zoopGAM.R

jadams

2023-03-22

```
#GAM analysis for FASTR zooplankton

library(tidyverse)
library(lubridate)
library(scales)
library(knitr)
library(mgcv)
library(lme4)
library(car)
library(emmeans)
library(gratia)
library(here)
library(forcats)
# Source functions
source(here("global_ndfa_funcs.R"))
source(here("Water_Quality/global_wq_funcs.R"))

#zooplankton data from Mallory and Nicole (biomass)
zoopNDFAv2<-read.csv("Zoop_code/zoop_NDFA_v2.csv", stringsAsFactors = FALSE)
#data organization and cleanup
#change to factors and organize sample period levels
zoopNDFAv2$SamplePeriod <- factor(zoopNDFAv2$SamplePeriod,levels = c("Before","During","After"
))
zoopNDFAv2$StationCode <- factor(zoopNDFAv2$StationCode)

#create regions for station groups
zoopNDFAv2$Region <- fct_collapse(zoopNDFAv2$StationCode,UpperYolo=c("RD22","I80"),
                                  MiddleSacRiver=c("SHR"),
                                  LowerYolo=c("LIS","STTD"),
                                  ColusaDrainRCS=c("RMB","RCS"),
                                  CacheSloughComplex=c("BL5","LIB"),
                                  LowerSac=c("RYI","RVB"))

#organize regions from north to south for facet plotting
zoopNDFAv2$Region <- factor(zoopNDFAv2$Region,levels = c("ColusaDrainRCS","UpperYolo","LowerYo
lo","CacheSloughComplex","MiddleSacRiver","LowerSac"))


#create regions for station groups
zoopNDFAv2$Regions2 <- fct_collapse(zoopNDFAv2$StationCode,Upstream=c("RCS","RD22","I80","LIS"
,"STTD"),
                                    Downstream=c("BL5","LIB","RYI","RVB"))

#organize regions from north to south for facet plotting
zoopNDFAv2$Regions2 <- factor(zoopNDFAv2$Regions2,levels = c("Upstream","Downstream"))
```

```
#remove macrozooplankton (incomplete dataset and not targeted by our gear)
zoopNDFAv2 <- zoopNDFAv2%>%filter(Classification!="Macrozooplankton")
#glimpse(zoopNDFAv2)

#read in data with additional flow parameters and create new data table joined with zoop data
flow_magnitude<-read.csv("Zoop_code/flow_magnitude.csv", stringsAsFactors = FALSE, na.strings=
"",header = TRUE)
flow_dates<-read.csv("Zoop_code/FlowDatesDesignations.csv", stringsAsFactors = FALSE, na.strin
gs="",header = TRUE)
flow_dates$PreFlowStart <- format(as.Date(flow_dates$PreFlowStart, format = "%m/%d/%Y"), "%Y-%
m-%d")
flow_dates$PreFlowEnd <- format(as.Date(flow_dates$PreFlowEnd, format = "%m/%d/%Y"), "%Y-%m-%d
")
flow_dates$PostFlowStart <- format(as.Date(flow_dates$PostFlowStart, format = "%m/%d/%Y"), "%Y
-%m-%d")
flow_dates$PostFlowEnd <- format(as.Date(flow_dates$PostFlowEnd, format = "%m/%d/%Y"), "%Y-%m-
%d")
flow_dates <- flow_dates %>% filter(Year!="2011")
flow_dates <- flow_dates %>% filter(Year!="2012")
flow_dates <- flow_dates %>% filter(Year!="2013")

zoopNDFAv3 <- zoopNDFAv2
zoopNDFAv3$Year <- as.character(zoopNDFAv3$Year)
flow_magnitude$Year<-as.character(flow_magnitude$Year)

zoopNDFAv3<-left_join(flow_magnitude,zoopNDFAv3)
```

```
## Joining, by = "Year"
```

```
#Remove years 2011 and 2012 with incomplete sampling and remove Sherwood (outside study area)
and Rominger Bridge (too few samples)
zoopNDFAv3 <-  zoopNDFAv3%>% filter(Year!="2011")
zoopNDFAv3 <-  zoopNDFAv3%>% filter(Year!="2012")
zoopNDFAv3 <- zoopNDFAv3 %>% filter(StationCode!="SHR")
zoopNDFAv3 <- zoopNDFAv3 %>% filter(StationCode!="RMB")

zoopNDFA4 <- zoopNDFAv3[,c("Year","Date","SamplePeriod","Region","Regions2","StationCode","CPU
EZoop")] #new table with relevant columns
zoopNDFA4$SamplePeriod <- as.character(zoopNDFA4$SamplePeriod)
zoopNDFA4$Regions2 <- as.character(zoopNDFA4$Regions2)
zoopNDFA4$StationCode <- as.character(zoopNDFA4$StationCode)



zoopNDFA4$scaleCPUE = scale(zoopNDFA4$CPUE) #may need to rescale data for certain analyses

#NOTE: added classification for individual taxa group analysis, but you need to remove this fo
r the original total zoop analysis

zoopNDFA4 <- zoopNDFA4 %>% group_by(Date,StationCode,Year, Regions2,SamplePeriod) %>%
  summarise(cpue=sum(CPUEZoop),
            scaled=sum(scaleCPUE))
```

```
## `summarise()` has grouped output by 'Date', 'StationCode', 'Year', 'Regions2'. You can over
ride using the `.groups` argument.
```

```
#2013 removed because only STTD sampled
zoopNDFA7=zoopNDFA4 %>% filter(Year!=2013) # this dataset excludes 2011-2013 and groups total
CPUE biomass for each sample

#remove data with bad flowmeter data affecting CPUE
zoopNDFA7 <- zoopNDFA7 %>%  filter(Date!="2016-07-07" | StationCode!="RVB")
zoopNDFA7 <- zoopNDFA7 %>% filter(Date!="2016-01-06" | StationCode!="STTD")

zoopNDFA7$SamplePeriod <- factor(zoopNDFA7$SamplePeriod,levels=c("Before","During","After"))


#the following is the model and post hoc from the original report#
#######Two-way interactive model and station code as a random effect--THIS IS THE MODEL we ult
imately chose#####
model4.1 <- lmer(log(cpue) ~ Regions2*Year+Year*SamplePeriod+SamplePeriod*Regions2+(1|StationC
ode),data = zoopNDFA7,REML = TRUE)
summary(model4.1)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log(cpue) ~ Regions2 * Year + Year * SamplePeriod + SamplePeriod *      Regions2 +
 (1 | StationCode)
##    Data: zoopNDFA7
##
## REML criterion at convergence: 1219.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.5194 -0.5012 -0.0072  0.6401  3.0745
##
## Random effects:
##  Groups      Name        Variance Std.Dev.
##  StationCode (Intercept) 0.3042   0.5515
##  Residual                1.5325   1.2379
## Number of obs: 372, groups:  StationCode, 9
##
## Fixed effects:
##                                  Estimate Std. Error t value
## (Intercept)                       8.58720    0.45370  18.927
## Regions2Upstream                 -0.48107    0.53243  -0.904
## Year2015                         -0.56854    0.48288  -1.177
## Year2016                          1.95111    0.50218   3.885
## Year2017                         -0.87450    0.46314  -1.888
## Year2018                         -0.31307    0.42997  -0.728
## Year2019                         -1.30615    0.48972  -2.667
## SamplePeriodDuring                0.13263    0.53483   0.248
## SamplePeriodAfter                -0.16325    0.40940  -0.399
## Regions2Upstream:Year2015        -0.59910    0.46291  -1.294
## Regions2Upstream:Year2016        -0.13756    0.47208  -0.291
```

```
## Regions2Upstream:Year2017            1.46255    0.48820   2.996
## Regions2Upstream:Year2018           -0.29557    0.43022  -0.687
## Regions2Upstream:Year2019            1.17735    0.47956   2.455
## Year2015:SamplePeriodDuring          0.76600    0.63209   1.212
## Year2016:SamplePeriodDuring          0.01746    0.72402   0.024
## Year2017:SamplePeriodDuring          0.06818    0.62687   0.109
## Year2018:SamplePeriodDuring          0.17568    0.60151   0.292
## Year2019:SamplePeriodDuring         -0.56238    0.64912  -0.866
## Year2015:SamplePeriodAfter           0.66558    0.53275   1.249
## Year2016:SamplePeriodAfter          -0.91570    0.53568  -1.709
## Year2017:SamplePeriodAfter          -0.58336    0.62517  -0.933
## Year2018:SamplePeriodAfter          -0.45732    0.48210  -0.949
## Year2019:SamplePeriodAfter          -0.21785    0.55292  -0.394
## Regions2Upstream:SamplePeriodDuring -0.51267    0.33169  -1.546
## Regions2Upstream:SamplePeriodAfter   0.24541    0.31155   0.788
```

```
##
## Correlation matrix not shown by default, as p = 26 > 12.
## Use print(x, correlation=TRUE)  or
##     vcov(x)        if you need it
```

```
model4.1
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log(cpue) ~ Regions2 * Year + Year * SamplePeriod + SamplePeriod *      Regions2 +
 (1 | StationCode)
##    Data: zoopNDFA7
## REML criterion at convergence: 1219.928
## Random effects:
##  Groups      Name        Std.Dev.
##  StationCode (Intercept) 0.5515
##  Residual                1.2379
## Number of obs: 372, groups:  StationCode, 9
## Fixed Effects:
##                         (Intercept)                     Regions2Upstream
          Year2015
##                             8.58720                             -0.48107
          -0.56854
##                           Year2016                             Year2017
          Year2018
##                             1.95111                             -0.87450
          -0.31307
##                           Year2019                   SamplePeriodDuring
 SamplePeriodAfter
##                           -1.30615                              0.13263
          -0.16325
##           Regions2Upstream:Year2015           Regions2Upstream:Year2016            Regions
2Upstream:Year2017
##                           -0.59910                             -0.13756
           1.46255
##           Regions2Upstream:Year2018           Regions2Upstream:Year2019         Year2015:
SamplePeriodDuring
```

```
##                            -0.29557                                1.17735
          0.76600
##         Year2016:SamplePeriodDuring          Year2017:SamplePeriodDuring          Year2018:
SamplePeriodDuring
##                             0.01746                                0.06818
          0.17568
##         Year2019:SamplePeriodDuring           Year2015:SamplePeriodAfter           Year2016
:SamplePeriodAfter
##                            -0.56238                                0.66558
         -0.91570
##           Year2017:SamplePeriodAfter           Year2018:SamplePeriodAfter           Year2019
:SamplePeriodAfter
##                            -0.58336                               -0.45732
         -0.21785
## Regions2Upstream:SamplePeriodDuring   Regions2Upstream:SamplePeriodAfter
##                            -0.51267                                0.24541
```

```
modtab <- Anova(model4.1, type = 3, test.statistic = "F") #this runs- year, regions:year, year
:sampleperiod, regions:sampleperiod are all significant
kable(modtab)
```

	F	Df	Df.res	Pr(>F)
(Intercept)	358.2249166	1	38.81165	0.0000000
Regions2	0.8163806	1	23.30296	0.3754865
Year	10.0582378	5	339.27879	0.0000000
SamplePeriod	0.1910946	2	339.02138	0.8261433
Regions2:Year	6.5141829	5	339.36170	0.0000084
Year:SamplePeriod	1.5048626	10	339.30267	0.1358657
Regions2:SamplePeriod	2.6001219	2	339.13176	0.0757446

```
summary(modtab)
```

```
##        F                   Df             Df.res          Pr(>F)
##  Min.   :  0.1911   Min.   : 1.000   Min.   : 23.3   Min.   :0.0000000
##  1st Qu.:  1.1606   1st Qu.: 1.500   1st Qu.:188.9   1st Qu.:0.0000042
##  Median :  2.6001   Median : 2.000   Median :339.1   Median :0.0757446
##  Mean   : 54.2728   Mean   : 3.714   Mean   :251.2   Mean   :0.2018926
##  3rd Qu.:  8.2862   3rd Qu.: 5.000   3rd Qu.:339.3   3rd Qu.:0.2556761
##  Max.   :358.2249   Max.   :10.000   Max.   :339.4   Max.   :0.8261433
```

```
#####Post-hoc with emmeans Sidak method - THIS IS THE POST HOC USED######
#use emmeans instead to get p value
lmer_emm <- emmeans(model4.1, specs = pairwise ~Regions2:Year,adjust="sidak")#post hoc test on
 region and year (significant from anova) shows no significant differences of individual contr
asts within a year but significant differences between years
```

```
## Warning: You may have generated more contrasts than you really wanted. In the future,
## we suggest you avoid things like 'pairwise ~ fac1*fac2' when you have
## more than one factor. Instead, call emmeans() with just '~ fac1*fac2' and do the
## contrasts you need in a later step. See vignette("QuickStart", "emmeans").
```

```
phoc <- print(test(lmer_emm)$contrasts)
```

```
##  contrast                                  estimate    SE    df t.ratio p.value
##  Downstream Year2014 - Upstream Year2014     0.5702 0.502  18.5   1.136  1.0000
##  Downstream Year2014 - Downstream Year2015   0.0913 0.358 339.0   0.255  1.0000
##  Downstream Year2014 - Upstream Year2015     1.2606 0.501  18.3   2.518  0.7585
##  Downstream Year2014 - Downstream Year2016  -1.6517 0.370 339.1  -4.467  0.0007
##  Downstream Year2014 - Upstream Year2016    -0.9440 0.521  21.4  -1.812  0.9970
##  Downstream Year2014 - Downstream Year2017   1.0462 0.379 339.1   2.762  0.3303
##  Downstream Year2014 - Upstream Year2017     0.1538 0.521  21.3   0.295  1.0000
##  Downstream Year2014 - Downstream Year2018   0.4069 0.327 339.8   1.244  1.0000
##  Downstream Year2014 - Upstream Year2018     1.2727 0.495  17.4   2.573  0.7267
##  Downstream Year2014 - Downstream Year2019   1.5662 0.367 339.0   4.264  0.0017
##  Downstream Year2014 - Upstream Year2019     0.9590 0.509  19.6   1.884  0.9940
##  Upstream Year2014 - Downstream Year2015    -0.4788 0.498  17.9  -0.962  1.0000
##  Upstream Year2014 - Upstream Year2015       0.6904 0.312 339.1   2.215  0.8405
##  Upstream Year2014 - Downstream Year2016    -2.2219 0.512  20.1  -4.338  0.0208
##  Upstream Year2014 - Upstream Year2016      -1.5141 0.340 339.4  -4.459  0.0007
##  Upstream Year2014 - Downstream Year2017     0.4761 0.513  20.2   0.928  1.0000
##  Upstream Year2014 - Upstream Year2017      -0.4163 0.346 339.5  -1.202  1.0000
##  Upstream Year2014 - Downstream Year2018    -0.1632 0.479  15.3  -0.341  1.0000
##  Upstream Year2014 - Upstream Year2018       0.7025 0.302 339.4   2.330  0.7437
##  Upstream Year2014 - Downstream Year2019     0.9961 0.506  19.1   1.968  0.9871
##  Upstream Year2014 - Upstream Year2019       0.3889 0.326 339.0   1.191  1.0000
##  Downstream Year2015 - Upstream Year2015     1.1693 0.485  16.2   2.409  0.8488
##  Downstream Year2015 - Downstream Year2016  -1.7430 0.353 339.1  -4.931  0.0001
##  Downstream Year2015 - Upstream Year2016    -1.0353 0.504  18.8  -2.055  0.9745
##  Downstream Year2015 - Downstream Year2017   0.9549 0.357 339.1   2.674  0.4062
##  Downstream Year2015 - Upstream Year2017     0.0625 0.507  19.2   0.123  1.0000
##  Downstream Year2015 - Downstream Year2018   0.3156 0.304 339.9   1.037  1.0000
##  Downstream Year2015 - Upstream Year2018     1.1813 0.478  15.3   2.469  0.8213
##  Downstream Year2015 - Downstream Year2019   1.4749 0.346 339.0   4.262  0.0017
##  Downstream Year2015 - Upstream Year2019     0.8677 0.494  17.4   1.757  0.9988
##  Upstream Year2015 - Downstream Year2016    -2.9123 0.498  17.9  -5.850  0.0010
##  Upstream Year2015 - Upstream Year2016      -2.2046 0.320 339.2  -6.883  <.0001
##  Upstream Year2015 - Downstream Year2017    -0.2144 0.500  18.2  -0.429  1.0000
##  Upstream Year2015 - Upstream Year2017      -1.1068 0.325 339.4  -3.401  0.0484
##  Upstream Year2015 - Downstream Year2018    -0.8537 0.464  13.6  -1.838  0.9977
##  Upstream Year2015 - Upstream Year2018       0.0121 0.279 339.3   0.043  1.0000
##  Upstream Year2015 - Downstream Year2019     0.3056 0.493  17.2   0.620  1.0000
##  Upstream Year2015 - Upstream Year2019      -0.3016 0.306 339.2  -0.987  1.0000
##  Downstream Year2016 - Upstream Year2016     0.7077 0.503  18.6   1.408  1.0000
##  Downstream Year2016 - Downstream Year2017   2.6979 0.376 339.1   7.182  <.0001
##  Downstream Year2016 - Upstream Year2017     1.8055 0.517  20.8   3.491  0.1355
##  Downstream Year2016 - Downstream Year2018   2.0586 0.322 339.7   6.392  <.0001
##  Downstream Year2016 - Upstream Year2018     2.9244 0.491  17.0   5.951  0.0010
```

```
##  Downstream Year2016 - Downstream Year2019   3.2179 0.363 339.1   8.858  <.0001
##  Downstream Year2016 - Upstream Year2019     2.6107 0.506  19.1   5.159  0.0036
##  Upstream Year2016 - Downstream Year2017     1.9902 0.518  20.9   3.841  0.0610
##  Upstream Year2016 - Upstream Year2017       1.0978 0.354 339.4   3.097  0.1306
##  Upstream Year2016 - Downstream Year2018     1.3509 0.485  16.1   2.786  0.5829
##  Upstream Year2016 - Upstream Year2018       2.2167 0.310 339.4   7.142  <.0001
##  Upstream Year2016 - Downstream Year2019     2.5102 0.512  20.0   4.901  0.0057
##  Upstream Year2016 - Upstream Year2019       1.9030 0.336 339.6   5.669  <.0001
##  Downstream Year2017 - Upstream Year2017    -0.8924 0.507  19.3  -1.759  0.9986
##  Downstream Year2017 - Downstream Year2018  -0.6393 0.329 339.9  -1.945  0.9717
##  Downstream Year2017 - Upstream Year2018     0.2265 0.494  17.4   0.458  1.0000
##  Downstream Year2017 - Downstream Year2019   0.5200 0.368 339.1   1.413  1.0000
##  Downstream Year2017 - Upstream Year2019    -0.0872 0.510  19.6  -0.171  1.0000
##  Upstream Year2017 - Downstream Year2018     0.2531 0.487  16.4   0.519  1.0000
##  Upstream Year2017 - Upstream Year2018       1.1188 0.313 339.0   3.572  0.0264
##  Upstream Year2017 - Downstream Year2019     1.4124 0.515  20.4   2.743  0.5607
##  Upstream Year2017 - Upstream Year2019       0.8052 0.340 339.8   2.370  0.7056
##  Downstream Year2018 - Upstream Year2018     0.8657 0.457  12.8   1.893  0.9963
##  Downstream Year2018 - Downstream Year2019   1.1593 0.317 339.9   3.661  0.0190
##  Downstream Year2018 - Upstream Year2019     0.5521 0.474  14.7   1.166  1.0000
##  Upstream Year2018 - Downstream Year2019     0.2935 0.487  16.4   0.603  1.0000
##  Upstream Year2018 - Upstream Year2019      -0.3136 0.296 339.8  -1.061  1.0000
##  Downstream Year2019 - Upstream Year2019    -0.6072 0.502  18.5  -1.210  1.0000
##
## Results are averaged over the levels of: SamplePeriod
## Degrees-of-freedom method: kenward-roger
## Results are given on the log (not the response) scale.
## P value adjustment: sidak method for 66 tests
```

```
kable(phoc)
```

contrast	estimate	SE	df	t.ratio	p.value
Downstream Year2014 - Upstream Year2014	0.5701628	0.5019185	18.51625	1.1359667	1.0000000
Downstream Year2014 - Downstream Year2015	0.0913436	0.3575355	339.00281	0.2554814	1.0000000
Downstream Year2014 - Upstream Year2015	1.2606056	0.5006444	18.32730	2.5179662	0.7585180
Downstream Year2014 - Downstream Year2016	-1.6516994	0.3697902	339.05784	-4.4665850	0.0007153
Downstream Year2014 - Upstream Year2016	-0.9439766	0.5208989	21.37195	-1.8122069	0.9969507
Downstream Year2014 - Downstream Year2017	1.0462248	0.3787772	339.06683	2.7621110	0.3302858
Downstream Year2014 - Upstream Year2017	0.1538404	0.5208343	21.34490	0.2953730	1.0000000
Downstream Year2014 - Downstream Year2018	0.4069489	0.3271289	339.83716	1.2440016	0.9999999
Downstream Year2014 - Upstream Year2018	1.2726796	0.4945718	17.44867	2.5732957	0.7267026
Downstream Year2014 - Downstream Year2019	1.5662254	0.3673432	339.00224	4.2636569	0.0017216
Downstream Year2014 - Upstream Year2019	0.9590338	0.5089682	19.55591	1.8842705	0.9939549
Upstream Year2014 - Downstream Year2015	-0.4788191	0.4975549	17.89296	-0.9623443	1.0000000

Upstream Year2014 - Upstream Year2015	0.6904428	0.3117172	339.06988	2.2149652	0.8404582
Upstream Year2014 - Downstream Year2016	-2.2218622	0.5122246	20.04520	-4.3376715	0.0207744
Upstream Year2014 - Upstream Year2016	-1.5141394	0.3396003	339.43563	-4.4585932	0.0007407
Upstream Year2014 - Downstream Year2017	0.4760620	0.5130859	20.18008	0.9278407	1.0000000
Upstream Year2014 - Upstream Year2017	-0.4163224	0.3464147	339.51689	-1.2018035	1.0000000
Upstream Year2014 - Downstream Year2018	-0.1632139	0.4786761	15.33345	-0.3409693	1.0000000
Upstream Year2014 - Upstream Year2018	0.7025168	0.3015708	339.43212	2.3295255	0.7437222
Upstream Year2014 - Downstream Year2019	0.9960626	0.5061594	19.13570	1.9678834	0.9870576
Upstream Year2014 - Upstream Year2019	0.3888710	0.3264173	339.03685	1.1913307	1.0000000
Downstream Year2015 - Upstream Year2015	1.1692619	0.4853666	16.22648	2.4090286	0.8487903
Downstream Year2015 - Downstream Year2016	-1.7430431	0.3534745	339.06008	-4.9311703	0.0000846
Downstream Year2015 - Upstream Year2016	-1.0353203	0.5037293	18.74824	-2.0553107	0.9744559
Downstream Year2015 - Downstream Year2017	0.9548811	0.3571456	339.09098	2.6736469	0.4062069
Downstream Year2015 - Upstream Year2017	0.0624967	0.5072576	19.25145	0.1232051	1.0000000
Downstream Year2015 - Downstream Year2018	0.3156053	0.3042419	339.94240	1.0373498	1.0000000
Downstream Year2015 - Upstream Year2018	1.1813359	0.4783887	15.30599	2.4694062	0.8213077
Downstream Year2015 - Downstream Year2019	1.4748817	0.3460661	339.00176	4.2618496	0.0017348
Downstream Year2015 - Upstream Year2019	0.8676901	0.4938313	17.37537	1.7570578	0.9987674
Upstream Year2015 - Downstream Year2016	-2.9123050	0.4978464	17.92496	-5.8498062	0.0010297
Upstream Year2015 - Upstream Year2016	-2.2045822	0.3203042	339.21474	-6.8827775	0.0000000
Upstream Year2015 - Downstream Year2017	-0.2143808	0.5000887	18.25325	-0.4286856	1.0000000
Upstream Year2015 - Upstream Year2017	-1.1067652	0.3254357	339.35885	-3.4008726	0.0484281
Upstream Year2015 - Downstream Year2018	-0.8536567	0.4643304	13.59224	-1.8384682	0.9977012
Upstream Year2015 - Upstream Year2018	0.0120740	0.2785586	339.28180	0.0433445	1.0000000
Upstream Year2015 - Downstream Year2019	0.3056198	0.4929444	17.24560	0.6199883	1.0000000
Upstream Year2015 - Upstream Year2019	-0.3015718	0.3055244	339.18389	-0.9870629	1.0000000
Downstream Year2016 - Upstream Year2016	0.7077228	0.5026290	18.58066	1.4080421	0.9999971
Downstream Year2016 - Downstream Year2017	2.6979242	0.3756263	339.09346	7.1824690	0.0000000
Downstream Year2016 - Upstream Year2017	1.8055398	0.5171480	20.76313	3.4913408	0.1355397
Downstream Year2016 - Downstream Year2018	2.0586483	0.3220754	339.73606	6.3918212	0.0000000
Downstream Year2016 - Upstream Year2018	2.9243790	0.4913867	17.01186	5.9512784	0.0010384

Downstream Year2016 - Downstream Year2019	3.2179248	0.3632708	339.05486	8.8581980	0.0000000
Downstream Year2016 - Upstream Year2019	2.6107332	0.5060368	19.11602	5.1591761	0.0036074
Upstream Year2016 - Downstream Year2017	1.9902014	0.5181991	20.95434	3.8406112	0.0610166
Upstream Year2016 - Upstream Year2017	1.0978170	0.3544825	339.39265	3.0969571	0.1305946
Upstream Year2016 - Downstream Year2018	1.3509255	0.4849742	16.11877	2.7855616	0.5828760
Upstream Year2016 - Upstream Year2018	2.2166562	0.3103480	339.42315	7.1424859	0.0000000
Upstream Year2016 - Downstream Year2019	2.5102020	0.5121903	20.01027	4.9009166	0.0056748
Upstream Year2016 - Upstream Year2019	1.9030104	0.3356931	339.58456	5.6688986	0.0000020
Downstream Year2017 - Upstream Year2017	-0.8923844	0.5074338	19.30500	-1.7586222	0.9985707
Downstream Year2017 - Downstream Year2018	-0.6392759	0.3286152	339.88646	-1.9453630	0.9716503
Downstream Year2017 - Upstream Year2018	0.2264548	0.4941785	17.41209	0.4582449	1.0000000
Downstream Year2017 - Downstream Year2019	0.5200006	0.3680927	339.07835	1.4126893	0.9999888
Downstream Year2017 - Upstream Year2019	-0.0871910	0.5095094	19.63214	-0.1711274	1.0000000
Upstream Year2017 - Downstream Year2018	0.2531085	0.4874433	16.42867	0.5192574	1.0000000
Upstream Year2017 - Upstream Year2018	1.1188392	0.3132274	339.02896	3.5719708	0.0264179
Upstream Year2017 - Downstream Year2019	1.4123850	0.5148746	20.40690	2.7431631	0.5607108
Upstream Year2017 - Upstream Year2019	0.8051934	0.3397747	339.81970	2.3697864	0.7055899
Downstream Year2018 - Upstream Year2018	0.8657307	0.4572303	12.77042	1.8934236	0.9962569
Downstream Year2018 - Downstream Year2019	1.1592765	0.3166198	339.87583	3.6614155	0.0190093
Downstream Year2018 - Upstream Year2019	0.5520849	0.4736620	14.71040	1.1655672	1.0000000
Upstream Year2018 - Downstream Year2019	0.2935458	0.4868504	16.40123	0.6029487	1.0000000
Upstream Year2018 - Upstream Year2019	-0.3136458	0.2955914	339.75741	-1.0610789	1.0000000
Downstream Year2019 - Upstream Year2019	-0.6071916	0.5018114	18.50230	-1.2099995	1.0000000

```
####GAM analysis####

#first some preliminary sample counts to assess study design

#counts by station, year and sample period
zoopNDFA7 %>%
  group_by(Year, SamplePeriod, StationCode) %>% summarise(n = n()) %>%
  arrange(StationCode) %>%
  pivot_wider(names_from = StationCode, values_from = n) %>%
  arrange(Year, SamplePeriod) %>%
  kable()
```

```
## `summarise()` has grouped output by 'Year', 'SamplePeriod'. You can override using the `.gr
oups` argument.
```

Year	SamplePeriod	BL5	I80	LIB	LIS	RCS	RD22	RVB	RYI	STTD
2014	Before	2	2	2	2	2	2	2	2	3
2014	During	1	1	1	1	1	1	1	1	1
2014	After	3	3	3	3	3	3	3	3	4
2015	Before	2	2	2	2	1	2	2	2	4
2015	During	3	2	3	2	2	2	3	3	3
2015	After	2	3	2	3	3	3	2	2	3
2016	Before	2	1	2	1	1	1	1	2	4
2016	During	1	1	1	1	1	1	1	1	1
2016	After	4	4	4	4		4	4	4	4
2017	Before	3	3	3	3	3	3	3	3	3
2017	During	2	2	2	2	2	2	2	2	2
2017	After	1		1				1	1	4
2018	Before	5	3	3	3	3	3	3	5	5
2018	During	4	2	2	2	2	2	2	4	4
2018	After	6	2	2	2	2	2	2	6	7
2019	Before	2	2	2	2	2	2	2	2	2
2019	During	2	2	2	2	2	2	2	2	2
2019	After	2	2	2	2	2	2	2	2	2

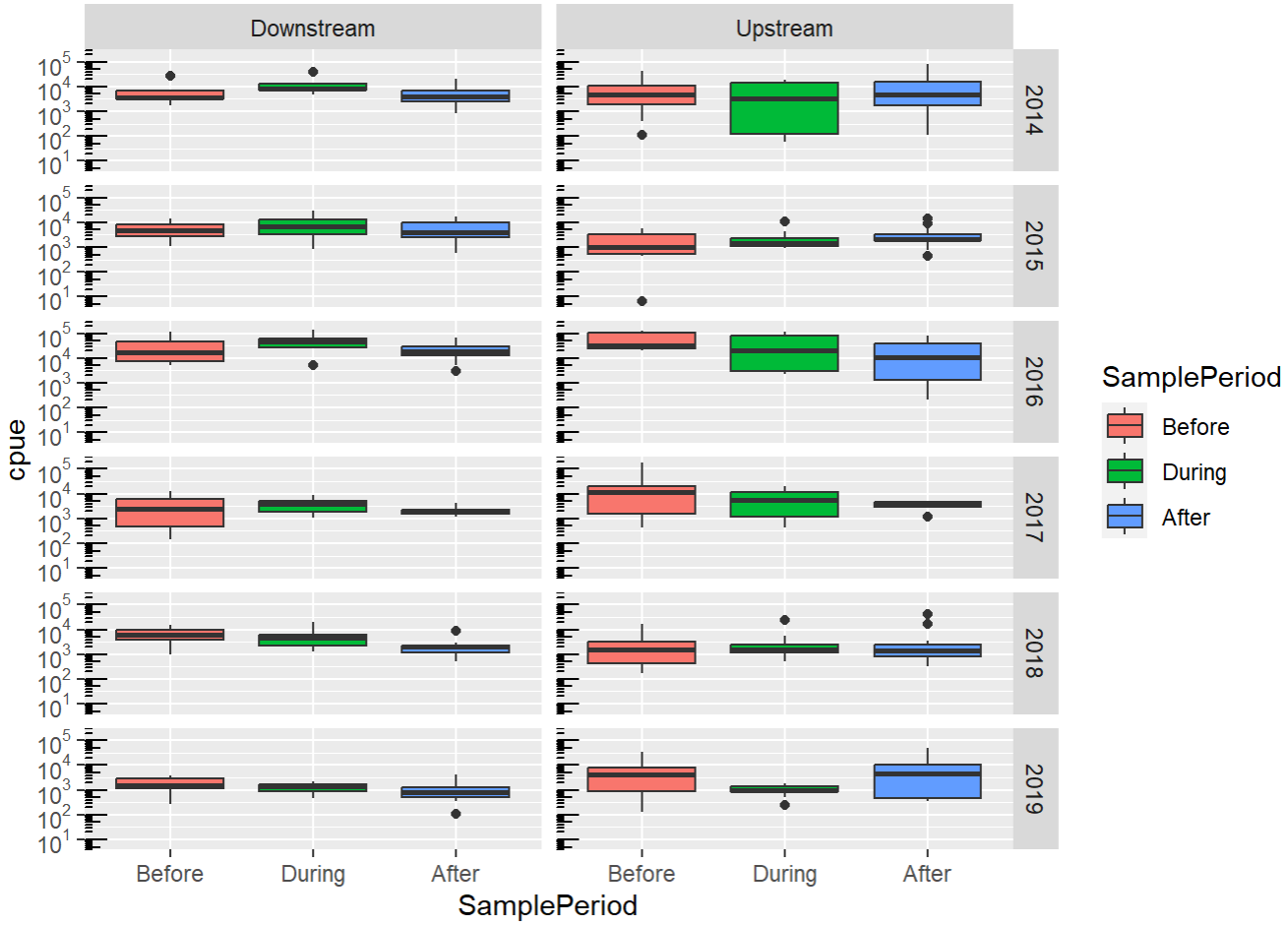
```
#counts by region, year and sample period
zoopNDFA7 %>%
  group_by(Year, SamplePeriod, Regions2) %>% summarise(n = n()) %>%
  pivot_wider(names_from = Regions2, values_from = n) %>%
  kable()
```

```
## `summarise()` has grouped output by 'Year', 'SamplePeriod'. You can override using the `.gr
oups` argument.
```

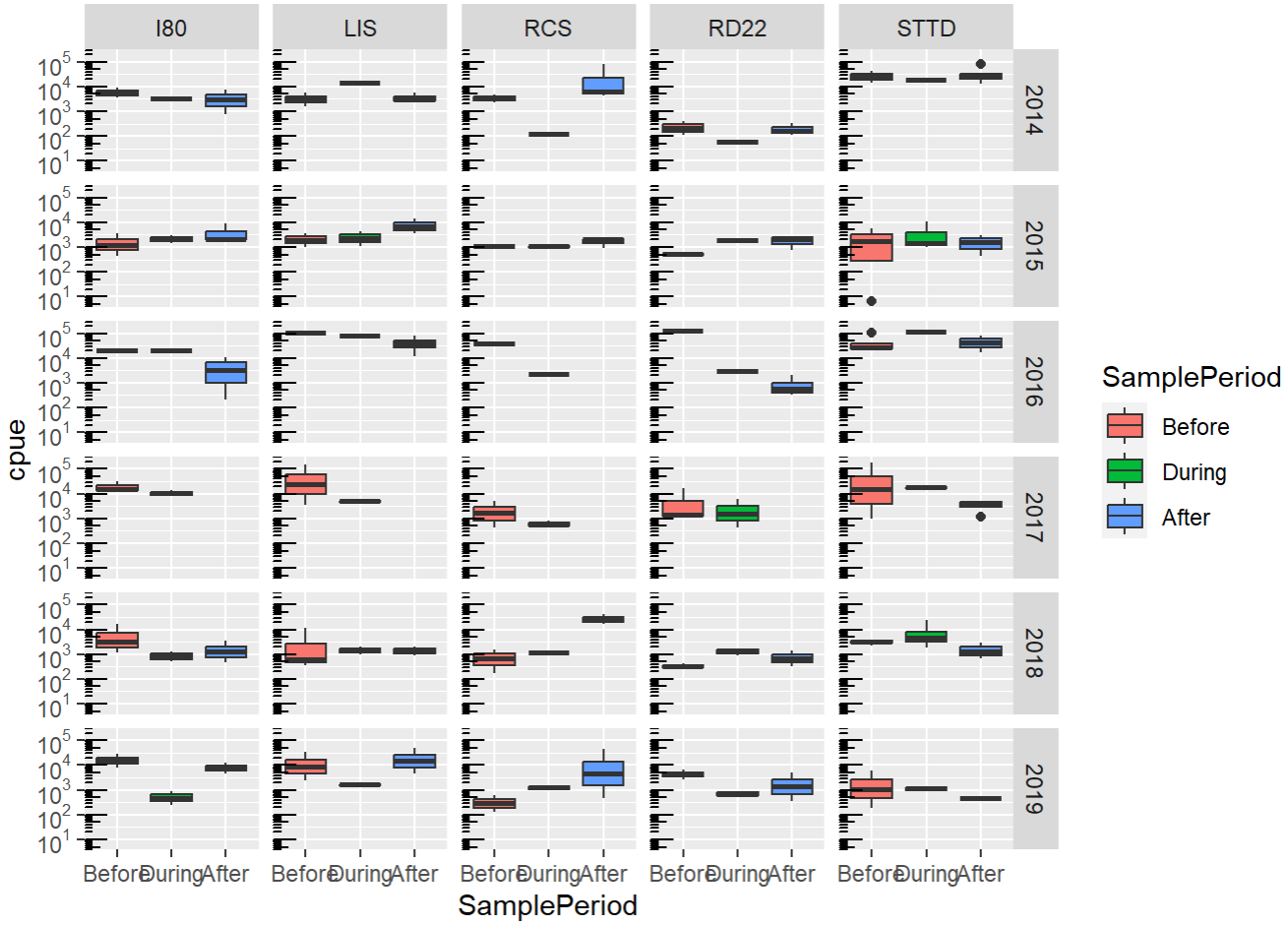
Year	SamplePeriod	Downstream	Upstream
2014	Before	8	11
2014	During	4	5
2014	After	12	16

2015	Before	8	11
2015	During	12	11
2015	After	8	15
2016	Before	7	8
2016	During	4	5
2016	After	16	16
2017	Before	12	15
2017	During	8	10
2017	After	4	4
2018	Before	16	17
2018	During	12	12
2018	After	16	15
2019	Before	8	10
2019	During	8	10
2019	After	8	10

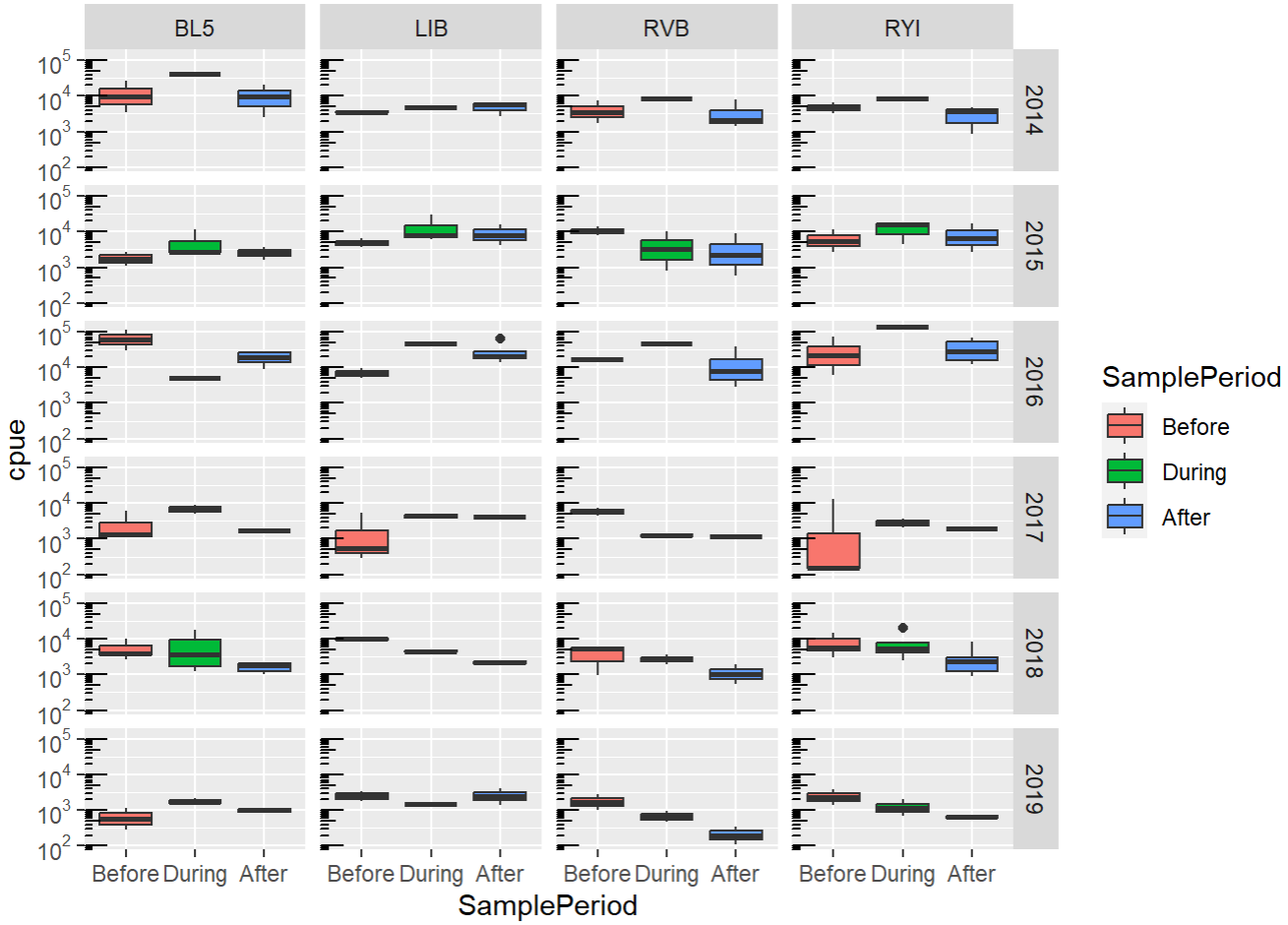
```
#boxplots by year and region

zoopNDFA7 %>%
  ggplot(aes(x = SamplePeriod, y = cpue, fill = SamplePeriod)) +
  geom_boxplot() +
  facet_grid(rows = vars(Year), cols = vars(Regions2)) +
  scale_y_log10(labels = trans_format("log10", math_format(10^.x))) +
  annotation_logticks(sides = "l")
```




```
#boxplots by station and region

zoopNDFA7 %>%
  filter(Regions2 == "Upstream") %>%
  ggplot(aes(x = SamplePeriod, y = cpue, fill = SamplePeriod)) +
  geom_boxplot() +
  facet_grid(rows = vars(Year), cols = vars(StationCode)) +
  scale_y_log10(labels = trans_format("log10", math_format(10^.x))) +
  annotation_logticks(sides = "l")
```



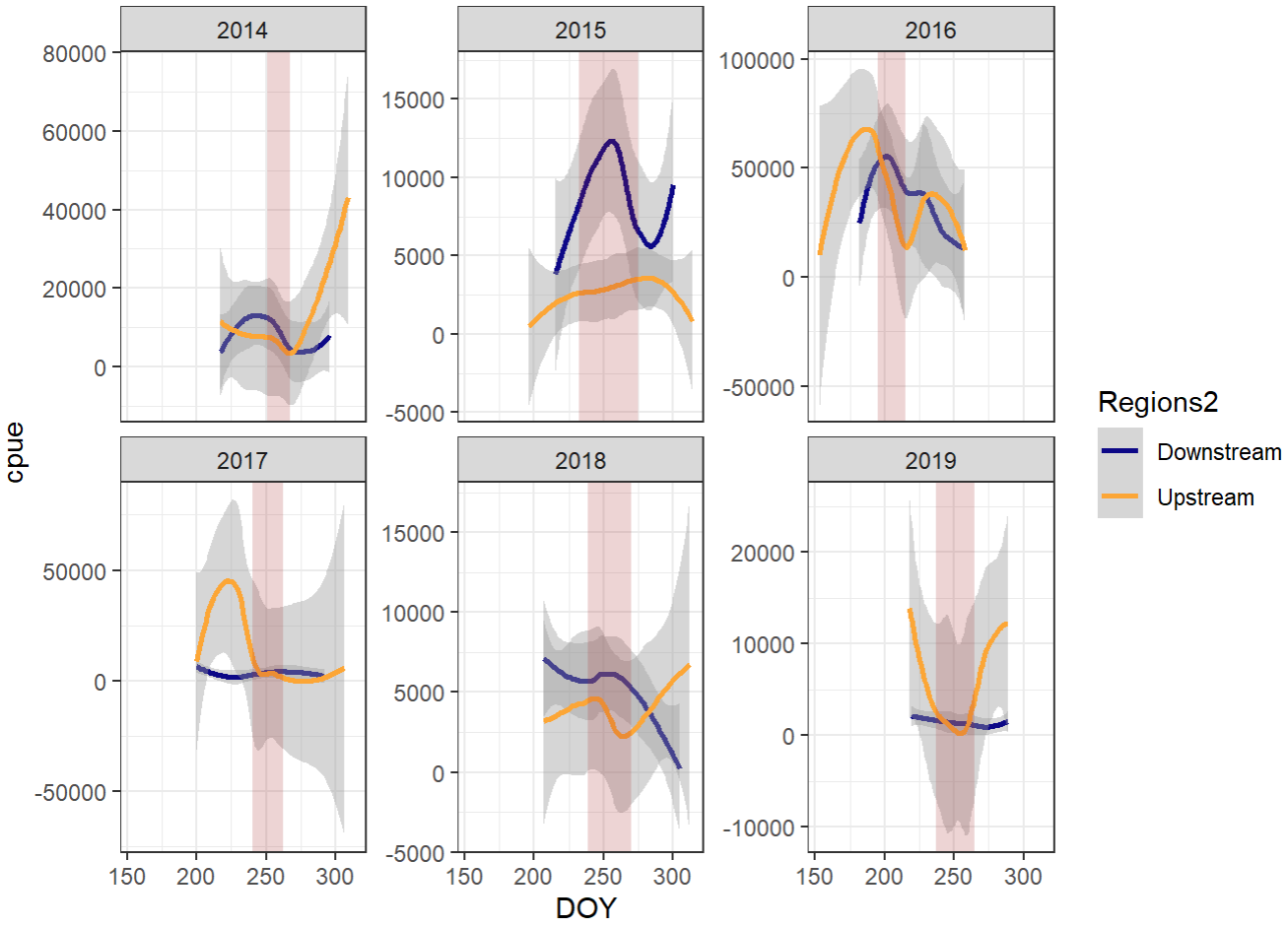

```
zoopNDFA7 %>%
  filter(Regions2 == "Downstream") %>%
  ggplot(aes(x = SamplePeriod, y = cpue, fill = SamplePeriod)) +
  geom_boxplot() +
  facet_grid(rows = vars(Year), cols = vars(StationCode)) +
  scale_y_log10(labels = trans_format("log10", math_format(10^.x))) +
  annotation_logticks(sides = "l")
```




```
#GAM smooth plots
zoopNDFA8 <- zoopNDFA7
zoopNDFA8$DOY <- yday(zoopNDFA8$Date)
zoopNDFA8$logcpue <- zoopNDFA8$cpue
zoopNDFA8$logcpue=log10(zoopNDFA8$cpue)
zoopNDFA8$StationCode <- factor(zoopNDFA8$StationCode)
zoopNDFA8$Regions2 <- factor(zoopNDFA8$Regions2)
zoopNDFA8$Year <- factor(zoopNDFA8$Year)


flow_dates$DOYpreEND <- yday(flow_dates$PreFlowEnd)
flow_dates$DOYpostSTART <- yday(flow_dates$PostFlowStart)

zoopNDFA8 %>%
  ggplot(aes(x = DOY, y = cpue, color = Regions2)) +
  geom_smooth() +
  scale_color_viridis_d(option = "plasma", end = 0.8) +
  facet_wrap(vars(Year), scales = "free_y") +
  geom_rect(
    data = flow_dates,
    aes(
      xmin = DOYpreEND,
      xmax = DOYpostSTART,
      ymin = -Inf,
      ymax = Inf
    ),
```

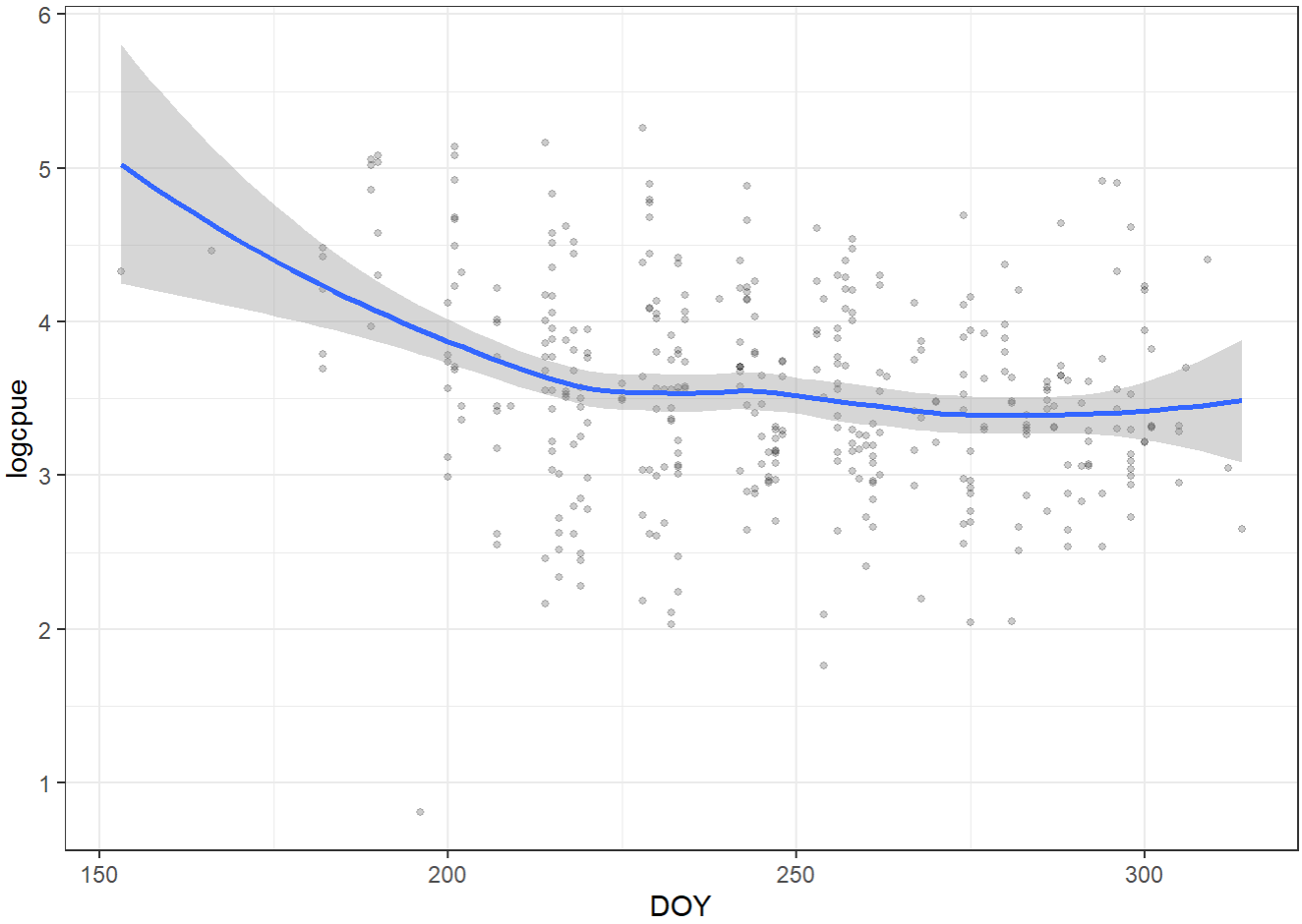
```
    inherit.aes = FALSE,
    alpha = 0.2,
    fill = "brown"
  ) +
   theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



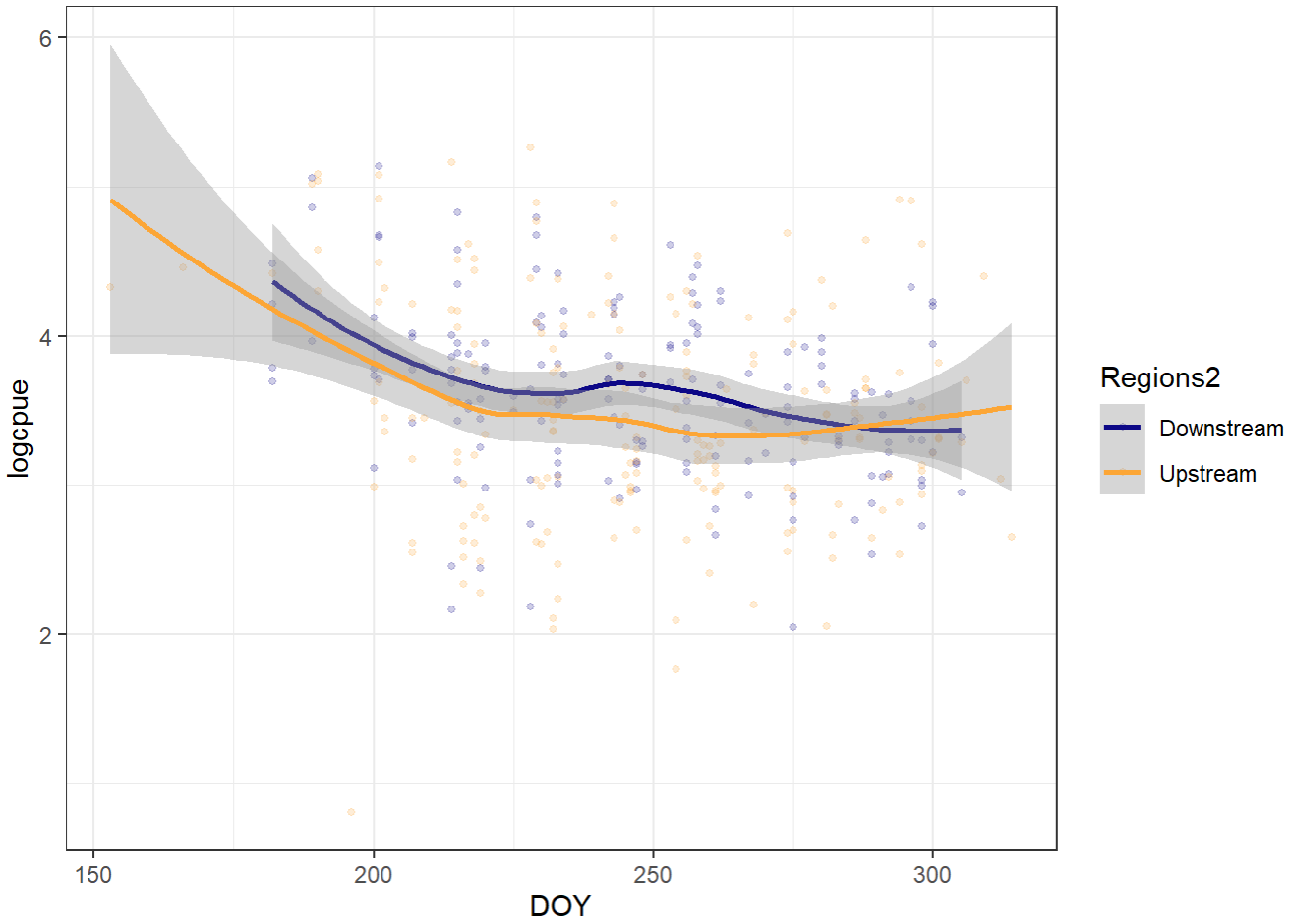

```
zoopNDFA8 %>%
  ggplot(aes(x = DOY, y = logcpue)) +
  geom_point(size = 1, alpha = 0.2) +
  geom_smooth() +
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
zoopNDFA8 %>%
  ggplot(aes(x = DOY, y = logcpue, color = Regions2)) +
  geom_point(size = 1, alpha = 0.2) +
  geom_smooth() +
  scale_color_viridis_d(option = "plasma", end = 0.8) +
  theme_bw()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

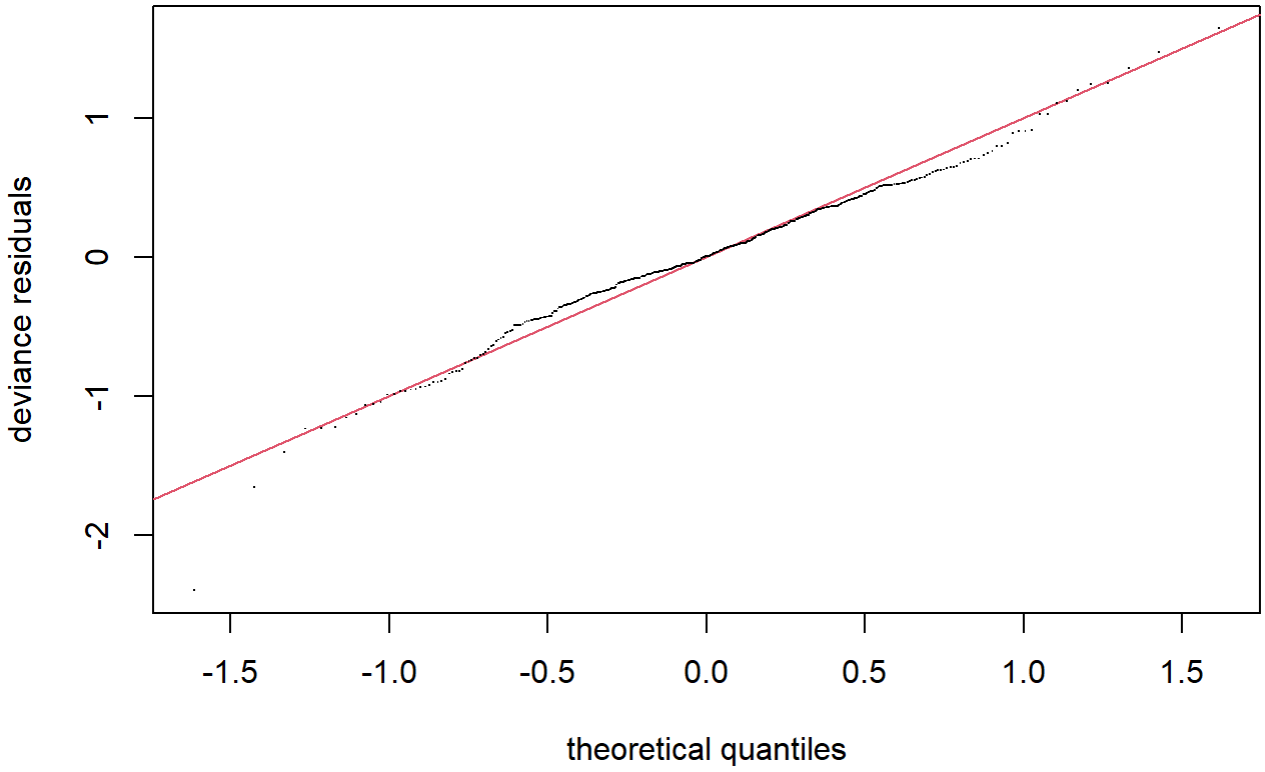


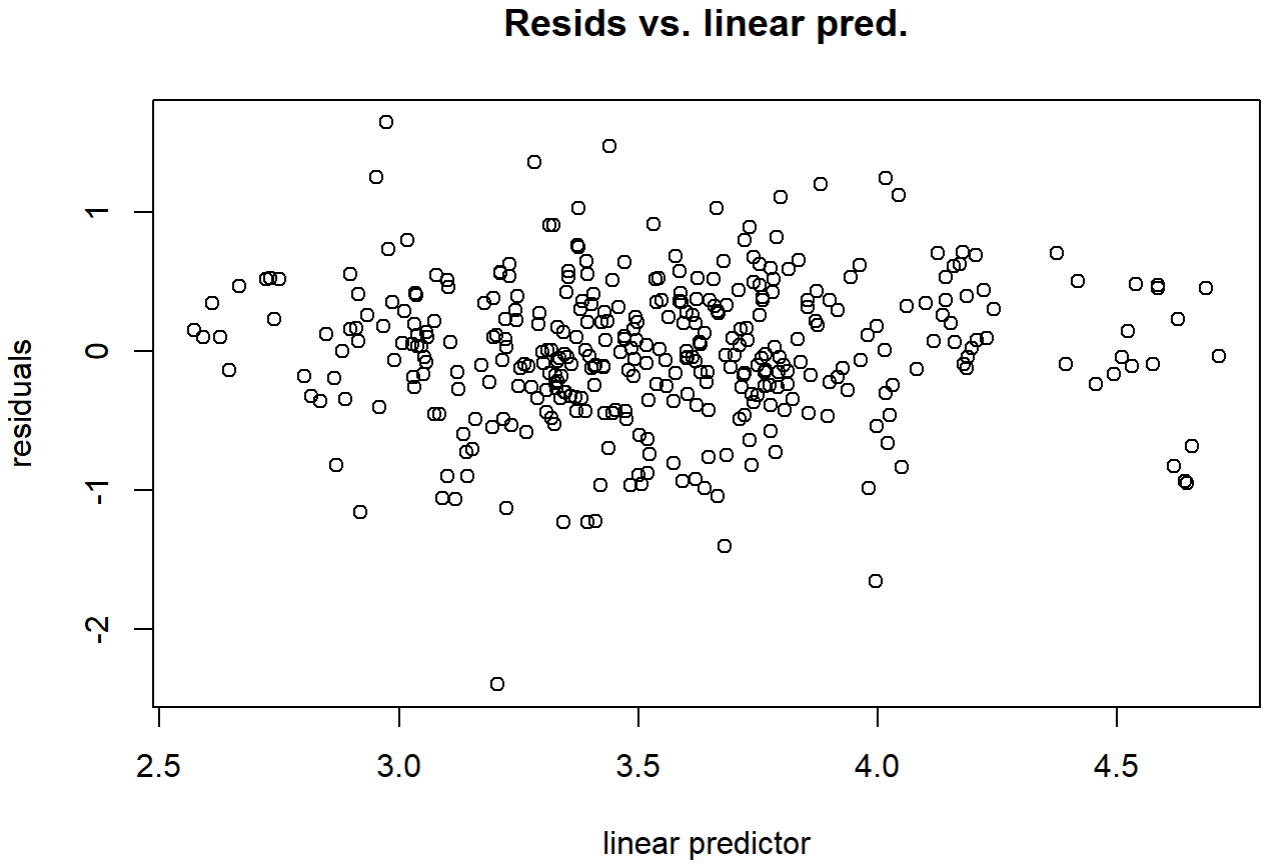

```
#model4.1 <- lmer(log(cpue) ~ Regions2*Year+Year*SamplePeriod+SamplePeriod*Regions2+(1|Station
Code),data = zoopNDFA7,REML = TRUE)

m_cpue_gam <- gam(
  logcpue ~ (Year+SamplePeriod+Regions2)^2 + s(DOY, k=20) + s(StationCode, bs = "re"),
  data = zoopNDFA8,
  method = "REML"
)

summary(m_cpue_gam)
```

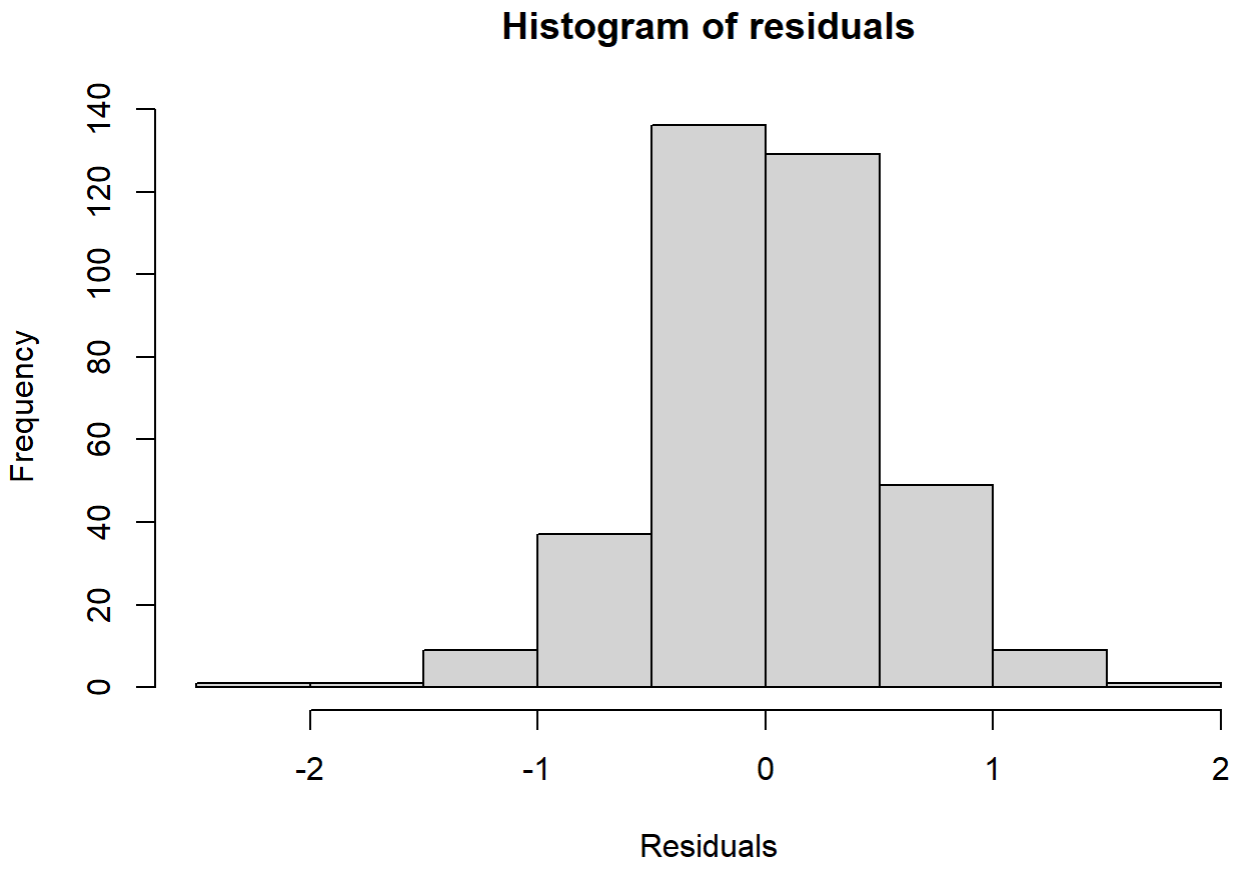
```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## logcpue ~ (Year + SamplePeriod + Regions2)^2 + s(DOY, k = 20) +
##     s(StationCode, bs = "re")
##
## Parametric coefficients:
##                                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       3.75784    0.20626  18.219  < 2e-16 ***
## Year2015                         -0.24456    0.21001  -1.165 0.245025
## Year2016                          0.90105    0.24625   3.659 0.000293 ***
## Year2017                         -0.36649    0.20334  -1.802 0.072372 .
```

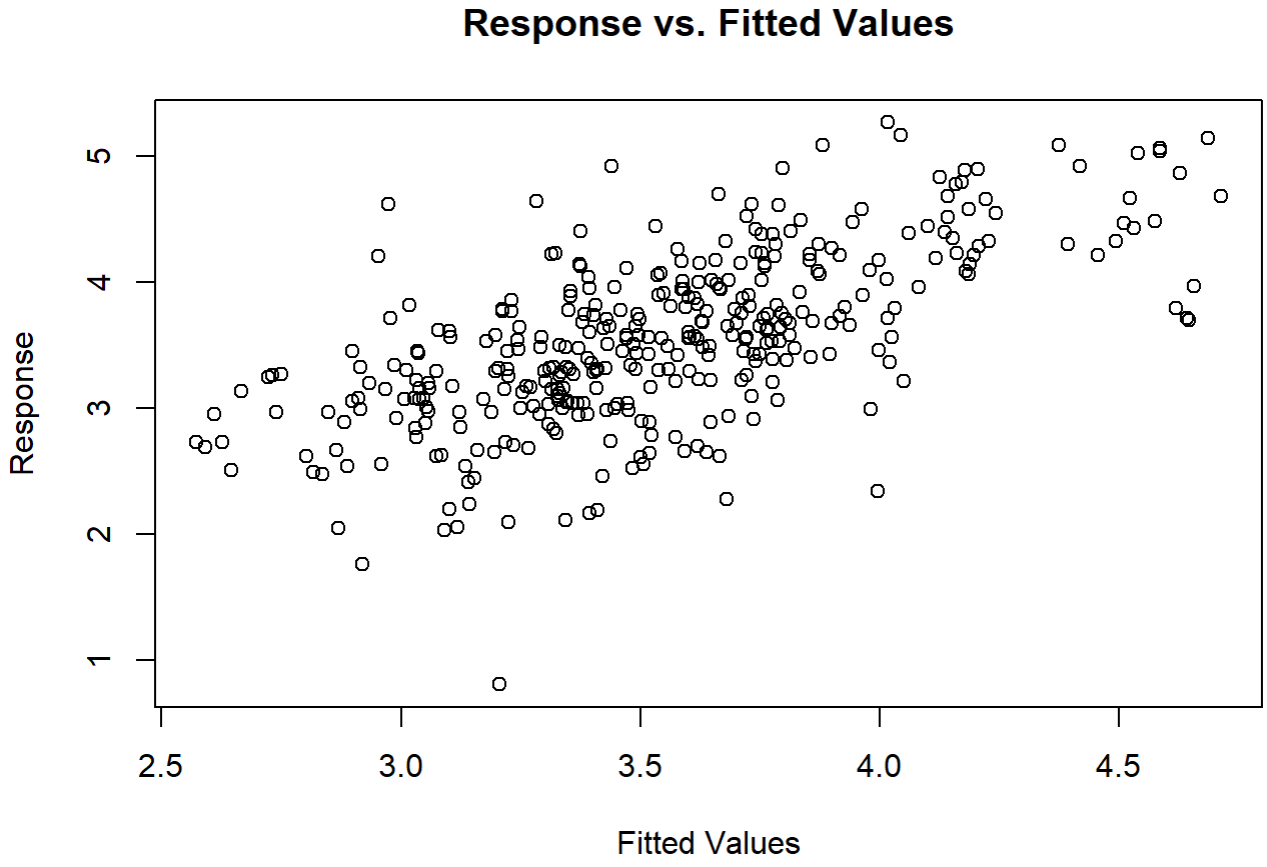
```
## Year2018                             -0.12904    0.18752  -0.688 0.491825
## Year2019                             -0.56858    0.21294  -2.670 0.007948 **
## SamplePeriodDuring                    0.02186    0.24461   0.089 0.928859
## SamplePeriodAfter                    -0.14299    0.23477  -0.609 0.542869
## Regions2Upstream                     -0.21037    0.23141  -0.909 0.363958
## Year2015:SamplePeriodDuring           0.32353    0.27551   1.174 0.241107
## Year2016:SamplePeriodDuring           0.01966    0.31586   0.062 0.950416
## Year2017:SamplePeriodDuring           0.02043    0.27325   0.075 0.940437
## Year2018:SamplePeriodDuring           0.06920    0.26196   0.264 0.791825
## Year2019:SamplePeriodDuring          -0.24458    0.28223  -0.867 0.386767
## Year2015:SamplePeriodAfter            0.27110    0.23475   1.155 0.248965
## Year2016:SamplePeriodAfter           -0.39392    0.23311  -1.690 0.091981 .
## Year2017:SamplePeriodAfter           -0.27617    0.27610  -1.000 0.317901
## Year2018:SamplePeriodAfter           -0.21550    0.21266  -1.013 0.311595
## Year2019:SamplePeriodAfter           -0.09427    0.24040  -0.392 0.695196
## Year2015:Regions2Upstream            -0.25393    0.20171  -1.259 0.208927
## Year2016:Regions2Upstream            -0.05708    0.20533  -0.278 0.781179
## Year2017:Regions2Upstream             0.63769    0.21233   3.003 0.002870 **
## Year2018:Regions2Upstream            -0.12765    0.18706  -0.682 0.495445
## Year2019:Regions2Upstream             0.51408    0.20859   2.465 0.014215 *
## SamplePeriodDuring:Regions2Upstream -0.22268    0.14422  -1.544 0.123501
## SamplePeriodAfter:Regions2Upstream   0.10576    0.13547   0.781 0.435526
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                  edf Ref.df     F p-value
## s(DOY)         1.001  1.002 0.222   0.639
## s(StationCode) 6.210  7.000 7.720  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =   0.36   Deviance explained = 41.6%
## -REML =  322.8  Scale est. = 0.28971   n = 372
```

```
gam.check(m_cpue_gam)
```


deviance residuals
1
0
-1
-2
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5
theoretical quantiles

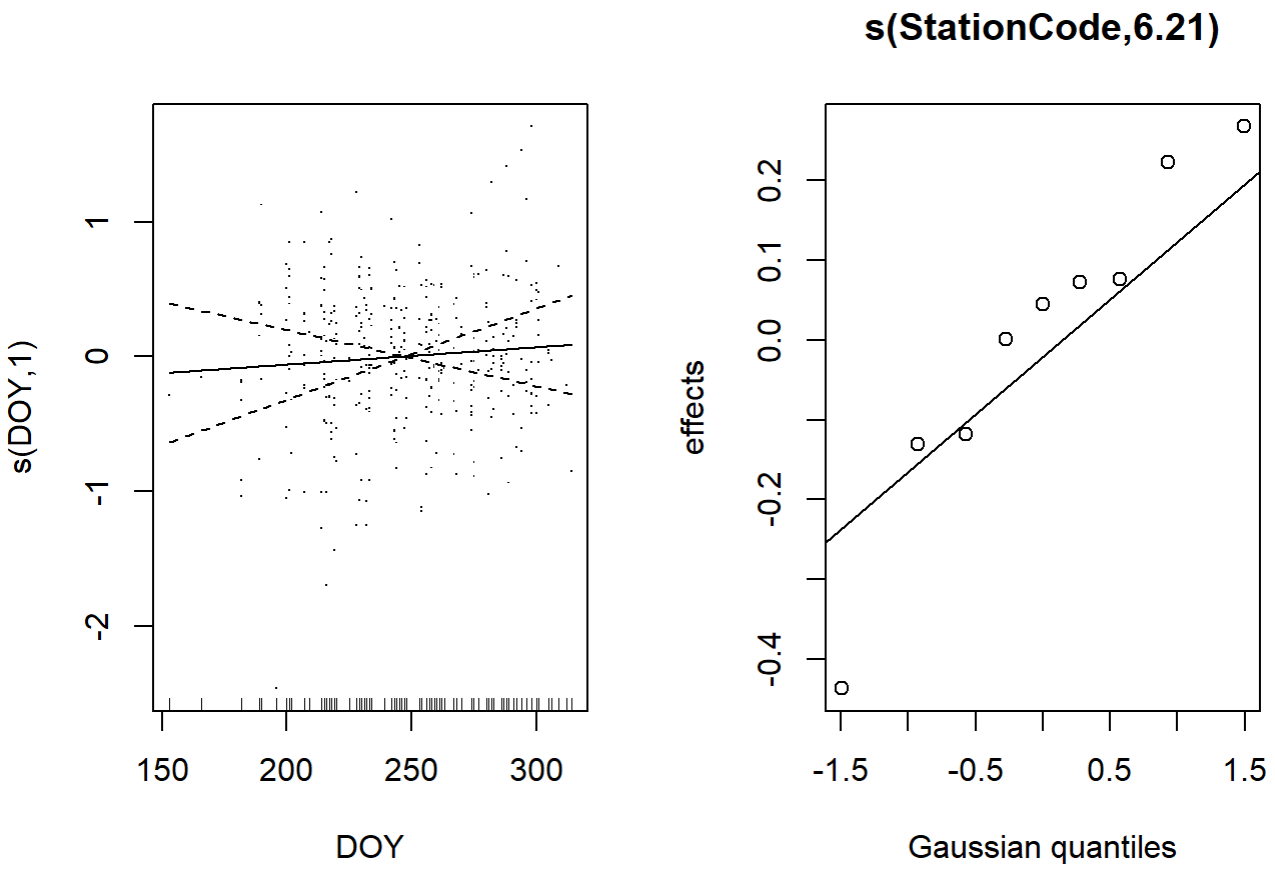

Resids vs. linear pred.
residuals
1
0
-1
-2
2.5
3.0
3.5
4.0
4.5
linear predictor


Histogram of residuals
Frequency
0
20
40
60
80
100
120
140
-2
-1
0
1
2
Residuals


Response vs. Fitted Values
Response
1
2
3
4
5
2.5
3.0
3.5
4.0
4.5
Fitted Values

```
##
## Method: REML   Optimizer: outer newton
## full convergence after 11 iterations.
## Gradient range [-0.0001881878,0.0002267279]
## (score 322.7969 & scale 0.2897056).
## Hessian positive definite, eigenvalue range [0.0001880989,172.5565].
## Model rank =  54 / 54
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                   k'   edf k-index p-value
## s(DOY)         19.00  1.00       1    0.47
## s(StationCode)  9.00  6.21      NA      NA
```

```
plot(m_cpue_gam, pages=1, residuals=TRUE)
```




```
#check out geom_quasirandom from the ggbeeswarm package
```