FMWT\_SMSCG

Rosemary Hartman

7/12/2019

In the Delta Smelt Resiliancy strategy, there is the idea we can change the operation of the Suisun Marsh Salinity control gates to improve habitat in the marsh for Delta Smelt. I was curious if there were any trends between historical gate operations and presence of Delta Smelt in the Marsh. Therefore, I decided to compare catch of Delta Smelt in the marsh during the fall (when the gates are most frequently operated) with gate operations.

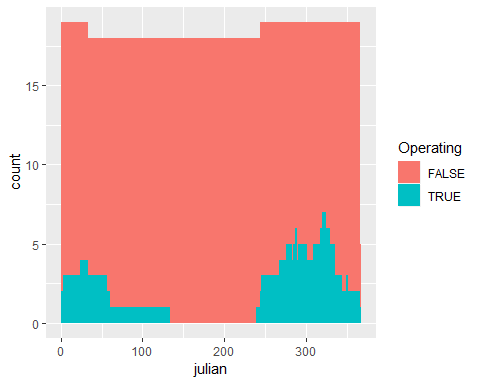
First I did some data manipulation to get the gate operation data lined up with the Delta Smelt Catch from the Fall Midwater Trawl, and I calculated CPUE.

FMWT data is avaialable here: <ftp://ftp.wildlife.ca.gov/TownetFallMidwaterTrawl/FMWT%20Data/> Michael Koohafman gave me the gate operation data.

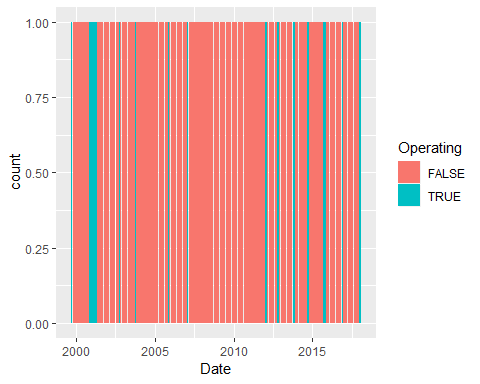
#uplaod the fish catch data  
   
FMWT <- read\_excel("FMWT 1967-2018 Catch Matrix\_updated.xlsx", sheet = "FlatFile",   
 col\_types = c("numeric","date", "numeric", "text", "date",   
 "numeric", "numeric", "numeric", "numeric", "numeric", "numeric",   
 "numeric", "numeric", "text", "text", "text", rep("numeric", times =112)))  
  
#put it in long format instead of wide  
FMWTl = gather(FMWT, key = "Species", value = "catch", `Aequorea spp.`:`Yellowfin Goby`)  
  
#rename all the stupid names  
#names(FMWTl)  
names(FMWTl) = c("Year" ,"Date","Survey","Station", "StartTime","Index",   
 "TopTemp", "TopEC","BottomEC","Turb",  
 "Secchi" , "Depthft","TowVolume","Tide","TowDirection",  
 "Weather","Microcystis","Wave", "Species" , "catch" )  
  
#Ideally, we'd do this analysis on CPUE instead of raw catch  
#They didn't caculate volumes until later, so I'll use the average  
#volume per station for the older tows  
  
#first replace any zero volumes with NAs, because zero volumes don't make sense  
FMWTl$TowVolume[which(FMWTl$TowVolume==0)] = NA  
  
#Calculate the average volume  
meanvol = group\_by(FMWTl, Station) %>% summarize(mvol = mean(TowVolume, na.rm = T))  
FMWTl2 = merge(FMWTl, meanvol)  
FMWTl2$TowVolume[which(is.na(FMWTl2$TowVolume))] = FMWTl2$mvol[which(is.na(FMWTl2$TowVolume))]  
  
#Calculate CPUE   
FMWTl2 = mutate(FMWTl2, CPUE = catch\*TowVolume)  
  
#For starters, I'll just look at Delta Smelt  
FMWT\_DS = filter(FMWTl2, Species == "Delta Smelt")  
  
#Just Delta Smelt from the stations in MOntezuma Slough  
FMWT\_DSm = filter(FMWT\_DS, Station == 605 |Station == 606| Station == 608 )  
  
#Let's filter it so that we just look at 1999-2011. The Delta Smelt catch from   
#2012-2017 was so low it's just going to throw things off  
  
FMWT\_DSm = filter(FMWT\_DSm, Year < 2012)  
  
#load the water quality and gate operations data  
load("~/salinity control gates/SMSCG/operations.RData")  
load("~/salinity control gates/SMSCG/waterquality.RData")  
  
#merge the gate operations with the fish data  
FMWT\_DSm$Date = as.Date(FMWT\_DSm$Date)  
op.daily$Date = as.Date(op.daily$Date)  
FMWT\_DSmg = merge(FMWT\_DSm, op.daily, by = "Date", all.x = T)  
  
#make a new variable for "day of the year"  
FMWT\_DSmg$julian = yday(FMWT\_DSmg$Date)  
  
#Operation is a factor, not a number  
FMWT\_DSmg$Operating = as.factor(FMWT\_DSmg$Operating)

Now for some quick exploritory plots of the data. Look at when the gates are usually operated and what the fish catch was like when they are or are not operated.

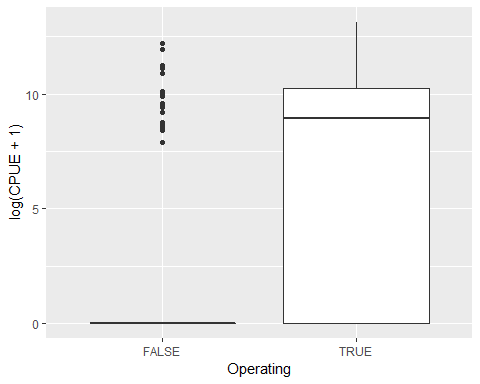
#when are the gates operated?  
op.daily$julian = yday(op.daily$Date)  
ggplot(op.daily, aes(x=julian, fill = Operating)) + geom\_bar(stat = "Count")



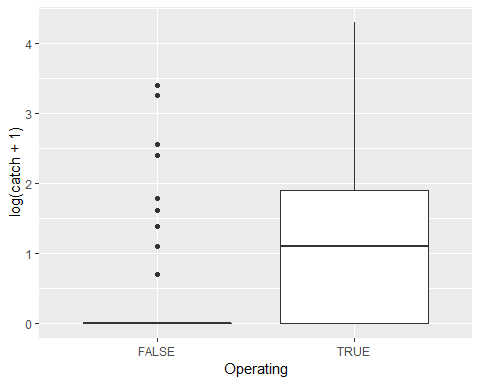
ggplot(op.daily, aes(x=Date, fill = Operating)) + geom\_bar(stat = "Count")



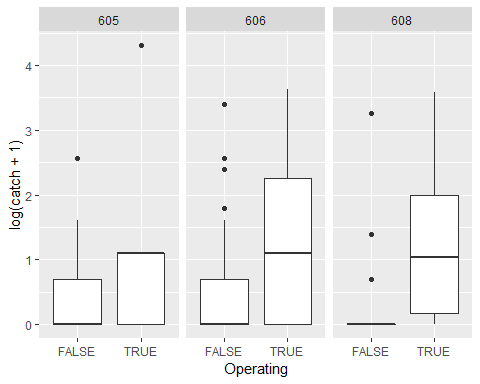
#now with fish, first just the time we have gate data, and just the fall because we have more trawls in the fall  
FMWT\_DSmg2 = filter(FMWT\_DSmg, !is.na(Operating), julian >200)  
#Log CPUE versus gate operations  
ggplot(FMWT\_DSmg2, aes(x = Operating, y = log(CPUE+1))) + geom\_boxplot()



#try log catch instead of CPUE  
ggplot(FMWT\_DSmg2, aes(x = Operating, y = log(catch+1))) + geom\_boxplot()



#seperate by station  
ggplot(FMWT\_DSmg2, aes(x = Operating, y = log(catch+1))) + geom\_boxplot() + facet\_wrap(~Station)



Let’s run some models to see whether there are statisticall more smelt when the gates are operating. Other things are probably involved too, such as year, water year type, salinity, station, day of the year, etc. I’ll run several models and rank them with AICc to see which is best.

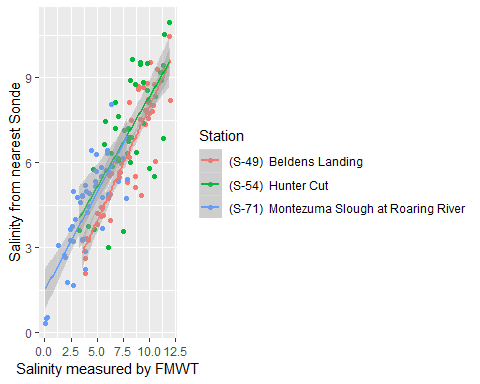
It took me a long time to figure out what type of model to run. Delta smelt catch data is “count data”, so theoretically it should follow a poisson distribution. However, my preliminary analysis showed it is highly overdisperssed and has WAY more zeros than a normal Poisson distribution. Therefore, after much discussion, research, statistics textbooks, and false starts, I settled on a zero-inflated negative binomial model.

I used the salinity from the nearest sonde rather than the salinity measured by FMWT, because some of the CDFW data was suspect (much higher or much lower than would be expected for that time of year). Michael Koohafman also gave me that data.

#first some data manipulation to get the sonde salinity organized  
histday = filter(historical.daily, Analyte == "Salinity")  
FMWT\_DSmg2 = mutate(FMWT\_DSmg2, salinity = TopEC\*0.64/1000, Datetime = Date)  
FMWT\_DSmg2$fishStation = FMWT\_DSmg2$Station  
FMWT\_DSmg2$Station[which(FMWT\_DSmg2$Station == 608)] = "(S-71) Montezuma Slough at Roaring River"  
FMWT\_DSmg2$Station[which(FMWT\_DSmg2$Station == 606)] = "(S-49) Beldens Landing"  
FMWT\_DSmg2$Station[which(FMWT\_DSmg2$Station == 605)] = "(S-54) Hunter Cut"   
FMWT\_DSmg2 = mutate(FMWT\_DSmg2, Datetime = Date)  
  
FMWTwSal2 = unique(merge(FMWT\_DSmg2, histday))  
  
ggplot(FMWTwSal2, aes(x=salinity, y= Mean, color = Station)) +   
 geom\_point() + ylab("Salinity from nearest Sonde") +  
 xlab("Salinity measured by FMWT") + geom\_smooth(method = lm)

## Warning: Removed 4 rows containing non-finite values (stat\_smooth).

## Warning: Removed 4 rows containing missing values (geom\_point).



#so it's close, but not great.  
  
############################################################################################  
  
dszip4a = zeroinfl(catch~ Station + Operating+julian +   
 Mean + Year, dist = "negbin", data = FMWTwSal2)  
dszip4b = zeroinfl(catch~ Station + julian + Mean +  
 Year, dist = "negbin", data = FMWTwSal2)  
dszip4c = zeroinfl(catch~ Station + Operating\*julian +   
 Mean, dist = "negbin", data = FMWTwSal2)  
dszip4e = zeroinfl(catch~ Station + Mean +  
 Year, dist = "negbin", data = FMWTwSal2)  
dszip4f = zeroinfl(catch~ Station + julian, dist = "negbin", data = FMWTwSal2)  
  
dszip4g = zeroinfl(catch~ Station + Mean, dist = "negbin", data = FMWTwSal2)  
  
dszip4h = zeroinfl(catch~ Year + Mean, dist = "negbin", data = FMWTwSal2)  
  
AIC( dszip4a, dszip4b,dszip4c, dszip4e,dszip4f, dszip4g, dszip4h)

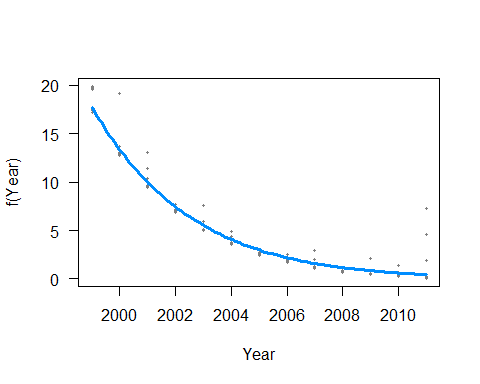
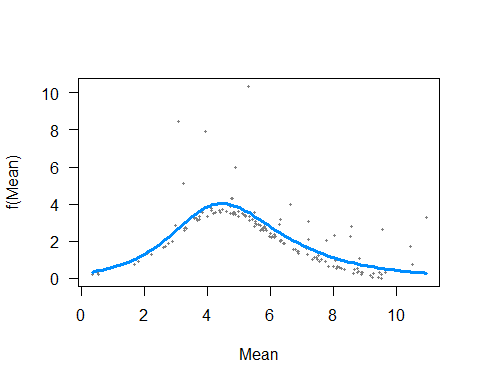
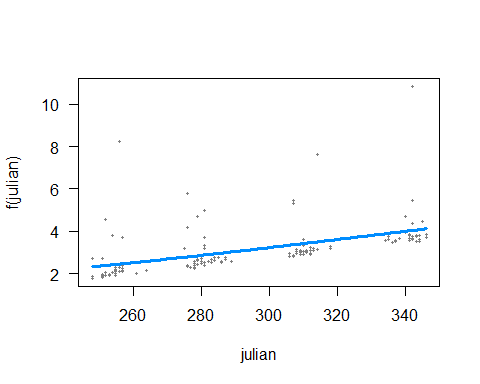
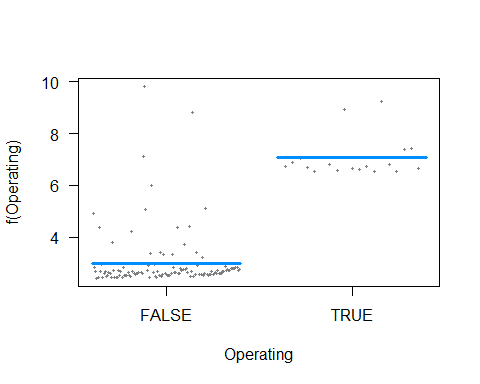
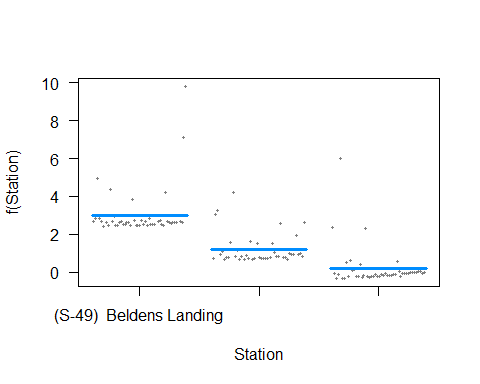
## df AIC  
## dszip4a 15 388.0778  
## dszip4b 13 387.9348  
## dszip4c 15 404.9390  
## dszip4e 11 386.2096  
## dszip4f 9 412.4029  
## dszip4g 9 415.0151  
## dszip4h 7 395.1446

#best model  
summary(dszip4a)

## Warning in sqrt(diag(object$vcov)): NaNs produced

##   
## Call:  
## zeroinfl(formula = catch ~ Station + Operating + julian + Mean +   
## Year, data = FMWTwSal2, dist = "negbin")  
##   
## Pearson residuals:  
## Min 1Q Median 3Q Max   
## -0.5819 -0.4390 -0.3392 -0.1680 6.8174   
##   
## Count model coefficients (negbin with log link):  
## Estimate Std. Error  
## (Intercept) 656.782750 2.020468  
## Station(S-54) Hunter Cut -0.480695 0.578243  
## Station(S-71) Montezuma Slough at Roaring River -2.787321 0.705828  
## OperatingTRUE 0.795825 0.397628  
## julian 0.007410 0.008201  
## Mean -0.492713 0.137823  
## Year -0.326618 NA  
## Log(theta) -0.983450 0.261339  
## z value Pr(>|z|)   
## (Intercept) 325.065 < 2e-16 \*\*\*  
## Station(S-54) Hunter Cut -0.831 0.405803   
## Station(S-71) Montezuma Slough at Roaring River -3.949 7.85e-05 \*\*\*  
## OperatingTRUE 2.001 0.045346 \*   
## julian 0.904 0.366231   
## Mean -3.575 0.000350 \*\*\*  
## Year NA NA   
## Log(theta) -3.763 0.000168 \*\*\*  
##   
## Zero-inflation model coefficients (binomial with logit link):  
## Estimate Std. Error  
## (Intercept) 403.29776 5.20098  
## Station(S-54) Hunter Cut 2.00695 1.40090  
## Station(S-71) Montezuma Slough at Roaring River -1.74581 1.76349  
## OperatingTRUE -1.12288 1.15832  
## julian 0.01517 0.02149  
## Mean -1.34627 0.21450  
## Year -0.20060 NA  
## z value Pr(>|z|)   
## (Intercept) 77.543 < 2e-16 \*\*\*  
## Station(S-54) Hunter Cut 1.433 0.152   
## Station(S-71) Montezuma Slough at Roaring River -0.990 0.322   
## OperatingTRUE -0.969 0.332   
## julian 0.706 0.480   
## Mean -6.276 3.47e-10 \*\*\*  
## Year NA NA   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1   
##   
## Theta = 0.374   
## Number of iterations in BFGS optimization: 95   
## Log-likelihood: -179 on 15 Df

#Here are the partial residual plots for the best model.   
#Partial residual plots show you the effect of each factor when the effects of the  
#other factors have been accounted for.   
visreg(dszip4a)



Those partial-residual plots look good!!! Though the “year” effect is still having problems. I’ve got negative values in my variance-covariance matrix which makes it impossible to calculate standard errors or p-values. Also, I’m not sure what all those computational singularities were about.