

GAMs to Analyze Plankton Community Based on Season and Station

Curtis C. Bohlen, Casco Bay Estuary Partnership

7/21/2022

Contents

| | |
|----------------------------------|-----------|
| Introduction | 2 |
| Load Libraries | 2 |
| Set Graphics Theme | 3 |
| Input Data | 3 |
| Folder References | 3 |
| Load Data | 3 |
| Complete Cases | 5 |
| Reduced Data | 6 |
| Models of Fish Abundance | 6 |
| Model 1 | 6 |
| Model on Reduced Data | 11 |
| Total Zooplankton Density | 16 |
| Model 1 | 16 |
| Model on Reduced Data | 20 |
| Shannon Diversity | 23 |
| Model 1 | 23 |
| Model on Reduced Data | 27 |
| Diagnostic Plots | 30 |

| | |
|---|-----------|
| Single Species Models | 31 |
| Automating Analysis of Separate Species | 31 |
| Acartia | 33 |
| Balanus | 38 |
| Eurytemora | 43 |
| Polychaete | 48 |
| Pseudocal | 53 |
| Temora | 58 |

Introduction

This notebook includes analyses using GAMs, based on analysis of model colinearity and concurvity.

In this Notebook, I emphasize Season and Station as factors to explain patterns in the zooplankton data. These terms are fairly highly correlated with Temperature and Salinity, so including terms for Season and Station more or less precludes including Salinity and Temperature in a GAM model.

Where I have compared models, the Season and Station models generally don't perform as well as the models based on quantitative environmental predictors (based on AIC), but I have not compared all pairs of such models.

The base model here includes the following Fixed Effects predictors:

- Season
- Station
- Temperature
- Salinity
- log(Turb)
- log(Chl)
- log1p(Fish)

Year is included in the model as a random effect largely to reduce unexplained variance in the model.

This means these models omit:

- Discharge (highly colinear with other predictors),
- Oxygen Saturation (Incomplete data; highly colinear with Temperature and thus Season and Station).
- Temperature (Highly correlated with Season and somewhat correlated with Station),
- Salinity (Highly correlated with Station and somewhat correlated with Season),
- Sample Event (inclusion as a random factor often led to overspecified models and it seldom proved important)

Load Libraries

```

library(tidyverse)
#> -- Attaching packages ----- tidyverse 1.3.1 --
#> v ggplot2 3.3.6      v purrr 0.3.4
#> v tibble 3.1.7       v dplyr 1.0.9
#> v tidyr 1.2.0        v stringr 1.4.0
#> v readr 2.1.2        v forcats 0.5.1
#> -- Conflicts ----- tidyverse_conflicts() --
#> x dplyr::filter() masks stats::filter()
#> x dplyr::lag()     masks stats::lag()
library(readxl)
library(mgcv) # for GAM models
#> Loading required package: nlme
#>
#> Attaching package: 'nlme'
#> The following object is masked from 'package:dplyr':
#>
#> collapse
#> This is mgcv 1.8-40. For overview type 'help("mgcv-package")'.
library(emmeans) # For extracting useful "marginal" model summaries

```

Set Graphics Theme

This sets `ggplot()` graphics for no background, no grid lines, etc. in a clean format suitable for (some) publications.

```
theme_set(theme_classic())
```

Input Data

Folder References

```
data_folder <- "Original_Data"
```

Load Data

```

filename.in <- "penob.station.data EA 3.12.20.xlsx"
file_path <- file.path(data_folder, filename.in)
station_data <- read_excel(file_path,
  sheet="Final", col_types = c("skip", "date",
    "numeric", "text", "numeric",
    "text", "skip", "skip",
    "skip",
    rep("numeric", 10),
    "text",
    rep("numeric", 47),
    "text",

```

```

      rep("numeric", 12))) %>%
rename_with(~ gsub(" ", "_", .x)) %>%
rename_with(~ gsub("\\.", "_", .x)) %>%
rename_with(~ gsub("\\?", "", .x)) %>%
rename_with(~ gsub("%", "pct", .x)) %>%
rename_with(~ gsub("_Abundance", "", .x)) %>%
filter(! is.na(date))
#> New names:
#> * `` -> `...61`

```

```

names(station_data)[10:12]
#> [1] "discharge_week_cftpersec" "discharg_day"
#> [3] "discharge_week_max"

```

```

names(station_data)[10:12] <- c('disch_wk', 'disch_day', 'disch_max')

```

Station names are arbitrary, and Erin previously expressed interest in renaming them from Stations 2, 4, 5 and 8 to Stations 1,2,3,and 4.

The `factor()` function by default sorts levels before assigning numeric codes, so a convenient way to replace the existing station codes with sequential numbers is to create a factor and extract the numeric indicator values with `as.numeric()`.

```

station_data <- station_data %>%
  mutate(station = factor(as.numeric(factor(station))))
head(station_data)
#> # A tibble: 6 x 76
#>   date                year month month_num season riv_km station station_num
#>   <dtm>              <dbl> <chr>      <dbl> <chr>   <dbl> <fct>      <dbl>
#> 1 2013-05-28 00:00:00 2013 May         5 Spring 22.6  1          1
#> 2 2013-05-28 00:00:00 2013 May         5 Spring 13.9  2          2
#> 3 2013-05-28 00:00:00 2013 May         5 Spring  8.12  3          3
#> 4 2013-05-28 00:00:00 2013 May         5 Spring  2.78  4          4
#> 5 2013-07-25 00:00:00 2013 July         7 Summer 22.6  1          1
#> 6 2013-07-25 00:00:00 2013 July         7 Summer 13.9  2          2
#> # ... with 68 more variables: depth <dbl>, disch_wk <dbl>, disch_day <dbl>,
#> #   disch_max <dbl>, tide_height <dbl>, Full_Moon <dbl>, Abs_Moon <dbl>,
#> #   Spring_or_Neap <chr>, ave_temp_c <dbl>, ave_sal_psu <dbl>,
#> #   ave_turb_ntu <dbl>, ave_do_mgperl <dbl>, ave_DO_Saturation <dbl>,
#> #   ave_chl_microgperl <dbl>, sur_temp <dbl>, sur_sal <dbl>, sur_turb <dbl>,
#> #   sur_do <dbl>, sur_chl <dbl>, bot_temp <dbl>, bot_sal <dbl>, bot_turb <dbl>,
#> #   bot_do <dbl>, bot_chl <dbl>, max_temp <dbl>, max_sal <dbl>, ...

```

Subsetting to Desired Data Columns

I base selection of predictor variables here on the ones used in the manuscript.

```

base_data <- station_data %>%
  rename(Date = date,
         Station = station,
         Year = year) %>%
  select(-c(month, month_num)) %>%

```

```

mutate(Month = factor(as.numeric(format(Date, format = '%m')),
                      levels = 1:12,
                      labels = month.abb),
      DOY = as.numeric(format(Date, format = '%j')),
      season = factor(season, levels = c('Spring', 'Summer', 'Fall')),
      is_sp_up = season == 'Spring' & Station == 1,
      Yearf = factor(Year)) %>%
rename(Season = season,
      Density = combined_density,
      Temp = ave_temp_c,
      Sal = ave_sal_psu,
      Turb = sur_turb,
      AvgTurb = ave_turb_ntu,
      DOsat = ave_DO_Saturation,
      Chl = ave_chl_microgperl,
      Fish = `___61`,
      RH = Herring
) %>%
select(Date, Station, Year, Yearf, Month, Season, is_sp_up, DOY, riv_km,
      disch_wk, disch_day, disch_max,
      Temp, Sal, Turb, AvgTurb, DOsat, Chl,
      Fish, RH,
      Density, H, SEI,
      Acartia, Balanus, Eurytemora, Polychaete, Pseudocal, Temora) %>%
arrange(Date, Station)
head(base_data)
#> # A tibble: 6 x 29
#>   Date           Station Year Yearf Month Season is_sp_up DOY riv_km
#>   <dtm>         <fct>   <dbl> <fct> <fct> <fct> <lgl>   <dbl> <dbl>
#> 1 2013-05-28 00:00:00 1      2013 2013 May   Spring TRUE    148 22.6
#> 2 2013-05-28 00:00:00 2      2013 2013 May   Spring FALSE   148 13.9
#> 3 2013-05-28 00:00:00 3      2013 2013 May   Spring FALSE   148 8.12
#> 4 2013-05-28 00:00:00 4      2013 2013 May   Spring FALSE   148 2.78
#> 5 2013-07-25 00:00:00 1      2013 2013 Jul    Summer FALSE   206 22.6
#> 6 2013-07-25 00:00:00 2      2013 2013 Jul    Summer FALSE   206 13.9
#> # ... with 20 more variables: disch_wk <dbl>, disch_day <dbl>, disch_max <dbl>,
#> #   Temp <dbl>, Sal <dbl>, Turb <dbl>, AvgTurb <dbl>, DOsat <dbl>, Chl <dbl>,
#> #   Fish <dbl>, RH <dbl>, Density <dbl>, H <dbl>, SEI <dbl>, Acartia <dbl>,
#> #   Balanus <dbl>, Eurytemora <dbl>, Polychaete <dbl>, Pseudocal <dbl>,
#> #   Temora <dbl>

```

```
rm(station_data)
```

Complete Cases

This drops only two samples, one for missing Zooplankton data, one for missing fish data. We need this reduced data set to run The `step()` function. It makes little sense to try stepwise model selection if each time you add or remove a variable, the sample you are studying changes. Since fish is never an important predictor, we will want need to refit models after stepwise elimination to use the most complete possible data set.

```
complete_data <- base_data %>%
  select(Season, Station, Yearf,
         is_sp_up, Temp, Sal, Turb, Chl, Fish, RH,
         Density, H,
         Acartia, Balanus, Eurytemora, Polychaete, Pseudocal, Temora) %>%
  filter(complete.cases(.))
```

Reduced Data

The low salinity spring samples are doing something rather different, and they complicate model fitting. Models are far better behaved if we exclude a few extreme samples. These are low salinity low zooplankton samples. We have two complementary ways to specify which samples to omit, without just omitting “outliers”. The first is to restrict modeling to “marine” samples over a certain salinity, and the other is to omit spring upstream samples, which include most of the problematic samples.

```
drop_low <- complete_data %>%
  filter(Sal > 10) # Pulls three samples, including one fall upstream sample
                  # a fourth low salinity sample lacks zooplankton data
#drop_sp_up <- complete_data %>%
# filter(! is_sp_up) # drops four samples
```

Models of Fish Abundance

Model 1

```
fish_gam_ss <- gam(log1p(Fish) ~
  Season +
  Station +
  s(log(Turb), bs="ts", k = 5) +
  s(log(Chl), bs="ts", k = 5) +
  s(log1p(Density), bs="ts", k = 5) +
  s(Yearf, bs = 're'),
  data = complete_data, family = 'gaussian')
summary(fish_gam_ss)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Fish) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Density), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   5.8008      0.7667    7.565 1.66e-09 ***
#> SeasonSummer -0.7653      0.7767   -0.985  0.3298
#> SeasonFall   -1.3418      0.7212   -1.861  0.0695 .
#> Station2     -0.8101      0.6557   -1.235  0.2232
```

```

#> Station3      -1.2246      0.6574 -1.863  0.0692 .
#> Station4      -1.1382      0.7339 -1.551  0.1281
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df      F p-value
#> s(log(Turb))    1.586e+00    4 0.339 0.42372
#> s(log(Chl))     2.758e+00    4 1.149 0.17350
#> s(log1p(Density)) 3.426e+00    4 3.259 0.00906 **
#> s(Yearf)        2.359e-10    5 0.000 0.64612
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.209   Deviance explained = 38.7%
#> GCV = 3.3481   Scale est. = 2.5532      n = 58

```

Note that the model explains on the order of 39% of the variance. That is actually slightly better than the model based on quantitative environmental predictors (in spite of the fact that neither Season nor Station is significant by ANOVA).

```

anova(fish_gam_ss)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Fish) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Density), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>              df      F p-value
#> Season      2 1.912  0.160
#> Station     3 1.268  0.297
#>
#> Approximate significance of smooth terms:
#>              edf   Ref.df      F p-value
#> s(log(Turb))    1.586e+00 4.000e+00 0.339 0.42372
#> s(log(Chl))     2.758e+00 4.000e+00 1.149 0.17350
#> s(log1p(Density)) 3.426e+00 4.000e+00 3.259 0.00906
#> s(Yearf)        2.359e-10 5.000e+00 0.000 0.64612

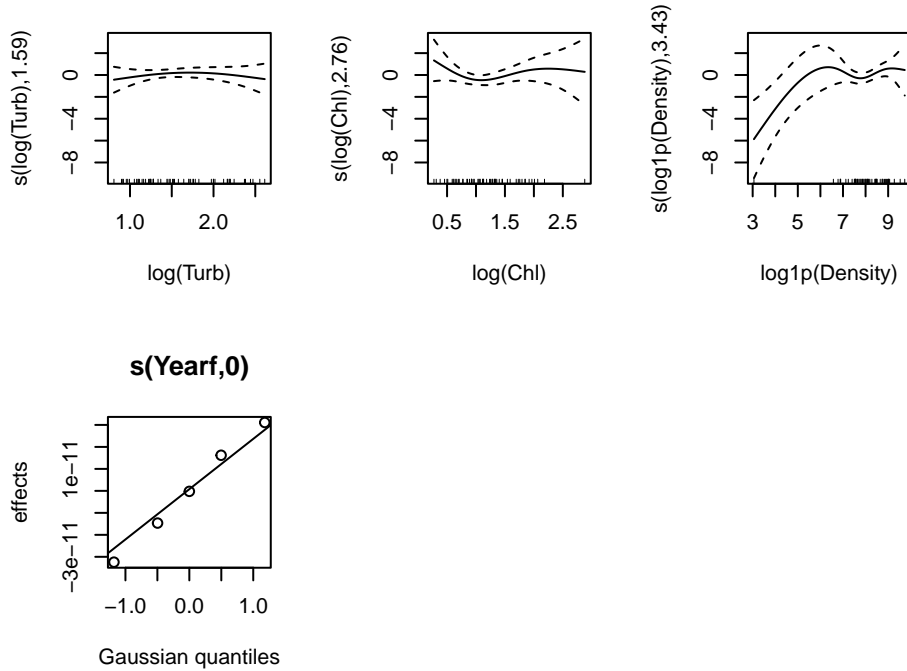
```

Plot GAM Results

```

oldpar <- par(mfrow = c(2,3))
plot(fish_gam_ss)
par(oldpar)

```



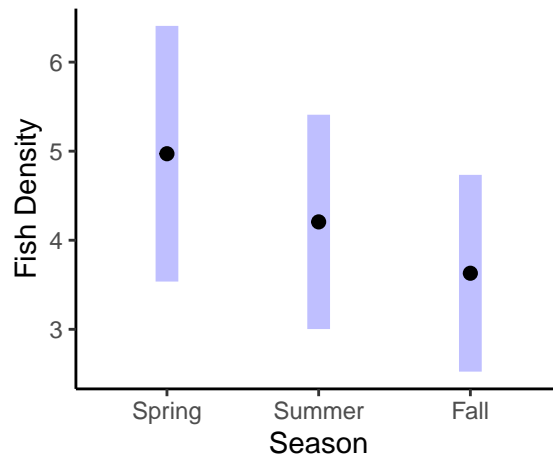
The GAM fit continues to be highly influenced by a low Density sample with very few fish.

Plot Season and Station

Seasons don't differ consistently enough to rise above other sources of variation.

```
emms <- emmeans(fish_gam_ss, ~Season)
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer    0.765 0.777 44.2   0.985 0.5899
#> Spring - Fall      1.342 0.721 44.2   1.861 0.1622
#> Summer - Fall       0.576 0.537 44.2   1.073 0.5357
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates

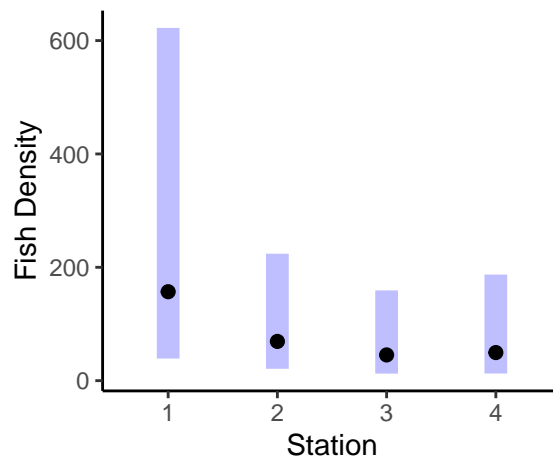
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```

Neither do Stations,

```
emms <- emmeans(fish_gam_ss, ~Station, type = 'response')
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate      SE    df t.ratio p.value
#> Station1 - Station2  0.8101 0.656 44.2   1.235 0.6080
#> Station1 - Station3  1.2246 0.657 44.2   1.863 0.2587
#> Station1 - Station4  1.1382 0.734 44.2   1.551 0.4167
#> Station2 - Station3  0.4145 0.656 44.2   0.632 0.9211
#> Station2 - Station4  0.3281 0.693 44.2   0.474 0.9645
#> Station3 - Station4 -0.0864 0.627 44.2  -0.138 0.9991
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates

plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



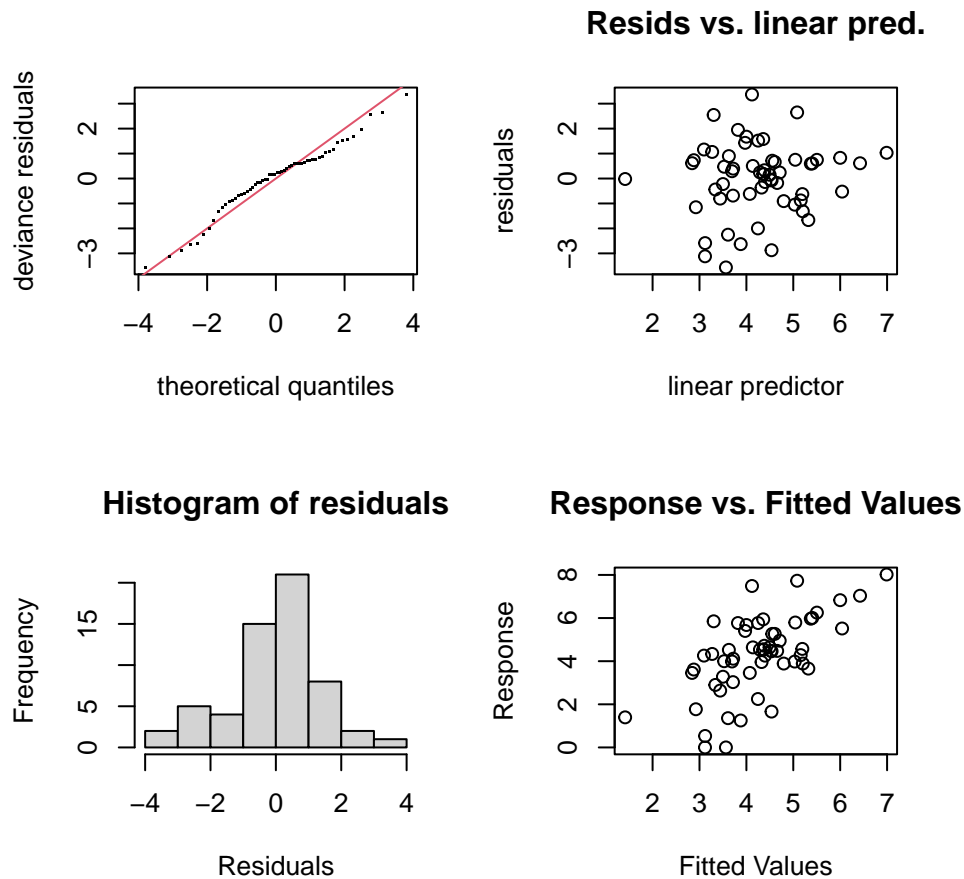
Concurvity Analysis

A common recommendation is that values over 0.8 are problematic. We have no problems here.

```
concurvity(fish_gam_ss)
#>      para s(log(Turb)) s(log(Chl)) s(log1p(Density)) s(Yearf)
#> worst      1      0.5073677      0.7512989      0.7827087 1.0000000
#> observed    1      0.4190355      0.4611581      0.4738400 0.2154188
#> estimate    1      0.4562358      0.6123281      0.5243133 0.4727051
```

Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(fish_gam_ss)
```



```
#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 17 iterations.
#> The RMS GCV score gradient at convergence was 3.656833e-07 .
#> The Hessian was positive definite.
```

```
#> Model rank = 23 / 23
#>
#> 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(log(Turb))  4.00e+00 1.59e+00   0.90   0.21
#> s(log(Chl))   4.00e+00 2.76e+00   1.06   0.72
#> s(log1p(Density)) 4.00e+00 3.43e+00   1.05   0.56
#> s(Yearf)      5.00e+00 2.36e-10    NA    NA
par(oldpar)
```

The model is pretty good, with slightly skewed and slightly heavy tails to the residuals.

Model on Reduced Data

We refit to data that omits samples where salinity was below 10 PSU.

This reduced data set drops the lowest fish abundance sample in the data set, because it happened to coincide with a low salinity (and low plankton density sample).

```
fish_gam_ss_no_low <- gam(log1p(Fish) ~
  Season +
  Station +
  s(log(Turb), bs="ts", k = 5) +
  s(log(Chl), bs="ts", k = 5) +
  s(log1p(Density), bs="ts", k = 5) +
  s(Yearf, bs = 're'),
  data = drop_low, family = 'gaussian')
summary(fish_gam_ss_no_low)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Fish) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Density), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   6.3384      0.8487   7.469 2.95e-09 ***
#> SeasonSummer -1.2409      0.7957  -1.559  0.1263
#> SeasonFall   -1.8724      0.7471  -2.506  0.0161 *
#> Station2     -0.8761      0.6984  -1.254  0.2166
#> Station3     -1.3048      0.6649  -1.962  0.0563 .
#> Station4     -1.2543      0.7395  -1.696  0.0972 .
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df      F p-value
#> s(log(Turb))   5.312e-09     4 0.000  0.5950
```

```

#> s(log(Chl))          2.887e+00      4 1.543 0.0978 .
#> s(log1p(Density)) 3.706e+00      4 2.308 0.0564 .
#> s(Yearf)           1.155e-10      5 0.000 0.7247
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.262   Deviance explained = 42%
#> GCV = 3.1124   Scale est. = 2.3998      n = 55

```

This model explains 42% of the variance, quite a bit better than the equivalent model based only on the quantitative environmental predictors. Fish abundance (in our true “marine” sample) does appear to be affected by time of year and possibly by zooplanktn density.

```

anova(fish_gam_ss_no_low)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Fish) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Density), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>      df      F p-value
#> Season  2 3.305 0.0463
#> Station 3 1.417 0.2512
#>
#> Approximate significance of smooth terms:
#>      edf   Ref.df    F p-value
#> s(log(Turb))    5.312e-09 4.000e+00 0.000 0.5950
#> s(log(Chl))     2.887e+00 4.000e+00 1.543 0.0978
#> s(log1p(Density)) 3.706e+00 4.000e+00 2.308 0.0564
#> s(Yearf)        1.155e-10 5.000e+00 0.000 0.7247

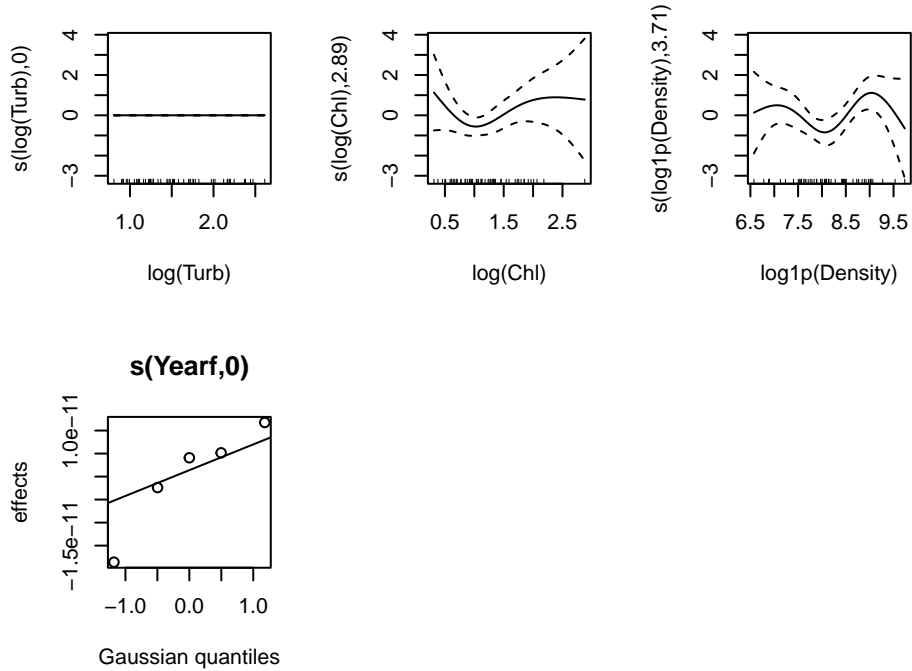
```

Plot GAM Results

```

oldpar <- par(mfrow = c(2,3))
plot(fish_gam_ss_no_low)
par(oldpar)

```

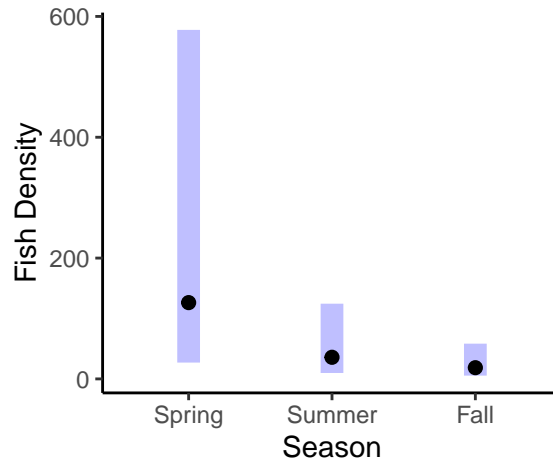


Plot Season and Station

While Seasons differ by ANOVA, no pairwise comparisons are significant. Noise in this relationship is fairly high, obscuring patterns.

```
emms <- emmeans(fish_gam_ss_no_low, ~Season, type = 'response')
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE    df t.ratio p.value
#> Spring - Summer    1.241 0.796 42.4   1.559 0.2742
#> Spring - Fall      1.872 0.747 42.4   2.506 0.0418
#> Summer - Fall       0.632 0.527 42.4   1.199 0.4602
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates

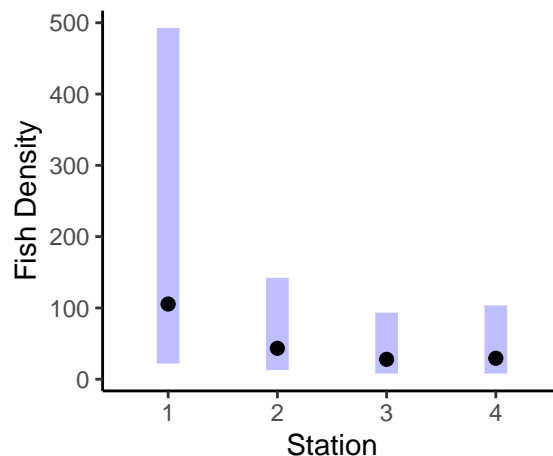
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



A similar finding for Station: while there are some suggestive patterns, none rise above the noise.

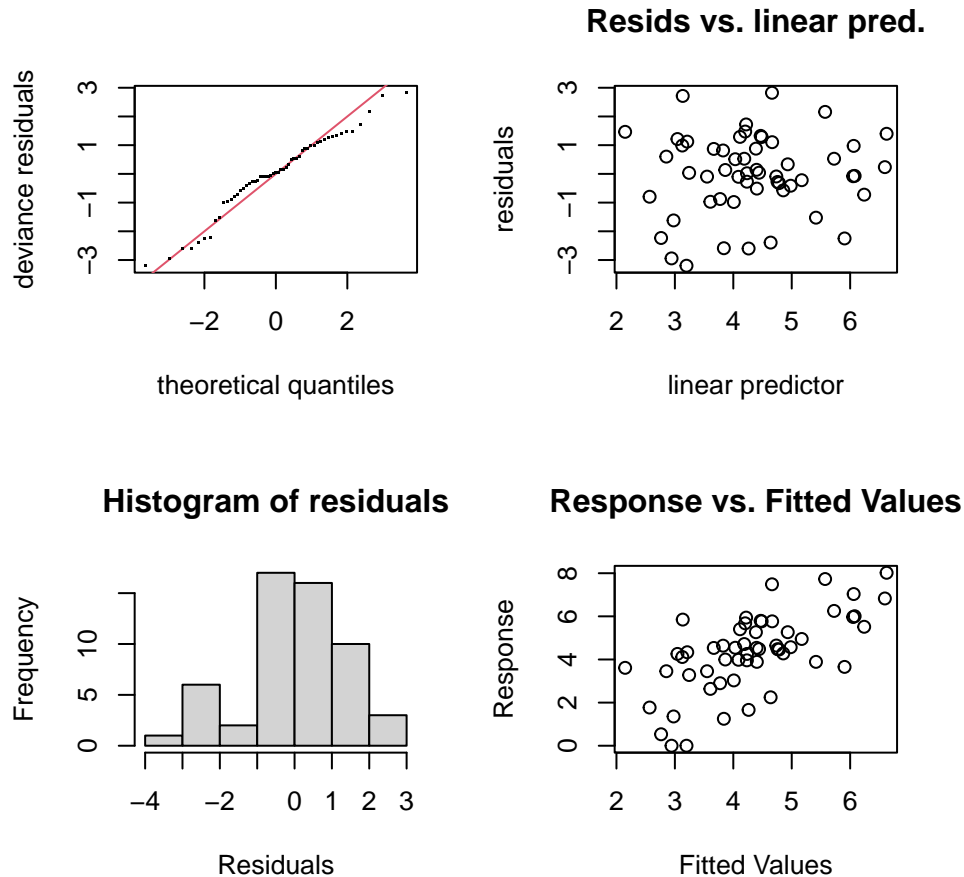
```
emms <- emmeans(fish_gam_ss_no_low, ~Station, type = 'response')
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate      SE    df t.ratio p.value
#> Station1 - Station2  0.8761 0.698 42.4   1.254 0.5965
#> Station1 - Station3  1.3048 0.665 42.4   1.962 0.2183
#> Station1 - Station4  1.2543 0.740 42.4   1.696 0.3384
#> Station2 - Station3  0.4288 0.596 42.4   0.719 0.8889
#> Station2 - Station4  0.3782 0.606 42.4   0.624 0.9239
#> Station3 - Station4 -0.0506 0.608 42.4  -0.083 0.9998
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates

plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(fish_gam_ss_no_low)
```



```
#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 27 iterations.
#> The RMS GCV score gradient at convergence was 2.837831e-07 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb))  4.00e+00 5.31e-09  0.87  0.14
#> s(log(Chl))   4.00e+00 2.89e+00  1.08  0.59
#> s(log1p(Density)) 4.00e+00 3.71e+00  1.11  0.78
#> s(Yearf)      5.00e+00 1.15e-10   NA   NA
par(oldpar)
```

That model is fairly robust, although it shows some signs of kurtosis again

Total Zooplankton Density

I fit the simplified model without Station. The full model has the same concurvity problems as before, and here the model fails to converge. While I could alter the convergence criteria to search for a solution, we know the model that includes Station will have concurvity problems, so there is little point.

Model 1

```
density_gam<- gam(log(Density) ~
  Season +
  Station +
  s(log(Turb), bs="ts", k = 5) +
  s(log(Chl), bs="ts", k = 5) +
  s(log1p(Fish),bs="ts", k = 5) +
  s(Yearf, bs = 're'),
  data = complete_data, family = 'gaussian')
summary(density_gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   7.945722   0.420321  18.904   <2e-16 ***
#> SeasonSummer -0.004627   0.371614  -0.012   0.990
#> SeasonFall    0.021791   0.355962   0.061   0.951
#> Station2      0.085937   0.323406   0.266   0.792
#> Station3      0.285820   0.321314   0.890   0.378
#> Station4     -0.155957   0.356717  -0.437   0.664
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df      F p-value
#> s(log(Turb))   0.4320      4 0.255 0.13866
#> s(log(Chl))    3.1449      4 2.403 0.11719
#> s(log1p(Fish)) 0.8286      4 1.538 0.01296 *
#> s(Yearf)       2.8534      4 2.710 0.00879 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.323   Deviance explained = 46.9%
#> GCV = 0.83252   Scale est. = 0.6422      n = 58
```

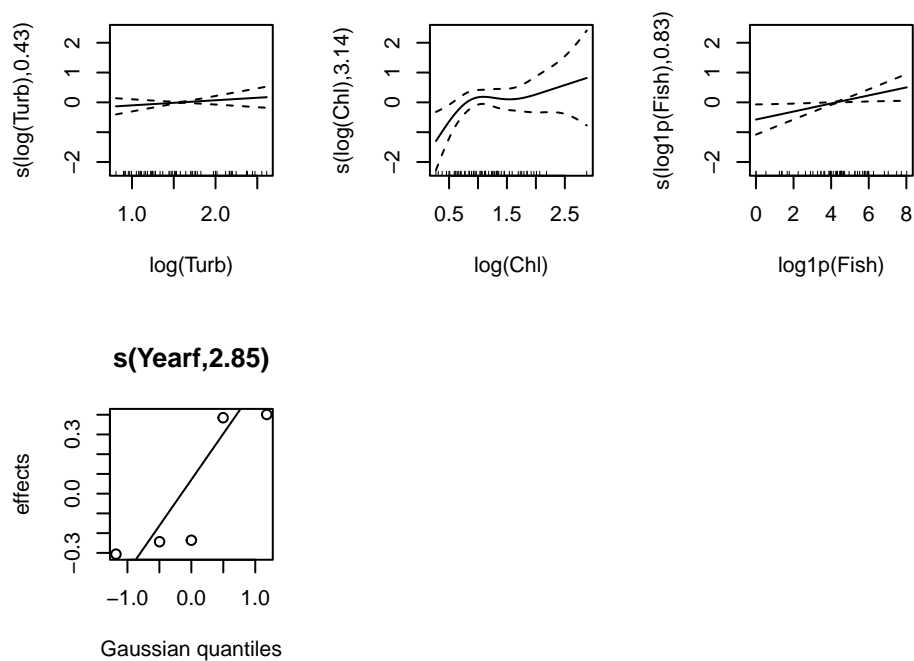

This model explains a far lower proportion of the variance than the model based solely on quantitative predictors.

```
anova(density_gam)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>      df      F p-value
#> Season  2 0.005  0.995
#> Station 3 0.732  0.538
#>
#> Approximate significance of smooth terms:
#>      edf Ref.df      F p-value
#> s(log(Turb))  0.4320 4.0000 0.255 0.13866
#> s(log(Chl))   3.1449 4.0000 2.403 0.11719
#> s(log1p(Fish)) 0.8286 4.0000 1.538 0.01296
#> s(Yearf)      2.8534 4.0000 2.710 0.00879
```

The only significant pattern is with fish abundance. Zooplankton density tends to increase along with fish abundance.

Plot the GAM

```
oldpar <- par(mfrow = c(2,3))
plot(density_gam)
par(oldpar)
```

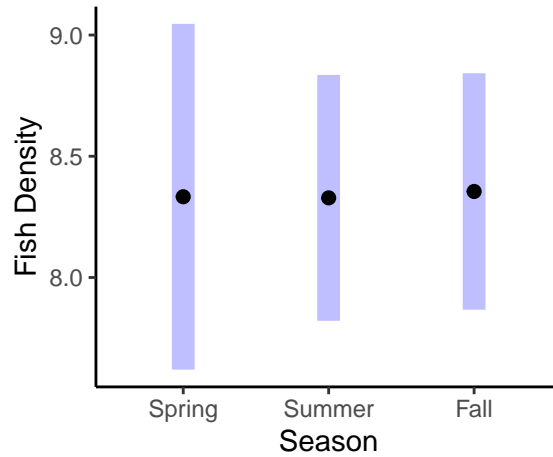


Plot Season and Station

We see absolutely no pattern with either Season or Station.

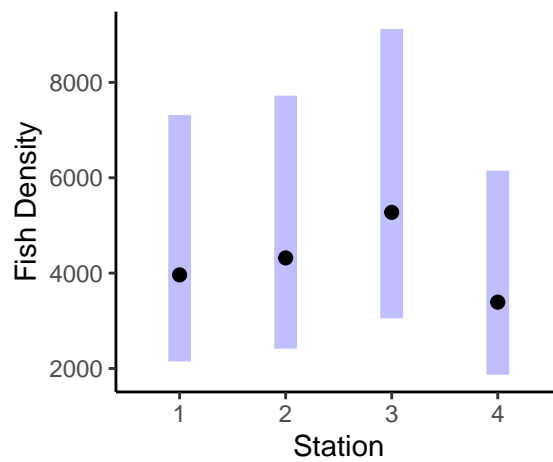
```
emms <- emmeans(density_gam, ~Season)
pairs(emms)
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer  0.00463 0.372 44.7   0.012  0.9999
#> Spring - Fall   -0.02179 0.356 44.7  -0.061  0.9979
#> Summer - Fall   -0.02642 0.269 44.7  -0.098  0.9947
#>
#> Results are averaged over the levels of: Station, Yearf
#> Results are given on the log (not the response) scale.
#> P value adjustment: tukey method for comparing a family of 3 estimates

plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



```
emms <- emmeans(density_gam, ~Station, type = 'response')
pairs(emms)
#> contrast      ratio    SE   df null t.ratio p.value
#> Station1 / Station2 0.918 0.297 44.7    1  -0.266 0.9933
#> Station1 / Station3 0.751 0.241 44.7    1  -0.890 0.8103
#> Station1 / Station4 1.169 0.417 44.7    1   0.437 0.9717
#> Station2 / Station3 0.819 0.251 44.7    1  -0.653 0.9139
#> Station2 / Station4 1.274 0.407 44.7    1   0.757 0.8734
#> Station3 / Station4 1.555 0.485 44.7    1   1.418 0.4950
#>
#> Results are averaged over the levels of: Season, Yearf
#> P value adjustment: tukey method for comparing a family of 4 estimates
#> Tests are performed on the log scale

plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Model on Reduced Data

```
density_gam_ss_no_low<- gam(log(Density) ~
                             Season +
                             Station +
                             s(log(Turb), bs="ts", k = 5) +
                             s(log(Chl), bs="ts", k = 5) +
                             s(log1p(Fish),bs="ts", k = 5) +
                             s(Yearf, bs = 're'),
                             data = drop_low, family = 'gaussian')
summary(density_gam_ss_no_low)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   8.7778      0.3521  24.929 < 2e-16 ***
#> SeasonSummer -0.3430      0.2125  -1.614  0.11378
#> SeasonFall   -0.2746      0.1873  -1.466  0.14987
#> Station2     -0.5458      0.1907  -2.863  0.00644 **
#> Station3     -0.3498      0.1936  -1.807  0.07773 .
#> Station4     -0.7252      0.2179  -3.328  0.00179 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df      F p-value
#> s(log(Turb))   8.439e-01     4  3.092 0.00764 **
#> s(log(Chl))    9.508e-01     4 13.176 0.00338 **
#> s(log1p(Fish)) 2.448e-10     4  0.000 0.78663
#> s(Yearf)       3.782e+00     4 15.672 < 2e-16 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.625 Deviance explained = 69.9%
#> GCV = 0.24733 Scale est. = 0.19527 n = 55
```

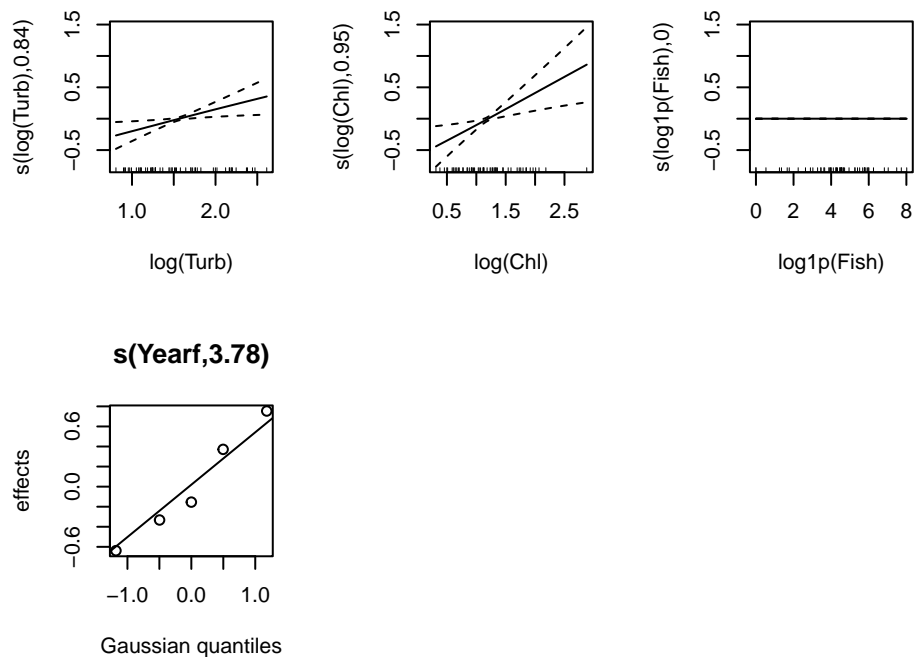
```
anova(density_gam_ss_no_low)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
```

```
#> Parametric Terms:
#>      df      F p-value
#> Season  2 1.389 0.26024
#> Station 3 4.358 0.00907
#>
#> Approximate significance of smooth terms:
#>      edf   Ref.df      F p-value
#> s(log(Turb)) 8.439e-01 4.000e+00  3.092 0.00764
#> s(log(Chl))  9.508e-01 4.000e+00 13.176 0.00338
#> s(log1p(Fish)) 2.448e-10 4.000e+00  0.000 0.78663
#> s(Yearf)      3.782e+00 4.000e+00 15.672 < 2e-16
```

When we eliminate a few low salinity samples (Salinity < 10 PSU), we uncover several patterns that were previously obscured. Zooplankton tend to be higher under higher chlorophyll and higher under higher Turbidity.

Plot the GAM

```
oldpar <- par(mfrow = c(2,3))
plot(density_gam_ss_no_low)
par(oldpar)
```



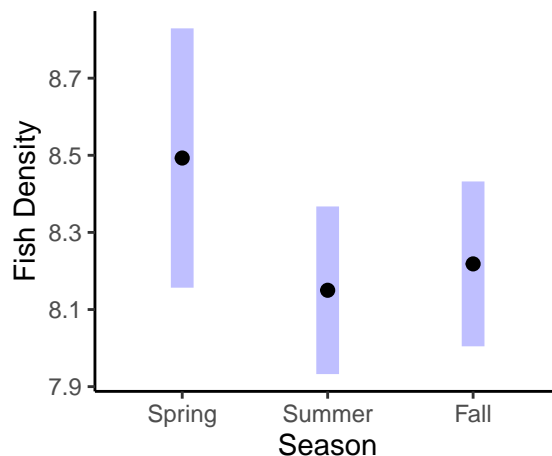
Plot Season and Station

```

emms <- emmeans(density_gam_ss_no_low, ~Season)
pairs(emms)
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer  0.3430 0.213 43.4   1.614 0.2507
#> Spring - Fall    0.2746 0.187 43.4   1.466 0.3171
#> Summer - Fall    -0.0684 0.148 43.4  -0.462 0.8891
#>
#> Results are averaged over the levels of: Station, Yearf
#> Results are given on the log (not the response) scale.
#> P value adjustment: tukey method for comparing a family of 3 estimates

plot(emms) +
  coord_flip() +
  xlab('Fish Density')

```



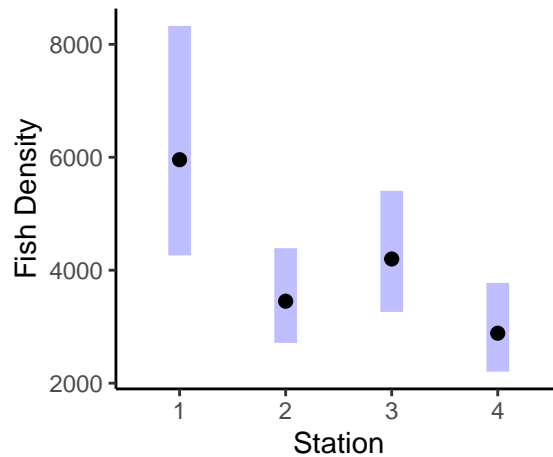
And spring emerges as having higher zooplankton abundance (except for those very low salinity samples.)

```

emms <- emmeans(density_gam_ss_no_low, ~Station, type = 'response')
pairs(emms)
#> contrast      ratio    SE   df null t.ratio p.value
#> Station1 / Station2 1.726 0.329 43.4   1   2.863 0.0315
#> Station1 / Station3 1.419 0.275 43.4   1   1.807 0.2841
#> Station1 / Station4 2.065 0.450 43.4   1   3.328 0.0093
#> Station2 / Station3 0.822 0.139 43.4   1  -1.156 0.6571
#> Station2 / Station4 1.196 0.218 43.4   1   0.985 0.7588
#> Station3 / Station4 1.456 0.249 43.4   1   2.193 0.1412
#>
#> Results are averaged over the levels of: Season, Yearf
#> P value adjustment: tukey method for comparing a family of 4 estimates
#> Tests are performed on the log scale

plot(emms) +
  coord_flip() +
  xlab('Fish Density')

```



Shannon Diversity

Model 1

```
shannon_gam_ss <- gam(H ~
  Season +
  Station +
  s(log(Turb), bs="ts", k = 5) +
  s(log(Chl), bs="ts", k = 5) +
  s(loglp(Fish), bs="ts", k = 5) +
  s(Yearf, bs = 're'),
  data = complete_data, family = 'gaussian')
summary(shannon_gam_ss)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> H ~ Season + Station + s(log(Turb), bs = "ts", k = 5) + s(log(Chl),
#>      bs = "ts", k = 5) + s(loglp(Fish), bs = "ts", k = 5) + s(Yearf,
#>      bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   1.2717      0.2367   5.372 2.39e-06 ***
#> SeasonSummer -0.2322      0.2208  -1.052  0.2983
#> SeasonFall   -0.2802      0.2073  -1.352  0.1829
#> Station2      0.3533      0.1946   1.815  0.0759 .
#> Station3      0.3781      0.1894   1.997  0.0517 .
#> Station4      0.2963      0.2049   1.446  0.1549
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df    F p-value
```

```

#> s(log(Turb))    4.607e-10    4 0.000    0.819
#> s(log(Chl))     2.656e+00    4 1.564    0.170
#> s(log1p(Fish))  3.740e-09    4 0.000    0.983
#> s(Yearf)        2.504e+00    4 1.674    0.041 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.197   Deviance explained =   34%
#> GCV = 0.29856   Scale est. = 0.24111    n = 58

```

Again, the variance explained is quite small compared to the models based on quantitative predictors.

```

anova(shannon_gam_ss)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> H ~ Season + Station + s(log(Turb), bs = "ts", k = 5) + s(log(Chl),
#>      bs = "ts", k = 5) + s(log1p(Fish), bs = "ts", k = 5) + s(Yearf,
#>      bs = "re")
#>
#> Parametric Terms:
#>      df      F p-value
#> Season  2 0.922   0.405
#> Station 3 1.578   0.207
#>
#> Approximate significance of smooth terms:
#>      edf   Ref.df    F p-value
#> s(log(Turb))  4.607e-10 4.000e+00 0.000   0.819
#> s(log(Chl))   2.656e+00 4.000e+00 1.564   0.170
#> s(log1p(Fish)) 3.740e-09 4.000e+00 0.000   0.983
#> s(Yearf)       2.504e+00 4.000e+00 1.674   0.041

```

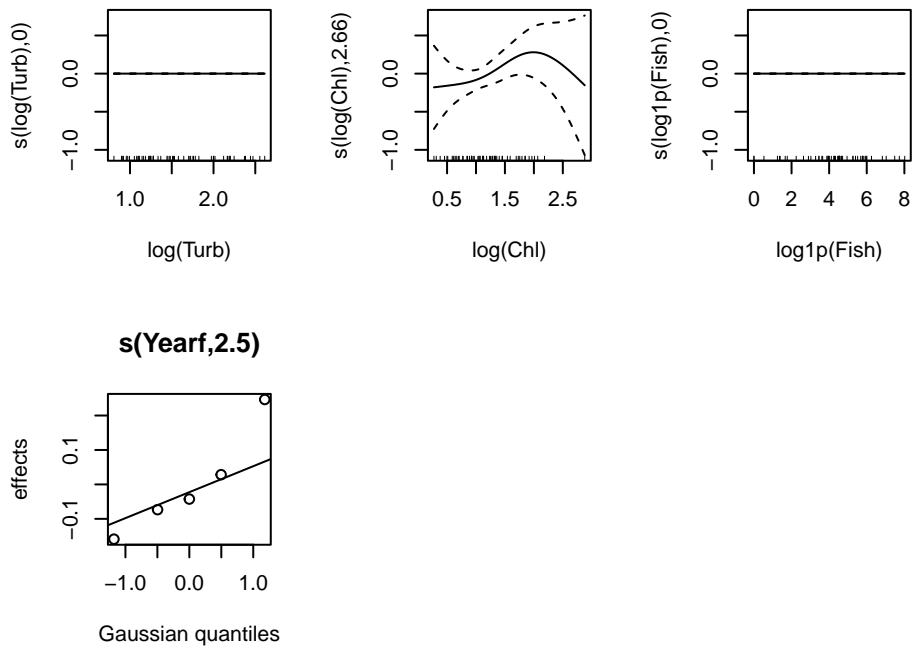
Nothing appears statistically significant. The error bands are huge...

Plot the GAM

```

oldpar <- par(mfrow = c(2,3))
plot(shannon_gam_ss)
par(oldpar)

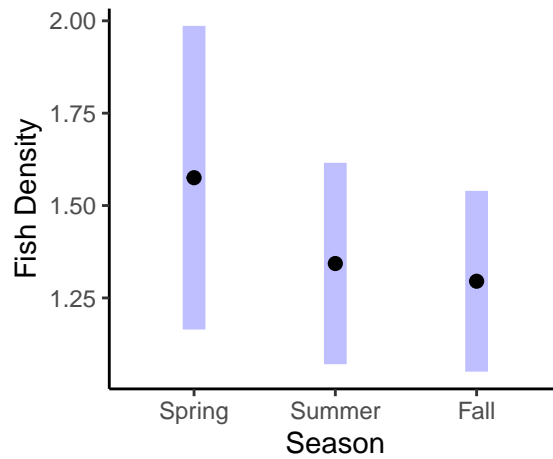
```

Plot Season and Station

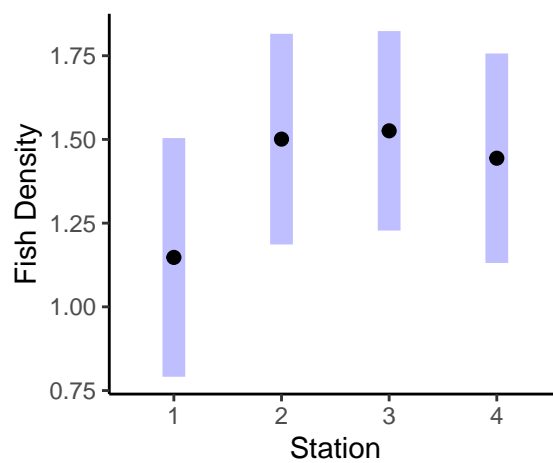
```
emms <- emmeans(shannon_gam_ss, ~Season)
pairs(emms)
#> contrast      estimate    SE    df t.ratio p.value
#> Spring - Summer  0.232 0.221 46.8   1.052  0.5485
#> Spring - Fall   0.280 0.207 46.8   1.352  0.3743
#> Summer - Fall   0.048 0.163 46.8   0.295  0.9532
#>
#> Results are averaged over the levels of: Station, Yearf
#> P value adjustment: tukey method for comparing a family of 3 estimates

plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



```
emms <- emmeans(shannon_gam_ss, ~Station, type = 'response')
pairs(emms)
#> contrast      estimate    SE   df t.ratio p.value
#> Station1 - Station2 -0.3533 0.195 46.8  -1.815  0.2791
#> Station1 - Station3 -0.3781 0.189 46.8  -1.997  0.2039
#> Station1 - Station4 -0.2963 0.205 46.8  -1.446  0.4779
#> Station2 - Station3 -0.0248 0.182 46.8   -0.136  0.9991
#> Station2 - Station4  0.0570 0.185 46.8    0.308  0.9897
#> Station3 - Station4  0.0818 0.189 46.8    0.433  0.9724
#>
#> Results are averaged over the levels of: Season, Yearf
#> P value adjustment: tukey method for comparing a family of 4 estimates

plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Model on Reduced Data

```
shannon_gam_ss_no_low <- gam(H ~
  Season +
  Station +
  s(log(Turb), bs="ts", k = 5) +
  s(log(Chl), bs="ts", k = 5) +
  s(log1p(Fish), bs="ts", k = 5) +
  s(Yearf, bs = 're'),
  data = drop_low, family = 'gaussian')
summary(shannon_gam_ss_no_low)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> H ~ Season + Station + s(log(Turb), bs = "ts", k = 5) + s(log(Chl),
#>      bs = "ts", k = 5) + s(log1p(Fish), bs = "ts", k = 5) + s(Yearf,
#>      bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   0.96383    0.24911   3.869 0.000380 ***
#> SeasonSummer -0.09453    0.20475  -0.462 0.646709
#> SeasonFall   -0.11474    0.19554  -0.587 0.560516
#> Station2      0.51231    0.17923   2.858 0.006638 **
#> Station3      0.63489    0.17390   3.651 0.000727 ***
#> Station4      0.51001    0.18976   2.688 0.010321 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df    F p-value
#> s(log(Turb))   1.082e-08    4 0.000 0.98267
#> s(log(Chl))    3.882e+00    4 6.230 0.01821 *
#> s(log1p(Fish)) 7.174e-01    4 0.868 0.04905 *
#> s(Yearf)       3.021e+00    4 3.166 0.00597 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.455  Deviance explained = 58.3%
#> GCV = 0.20476  Scale est. = 0.15405  n = 55
```

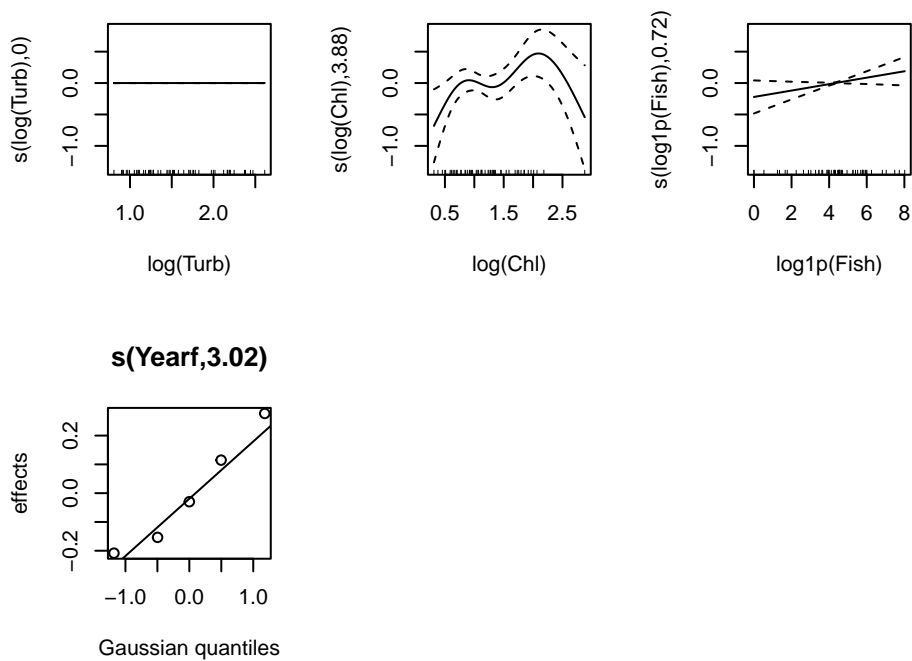
This model actually performs slightly better than the equivalent model based on quantitative predictors.

```
anova(shannon_gam_ss_no_low)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> H ~ Season + Station + s(log(Turb), bs = "ts", k = 5) + s(log(Chl),
#>      bs = "ts", k = 5) + s(log1p(Fish), bs = "ts", k = 5) + s(Yearf,
```

```
#>      bs = "re")
#>
#> Parametric Terms:
#>      df      F p-value
#> Season   2 0.172 0.84239
#> Station   3 4.601 0.00721
#>
#> Approximate significance of smooth terms:
#>      edf   Ref.df      F p-value
#> s(log(Turb))  1.082e-08 4.000e+00 0.000 0.98267
#> s(log(Chl))   3.882e+00 4.000e+00 6.230 0.01821
#> s(log1p(Fish)) 7.174e-01 4.000e+00 0.868 0.04905
#> s(Yearf)      3.021e+00 4.000e+00 3.166 0.00597
```

Plot the GAM

```
oldpar <- par(mfrow = c(2,3))
plot(shannon_gam_ss_no_low)
par(oldpar)
```

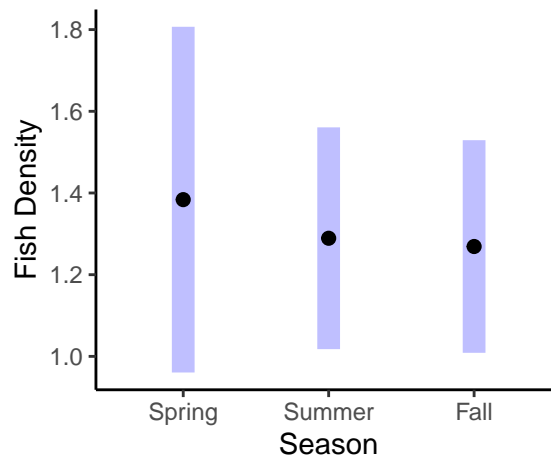


Plankton diversity is highest at intermediate levels and at low fish abundance. (But compare results for other models...).

Plot Season and Station

```
emms <- emmeans(shannon_gam_ss_no_low, ~Season)
pairs(emms)
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer  0.0945 0.205 41.4   0.462  0.8895
#> Spring - Fall    0.1147 0.196 41.4   0.587  0.8279
#> Summer - Fall    0.0202 0.134 41.4   0.151  0.9876
#>
#> Results are averaged over the levels of: Station, Yearf
#> P value adjustment: tukey method for comparing a family of 3 estimates

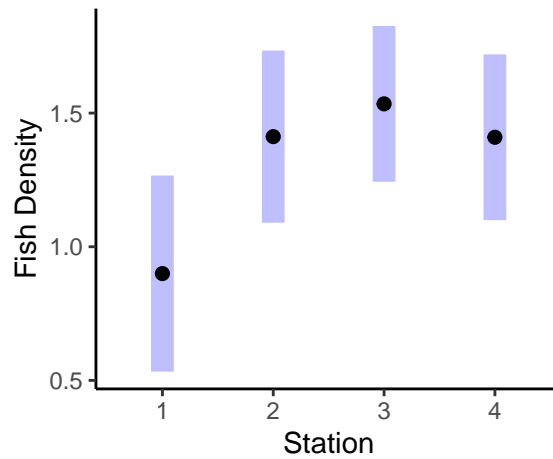
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Upstream has lower zooplankton diversity than downstream.

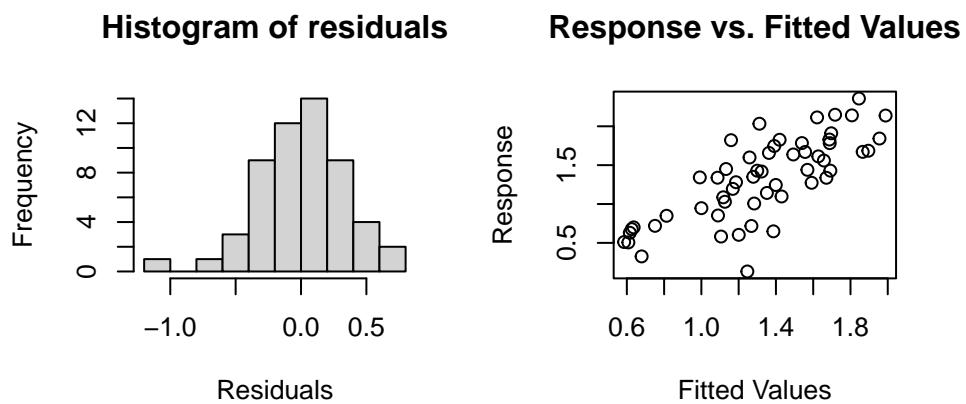
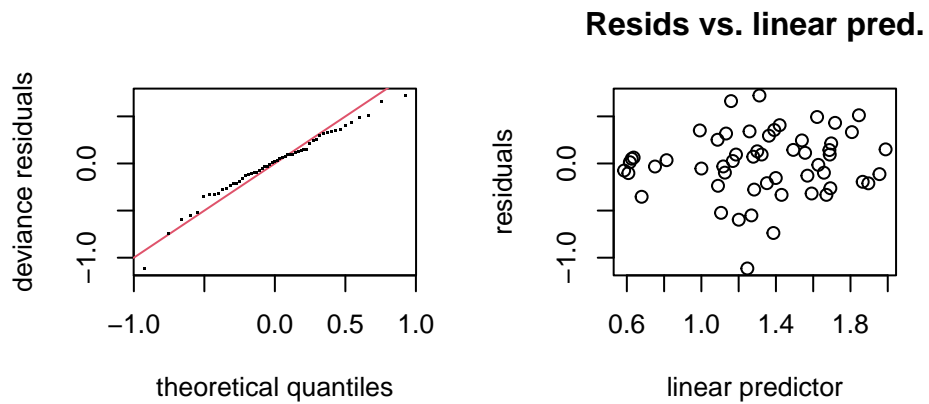
```
emms <- emmeans(shannon_gam_ss_no_low, ~Station, type = 'response')
pairs(emms)
#> contrast      estimate    SE   df t.ratio p.value
#> Station1 - Station2 -0.5123 0.179 41.4  -2.858  0.0323
#> Station1 - Station3 -0.6349 0.174 41.4  -3.651  0.0039
#> Station1 - Station4 -0.5100 0.190 41.4  -2.688  0.0487
#> Station2 - Station3 -0.1226 0.150 41.4  -0.820  0.8448
#> Station2 - Station4  0.0023 0.150 41.4   0.015  1.0000
#> Station3 - Station4  0.1249 0.153 41.4   0.814  0.8476
#>
#> Results are averaged over the levels of: Season, Yearf
#> P value adjustment: tukey method for comparing a family of 4 estimates

plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Diagnostic Plots

```
oldpar <- par(mfrow = c(2,2))
gam.check(shannon_gam_ss_no_low)
```



```

#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 19 iterations.
#> The RMS GCV score gradient at convergence was 5.69925e-08 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb)) 4.00e+00 1.08e-08   1.24   0.94
#> s(log(Chl))  4.00e+00 3.88e+00   1.05   0.61
#> s(log1p(Fish)) 4.00e+00 7.17e-01   0.93   0.27
#> s(Yearf)     5.00e+00 3.02e+00    NA    NA
par(oldpar)

```

Model looks excellent.

Single Species Models

Automating Analysis of Separate Species

I'm going to automate analysis of all selected species by using a “nested” Tibble. This is a convenient alternative to writing a “for” loop to run multiple identical analyses.

I create a “long” data source, based on the reduced data set that omits low salinity samples.

```

spp_data <- drop_low %>%
  select(Yearf, Season, Station, Temp,
         Sal, Turb, Chl, Fish,
         Acartia, Balanus, Eurytemora, Polychaete, Pseudocal, Temora) %>%
  pivot_longer(-c(Yearf:Fish), names_to = 'Species', values_to = 'Density')

```

Next, I create a function to run the analysis. This function takes a data frame or tibble as an argument. The tibble must have data columns with the correct names.

The initial model fits for some species had a lot of wiggles in them, to an extent that I thought did not make much scientific sense, so I decided to reduce the dimensionality of the GAM smoothers, by adding the parameter `k= 4`. Lower numbers constrain the GAM to fit smoother lines.

```

my_gam <- function(.dat) {
  gam(log1p(Density) ~
    Season +
    Station +
    s(log(Turb), bs="ts", k = 5) +
    s(log(Chl), bs="ts", k = 5) +
    s(log1p(Fish), bs="ts", k = 5) +
    s(Yearf, bs = 're'),
    data = .dat, family = "gaussian")
}

```

Next, I create the nested tibble, and conduct the analysis on each species...

```
spp_analysis <- spp_data %>%  
  group_by(Species) %>%  
  nest() %>%  
  mutate(gam_mods = map(data, my_gam))
```

and finally, output the model results. I can do that in a “for” loop, but it’s Awkward to look through a long list of output, so I step through each species in turn.

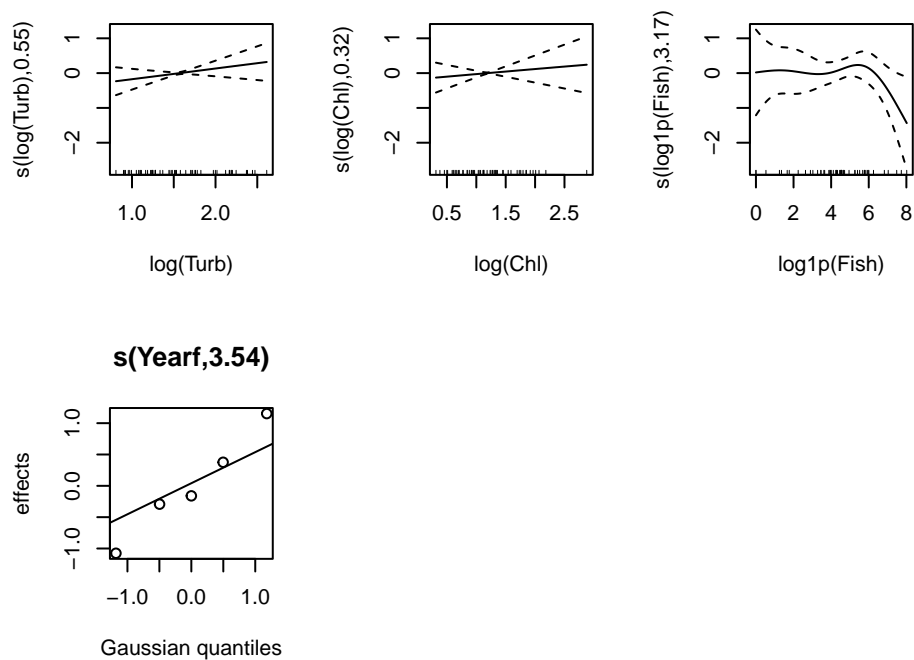
Acartia

```
spp = 'Acartia'
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>   s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>   k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>               Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   5.22540     0.64351   8.120 4.22e-10 ***
#> SeasonSummer  2.24280     0.41888   5.354 3.47e-06 ***
#> SeasonFall    2.39985     0.41403   5.796 8.16e-07 ***
#> Station2      -0.63673     0.44671  -1.425   0.162
#> Station3      -0.07111     0.44848  -0.159   0.875
#> Station4      -0.35615     0.48603  -0.733   0.468
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>               edf Ref.df      F  p-value
#> s(log(Turb))    0.5499     4 0.447   0.115
#> s(log(Chl))     0.3219     4 0.161   0.291
#> s(log1p(Fish))  3.1678     4 1.901   0.151
#> s(Yearf)        3.5446     4 6.967 6.42e-05 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.654  Deviance explained = 73.5%
#> GCV = 1.3658  Scale est. = 1.0285    n = 55
cat('\n\nANOVA')
#>
#>
#> ANOVA
anova(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>   s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>   k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>               df      F  p-value
#> Season        2 18.789 1.56e-06
```

```
#> Station 3 0.941 0.429
#>
#> Approximate significance of smooth terms:
#>           edf Ref.df      F  p-value
#> s(log(Turb)) 0.5499 4.0000 0.447 0.115
#> s(log(Chl)) 0.3219 4.0000 0.161 0.291
#> s(log1p(Fish)) 3.1678 4.0000 1.901 0.151
#> s(Yearf) 3.5446 4.0000 6.967 6.42e-05
```

Plot GAM

```
oldpar <- par(mfrow = c(2,3))
plot(mod)
par(oldpar)
```

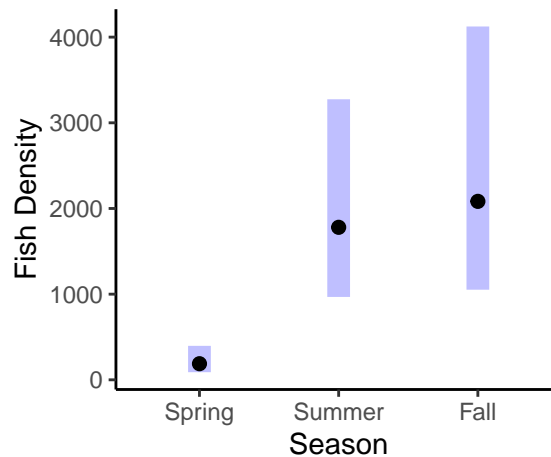


Plot Season and Station

```
emms <- emmeans(mod, ~Season, type = 'response',
                 data = spp_analysis$data[spp_analysis$Species == spp][[1]] )
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate      SE    df t.ratio p.value
#> Spring - Summer -2.243 0.419 41.4 -5.354 <.0001
#> Spring - Fall -2.400 0.414 41.4 -5.796 <.0001
```

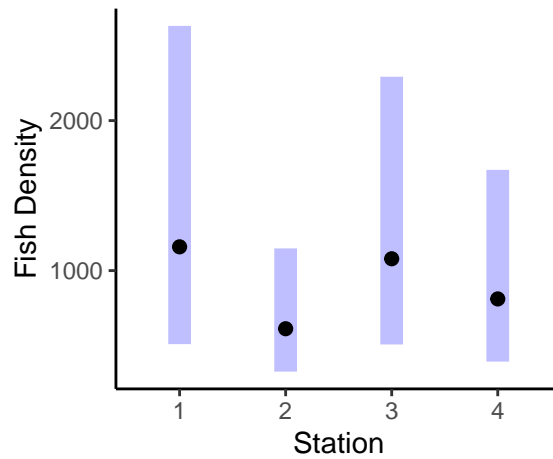
```
#> Summer - Fall      -0.157 0.340 41.4  -0.462  0.8892
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



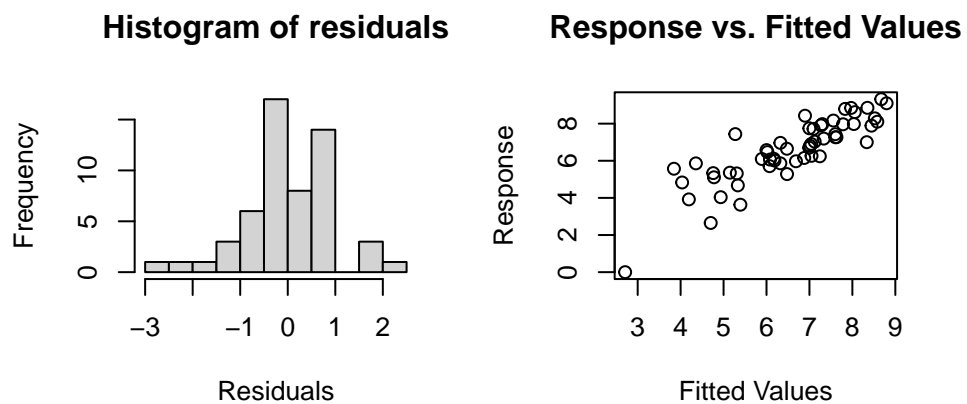
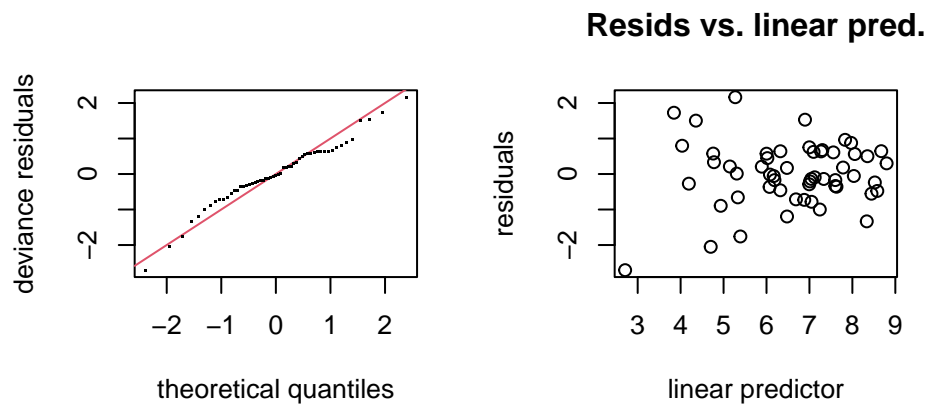
```
emms <- emmeans(mod, ~Station, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Station1 - Station2  0.6367 0.447 41.4   1.425  0.4911
#> Station1 - Station3  0.0711 0.448 41.4   0.159  0.9986
#> Station1 - Station4  0.3561 0.486 41.4   0.733  0.8834
#> Station2 - Station3 -0.5656 0.401 41.4  -1.410  0.5005
#> Station2 - Station4 -0.2806 0.407 41.4  -0.690  0.9005
#> Station3 - Station4  0.2850 0.394 41.4   0.723  0.8873
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(mod)
```



```

#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 54 iterations.
#> The RMS GCV score gradient at convergence was 1.015266e-06 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb)) 4.000 0.550   1.10   0.70
#> s(log(Chl))  4.000 0.322   0.82   0.08 .
#> s(log1p(Fish)) 4.000 3.168   1.00   0.49
#> s(Yearf)      5.000 3.545    NA    NA
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(oldpar)

```

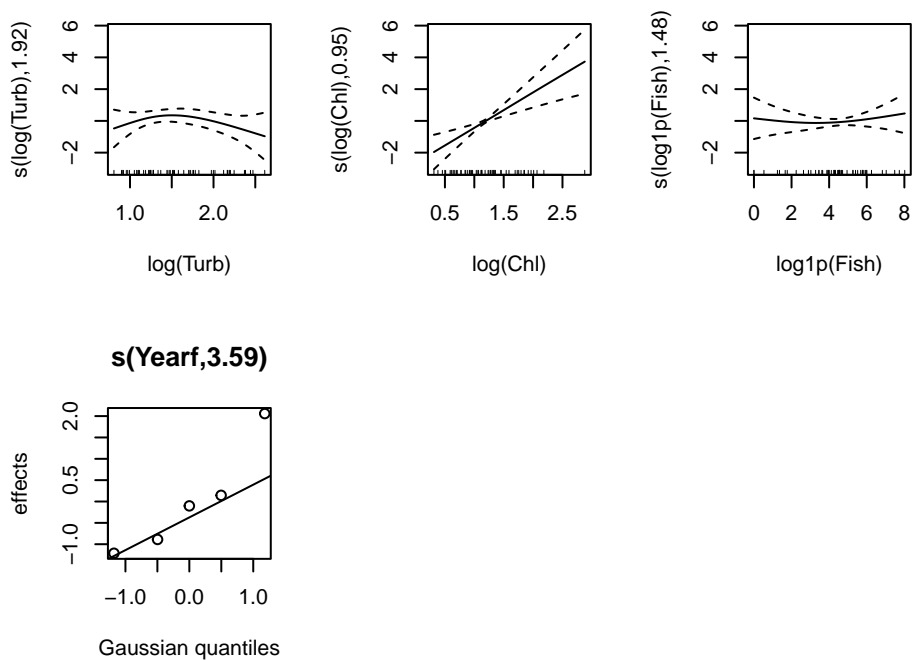
Balanus

```
spp = 'Balanus'
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)    4.2549      1.0561   4.029 0.000236 ***
#> SeasonSummer  -1.2623      0.7326  -1.723 0.092396 .
#> SeasonFall    -1.8585      0.6699  -2.774 0.008291 **
#> Station2       0.3506      0.6562   0.534 0.596017
#> Station3       0.9263      0.6762   1.370 0.178178
#> Station4       0.7601      0.7695   0.988 0.328991
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df      F p-value
#> s(log(Turb))   1.9171     4  1.176 0.183185
#> s(log(Chl))    0.9467     4 12.814 0.000201 ***
#> s(log1p(Fish)) 1.4791     4  0.267 0.529217
#> s(Yearf)       3.5947     4  8.045 2.17e-05 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.593   Deviance explained = 69%
#> GCV = 2.8665   Scale est. = 2.1401     n = 55
cat('\n\nANOVA')
#>
#>
#> ANOVA
anova(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>              df      F p-value
#> Season       2 3.992  0.026
```

```
#> Station 3 0.704 0.555
#>
#> Approximate significance of smooth terms:
#>           edf Ref.df      F p-value
#> s(log(Turb)) 1.9171 4.0000 1.176 0.183185
#> s(log(Chl)) 0.9467 4.0000 12.814 0.000201
#> s(log1p(Fish)) 1.4791 4.0000 0.267 0.529217
#> s(Yearf) 3.5947 4.0000 8.045 2.17e-05
```

Plot GAM

```
oldpar <- par(mfrow = c(2,3))
plot(mod)
par(oldpar)
```

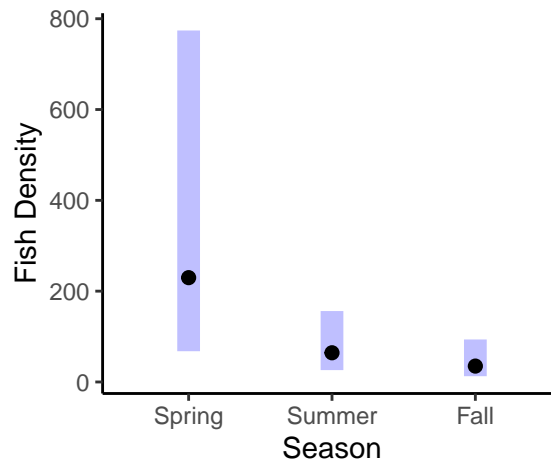


Plot Season and Station

```
emms <- emmeans(mod, ~Season, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer  1.262 0.733 41.1   1.723 0.2089
#> Spring - Fall   1.858 0.670 41.1   2.774 0.0221
```

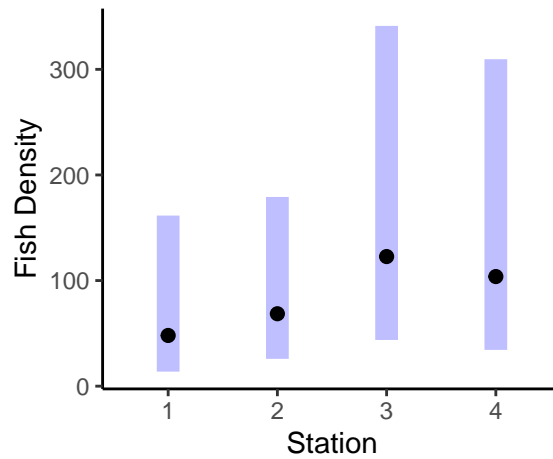
```
#> Summer - Fall      0.596 0.500 41.1  1.193  0.4642
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



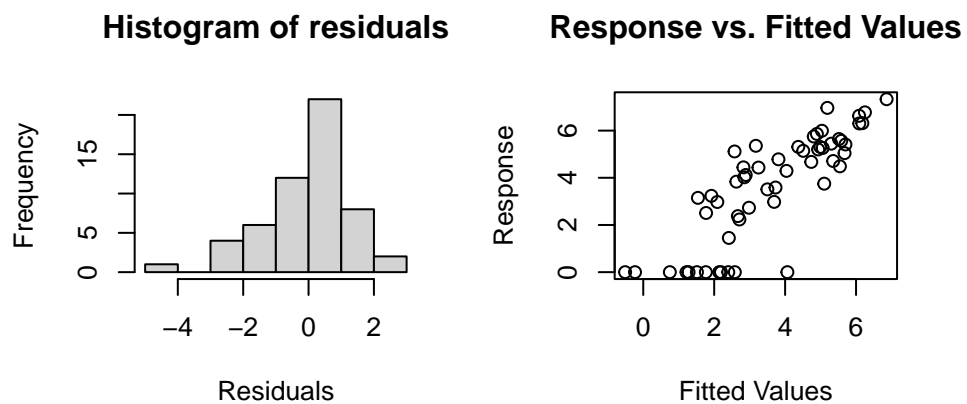
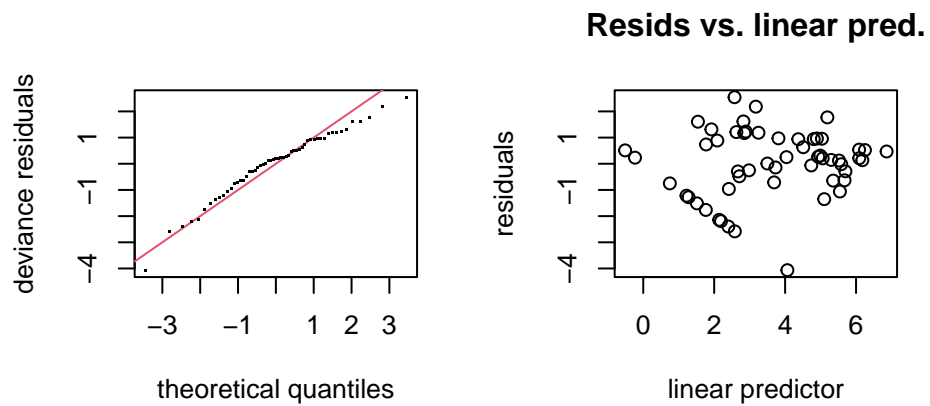
```
emms <- emmeans(mod, ~Station, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Station1 - Station2 -0.351 0.656 41.1 -0.534 0.9502
#> Station1 - Station3 -0.926 0.676 41.1 -1.370 0.5250
#> Station1 - Station4 -0.760 0.769 41.1 -0.988 0.7571
#> Station2 - Station3 -0.576 0.570 41.1 -1.010 0.7447
#> Station2 - Station4 -0.410 0.618 41.1 -0.663 0.9104
#> Station3 - Station4  0.166 0.574 41.1  0.290 0.9914
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```

Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(mod)
```



```

#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 8 iterations.
#> The RMS GCV score gradient at convergence was 7.850988e-05 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb)) 4.000 1.917   0.80  0.07 .
#> s(log(Chl))  4.000 0.947   0.99  0.41
#> s(log1p(Fish)) 4.000 1.479   0.99  0.40
#> s(Yearf)      5.000 3.595    NA    NA
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(oldpar)

```

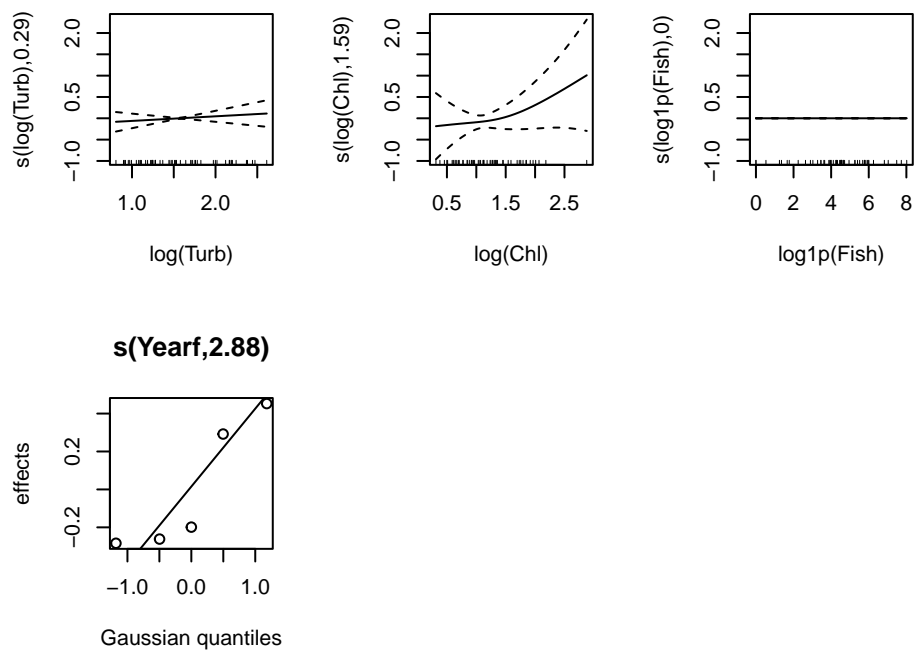
Eurytemora

```
spp = "Eurytemora"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>   s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>   k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>               Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   7.7441      0.4782  16.195 < 2e-16 ***
#> SeasonSummer -0.6914      0.4043  -1.710 0.094221 .
#> SeasonFall   -0.4968      0.3657  -1.359 0.181179
#> Station2     -0.3502      0.3560  -0.984 0.330552
#> Station3     -1.0909      0.3512  -3.106 0.003306 **
#> Station4     -1.5618      0.3871  -4.035 0.000213 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>               edf Ref.df      F p-value
#> s(log(Turb))   2.894e-01     4 0.140  0.1846
#> s(log(Chl))    1.592e+00     4 0.999  0.2398
#> s(log1p(Fish)) 7.975e-11     4 0.000  0.9535
#> s(Yearf)       2.878e+00     4 2.281  0.0208 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.383   Deviance explained = 49.5%
#> GCV = 0.83485   Scale est. = 0.67153   n = 55
cat('\n\nANOVA')
#>
#>
#> ANOVA
anova(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>   s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>   k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>               df      F p-value
#> Season        2 1.469 0.241088
```

```
#> Station 3 7.617 0.000327
#>
#> Approximate significance of smooth terms:
#>          edf    Ref.df      F p-value
#> s(log(Turb)) 2.894e-01 4.000e+00 0.140 0.1846
#> s(log(Chl))  1.592e+00 4.000e+00 0.999 0.2398
#> s(log1p(Fish)) 7.975e-11 4.000e+00 0.000 0.9535
#> s(Yearf)      2.878e+00 4.000e+00 2.281 0.0208
```

Plot GAM

```
oldpar <- par(mfrow = c(2,3))
plot(mod)
par(oldpar)
```

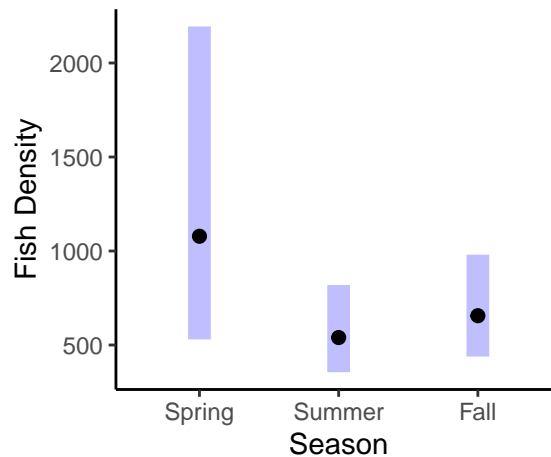


Plot Season and Station

```
emms <- emmeans(mod, ~Season, type = 'response',
                 data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer  0.691 0.404 44.2   1.710 0.2127
#> Spring - Fall    0.497 0.366 44.2   1.359 0.3712
```

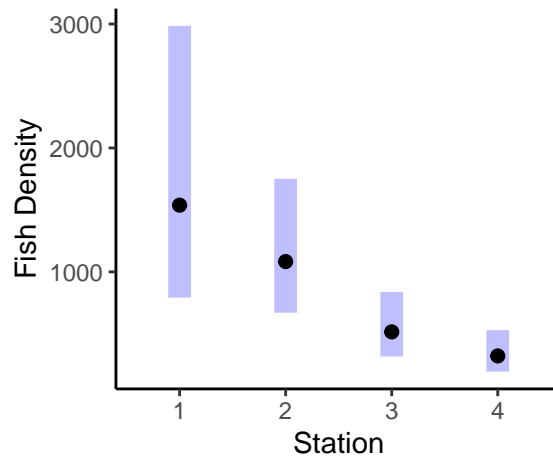
```
#> Summer - Fall      -0.195 0.275 44.2 -0.707 0.7604
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



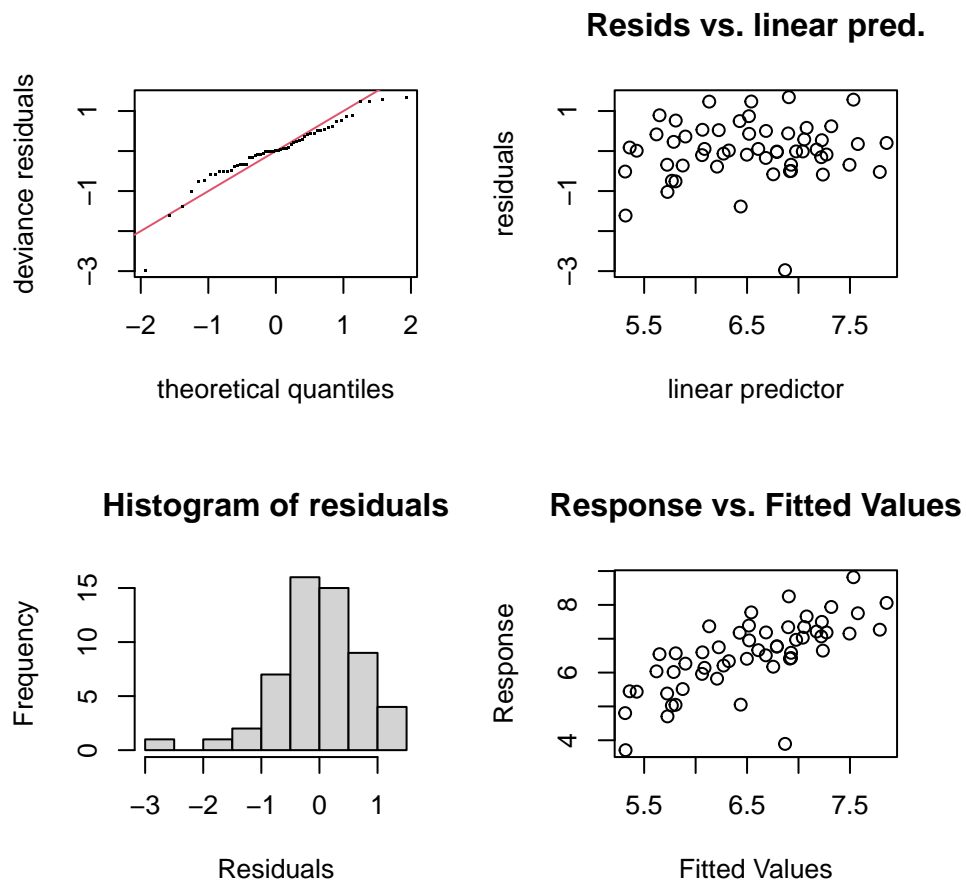
```
emms <- emmeans(mod, ~Station, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Station1 - Station2    0.350 0.356 44.2   0.984 0.7593
#> Station1 - Station3    1.091 0.351 44.2   3.106 0.0168
#> Station1 - Station4    1.562 0.387 44.2   4.035 0.0012
#> Station2 - Station3    0.741 0.307 44.2   2.416 0.0887
#> Station2 - Station4    1.212 0.319 44.2   3.799 0.0024
#> Station3 - Station4    0.471 0.316 44.2   1.491 0.4515
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(mod)
```



```

#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 16 iterations.
#> The RMS GCV score gradient at convergence was 8.747507e-08 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb)) | 4.00e+00 | 2.89e-01 | 0.86 | 0.10 |
| #> s(log(Chl)) | 4.00e+00 | 1.59e+00 | 1.02 | 0.47 |
| #> s(log1p(Fish)) | 4.00e+00 | 7.97e-11 | 0.99 | 0.38 |
| #> s(Yearf) | 5.00e+00 | 2.88e+00 | NA | NA |

```

par(oldpar)

```

Polychaete

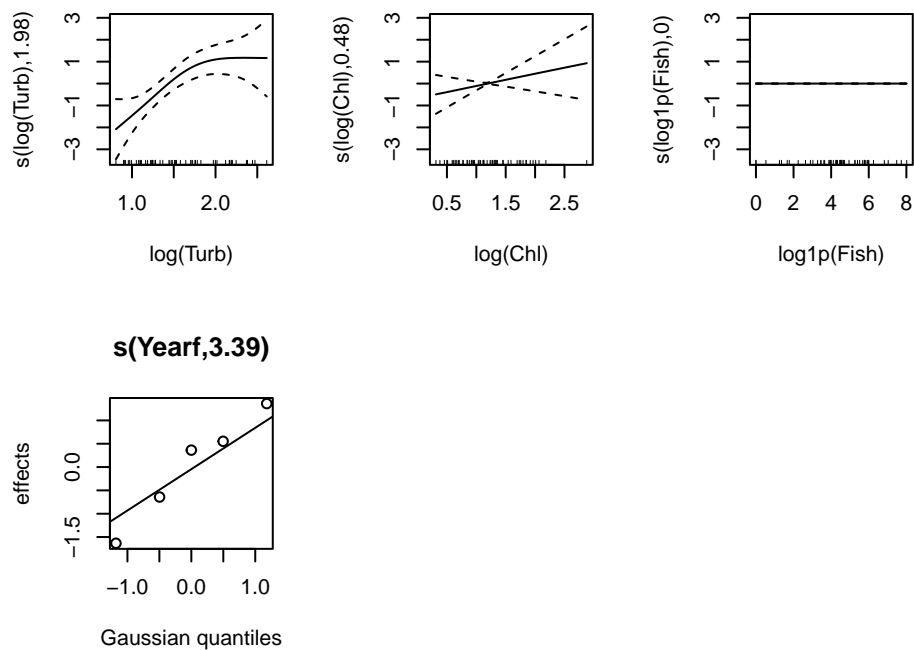
```
spp = "Polychaete"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   3.5524      1.0087   3.522  0.00103 **
#> SeasonSummer -3.3010      0.7331  -4.503 5.04e-05 ***
#> SeasonFall    -2.2816      0.6700  -3.405  0.00144 **
#> Station2       1.4572      0.7170   2.032  0.04830 *
#> Station3       2.0591      0.7423   2.774  0.00815 **
#> Station4       2.1983      0.8199   2.681  0.01035 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df      F p-value
#> s(log(Turb))   1.983e+00     4 5.860 0.000680 ***
#> s(log(Chl))    4.812e-01     4 0.631 0.106998
#> s(log1p(Fish)) 2.941e-10     4 0.000 0.714175
#> s(Yearf)       3.388e+00     4 4.908 0.000644 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.563   Deviance explained = 65.1%
#> GCV = 3.7452   Scale est. = 2.9381     n = 55
cat('\n\nANOVA')
#>
#>
#> ANOVA
anova(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>              df      F p-value
#> Season       2 10.257 0.000227
```



```
#> Station 3 3.018 0.040001
#>
#> Approximate significance of smooth terms:
#>          edf    Ref.df      F  p-value
#> s(log(Turb))  1.983e+00 4.000e+00 5.860 0.000680
#> s(log(Chl))   4.812e-01 4.000e+00 0.631 0.106998
#> s(log1p(Fish)) 2.941e-10 4.000e+00 0.000 0.714175
#> s(Yearf)       3.388e+00 4.000e+00 4.908 0.000644
```

Plot GAM

```
oldpar <- par(mfrow = c(2,3))
plot(mod)
par(oldpar)
```

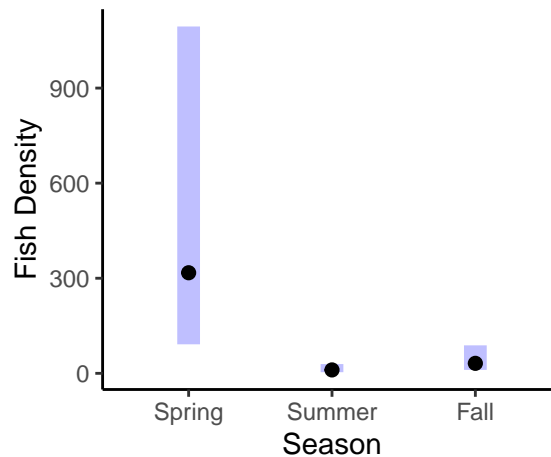


Plot Season and Station

```
emms <- emmeans(mod, ~Season, type = 'response',
                 data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer    3.30 0.733 43.1   4.503 0.0001
#> Spring - Fall      2.28 0.670 43.1   3.405 0.0040
```

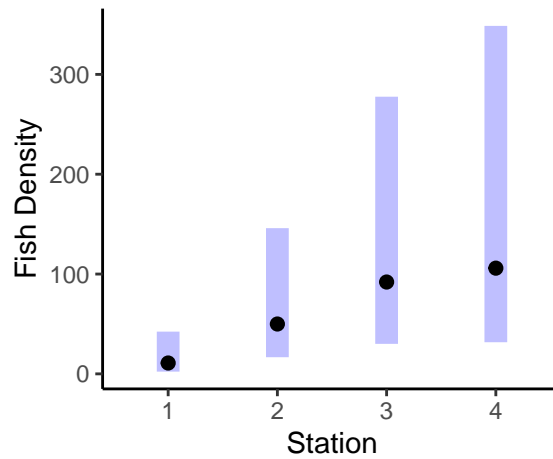
```
#> Summer - Fall      -1.02 0.568 43.1 -1.796 0.1831
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



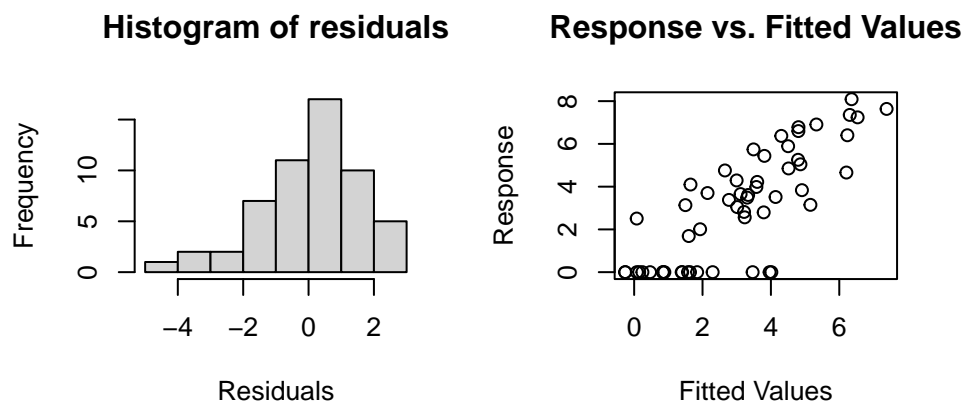
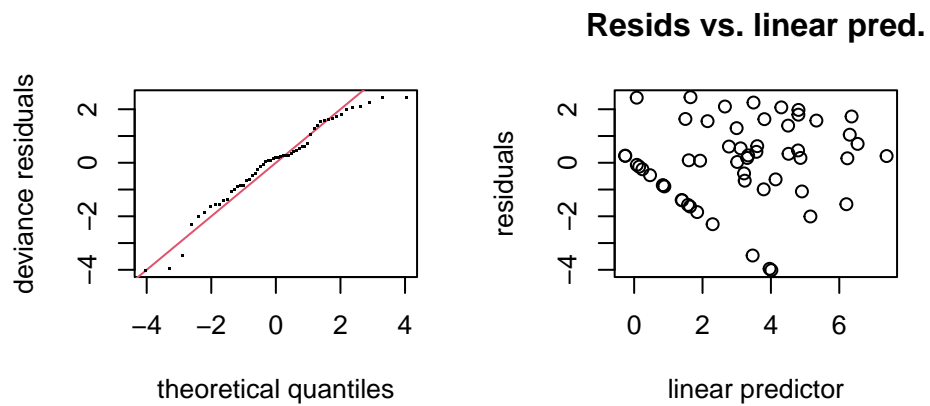
```
emms <- emmeans(mod, ~Station, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Station1 - Station2 -1.457 0.717 43.1 -2.032 0.1922
#> Station1 - Station3 -2.059 0.742 43.1 -2.774 0.0392
#> Station1 - Station4 -2.198 0.820 43.1 -2.681 0.0490
#> Station2 - Station3 -0.602 0.662 43.1 -0.909 0.8004
#> Station2 - Station4 -0.741 0.716 43.1 -1.035 0.7299
#> Station3 - Station4 -0.139 0.658 43.1 -0.212 0.9966
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(mod)
```



```

#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 15 iterations.
#> The RMS GCV score gradient at convergence was 1.639085e-07 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb)) 4.00e+00 1.98e+00 0.97 0.330
#> s(log(Chl)) 4.00e+00 4.81e-01 0.73 0.015 *
#> s(log1p(Fish)) 4.00e+00 2.94e-10 1.02 0.510
#> s(Yearf)      5.00e+00 3.39e+00    NA    NA
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
par(oldpar)

```

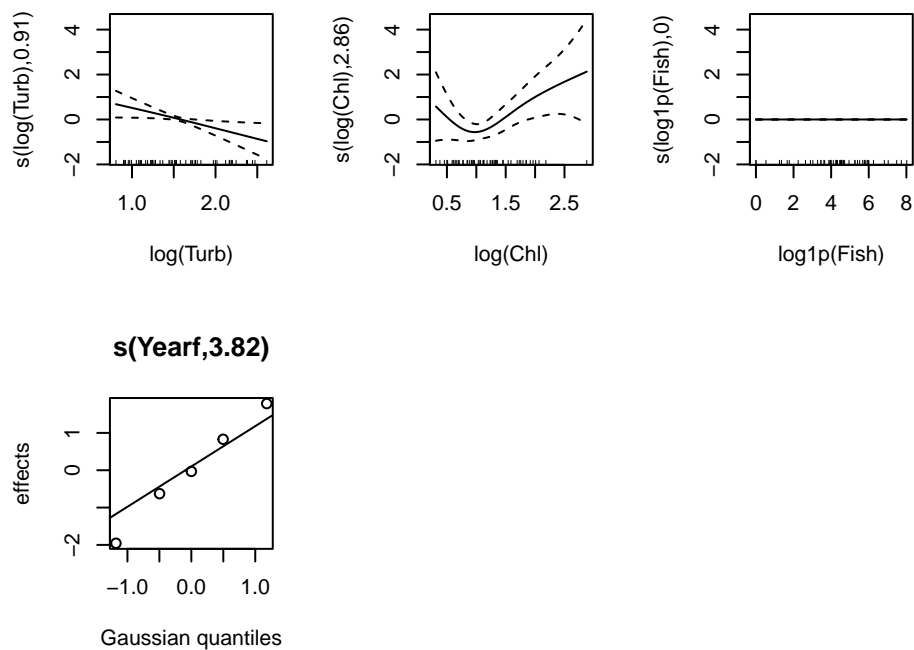
Pseudocal

```
spp = "Pseudocal"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>   s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>   k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>               Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   3.04774     1.05116   2.899  0.00596 **
#> SeasonSummer -0.06804     0.62108  -0.110  0.91330
#> SeasonFall   -1.01624     0.57450  -1.769  0.08427 .
#> Station2      2.59088     0.51615   5.020 1.03e-05 ***
#> Station3      2.95918     0.52315   5.656 1.29e-06 ***
#> Station4      1.73504     0.59594   2.911  0.00577 **
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>               edf Ref.df      F  p-value
#> s(log(Turb))   9.071e-01     4  2.145   0.0119 *
#> s(log(Chl))    2.857e+00     4  9.716   0.0760 .
#> s(log1p(Fish)) 8.479e-10     4  0.000   0.4257
#> s(Yearf)       3.824e+00     4 12.050 8.98e-07 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) =  0.702   Deviance explained = 77.1%
#> GCV = 1.7994   Scale est. = 1.3548     n = 55
cat('\n\nANOVA')
#>
#>
#> ANOVA
anova(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>   s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>   k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>               df      F  p-value
#> Season        2  3.669   0.0341
```

```
#> Station 3 13.246 3.34e-06
#>
#> Approximate significance of smooth terms:
#>          edf    Ref.df      F  p-value
#> s(log(Turb))  9.071e-01  4.000e+00  2.145  0.0119
#> s(log(Chl))   2.857e+00  4.000e+00  9.716  0.0760
#> s(log1p(Fish)) 8.479e-10  4.000e+00  0.000  0.4257
#> s(Yearf)      3.824e+00  4.000e+00 12.050 8.98e-07
```

Plot GAM

```
oldpar <- par(mfrow = c(2,3))
plot(mod)
par(oldpar)
```

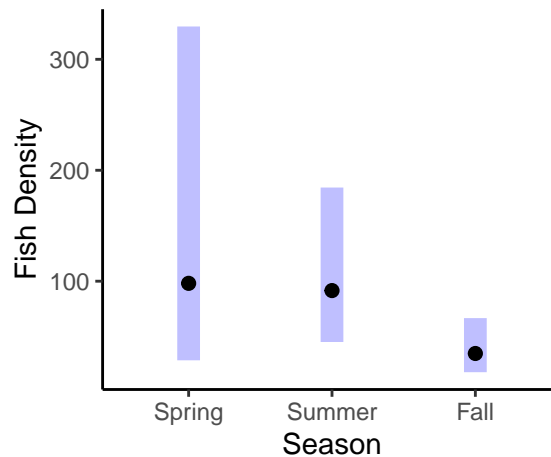


Plot Season and Station

```
emms <- emmeans(mod, ~Season, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer  0.068 0.621 41.4   0.110  0.9934
#> Spring - Fall   1.016 0.574 41.4   1.769  0.1926
```

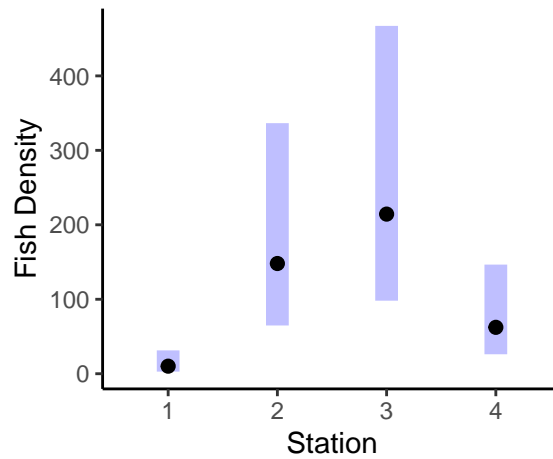
```
#> Summer - Fall      0.948 0.396 41.4  2.394  0.0544
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



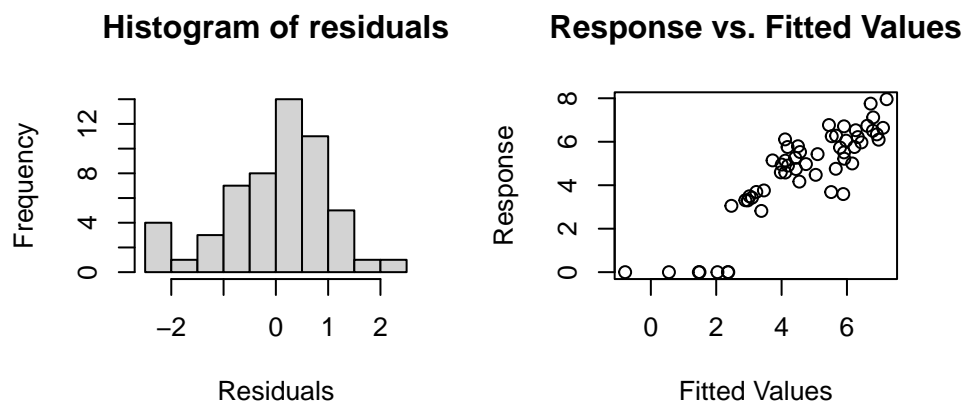
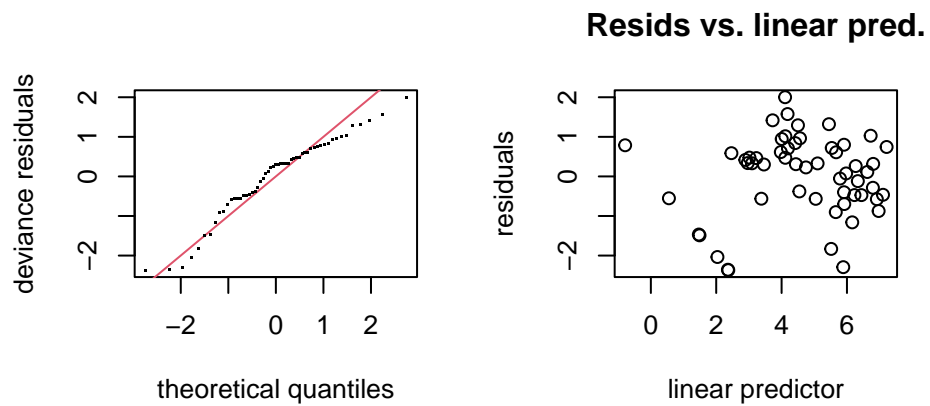
```
emms <- emmeans(mod, ~Station, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE  df t.ratio p.value
#> Station1 - Station2 -2.591 0.516 41.4 -5.020 0.0001
#> Station1 - Station3 -2.959 0.523 41.4 -5.656 <.0001
#> Station1 - Station4 -1.735 0.596 41.4 -2.911 0.0284
#> Station2 - Station3 -0.368 0.454 41.4 -0.811 0.8489
#> Station2 - Station4  0.856 0.491 41.4  1.744 0.3145
#> Station3 - Station4  1.224 0.459 41.4  2.669 0.0509
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(mod)
```




```

#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 20 iterations.
#> The RMS GCV score gradient at convergence was 1.725488e-07 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb)) | 4.00e+00 | 9.07e-01 | 1.07 | 0.60 |
| #> s(log(Chl)) | 4.00e+00 | 2.86e+00 | 1.14 | 0.82 |
| #> s(log1p(Fish)) | 4.00e+00 | 8.48e-10 | 1.00 | 0.47 |
| #> s(Yearf) | 5.00e+00 | 3.82e+00 | NA | NA |

```

par(oldpar)

```

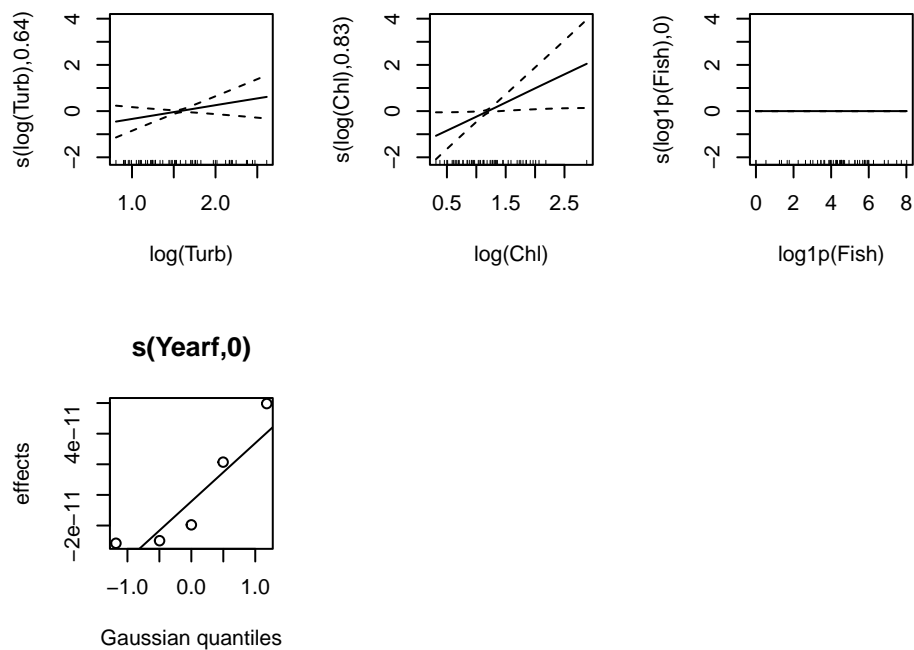
Temora

```
spp = "Temora"
mod <- spp_analysis$gam_mods[spp_analysis$Species == spp][[1]]
summary(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   0.34825    0.84090   0.414 0.680632
#> SeasonSummer  0.01713    0.75836   0.023 0.982070
#> SeasonFall    0.05654    0.68971   0.082 0.935013
#> Station2      1.00995    0.72696   1.389 0.171225
#> Station3      2.77847    0.72961   3.808 0.000401 ***
#> Station4      1.96495    0.79413   2.474 0.016972 *
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Approximate significance of smooth terms:
#>              edf Ref.df    F p-value
#> s(log(Turb))   6.375e-01    4 0.442  0.1005
#> s(log(Chl))    8.324e-01    4 1.187  0.0189 *
#> s(log1p(Fish)) 2.319e-11    4 0.000  0.5568
#> s(Yearf)       3.076e-10    5 0.000  0.4572
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> R-sq.(adj) = 0.295   Deviance explained = 37.9%
#> GCV = 3.4527   Scale est. = 2.9838    n = 55
cat('\n\nANOVA')
#>
#>
#> ANOVA
anova(mod)
#>
#> Family: gaussian
#> Link function: identity
#>
#> Formula:
#> log1p(Density) ~ Season + Station + s(log(Turb), bs = "ts", k = 5) +
#>      s(log(Chl), bs = "ts", k = 5) + s(log1p(Fish), bs = "ts",
#>      k = 5) + s(Yearf, bs = "re")
#>
#> Parametric Terms:
#>              df      F p-value
#> Season       2 0.005 0.99550
```

```
#> Station 3 5.327 0.00302
#>
#> Approximate significance of smooth terms:
#>          edf    Ref.df      F p-value
#> s(log(Turb)) 6.375e-01 4.000e+00 0.442 0.1005
#> s(log(Chl))  8.324e-01 4.000e+00 1.187 0.0189
#> s(log1p(Fish)) 2.319e-11 4.000e+00 0.000 0.5568
#> s(Yearf)      3.076e-10 5.000e+00 0.000 0.4572
```

Plot GAM

```
oldpar <- par(mfrow = c(2,3))
plot(mod)
par(oldpar)
```

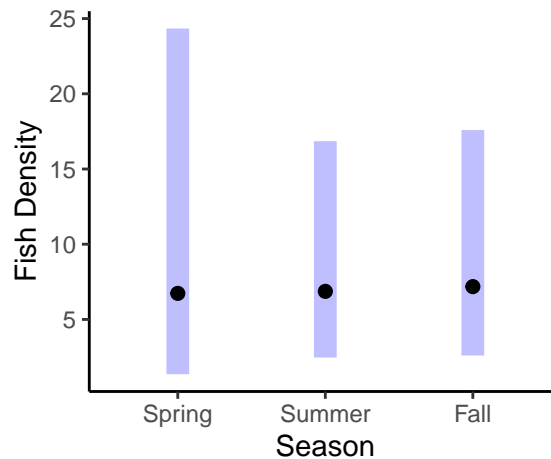


Plot Season and Station

```
emms <- emmeans(mod, ~Season, type = 'response',
                 data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Spring - Summer -0.0171 0.758 47.5 -0.023 0.9997
#> Spring - Fall   -0.0565 0.690 47.5 -0.082 0.9963
```

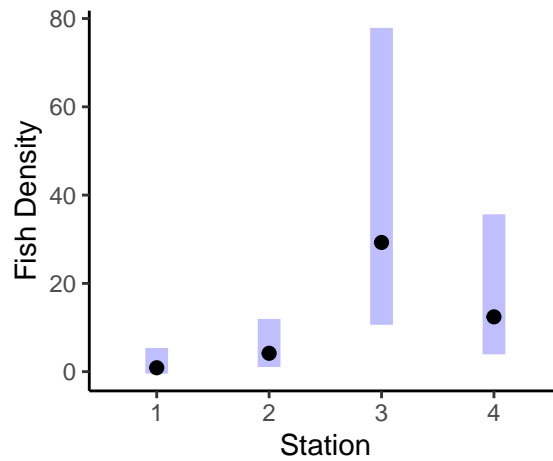
```
#> Summer - Fall    -0.0394 0.569 47.5  -0.069  0.9974
#>
#> Results are averaged over the levels of: Station, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 3 estimates
```

```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



```
emms <- emmeans(mod, ~Station, type = 'response',
  data = spp_analysis$data[spp_analysis$Species == spp][[1]])
pairs(emms)
#> Note: Use 'contrast(regrid(object), ...)' to obtain contrasts of back-transformed estimates
#> contrast      estimate    SE   df t.ratio p.value
#> Station1 - Station2  -1.010 0.727 47.5  -1.389  0.5122
#> Station1 - Station3  -2.778 0.730 47.5  -3.808  0.0022
#> Station1 - Station4  -1.965 0.794 47.5  -2.474  0.0771
#> Station2 - Station3  -1.769 0.654 47.5  -2.703  0.0453
#> Station2 - Station4  -0.955 0.688 47.5  -1.388  0.5132
#> Station3 - Station4   0.814 0.659 47.5   1.235  0.6081
#>
#> Results are averaged over the levels of: Season, Yearf
#> Note: contrasts are still on the log1p scale
#> P value adjustment: tukey method for comparing a family of 4 estimates
```

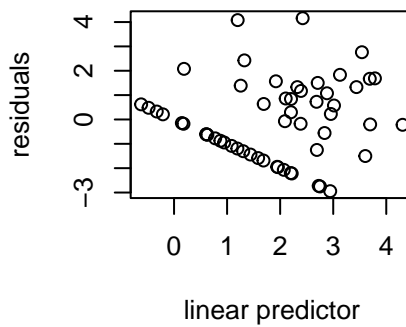
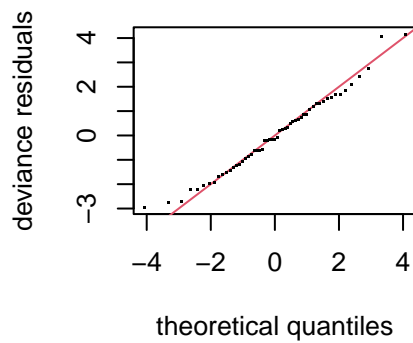
```
plot(emms) +
  coord_flip() +
  xlab('Fish Density')
```



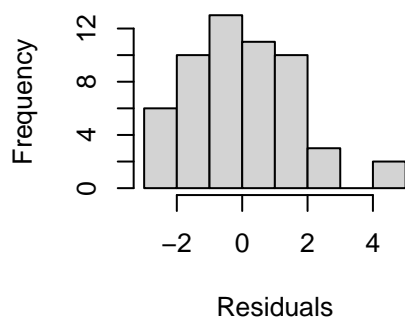
Model Diagnostics

```
oldpar <- par(mfrow = c(2,2))
gam.check(mod)
```

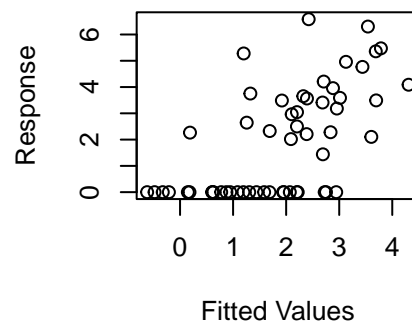
Resids vs. linear pred.



Histogram of residuals



Response vs. Fitted Values



```

#>
#> Method: GCV   Optimizer: magic
#> Smoothing parameter selection converged after 15 iterations.
#> The RMS GCV score gradient at convergence was 1.816221e-07 .
#> The Hessian was positive definite.
#> Model rank = 23 / 23
#>
#> 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(log(Turb)) | 4.00e+00 | 6.37e-01 | 1.10 | 0.72 |
| #> s(log(Chl)) | 4.00e+00 | 8.32e-01 | 1.05 | 0.62 |
| #> s(log1p(Fish)) | 4.00e+00 | 2.32e-11 | 1.08 | 0.68 |
| #> s(Yearf) | 5.00e+00 | 3.08e-10 | NA | NA |

```

par(oldpar)

```