl\_sr\_CPUE\_z   
#3.382381 26.180925 9.398553 29.221926 1.098351   
#Removed OMR flow due to excess collinearity (VIF > 10)  
  
sr\_full\_model<-glm.nb(spring\_expanded\_salvage~ sac\_flow\_z + san\_joaquin\_flow\_z + export\_z + sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
car::vif(sr\_full\_model)

## sac\_flow\_z san\_joaquin\_flow\_z export\_z sac\_trawl\_sr\_CPUE\_z   
## 3.288407 3.084517 1.124690 1.097063

#sac\_flow\_z san\_joaquin\_flow\_z export\_z sac\_trawl\_sr\_CPUE\_z   
#3.288407 3.084517 1.124690 1.097063   
  
  
sr\_full\_model<-glm.nb(spring\_expanded\_salvage~ san\_joaquin\_flow\_z + export\_z + sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
car::vif(sr\_full\_model)

## san\_joaquin\_flow\_z export\_z sac\_trawl\_sr\_CPUE\_z   
## 1.110267 1.056263 1.077446

#san\_joaquin\_flow\_z export\_z sac\_trawl\_sr\_CPUE\_z   
#1.110267 1.056263 1.077446   
  
# VIF values for both models look good ~1 after using VIF threshold of 3 instead

### Model selection

For both Chinook Salmon races, the model selection process included all possible additive combination of covariates, as well as addition combination that involves at least one interaction between a continuous variable and the monthly categorical variable. This resulted in 26 possible models (including null) for each Chinook Salmon race, and the top performing model was determined by Akaike Information Criterion for small sample size (AICc). The top model identified through this model selection process was then further evaluated by using leave-one-out cross validation (LOOCV). This was done to provide a measure for 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. Ordinary least squares linear regression is used to compare the relationship between observed and predicted salvage counts and the resulting from this regression is a measure agreement between observed and predicted observations.

## Results

##############################  
#Model selection for winter-run  
  
Cand.set.WR <- list( )  
  
Cand.set.WR[[1]] <- glm.nb(winter\_expanded\_salvage~NULL, data=data\_WR)  
Cand.set.WR[[2]] <- glm.nb(winter\_expanded\_salvage~month\_factor, data=data\_WR)  
Cand.set.WR[[3]] <- glm.nb(winter\_expanded\_salvage~san\_joaquin\_flow\_z, data=data\_WR)  
Cand.set.WR[[4]] <- glm.nb(winter\_expanded\_salvage~export\_z, data=data\_WR)  
Cand.set.WR[[5]] <- glm.nb(winter\_expanded\_salvage~sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[6]] <- glm.nb(winter\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z, data=data\_WR)  
Cand.set.WR[[7]] <- glm.nb(winter\_expanded\_salvage~month\_factor+export\_z, data=data\_WR)  
Cand.set.WR[[8]] <- glm.nb(winter\_expanded\_salvage~month\_factor+sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[9]] <- glm.nb(winter\_expanded\_salvage~san\_joaquin\_flow\_z+export\_z, data=data\_WR)  
Cand.set.WR[[10]] <- glm.nb(winter\_expanded\_salvage~san\_joaquin\_flow\_z+sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[11]] <- glm.nb(winter\_expanded\_salvage~export\_z+sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[12]] <- glm.nb(winter\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z+export\_z, data=data\_WR)  
Cand.set.WR[[13]] <- glm.nb(winter\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z+sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[14]] <- glm.nb(winter\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z+sac\_trawl\_wr\_CPUE\_z+export\_z, data=data\_WR)  
Cand.set.WR[[15]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z, data=data\_WR)  
Cand.set.WR[[16]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z+export\_z, data=data\_WR)  
Cand.set.WR[[17]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z+sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[18]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z+sac\_trawl\_wr\_CPUE\_z+export\_z, data=data\_WR)  
Cand.set.WR[[19]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*export\_z, data=data\_WR)  
Cand.set.WR[[20]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*export\_z+sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[21]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*export\_z+san\_joaquin\_flow\_z, data=data\_WR)  
Cand.set.WR[[22]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*export\_z+sac\_trawl\_wr\_CPUE\_z+san\_joaquin\_flow\_z, data=data\_WR)  
Cand.set.WR[[23]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*sac\_trawl\_wr\_CPUE\_z, data=data\_WR)  
Cand.set.WR[[24]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*sac\_trawl\_wr\_CPUE\_z+san\_joaquin\_flow\_z, data=data\_WR)  
Cand.set.WR[[25]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*sac\_trawl\_wr\_CPUE\_z+export\_z, data=data\_WR)  
Cand.set.WR[[26]] <- glm.nb(winter\_expanded\_salvage~month\_factor\*sac\_trawl\_wr\_CPUE\_z+export\_z+san\_joaquin\_flow\_z, data=data\_WR)  
  
  
##create a vector of names to trace back models in set  
Modnames <- paste("mod","WinterRun", 1:length(Cand.set.WR), sep = "\_")  
  
##generate AICc table  
aictab(cand.set = Cand.set.WR, modnames = Modnames, sort = TRUE)

##   
## Model selection based on AICc:  
##   
## K AICc Delta\_AICc AICcWt Cum.Wt LL  
## mod\_WinterRun\_26 13 710.08 0.00 0.74 0.74 -340.60  
## mod\_WinterRun\_14 9 713.20 3.12 0.15 0.89 -346.91  
## mod\_WinterRun\_25 12 714.07 3.98 0.10 0.99 -343.81  
## mod\_WinterRun\_18 13 721.31 11.23 0.00 0.99 -346.21  
## mod\_WinterRun\_22 13 721.32 11.24 0.00 1.00 -346.22  
## mod\_WinterRun\_12 8 722.25 12.16 0.00 1.00 -352.57  
## mod\_WinterRun\_20 12 725.36 15.28 0.00 1.00 -349.45  
## mod\_WinterRun\_7 7 726.31 16.22 0.00 1.00 -355.73  
## mod\_WinterRun\_16 12 729.62 19.53 0.00 1.00 -351.58  
## mod\_WinterRun\_21 12 730.00 19.91 0.00 1.00 -351.77  
## mod\_WinterRun\_19 11 734.20 24.11 0.00 1.00 -355.07  
## mod\_WinterRun\_24 12 773.72 63.64 0.00 1.00 -373.63  
## mod\_WinterRun\_23 11 776.12 66.04 0.00 1.00 -376.03  
## mod\_WinterRun\_13 8 776.96 66.88 0.00 1.00 -379.93  
## mod\_WinterRun\_8 7 779.39 69.31 0.00 1.00 -382.27  
## mod\_WinterRun\_17 12 782.80 72.72 0.00 1.00 -378.17  
## mod\_WinterRun\_6 7 785.70 75.61 0.00 1.00 -385.42  
## mod\_WinterRun\_4 3 786.78 76.70 0.00 1.00 -390.30  
## mod\_WinterRun\_11 4 787.71 77.63 0.00 1.00 -389.71  
## mod\_WinterRun\_9 4 788.10 78.02 0.00 1.00 -389.90  
## mod\_WinterRun\_2 6 789.28 79.20 0.00 1.00 -388.33  
## mod\_WinterRun\_15 11 791.28 81.20 0.00 1.00 -383.61  
## mod\_WinterRun\_3 3 819.40 109.31 0.00 1.00 -406.61  
## mod\_WinterRun\_10 4 820.92 110.84 0.00 1.00 -406.31  
## mod\_WinterRun\_1 2 821.10 111.01 0.00 1.00 -408.51  
## mod\_WinterRun\_5 3 822.01 111.93 0.00 1.00 -407.92

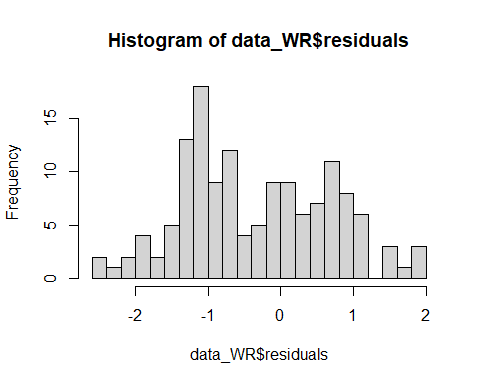
summary(Cand.set.WR[[26]])

##   
## Call:  
## glm.nb(formula = winter\_expanded\_salvage ~ month\_factor \* sac\_trawl\_wr\_CPUE\_z +   
## export\_z + san\_joaquin\_flow\_z, data = data\_WR, init.theta = 1.388142752,   
## link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.5543 -1.0966 -0.5051 0.4369 1.9167   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.269020 0.267923 1.004 0.31533   
## month\_factor2 1.057464 0.331616 3.189 0.00143 \*\*   
## month\_factor3 2.251595 0.342819 6.568 5.1e-11 \*\*\*  
## month\_factor4 16.537650 6.032028 2.742 0.00611 \*\*   
## month\_factor12 -0.486463 0.361555 -1.345 0.17847   
## sac\_trawl\_wr\_CPUE\_z 0.369933 0.190451 1.942 0.05209 .   
## export\_z 1.026055 0.115959 8.848 < 2e-16 \*\*\*  
## san\_joaquin\_flow\_z -0.307430 0.110588 -2.780 0.00544 \*\*   
## month\_factor2:sac\_trawl\_wr\_CPUE\_z 0.177557 0.281007 0.632 0.52748   
## month\_factor3:sac\_trawl\_wr\_CPUE\_z -0.883109 0.336856 -2.622 0.00875 \*\*   
## month\_factor4:sac\_trawl\_wr\_CPUE\_z 19.654679 7.969817 2.466 0.01366 \*   
## month\_factor12:sac\_trawl\_wr\_CPUE\_z 0.007617 0.260044 0.029 0.97663   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(1.3881) family taken to be 1)  
##   
## Null deviance: 382.96 on 139 degrees of freedom  
## Residual deviance: 155.18 on 128 degrees of freedom  
## AIC: 707.19  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 1.388   
## Std. Err.: 0.231   
##   
## 2 x log-likelihood: -681.194

# Model coefficients  
est <- cbind(Estimate = coef(Cand.set.WR[[26]]), confint(Cand.set.WR[[26]]))  
exp(est)

## Estimate 2.5 % 97.5 %  
## (Intercept) 1.308681e+00 0.7678537 2.227810e+00  
## month\_factor2 2.879059e+00 1.5038708 5.548822e+00  
## month\_factor3 9.502880e+00 4.8165797 1.917173e+01  
## month\_factor4 1.521283e+07 65.8580580 2.067804e+13  
## month\_factor12 6.147973e-01 0.3047836 1.236488e+00  
## sac\_trawl\_wr\_CPUE\_z 1.447638e+00 0.9847081 2.168925e+00  
## export\_z 2.790037e+00 2.2098983 3.553799e+00  
## san\_joaquin\_flow\_z 7.353346e-01 0.5782151 9.298824e-01  
## month\_factor2:sac\_trawl\_wr\_CPUE\_z 1.194297e+00 0.6838744 2.107354e+00  
## month\_factor3:sac\_trawl\_wr\_CPUE\_z 4.134955e-01 0.2082561 8.784769e-01  
## month\_factor4:sac\_trawl\_wr\_CPUE\_z 3.434934e+08 28.7268657 3.925356e+16  
## month\_factor12:sac\_trawl\_wr\_CPUE\_z 1.007646e+00 0.6033949 1.694165e+00

# Model residuals  
data\_WR$residuals <- resid(Cand.set.WR[[26]])  
  
hist(data\_WR$residuals, breaks = 20)



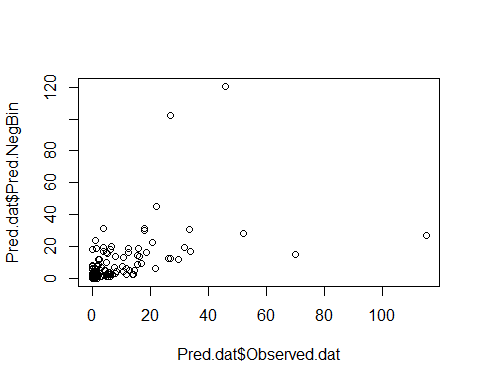
# Looping prediction of best model  
  
Pred.dat <- data.frame(Observed.dat = rep(NA, nrow(data\_WR)), Pred.NegBin = rep(NA, nrow(data\_WR)))  
head(Pred.dat)

## Observed.dat Pred.NegBin  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA

for(i in 1:nrow(data\_WR)){  
 Hat.dat <- data\_WR[-i, ]  
 CV.dat <- data\_WR[i, ]  
 Temp.mod1 <- glm.nb(winter\_expanded\_salvage~month\_factor\*sac\_trawl\_wr\_CPUE\_z+export\_z+san\_joaquin\_flow\_z, data=Hat.dat)  
 Pred.dat[i, "Observed.dat"] <- CV.dat[, "winter\_expanded\_salvage"]  
 Pred.dat[i, "Pred.NegBin"] <- predict(Temp.mod1, newdata = CV.dat, type="response")  
}  
  
head(Pred.dat)

## Observed.dat Pred.NegBin  
## 1 18.891935 16.275644  
## 2 12.607143 19.173000  
## 3 2.580645 5.525451  
## 4 1.133333 24.210047  
## 5 14.000000 2.426332  
## 6 1.935484 1.417661

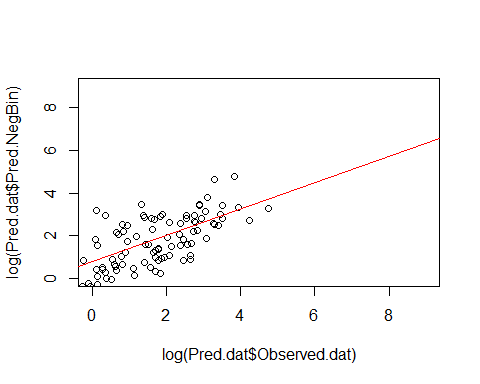
plot(Pred.dat$Observed.dat, Pred.dat$Pred.NegBin)



loocv.lm <- lm(log(Pred.NegBin+1)~log(Observed.dat+1), data=Pred.dat)  
summary(loocv.lm)

##   
## Call:  
## lm(formula = log(Pred.NegBin + 1) ~ log(Observed.dat + 1), data = Pred.dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2256 -0.5565 -0.1666 0.4619 2.0972   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.79384 0.09611 8.259 1.07e-13 \*\*\*  
## log(Observed.dat + 1) 0.61417 0.05330 11.522 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7345 on 138 degrees of freedom  
## Multiple R-squared: 0.4903, Adjusted R-squared: 0.4866   
## F-statistic: 132.8 on 1 and 138 DF, p-value: < 2.2e-16

#Multiple R-squared: 0.4844, Adjusted R-squared: 0.4807   
  
plot(log(Pred.dat$Observed.dat), log(Pred.dat$Pred.NegBin), xlim=c(0, 9), ylim=c(0, 9))  
abline(loocv.lm, col="red")



# Redo model with standard (original) values and ensure that results are essentially the same  
model\_wr\_final<-glm.nb(winter\_expanded\_salvage~month\_factor\*sac\_trawl\_wr\_cpue+export+san\_joaquin\_flow, data=data\_WR)  
  
# Saving final winter-run model  
saveRDS(model\_wr\_final, file = "model\_winter\_run\_final.rda")

##############################  
#Model selection for spring-run  
  
Cand.set.SR <- list( )  
  
Cand.set.SR[[1]] <- glm.nb(spring\_expanded\_salvage~NULL, data=data\_SR)  
Cand.set.SR[[2]] <- glm.nb(spring\_expanded\_salvage~month\_factor, data=data\_SR)  
Cand.set.SR[[3]] <- glm.nb(spring\_expanded\_salvage~san\_joaquin\_flow\_z, data=data\_SR)  
Cand.set.SR[[4]] <- glm.nb(spring\_expanded\_salvage~export\_z, data=data\_SR)  
Cand.set.SR[[5]] <- glm.nb(spring\_expanded\_salvage~sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[6]] <- glm.nb(spring\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z, data=data\_SR)  
Cand.set.SR[[7]] <- glm.nb(spring\_expanded\_salvage~month\_factor+export\_z, data=data\_SR)  
Cand.set.SR[[8]] <- glm.nb(spring\_expanded\_salvage~month\_factor+sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[9]] <- glm.nb(spring\_expanded\_salvage~san\_joaquin\_flow\_z+export\_z, data=data\_SR)  
Cand.set.SR[[10]] <- glm.nb(spring\_expanded\_salvage~san\_joaquin\_flow\_z+sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[11]] <- glm.nb(spring\_expanded\_salvage~export\_z+sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[12]] <- glm.nb(spring\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z+export\_z, data=data\_SR)  
Cand.set.SR[[13]] <- glm.nb(spring\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z+sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[14]] <- glm.nb(spring\_expanded\_salvage~month\_factor+san\_joaquin\_flow\_z+sac\_trawl\_sr\_CPUE\_z+export\_z, data=data\_SR)  
Cand.set.SR[[15]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z, data=data\_SR)  
Cand.set.SR[[16]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z+export\_z, data=data\_SR)  
Cand.set.SR[[17]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z+sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[18]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z+sac\_trawl\_sr\_CPUE\_z+export\_z, data=data\_SR)  
Cand.set.SR[[19]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*export\_z, data=data\_SR)  
Cand.set.SR[[20]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*export\_z+sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[21]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*export\_z+san\_joaquin\_flow\_z, data=data\_SR)  
Cand.set.SR[[22]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*export\_z+sac\_trawl\_sr\_CPUE\_z+san\_joaquin\_flow\_z, data=data\_SR)  
Cand.set.SR[[23]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*sac\_trawl\_sr\_CPUE\_z, data=data\_SR)  
Cand.set.SR[[24]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*sac\_trawl\_sr\_CPUE\_z+san\_joaquin\_flow\_z, data=data\_SR)  
Cand.set.SR[[25]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*sac\_trawl\_sr\_CPUE\_z+export\_z, data=data\_SR)  
Cand.set.SR[[26]] <- glm.nb(spring\_expanded\_salvage~month\_factor\*sac\_trawl\_sr\_CPUE\_z+export\_z+san\_joaquin\_flow\_z, data=data\_SR)  
  
  
##create a vector of names to trace back models in set  
Modnames <- paste("mod","SpringRun", 1:length(Cand.set.SR), sep = "\_")  
  
##generate AICc table  
aictab(cand.set = Cand.set.SR, modnames = Modnames, sort = TRUE)

##   
## Model selection based on AICc:  
##   
## K AICc Delta\_AICc AICcWt Cum.Wt LL  
## mod\_SpringRun\_16 10 1018.17 0.00 0.77 0.77 -498.00  
## mod\_SpringRun\_18 11 1020.61 2.44 0.23 1.00 -497.98  
## mod\_SpringRun\_12 7 1030.50 12.33 0.00 1.00 -507.71  
## mod\_SpringRun\_14 8 1032.74 14.57 0.00 1.00 -507.67  
## mod\_SpringRun\_21 10 1035.23 17.05 0.00 1.00 -506.52  
## mod\_SpringRun\_26 11 1037.34 19.17 0.00 1.00 -506.35  
## mod\_SpringRun\_22 11 1037.67 19.50 0.00 1.00 -506.52  
## mod\_SpringRun\_15 9 1051.32 33.15 0.00 1.00 -515.78  
## mod\_SpringRun\_17 10 1053.50 35.33 0.00 1.00 -515.66  
## mod\_SpringRun\_19 9 1062.50 44.33 0.00 1.00 -521.37  
## mod\_SpringRun\_20 10 1064.62 46.44 0.00 1.00 -521.22  
## mod\_SpringRun\_25 10 1069.03 50.86 0.00 1.00 -523.43  
## mod\_SpringRun\_6 6 1072.29 54.12 0.00 1.00 -529.75  
## mod\_SpringRun\_13 7 1074.55 56.38 0.00 1.00 -529.74  
## mod\_SpringRun\_7 6 1077.25 59.07 0.00 1.00 -532.22  
## mod\_SpringRun\_24 10 1080.21 62.04 0.00 1.00 -529.02  
## mod\_SpringRun\_2 5 1107.05 88.88 0.00 1.00 -548.24  
## mod\_SpringRun\_8 6 1108.22 90.05 0.00 1.00 -547.71  
## mod\_SpringRun\_23 9 1114.33 96.16 0.00 1.00 -547.28  
## mod\_SpringRun\_10 4 1119.94 101.76 0.00 1.00 -555.78  
## mod\_SpringRun\_5 3 1129.05 110.88 0.00 1.00 -561.41  
## mod\_SpringRun\_11 4 1130.27 112.10 0.00 1.00 -560.95  
## mod\_SpringRun\_3 3 1134.55 116.38 0.00 1.00 -564.16  
## mod\_SpringRun\_9 4 1135.69 117.52 0.00 1.00 -563.66  
## mod\_SpringRun\_1 2 1140.89 122.72 0.00 1.00 -568.39  
## mod\_SpringRun\_4 3 1142.00 123.82 0.00 1.00 -567.89

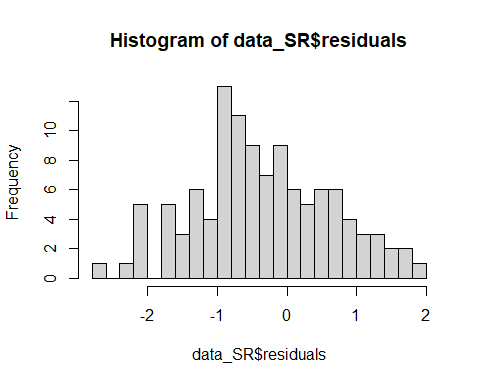
summary(Cand.set.SR[[16]])

##   
## Call:  
## glm.nb(formula = spring\_expanded\_salvage ~ month\_factor \* san\_joaquin\_flow\_z +   
## export\_z, data = data\_SR, init.theta = 0.8408519393, link = log)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.6962 -0.9707 -0.4772 0.3482 1.8640   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 2.7791 0.2377 11.692 < 2e-16 \*\*\*  
## month\_factor4 2.8805 0.3396 8.482 < 2e-16 \*\*\*  
## month\_factor5 1.9990 0.3712 5.385 7.25e-08 \*\*\*  
## month\_factor6 -2.7143 0.3887 -6.982 2.91e-12 \*\*\*  
## san\_joaquin\_flow\_z 1.3104 0.2098 6.246 4.20e-10 \*\*\*  
## export\_z 1.1052 0.1640 6.739 1.60e-11 \*\*\*  
## month\_factor4:san\_joaquin\_flow\_z -1.1572 0.2777 -4.168 3.08e-05 \*\*\*  
## month\_factor5:san\_joaquin\_flow\_z -0.5154 0.3124 -1.650 0.0989 .   
## month\_factor6:san\_joaquin\_flow\_z 0.6644 0.4228 1.571 0.1161   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for Negative Binomial(0.8409) family taken to be 1)  
##   
## Null deviance: 358.30 on 111 degrees of freedom  
## Residual deviance: 121.56 on 103 degrees of freedom  
## AIC: 1016  
##   
## Number of Fisher Scoring iterations: 1  
##   
##   
## Theta: 0.841   
## Std. Err.: 0.111   
##   
## 2 x log-likelihood: -995.995

# Model coefficients  
est <- cbind(Estimate = coef(Cand.set.SR[[16]]), confint(Cand.set.SR[[16]]))  
exp(est)

## Estimate 2.5 % 97.5 %  
## (Intercept) 16.10407840 10.33299447 26.4366187  
## month\_factor4 17.82364913 9.23432605 34.2769725  
## month\_factor5 7.38195404 3.53197178 15.1887904  
## month\_factor6 0.06625441 0.03044602 0.1407236  
## san\_joaquin\_flow\_z 3.70783222 2.30144474 6.1841055  
## export\_z 3.01986121 2.13665194 4.3135319  
## month\_factor4:san\_joaquin\_flow\_z 0.31435537 0.16741814 0.6006191  
## month\_factor5:san\_joaquin\_flow\_z 0.59723433 0.28325712 1.3277487  
## month\_factor6:san\_joaquin\_flow\_z 1.94330704 0.79572295 5.0248423

# Model residuals  
data\_SR$residuals <- resid(Cand.set.SR[[16]])  
  
hist(data\_SR$residuals, breaks = 20)



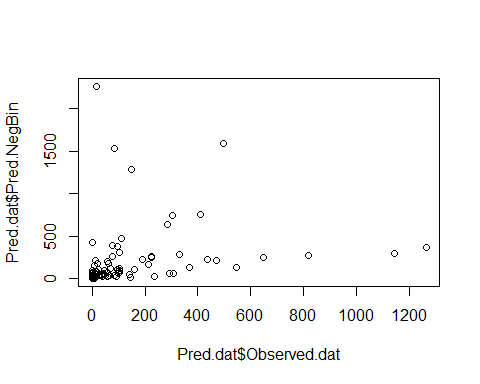
# Looping prediction of best model  
  
Pred.dat <- data.frame(Observed.dat = rep(NA, nrow(data\_SR)), Pred.NegBin = rep(NA, nrow(data\_SR)))  
head(Pred.dat)

## Observed.dat Pred.NegBin  
## 1 NA NA  
## 2 NA NA  
## 3 NA NA  
## 4 NA NA  
## 5 NA NA  
## 6 NA NA

for(i in 1:nrow(data\_SR)){  
 Hat.dat <- data\_SR[-i, ]  
 CV.dat <- data\_SR[i, ]  
 Temp.mod1 <- glm.nb(spring\_expanded\_salvage~month\_factor\*san\_joaquin\_flow\_z+export\_z, data=Hat.dat)  
 Pred.dat[i, "Observed.dat"] <- CV.dat[, "spring\_expanded\_salvage"]  
 Pred.dat[i, "Pred.NegBin"] <- predict(Temp.mod1, newdata = CV.dat, type="response")  
}  
  
head(Pred.dat)

## Observed.dat Pred.NegBin  
## 1 1.6774194 15.707391  
## 2 110.1333333 467.620008  
## 3 139.7741935 47.125751  
## 4 0.9333333 0.262235  
## 5 7.4193548 6.351538  
## 6 100.9266667 82.835331

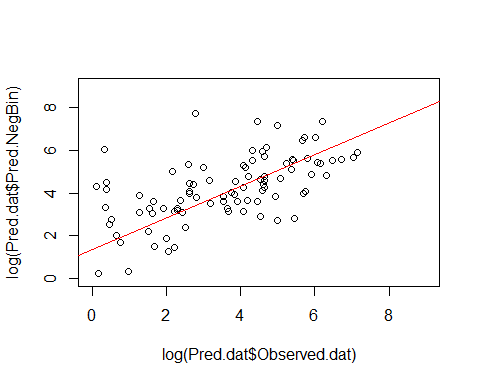
plot(Pred.dat$Observed.dat, Pred.dat$Pred.NegBin)



loocv.lm <- lm(log(Pred.NegBin+1)~log(Observed.dat+1), data=Pred.dat)  
summary(loocv.lm)

##   
## Call:  
## lm(formula = log(Pred.NegBin + 1) ~ log(Observed.dat + 1), data = Pred.dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.5173 -0.9997 -0.1086 0.7873 4.2925   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.32840 0.20971 6.334 5.35e-09 \*\*\*  
## log(Observed.dat + 1) 0.74293 0.05725 12.977 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.25 on 110 degrees of freedom  
## Multiple R-squared: 0.6049, Adjusted R-squared: 0.6013   
## F-statistic: 168.4 on 1 and 110 DF, p-value: < 2.2e-16

#Multiple R-squared: 0.6049, Adjusted R-squared: 0.6013   
  
plot(log(Pred.dat$Observed.dat), log(Pred.dat$Pred.NegBin), xlim=c(0, 9), ylim=c(0, 9))  
abline(loocv.lm, col="red")



# Redo model with standard (original) values and ensure that results are essentially the same  
model\_sr\_final<-glm.nb(spring\_expanded\_salvage~month\_factor\*san\_joaquin\_flow+export, data=data\_SR)  
  
# Saving final spring-run model  
saveRDS(model\_sr\_final, file = "model\_spring\_run\_final.rda")